US 20030129176A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2003/0129176 A1

Jones et al. (43) Pub. Date: Jul. 10, 2003

(54) ATHEROSCLEROSIS-ASSOCIATED GENES

(75) Inventors: **Karen Anne Jones**, Saffron Walden (GB); **Wayne Volkmuth**, Calabasas, CA (US); **Michael G. Walker**, Sunnyvale, CA (US); **Lynn E. Murry**, Fayetteville, AR (US)

Correspondence Address:
INCYTE GENOMICS, INC.
3160 PORTER DRIVE
PALO ALTO, CA 94304 (US)

(73) Assignee: **Incyte Genomics, Inc.**, Palo Alto, CA (US)

(21) Appl. No.: **10/219,664**

(22) Filed: **Aug. 14, 2002**

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/349,015, filed on Jul. 7, 1999, now abandoned.

Publication Classification

(51) **Int. Cl.**[7] **A61K 38/43**; C12Q 1/68; C07H 21/04; C12N 9/00; C12P 21/02; C12N 5/06

(52) **U.S. Cl.** **424/94.1**; 435/6; 435/69.1; 435/183; 435/320.1; 435/325; 536/23.2

(57) **ABSTRACT**

The present invention relates to a combination comprising a plurality of cDNAs which are differentially expressed in cardiovascular diseases. The combination and compositions can be used entirely or in part to diagnose, to stage, to treat, or to monitor the progression or treatment of disorders associated with atherosclerosis.

```
                9                  18                  27                  36                  45                  54
5'  GCT GCC TGC CGG TGC TCT TCG TGG CTC TGG GCA TGG CCT CGG ACC CCA TCT TCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                63                  72                  81                  90                  99                 108
    CGC TGG CGC CCC CGC TGC ATT GCC ACT ACG GGG CCT TCC CCC CTA ATG CCT CTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

               117                 126                 135                 144                 153                 162
    GCT GGG AGC AGC CTC CCA ATG CCA GCG GCG TCA GCG TCG CCA GCG CTG CCC TAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

               171                 180                 189                 198                 207                 216
    CAG CCA GCG CCG CCA GCC GTG TCG CCA CCA GTA CCG ACC CCT CGT GCA GCG GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

               225                 234                 243                 252                 261                 270
    TCG CCC CGC CGG ACT TCA ACC ATT GCC CTC AAG GAT TGG GAC TAT AAT GGC CTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

               279                 288                 297                 306                 315                 324
    CCT GTG CTC ACC ACC AAC GCC ATC GGC CAG TGG GAT CTG GTG TGT GAC CTG GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

               333                 342                 351                 360                 369                 378
    TGG CAG GTG ATC CTG GAG CAG ATC CTC TTC ATC TTG GGC TTT GCC TCC GGC TAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
```

FIGURE 1A

```
        387         396         405         414         423         432
CTG TTC CTG GGT TAC CCC GCA GAC AGA TTT GGC CGT CGC GGG ATT GTG CTG CTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

        441         450         459         468         477         486
ACC TTG GGG CTG GTG GGC CCC TGT GGA GTA GGA GGG GCT GCT GCA GGC TCC TCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

        495         504         513         522         531         540
ACA GGC GTC ATG GCC CTC CGA TTC CTC TTG GGC TTT CTG CTT GCC GGT GTT GAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
             M   A   L   R   F   L   L   G   F   L   L   A   G   V   D

        549         558         567         576         585         594
CTG GGT GTC TAC CTG ATG CGC CTG GAG CTG TGC GAC CCA ACC CAG AGG CTT CGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   G   V   Y   L   M   R   L   E   L   C   D   P   T   Q   R   L   R

        603         612         621         630         639         648
GTG GCC CTG GCA GGG GAG TTG GTG GGG GTG GGA GGG CAC TTC CTG TTC CTG GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   A   L   A   G   E   L   V   G   V   G   G   H   F   L   F   L   G

        657         666         675         684         693         702
CTG GCC CTT GTC TCT AAG GAT TGG CGA TTC CTA CAG CGA ATG ATC ACC GCT CCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   A   L   V   S   K   D   W   R   F   L   Q   R   M   I   T   A   P
```

FIGURE 1B

```
                711                 720                 729                 738                 747                 756
TGC  ATC  CTC  TTC  CTG  TTT  TAT  GGC  TGG  CCT  GGT  TTG  TTC  CTG  GAG  TCC  GCA  CGG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 C    I    L    F    L    F    Y    G    W    P    G    L    F    L    E    S    A    R

                765                 774                 783                 792                 801                 810
TGG  CTG  ATA  GTG  AAG  CGG  CAG  ATT  GAG  GAG  GCT  CAG  TCT  GTG  CTG  AGG  ATC  CTG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 W    L    I    V    K    R    Q    I    E    E    A    Q    S    V    L    R    I    L

                819                 828                 837                 846                 855                 864
GCT  GAG  CGA  AAC  CGG  CCC  CAT  GGG  CAG  ATG  CTG  GGG  GAG  GAG  GCC  CAG  GAG  GCC
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 A    E    R    N    R    P    H    G    Q    M    L    G    E    E    A    Q    E    A

                873                 882                 891                 900                 909                 918
CTG  CAG  GAC  CTG  GAG  AAT  ACC  TGC  CCT  CTC  CCT  GCA  ACA  TCC  TCC  TTT  TCC  TTT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 L    Q    D    L    E    N    T    C    P    L    P    A    T    S    S    F    S    F

                927                 936                 945                 954                 963                 972
GCT  TCC  CTC  CTC  AAC  TAC  CGC  AAC  ATC  TGG  AAA  AAT  CTG  CTT  ATC  CTG  GGC  TTC
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 A    S    L    L    N    Y    R    N    I    W    K    N    L    L    I    L    G    F
```

FIGURE 1C

```
         981         990         999        1008        1017        1026
ACC AAC TTC ATT GCC CAT GCC ATT CGC CAC TGC TAC CAG CCT GTG GGA GGA GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   N   F   I   A   H   A   I   R   H   C   Y   Q   P   V   G   G   G

        1035        1044        1053        1062        1071        1080
GGG AGC CCA TCG GAC TTC TAC CTG TGC TCT CTG CTG GCC AGC GGC ACC GCA GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   S   P   S   D   F   Y   L   C   S   L   L   A   S   G   T   A   A

        1089        1098        1107        1116        1125        1134
CTG GCC TGT GTC TTC CTG GGG GTC ACC GTG GAC CGA TTT GGC CGC CGG GGC ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   A   C   V   F   L   G   V   T   V   D   R   F   G   R   R   G   I

        1143        1152        1161        1170        1179        1188
CTT CTT CTC TCC ATG ACC CTT ACC GGC ATT GCT TCC CTG GTC CTG CTG GGC CTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   L   L   S   M   T   L   T   G   I   A   S   L   V   L   L   G   L

        1197        1206        1215        1224        1233        1242
TGG GAT TAT CTG AAC GAG GCT GCC ATC ACC ACT TTC TCT GTC CTT GGG CTC TTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 W   D   Y   L   N   E   A   A   I   T   T   F   S   V   L   G   L   F
```

FIGURE 1D

```
        1251                1260                1269                1278                1287                1296
TCC TCC CAA GCT GCC GCC ATC CTC AGC ACC CTC CTT GCT GCT GAG GTC ATC CCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   S   Q   A   A   A   I   L   S   T   L   L   A   A   E   V   I   P

        1305                1314                1323                1332                1341                1350
ACC ACT GTC CGG GGC CGT GGC CTG GGC CTG ATC ATG GCT CTA GGG GCG CTT GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   T   V   R   G   R   G   L   G   L   I   M   A   L   G   A   L   G

        1359                1368                1377                1386                1395                1404
GGA CTG AGC GGC CCG GCC CAG CGC CTC CAC ATG GGC CAT GGA GCC TTC CTG CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   L   S   G   P   A   Q   R   L   H   M   G   H   G   A   F   L   Q

        1413                1422                1431                1440                1449                1458
CAC GTG GTG CTG GCG GCC TGC GCC CTC CTC TGC ATT CTC AGC ATT ATG CTG CTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 H   V   V   L   A   A   C   A   L   L   C   I   L   S   I   M   L   L

        1467                1476                1485                1494                1503                1512
CCG GAG ACC AAG CGC AAG CTC CTG CCC GAG GTG CTC CGG GAC GGG GAG CTG TGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 P   E   T   K   R   K   L   L   P   E   V   L   R   D   G   E   L   C
```

FIGURE 1E

```
                  1521                1530                1539                1548                1557                1566
CGC CGG CCT TCC CTG CTG CGG CAG CCA CCC CCT ACC CGC TGT GAC CAC GTC CCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   R   P   S   L   L   R   Q   P   P   P   T   R   C   D   H   V   P

                  1575                1584                1593                1602                1611                1620
CTG CTT GCC ACC CCC AAC CCT GCC CTC TGA GCG GCC TCT GAG TAC CCT GGC GGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   L   A   T   P   N   P   A   L

                  1629                1638                1647                1656                1665                1674
AGG CTG GCC CAC ACA GAA AGG TGG CAA GAA GAT CGG GAA GAC TGA GTA GGG AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                  1683                1692                1701                1710                1719                1728
GCA GGG CTG CCC AGA AGT CTC AGA GGC ACC TCA CGC CAG CCA TCG CGG AGA GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                  1737                1746                1755                1764                1773                1782
CAG AGG GCC GTC CCC ACC CTG CCT CCT CCC TGC TGC TTT GCA TTC ACT TCC TTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                  1791                1800                1809                1818                1827                1836
GCC AGA GTC AGG GGA CAG GGA GAG AGC TCC ACA CTG TAA CCA CTG GGT CTG GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
```

FIGURE 1F

```
        1845        1854        1863        1872        1881        1890
TCC ATC CTG CGC CCA AAG ACA TCC ACC CAG ACC TCA TTA TTT CTT GCT CTA TCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

        1899        1908        1917        1926
TTC TGT TTC AAT AAA GAC ATT TGG AAT AAA AAA AAA AAA 3'
--- --- --- --- --- --- --- --- --- --- --- --- ---
```

FIGURE 1G

ATHEROSCLEROSIS-ASSOCIATED GENES

FIELD OF THE INVENTION

[0001] The invention relates to a combination of isolated cDNAs that are significantly co-expressed with one or more known atherosclerosis-associated genes. The invention also relates to the use of the combination in the diagnosis, prognosis, treatment, and evaluation of therapies for disorders associated with atherosclerosis.

BACKGROUND OF THE INVENTION

[0002] Atherosclerosis is a disorder characterized by cellular changes in the arterial intima and the formation of arterial plaques containing intracellular and extracellular deposits of lipids. The thickening of artery walls and the narrowing of the arterial lumen underlies the pathologic condition in most cases of coronary artery disease, aortic aneurysm, peripheral vascular disease, and stroke. A number of metabolic pathways and a cascade of molecular events is involved in the cellular morphogenesis, proliferation, and cellular migration that results in atherogenesis (Libby et al. (1997) Int J Cardiol 62 (S2):23-29).

[0003] The artery walls consist of three layers: the intima (innermost), the media, and the adventitia (outermost). The intima consists of a layer of endothelial cells lining the lumen of arteries and arterioles. Endothelial cells form a barrier against the indiscriminate entry of substances from the blood into the artery. Specific transporter proteins expressed by endothelial cells facilitate barrier function. Endothelial cells also secrete a number of substances which help regulate downstream vascular contractility blood coagulation, and other aspects of vascular biology. The medial layer of the arterial wall contains smooth muscle cells in a matrix of collagen and elastic fibers produced by the smooth muscle cells. Contraction and relaxation of the smooth muscle layer allows arteries and arterioles to modulate blood pressure and blood flow. The outermost layer of the arterial wall, the adventitia, is a mixture of collagen bundles, elastic fibers, some smooth muscle cells, fibroblasts and nerve cells. The adventitia provides structural integrity to the blood vessel and acts as a support matrix for the media and intima.

[0004] Initiation of an atherosclerotic lesion often occurs following vascular endothelial cell injury often attributable to hypertension, diabetes mellitus, hyperlipidemia, fluctuating shear stress, smoking, or transplant rejection. Nitric oxide and superoxide anions are released and react to form cytodestructive peroxynitrite radicals that cause injury to the endothelium and myocytes of the intima and lead to expression of a variety of molecules that produce local and systemic effects. These effects include the release of mediators of inflammation such as cytokines, complement components, prostaglandins, and downstream transcription factors. Such mediators promote monocyte infiltration of the vascular intima and lead to the upregulation of adhesion molecules which encourages attachment of the monocytes to the damaged endothelial cells. Simultaneously, components of the extracellular matrix including collagens, fibrinogens, and matrix Gla protein are induced and provide sites for monocyte attachment, and annexins, plasminogen activator inhibitor 1, and nitric oxide synthases are induced to counteract these effects.

[0005] Monocytes that infiltrate the lesion accumulate modified low density lipoprotein through scavenger receptors such as CD36 and macrophage scavenger receptor type I. The abundance of modified lipids is a factor in atherogenesis and is influenced by modifying enzymes such as lipoprotein lipase, carboxyl ester lipase, serum amyloid P component, LDL-receptor related protein, microsomal triglyceride transfer protein, and serum esterases such as paraoxonase. Lipid metabolism is governed by cholesterol biosynthesis enzymes such as 3-hydroxy-3-methylglutaryl coenzyme A synthase, and products of the apolipoprotein genes. Modified lipid stabilization and accumulation is aided by perilipin and alpha-2-macroglobulin.

[0006] As monocytes accumulate in the lesion, they can rupture and release free cholesterol, cytokines, and procoagulants into the surrounding environment. This is the process that leads to plaque development; plaque consists of a mass of lipid-engorged monocytes and a lipid-rich necrotic core covered by a fibrous cap. The gradual progression of plaque growth is punctuated by thrombus formation and leads to clinical symptoms such as unstable angina, myocardial infarction, or stroke. Thrombus formation is initiated by episodic plaque rupture which exposes flowing blood to tissue factors, which induce coagulation, and collagen, which activates platelets. After initiation of the atherosclerotic lesion, enzymes that degrade extracellular matrix components (ECM) such as matrix metalloproteinases and cathepsin K are up-regulated, and inhibitors of ECM are down-regulated. This results in destabilization of the atherosclerotic lesion and subsequent complications including myocardial infarction, angina, and stroke. Further arterial occlusion and infiltration increase with the expression of coagulation factors and down-regulation of their inhibitors, antithrombin III, and lipoprotein-associated coagulation inhibitor.

[0007] Smooth muscle cells build up in the arterial media and constitute one of the principal cell types in atherosclerotic and restenotic lesions. They show a high degree of plasticity and are able to shift between a differentiated, contractile phenotype and a less differentiated, synthetic phenotype. This modulation occurs as a response to factors secreted from cells at the site of vascular injury and results in structural reorganization with a loss of myofilaments and the formation of an extensive endoplasmic reticulum and a large Golgi complex. Genes encoding secreted protein, acidic and rich in cysteine (SPARC) and endothelin-1 contribute to these changes. At the same time, the expression of cytoskeletal proteins such as calponin, myosin, desmin, and other gene products in the cells is altered. As a result, the smooth muscle cells lose their contractility and become able to migrate from the media to the intima, to proliferate, and to secrete extracellular matrix components which contribute to arterial intimal thickening.

[0008] The initiation and progression of atherosclerotic lesion development requires the interplay of various molecular pathways Many genes that participate in these processes are known, and some of them have been shown to have a direct role in atherosclerosis pathogenesis by animal model experiments, in vitro assays, and epidemiological studies (Krettek et al. (1997) Arterioscler Thromb Vasc Biol 17:2897-2903; Fisher et al. (1997) Atherosclerosis 135:145-159; Shih et al. (1998) Circulation 95:2684-2693; and Bocan et al. (1998) Atherosclerosis 139:21-30).

[0009] The present invention satisfies a need in the art by providing a combination comprising a plurality of cDNAs that are useful for diagnosis, prognosis, treatment, and evaluation of therapies for disorders associated with atherosclerosis.

SUMMARY OF THE INVENTION

[0010] The invention provides a combination of isolated cDNAs that are significantly co-expressed with one or more known atherosclerosis-associated genes. The combination comprises the isolated cDNAs having the nucleic acid sequences of SEQ ID NOs:1-25 and the complements of SEQ ID NOs:1-25. In one embodiment, the combination is placed on a substrate. In another embodiment, the substrate is a microarray.

[0011] The invention also provides a method for using the combination to detect gene expression in a sample containing nucleic acids, the method comprising hybridizing the substrate containing the combination to the nucleic acids of the sample under conditions for formation of one or more hybridization complexes and detecting hybridization complex formation, wherein complex formation indicates gene expression in the sample. In one embodiment, the sample is from artery or is obtained during microsurgery to open a blocked artery. In another embodiment, complex formation is compared to standards and is diagnostic of a disorder associated with atherosclerosis.

[0012] The invention further provides a method of using a combination to screen a plurality of molecules to identify at least one ligand which specifically binds a cDNA of the combination, the method comprising combining the substrate containing the combination with molecules under conditions to allow specific binding; and detecting specific binding, thereby identifying a ligand which specifically binds at least one cDNA of the combination. In one embodiment, the molecules are selected from DNA molecules, peptides, proteins, RNA molecules, and transcription factors.

[0013] The invention provides an isolated cDNA comprising a polynucleotide having the nucleic acid sequence of SEQ ID NO:8 and the complement of SEQ ID NO:8. In different embodiments, the cDNA is used as a probe, in an expression vector, and in assays for diagnosis, prognosis, and treatment of disorders associated with atherosclerosis. The invention also provides a composition comprising the cDNA and a labeling moiety. The invention further provides a method for using the cDNA to screen a plurality of molecules to identify a ligand which specifically binds the cDNA, the method comprising combining the cDNA with a sample under conditions to allow specific binding; recovering the bound cDNA; and separating the ligand from the bound cDNA, thereby obtaining purified ligand. In one embodiment, the molecules to be screened are selected from DNA molecules, peptides, proteins, RNA molecules, and transcription factors. The invention yet further provides a method for using a cDNA to detect gene expression in a sample containing nucleic acids, the method comprising hybridizing the cDNA to nucleic acids of a sample under conditions for formation of one or more hybridization complexes; and detecting hybridization complex formation, wherein complex formation indicates gene expression in the sample. In one embodiment, the cDNA is attached to a substrate. In another embodiment, gene expression when compared to standards is diagnostic of a disorder associated with atherosclerosis.

[0014] The invention provides a vector containing the cDNA, and a host cell containing the vector. The invention also provides a method for producing a peptide or protein, the method comprising culturing the host cell under conditions for expression of the peptide or protein; and recovering the peptide or protein so produced from cell culture.

[0015] The invention provides a purified peptide or protein comprising an amino acid sequence expressed by a cDNA of the invention. In one embodiment, the protein comprises the amino acid sequence of SEQ ID NO:26. The invention additionally provides a composition comprising the protein and a pharmaceutical carrier. The invention also provides a method for using a peptide or protein to screen a plurality of molecules to identify at least one ligand which specifically binds the protein. In one embodiment, the molecules to be screened are selected from agonists, antagonists, antibodies, DNA molecules, transcription factors, RNA molecules, and small drug molecules or compounds. The invention further provides a method of using a peptide or protein to purify a ligand.

[0016] The invention provides a method for using the peptide or protein to produce an antibody which specifically binds the protein. A method for preparing polyclonal antibodies comprises immunizing a animal with peptide or protein under conditions to elicit an antibody response, isolating animal antibodies, attaching the peptide or protein to a substrate, contacting the substrate with isolated antibodies under conditions to allow specific binding to the peptide or protein, dissociating the antibodies from the peptide or protein, thereby obtaining purified polyclonal antibodies. A method for preparing monoclonal antibodies comprises immunizing a animal with a peptide or protein under conditions to elicit an antibody response, isolating antibody producing cells from the animal, fusing the antibody producing cells with immortalized cells in culture to form monoclonal antibody producing hybridoma cells, culturing the hybridoma cells, and isolating monoclonal antibodies from culture.

[0017] The invention provides purified antibodies which bind specifically to a peptide or protein. The invention also provides a method for using an antibody to detect expression of a peptide or protein in a sample, the method comprising combining the antibody with a sample under conditions for formation of antibody:peptide or protein complexes, and detecting complex formation, wherein complex formation indicates expression of the peptide or protein in the sample. In one aspect, the amount of complex formation when compared to standards is diagnostic of a disorder of the nervous system.

[0018] The invention provides a method for immunopurification of a protein comprising attaching an antibody to a substrate, exposing the antibody to a sample containing protein under conditions to allow antibody:protein complexes to form, dissociating the protein from the complex, and collecting purified protein. The invention also provides an array upon which a cDNA encoding a protein, the protein, or an antibody which specifically binds the protein are immobilized. The invention also provides a composition comprising a cDNA, a protein, an antibody, or a ligand which has agonistic or antagonistic activity.

[0019] The invention provides an antibody comprising an antigen binding site, wherein the antigen binding site specifically binds to the protein The invention also provides a method for treating a disorder associated with the differential expression of a cDNA that is coexpressed with one or more known atherosclerosis-associated genes in a subject in need, the method comprising the step of administering to the subject in need the antibody in an amount effective for treating the disorder. The invention further provides an immunoconjugate comprising the antigen binding site of the antibody or joined to a therapeutic agent. The invention additionally provides a method for treating a disorder associated with the differential expression of a cDNA that is coexpressed with one or more known atherosclerosis-associated genes in a subject in need, the method comprising the step of administering to the subject in need the immunoconjugate in an amount effective for treating the disorder.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING, FIGURE AND TABLES

[0020] The Sequence Listing provides exemplary cDNAs associated with atherosclerosis including polynucleotide sequences SEQ ID NOs:1-25 and the polypeptide sequence, SEQ ID NO:26. Each sequence is identified by a sequence identification number (SEQ ID NO).

[0021] **FIG. 1** shows the cDNA having the nucleic acid sequence of SEQ ID NO:8 which encodes the protein having the amino acid sequence of SEQ ID NO:26. The alignment was produced using MAcDNASIS,PRO software (Hitachi Software Engineering, South San Francisco Calif.).

[0022] Table 4 shows the-co-expression values (-log p) between the known atherosclerosis-associated genes (abbreviations as shown in Table 3) and the cDNAs of the invention (SEQ IDs).

[0023] Table 5 summarizes the highly significantly coexpression between each cDNA (SEQ ID) and two known atherosclerosis-associated genes (Gene 1 and Gene 2). P-value and the function or importance of the known atherosclerosis-associated gene are derived from Tables 3 and 4.

[0024] Table 6 shows transcript images for several of the cDNAs (SEQ ID) of the invention. Column 1 shows the SEQ ID; column 2, the library name; column 3, the number of cDNAs in the library; column 4, the description of the sample from which the library was constructed; column 5, transcript abundance; and column 6, percent transcript abundance. These sets demonstrate differential expression within experiments using cardiovascular tissues.

[0025] Table 7 shows microarray data for several of the cDNAs (SEQ ID) of the invention. Column 1 shows the SEQ ID; column 2, the name of the microarray (GEM) used for the experiment; column 3, the log2 (C5/Cy3 ratio); column 4, the description of the Cy3 sample; and column 5, the description of the Cy5 sample. These data demonstrate differential expression of the cDNAs in experiments using cardiovascular samples.

DESCRIPTION OF THE INVENTION

[0026] It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include the plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

[0027] Definitions

[0028] "Antibody" refers to intact immunoglobulin molecule, a polyclonal antibody, a monoclonal antibody, a chimeric antibody, a recombinant antibody, a humanized antibody, single chain antibodies, a Fab fragment, an $F(ab')_2$ fragment, an Fv fragment; and an antibody-peptide fusion protein.

[0029] "Antigenic determinant" refers to an antigenic or immunogenic epitope, structural feature, or region of an oligopeptide, peptide, or protein which is capable of inducing formation of an antibody which specifically binds the protein. Biological activity is not a prerequisite for immunogenicity.

[0030] "Array" refers to an ordered arrangement of at least two cDNAs, proteins, or antibodies on a substrate. At least one of the cDNAs, proteins, or antibodies represents a control or standard, and the other represents a cDNA, protein, or antibody of diagnostic or therapeutic interest. The arrangement of at least two and up to about 40,000 cDNAs, proteins, or antibodies on the substrate assures that the size and signal intensity of each labeled complex, formed between each cDNA and at least one nucleic acid, each protein and at least one ligand or antibody, or each antibody and at least one protein to which the antibody specifically binds, is individually distinguishable.

[0031] "Atherosclerosis-associated cDNA" refers to isolated polynucleotide that exhibits a statistically significant co-expression pattern with known atherosclerosis-associated genes. The cDNAs are represented by SEQ ID NOs:1-25 of the Sequence Listing and the complements of SEQ ID NOs:1-25. They may be of recombinant or synthetic origin, used in its double-stranded or single-stranded form, and combined with vitamins, minerals, carbohydrates, lipids, proteins, other nucleic acids, a pharmaceutical carrier or a labeling moiety to perform a particular activity or form a useful composition.

[0032] An "atherosclerotic sample or tissue" may be taken using needles, catheters, or scapels and include vessels including the aorta, arteries, arterioles, endothelial cells, plaque, and blood. The sample may contain nucleic acids, proteins, antibodies, and the like. Additionally the sample may comprise the soluble fraction of a cell preparation; an aliquot of media in which cells were grown; a chromosome, an organelle, or membrane isolated or extracted from a cell; genomic DNA, RNA, or cDNA in solution or bound to a substrate.

[0033] A "combination" comprises at least two sequences selected from SEQ ID NOs:1-25 as presented in the Sequence Listing and the complements of SEQ ID NOs:1-25.

[0034] "Differential expression" refers to an increased or up-regulated or a decreased or down-regulated expression as detected by absence, presence, or at least two-fold change in the amount of transcribed messenger RNA or translated protein in a sample.

[0035] "Disorders associated with atherosclerosis" include angina pectoris, coronary artery disease, myocardial infarction, hypertension, transient cerebral ischemia, mesenteric ischemia, peripheral vascular disease, renal artery stenosis, and stroke.

[0036] An "expression profile" is a representation of gene expression in a sample. A nucleic acid expression profile is produced using sequencing, hybridization, or amplification technologies and mRNAs or cDNAs from a sample. A protein expression profile, although time delayed, mirrors the nucleic acid expression profile and uses two-dimensional polyacrylamide electrophoresis (2D-PAGE) and mass spectrophotometry (MS) or western analysis, enzyme-linked immunosorbent assays (ELISAs), fluorescence activated cell sorting (FACS), radioimmunoassays (RIAs), or arrays and labeling moieties or antibodies to detect expression in a sample. The nucleic acids, proteins, or antibodies may be used in solution or attached to a substrate, and their detection is based on methods and labeling moieties well known in the art.

[0037] A "hybridization complex" is formed between a cDNA of the invention and a nucleic acid of a sample when the purines of one molecule hydrogen bond with the pyrimidines of the complementary molecule, e.g., 5'-A-G-T-C-3' base pairs with its complete complement, 3'-T-C-A-G-5'. The degree of complementarity and the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

[0038] "Identity" as applied to sequences, refers to the quantification (usually percentage) of nucleotide or residue matches between at least two sequences aligned using a standardized algorithm such as Smith-Waterman alignment (Smith and Waterman (1981) J Mol Biol 147:195-197), CLUSTALW (Thompson et al. (1994) Nucleic Acids Res 22:4673-4680), or BLAST2 (Altschul et al. (1997) Nucleic Acids Res 25:3389-340). BLAST2 may be used in a standardized and reproducible way to insert gaps in one of the sequences in order to optimize alignment and to achieve a more meaningful comparison between them.

[0039] "Similarity" as applied to proteins uses the same algorithms but takes into account conservative substitutions of nucleotides or residues.

[0040] "Isolated or purified" refers to a cDNA or protein that is removed from its natural environment and that is separated from other components with which it is naturally present.

[0041] "Genes known to be associated with atherosclerosis" include human 22 kDa smooth muscle protein, calponin (CNN1), pro alpha 1 (I) collagen (COLIA1), collagen alpha-2 type I (COLIA2), collagen alpha-6 type I (COL6A1), procollagen alpha 2(V) (COL5A2), collagen VI alpha-2 (COL6A2), type VI collagen alpha 3 (COL6A3), pro-alpha-i type 3 collagen (COL3A1), pro-alpha-1(V) collagen (COLSA1), matrix Gla protein (MGP), cathepsin K (CSTK), fibrinogen beta chain gene (FBG), pre-pro-von Willebrand factor (VWF), platelet endothelial cell adhesion molecule (PECAM-1), antithrombin III variant (AT3), lipoprotein lipase (LPL), alpha-2-macroglobulin (A2M), apolipoprotein AI (APOA1), apolipoprotein AII $(APOA1)_2$, apolipoprotein B-100 (APOB), lipoprotein apoCII (APOC2), pre-apolipoprotein CIII (APOC3), apolipoprotein apo C-IV (APOC4), macrophage scavenger receptor type I (MSR1), human antigen CD36 gene (CD36), serum amyloid P component (SAP), carboxyl ester lipase gene (CEL), paraoxonase 1 (PONI), paraoxonase 2 (PON2), paraoxonase 3 (PON3), perilipin (PLIN), prostaglandin D2 synthase (PTGDS), annexin II/lipocortin II(ANX2), annexin I/lipocortin (ANX1), and secreted protein, acidic and rich in cysteine (SPARC).

[0042] "Labeling moiety" refers to any reporter molecule whether a visible or radioactive label, stain or dye that can be attached to or incorporated into a cDNA or protein. Visible labels and dyes include but are not limited to anthocyanins, B glucuronidase, BIODIPY, Coomassie blue, Cy3 and Cy5, digoxigenin, FITC, green fluorescent protein, luciferase, spyro red, silver, and the like. Radioactive markers include radioactive forms of hydrogen, iodine, phosphorous, sulfur, and the like.

[0043] "Ligand" refers to any agent, molecule, or compound which will bind specifically to a complementary site on a cDNA molecule, a polynucleotide, an epitope or a protein. Such ligands stabilize or modulate the activity of polynucleotides or proteins and may be composed of inorganic or organic substances including nucleic acids, proteins, carbohydrates, fats, and lipids.

[0044] "Markers for disorders associated with atherosclerosis" refers to cDNAs, peptides or proteins, and antibodies which are useful in the diagnosis, prognosis, treatment, selection or evaluation of therapies for disorders associated with atherosclerosis. These markers are differentially expressed in samples from subjects predisposed to or manifesting one of these disorders. The known atherosclerosis-associated genes and their contribution and/or function to disorders associated with atherosclerosis are listed in TABLE3.

[0045] "Probe" refers to a cDNA of the invention that hybridizes to at least one nucleic acid in a sample. Where targets are single stranded, probes are complementary single strands. Probes can be labeled for use in hybridization reactions including Southern, northern, in situ, dot blot, array, and like technologies or in screening assays.

[0046] "Protein" refers to a polypeptide or any portion thereof. An "oligopeptide" is an amino acid sequence from about five residues to about 15 residues that is used as part of a fusion protein to produce an antibody that specifically binds the protein.

[0047] "Specific binding" refers to a special and precise interaction between two molecules which is dependent upon their structure, particularly their molecular side groups. For example, the intercalation of a regulatory protein into the major groove of a DNA molecule, the hydrogen bonding along the backbone between two single stranded nucleic acids, or the binding between an epitope of a protein and an agonist, antagonist, or antibody.

[0048] "Substrate" refers to any rigid or semi-rigid support to which cDNAs or proteins are bound and includes membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, capillaries or other tubing, plates, polymers, and microparticles with a variety of surface forms including wells, trenches, pins, channels and pores.

[0049] A "transcript image" (TI) is a profile of gene transcription activity in a particular tissue at a particular

time. TI provides assessment of the relative abundance of expressed transcripts in the cDNA libraries of an EST database as described in U.S. Pat. No. 5,840,484, incorporated herein by reference.

[0050] "Variant" refers to molecules that are recognized variations of a polynucleotide or a protein. Splice variants may be determined by BLAST score, wherein the score is at least 100, and most preferably at least 400. Allelic variants have a high percent identity to a polynucleotide and may differ by about three bases per hundred bases. "Single nucleotide polymorphism" (SNP) refers to a change in a single base as a result of a substitution, insertion or deletion. The change may be conservative (purine for purine) or non-conservative (purine to pyrimidine) and may or may not result in a change in an encoded amino acid.

[0051] The Method

[0052] The present invention encompasses a method for identifying cDNAs that are significantly co-expressed with known atherosclerosis-associated genes. In particular, the method identifies a combination of cDNAs useful in diagnosis, prognosis, treatment, and evaluation of therapies for disorders associated with atherosclerosis.

[0053] The method involves identifying cDNAs that are expressed in a plurality of cDNA libraries. These cDNAs include genes of known or unknown function whose expression patterns are compared with genes having a known function and disease association to determine whether a specified coexpression probability threshold is met. Through this comparison, a subset of the cDNAs having a high coexpression probability with the known genes can be identified.

[0054] The cDNAs may originate from cDNA libraries derived from a variety of sources including, but not limited to, eukaryotes such as human, mouse, rat, dog, monkey, plant, and yeast; prokaryotes such as bacteria; and viruses. The cDNAs can also be selected from a variety of sequence types including, but not limited to, expressed sequence tags (ESTs), assembled polynucleotide sequences, full length gene coding regions, promoters, introns, enhancers, 5' untranslated regions, and 3' untranslated regions. To have statistically significant analytical results, the cDNAs need to be expressed in at least three cDNA libraries.

[0055] The cDNA libraries used in the coexpression analysis of the present invention can be obtained from adrenal gland, biliary tract, bladder, blood cells, blood vessels, bone marrow, brain, bronchus, cartilage, chromaffin system, colon, connective tissue, cultured cells, embryonic stem cells, endocrine glands, epithelium, esophagus, fetus, ganglia, heart, hypothalamus, immune system, intestine, islets of Langerhans, kidney, larynx, liver, lung, lymph, muscles, neurons, ovary, pancreas, penis, peripheral nervous system, phagocytes, pituitary, placenta, pleurus, prostate, salivary glands, seminal vesicles, skeleton, spleen, stomach, testis, thymus, tongue, ureter, uterus, and the like. The number of cDNA libraries selected can range from as few as 3 to greater than 10,000. Preferably, the number of the cDNA libraries is greater than 500.

[0056] In a preferred embodiment, the cDNAs are assembled from related sequences, such as assembled sequence fragments derived from a single transcript. Assembly of the sequences can be performed using sequences of various types including, but not limited to, ESTs, extensions, or shotgun sequences. In a most preferred embodiment, the cDNAs are derived from human sequences that have been assembled using the algorithm disclosed in U.S. Ser. No. 09/276,534, filed Mar. 25, 1999, incorporated herein by reference.

[0057] Experimentally, differential expression of the cDNAs can be evaluated by methods including, but not limited to, differential display by spatial immobilization or by gel electrophoresis, genome mismatch scanning, representational difference analysis, and transcript imaging. Additionally, differential expression can be assessed by microarray technology. These methods may be used alone or in combination.

[0058] The known atherosclerosis-associated genes were selected based on their function in pathways associated with atherogenesis, their use as diagnostic or prognostic markers, their behavior in model systems or their use as therapeutic targets.

[0059] The procedure for identifying a cDNA that exhibits a statistically significant coexpression pattern with known atherosclerosis-associated genes is as follows. First, the presence or absence of a gene in a cDNA library is defined: a gene is present in a cDNA library when at least one cDNA fragment corresponding to that gene is detected in a cDNA sample taken from the library, and a gene is absent from a library when no corresponding cDNA fragment is detected in the sample.

[0060] Second, the significance of gene coexpression is evaluated using a probability method to measure a due-to-chance probability of the coexpression. The probability method can be the Fisher exact test, the chi-squared test, or the kappa test. These tests and examples of their applications are well known in the art and can be found in standard statistics texts (Agresti (1990) *Categorical Data Analysis*, John Wiley & Sons, New York N.Y.; Rice (1988) *Mathematical Statistics and Data Analysis*, Duxbury Press, Pacific Grove Calif.). A Bonferroni correction (Rice, supra, p. 384) can also be applied in combination with one of the probability methods for correcting statistical results of one gene vetsus multiple other genes. In a preferred embodiment, the due-to-chance probability is measured by a Fisher exact test, and the threshold of the due-to-chance probability is set preferably to less than 0.001, more preferably to less than 0.00001.

[0061] To determine whether two genes, A and B, have similar coexpression patterns, occurrence data vectors can be generated as illustrated in Table 1. The presence of a gene occurring at least once in a library is indicated by a one, and its absence from the library, by a zero.

TABLE 1

Occurrence data for genes A and B

	Library 1	Library 2	Library 3	. . .	Library N
gene A	1	1	0	. . .	0
gene B	1	0	1	. . .	0

[0062] For a given pair of genes, the occurrence data in Table 1 can be summarized in a 2×2 contingency table.

TABLE 2

Contingency table for co-occurrences of genes A and B

	Gene A present	Gene A absent	Total
Gene B present	8	2	10
Gene B absent	2	18	20
Total	10	20	30

[0063] Table 2 presents co-occurrence data for gene A and gene B in a total of 30 libraries. Both gene A and gene B occur 10 times in the libraries. Table 2 summarizes and presents: 1) the number of times gene A and B are both present in a library; 2) the number of times gene A and B are both absent in a library; 3) the number of times gene A is present, and gene B is absent; and 4) the number of times gene B is present, and gene A is absent. The upper left entry is the number of times the two genes co-occur in a library, and the middle right entry is the number of times neither gene occurs in a library. The off diagonal entries are the number of times one gene occurs, and the other does not. Both A and B are present eight times and absent 18 times. Gene A is present, and gene B is absent, two times; and gene B is present, and gene A is absent, two times. The probability ("p-value") that the above association occurs due to chance as calculated using a Fisher exact test is 0.0003. Associations are generally considered significant if a p-value is less than 0.01 (Agresti, supra; Rice, supra).

[0064] This method of estimating the probability for co-expression of two genes makes several assumptions. The method assumes that the libraries are independent and are identically sampled. However, in practical situations, the selected cDNA libraries are not entirely independent, because more than one library may be obtained from a single subject or tissue. Nor are they entirely identically sampled, because different numbers of cDNAs may be sequenced from each library. The number of cDNAs sequenced typically ranges from 5,000 to 10,000 cDNAs per library. In addition, because a Fisher exact co-expression probability is calculated for each gene versus 45,233 other assembled genes, a Bonferroni correction for multiple statistical tests is used.

[0065] The Invention

[0066] The present invention identifies 25 atherosclerosis-associated cDNAs that exhibit strong association with genes known to be specifically expressed in atherosclerosis. The results presented in Tables 4 and 5 show that the expression of the 25 novel atherosclerosis-associated cDNAs have direct association with the expression of known atherosclerosis-associated genes as described in Table 3 and in the background of the invention. Therefore, the novel atherosclerosis-associated cDNAs can potentially be used in diagnosis, prognosis, treatment or evaluation of therapies fro disorders associated with atherosclerosis. Further, the gene products of the 25 novel atherosclerosis-associated cDNAs are either potential therapeutics or targets for the development of therapeutics against disorders associated with atherosclerosis.

[0067] Therefore, in one embodiment, the present invention encompasses a combination comprising a plurality of cDNAs having the nucleic acid sequences of SEQ ID NOs:1-25 or the complements of SEQ ID NOs:1-25. These 25 cDNAs have been shown by the method of the present invention to have statistically significant co-expression with known atherosclerosis-associated genes and with each other. The invention also encompasses a cDNA comprising a polynucleotide having the nucleic acid sequence of SEQ ID NO:8 and the complement thereof. As shown in **FIG. 1**, SEQ ID NO:8 encodes the protein of SEQ ID NO:26. The invention further encompasses a protein comprising the polypeptide having the amino acid sequence of SEQ ID NO:26.

[0068] The protein encoded by SEQ ID NO:8 has 366 amino acids. Motif analyses of SEQ ID NO:26 shows one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at residue S343, two potential casein kinase II phosphorylation sites at residues S179 and T351, and four potential protein kinase C phosphorylation sites at residues T29, S85, T269, and T324. Additionally, SEQ ID NO:26 contains a potential sugar transport protein signature sequence from residues L201 to S217.

[0069] cDNAs and Their Uses

[0070] cDNAs can be prepared by a variety of synthetic or enzymatic methods well known in the art. cDNAs can be synthesized, in whole or in part, using chemical methods well known in the art (Caruthers et al. (1980) Nucleic Acids Symp Ser (7):215-233). Alternatively, cDNAs can be produced enzymatically or recombinantly, by in vitro or in vivo transcription.

[0071] Nucleotide analogs can be incorporated into cDNAs by methods well known in the art. The only requirement is that the incorporated analog must base pair with native purines or pyrimidines. For example, 2,6-diaminopurine can substitute for adenine and form stronger bonds with thymidine than those between adenine and thymidine. A weaker pair is formed when hypoxanthine is substituted for guanine and base pairs with cytosine. Additionally, cDNAs can include nucleotides that have been derivatized chemically or enzymatically.

[0072] cDNAs can be synthesized on a substrate. Synthesis on the surface of a substrate may be accomplished using a chemical coupling procedure and a piezoelectric printing apparatus as described by Baldeschweiler et al. (PCT publication WO95/251116). Alternatively, the cDNAs can be synthesized on a substrate surface using a self-addressable electronic device that controls when reagents are added as described by Heller et al. (U.S. Pat. No. 5,605,662). cDNAs can be synthesized directly on a substrate by sequentially dispensing reagents for their synthesis on the substrate surface or by dispensing preformed DNA fragments to the substrate surface. Typical dispensers include a micropipette delivering solution to the substrate with a robotic system to control the position of the micropipette with respect to the substrate. There can be a multiplicity of dispensers so that reagents can be delivered to the reaction regions efficiently.

[0073] cDNAs can be immobilized on a substrate by covalent means such as by chemical bonding procedures or UV irradiation. In one method, a cDNA is bound to a glass surface which has been modified to contain epoxide or

aldehyde groups. In another method, a cDNA is placed on a polylysine coated surface and UV cross-linked to it as described by Shalon et al. (WO95/35505). In yet another method, a cDNA is actively transported from a solution to a given position on a substrate by electrical means (Heller, supra). cDNAs do not have to be directly bound to the substrate, but rather can be bound to the substrate through a linker group. The linker groups are typically about 6 to 50 atoms long to provide exposure of the attached cDNA. Preferred linker groups include ethylene glycol oligomers, diamines, diacids and the like. Reactive groups on the substrate surface react with a terminal group of the linker to bind the linker to the substrate. The other terminus of the linker is then bound to the cDNA. Alternatively, polynucleotides, plasmids or cells can be arranged on a filter. In the latter case, cells are lysed, proteins and cellular components degraded, and the DNA is coupled to the filter by UV cross-linking.

[0074] The cDNAs may be used for a variety of purposes. For example, the combination of the invention may be used on an array. The array, in turn, can be used in high-throughput methods for detecting a related polynucleotide in a sample, screening a plurality of molecules or compounds to identify a ligand, diagnosing a disorder such as diabetes, or inhibiting or inactivating a therapeutically relevant gene related to the cDNA.

[0075] When the cDNAs of the invention are employed on an array, the cDNAs are arranged in an ordered fashion so that each cDNA is present at a specified location. Because the cDNAs are at specified locations on the substrate, the hybridization patterns and intensities, which together create a unique, can be interpreted in terms of expression levels of particular genes and can be correlated with a particular metabolic process, condition, disorder, disease, stage of disease, or treatment.

[0076] Hybridization

[0077] The cDNAs or fragments or complements thereof may be used in various hybridization technologies. The cDNAs may be labeled using a variety of reporter molecules by either PCR, recombinant, or enzymatic techniques. For example, a commercially available vector containing the cDNA is transcribed in the presence of an appropriate polymerase, such as T7 or SP6 polymerase, and at least one labeled nucleotide. Commercial kits are available for labeling and cleanup of such cDNAs. Radioactive (Amersham Biosciences (APB), Piscataway N.J.), fluorescent (Qiagen-Operon, Alameda Calif.), and chemiluminescent labeling (Promega, Madison Wis.) are well known in the art.

[0078] A cDNA may represent the complete coding region of an mRNA or be designed or derived from unique regions of the mRNA or genomic molecule, an intron, a 3' untranslated region, or from a conserved motif. The cDNA is at least 18 contiguous nucleotides in length and is usually single stranded. Such a cDNA may be used under hybridization conditions that allow binding only to an identical sequence, a naturally occurring molecule encoding the same protein, or an allelic variant. Discovery of related human and mammalian sequences may also be accomplished using a pool of degenerate cDNAs and appropriate hybridization conditions. Generally, a cDNA for use in Southern or northern hybridizations may be from about 400 to about 6000 nucleotides long. Such cDNAs have high binding specificity in solution-based or substrate-based hybridizations. An oligonucleotide, a fragment of the cDNA, may be used to detect a polynucleotide in a sample using PCR.

[0079] The stringency of hybridization is determined by G+C content of the cDNA, salt concentration, and temperature. In particular, stringency is increased by reducing the concentration of salt or raising the hybridization temperature. In solutions used for some membrane based hybridizations, addition of an organic solvent such as formamide allows the reaction to occur at a lower temperature. Hybridization may be performed with buffers, such as 5×saline sodium citrate (SSC) with 1% sodium dodecyl sulfate (SDS) at 60° C., that permit the formation of a hybridization complex between nucleic acid sequences that contain some mismatches. Subsequent washes are performed with buffers such as 0.2×SSC with 0.1% SDS at either 45° C. (medium stringency) or 65°-68° C. (high stringency). At high stringency, hybridization complexes will remain stable only where the nucleic acids are completely complementary. In some membrane-based hybridizations, preferably 35% or most preferably 50%, formamide may be added to the hybridization solution to reduce the temperature at which hybridization is performed. Background signals may be reduced by the use of detergents such as Sarkosyl or TRITON X-100 (Sigma-Aldrich, St. Louis Mo.) and a blocking agent such as denatured salmon sperm DNA. Selection of components and conditions for hybridization are well known to those skilled in the art and are reviewed in Ausubel et al. (1997*, Short Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., Units 2.8-2.11, 3.18-3.19 and 4.6-4.9).

[0080] Dot-blot, slot-blot, low density and high density arrays are prepared and analyzed using methods known in the art. cDNAs from about 18 consecutive nucleotides to about 5000 consecutive nucleotides in length are contemplated by the invention and used in array technologies. The preferred number of cDNAs on an array is at least about 100,000, a more preferred number is at least about 40,000, an even more preferred number is at least about 10,000, and a most preferred number is at least about 600 to about 800. The array may be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and SNPs. Such information may be used to determine gene function; to understand the genetic basis of a disorder; to diagnose a disorder; and to develop and monitor the activities of therapeutic agents being used to control or cure a disorder. (See, e.g., U.S. Pat. No. 5,474,796; WO95/11995; WO95/35505; U.S. Pat. No. 5,605,662; and U.S. Pat. No. 5,958,342.)

[0081] Screening and Purification Assays Using cDNAs

[0082] A cDNA may be used to screen a library or a plurality of molecules or compounds for a ligand which specifically binds the cDNA. Ligands may be DNA molecules, RNA molecules, peptide nucleic acid molecules, peptides, proteins such as transcription factors, promoters, enhancers, repressors, and other proteins that regulate replication, transcription, or translation of the polynucleotide in the biological system. The assay involves combining the cDNA or a fragment thereof with the molecules or compounds under conditions that allow specific binding and detecting the bound cDNA to identify at least one ligand that specifically binds the cDNA.

[0083] In one embodiment, the cDNA may be incubated with a library of isolated and purified molecules or compounds and binding activity determined by methods such as a gel-retardation assay (U.S. Pat. No. 6,010,849) or a reticulocyte lysate transcriptional assay. In another embodiment, the cDNA may be incubated with nuclear extracts from biopsied and/or cultured cells and tissues. Specific binding between the cDNA and a molecule or compound in the nuclear extract is initially determined by gel shift assay. Protein binding may be confirmed by raising antibodies against the protein and adding the antibodies to the gel-retardation assay where specific binding will cause a supershift in the assay.

[0084] In another embodiment, the cDNA may be used to purify a molecule or compound using affinity chromatography methods well known in the art. In one embodiment, the cDNA is chemically reacted with cyanogen bromide groups on a polymeric resin or gel. Then a sample is passed over and reacts with or binds to the cDNA. The molecule or compound which is bound to the cDNA may be released from the cDNA by increasing the salt concentration of the flow-through medium and collected.

[0085] The cDNA may be used to purify a ligand from a sample. A method for using a cDNA to purify a ligand would involve combining the cDNA or a fragment thereof with a sample under conditions to allow specific binding, recovering the bound cDNA, and using an appropriate agent to separate the cDNA from the purified ligand.

[0086] Protein Production and Uses

[0087] The full length cDNAs or fragments thereof may be used to produce purified proteins using recombinant DNA technologies described herein and taught in Ausubel (supra; Units 16.1-16.62). One of the advantages of producing proteins by these procedures is the ability to obtain highly-enriched sources of the proteins thereby simplifying purification procedures.

[0088] The proteins may contain amino acid substitutions, deletions or insertions made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. Such substitutions may be conservative in nature when the substituted residue has structural or chemical properties similar to the original residue (e.g., replacement of leucine with isoleucine or valine) or they may be nonconservative when the replacement residue is radically different (e.g., a glycine replaced by a tryptophan). Computer programs included in LASERGENE software (DNASTAR, Madison Wis.) and algorithms included in RasMol software (University of Massachusetts, Amherst Mass.) may be used to help determine which and how many amino acid residues in a particular portion of the protein may be substituted, inserted, or deleted without abolishing biological or immunological activity.

[0089] Expression of Encoded Proteins

[0090] Expression of a particular cDNA may be accomplished by cloning the cDNA into a vector and transforming this vector into a host cell. The cloning vector used for the construction of cDNA libraries in the LIFESEQ databases (Incyte Genomics, Palo Alto Calif.) may also be used for expression. Such vectors usually contain a promoter and a polylinker useful for cloning, priming, and transcription. An exemplary vector may also contain the promoter for β-galactosidase, an amino-terminal methionine and the subsequent seven amino acid residues of β-galactosidase. The vector may be transformed into competent *E. coli* cells. Induction of the isolated bacterial strain with isopropylthiogalactoside (IPTG) using standard methods will produce a fusion protein that contains an N terminal methionine, the first seven residues of β-galactosidase, about 15 residues of linker, and the protein encoded by the cDNA.

[0091] The cDNA may be shuttled into other vectors known to be useful for expression of protein in specific hosts. Oligonucleotides containing cloning sites and fragments of DNA sufficient to hybridize to stretches at both ends of the cDNA may be chemically synthesized by standard methods. These primers may then be used to amplify the desired fragments by PCR. The fragments may be digested with appropriate restriction enzymes under standard conditions and isolated using gel electrophoresis. Alternatively, similar fragments are produced by digestion of the cDNA with appropriate restriction enzymes and filled in with chemically synthesized oligonucleotides. Fragments of the coding sequence from more than one gene may be ligated together and expressed.

[0092] Signal sequences that dictate secretion of soluble proteins are particularly desirable as component parts of a recombinant sequence. For example, a chimeric protein may be expressed that includes one or more additional purification-facilitating domains. Such domains include, but are not limited to, metal-chelating domains that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex, Seattle Wash.). The inclusion of a cleavable-linker sequence such as ENTEROKINASEMAX (Invitrogen, San Diego Calif.) between the protein and the purification domain may also be used to recover the protein.

[0093] Suitable host cells may include, but are not limited to, mammalian cells such as Chinese Hamster Ovary (CHO) and human 293 cells, insect cells such as Sf9 cells, plant cells such as *Nicotiana tabacum*, yeast cells such as *Saccharomyces cerevisiae*, and bacteria such as *E. coli*. For each of these cell systems, a useful vector may also include an origin of replication and one or two selectable markers to allow selection in bacteria as well as in a transformed eukaryotic host. Vectors for use in eukaryotic host cells may require the addition of 3' poly(A) tail if the cDNA lacks poly(A).

[0094] Additionally, the vector may contain promoters or enhancers that increase gene expression. Many promoters are known and used in the art. Most promoters are host specific and exemplary promoters includes SV40 promoters for CHO cells; T7 promoters for bacterial hosts; viral promoters and enhancers for plant cells; and PGH promoters for yeast. Adenoviral vectors with the rous sarcoma virus enhancer or retroviral vectors with long terminal repeat promoters may be used to drive protein expression in mammalian cell lines. Once homogeneous cultures of recombinant cells are obtained, large quantities of secreted soluble protein may be recovered from the conditioned medium and analyzed using chromatographic methods well known in the art. An alternative method for the production of large amounts of secreted protein involves the transfor-

mation of mammalian embryos and the recovery of the recombinant protein from milk produced by transgenic cows, goats, sheep, and the like.

[0095] In addition to recombinant production, proteins or portions thereof may be produced manually, using solid-phase techniques (Stewart et al. (1969) *Solid-Phase Peptide Synthesis*, W H Freeman, San Francisco Calif.; Merrifield (1963) J Am Chem Soc 5:2149-2154), or using machines such as the 431A peptide synthesizer (Applied Biosystems (ABI), Foster City Calif.). Proteins produced by any of the above methods may be used as pharmaceutical compositions to treat disorders associated with null or inadequate expression of the genomic sequence.

[0096] Screening and Purification Assays Using Proteins

[0097] A protein or a portion thereof encoded by the cDNA may be used to screen a library or a plurality of molecules or compounds for a ligand with specific binding affinity or to purify a molecule or compound from a sample. The protein or portion thereof employed in such screening may be free in solution, affixed to an abiotic or biotic substrate, or located intracellularly. For example, viable or fixed prokaryotic host cells that are stably transformed with recombinant nucleic acids that have expressed and positioned a protein on their cell surface can be used in screening assays. The cells are screened against a library or a plurality of ligands and the specificity of binding or formation of complexes between the expressed protein and the ligand may be measured. The ligands may be agonists, antagonists, antibodies, DNA molecules, enhancers, small drug molecules, immunoglobulins, inhibitors, mimetics, peptide nucleic acid molecules, peptides, pharmaceutical agents, proteins, and regulatory proteins, repressors, RNA molecules, ribozymes, transcription factors, or any other test molecule or compound that specifically binds the protein. An exemplary assay involves combining the mammalian protein or a portion thereof with the molecules or compounds under conditions that allow specific binding and detecting the bound protein to identify at least one ligand that specifically binds the protein.

[0098] This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound capable of binding to the protein or oligopeptide or fragment thereof. One method for high throughput screening using very small assay volumes and very small amounts of test compound is described in U.S. Pat. No. 5,876,946. Molecules or compounds identified by screening may be used in a model system to evaluate their toxicity, diagnostic, or therapeutic potential.

[0099] The protein may be used to purify a ligand from a sample. A method for using a protein to purify a ligand would involve combining the protein or a portion thereof with a sample under conditions to allow specific binding, recovering the bound protein, and using an appropriate chaotropic agent to separate the protein from the purified ligand.

[0100] Production of Antibodies

[0101] A protein encoded by a cDNA of the invention may be used to produce specific antibodies. Antibodies may be produced using an oligopeptide or a portion of the protein with inherent immunological activity. Methods for producing antibodies include: 1) injecting an animal, usually goats, rabbits, or mice, with the protein, or an antigenically-effective portion or an oligopeptide thereof, to induce an immune response; 2) engineering hybridomas to produce monoclonal antibodies; 3) inducing in vivo production in the lymphocyte population; or 4) screening libraries of recombinant immunoglobulins. Recombinant immunoglobulins may be produced as taught in U.S. Pat. No. 4,816,567.

[0102] Antibodies produced using the proteins of the invention are useful for the diagnosis of prepathologic disorders as well as the diagnosis of chronic or acute diseases characterized by abnormalities in the expression, amount, or distribution of the protein. A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies specific for proteins are well known in the art. Immunoassays typically involve the formation of complexes between a protein and its specific binding molecule or compound and the measurement of complex formation. Immunoassays may employ a two-site, monoclonal-based assay that utilizes monoclonal antibodies reactive to two noninterfering epitopes on a specific protein or a competitive binding assay (Pound (1998) *Immunochemical Protocols*, Humana Press, Totowa N.J.).

[0103] Immunoassay procedures may be used to quantify expression of the protein in cell cultures, in subjects with a particular disorder or in model animal systems under various conditions. Increased or decreased production of proteins as monitored by immunoassay may contribute to knowledge of the cellular activities associated with developmental pathways, engineered conditions or diseases, or treatment efficacy. The quantity of a given protein in a given tissue may be determined by performing immunoassays on freeze-thawed detergent extracts of biological samples and comparing the slope of the binding curves to binding curves generated by purified protein.

[0104] Antibody Arrays

[0105] In an alternative to yeast two hybrid system analysis of proteins, an antibody array can be used to study protein-protein interactions and phosphorylation. A variety of protein ligands are immobilized on a membrane using methods well known in the art. The array is incubated in the presence of cell lysate until protein:antibody complexes are formed. Proteins of interest are identified by exposing the membrane to an antibody specific to the protein of interest. In the alternative, a protein of interest is labeled with digoxigenin (DIG) and exposed to the membrane; then the membrane is exposed to anti-DIG antibody which reveals where the protein of interest forms a complex. The identity of the proteins with which the protein of interest interacts is determined by the position of the protein of interest on the membrane.

[0106] Antibody arrays can also be used for high-throughput screening of recombinant antibodies. Bacteria containing antibody genes are robotically-picked and gridded at high density (up to 18,342 different double-spotted clones) on a filter. Up to 15 antigens at a time are used to screen for clones to identify those that express binding antibody fragments. These antibody arrays can also be used to identify proteins which are differentially expressed in samples (de Wildt et al. (2000) Nature Biotechnol 18:989-94).

[0107] Labeling of Molecules for Assay

[0108] A wide variety of reporter molecules and conjugation techniques are known by those skilled in the art and may be used in various cDNA, polynucleotide, protein, peptide or antibody assays. Synthesis of labeled molecules may be achieved using commercial kits for incorporation of a labeled nucleotide such as ^{32}P-dCTP, Cy3-dCTP or Cy5-dCTP or amino acid such as ^{35}S-methionine. Polynucleotides, cDNAs, proteins, or antibodies may be directly labeled with a reporter molecule by chemical conjugation to amines, thiols and other groups present in the molecules using reagents such as BIODIPY or FITC (Molecular Probes, Eugene Oreg.).

[0109] The proteins and antibodies may be labeled for purposes of assay by joining them, either covalently or noncovalently, with a reporter molecule that provides for a detectable signal. A wide variety of labels and conjugation techniques are known and have been reported in the scientific and patent literature including, but not limited to U.S. Pat. No. 3,817,837; U.S. Pat. No. 3,850,752; U.S. Pat. No. 3,939,350; U.S. Pat. No. 3,996,345; U.S. Pat. No. 4,277,437; U.S. Pat. No. 4,275,149; and U.S. Pat. No. 4,366,241.

[0110] Diagnostics

[0111] The cDNAs, or fragments thereof, may be used to detect and quantify differential gene expression; absence, presence, or excess expression of mRNAs; or to monitor mRNA levels during therapeutic intervention. Disorders associated with atherosclerosis include angina pectoris, coronary artery disease, myocardial infarction, hypertension, transient cerebral ischemia, mesenteric ischemia, peripheral vascular disease, renal artery stenosis, and stroke. These cDNAs can also be utilized as markers of treatment efficacy against the disorders noted above and other disorders, conditions, and diseases over a period ranging from several days to months. The diagnostic assay may use hybridization or amplification technology to compare gene expression in a biological sample from a patient to standard samples in order to detect altered gene expression. Qualitative or quantitative methods for this comparison are well known in the art.

[0112] For example, the cDNA may be labeled by standard methods and added to a biological sample from a patient under conditions for hybridization complex formation. After an incubation period, the sample is washed and the amount of label (or signal) associated with hybridization complexes is quantified and compared with a standard value. If the amount of label in the patient sample is significantly altered in comparison to the standard value, then the presence of the associated condition, disease or disorder is indicated.

[0113] In order to provide a basis for the diagnosis of a condition, disease or disorder associated with gene expression, a normal or standard expression profile is established. This may be accomplished by combining a biological sample taken from normal subjects, either animal or human, with a probe under conditions for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified target sequence is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular condition is used to diagnose that condition.

[0114] Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies and in clinical trial or to monitor the treatment of an individual patient. Once the presence of a condition is established and a treatment protocol is initiated, diagnostic assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in a normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

[0115] Gene Expression Profiles

[0116] A gene expression profile comprises a plurality of cDNAs and a plurality of detectable hybridization complexes, wherein each complex is formed by hybridization of one or more probes to one or more complementary nucleic acids in a sample. The cDNAs of the invention are used as elements on a array to analyze gene expression profiles. In one embodiment, the array is used to monitor the progression of disease. Researchers or clinicians can catalog the differences in gene expression between healthy and diseased tissues or cells. By analyzing changes in patterns of gene expression, disease can be diagnosed at earlier stages before the patient is symptomatic. The invention can be used to formulate a prognosis and to design a treatment regimen. The invention can also be used to monitor the efficacy of treatment. For treatments with known side effects, the array is employed to improve the treatment regimen. A dosage is established that causes a change in genetic expression patterns indicative of successful treatment. Expression patterns associated with the onset of undesirable side effects are avoided. This approach may be more sensitive and rapid than waiting for the patient to show inadequate improvement, or to manifest side effects, before altering the course of treatment.

[0117] Experimentally, expression profiles can also be evaluated by methods including, but not limited to, differential display by spatial immobilization or by gel electrophoresis, labeling with radionuclide and quantification using a scintillation counter, genome mismatch scanning, representational difference analysis, transcript imaging, quantitative PCR, and by protein or antibody arrays. Expression profiles produced by these methods may be contrasted with expression profiles produced using normal or diseased tissues. Of note is the correspondence between mRNA and protein expression has been discussed by Zweiger (2001, *Transducing the Genome*. McGraw-Hill, San Francisco, Calif.) and Glavas et al. (2001; T cell activation upregulates cyclic nucleotide phosphodiesterases 8A1 and 7A3, Proc Natl Acad Sci 98:6319-6342) among others.

[0118] In another embodiment, animal models which mimic a human disease can be used to produce expression profiles associated with a particular condition, disorder or disease; or treatment of the condition, disorder or disease. Novel treatment regimens may be tested in these animal models using arrays to establish and then follow expression profiles over time. In addition, arrays may be used with cell cultures or tissues removed from animal models to rapidly screen large numbers of candidate drug molecules, looking

for ones that produce an expression profile similar to those of known therapeutic drugs, with the expectation that molecules with the same expression profile will likely have similar therapeutic effects. Thus, the invention provides the means to rapidly determine the molecular mode of action of a drug.

[0119] Assays Using Antibodies

[0120] Antibodies directed against antigenic determinants of a protein encoded by a cDNA of the invention may be used in assays to quantify the amount of protein found in a particular human cell. Such assays include methods utilizing the antibody and a label to detect expression level under normal or disease conditions. The antibodies may be used with or without modification, and labeled by joining them, either covalently or noncovalently, with a labeling moiety.

[0121] Protocols for detecting and measuring protein expression using either polyclonal or monoclonal antibodies are well known in the art. Examples include, but are not limited to, western analysis, ELISA, RIA, FACS, and arrays. Such immunoassays typically involve the formation of complexes between the protein and its specific antibody and the measurement of such complexes. These assays are specifically described in Pound (supra).

[0122] Therapeutics

[0123] The cDNAs and fragments thereof can be used in gene therapy. cDNAs can be delivered ex vivo to target cells, such as cells of bone marrow. Once stable integration and transcription and or translation are confirmed, the bone marrow may be reintroduced into the subject. Expression of the protein encoded by the cDNA may correct a disorder associated with mutation of a normal sequence, reduction or loss of an endogenous target protein, or overepression of an endogenous or mutant protein. Alternatively, cDNAs may be delivered in vivo using vectors such as retrovirus, adenovirus, adeno-associated virus, herpes simplex virus, and bacterial plasmids. Non-viral methods of gene delivery include cationic liposomes, polylysine conjugates, artificial viral envelopes, and direct injection of DNA (Anderson (1998) Nature 392:25-30; Dachs et al. (1997) Oncol Res 9:313-325; Chu et al. (1998) J Mol Med 76(3-4):184-192; Weiss et al. (1999) Cell Mol Life Sci 55(3):334-358; Agrawal (1996) *Antisense Therapeutics*, Humana Press, Totowa N.J.; and August et al. (1997) *Gene Therapy* (*Advances in Pharmacology*, Vol. 40), Academic Press, San Diego Calif.).

[0124] In addition, expression of a particular protein can be regulated through the specific binding of a fragment of a cDNA to a genomic sequence or an mRNA which encodes the protein or directs its transcription or translation. The cDNA can be modified or derivatized to any RNA-like or DNA-like material including peptide nucleic acids, branched nucleic acids, and the like. These sequences can be produced biologically by transforming an appropriate host cell with a vector containing the sequence of interest.

[0125] Molecules which regulate the activity of the cDNA or encoded protein are useful as therapeutics for diabetes mellitus, obesity, hypertension, atherosclerosis, polycystic ovarian syndrome, and cancers including breast, prostate, and colon. Such molecules include agonists which increase the expression or activity of the polynucleotide or encoded protein, respectively; or antagonists which decrease expression or activity of the polynucleotide or encoded protein, respectively. In one aspect, an antibody which specifically binds the protein may be used directly as an antagonist or indirectly as a delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express the protein.

[0126] Additionally, any of the proteins, or their ligands, or complementary nucleic acid sequences may be administered as pharmaceutical compositions or in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to affect the treatment or prevention of the conditions and disorders associated with an immune response. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects. Further, the therapeutic agents may be combined with pharmaceutically-acceptable carriers including excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration used by doctors and pharmacists may be found in the latest edition of Remington's Pharmaceutical Sciences (Mack Publishing, Easton Pa.).

[0127] Model Systems

[0128] Animal models may be used as bioassays where they exhibit a phenotypic response similar to that of humans and where exposure conditions are relevant to human exposures. Mammals are the most common models, and most infectious agent, cancer, drug, and toxicity studies are performed on rodents such as rats or mice because of low cost, availability, lifespan, reproductive potential, and abundant reference literature. Inbred and outbred rodent strains provide a convenient model for investigation of the physiological consequences of underexpression or overexpression of genes of interest and for the development of methods for diagnosis and treatment of diseases. A mammal inbred to overexpress a particular gene (for example, secreted in milk) may also serve as a convenient source of the protein expressed by that gene.

[0129] Transgenic Animal Models

[0130] Transgenic rodents that overexpress or underexpress a gene of interest may be inbred and used to model human diseases or to test therapeutic or toxic agents. (See, e.g., U.S. Pat. No. 5,175,383 and U.S. Pat. No. 5,767,337.) In some cases, the introduced gene may be activated at a specific time in a specific tissue type during fetal or postnatal development. Expression of the transgene is monitored by analysis of phenotype, of tissue-specific mRNA expression, or of serum and tissue protein levels in transgenic animals before, during, and after challenge with experimental drug therapies.

[0131] Embryonic Stem Cells

[0132] Embryonic (ES) stem cells isolated from rodent embryos retain the potential to form embryonic tissues. When ES cells such as the mouse 129/SvJ cell line are placed in a blastocyst from the C57BL/6 mouse strain, they resume normal development and contribute to tissues of the live-born animal. ES cells are preferred for use in the creation of experimental knockout and knockin animals. The method for this process is well known in the art and the steps

are: the cDNA is introduced into a vector, the vector is transformed into ES cells, transformed cells are identified and microinjected into mouse cell blastocysts, blastocysts are surgically transferred to pseudopregnant dams. The resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains.

[0133] Knockout Analysis

[0134] In gene knockout analysis, a region of a gene is enzymatically modified to include a non-natural intervening sequence such as the neomycin phosphotransferase gene (neo; Capecchi (1989) Science 244:1288-1292). The modified gene is transformed into cultured ES cells and integrates into the endogenous genome by homologous recombination. The inserted sequence disrupts transcription and translation of the endogenous gene.

[0135] Knockin Analysis

[0136] ES cells can be used to create knockin humanized animals or transgenic animal models of human diseases. With knockin technology, a region of a human gene is injected into animal ES cells, and the human sequence integrates into the animal cell genome. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on the progression and treatment of the analogous human condition.

[0137] As described herein, the uses of the cDNAs, provided in the Sequence Listing of this application, and their encoded proteins are exemplary of known techniques and are not intended to reflect any limitation on their use in any technique that would be known to the person of average skill in the art. Furthermore, the cDNAs provided in this application may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known to the person of ordinary skill in the art, e.g., the triplet genetic code, specific base pair interactions, and the like. Likewise, reference to a method may include combining more than one method for obtaining or assembling full length cDNA sequences that will be known to those skilled in the art. It is also to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary. It is also understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

EXAMPLES

[0138] It is to be understood that this invention is not limited to the particular devices, machines, materials and methods described. Although particular embodiments are described, equivalent embodiments may be used to practice the invention. The described embodiments are provided to illustrate the invention and are not intended to limit the scope of the invention which is limited only by the appended claims.

[0139] I cDNA Library Construction

[0140] The cDNA library SMCCNOS01 was selected as an example to demonstrate the construction of cDNA libraries from which the cDNAs co-expressed with known atherosclerosis-associated genes were derived. The SMCCNOS01 subtracted coronary artery smooth muscle cell library was constructed using 7.56×10^{6} clones from the SMCCNOT02 library and was subjected to two rounds of subtraction hybridization for 48 hours with 6.12×10^{6} clones from SMCCNOT01.

[0141] The SMCCNOT02 library was constructed using RNA isolated from coronary artery smooth muscle cells removed from a 3-year-old Caucasian male. The cells were treated for 20 hours with TNFα and IL-1β at 10 ng/ml each. The SMCCNOT01 was constructed using RNA isolated from untreated coronary artery smooth muscle cells from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (1991; Nucleic Acids Res 19:1954) and Bonaldo et al. (1996; Genome Research 6:791).

[0142] For both cDNA libraries, SMCCNOT01 and SMCCNOT02, the frozen coronary artery smooth muscle cells (50-100 mg) were homogenized in GTC buffer (4.0M guanidine thiocyanate, 0.1M Tris-HCl pH 7.5, 1% 2-mercaptoethanol). Two volumes of binding buffer (0.4M LiCl, 0.1M Tris-HCl pH 7.5, 0.02M EDTA) were added, and the resulting mixture was vortexed at 13,000 rpm. The supernatant was removed and combined with oligo $d(T)_{25}$ bound streptavidin particles (MPG). After rotation at room temperature, the mRNA-oligo $d(T)_{25}$ bound streptavidin particles were separated from the supernatant, washed twice with hybridization buffer 1 (0.15M NaCl, 0.01M Tris-HCl pH 8.0, 1 mM EDTA, 0.1% lauryl sarcosinate) using magnetic separation at each step to remove the supernatant from the particles. Bound mRNA was eluted from the particles with release solution and heated to 65° C. The supernatant containing eluted mRNA was magnetically separated from the particles and used to construct the cDNA libraries.

[0143] The RNA was used according to the recommended protocols in the SUPERSCRIPT plasmid system (Invitrogen). The cDNAs were fractionated on a SEPHAROSE CLIB column (APB), and those cDNAs exceeding 400 bp were ligated into pINCY plasmid (Incyte Genomics, Palo Alto Calif.). Recombinant plasmids were transformed into DH5α competent cells or ELECTROMAX cells (Invitrogen).

[0144] II Isolation and Sequencing of cDNA Clones

[0145] Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Qiagen, Valencia Calif.). The recommended protocol was employed except for the following changes: 1) the bacteria were inoculated into 1 ml of sterile TERRIFIC BROTH (BD Biosciences, San Jose Calif.) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) the cells were cultured for 19 hours and then lysed with 0.3 ml of lysis buffer; 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml distilled water, and 4) the samples were transferred to a 96-well block for storage at 4° C.

[0146] The cDNAs were prepared using a MICROLAB 2200 system (Hamilton, Reno Nev.) in combination with the DNA ENGINE thermal cycler (MJ Research, Watertown Mass.). cDNAs were sequenced by the method of Sanger et al. (1975, J Mol Biol 94:441-446) using PRISM 377 (ABI) or MEGABACE 1000 sequencing systems (APB).

[0147] Most of the sequences were sequenced using standard protocols and kits (ABI) at solution volumes of 0.25×-1.0×concentrations. In the alternative, some of the sequences disclosed herein were sequenced using solutions and dyes from APB.

[0148] III Selection, Assembly, and Characterization of Sequences

[0149] The sequences used for co-expression analysis were assembled from EST sequences, 5' and 3' longread sequences, and full length coding sequences. The cDNAs claimed herein were expressed in at least three libraries.

[0150] The assembly process is described as follows. EST sequence chromatograms were processed and verified. Quality scores were obtained using PHRED (Ewing et al. (1998) Genome Res 8:175-185; Ewing and Green (1998) Genome Res 8:186-194), and edited sequences were loaded into a relational database management system (RDBMS). The sequences were clustered using BLAST with a product score of 50. All clusters of two or more sequences created a bin which represents one transcribed gene.

[0151] Assembly of the component sequences within each bin was performed using a modification of Phrap, a publicly available program for assembling DNA fragments (Green, P. University of Washington, Seattle Wash.). Bins that showed 82% identity from a local pair-wise alignment between any of the consensus sequences were merged.

[0152] Bins were annotated by screening the consensus sequence in each bin against public databases, such as GBpri and GenPept from NCBI. The annotation process involved a FASTn screen against the GBpri database in GenBank. Those hits with a percent identity of greater than or equal to 75% and an alignment length of greater than or equal to 100 base pairs were recorded as homolog hits. The residual unannotated sequences were screened by FASTx against GenPept. Those alignments with an E value of less than or equal to 10^{-8} were recorded as homologs.

[0153] Sequences were then reclustered using BLASTn and Cross-Match, a program for rapid amino acid and nucleic acid sequence comparison and database search (Green, supra), sequentially. Any BLAST alignment between a sequence and a consensus sequence with a score greater than 150 was realigned using cross-match. The sequence was added to the bin whose consensus sequence gave the highest Smith-Waterman score (Smith et al. (1992) Protein Engineering 5:35-51) amongst local alignments with at least 82% identity. Non-matching sequences were moved into new bins, and assembly processes were repeated.

[0154] IV Coexpression Analyses of Atherosclerosis-Associated Genes

[0155] Known atherosclerosis-associated genes were selected to identify the cDNAs that are closely associated with atherosclerosis. The known atherosclerosis-associated genes which were used in this analysis all occur within the LIFESEQ Gold database (Incyte Genomics), and brief descriptions of their functions as they have been reported in the literature are listed in Table 3.

TABLE 3

Descriptions of Known Atherosclerosis-Associated Genes

GENE	DESCRIPTION AND REFERENCES
Human 22kDa smooth muscle protein (SM22)	Smooth muscle cell-specific gene which is down-regulated during smooth muscle cell dedifferentiation as part of atherogenic process (Sobue et al. (1998) Horm Res 50(S2):15–24; Sobue et al. (1999) Mol Cell Biochem 190:105–18)
calponin (CNN1)	Calponin is smooth muscle-specific and may mediate smooth muscle contractility through binding of the amino-terminal end of the myosin regulatory light chain. Involved in phenotypic modulation of smooth muscle cells, a feature of atherosclerosis (Szymanski et al.(1999) Biochemistry 38:3778–84)
pro alpha 1(I) collagen (COL1A1)	Member of family of fibrous structural proteins. Most abundant structural component of the extracellular matrix. Secreted as procollagen and converted to collagen by matrix metalloproteinases. Collagens are important in atherosclerosis for promoting platelet aggregation and for providing sites for platelet adhesion to the vessel wall (Wen et al. (1999) Arterioscler Thromb Vasc Biol 19:519–24)
collagen alpha-2 type I (COL1A2)	see COL1A1 above
COL6A1	see COL1A1 above
procollagen alpha 2(V) (COL5A2)	see COL1A1 above
collagen VI alpha-2 (COL6A2)	see COL1A1 above
type VI collagen alpha3 (COL6A3)	see COL1A1 above
pro-alpha-1 type 3 collagen (COL3A1)	see COL1A1 above
pro-alpha-1 (V) collagen (COL3A1)	see COL1A1 above
matrix Gla protein (MGP)	Role in active calcification of vascular smooth muscle cells, suggested by expression study on VSMC in vitro differentiation study. Calcifying phenotype associated with high MGP levels. MGP knockout mice develop to term, but die up to 2 months after birth due to extensive calcification of the arteries, causing blood vessel rupture (Luo et al. (1997) Nature 386:78–81; Mori et al. (1998) FEES Lett 433:19–22)

TABLE 3-continued

Descriptions of Known Atherosclerosis-Associated Genes

GENE	DESCRIPTION AND REFERENCES
cathepsin K (CTSK)	Nonmetalloenzyme, potent elastase present in advanced atherosclerotic plaques. Contributes to the breakdown of components of vascular extracellular matrix, reducing tensile strength, increasing plaque vulnerability (Sukhova et al.(1998) J Clin Invest 102:576–83)
fibrinogen beta chain gene (FGB)	Component of fibrin in the extracellular matrix. Fibrin deposition is an integral part of advanced atherosclerotic lesion development. Variation at the beta fibrinogen locus associated with peripheral atherosclerosis (Sueishi et al. (1998) Semin Thromb Hemost 24:255–260; Fowkes et al.(1992) Lancet 339:693–696)
pre-pro-von Willebrand factor (VWF)	Blood glycoprotein involved in normal hemostasis. Mediates adhesion of platelets to sites of vascular damage. Also acts as a cofactor in factor VIII activity in blood coagulation. Increased levels of VWF are found in atherosclerosis and in several of its major risk factors, including hypercholesterolemia, diabetes, obesity, hypertension. Levels serve as a predictor of adverse clinical outcome following vascular surgery, possibly as an indicator of thrombus formation (Sadler (1998) Annu Rev Biochem 67:395–424 Blann et al. (1994) Eur J Vasc Surg 8:10–15; Kessler et al. (1998) Diabetes Metab 24:327–36; Folsom et al. (1997) Circulation 96:1102–1108)
platelet endothelial cell adhesion molecule (PECAM-1)	Signaling molecule in the migration of cells as part of the pathophysiology of vascular occulsive diseases such as atherosclerosis. Analysis of endothelial/monocyte co-cultures indicates oxidative stress induces transendothelial migration of monocytes as a result of phosphorylation of PECAM-1 (Rattan et al. (1997) Am J Physiol 273:E453–61)
antithrombin III variant (AT3)	ATIII is the sole blood component through which heparin exerts its anti-coagulation effect. Deficiency in ATIII causes recurrent venous thrombosis and pulmonary embolism and can be inherited in autosomal dominant fashion (Hultin et al. (1988) Thromb Haemost 59:468–73; Lane et al. (1996) Blood Rev 10:59–74)
lipoprotein lipase (LPL)	Hydrolyses triglyceride in chylomicrons and therefore regulates metabolism of circulating lipoproteins. Appears to have an atherogenic effect on the arterial wall due to its ability to alter the properties of LDL. Increased activity of LPL is found in atherosclerotic arteries when compared to normal. Expressed by macrophages in atherosclerotic lesions. Mutations in LPL responsible for familial hypercholesterolemia and premature atherosclerosis (Fisher et al. (1997) Atherosclerosis 135:145–159; Goldberg (1996) J Lipid Res 37:693–707; Gerdes et al. (1997) Circulation 96:733–740)
alpha-2-macro-globulin (A2M)	Foam cell formation—retains LDL cholesterol in the lipid core of atherosclerotic plaque (Llorente et al. (1998) Rev Esp Cardiol 51:633–641)
apolipoprotein AI (APOA1)	Participates in reverse cholesterol transport from tissues to the liver. Promotes cholesterol efflux from tissues and acts as a cofactor for lecithin cholesterol acyltransferase (LCAT). Mutations in ApoA1 and of ApoAI/CIII/AIV gene cluster assoc with atherosclerosis. Transgenic mice expressing high plasma APOAI levels are protected from fatty streak development with a high atherogenic diet (Gordon et al. (1989) Circulation 79:8–15; Rubin et al. (1991) Nature 353:265–7; Karathanasis et al. (1987) Proc Natl Acad Sci 84:7198–7202)
apolipoprotein AII (APOA2)	Major component of HDL. Appears to have an opposite effect to that of APOAI, though exact function unknown. APOAII may have ability to convert HDL from an anti- to a pro-inflammatory particle, with paraoxonase having a role in this transformation process. Plasma APOAII levels significantly associated with plasma free fatty acid levels. Transgenic mice expressing varying levels of APOAII show increased atherosclerotic lesions than wt when fed an atherogenic diet. Possible interaction between diet/genotype and atherogenic potential (Escola-Gil et al. (1998) J Lipid Res 39:457–462; Warden et al. (1993) Proc Natl Acad Sci 90:10886–10890)
apolipoprotein B-100 (APOB)	Main apolipoprotein of chylomicrons and low density lipoproteins. Mutations in APOB100 underly familial defective apolipoprotein B-100 in which patients suffer from premature atherosclerosis. Mutations result in defect in binding of LDL to LDL receptor, and accumulation of plasma LDL. High-expressing APOB transgenic mice exibit elevated VLDL-LDL cholesterol and atherogenic lesions (Callow et al. (1995) J Clin Invest 96:1639–1646; Brasaemle et al.(1997) J Biol Chem 272:9378–9387)
lipoprotein apoCII (APOC2)	Role in lipoprotein metabolism. Cofactor in the activity of lipoprotein lipase the enzyme that hydrolyzes triglycerides in plasma and transfers the fatty acids to tissues. Mutations in APOC2 responsible for hyperlipoproteinemia 1B, similar to lipoprotein lipase deficiency (Cox et al. (1978) N Engl J Med 299:1421–1424; Arimoto et al. (1998) J Lipid Res 39:143–151)
pre-apolipoprotein CIII (APOC3)	Inhibits lipoprotein lipase and hepatic lipase, decreases uptake of lymph chylomicrons by hepatic cells. APOA3 possibly delays breakdown of triglyceride rich particles. SstI RFLP in apoCIII is associated with plasma triglyceride and apoCIII levels and hyperlipidemic phenotypes (Henderson et al.(1987) Hum Genet 75:62–65)

TABLE 3-continued

Descriptions of Known Atherosclerosis-Associated Genes

GENE	DESCRIPTION AND REFERENCES
apolipoprotein apoC-IV (APOC4)	APOC4 is a lipid-binding protein that has the potential to alter lipid metabolism. Human APOC4 transgenic mice are hypertriglyceridaemic compared to normal controls (Allan et al. (1996) J Lipid Res 37:1510–1518)
macrophage scavenger receptor type I (MSR1)	Mediates binding, internalisation and processing of negatively-charged macro-molecules. Implicated in the pathological deposition of cholesterol in arterial walls during atherogenesis (Han et al. (1998) Hum Mol Genet 7:1039–1046)
Human antigen CD36 gene (CD36)	Acts as a scavenger receptor for oxidised LDL. Transient regulation under control of M-CSF during monocyte-macrophage differentiation increases foam cell accumulation, Possible role in atherogenesis: increased M-CSF levels detected in atherosclerotic lesions in rabbits and humans (Huh et al. (1996) Blood 87:2020–2028; Aitman et al. (1999) Nat Genet 21:76–83)
serum amyloid P component (SAP)	Plasma glycoprotein expressed in atherosclerotic lesions. Interacts with lipoproteins in specific manner (Li et al. (1995) Arterioscler Thromb Vasc Biol 15:252–257; Li et al. (1998) Biochem Biophys Res Commun 244:249–252)
carboxyl ester lipase gene (CEL)	CEL gene expression increases in presence of oxidised and native LDL in vitro. It is expressed in the vessel wall and in aortic extracts - may interact with cholesterol to modulate progression of atherosclerosis (Li et al. (1998) Biochem J 329:675–679)
paraoxonase 1 (PON1)	Serum esterase exclusively associated with high-density lipoproteins; it might confer protection against coronary artery disease by destroying pro-inflammatory oxidized lipids in oxidized low-density lipoproteins. PON1 gln192-to-arg polymorphism associated with coronary artery disease. Association between PON1 genetic variation and plasma LDL, HDL and non-HDL and apoB levels in genetically isolated Alberta Hutterite population. When fed on a high-fat, high-cholesterol diet, PON1-null mice were more susceptible to atherosclerosis than wild-type (Serrato et al. (1995) J Clin Invest 96:3005–3008; Boright et al. (1998) Atherosclerosis 139:131–136; Shih et al. (1998) Nature 394:284–287)
paraoxonase 2 (PON2)	Serum esterase exclusively associated with high-density lipoproteins; it might confer protection against coronary artery disease by destroying pro-inflammatory oxidized lipids in oxidized low-density lipoproteins. Common polymorphism at codon 311 (cys-ser) in PON2 associated with CHD alone and synergistically with the 192 polymorphism in PON1 in Asian Indians. Association between genetic variation in PON2 and plasma cholesterol and apolipoprotein A1 in genetically isolated Alberta Hutterite population (Sanghera et al. (1998) Am J Hum Genet 62:36–44; Boright supra)
paraoxonase 3 (PON3)	Serum esterase exclusively associated with high-density lipoproteins; it might confer protection against coronary artery disease by destroying pro-inflammatory oxidized lipids in oxidized low-density lipoproteins. Other members PON2, 3 associated with CHD and cholesterol levels (Laplaud et al. (1998) Clin Chem Lab Med 36:431–441)
perilipin (PLIN)	Lipid storage droplets of steroidogenic cells are surrounded by perilipins, family of phosphorylated proteins encoded by a singlegene, detected in adipocytes and steroidogenic cells. Possible role in lipid metabolism (Brasaemle et al. (1997) J Biol Chem 272:9378–9387)
Prostaglandin D2 synthase (PTGDS)	Catalyses conversion of PGH2 to PGD2, a prostaglandin important in smooth muscle contraction/relaxation and potent inhibitor of platelet aggregation. Northern analysis shows strong specific expression in heart. Immunocyto-chemical localization to myocardial and atrio endocardial cells, and accumulates in end-stage atherosclerotic plaques. High plasma levels detected in severe angina patients (Eguchi et al. (1997) Proc Natl Acad Sci 94:14689–14694)
Annexin II/lipocortinII(ANX2)	Inhibits phospholipase A2 activity and the production of arachidonic acid, the precursor of the inflammatory mediators prostaglandins and leukotrienes. ANX2 is an important anti-inflammatory molecule that binds plasminogen and t-PA and is suspected of having a role in atherogenesis. Binding of plasminoger to ANX2 is specifically inhibited by the excess atherogenic Lp(a) (Hajjar et al. (1998) J Investig Med 46:364–369)
Annexin I/lipocortin(ANX1)	Inhibits phospholipase A2 activity and production of arachidonic acid, the precursor of the inflammatory mediators prostaglandins and leukotrienes. ANXI is an important anti-inflammatory molecule (Wallner et al. (1986) Nature 320:77–81)
Secreted protein, acidic and rich in cysteine (SPARC)	Extracellular glycoprotein secreted by endothelial cells which has a suspected role in calcification of atherosclerotic plaques. Interacts with PDGF-B containing dimers and inhibits binding to its receptors. Expression of SPARC and PDGF is minimal in most adult tissues, but is enhanced following injury and advanced atherosclerotic lesions. Selective expression of SPARC causes rounding of adherent endothelial cells and influences extravasation of macromolecules (Raines et al. (1992) Proc Natl Acad Sci 89:1281–1285; Goldblum et al. (1994) Proc Natl Acad Sci 91:3448–3452)

[0156] From a total of 45,233 assembled gene sequences, 25 cDNAs were identified (SEQ ID NOs:1-25 of the Sequence Listing) that show strong association with the known atherosclerosis-associated genes.

[0157] Initially, the degree of association was measured by probability values using a cutoff p-value less than 0.00001. The sequences were further examined to ensure that the genes that passed the probability test had strong association with known atherosclerosis-associated genes. Details of the co-expression patterns for the known genes and co-expressed cDNAs are presented in Table 4. The entries in Table 4 are the negative log of the p-value (−log p) for the coexpression of the two genes. Table 5 summarizes the highly significant co-expression relationships between each cDNA, two marker genes and their functions.

[0158] V Atherosclerosis-Associated cDNAs

[0159] Using the co-expression analysis method, cDNAs comprising the polynucleotides of SEQ ID NOs:1-25 and their complements, were identified by their highly significant co-expression with known atherosclerosis-associated genes.

[0160] BLAST and other motif searches were performed for SEQ ID NOs:1-25 according to Example VII. SEQ ID NO:8 was determined to be full length and translated as shown in **FIG. 1**.

[0161] VI Transcript Imaging

[0162] Transcript images were performed for several of the cDNAs of the invention using the LIFESEQ GOLD database (July 02 release, Incyte Genomics). This process allowed assessment of the relative abundance of the expressed polynucleotides in all of the cDNA libraries, but those in the cardiovascular category are specifically emphasized. Criteria for transcript imaging can be selected from category, number of cDNAs per library, library description, disease indication, clinical relevance of sample, and the like.

[0163] All sequences and cDNA libraries in the LIFESEQ database have been categorized by system, organ/tissue and cell type. For each category, the number of libraries in which the sequence was expressed were counted and shown over the total number of libraries in that category. For each library, the number of cDNAs were counted and shown over the total number of cDNAs in that library. In some transcript images, all normalized or subtracted libraries, which have high copy number sequences removed prior to processing, and all mixed or pooled tissues, which are considered non-specific in that they contain more than one tissue type or more than one subject's tissue, can be excluded from the analysis. Treated and untreated cell lines and/or fetal tissue data can also be excluded where clinical relevance is emphasized. Conversely, fetal tissue can be emphasized wherever elucidation of inherited disorders or differentiation of particular adult or embryonic stem cells into tissues or organs such as heart, kidney, nerves or pancreas would be aided by removing clinical samples from the analysis. Transcript imaging can also be used to support data from other methodologies such as guilt-by-association and hybridization analyses.

[0164] The transcript images for SEQ ID NOs:3-6, 8,13, 15-20, and 22 are shown in Table 6. The first column shows library name; the second column, the number of cDNAs sequenced in that library; the third column, the description of the library; the fourth column, absolute abundance of the transcript in the library; and the fifth column, percentage abundance of the transcript in the library. In some cases, the normal library shows differential expression of the cDNA; in others, the induced or diseased library shows differential expression. For example, for SEQ ID NO:8, the endothelial cells treated with growth factors (VEGF and EF) show 2-8×higher expression than untreated endothelial cells in the same experiment.

[0165] These data confirm the differential expression of SEQ ID NOs:3-6, 8,13, 15-20, and 22 under conditions that correlate with disorders associated with atherosclerosis. VII Homology Searching for Atherosclerosis-Associated cDNAs and Polypeptides The polynucleotide sequences, SEQ ID NO:1-25, and polypeptide sequence, SEQ ID NO:26, were queried against databases derived from sources such as GenBank and SwissProt. These databases, which contain previously identified and annotated sequences, were searched for regions of similarity using BLAST (Altschul, supra). BLAST searched for matches and reported only those that satisfied the probability thresholds of 10-25 or less for nucleotide sequences and 10-8 or less for polypeptide sequences.

[0166] The polypeptide sequence was also analyzed for known motif patterns using MOTIFS, SPSCAN, BLIMPS, and HMM-based protocols. MOTIFS (Genetics Computer Group, Madison Wis.) searches polypeptide sequences for patterns that match those defined in the Prosite Dictionary of Protein Sites and Patterns (Bairoch, supra) and displays the patterns found and their corresponding literature abstracts. SPSCAN (Genetics Computer Group) searches for potential signal peptide sequences using a weighted matrix method (Nielsen et al. (1997) Protein Engineering 10:1-6). Hits with a score of 5 or greater were considered. BLIMPS uses a weighted matrix analysis algorithm to search for sequence similarity between the polypeptide sequences and those contained in BLOCKS, a database consisting of short amino acid segments, or blocks of 3-60 amino acids in length, compiled from the PROSITE database (Henikoff; supra; Bairoch, supra), and those in PRINTS, a protein fingerprint database based on non-redundant sequences obtained from sources such as SwissProt, GenBank, PIR, and NRL-3D (Attwood et al. (1997) J Chem Inf Comput Sci 37:417-424). For the purposes of the present invention, the BLIMPS searches reported matches with a cutoff score of 1000 or greater and a cutoff probability value of 1.0×10^{-3}. HMM-based protocols were based on a probabilistic approach and searched for consensus primary structures of gene families in the protein sequences (Eddy, supra; Sonnhammer, supra). More than 500 known protein families with cutoff scores ranging from 10 to 50 bits were selected for use in this invention.

[0167] VIII Hybridization Technologies: Selection of Sequences, Microarray Preparation and Use

[0168] SEQ ID NO:1-25 are represented among the template sequences in the LIFESEQ GOLD database (Incyte Genomics). Several of these sequences, specifically SEQ ID NOs:1, 2, 10, 12, 18, and 32-34 have been used on microarrays in experiments investigating differential gene expression in cardiovascular samples. Table 7 presents the results

of these experiments; results were significant if the log2 Cy/Cy5 ratio exceeded ±1.00 in either the normal or the induced or disease state.

[0169] Exemplary Experimental Materials and Protocols

[0170] Activation of the vascular endothelium is considered to be a central event in a wide range of both physiological and pathophysiological processes, such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation.

[0171] HAEC, human aortic endothelial cells, are primary cells derived from the endothelium of a human aorta. They have been used as an experimental model for investigating the role of the endothelium in human vascular biology in vitro. HAECs were grown to 85% confluency, split into two samples, one of which was then treated with growth factors, LDL, cytokines, O2, and the like for variable time periods.

[0172] ECV304, HUAEC, and HUVEC are endothelial cell lines derived from the endothelium of the human umbilical artery or vein. This cell model has been extensively used to study the functional biology of human endothelial cells. These cells were also grown to about 85% confluency, split into samples, one of which was then treated with growth factors, LDL, cytokines, O2, and the like for variable time periods.

[0173] The experimental treatments and time of exposure are shown by the tissue description in Table 7.

[0174] Exemplary Activators and Inducers

[0175] TNF-α is a pleiotropic cytokine that is known to play a central role in the mediation of inflammatory responses through activation of multiple signal transduction pathways. TNF-α is produced by activated lymphocytes, macrophages, and other white blood cells, and is known to activate endothelial cells. Monitoring the endothelial cells' response to TNF-α at the level of the mRNA expression can provide information necessary for better understanding of both TNF-α signaling pathways and endothelial cell biology.

[0176] PMA is a broad activator of the protein kinase C-dependent pathways. lonomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium concentration. The combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In T cells, the combination of PMA and ionomycin mimics the type of secondary signaling events elicited during optimal B cell activation.

[0177] Microarrays

[0178] The HUMAN GENOME GEM series 1-5 microarrays (Incyte Genomics) contain 45,320 array elements which represent 22,632 annotated clusters and 22,688 unannotated clusters. For the UNIGEM series microarrays (Incyte Genomics), Incyte clones were mapped to non-redundant Unigene clusters (Unigene database (build 46), NCBI; Shuler (1997) J Mol Med 75:694-698), and the 5' clone with the strongest BLAST alignment (at least 90% identity and 100 bp overlap) was chosen, verified, and used in the construction of the microarray. The UNIGEM V 2.0 microarray (Incyte Genomics) contains 8,502 array elements which represent 8,372 annotated genes and 130 unannotated clusters.

[0179] To construct microarrays, cDNAs were amplified from bacterial cells using primers complementary to vector sequences flanking the cDNA insert. Thirty cycles of PCR increased the initial quantity of cDNAs from 1-2 ng to a final quantity greater than 5 μg. Amplified cDNAs were then purified using SEPHACRYL-400 columns (APB). Purified cDNAs were immobilized on polymer-coated glass slides. Glass microscope slides (Corning, Corning N.Y.) were cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides were etched in 4% hydrofluoric acid (VWR Scientific Products, West Chester Pa.), washed thoroughly in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich) in 95% ethanol. Coated slides were cured in a 110° C. oven. cDNAs were applied to the coated glass substrate using a procedure described in U.S. Pat. No. 5,807,522. One microliter of the cDNA at an average concentration of 100 ng/μl was loaded into the open capillary printing element by a high-speed robotic apparatus which then deposited about 5 nl of cDNA per slide.

[0180] Microarrays were UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene), and then washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites were blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (Tropix, Bedford Mass.) for 30 minutes at 60° C. followed by washes in 0.2% SDS and distilled water as before.

[0181] Isolation and Labeling of Sample cDNAs

[0182] Cells were harvested and lysed in 1 ml of TRIZOL reagent (5×10^6 cells/ml; Invitrogen). The lysates were vortexed thoroughly and incubated at room temperature for 2-3 minutes and extracted with 0.5 ml chloroform. The extract was mixed, incubated at room temperature for 5 minutes, and centrifuged at 16,000×g for 15 minutes at 4° C. The aqueous layer was collected, and an equal volume of isopropanol was added. Samples were mixed, incubated at room temperature for 10 minutes, and centrifuged at 16,000×g for 20 minutes at 4° C. The supernatant was removed, and the RNA pellet was washed with 1 ml of 70% ethanol, centrifuged at 16,000×g at 4° C., and resuspended in RNAse-free water. The concentration of the RNA was determined by measuring the optical density at 260 nm.

[0183] Poly(A) RNA was prepared using an OLIGOTEX mRNA kit (Qiagen) with the following modifications: OLIGOTEX beads were washed in tubes rather than spin columns, resuspended in elution buffer, and then loaded onto spin columns to recover the mRNA. To obtain maximum yield, the mRNA was eluted twice.

[0184] Each poly(A) RNA sample was reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-d(T) primer (21mer), 1×first strand buffer, 0.03 units/ul RNAse inhibitor, 500 uM dATP, 500 uM dGTP, 500 uM dTTP, 40 uM dCTP, and 40 uM either dCTP-Cy3 or dCTP-Cy5 (APB). The reverse transcription reaction was performed in a 25 ml volume containing 200 ng poly(A) RNA using the GEMBRIGHT kit (Incyte Genomics). Specific control poly(A) RNAs (YCFRO6, YCFR45, YCFR67, YCFR85, YCFR43, YCFR22, YCFR23, YCFR25, YCFR44, YCFR26) were synthesized by in vitro transcription from non-coding yeast genomic DNA (W. Lei, unpublished). As quantitative controls, control mRNAs (YCFRO6, YCFR45,

YCFR67, and YCFR85) at 0.002 ng, 0.02 ng, 0.2 ng, and 2 ng were diluted into reverse transcription reaction at ratios of 1:100,000, 1:10,000, 1:1000, 1:100 (w/w) to sample mRNA, respectively. To sample differential expression patterns, control mRNAs (YCFR43, YCFR22, YCFR23, YCFR25, YCFR44, YCFR26) were diluted into reverse transcription reaction at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, 25:1 (w/w) to sample mRNA. Reactions were incubated at 37° C. for 2 hr, treated with 2.5 ml of 0.5M sodium hydroxide, and incubated for 20 minutes at 85° C. to the stop the reaction and degrade the RNA.

[0185] cDNAs were purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech). Cy3- and Cy5-labeled reaction samples were combined as described below and ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The cDNAs were then dried to completion using a SpeedVAC system (Savant Instruments, Holbrook N.Y.) and resuspended in 14 μl 5×SSC, 0.2% SDS.

[0186] Hybridization and Detection

[0187] Hybridization reactions contained 9 μl of sample mixture containing 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5×SSC, 0.2% SDS hybridization buffer. The mixture was heated to 65° C. for 5 minutes and was aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The microarrays were transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber was kept at 100% humidity internally by the addition of 140 μl of 5×SSC in a corner of the chamber. The chamber containing the microarrays was incubated for about 6.5 hours at 60° C. The microarrays were washed for 10 min at 45° C. in low stringency wash buffer (1×SSC, 0.1% SDS), three times for 10 minutes each at 45° C. in high stringency wash buffer (0.1×SSC), and dried.

[0188] Reporter-labeled hybridization complexes were detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Santa Clara Calif.) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light was focused on the microarray using a 20×microscope objective (Nikon, Melville N.Y.). The slide containing the microarray was placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm×1.8 cm microarray used in the present example was scanned with a resolution of 20 micrometers.

[0189] In two separate scans, the mixed gas multiline laser excited the two fluorophores sequentially. Emitted light was split, based on wavelength, into two photomultiplier tube detectors (PMT R1477; Hamamatsu Photonics Systems, Bridgewater N.J.) corresponding to the two fluorophores. Appropriate filters positioned between the microarray and the photomultiplier tubes were used to filter the signals. The emission maxima of the fluorophores used were 565 nm for Cy3 and 650 nm for Cy5. Each microarray was typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus was capable of recording the spectra from both fluorophores simultaneously.

[0190] The sensitivity of the scans was calibrated using the signal intensity generated by a cDNA control species. Samples of the calibrating cDNA were separately labeled with the two fluorophores and identical amounts of each were added to the hybridization mixture. A specific location on the microarray contained a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000.

[0191] The output of the photomultiplier tube was digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Norwood, Mass.) installed in an IBM-compatible PC computer. The digitized data were displayed as an image where the signal intensity was mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data was also analyzed quantitatively. Where two different fluorophores were excited and measured simultaneously, the data were first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

[0192] A grid was superimposed over the fluorescence signal image such that the signal from each spot was centered in each element of the grid. The fluorescence signal within each element was then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis was the GEMTOOLS gene expression analysis program (Incyte Genomics). Significance was defined as signal to background ratio exceeding 2×and area hybridization exceeding 40%.

[0193] IX Further Characterization of Differentially Expressed cDNAs and Proteins

[0194] Clones were aligned against the LIFESEQ Gold 5.1 database (Incyte Genomics) and an Incyte template and its sequence variants were chosen for each clone. The template and variant sequences were aligned against the GenBank nucleotide sequence databases using BLASTn (vers. 2.0, NCBI) to acquire annotation. The template and variant sequences were translated into amino acid sequences which were aligned against GenPept and other protein databases using BLASTp (vers. 2.0, NCBI) to acquire annotation and characterization, i.e., structural motifs. Table 3 shows the GenBank annotations (where available) for SEQ ID NOs:1-25 of this invention as produced by BLAST analysis.

[0195] Percent sequence identity can be determined electronically for two or more amino acid or nucleic acid sequences using the MEGALIGN program, a component of LASERGENE software (DNASTAR). The percent identity between two amino acid sequences is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage identity.

[0196] Sequences with conserved protein motifs may be searched using the BLOCKS search program. This program analyses sequence information contained in the Swiss-Prot and PROSITE databases and is useful for determining the classification of uncharacterized proteins translated from genomic or cDNA sequences (Bairoch, supra; Attwood, supra). PROSITE database is a useful source for identifying

functional or structural domains that are not detected using motifs due to extreme sequence divergence. Using weight matrices, these domains are calibrated against the SWISS-PROT database to obtain a measure of the chance distribution of the matches.

[0197] The PRINTS database can be searched using the BLIMPS search program to obtain protein family "fingerprints". The PRINTS database complements the PROSITE database by exploiting groups of conserved motifs within sequence alignments to build characteristic signatures of different protein families. For both BLOCKS and PRINTS analyses, the cutoff scores for local similarity were:

[0198] >1300=strong, 1000-1300=suggestive; for global similarity were: p<exp−3; and for strength (degree of correlation) were: >1300=strong, 1000-1300=weak.

[0199] X Other Hybridization Technologies and Analyses

[0200] Other hybridization technologies utilize a variety of substrates such as nylon membranes, capillary tubes, etc. Arranging cDNAs on polymer coated slides is described in Example V; sample cDNA preparation and hybridization and analysis using polymer coated slides is described in examples VI and VII, respectively.

[0201] The cDNAs are applied to a membrane substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer. Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on LB agar containing selective agent (carbenicillin, kanamycin, ampicillin, or chloramphenicol depending on the vector used) and incubated at 37° C. for 16 hr. The membrane is removed from the agar and consecutively placed colony side up in 10% SDS, denaturing solution (1.5 M NaCl, 0.5 M NaOH), neutralizing solution (1.5 M NaCl, 1 M Tris, pH 8.0), and twice in 2×SSC for 10 min each. The membrane is then UV irradiated in a STRATALINKER UV-crosslinker (Stratagene).

[0202] In the second method, cDNAs are amplified from bacterial vectors by thirty cycles of PCR using primers complementary to vector sequences flanking the insert. PCR amplification increases a starting concentration of 1-2 ng nucleic acid to a final quantity greater than 5 μg. Amplified nucleic acids from about 400 bp to about 5000 bp in length are purified using SEPHACRYL-400 beads (APB). Purified nucleic acids are arranged on a nylon membrane manually or using a dot/slot blotting manifold and suction device and are immobilized by denaturation, neutralization, and UV irradiation as described above.

[0203] Hybridization probes derived from cDNAs of the Sequence Listing are employed for screening cDNAs, mRNAs, or genomic DNA in membrane-based hybridizations. Probes are prepared by diluting the cDNAs to a concentration of 40-50 ng in 45 IL TE buffer, denaturing by heating to 100° C. for five min and briefly centrifuging. The denatured cDNA is then added to a REDIPRIME tube (APB), gently mixed until blue color is evenly distributed, and briefly centrifuged. Five microliters of [^{32}P]dCTP is added to the tube, and the contents are incubated at 37° C. for 10 min. The labeling reaction is stopped by adding 5 μl of 0.2M EDTA, and probe is purified from unincorporated nucleotides using a PROBEQUANT G-50 microcolumn (APB). The purified probe is heated to 100° C. for five min and then snap cooled for two min on ice.

[0204] Membranes are pre-hybridized in hybridization solution containing 1% Sarkosyl and Ix high phosphate buffer (0.5 M NaCl, 0.1 M Na_2HPO_4, 5 mM EDTA, pH 7) at 55° C. for two hr. The probe, diluted in 15 ml fresh hybridization solution, is then added to the membrane. The membrane is hybridized with the probe at 55° C. for 16 hr. Following hybridization, the membrane is washed for 15 min at 25° C. in imM Tris (pH 8.0), 1% Sarkosyl, and four times for 15 min each at 25° C. in 1 mM Tris (pH 8.0). To detect hybridization complexes, XOMAT-AR film (Eastman Kodak, Rochester N.Y.) is exposed to the membrane overnight at −70° C., developed, and examined.

[0205] XI Expression of the Encoded Protein

[0206] Expression and purification of a protein encoded by a cDNA of the invention is achieved using bacterial or virus-based expression systems. For expression in bacteria, cDNA is subcloned into a vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the trp-lac (tac) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the lac operator regulatory element. Recombinant vectors are transformed into bacterial hosts, such as BL21(DE3). Antibiotic resistant bacteria express the protein upon induction with IPTG. Expression in eukaryotic cells is achieved by infecting *Spodoptera frugiperda* (Sf9) insect cells with recombinant baculovirus, *Autographica californica* nuclear polyhedrosis virus. The polyhedrin gene of baculovirus is replaced with the cDNA by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of transcription.

[0207] For ease of purification, the protein is synthesized as a fusion protein with glutathione-S-transferase (GST; APB) or a similar alternative such as FLAG. The fusion protein is purified on immobilized glutathione under conditions that maintain protein activity and antigenicity. After purification, the GST moiety is proteolytically cleaved from the protein with thrombin. A fusion protein with FLAG, an 8-amino acid peptide, is purified using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak, Rochester N.Y.).

[0208] XII Production of Antibodies

[0209] A denatured protein from a reverse phase HPLC separation is obtained in quantities up to 75 mg. This denatured protein is used to immunize mice or rabbits following standard protocols. About 100 μg is used to immunize a mouse, while up to 1 mg is used to immunize a rabbit. The denatured protein is radioiodinated and incubated with murine B-cell hybridomas to screen for monoclonal antibodies. About 20 mg of protein is sufficient for labeling and screening several thousand clones.

[0210] In another approach, the amino acid sequence translated from a cDNA of the invention is analyzed using PROTEAN software (DNASTAR) to determine antigenic

determinants of the protein. The optimal sequences for immunization are usually at the C-terminus, the N-terminus, and those intervening, hydrophilic regions of the protein that are likely to be exposed to the external environment when the protein is in its natural conformation. Typically, oligopeptides about 15 residues in length are synthesized using an 431 peptide synthesizer (ABI) using Fmoc-chemistry and then coupled to keyhole limpet hemocyanin (KLH; Sigma-Aldrich) by reaction with M-maleimidobenzoyl-N-hydroxysuccinimide ester. If necessary, a cysteine may be introduced at the N-terminus of the peptide to permit coupling to KLH. Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated goat anti-rabbit IgG.

[0211] Hybridomas are prepared and screened using standard techniques. Hybridomas of interest are detected by screening with radioiodinated protein to identify those fusions producing a monoclonal antibody specific for the protein. In a typical protocol, wells of 96 well plates (FAST, Becton-Dickinson, Palo Alto Calif.) are coated with affinity-purified, specific rabbit-anti-mouse (or suitable anti-species Ig) antibodies at 10 mg/ml. The coated wells are blocked with 1% BSA and washed and exposed to supernatants from hybridomas. After incubation, the wells are exposed to radiolabeled protein at 1 mg/ml. Clones producing antibodies bind a quantity of labeled protein that is detectable above background.

[0212] Such clones are expanded and subjected to 2 cycles of cloning at 1 cell/3 wells. Cloned hybridomas are injected into pristane-treated mice to produce ascites, and monoclonal antibody is purified from the ascitic fluid by affinity chromatography on protein A (APB). Monoclonal antibodies with affinities of at least 10^8 M^-, preferably 10^9 to 1010 M^{-1} or stronger, are made by procedures well known in the art.

[0213] XIII Purification of Naturally Occurring Protein Using Antibodies

[0214] Naturally occurring or recombinant protein is purified by immunoaffinity chromatography using antibodies specific for the protein. An immunoaffinity column is constructed by covalently coupling the antibody to CNBr-activated SEPHAROSE resin (APB). Media containing the protein is passed over the immunoaffinity column, and the column is washed using high ionic strength buffers in the presence of detergent to allow preferential absorbance of the protein. After coupling, the protein is eluted from the column using a buffer of pH 2-3 or a high concentration of urea or thiocyanate ion to disrupt antibody/protein binding, and the protein is collected.

[0215] XIV Screening for Molecules That Specifically Bind the cDNA or Protein

[0216] The cDNA or fragments thereof and the protein or portions thereof are labeled with ^{32}P-dCTP, Cy3-dCTP, Cy5-dCTP (APB), or BIODIPY or FITC (Molecular Probes), respectively. Candidate molecules or compounds previously arranged on a substrate are incubated in the presence of labeled nucleic or amino acid. After incubation under conditions for either a cDNA or a protein, the substrate is washed, and any position on the substrate retaining label, which indicates specific binding or complex formation, is assayed. The binding molecule is identified by its arrayed position on the substrate. Data obtained using different concentrations of the nucleic acid or protein are used to calculate affinity between the labeled nucleic acid or protein and the bound molecule. High throughput screening using very small assay volumes and very small amounts of test compound is fully described in U.S. Pat. No. 5,876,946.

[0217] All patents and publications mentioned in the specification are incorporated herein by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

TABLE 4

Co-expression of 25 cDNAs with known atherosclerosis genes (- log p).

SEQ ID	CNN1	COL1A1	COL1A2	COL6A1	COL5A2	COL6A2	COL6A3	COL3A1	COL5A1	MGP	CTSK	FGB	VWF
1	0	1	0	0	0	0	0	0	0	0	0	4	1
2	1	1	3	2	2	2	4	3	4	5	3	1	6
3	0	1	0	1	1	1	0	0	1	1	1	4	0
4	1	0	1	1	1	0	1	1	1	0	1	7	0
5	5	9	12	13	3	14	11	12	4	10	13	1	3
6	6	4	7	9	4	13	5	6	8	14	6	1	12
7	5	4	5	5	2	7	5	6	5	10	1	2	1
8	1	1	0	2	1	1	0	1	0	1	0	1	1
9	1	1	4	2	1	6	4	4	5	6	1	0	10
10	3	4	3	3	3	4	4	3	1	4	1	0	5
11	1	1	1	1	1	1	1	0	1	1	2	1	1
12	11	4	7	9	2	9	11	10	7	11	3	0	5
13	2	6	5	6	9	8	4	6	7	1	3	1	1
14	0	0	0	0	0	0	0	1	0	1	1	4	1
15	17	7	13	17	2	25	12	11	6	21	6	3	7
16	13	22	27	27	15	33	24	34	15	20	18	0	8
17	2	1	3	3	0	2	6	3	1	9	1	0	7

TABLE 4-continued

Co-expression of 25 cDNAs with known atherosclerosis genes (- log p).

18	6	5	6	7	3	7	10	6	2	14	5	0	5
19	0	1	1	2	1	1	0	1	0	0	1	7	0
20	8	7	8	12	3	9	11	10	3	11	6	1	8
21	0	2	4	2	4	2	5	5	3	2	1	0	4
22	1	1	0	1	3	1	2	1	1	1	1	4	1
23	1	1	0	1	0	1	0	1	0	0	0	7	0
24	0	1	2	0	0	0	0	1	0	1	0	8	1
25	48	11	17	24	3	26	18	17	13	20	6	1	8

SEQ ID	PECAM1	AT3	LPL	A2M	APOA1	APOA2	APOB	APOC2	APOC3	APOC4	MSR1	CD36	SAP	CEL
1	0	5	0	1	3	4	2	2	5	4	0	0	6	0
2	7	0	1	4	0	1	0	0	0	0	2	3	0	1
3	0	8	1	1	5	4	3	3	5	7	0	0	8	1
4	0	4	1	2	4	5	4	4	6	2	0	1	3	0
5	2	0	4	9	1	0	0	1	0	0	4	2	0	2
6	6	1	5	7	0	1	1	0	1	0	7	3	0	2
7	4	3	3	8	1	1	2	0	2	1	2	7	2	0
8	0	0	3	3	0	0	0	2	0	0	1	1	1	0
9	8	0	5	6	1	1	1	2	0	0	5	6	0	1
10	3	0	4	4	0	0	0	0	0	0	9	5	0	0
11	1	2	1	0	3	1	2	2	1	0	0	2	3	6
12	2	1	3	6	2	0	1	0	0	0	4	3	0	3
13	0	0	1	1	3	2	1	2	1	0	1	2	1	3
14	0	5	1	1	4	5	5	5	3	1	0	0	2	1
15	5	1	1	13	1	1	1	0	1	0	10	1	1	0
16	7	0	3	12	1	0	0	1	1	0	9	5	0	1
17	5	0	6	7	0	0	0	1	0	0	5	9	0	1
18	4	0	6	10	1	0	1	0	0	0	8	5	1	1
19	1	8	1	3	5	8	4	6	8	9	1	1	11	0
20	2	0	12	10	1	1	0	1	0	0	9	14	0	0
21	4	0	8	5	0	1	0	1	1	0	5	11	1	1
22	4	9	0	2	6	8	7	6	4	3	1	1	7	0
23	0	6	0	1	6	7	7	4	8	1	0	1	7	1
24	0	7	0	2	8	10	10	7	8	4	0	0	5	1
25	2	0	2	20	1	0	3	0	0	0	10	4	0	1

SEQ ID	PON1	PON2	PON3	PLIN	PTGDS	ANX2	ANX1	SPARC	SM22
1	5	1	2	0	0	2	0	1	0
2	1	6	1	1	2	3	4	8	3
3	5	2	2	0	0	1	0	1	0
4	2	0	1	0	0	1	0	1	1
5	1	1	1	6	4	6	3	9	9
6	1	4	0	11	7	6	4	11	14
7	1	4	5	2	1	11	7	7	8
8	1	5	0	1	23	1	1	4	2
9	0	7	1	6	3	3	2	7	4
10	0	1	0	15	2	2	1	2	4
11	6	2	7	4	0	0	2	1	0
12	1	3	1	3	3	6	3	7	9
13	2	3	2	2	1	1	3	4	2
14	2	2	1	1	0	0	0	1	0
15	0	2	1	1	12	7	9	14	27
16	1	2	2	7	6	12	9	22	20
17	0	1	0	6	2	3	1	6	3
18	1	2	3	8	6	3	10	6	11
19	5	1	2	0	1	0	1	0	1
20	0	0	0	19	4	4	2	7	10
21	0	1	0	8	0	2	2	3	3
22	4	2	2	1	2	1	3	1	1
23	4	1	3	1	0	0	0	1	0
24	3	1	1	0	2	1	1	1	0
25	1	1	1	3	11	9	6	9	30

[0218]

TABLE 5

Summary of coexpressed genes and their function in atherosclerosis-associated disorders

SEQ ID	Pvalue	Gene 1	Gene function from Table 3	Pvalue	Gene 2	Gene function from Table 3
1	6	SAP	associated with amyloid P in lesions	5	APOC3	associated with plasma triglyceride and hyperlipidemia
2	8	SPARC	calcification of plaques	7	PECAM-	implicated in migration of cells under O2 stress
3	8	AT3	deficiency causes recurrent venous thrombosis	8	SAP	associated with amyloid P in lesions
4	7	FGB	fibrin deposition in plaque	6	APOC3	apolipoprotein CIII
5	14	COL6A2	promoting platelet aggregation	12	CTSK	present in advanced plaques
6	14	SM22	downregulated during atherogenesis	14	MGP	calcificaton associated with high expression
7	11	ANX2	role in atherogenesis	10	MGP	calcificaton associated with high expression
8	23	PTDGS	accumulates in end-stage atherosclerotic plaques	15	AGT	associated with higher total cholesterol levels
9	10	VWF	Increased levels found in atherosclerosis	8	PECAM-	implicated in migration of cells under O2 stress
10	15	PLIN	functions in lipid metabolism	9	MSR1	deposits cholesterol during atherogenesis
11	7	PON3	associated with coronary heart disease and cholesterol levels	6	CEL	implicated in the progression of atherosclerosis
12	11	COL6A3	promoting platelet aggregation	11	MGP	calcificaton associated with high expression
13	9	COL5A2	promoting platelet aggregation	8	COL6A2	promoting platelet aggregation
14	5	APOA/B	accumulation of plasma LDL in atherogenesis	5	AT3	deficiency causes recurrent venous thrombosis
15	27	SM22	downregulated during atherogenesis	25	COL6A2	promoting platelet aggregation
16	34	COL3A1	promoting platelet aggregation	22	SPARC	calcification of plaques
17	9	MGP	calcificaton associated with high expression	9	CD36	receptor for oxidised LDL
18	11	SM22	downregulated during atherogenesis	10	A2M	retains LDL in core of plaque
19	11	SAP	associated with amyloid P in lesions	9	APOC4	alters lipid metabolism toward hypertriglyceridaemia
20	19	PLIN	functions in lipid metabolism	14	CD36	receptor for oxidised LDL
21	11	CD36	receptor for oxidised LDL	8	PLIN	functions in lipid metabolism
22	9	AT3	deficiency causes recurrent venous thrombosis	8	APOA2	associated with increased atherosclerotic lesions
23	8	APOC3	associated with plasma triglyceride and hyperlipidemia	7	SAP	associated with amyloid P in lesions
24	10	APOA/B	accumulation of plasma LDL in atherogenesis	8	FGB	fibrin deposition in plaque
25	48	CNN1	modulation of SMC, a feature of atherosclerosis	26	COL6A2	promoting platelet aggregation

[0219]

TABLE 6

Cardiovascular transcript images for the coexpressed cDNAs

SEQ ID	Library	cDNAs	Description of Sample	Abundance	% Abundance
3	SMCRUNT01	3472	renal vein, smooth muscle cells, 57M, Untx	1	0.0288
	ENDVUNT01	5215	microvascular, dermal, endothelial cells, 22F, Untx	1	0.0192
	HEAANOT01	12578	heart, coronary artery, CAD, 46M	1	0.0080
4	MONOTXN05	2709	periph blood, monocytes, 42F, t/IL-10, LPS, NORM	1	0.0369
	MONOTXT02	3554	periph blood, monocytes, 42F, t/IL-10, LPS	1	0.0281
	MCLRUNT01	6149	periph blood mononuclear cells, 60M, untreated	1	0.0163
5	ARTANOT06	6311	aorta, adventitia, 48M	2	0.0317
	HEAONOT02	3482	heart, aorta, 10M	1	0.0287
	HEAONOT04	4002	heart, aorta, 12F	1	0.0250
	ARTANOT07	5716	aorta, adventitia, 65F	1	0.0175
6	HEAONOE01	3639	heart, aorta, 39M, 5RP	10	0.2748
	HEAONOT03	3720	heart, aorta, aw/cerebral agenesis, 27F	3	0.0806
	HEAPNOT01	3502	heart, coronary artery, plaque, pool	2	0.0571

TABLE 6-continued

Cardiovascular transcript images for the coexpressed cDNAs

SEQ ID	Library	cDNAs	Description of Sample	Abundance	% Abundance
8	ENDVTXT01	1882	microvascular, dermal, endothelial cells, 22F, t/bFGF, EF	3	0.1594
	ENDVTXT02	1876	microvascular, dermal, endothelial cells, 22F, t/VEGE, EF	1	0.0533
	ENDVUNT01	5215	microvascular, dermal, endothelial cells, 22F, Untx	1	0.0192
13	SMCCNOS01	3494	coronary artery, smooth muscle cells, 3M, t/TNF, IL-1, SUB	1	0.0286
	SMCANOT01	7327	aortic smooth muscle line, M	1	0.0136
15	SMCRTXT01	3453	renal vein, smooth muscle cells, 57M t/TNF, IL1	2	0.0579
	SMCRUNT01	3472	renal vein, smooth muscle cells, 57M, Untx	1	0.0288
16	SMCCNOT01	4266	coronary artery, smooth muscle cells, 3M	5	0.1172
	SMCCNOT02	3980	coronary artery, smooth muscle cells, 3M, t/TNF, IL-1	2	0.0503
	SMCCNOS01	3494	coronary artery, smooth muscle cells, 3M, t/TNF, IL-1, SUB	1	0.0286
17	ENDIUNT01	3582	iliac artery, endothelial cells, F, control, untreated	26	0.7259
	ENDITXT01	3464	iliac artery, endothelial cells, F, t/1% oxygen 24 hr	21	0.6062
	ENDINOT02	3208	iliac artery, endothelial cells, F, t/TNF, IL-1 20 hr	10	0.3117
18	ENDVTXT01	1882	microvascular, dermal, endothelial cells, 22F, t/bFGF, EF	1	0.0531
	ENDVNOT01	4955	microvascular, dermal, endothelial cells, 18F, untreated	1	0.0202
19	ENDATXP01	1877	aorta, endothelial cells, t/TNF, TIGR	1	0.0533
	ENDVTXT02	1876	microvascular, dermal, endothelial cells, 22F, t/VEGF, EF	1	0.0533
	ENDVTXT01	1882	microvascular, dermal, endothelial cells, 22F, t/bFGF, EF	1	0.0531
	ENDVNOT01	4955	microvascular, dermal, endothelial cells, 18F, untreated	1	0.0202
20	HEAANOT01	12578	heart, coronary artery, CAD, 46M	13	0.1034
	HEAONOT05	3959	heart, aorta, 17F	3	0.0758
	HEAONOT04	4002	heart, aorta, 12F	1	0.0250
22	ENDITXT01	3464	iliac artery, endothelial cells, F, t/1% oxygen 24 hr	1	0.0289
	HEAONOE01	3639	heart, aorta, 39M, 5RP	1	0.0275
	HEAANOT01	12578	heart, coronary artery, CAD, 46M	1	0.0080

[0220]

TABLE 7

Microarray data from cardiovascular experiments

SEQ ID	GEM	Log2 (Cy5/Cy3)	Cy3 Sample	Cy5 Sample
1	LG1	−1.01	ECV304 Line, Untx	ECV304 Line, t/PMA + Iono
2	HG1	−1.03	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 1 hr, Nrml
2	HG1	−1.03	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 24 hr, Nrml
2	HG1	−1.03	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/IL1b + TNFa 10 ng/mL, 10 ng/mL 2 4hr, 4 hr, Nrml
2	HG1	−1.07	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/PMA + TNFa 10 nM, 10 ng/mL 24 hr, 1 hr, Nrml
2	HG1	−1.07	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 48 hr, Nrml
2	HG1	−1.10	HUVEC Cells, Untx	HUVEC Cells, t/IL4 10 ng/mL 24 hr
2	HG1	−1.10	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 2 hr, Nrml
2	HG1	−1.12	HUVEC Cells, Untx, 48 hr	HUVEC Cells, t/IL4 + TNFa 10 ng/mL, 10 ng/mL 24 hr, 1 hr
2	HG1	−1.12	HUVEC Cells, Untx	HUVEC Cells, t/IL10 + TNFa 10 ng/mL, 10 ng/mL 24 hr,4 hr
2	HG1	−1.13	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 4 hr, Nrml
2	HG1	−1.14	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/CHX + TNFa 10 mcg/mL, 10 ng/mL 30 min, 24 hr, Nrml
2	HG1	−1.14	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/PMA 10 nM 24 hr, Nrml
2	HG1	−1.15	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 24 hr, Nrml
2	HG1	−1.17	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 24 hr, Nrml
2	HG1	−1.18	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/8ClcAMP + TNFa 7.5 microM, 10 ng/mL 24 hr, 4 hr, Nrml
2	HG1	−1.19	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 24 hr, Nrml
2	HG1	−1.21	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TL1b + TNFa 10 ng/mL, 10 ng/mL 2 4hr, 24 hr, Nrml

TABLE 7-continued

Microarray data from cardiovascular experiments

SEQ ID	GEM	Log2 (Cy5/Cy3)	Cy3 Sample	Cy5 Sample
2	HG1	−1.22	HIAEC Cells, Untx, Nrml	HIAEC Cells, t/TNFa 10 ng/mL 24 hr, Nrml
2	HG1	−1.22	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/TNFa + TNFa .1 ng/ml, 10 ng/ml 24 hr, 24 hr, Nrml
2	HG1	−1.23	HUVEC Cells, Untx	HUVEC Cells, t/TNFa 10 ng/mL 4 hr"
2	HG1	−1.26	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/8ClcAMP + TNFa 7.5 microM, 10 ng/mL 24 hr, 24 hr, Nrml
2	HG1	−1.28	HUVEC Cells, Untx	HUVEC Cells, t/TNFa .1 ng/mL 4 hr
2	HG1	−1.29	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 4 hr, Nrml
2	HG1	−1.29	HIAEC Cells, Untx, Nrml	HIAEC Cells, t/TNFa 10 ng/mL 4 hr, Nrml
2	HG1	−1.30	HUVEC Cells, Untx	HUVEC Cells, t/IL10 + TNFa 10 ng/mL, 10 ng/mL 24 hr, 24 hr
2	HG1	−1.31	HAEC Cells, Untx, Nrml	HAEC Cells, t/TNFa 10 ng/ml 10 hr, Nrml
2	HG1	−1.31	HAEC Cells, Untx, Nrml	HAEC Cells, t/TNFa 10 ng/ml 24 hr, Nrml
2	HG1	−1.34	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/IL4 + TNFa 10 ng/ml, 10 ng/ml 24 hr, 4 hr, Nrml
2	HG1	−1.36	HAEC Cells, Untx, Nrml	HAEC Cells, t/TNFa 10 ng/ml 4 hr, Nrml
2	HG1	−1.40	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/IL4 + TNFa 10 ng/ml, 10 ng/ml 24 hr, 24 hr, Nrml
2	HG1	−1.43	HUVEC Cells, Untx	HUVEC Cells, t/TNFa 10 ng/mL 24 hr
2	HG1	−1.44	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/IL1b 10 ng/mL 24 hr, Nrml
2	HG1	−1.48	HUVEC Cells, Untx	HUVEC Cells, t/TNFa 10 ng/mL 24 hr"
2	HG1	−1.52	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 8 hr, Nrml
2	HG1	−1.53	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 4 hr, Nrml
2	HG1	−1.53	HUVEC Cells, Untx	HUVEC Cells, t/CHX + TNFa 10 mcg/ml, 10 ng/ml 30 min, 23.5 hr
2	HG1	−1.53	HUVEC Cells, Untx	HUVEC Cells, t/TNFa 1 ng/mL 24 hr
2	HG1	−1.55	HAEC Cells, Untx, Nrml	HAEC Cells, t/TNFa 10 ng/ml 8 hr, Nrml
2	HG1	−1.58	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/PD98059 + TNFa 50 microM, 10 ng/mL 24 hr,24 hr, Nrml
2	HG1	−1.59	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 3 d, Nrml
2	HG1	−1.59	HAEC Cells, Untx, Nrml	HAEC Cells, t/TNFa 10 ng/ml 6 hr, Nrml
2	HG1	−1.61	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/IL1b + TNFa 10 ng/mL, 10 ng/mL 24 hr, 1 hr, Nrml
2	HG1	−1.62	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/TNFa 10 ng/ml 24 hr, Nrml
2	HG1	−1.72	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 8 hr, Nrml
2	HG1	−1.80	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/PD98059 + TNFa 50 microM, 10 ng/mL 24 hr, 4 hr, Nrml
2	HG1	−1.83	HIAEC Cells, Untx, Nrml	HIAEC Cells, t/TNFa 10 ng/mL 8 hr, Nrml
2	HG1	−2.07	HUVEC Cells, Untx, 48 hr	HUVEC Cells, t/IL4 + TNFa 10 ng/mL, 10 ng/mL 24 hr, 24 hr
2	HG1	−2.15	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 8 hr, Nrml
2	HG1	−2.18	HUVEC Cells, Untx, 48 hr	HUVEC Cells, t/IL4 + TNFa 10 ng/mL, 10 ng/mL 24 hr,4 hr
10	UG1	3.38	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/CHX + TNFa 10 mcg/mL, 10 ng/mL 30 min, 4 hr, Nrml
10	UG1	3.16	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/CHX + TNFa 10 mcg/mL, 10 ng/mL 30 min, 1 hr, Nrml
10	UG1	2.85	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 2 hr, Nrml
10	UG1	2.68	HPAEC Cells, Untx, Nrml	HPAEC Cells, t/TNFa 10 ng/mL 1 hr, Nrml
10	UG1	2.64	HPAEC Cells, Untx, Nrml	HPAEC Cells, t/TNFa 10 ng/mL 2 hr, Nrml

TABLE 7-continued

Microarray data from cardiovascular experiments				
SEQ ID	GEM	Log2 (Cy5/Cy3)	Cy3 Sample	Cy5 Sample
10	UG1	2.60	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/PMA + TNFa 100 nM, 10 ng/mL 24 hr, 1 hr, Nrml
10	UG1	2.55	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/PMA + TNFa 100 nM, 10 ng/mL 24 hr, 24 hr, Nrml
10	UG1	2.47	HUVEC Cells, Untx, 0 hr, Nrml	HUVEC Cells, t/Dex + TNFa 100 nM, 10 ng/mL 24 hr, 1 hr Nrml
10	UG1	2.40	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 48 hr, Nrml
10	UG1	2.39	HUVEC Cells, Unix, Nrml, 0 hr	HUVEC Cells, t/Dex + TNFa 100 nM, 10 ng/mL 24 hr, 24 hr Nrml
10	UG1	2.27	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/IL10 + TNFa 10 ng/mL, 10 ng/mL 24 hr, 24 hr, Nrml
10	UG1	2.24	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 4 hr, Nrml
10	UG1	2.18	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 1 hr, Nrml
10	UG1	2.17	HUVEC Cells, Untx, Nrml, 0 hr	HUVEC Cells, t/Dex + TNFa 10 nM, 10 ng/mL 24 hr, 24 hr Nrml
10	UG1	2.14	HUVEC Cells, Untx, 24 hr	HUVEC Cells, t/Lipom + TNFa 8 mcg/mL, 10 ng/ml 4 hr, 4 hr, 1 hr
10	UG1	2.11	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/ILlb + TNFa 10 ng/mL, 10 ng/mL 24 hr, 24 hr, Nrml
10	UG1	2.04	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/IFNg + TNFa 200 ng/mL, 10 ng/mL 24 hr, 24 hr, Nrml
10	UG1	1.95	HUVEC Cells, Untx, Nrml, 0 hr	HUVEC Cells, t/Dex + TNFa 10 nM, 10 ng/mL 24 hr, 1 hr Nrml
10	UG1	1.91	HUVEC Cells, Untx, 24 hr	HUVEC Cells, t/Lipom + asOligo + TNFa 8 mcg/mL, 100 nM, 50 nM, 10 ng/ml 4 hr, 4 hr, 1 hr
10	UG1	1.91	HUVEC Cells, Untx, Nrml, 24 hr	HUVEC Cells, t/PD98059 + TNFa 50 microM, 10 ng/mL 24 hr, 1 hr, Nrml
10	UG1	1.88	HUVEC Cells, Untx, Nrml, 0 hr	HUVEC Cells, t/Dex + TNFa 100 nM, 10 ng/mL 24 hr, 4 hr Nrml
10	UG1	1.88	HUVEC Cells, Untx, Nrml, 0 hr	HUVEC Cells, t/Dex + TNFa 10 nM, 10 ng/mL 24 hr, 4 hr Nrml
10	UG1	1.86	HUVEC Cells, Untx, 48 hr	HUVEC Cells, t/IL4 + TNFa 10 ng/mL, 10 ng/mL 24 hr, 24 hr
10	UG1	1.82	HUVEC Cells, Untx, 24 hr	HUVEC Cells, t/Lipom + TNFa 8 mcg/mL, 10 ng/ml 4 hr, 4 hr, 24 hr
10	UG1	1.82	HUVEC Cells, Untx, 48 hr	HUVEC Cells, t/IL4 + TNFa 10 ng/mL, 10 ng/mL 24 hr, 1 hr
10	UG1	1.76	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/IL1b 10 ng/mL 24 hr, Nrml
10	UG1	1.74	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/TNFa 10 ng/mL 24 hr, Nrml
10	HG3	1.20	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 1 hr, Nrml
10	HG3	1.04	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 2 hr, Nrml
12	HG5	1.02	ECV304 Line, t/TNFa 10 ng/mL 3 d, Nrml	ECV304 Line, Untx, Nrml
12	HG5	−1.07	HUVEC Cells, Untx	HUVEC Cells, t/IL10 + TNFa 10 ng/mL, 10 ng/mL 24 hr, 4 hr
12	HG5	−1.16	HUAEC Cells, Untx, Nrml	HUAEC Cells, t/TNFa 10 ng/mL 8 hr, Nrml
12	HG5	−1.18	HMVEC Cells, Untx, Nrml	HMVEC Cells, t/TNFa 10 ng/mL 24 hr, Nrml
12	HG5	−1.24	HMVEC Cells, Untx, Nrml	HMVEC Cells, t/TNFa 10 ng/mL 8 hr, Nrml
12	HG5	−1.50	HIAEC Cells, Untx, Nrml	HIAEC Cells, t/TNFa 10 ng/mL 8 hr, Nrml
18	HG2	−2.38	HUVEC Cells, Untx, Nrml	HUVEC Cells, t/IFNg + TNFa 10 ng/mL, 10 ng/mL 24 hr, 24 hr, Nrml
32	HG3	−1.13	HUVEC Cells, Untx 24 hr, Nrml	HUVEC Cells, t/PMA + TNFa 10 nM, 10 ng/mL 24 hr, 24 hr, Nrml
33	LG1	−1.17	Tangier	Fibroblast Cells, t/LDL Cholesterol, Nrml
34	UG1	−1.04	ECV304 Line, Untx	ECV304 Line, t/PMA + Iono 1 microM, 1 meg/ml 4 hr
34	UG1	−1.11	ECV304 Line, Untx	ECV304 Line, t/PMA + Iono 1 microM, 1 meg/ml 5 hr

[0221]

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 26

<210> SEQ ID NO 1
<211> LENGTH: 1334
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 674, 735, 788
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 1

```
aggcctccct ccacctgtct tctcagagca gataatggca agcatggctg ccgtgctcac     60
ctgggctctg gctcttcttt cagcgttttc ggccacccag gcacggaaag gcttctggga    120
ctacttcagc cagaccagcg gggacaaagg cagggtggag cagatccatc agcagaagat    180
ggctcgcgag cccgcgaccc tgaaagacag ccttgagcaa gacctcaaca atatgaacaa    240
gttcctggaa aagctgaggc ctctgagtgg gagcgaggct cctcggctcc cacaggaccc    300
ggtgggcatg cggcggcagc tgcaggagga gttggaggag gtgaaggctc gcctccagcc    360
ctacatggca gaggcgcacg agctggtggg ctggaatttg gagggcttgc ggcagcaact    420
gaagccctac acgatggatc tgatggagca ggtggccctg cgcgtgcagg agctgcagga    480
gcagttgcgc gtggtggggg aagacaccaa ggcccagttg ctgggggcgg tggacgaggc    540
ttgggctttg ctgcagggac tgcagagccg cgtggtgcac cacaccggcc gcttcaaaga    600
gctcttccac ccatacgccg agagcctggt gagcggcatc gggcgccacg tgcaggagct    660
gcaccgcagt gtgntccgca cgcccccgcc agccccgcgc gcctcagtcg ctgcgtgcag    720
gtgctctccc ggaantcacg ctcaaggcca aggccctgca cgcacgcatc cagcagaacc    780
tggaccantg cgcgaagagc tcagcagagc ctttgcaggc actgggactg aggaaggggc    840
cggcccggac cccagatgc tctccgagga ggtgcgccag cgacttcagg ctttccgcca    900
ggacacctac ctgcagatag ctgccttcac tcgcgccatc gaccaggaga ctgaggaggt    960
ccagcagcag ctggcgccac ctccaccagg ccacagtgcc ttcgccccag agtttcaaca   1020
aacagacagt ggcaaggttc tgagcaagct gcaggcccgt ctggatgacc tgtgggaaga   1080
catcactcac agccttcatg accagggcca cagccgtctg ggggacccct gaggatctac   1140
ctgcccaggc ccattcccag cttcttgtct ggggagcctt ggctctgagc ctctagcatg   1200
gttcagtcct tgaaagtggc ctgttgggtg gagggtggaa ggtcctgtgc aggacaggga   1260
ggccaccaaa ggggctgctg tctcctgcat atccagcctc ctgcgactcc ccaatgcagg   1320
atgcattcat tcac                                                     1334
```

<210> SEQ ID NO 2
<211> LENGTH: 1702
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 2

-continued

```
cgttcccact gcaccctgga gaacgagcct ttgcggggtt tctcctggct gtcctccgac      60
cccggcggtc tcgaaagcga cacgctgcag tgggtggagg agccccaacg ctcctgcacc     120
gcgcggagat gcgcggtact ccaggccacc ggtggggtcg agcccgcagg ctggaaggag     180
atgcgatgcc acctgcgcgc caacggctac ctgtgcaagt accagtttga ggtcttgtgt     240
cctgcgccgc gccccggggc cgcctctaac ttgagctatc gcgcgccctt ccagctgcac     300
agcgccgctc tggacttcag tccacctggg accgaggtga gtgcgctctg ccggggacag     360
ctcccgatct cagttacttg catcgcggac gaaatcggcg ctcgctggga caaactctcg     420
ggcgatgtgt tgtgtccctg ccccgggagg tacctccgtg ctggcaaatg cgcagagctc     480
cctaactgcc tagacgactt gggaggcttt gcctgcgaat gtgctacggg cttcgagctg     540
gggaaggacg gccgctcttg tgtgaccagt ggggaaggac agccgaccct tgggggggacc    600
ggggtgccca ccaggcgccc gccggccact gcaaccagcc ccgtgccgca gagaacatgg     660
ccaatcaggg tcgacgagaa gctgggagag acaccacttg tccctgaaca agacaattca     720
gtaacatcta ttcctgagat tcctcgatgg ggatcacaga gcacgatgtc taccctcaa      780
atgtcccttc aagccgagtc aaaggccact atcaccccat cagggagcgt gatttccaag     840
tttaattcta cgacttcctc tgccactcct caggctttcg actcctcctc tgccgtggtc     900
ttcatatttg tgagcacagc agtagtagtg ttggtgatct tgaccatgac agtactgggg     960
cttgtcaagc tctgctttca cgaaagcccc tcttcccagc caaggaagga gtctatgggc    1020
ccgccgggcc tggagagtga tcctgagccc gctgctttgg gctccagttc tgcacattgc    1080
acaaacaatg ggtgaaagt cggggactgt gatctgcggg acagagcaga gggtgccttg     1140
ctggcggagt cccctcttgg ctctagtgat gcatagggaa acaggggaca tgggcactcc    1200
tgtgaacagt tttcacttt tgatgaaacg gggaaccaag aggaacttac ttgtgtaact     1260
gacaatttct gcagaaatcc cccttcctct aaattccctt tactccactg aggagctaaa    1320
tcagaactgc acactccttc cctgatgata gaggaagtgg aagtgccttt aggatggtga    1380
tactggggga ccgggtagtg ctggggagag atattttctt atgtttattc ggagaatttg    1440
gagaagtgat tgaacttttc aagacattgg aaacaaatag aacacaatat aatttacatt    1500
aaaaaataat ttctaccaaa atggaaagga aatgttctat gttgttcagg ctaggagtat    1560
attggttcga aatcccaggg aaaaaaataa aaataaaaaa ttaaaggatt gttgataaaa    1620
aaaaaaaaaa aaaaagatct ttaattaagc ggcccaagct tattcccttt agtgagggtt    1680
aattttagct tgcactggcc ac                                             1702
```

```
<210> SEQ ID NO 3
<211> LENGTH: 586
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 48, 66, 560, 574, 577, 580
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 3

tcgaggactc cgccaactac agctgcgtct acgtggacct gaagccgnct ttcgggggct      60
acgcgnccag cgagcgcttg gagctgcacg tggacggacc ccctcccagg cctcagctcc     120
```

-continued

```
gggcgacgtg gagtggggcg gtcctggcgg gccgagatgc cgtcctgcgc tgcgagggac    180

ccatccccga cgtcaccttc gagctgctgc gcgagggcga gacgaaggcc gtgaagacgg    240

tccgcacccc cggggccgcg gcgaacctcg agctgatctt cgtggggccc cagcacgccg    300

gcaactacag gtgccgctac cgctcctggg tgccccacac cttcgaatcg gagctcagcg    360

accctgtgga gctcctggtg gcagaaagct gatgcagccg cgggcccagg gtgctgttgg    420

tgtcctcaga agtgccgggg attctggact ggctccctcc cctcctgttg cagcacaagg    480

ccggggtctc tggggggctg gagaagcctc cctcattcct cccaggaatt aataaatgtg    540

aagagagctc tgtttaaaan aaaaaaaaag aaanaanaan aaccaa                   586
```

```
<210> SEQ ID NO 4
<211> LENGTH: 433
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 4

ctcaagaccc agcagtggga cagccagaca gacggcacga tggcactgag ctcccagatc     60

tgggccgctt gcctcctgct cctcctcctc ctcgccagcc tgaccagtgg ctctgttttc    120

ccacaacaga cgggacaact tgcagagctg caaccccagg acagagctgg agccagggcc    180

agctggatgc ccatgttcca gaggcgaagg aggcgagaca cccacttccc catctgcatt    240

ttctgctgcg gctgctgtca tcgatcaaag tgtgggatgt gctgcaagac gtagaaccta    300

cctgccctgc ccccgtcccc tcccttcctt atttattcct gctgccccag aacataggtc    360

ttggaataaa atggctggtt cttttgtttt ccaaaaaaaa aaaaaaaaaa aaaaaaaaaa    420

aaaaaaaaaa aaa                                                        433
```

```
<210> SEQ ID NO 5
<211> LENGTH: 3111
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 44
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 5

cgagggcgga cgcaaagaac gcggaggacc tctgggtgcc tgcnggggag ctgctccagc     60

cgggccgccg ggagcggtgg ggagagcatc gcgcagccgc ccctccacgc gcccgcccag    120

ccgcgttcgc ccactgggct ctcccggctg cagtgccagg gcgcaggacg cggccgatct    180

cccgctcccg ccacctccgc caccatgctg ctcccccagc tctgctggct gccgctgctc    240

gctgggctgc tcccgccggt gcccgctcag aagttctcgg cgctcacgtt tttgagagtg    300

gatcaagata aagacaagga ttgtagcttg gactgtgcgg gttcgcccca gaaacctctc    360

tgcgcatctg acggaaggac cttcctttcc cgttgtgaat ttcaacgtgc caagtgcaaa    420

gatccccagc tagagattgc atatcgagga aactgcaaag acgtgtccag gtgtgtggcc    480

gaaaggaagt atacccagga gcaagcccgg aaggagtttc agcaagtgtt cattcctgag    540

tgcaatgacg acggcaccta cagtcaggtc cagtgtcaca gctacacggg atactgctgg    600
```

-continued

```
tgcgtcacgc ccaacgggag gcccatcagc ggcactgccg tggcccacaa gacgccccgg    660
tgcccgggtt ccgtaaatga aaagttaccc caacgcgaag gcacaggaaa aacagatgat    720
gccgcagctc cagcgttgga gactcagcct caaggagatg aagaagatat tgcatcacgt    780
taccctaccc tttggactga acaggttaaa agtcggcaga acaaaaccaa taagaattca    840
gtgtcatcct gtgaccaaga gcaccagtct gccctggagg aagccaagca gcccaagaac    900
gacaatgtgg tgatccctga gtgtgcgcac ggcggcctct acaagccagt gcagtgccac    960
ccctccacgg ggtactgctg gtgcgtcctg gtggacacgg ggcgccccat tcccggcaca   1020
tccacaaggt acgagcagcc gaaatgtgac aacacgggcc agggcccacc cagccaaagc   1080
ccgggacctg tacaagggcc gccagctaca aggttgtccg ggtgccaaaa agcatgagtt   1140
tctgaccagc gttctggacg cgctgtccac ggacatggtc cacgccgcct ccgacccctc   1200
ctcctcgtca ggcaggctct cagaacccga ccccagccat accctagagg agcgggtggt   1260
gcactggtac ttcaaactac tggataaaaa ctccagtgga gacatcggca aaaaggaaat   1320
caaacccttc aagaggttcc ttcgcaaaaa atcaaagccc aaaaaatgtg tgaagaagtt   1380
tgttgaatac tgtgacgtga ataatgacaa atccatctcc gtacaagaac tgatgggctg   1440
cctgggcgtg gcgaaagagg acggcaaagc ggacaccaag aaacgccaca cccccagagg   1500
tcatgctgaa agtacgtcta atagacagcc aaggaaacaa ggataaatgg ctcatacccc   1560
gaaggcagtt cctagacaca tgggaaattt ccctcaccaa agagcaatta agaaaacaaa   1620
aacagaaaca catagtattt gcactttgta ctttaaatgt aaattcactt tgtagaaatg   1680
agctatttaa acagactgtt ttaatctgtg aaaatggaga gctggcttca gaaaattaat   1740
cacataccaa tgtatgtgtc ctcttttgac cttggaaatc tgtatgtggt ggagaagtat   1800
ttgaatgcat ttaggcttaa tttcttcgcc ttccacatgt taacagtaga gctctatgca   1860
ctccggctgc aatcgtatgg ctttctctaa cccctgcagt cacttccaga tgcctgtgct   1920
tacagcattg tggaatcatg ttggaagctc cacatgtcca tggaagtttg tgatgtacgg   1980
ccgaccctac aggcagttaa catgcatggg ctggtttgtt tcttgggatt ttctgttagt   2040
ttgtcttgtt ttgctttcca gagatcttgc tcatacaatg aatcacgcaa ccactaaagc   2100
tatccagtta agtgcaggta gttcccctgg aggaaataat attttcaaac tgtcgttggt   2160
gtgatacttt ggctcaaagg atctttgctt ttccatttta agcttctgtt ttgagttttg   2220
ccctggggct tgaatgagtc ccagagagtc gttcggatgg tgggaggctg cctaggaggc   2280
agtaaatcca gttcacagtg cctgggaggg gcccatcctt ccaaaatgta aatccagttc   2340
gcggtgtgac cgagctgggc taacaggctt gtctgcctgg ttttcctacc tacacgtgga   2400
cattattctc ctgatcctcc tacctggttc caccccaggg ctaccggaag gtaaaatctt   2460
cacctgaacc aattatgagc agtctcctta ctgaaggtac agccggatac gtggtgcccc   2520
cggggctggt gttggcagcc ggggggaggt gcctgagggt ccccacggtt cctttctgct   2580
tttctgaatg catcaagggt acgagaactt gccaatggga aattcatccg agtggcactg   2640
gcagagaagg ataggagtgg aatgcccaca cagtgaccaa cagaactggt ctgcgtgcat   2700
aaccagctgc caccctcagg cctgggcccc agagctcagg gcacccagtg tcttaaggaa   2760
ccatttggag gacagtctga gagcaggaac ttcaagctgt gattctatct cggctcagac   2820
ttttggttgg aaaaagatct tcatggcccc aaatcccctg agacatgcct tgtagaatga   2880
```

-continued

```
ttttgtgatg ttgtgatgct tgtggagcat cgcgtaaggc ttcttgctta tttaaactgt   2940

gcaaggtaaa aatcaagcct ttggagccac agaaccagct caagtacatg ccaatgttgt   3000

ttaagaaaca gttatgatcc taaacttttt ggataatctt ttatatttct gacctttgaa   3060

tttaatcatt gttcttagat taaaataaaa tatgctattg aaactaaaaa a            3111


<210> SEQ ID NO 6
<211> LENGTH: 2311
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 474-544, 2288, 2295-2299
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 6

gccgctcgcc cacggactcc gacgtgtccc tcgactccga ggactccggg gctaagtctc     60

caggcatcct gggctacaat atctgtcccc gcgggtggaa tggcagcctt cggctcaagc    120

gtggcagcct ccccgccgag gcctcctgca ccacctagag ccccaccccc gaccccaccc    180

cgggagggca gagccagaag aaggctcatt agacctgggg gacccaaagg gtctggcctc    240

tttgggcagc cccagagatg aggggtcagc agaggagagc tctggggttg gggatgggtt    300

agggacgcaa gcttgagttc tagcccttgc tctcattcag ctgttgtgtg accctgggta    360

agacccttcc ttgtttgacc ctcagctttc ccatctgttt aatggtggct ttggccaagg    420

caatccacaa acgtcaaaat tccccttccc atcagtacac acaccgatgc acannnnnnn    480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    540

nnnntagtta gtgccttgga tgaggcgggg cagtgtgtat atggacccct ggacttgcta    600

ccttcagggt tccatactcg tccctcccct cctggctctg ctgtctggag tctggcaagc    660

ggggtgtgtt cagaaggtcc taggcctgtg tcgcatgtcc aggcactggc ctgaccatcc    720

ggctccctgg gcaccaagtc ccagggcagg agcagctgtt ttccatccct tcccagacaa    780

gctctatttt tatcacaatg acctttagag aggtctccca ggccagctca aggtgtccca    840

ctatcccctc tggagggaag aggcaggaaa attctccccg ggtccctgtc atgctacttt    900

ctccatccca gttcagactg tccaggacat cttatctgca gccataagag aattataagg    960

cagtgatttc ccttaggccc aggacttggg cctccagctc atctgttcct tctgggccca   1020

ttcatgggca ggttctgggc tcaaagctga actggggaga gaagagatac agagctacca   1080

tgtgacttta cctgattgcc ctcagtttgg ggttgcttat tgggaaagag agagacaaag   1140

agttacttgt tacgggaaat atgaaaagca tggccaggat gcatagagga gattctagca   1200

ggggacagga ttggctcaga tgacccctga gggctcttcc agtcttgaaa tgcattccat   1260

gatattagga agtcgggggt gggtggtggt ggtgggctag ttgggcttga atttaggggc   1320

cgatgagctt gggtacgtga gcagggtgtt aagttagggt ctgcctgtat ttctggtccc   1380

cttgggaaat gtccccttct tcagtgtcag acctcagtcc cagtgtccat atcgtgccca   1440

gaaaagtaga cattatcctg ccccatccct tccccagtgc actctgacct agctagtgcc   1500

tggtgcccag tgacctgggg gagcctggct gcaggccctc actggttccc taaaccttgg   1560

tggctgtgat tcaggtcccc agggggggact cagggaggaa tatggctgag ttctgtagtt   1620
```

-continued

```
tccagagttg ggctggtaga gctttctaga ggttcagaat attagcttca ggatcagctg   1680

ggggtatgga attggctgag gatcaaacgt atgtaggtga aaggatacca ggatgttgct   1740

aaaggtgagg gacagtttgg gtttgggact taccggggtg atgttagatc tggaaccccc   1800

aagtgaggct ggagggagtt aaggtcagta tggaagatag ggttgggaca gggtgctttg   1860

gaatgaaaga gtgaccttag agggctcctt gggcctcagg aatgctcctg ctgctgtgaa   1920

gatgagaagg tgctcttact cagttaatga tgagtgacta tatttaccaa agcccctacc   1980

tgctgctggg tcccttgtag cacaggagac tggggctaag ggcccctccc agggaaggga   2040

caccatcagg cctctggctg aggcagtagc atagaggatc catttctacc tgcatttccc   2100

agaggactag caggaggcag ccttgagaaa ccggcagttc ccaagccagc gcctggctgt   2160

tctctcattg tcactgccct ctccccaacc tctcctctaa cccactagag attgcctgtg   2220

tcctgcctct tgcctcttgt agaatgcagc tctggccctc aataaatgct tcctgcattc   2280

taaaaaanaa aaaannnnna aaaaaaaaag g                                  2311
```

```
<210> SEQ ID NO 7
<211> LENGTH: 1866
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 7

agcttttgtt cacactttaa atagcagtcc cagaatgatt tcactacaga ctctctggaa     60

agcctgggag ctgaattccg gaagatcccc acatcgatga aagcaaagcg aagccaccaa    120

gccatcatca tgtccacgtc gctacgagtc agcccatcca tccatggcta ccacttcgac    180

acagcctctc gtaagaaagc cgtgggcaac atctttgaaa acacagacca agaatcacta    240

gaaaggctct tcagaaactc tggagacaag aaagcagagg agagagccaa gatcattttt    300

gccatagatc aagatgtgga ggagaaaacg cgtgccctga tggccttgaa gaagaggaca    360

aaagacaagc ttttccagtt tctgaaactg cggaaatatt ccatcaaagt tcactgaaga    420

gaagaggatg gataaggacg ttatccaaga atggacattc aaagaccaag tgagtttgtg    480

agattctaac agatgcagca ttttgctgct accttacaag cttctcttct gtcaggactc    540

cagaggctgg aaagggaccg ggactggaaa gggaccagga ctgaacagac tggttacaaa    600

gactccaaac aatttcatgc cctgtgctgt tacagaggag aacaaaatgc tttcagcaag    660

gatttgaaaa ctcttccgtc cctgcaggaa aggattgatg ctgatagaag agcctggaca    720

gatgtaatga gaactaaaga aaacagatgg ctggagatga catttatcca gggtcacttt    780

gtcaggccct aggacttaaa tcgaagttga actttttttt ttttttaacc aaatagatag    840

gggaagggag gagggagagg gaggacaggg agagaaaata ccatgcataa attgtttact    900

gaatttttat atctgagtgt tcaaaatatt tccaagcctg agtattgtct attggtatag    960

atttttagaa atcaataatt gattatttat ttgcacttat tacaatgcct gaaaaagtgc   1020

accacatgga tgttaagtag aaattcaaga aagtaagatg tcttcagcaa ctcagtaaaa   1080

ccttacgcca ccttttggtt tgtaaaaggt tttttataca tttcaaacag gttgcacaaa   1140

agttaaaata atggggtctt ttataaatcc aaagtactgt gaaaacattt tacatatttt   1200

ttaaatcttc tgactaatgc taaaacgtaa tctaattaaa tttcatacag ttactgcagt   1260
```

-continued

```
aagcattagg aagtgaatat gatatacaaa atagtttata aagactctat agtttctata   1320

atttatttta ctggcaaatg tcatgcaaca ataataaatt attgtaaact ttgtggcttt   1380

tggtctgtga tgcttggtct caaaggaaaa aataagatgg taaatgttga tatttacaaa   1440

cttttctaaa gatgtgtctc taacaataaa agttaatttt agagtagttt tatattaatt   1500

accaaacttt ttcaaaacaa attcttacgt caaatatctg ggaagtttct ctgtcccaat   1560

cttaaaatat aaaatataga tatagaagtt catagattga ctccttggca tttctattta   1620

tgtatccatt aaggatgagt tttaaaaggc tttctcttca tacttttgaa aaatttcttc   1680

tatgattaca gtagctatgt acatgtgtac atctattttt cccaagcaat atgttttggg   1740

tttagagtct gagtgatgac caagattctg tgtgttacta ctgtttgttt aataggaaca   1800

aatatagaaa taatattatc tctttgctta tttcccgtta aaactataat aaaatgtttc   1860

taggaa                                                              1866
```

```
<210> SEQ ID NO 8
<211> LENGTH: 1929
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 8

gctgcctgcc ggtgctcttc gtggctctgg gcatggcctc ggaccccatc ttcacgctgg     60

cgcccccgct gcattgccac tacggggcct tccccccaa tgcctctggc tgggagcagc     120

ctcccaatgc cagcggcgtc agcgtcgcca gcgctgccct agcagccagc gccgccagcc    180

gtgtcgccac cagtaccgac ccctcgtgca gcggcttcgc cccgccggac ttcaaccatt    240

gccctcaagg attgggacta taatggcctt cctgtgctca ccaccaacgc catcggccag    300

tgggatctgg tgtgtgacct gggctggcag gtgatcctgg agcagatcct cttcatcttg    360

ggctttgcct ccggctacct gttcctgggt taccccgcag acagatttgg ccgtcgcggg    420

attgtgctgc tgaccttggg gctggtgggc ccctgtggag taggaggggc tgctgcaggc    480

tcctccacag gcgtcatggc cctccgattc ctcttgggct ttctgcttgc cggtgttgac    540

ctgggtgtct acctgatgcg cctggagctg tgcgacccaa cccagaggct tcgggtggcc    600

ctggcagggg agttggtggg ggtgggaggg cacttcctgt tcctgggcct ggcccttgtc    660

tctaaggatt ggcgattcct acagcgaatg atcaccgctc cctgcatcct cttcctgttt    720

tatggctggc ctggtttgtt cctggagtcc gcacggtggc tgatagtgaa gcggcagatt    780

gaggaggctc agtctgtgct gaggatcctg gctgagcgaa accggcccca tgggcagatg    840

ctgggggagg aggcccagga ggccctgcag gacctggaga atacctgccc tctccctgca    900

acatcctcct tttcctttgc ttccctcctc aactaccgca acatctggaa aaatctgctt    960

atcctgggct tcaccaactt cattgcccat gccattcgcc actgctacca gcctgtggga   1020

ggaggaggga gcccatcgga cttctacctg tgctctctgc tggccagcgg caccgcagcc   1080

ctggcctgtg tcttcctggg ggtcaccgtg gaccgatttg gccgccgggg catccttctt   1140

ctctccatga cccttaccgg cattgcttcc ctggtcctgc tgggcctgtg ggattatctg   1200

aacgaggctg ccatcaccac tttctctgtc cttgggctct tctcctccca agctgccgcc   1260

atcctcagca ccctccttgc tgctgaggtc atccccacca ctgtccgggg ccgtggcctg   1320
```

-continued

```
ggcctgatca tggctctagg ggcgcttgga ggactgagcg gcccggccca gcgcctccac   1380

atgggccatg gagccttcct gcagcacgtg gtgctggcgg cctgcgccct cctctgcatt   1440

ctcagcatta tgctgctgcc ggagaccaag cgcaagctcc tgcccgaggt gctccgggac   1500

ggggagctgt gtcgccggcc ttccctgctg cggcagccac cccctacccg ctgtgaccac   1560

gtcccgctgc ttgccacccc caaccctgcc ctctgagcgg cctctgagta ccctggcggg   1620

aggctggccc acacagaaag gtggcaagaa gatcgggaag actgagtagg gaaggcaggg   1680

ctgcccagaa gtctcagagg cacctcacgc cagccatcgc ggagagctca gagggccgtc   1740

cccaccctgc ctcctccctg ctgctttgca ttcacttcct tggccagagt caggggacag   1800

ggagagagct ccacactgta accactgggt ctgggctcca tcctgcgccc aaagacatcc   1860

acccagacct cattatttct tgctctatca ttctgtttca ataaagacat ttggaataaa   1920

aaaaaaaaa                                                           1929
```

<210> SEQ ID NO 9
<211> LENGTH: 1831
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 9

```
ctggagccgc cctgggtgtc agcggctcgg ctcccgcgca cgctccggcc gtcgcgcacc     60

tcgggcacct gcaggtccgt ggcgtcccgc ggctgggcgc ccctgactcc gtcccggcca    120

gggagggcca tgatttccct cccggggccc ctggtgacca acttgctgcg gtttttgttc    180

ctggggctga gtgccctcgc gccccccctcg cgggcccagc tgcaactgca cttgcccgcc   240

aaccggttgc aggcggtgga gggaggggaa gtggtgcttc cagcgtggta caccttgcac    300

ggggaggtgt cttcatccca gccatgggag gtgccctttg tgatgtggtt cttcaaacag    360

aaagaaaagg aggatcaggt gttgtcctac atcaatgggg tcacaacaag caaacctgga    420

gtatccttgg tctactccat gccctcccgg aacctgtccc tgcggctgga gggtctccag    480

gagaaagact ctggccccta cagctgctcc gtgaatgtgc aagacaaaca aggcaaatct    540

aggggccaca gcatcaaaac cttagaactc aatgtactgg ttcctccagc tcctccatcc    600

tgccgtctcc agggtgtgcc ccatgtgggg gcaaacgtga ccctgagctg ccagtctcca    660

aggagtaagc ccgctgtcca ataccagtgg gatcggcagc ttccatcctt ccagactttc    720

tttgcaccag cattagatgt catccgtggg tctttaagcc tcaccaacct ttcgtcttcc    780

atggctggag tctatgtctg caaggcccac aatgaggtgg gcactgccca atgtaatgtg    840

acgctggaag tgagcacagg tcagtgaggg ggcctggagc tgcagtggtt gctggagctg    900

ttgtgggtac cctggttgga ctggggttgc tggctgggct ggtcctcttg taccaccgcc    960

ggggcaaggc cctggaggag ccagccaatg atatcaagga ggatgccatt gctccccgga   1020

ccctgccctg gcccaagagc tcagacacaa tctccaagaa tgggaccctt tcctctgtca   1080

cctccgcacg agccctccgg ccaccccatg gccctcccag gcctggtgca ttgaccccca   1140

cgcccagtct ctccagccag gccctgccct caccaagact gcccacgaca gatggggccc   1200

accctcaacc aatatccccc atccctggtg ggtttcttc ctctggcttg agccgcatgg    1260

gtgctgtgcc tgtgatggtg cctgcccaga gtcaagctgg ctctctggta tgatgacccc   1320
```

-continued

```
accactcatt ggctaaagga tttggggtct ctccttccta taagggtcac ctctagcaca   1380

gaggcctgag tcatgggaaa gagtcacact cctgaccctt agtactctgc cccacctct   1440

ctttactgtg ggaaaaccat ctcagtaaga cctaagtgtc caggagacag aaggagaaga   1500

ggaagtggat ctggaattgg gaggagcctc cacccacccc tgactcctcc ttatgaagcc   1560

agctgctgaa attagctact caccaagagt gaggggcaga gacttccagt cactgagtct   1620

cccaggcccc cttgatctgt accccacccc tatctaacac cacccttggc tcccactcca   1680

gctccctgta ttgatataac ctgtcaggct ggcttggtta ggttttactg gggcagagga   1740

tagggaatct cttattaaaa ctaacatgaa atatgtgttg ttttcatttg caaatttaaa   1800

taaagataca taatgtttgt atgagataag a                                  1831
```

```
<210> SEQ ID NO 10
<211> LENGTH: 1453
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 903-935
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 10

ggagccaagt ggggccctcg gcctcttcct tcgttccagg cccatgattt tccctacact     60

tctccctggc ccaggctcca gccacaggca cctctcctgc ccccgcccac cctcctgacc    120

gcagctccca ggccctggag acctccaggc tttcctgccc tgggcagccc cacctcacag    180

ccagagtcaa tgccttcatg ggaagggctc ccagccacac ccagagtggc ccaaagctgt    240

tgaagtcagc atcctttgtc ccatcaggac cctcctgcct cctctccagg cccttgttcg    300

cctccccacc ctcctcagag gcccggggaa gggaagagca ggtcagtaca gaggttctgt    360

ctacagggag gggccctggg tctatgcaca gctggagctc tgagccttcc acagcccgtg    420

tgactgctag agggcagggg tgcagggctc agggggccg ggctggtcct ttggggctgg    480

tgttcctacg tcagtcccca cctggggaat aaactccagc ctctcctgct catacagaag    540

gaactggttg ggtttgcttt atgggatctt tgagaccaaa acagatgctc ctgtttgctg    600

ggggagggtg tgagcacgga gtatttctgt ccctcgtgaa gtcacgtcac acaggggaga    660

ggcgaggtcg atggaactgg ccacgcacag gctctggctc tggaaggagg gatgatgagt    720

gggcgttttc ccggcaggcc cccggggtcc tcagcctcag caacccaggg agaggacaga    780

aatgaaccga tggttgaggg attgtcacgg gaggaacatg acacccgaag ggactctagg    840

tgccctcgga gtgccacaca tgcccagacc ttctcacacc cacacaaata ggctctgccg    900

tgnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnttgtt cacactcaga acccaggaca    960

gccacagcca ccgcttaggg gaagccactg cagatgcccc tggaatgggc acagcacagc   1020

cagggcgctc ttccaggcag gcgaggataa cttgagagtt tcctagggca ccagggacag   1080

agctcagagg cccccgaggt gtgtgtagga ggcggaggcc cgcagagcac agagcaggag   1140

aagggcttgg gccctggagg agaaagccat tctggacacc aggggacctg gacggagggt   1200

ccccacagcc cgtgccccac gccgcctgga ggccagaggg gtcagtggcc ctgctgtccc   1260

ggctccatct tggttctagc cgccacctgt atgaacacag tggcccggct taacgcacta   1320
```

-continued

```
acccagcctc tccctgtgtc ccacagggag tagcaagacc caccccacac tgccttcacc   1380

atctacacca gtgacgccgc tgtgtgtctt agcatggaaa taaataaacc tgaatgcaaa   1440

aaaaaaaaaa agg                                                      1453


<210> SEQ ID NO 11
<211> LENGTH: 443
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 11

gatatagaca acttccagag tcaccagtgt gcaaatggag ccacctgcat tagtcatact     60

aatggctatt cttgcctctg ttttggaaat tttacaggaa aattttgcag acagagcaga    120

ttaccctcaa cagtctgtgg gaatgagaag acaaatctca cttgctacaa tggaggcaac    180

tgcacagagt tccagactga attaaaatgt atgtgccggc caggttttac tggagaatgg    240

tgtgaaaagg acattgatga gtgtgcctct gatccgtgtg tcaatggagg tctgtgccag    300

gacttactca acaaattcca gtgcctctgt gatgttgcct ttgctggcga gcgctgcgag    360

gtggacttgg cagatgactt gatctccgac attttcacca ctattggctc agtgactgtc    420

gccttgttac tgatcctctt gct                                            443


<210> SEQ ID NO 12
<211> LENGTH: 1537
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 284-285, 287
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 12

aaaaaaaaca acccggtagc attgtccctt ccccactgac aaacttatca aatccagaag     60

ctttagagtt tcgtctctaa ttatttttct cctgaacaaa attacccaag tcaaaacaaa    120

atgtattttt agaattacgg cagcatacga cctgaatttt gtgagtttcg tggctttatc    180

ttaaatcacc atttccctaa aaacggtttc tttctcctta gaaatgctgg tggcaacttg    240

atgaaacagc caaatgcacc agggcaggtc actttcccaa aaannanaag aaaaaaaact    300

cattgagata gctacagttc tataggttaa tttaaagcct cctttttcta ctcatttttg    360

aaagcaaaat tacattttac tattttacat aaccagtgaa aagacgttga aagcctacag    420

ctcactgttt ttggtgctct ggaaatgttg agggtgggtt tttaaccagt gatttttaac    480

gtgcagtgaa tttgttagac ttttaaacac cagctaaggt agtcaaactt gatccccatt    540

aaaaatcaag gaattagggg tcgggggagg gtttaggagt gatccagaat gacctcccag    600

aattactgtg cgtacaactt tatttttcag agttttcatt ggaatggtaa gagttttatg    660

aaagacagtt ttaaaactta ttctgagtta aatattaata ctttaaaaaa ttattgtact    720

agacttattg cagccttttg aaagtagcag agtttcatca taccacatat ataacagagc    780

ataaattttc tataatcagg caccttttgc tgcttttgag taagactgtt ttcctgttta    840

ggtgttaagc atcgccagac ataaaaaatct attctctcct ctcgattgta gcatagcctg    900
```

-continued

```
acagctctag atacagcatt tctatgatga aaaatgagta tccatcagga aatctagaag    960

actagccgtg ttttctcaga ctccaccttt gtttgcactc tgttgcctgt gaggagcttt   1020

ctggcatgtg attatttact tcaaaactag agttccaagc acctacatta attattttat   1080

attgtgtgca gaatagtata tcttttaatg tcagatatga tacactgcac atattgcttt   1140

tgcactctta aaatttttgt actaaataat agaaaatatt tatattcttt gagtgtgagc   1200

tttgaataga tggcattatc actttattgt ttttttaaca aaaacttttt ctcaattatt   1260

ctattgcaat gttattctga gcaagtccta tgccaaatat cttgtataat gtttgtatgg   1320

aagattaaat tttactcttg tgtggtaaga ctatttcagt tactgatttt atagttggaa   1380

tttgatattc cagcacaaag tccacagtgt attcagaaat ccaagttggt gtcatacatt   1440

tcattttgat gtgaactttt ctttgctttc ctttgttcta agactccatt ttgcaataaa   1500

cgttttgaca gtaaaaaaaa taaaaaagga aaaaaaa                            1537
```

```
<210> SEQ ID NO 13
<211> LENGTH: 972
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 13

acgcaaattc ggcacgaggg ttctaaaacc cagtttggtt tacgttgtct ttcacagtag     60

tatatttagc tcttctctgg aaagttgtgg gttaatataa ttcttaaaca tgaaaatgta    120

attaaacaca ccacgagaga acaatattcc aggagactta atagtgatta ctttcttcaa    180

tcaggaaatc gtttcagtgc ctcctttgta ggaatgcttt gttttgtgat gggttttctt    240

aaagaagagc acacctccgt ccaatctcct gagacagcca cgtctccgct gacatcccac    300

tgtgatgctt tcagatagtc agtgaatgtt tctgataacc ttcatccagt atctgaaaca    360

caatgtgaga gattatattg ttttagataa taacatccca tttagttgac taaaatcttc    420

caaactctga aagctgcaca ctgctactcc agagagtgca ggtcttagct cttctccttt    480

ctgacttcaa gatgaatctt tgggacgatg tttctggtgc ttggtccaca gtgattcact    540

tttgaaggag aggccacatg acatgaactg cctggtgtta caacctagct aacatatttg    600

atgctactcc tgttgtctgt actgcttatt caagtagtat tctaagttat gttactaaaa    660

aacatggtgg gtaaagcaca atcctaccca tcattgtcct ccaaaataat tgtatgacat    720

acacggccca gcccattgcc ctccctgcat ctctgtgctg ctttgccatt tccccttcta    780

cccagcctcc tcaaggggta ccttggtgga tatttcagta cttaaaacca gactgtaatc    840

ataacctccc tctgtgtggc atcaataaat agccaaactc aaaaaaaaaa aaaaaaaaaa    900

aaaaaaaaaa aaaaaatatc ggtcgcaagc ttattccctt tagtgagggt taattttagc    960

ttgcactgcc ta                                                        972
```

```
<210> SEQ ID NO 14
<211> LENGTH: 1544
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 14
```

-continued

```
tactttgact ttggatcatt tccctgactg ggctaatgtg acacatattg agacttagga     60
agagccacaa gaccacacac acagccctta ccctcctcag gactaccgaa ccttctggca    120
caccttgtac agagttttgg ggttcacacc ccaaaatgac ccaacgatgt ccacacacca    180
ccaaaaccca gccaatgggc cacctcttcc tccaagccca gatgcagaga tggacatggg    240
cagctggagg gtaggctcag aaatgaaggg aacccctcag tgggctgctg gacccatctt    300
tcccaagcct tgccattatc tctgtgaggg aggccaggta gccgagggat caggatgcag    360
gctgctgtac ccgctctgcc tcaagcatcc cccacacagg gctctggttt tcactcgctt    420
cgtcctagat agtttaaatg ggaatcagat cccctggttg agagctaaga caaccaccta    480
ccagtgccca tgtcccttcc agctcacctt gagcagcctc agatcatctc tgtcactctg    540
gaagggacac cccagccagg gacggaatgc ctggtcttga gcaacctccc actgctggag    600
tgcgagtggg aatcagagcc tcctgaagcc tctgggaact cctcctgtgg ccaccaccaa    660
aggatgagga atctgagttg ccaacttcag gacgacacct ggcttgccac ccacagtgca    720
ccacaggcca acctacgccc ttcatcactt ggttctgttt taatcgactg gccccctgtc    780
ccacctctcc agtgagcctc cttcaactcc ttggtcccct gttgtctggg tcaacatttg    840
ccgagacgcc ttggctggca ccctctgggg tccccctttt ctcccaggca ggtcatcttt    900
tctgggagat gcttcccctg ccatccccaa atagctagga tcacactcca agtatgggca    960
gtgatggcgc tctgggggcc acagtgggct atctaggtcc tccctcacct gaggcccaga   1020
gtggacacag ctgttaattt ccactggcta tgccacttca gagtctttca tgccagcgtt   1080
tgagctcctc tgggtaaaat cttccctttg ttgactggcc ttcacagcca tggctggtga   1140
caacagagga tcgttgagat tgagcagcgc ttggtgatct ctcagcaaac aacccctgcc   1200
cgtgggccaa tctacttgaa gttactcgga caaagacccc aaagtggggc aacaactcca   1260
gagaggctgt gggaatcttc agaagccccc ctgtaagaga cagacatgag agacaagcat   1320
cttctttccc ccgcaagtcc attttatttc cttcttgtgc tgctctggaa gagaggcagt   1380
agcaaagaga tgagctcctg gatggcattt tccagggcag gagaaagtat gagagcctca   1440
ggaaacccca tcaaggaccg agtatgtgtc tggttccttg ggtgggacga ttcctgacca   1500
cactgtccag ctcttgctct cattaaatgc tctgtctccc gcgg                    1544
```

<210> SEQ ID NO 15
<211> LENGTH: 1109
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 751-806, 884-885
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 15

```
cggacgcgtg gggccagcc tggaggccca gacgtggcgc agcgactcgg aggttcgcct     60
ccagcttgcg catcatctgc ggccgggtcc cgatgagcct cctgttgcct ccgctggcgc    120
tgctgctgct tctcgcggcg cttgtggccc cagccacagc cgccactgcc taccggccgg    180
actggaaccg tctgagcggc ctaacccgcg cccgggtaga gacctgcggg ggatgacagc    240
tgaaccgcct aaaggaggtg aaggctttcg tcacgcagga cattccattc tatcacaacc    300
```

-continued

```
tggtgatgaa acacctccct ggggccgacc ctgagctcgt gctgctgggc cgccgctacg     360

aggaactaga gcgcatccca ctcagtgaaa tgacccgcga agagatcaat gcgctagtgc     420

aggagctcgg cttctaccgc aaggcggcgc ccgacgcgca ggtgcccccc gagtacgtgt     480

gggcgcccgc gaagccccca gaggaaactt cggaccacgc tgacctgtag gtccgggggc     540

gcggcggagc tgggacctac ctgcctgagt cctggagaca gaatgaagcg ctcagcatcc     600

cgggaatact tctcttgctg agagccgatg cccgtccccg ggccagcagg gatggggttg     660

gggaggttct cccaacccca ctttcttcct tccccagctc cactaaattc cctcctgcct     720

taaaaaaaac aagaaaaacc aaacaaacaa nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     780

nnnnnnnnnn nnnnnnnnnn nnnnnntctt ctatagtgtc acctaaattc aattcactgg     840

ccgtcgtttt acaacgtcgt gactggcgac ggacaaagtt atcnntttaa tcgccttgca     900

gcacataccc ctttggccag ctggggtaat aggggaagcg ggccggaccc gatcggcctt     960

cccaaacagt tggggaagct tgaaatgcgc gacattgggc cgacggcctt ctatacggga    1020

ggatctctaa acgcggccgg ggtgttggtt gggttaaggc ggagtgtgac cccgcataat    1080

aacttttgca caggggccct atagggggcc                                     1109
```

```
<210> SEQ ID NO 16
<211> LENGTH: 1740
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 16

aagagaagtt accccgatga cttggtttgg aaggggttaa ggcaccagtg catcctcttc      60

taaagtgatt tatgatgatg tgtggagttt aaaaacttta ccccacccca aagaacagcc     120

ctctcactcc tcactgagtc cactctgaac gtgctaaaat gggaaggagg cggtgttttg     180

ctgatctgtt aaattcttag tgaagtttcc ttgatttcca gtggctgctg ttgtttgagt     240

ttggtttgga gcaaaactga ggtagtccta acatttctgg gactgaatcc aggcaagaga     300

aagaagaaaa agaagaagaa aaagaggagg aaaaaggtag ggagaaataa agggaggaga     360

gaagcacagt gaaagaaaaa aaaagtccct tttgcgacat cacattcctg tgttttccct     420

cagcctggaa aacatattaa tcccagtgct tttacgcccg gaaacaaaga gactaagcca     480

gactatgggg gaaagggaga taagaaggat cctggaactt taaagaggga aagagtgaga     540

ttcagaaatc gccaggactg gactttaagg gacgtcctgt gtcagcacaa gggactggca     600

cacacagaca cacgagaccg aggagaaact gcagacaaat ggagatacaa agacttagaa     660

ggacagctcc tttcacctca tcctacttgt ccagaaggta aaaagacaca gccagaaaga     720

aaaggcatcg gctcagctct cagatcagga caggctgtgg atctgtggcg gtactctgaa     780

agctggagct gcagcacacc ccttttgtat tgctcaccct cggtaaagag agagagggct     840

gggaggaaaa gtagttcatc taggaaactg tcctgggaac caaacttctg atttcttttg     900

caaccctctg cattccatct ctatgagcca ccattggatt acacaatgac atggagaatg     960

ggaccccgtt tcactatgct gttggccatg tggctagtgt gtggatcaga accccacccc    1020

catgccacta ttagaggcag ccacggagga cggaaagtgc ctttggtttc tccggacagc    1080

agtaggccag ctcggtttct gaggcacact gggaggtctc gcggaattga gagatccact    1140
```

-continued

```
ctggaggaac caaaccttca gcctctccag agaaggagga gtgtgcccgt gttgagacta   1200

gctcgcccaa cagagccgcc agcccgctcg gacatcaatg ggccgccgt gagacctgag   1260

caaagaccag cagccagggg ctctccgcgt gagatgatca gagatgaggg gtcctcagct   1320

cggtcaagaa tgttgcgttt cccttacggg gtccagctct cccaacatcc ttgccagctt   1380

tgcagggaag aacagagtat gggtcatctc agcccctcat gcctcggaag gctactaccg   1440

cctcatgatg agcctgctga aggacgatgt gtactgtgag ctggcggaga ggcacatcca   1500

acagattgtg ctcttccacc aggcaggtga ggaaggaggc aaggtgagaa ggatcaccag   1560

cgagggccag atcctggagc agcccctggg accctagcct catccctaag ctgatgagct   1620

tcctgaagct ggagaagggc aagtttggca tggtgctgct gaagaagacg ctgcaggtgg   1680

aggagcgcta tccatatccc gttaggctgg aagccatgta cgaggtcatc gaccaaggcc   1740
```

```
<210> SEQ ID NO 17
<211> LENGTH: 4467
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 971, 978, 1295-1522
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 17

gcgtcgcgct caccctgcgc gtgcccccgc ggcgcctggg cgtcttcctg gactacgagg     60

ccggagagct gtccttcttc aacgtgtccg acggctccca catcttcacc ttccacgaca    120

ccttctcggg cgcgctctgt gcgtacttca ggcccagggc ccacgacggc ggcgaacatc    180

cggatcccct gaccatctgc ccgctgccgg ttagagggac gcgcgtcccc gaagagaacg    240

acagtgacac ctggctacag ccctatgagc ccgcggatcc cgccctggac tggtggtgag    300

gcgccctcgt ggccgcggga ctggccccgg gggggccccc tggatcccag gccagcgctt    360

tgctctcctg ctccgtctga agggagcagg tgcaccagcc aaaatgtcag cgagggggac    420

aaagagaggg acctttgcct acgtagatgt gtatgtgtag tgcgattttc ttcaaggaaa    480

ggagacaagt ccaaagctcg tttgtggatt gtgggactga gcaaaggagt acaaatatat    540

ccacgtcgct cagagctggg gtgctcacgg tgggtggtgg gaaagaagcc agcatggaag    600

aaagaaggga gaaaactttg gtgactgcct tagagggatc agttaatttg tatagtttta    660

tatttttgt atatgtttgc tagctctaaa aaggtcgaga tgcaataaca cttcgtaagc    720

aacgagttca cctaagtaag gctcagatcc tagttttaaa aaccatttcc cattaaaatg    780

aagttggagg aacagctgct tctggagccg gggcaaaaaa tttcaaggtg agcctggagc    840

attgtgtgtg gtgaagtaaa ataaaggctc aaaacgtgac ggcaacccgg caaaagggta    900

gggagccagg ccgaagggcc tcactgacca attgtgggac aatttgaaca tcaggatgaa    960

taatgacagg ngaggttnta acacactgaa taaaaacata atccatgagt tcatgctgat   1020

actcaaattt ctttttaaaa aggagaaaca ggaaggtttc ttttggaggt gaaatctaat   1080

tattggtgag agtcttggag aacaggctgt ttccagtctc aaagcagtaa ccttatacac   1140

tacttataag tttgaaaggg gaaaggttac ctttacaatg gagacatcta ccagatgcat   1200

ccaagtgatt aaatttaaca tcatcaatga tgggaccaag gacattatta gtttgacaac   1260
```

-continued

```
tggggaaaga agtgttcttc accccctacc cccannnnnn nnnnnnnnnn nnnnnnnnnn   1320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1500
nnnnnnnnnn nnnnnnnnnn nnttcatttt cagagtgaga catttgtact gtggctatgt   1560
aggagaacat tcttgttctt agcaaacata ctgaagtttt tagatattaa ttaccacagt   1620
gtctgccact gaatttccag tgactaagtg gaaaaatata aaacatatga atataaagaa   1680
agaaagagac aagtcaaatg tagtaaaatg acaacacttg gtgactctag gtgactggtc   1740
gacagatgtt cattgtacta tcaatgtggc tttgctgtgg gtttgaaatt ttgcaaacta   1800
agagttgggt ggcgggggaga aggatacacc aaaaaactaa gtgattatct ttggatggga   1860
aaatgtttgg taattgcatt cttaaaatgt cttctttgta ttttttaatg ttcaataatg   1920
tatatgtatc agttctgtaa taaaggggaa aacacttttt ttaaatactc ataaaaaacc   1980
atccgtagga tcgagaagat caggcagaag ggctttgtcc agaaatgtaa ggcctctggt   2040
gtagagggcc aggtggtggc ggaggggaat gacggtggag gggggagcagg aaggccaagc   2100
ctgggcagcg agaagaagaa agaggaccca aggagagcac aagtcccacc aaccagagag   2160
agtcgggtga aggtcctgag aaaactggcc gccactgcac cagctttgcc ccaacctccc   2220
tcaaccccca gagccaccac ccttcctcct gccccaggcc acaacagtga ctcggtccac   2280
gtcccgggcg gtaacagttg ctgcaagacc tatgaccacc actgcctttc ccaccacggc   2340
agaggccctg gaccccctca ccctcccaca ggccccctac aaccactgag gtgatcactg   2400
ccaggagacc ctcagtttca gagaatcttt accctccatc ccggaaggat cagcacaggg   2460
agaggccaca gacaaccagg aggcccagca aggccaccag cttggagagc ttcacaaatg   2520
cccctcccac caccatctca gaacccagca caagggctgc tggcccaggc cgtttccggg   2580
acaaccgcat ggacaggcgg gaacatggcc accgagaccc aaatgtggtg ccaggtcctc   2640
ccaagccagc aaaggagaaa cctcccaaaa agaaggccca ggacaaaatt cttagtaatg   2700
agtatgagga gaagtatgac ctcagccggc ctactgcctc tcagctggag gacgagctgc   2760
aggtggggaa tgttcccctt aaaaaagcaa aggagtctaa aaagcatgaa aagcttgaga   2820
aaccagagaa ggagaagaaa aaaaagatga agaatgagaa cgcagacaag ttacttaaga   2880
gtgaaaagca aatgaagaag tctgagaaaa agagcaagca agagaaagag aagagcaaga   2940
agaaaaaagg aggtaaaaca gaacaggatg gctatcagaa acccaccaac aaacacttca   3000
cgcagagtcc caagaagtca gtggccgacc tgctggggtc ctttgaaggc aaacgaagac   3060
tccttctgat cactgctccc aaggctgaga acaatatgta tgtgcaacaa cgtgatgaat   3120
atctggaaag tttctgcaag atggctacca ggaaaatctc tgtgatcacc atcttcggcc   3180
ctgtcaacaa cagcaccatg aaaatcgacc actttcagct agataatgag aagcccatgc   3240
gagtggtgga tgatgaagac ttggtagacc aagcgtctca tcagcgagct gaggaaagag   3300
tacggaatga cctacaatga cttcttcatg gtgctaacag atgtggatct gagagtcaag   3360
caatactatg aggtaccaat aacaatgaag tctgtgtttg atctgatcga tactttccag   3420
tcccgaatca aagatatgga gaagcagaag aaggagggca ttgtttgcaa agaggacaaa   3480
aagcagtccc tggagaactt cctatccagg ttccggtgga ggaggaggtt gctggtgatc   3540
```

-continued

```
tctgctccta acgatgaaga ctgggcctat tcacagcagc tctctgccct cagtggtcag   3600

gcgtgcaatt ttggtctgcg ccacataacc attctgaagc ttttaggcgt tggagaggaa   3660

gttgggggag tgttagaact gttcccaatt aatgggagct ctgttgttga gcgagaagac   3720

gtaccagccc atttggtgaa agacattcgt aactattttc aagtgagccc ggagtacttc   3780

tccatgcttc tagtcggaaa agacggaaat gtcaaatcct ggtatccttc cccaatgtgg   3840

tccatggtga ttgtgtacga tttaattgat tcgatgcaac ttcggagaca ggaaatggcg   3900

attcagcagt cactggggat gcgctgccca gaagatgagt atgcaggcta tggttaccat   3960

agttaccacc aaggatacca ggatggttac caggatgact accgtcatca tgagagttat   4020

caccatggat acccttactg agcagaaata tgtaacctta gactcagcca gtttcctctg   4080

cagctgctaa aactacatgt ggccagctcc attcttccac actgcgtact acatttcctg   4140

ccttttctt tcagtgtttt tctaagacta aataaatagc aaactttcac ctattcatga   4200

gttattattg aaacctcaaa tcataaagac atttaaaaga attgtttttc taactggagg   4260

ggctctagtg ctaaataata gtactgaaaa ttgatattat tttccttttc ttatatgaag   4320

gaccttattt ggcatataaa attttataaa atatgtattt aaagcttttt cttattttt    4380

gtattaattg gtaagtgaaa actctgttaa agatcacacc acaatgtttt caagaaacat   4440

ctgaaaagat aaaacaaaga acaaata                                        4467
```

```
<210> SEQ ID NO 18
<211> LENGTH: 2965
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 1473-1777, 2948, 2951, 2961
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 18

aaacaaagtt caatttagct ggatttctga actatggttt tgaatgttta aagaagaatg     60

atgggtacag ttaggaaagt ttttttctta cacccgtgac ttgagggaaa cattgcttgt    120

ctttgagaaa ttgactgaca tactggaaga gaacaccatt ttatctcagg ttagtgaaga    180

atcagtgcag gtccctgact cttattttcc cagaggccat ggagctgaga ttgagactag    240

ccttgtggtt ttcacactaa agagtttcct tgttatgggc aacatgcatg acctaatgtc    300

ttgcaaaatc caatagaagt attgcagctt ccttctctgg ctcaagggct gagttaagtg    360

aaaggaaaaa cagcacaatg gtgaccactg ataaaggctt tattaggtat atctgaggaa    420

gtgggtcaca tgaaatgtaa aaagggaatg aggtttttgt tgttttttgg aagtaaaggc    480

aaacataaat attaccatga tgaattctag tgaaatgacc ccttgacttt gcttttctta    540

atacagatat ttactgagag gaactatttt tataacacaa gaaaaattta caattgatta    600

aaagtatcca tgtcttggat acatacgtat ctatagagct ggcatgtaat tcttcctcta    660

taaagaatag gtataggaaa gactgaataa aaatggaggg atatcccctt ggatttcact    720

tgcattgtgc aataagcaaa gaagggttga taaaagttct tgatcaaaaa gttcaaagaa    780

accagaattt tagacagcaa gctaaataaa tattgtaaaa ttgcactata ttaggttaag    840

tattatttag gtattataat atgctttgta aattttatat tccaaatatt gctcaatatt    900
```

-continued

```
tttcatctat taaattaatt tctagtataa ataagtagct tctatatctg tcttagtcta    960
ttataattgt aaggagtaaa attaaatgaa tagtctgcag gtataaattt gaacaatgca   1020
tagatgatcg aaaattacgg aaaatcatag ggcagagagg tgtgaagatt catcattatg   1080
tgaaatttgg atctttctca aatccttgct gaaatttagg atggttctca ctgtttttct   1140
gtgctgatag taccctttcc aaggtgacct tcaggggggat taaccttcct agctcaagca   1200
aggagctaaa aggagcctta tgcatgatct tcccacatat caaaataact aaaaggcact   1260
gagtttggca tttttctgcc tgctctgcta agaccttttt tttttttttac tttcattata   1320
acatattata catgacatta tacaaaaatg attaaaatat attaaaacaa catcaacaat   1380
ccaggatatt tttctataaa actttttaaa aataattgta tctatatatt caattttaca   1440
tcctttttca aaggctttgt ttttctaaag gcnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnatc tgggccttac gtaatatatt   1800
ttcttaatgg ctgcataata tcacatcaaa taggcatttt tcaaacctct ttccttatta   1860
aacatgtaga ctatatccat tttttactaa aataaataac atttcagata atatctttgc   1920
actgataatg ttgccaagcc atttctaaag tgaccttatc aatttaatta ccattggatg   1980
agggtgttgc tttcatcgca ccattgtaga ttgtcttttt tatttcaatt tgcgtttatt   2040
tataactggt tgcaaaggta cacagaacac acgctccttc aacttatctt tgataaaccc   2100
aagcaaggat acaaaaagtt ggacgacatt gagtagagtc atggtatacg gtgctgaccc   2160
tacagtatca gtggaaaaga taaggaaaat gtcactactc acctatgtta tgcaaaacag   2220
ttaggtgtgc tggggctgga tactgctctt ttacttgagc attggttgat taaagtttag   2280
gtaccatcca gggctggtct agagaagtct ttggagttaa ccatgctctt tttgttaaag   2340
aagagagtaa tgtgtttatc ctggctcata gtccgtcacc gaaaatagaa aatgccatcc   2400
ataggtaaaa tgctgaccta tagaaaaaaa tgaactctac ttttatagcc tagtaaaaat   2460
gctctacctg agtagttaaa agcaattcat gaagcctgaa gctaaagagc actctgatgg   2520
ttttggcata atagctgcat ttccagacct gacctttggc cccaaccaca agtgctccaa   2580
gccccaccag ctgaccaaag aaagcccaag ttctccttct gtccttccca caacctccct   2640
gctcccaaaa ctatgaaatt aatttgacca tattaacaca gctgactcct ccagtttact   2700
taaggtagaa agaatgagtt tacaacagat gaaaataagt gctttgggcg aactgtattc   2760
cttttaacag atccaaacta ttttacattt aaaaaaaaag ttaaactaaa cttctttact   2820
gctgatatgt ttcctgtatt ctagaaaaat ttttacactt tcacattatt tttgtacact   2880
ttccccatgt taagggatga tggcttttat aaatgtgtat tcattaaatg ttactttaaa   2940
aataaaanaa naaaaaaaaa naaaa                                          2965
```

<210> SEQ ID NO 19
<211> LENGTH: 1734
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:

-continued

<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 571-899
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 19

```
cgcctccgga aactgccccc cgggctgctg gccaacttca ccctcctgcg caccccttgac    60

cttggggaga accagttgga gaccttgcca cctgacctcc tgaggggtcc gctgcaatta   120

gaacggctac atctagaagg caacaaattg caagtactgg gaaaagatct cctcttgccg   180

cagccggacc tgcgctacct cttcctgaac ggcaacaagc tggccagggt ggcagccggt   240

gccttccagg gcctgcggca gctggacatg ctggacctct ccaataactc actggccagc   300

gtgcccgagg ggctctgggc atccctaggg cagccaaact gggacatgcg ggatggcttc   360

gacatctccg gcaacccctg gatctgtgac cagaacctga gcgacctcta tcgttggctt   420

caggcccaaa aagacaagat gttttcccag aatgacacgc gctgtgctgg gcctgaagcc   480

gtgaagggcc agacgctcct gggcagtggc caagtcccag tgagaccagg ggcttgggtt   540

gagggtgggg ggtctggtag aacactgcaa nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   600

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   660

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   720

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   780

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   840

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnna   900

taatcctgct tttacaggtg aaactcgggg ctgtccatag cggctgggac cccgtttcat   960

ccatccatgc ttcctagaac acacgatggg ctttccttac ccatgcccaa ggtgtgccct  1020

ccgtctggaa tgccgttccc tgtttcccag atctcttgaa ctctgggttc tcccagcccc  1080

ttgtccttcc ttccagctga gccctggcca cactggggct gcctttctct gactctgtct  1140

tccccaagtc agggggctct ctgagtgcag ggtctgatgc tgagtcccac ttagcttggg  1200

gtcagaacca aggggtttaa taaataaccc ttgaaaactg gatcggatga attggctttc  1260

attgtgttcc tagcatcttc tcaaatcaac ttcccaggac tccagggtga aggaggaaaa  1320

gaggcatggc ccaggccctg gggtgtggga tatggtctcc ctaggggatg acagttggga  1380

tcaatggcct gtgacttctc ctctcccttc ccccatcctg ggacctaact ggaaataaaa  1440

ccttgactgt tgcccgggtg tcattttacc agtggatttc tgccagggct tgtgtcctag  1500

gagaaggttt aagttaaacc agattgccca ggtctccaaa cgatttgtca tgctgacctg  1560

agatcatcga agggggcacc tgcccccggg caaggttgca ggggcaggat ggggctgaag  1620

ggatgagcag ggtcccgggc ccacctgctg atacagcatt ggccatgtgg gggctgcaat  1680

cggatttgga agaccctggg gcttgggggc atgtccattt tcccagtccc taaa        1734
```

<210> SEQ ID NO 20
<211> LENGTH: 4005
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 20

-continued

```
ggacaccgtc tgcagtggag tcactggtgc cgtaaatgtg gccaaggggg ccgtccagac     60
gggctgtaga cacggccaag accgtgctga ccggcaccaa ggacacagtc actactgggc    120
tcatgggggc agtgaatgtc gccaaaggga ccgtccagac cagtgtggac accaccaaga    180
ctgtcctaac tggtaccaag gacaccgtct gcagtggggt gaccggtgct gcgaatgtgg    240
ccaagggggc cgtccagggg ggcctggaca ctacaaagtc tgtcctgact ggcactaaag    300
acaccgtatc cactgggctc acaggggctg tgaacttggc caaagggact gtccagaccg    360
gcgtggacac cagcaagact gtcctgaccg gtaccaagga caccgtctgc agtggagtca    420
ctggtgccgt aaatgtggcc aaaggcaccg tccagacagg tgtggacaca gccaagacgg    480
tgctgagtgg cgctaaggat gcagtgacta ctggagtcac gggggcagtg aatgtggcca    540
aaggaaccgt gcagaccggc gtggacgcct ccaaggctgt gcttatgggt accaaggaca    600
ctgtcttcag tggggttacc ggtgccatga gcatggccaa aggggccgtc caggggggcc    660
tggacaccac caagacagtg ctgaccggaa ccaaagacgc agtgtccgct gggctcatgg    720
ggtcagggaa cgtggcgaca ggggccaccc acactggcct cagcaccttc cagaactggt    780
tacctagtac cccgcgccacc tcctggggtg gactcaccag ttccaggacc acagacaatg   840
gtggggagca gactgccctg agcccccaag aggccccgtt ctctggcatc tccacgcccc    900
cggatgtgct cagtgtaggc ccggagcctg cctgggaagc cgcagccact accaagggcc    960
ttgcgactga cgtggcgacg ttcacccaag gggccgcccc aggcagggag gacacggggc   1020
ttttggccac cacacacggc cccgaagaag ccccacgctt ggcaatgctg cagaatgagt   1080
tggaggggct gggggacatc ttccacccca tgaatgcgga ggagcaagct cagctggctg   1140
cctcccagcc cgggccaaag gtgctgtcgg cggaacaggg gagctacttc gttcgtttag   1200
gtgacctggg tcccagcttc cgccagcggg catttgaaca cgcggtgagc cacctgcagc   1260
acggccagtt ccaagccagg gacactctgg cccagctcca ggactgcttc aggctgattg   1320
aaaaggccca gcaggctcca gaagggcagc cacgtctgga ccagggctca ggtgccagtg   1380
cggaggacgc tgctgtccag gaggagcggg atgccggggt tctgtccagg gtctgcggcc   1440
ttctccggca gctgcacacg gcctacagtg gcctggtctc cagcctccag ggcctgcccg   1500
ccgagctcca gcagccagtg gggcgggcgc ggcacagcct ctgtgagctc tatggcatcg   1560
tggcctcagc tggctctgta gaggagctgc ccgcagagcg gctggtgcag agccgcgagg   1620
gtgtgcacca ggcttggcag gggttagagc agctgctgga gggcctacag cacaatcccc   1680
cgctcagctg gctggtaggg cccttcgcct tgcccgctgg cgggcagtag ctgtaggagc   1740
ctgcaggccc ggcgcggggt cgccctgctc tgtccaggga ggagctgcct cagaactttc   1800
tccccgcccc caaacctgga tcggttccct aaagccctag acctttgggg ctgcagctgg   1860
ctgagcgccg aggggctgcg gaggcagtga ccttcttaac tgagccaccc cacgccctgc   1920
tccgggcctg cctgcatctc ccacctcctc cccagcgctg cctgcccctc tcggagcctg   1980
gggtcactca gaccaccagc caagagcctt cccttgaagt ccccaagcaa gcactgcaat   2040
taggaaagag aaaaagcagc gtgcccagcc tggaagggca tctgtttgcc ccgctagcaa   2100
cccttttata tctagcaggg ctcttccagt cctgcagcac gggcccccag ctatcagcgg   2160
tgcaggcagt gctgtggcat cccaggctcc gggcagctcc gttctcatgc tgaaagtggg   2220
tctccggcct tagcacacac accttgaggg tcttaagaac cacattccct catagtagaa   2280
```

-continued

```
agtactagaa aaagcgacac tgccatcatc atcccaaggc aggctgctac tgcctttgct   2340

gacccccggg gtggcctcac ggtggggaca aagctgccag gagccacagc agccacagct   2400

ggggctttgc accagcctgg cttgagactg agcagtttgc agggggtggg gggtgcaaaa   2460

aacaagcaaa caggctgctg ctgcctccag ctgcccacca caggcctgcc ccaggcacct   2520

ggggctctga ggcccctggg gaggctgggc ccagcagctg cccctggaga acacagacaa   2580

aggacttccc cgcagggaac tgtgccctat ggagggatca gacagggctg ggaacagcca   2640

cagaggctgc gtgcctatgg cacagccctt cctccgccgc acactccccc tgggtcctca   2700

ggcccaccca agcgccgggc tgcagaggaa gcggggctgg ggaggctgca ggcatcagag   2760

acactggtgg tggcggaccc ggccgccggg ccccgtgctc tcaggctagc ccaggtcgtg   2820

gaggctggca ggctcaggtc gggtgtgaga cgtgccgtgg ctgcgctcag tccagcgggg   2880

aggagccgtt cagcccggcc tccccaggaa gccatatccc cactcacccg gtaagagaac   2940

cttgtcgtcc cctttccatg ctctcctagg acacgagccc aggaacccca gacccagggg   3000

gaggaagggt ggaggggccc caggggtcac catgtgcacc aggggccgtg aggggccggg   3060

gcattcagct cagctctgaa ccggggaagc tggcacggca aggactgcct caggtgacgg   3120

gccgtgagag gggacgggtc aggagccttc ccaagccttc tcctcagccc gacacccatg   3180

gccatcggag gctaggatgc cagacacagc catttgcaga aatcaggcac agtgactgca   3240

gctcacgtcc agccaaccaa gcatggggcc gcagctcagg aagtcccttc ccgccacacc   3300

acagcctaat tcttactggg acggaggcaa ctcggctacg ctgggcagga cgacaaacac   3360

gagacgccac tgtggaatga gcaacttcgg agcacggggt gacttgcttg ggaccgtgcc   3420

cacgtgacag ccccttatgc agaggaggaa agagaagccc cgagtgggag gggaacctgt   3480

ccaaagtcac acggtgtgtg ggtgacacag ctggggtgag tcgaggctgg cccctgaggc   3540

ccatgctccc tgaacgctgg agaccactgt cggctagcag cggctctcag ggaaggcctg   3600

gtctccaccc tcccagccta gcctcgcgga ccctcgtcct ccccacatcg gacctgctca   3660

cctgcctgga ccctgggctg ccagatgcag gaagcatcaa accccccagc ctcgtgggtg   3720

cggggcaggg cgcaggcagc acagcttaga tgccctggtt tgtccctctt gtctcctggg   3780

aagagcttgc tcccgcccag ctctcctgcc actggccttt cagggttggg ctgggcccag   3840

agtgcctttt agtcgcttct cacggtggcc tgatggctca acccagtccc aaacgggccc   3900

agtgacactg ccgactgcac cccagctcag gcccccactg caccagcaat gctagaaaac   3960

caagccaata aaagtgattt cttttttcat taaaaaaaaa aaaaa                   4005

<210> SEQ ID NO 21
<211> LENGTH: 846
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 21

caaaaatgag cggggtgtgg tggcccatgc ctgtagtccc agctgctcgg gagactgaat     60

ctcttgagcc tgggaagcag aggttgcagt gaactgagat cgcgtcactg cactccagcc    120

tggtgacaga gcgagattcc atctcaaaaa aaaaaaacag tatgcacgta caaatttctt    180

aacctgttat caatgtctga gctacataat tatctttcta gttggagttt gttttaggtg    240
```

-continued

```
tgtaccaact gacatttcag tttttctgtt tgaagtccaa tgtattagtg actctgtggc    300

tgctctcttc acctgcccct tgtggcctgt ctacaattct aaatggattt tgaactcaat    360

gtcgtcgctt ctggtttcct gcatatacca atagcattac ctatgacttt ttttttcctg    420

agctattttc actgagctga gctaatgaac taaaactgag ttatgtttaa tatttgtatc    480

aaatacataa aaggaatact gctttttcct tttgtggctc aaaggtagct gcattttaaa    540

atatttgtga aaataaaaac ttttgttatt agaaaaaaaa aaaaaaaaaa aaaaaaaaaa    600

aaaaaaaaaa agaaagacca aaaaggaaga gaagggaaaa agaagaagag aaacggaggg    660

acaacgggaa acacagagag cgagccggtg acgaaaagcg ggaaggccaa cgcaggagaa    720

gaaagagagg ggggcggcgt cgctcattgt gggagtgtcc tcagagttat gcgagtgggg    780

gatgatgggc aggagtgcta tgcgcccctt tgtatgaggg ggtgcctcaa ttgttgatgg    840

gccggg                                                                846
```

<210> SEQ ID NO 22
<211> LENGTH: 1740
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 216-390, 846-889
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 22

```
tacaaagttt taagaaagcc agcatctcag aaaggccttt caaacaagga cacttaatta     60

gccatcttat gtataagaaa agaaatataa agaacatgaa aatttaaaaa cagatttggc    120

agttttataa cagtctagga ggtggtgtta ttttttccta ttaagaatta gagggcaggt    180

taggaataaa taaaatacag tttgaaaata atgagnnnnn nnnnnnnnnn nnnnnnnnnn    240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gcacttcccc tactgattgc tgccttctct    420

gtggctacaa gggacccaca gaattacagg gaagttacag ggaagcaggt ttcatctcaa    480

tattgggaga gatttcaaac aatcacacct gcctgagaag gagtgggctg tcactaggaa    540

tttttattcc cagtccgtca ggaattttgt agaagggctt catgtgctgg taccaatagg    600

acaggaagat tttaatcagc tttactatct atgttttttt atggaaactg tgtgtatgta    660

tacatacatt ttccaaaaag aaaaattaaa tgattataga gattatgttt ttcagactac    720

tcacgtatct gcttttctta ctccccacct ctgctgataa ttcctagttc attggttttt    780

cccccacact ggaattacct ggggagctta aaaaaccctg atgcctgggt cccaccctca    840

gagatnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnt gatatgaatg    900

cagcctgggc actgggaaat ttaaaaactc cccagataat tactgtgaca gccaaggttg    960

agacctccta gtctagagct ttgctataca cagggtggtc cacaagccag gagcatcagc   1020

atcacttggc agcgcttaga gattcagacc ccagacccac tgaatctgac cctgcatttt   1080

cactagaccc caggtgatca gctgcactag actcaacctc taaaccaaga cctcccaccc   1140

tcacagtcta tgatccttta gtgaccctca gctgagtcct gtgctgaact gtgtttgttc   1200
```

-continued

```
tccttgagca catgcccgct gaccagggac agactggatg agcaagcaac ctgctggcta   1260

tggagaagag ccaggctggg taaatgtttg ctgtgactaa gccaggatca aagaactgcc   1320

tgttgcttgc actggctggc actgagcttg ccactctgtg aactgtgctt ccttcccctg   1380

catggacctg tgcctcagtc actattatcc gcaggccttc tccaagggca gccctctcct   1440

tgtttatccc tcttaagcct gcgtgcagga aggcacatta accctgtggc cccctgcagg   1500

caggagggtg ttgggtgccc ttacctacct tgcccttttt cttgtaccgt aggctgtgcc   1560

gtttatgagt aagtgatgtg tgtctgtgtg tgtgtctaga agtgctgcac tcaccttgtg   1620

ttattggagg ttgtgtaacc ccctagcttt gagcctggtc tcagatgttc cttttcccgt   1680

tctctgtcca gccgttaacg cccccagtct gtaataaaag cctatcagcc gtgcacttta   1740


<210> SEQ ID NO 23
<211> LENGTH: 499
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 23

ctgggatagc aataacctgt gaaaatgctc ccccggctaa tttgtatcaa tgattatgaa     60

caacatgcta aatcagtact tccaaagtct aaatatgact attacaggtc tggggcaaat    120

gatgaagaaa ctttggctga taatattgca gcattttcca gatggaagct gtatccaagg    180

atgctccgga atgttgctga aacagatctg tcgacttctg ttttaggaca gagggtcagc    240

atgccaatat gtgtgggggc tacggccatg cagcgcatgg ctcatgtgga cggcgagctt    300

gccactgtga gagcctgtca gtccctggga acgggcatga tgttgagttc ctgggccacc    360

tcctcaattg aagaagtggc ggaagctggt cctgaggcac ttcgttggct gctactgtat    420

atctacaagg accgagaagt caccaagaag ctagtgcggc aggcagagaa gatgggctac    480

aaggccatat ttgtgacag                                                 499


<210> SEQ ID NO 24
<211> LENGTH: 1774
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 1679, 1681, 1684-1686, 1691-1693, 1705-1708, 1710,
      1714, 1716, 1731, 1740, 1744
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 24

ctttcctaga caaggctgaa aggggccaac attatttctg aagacttcat tattggaatt     60

ctatgggagt gatctcactg agctattttg gaatagaaat gtggctagtt gcctgacctc    120

cctcaatggt ttcacgtggc tttcaaaggg aaggaagggc agtgctgact tttggtaaaa    180

tgggcgaaag ggtccatgcc agcaacacaa tcactcaaag tccagatgag ggatcagtaa    240

atacaacgtg cctgaaaggt ggcccttgag cacattcctc cggtagacat taacttatta    300

aattgattct gattacaaat ataaactttg cccccatctc acccagtaac aatgcaagag    360

ttgatgtcag tctataaaag gaagtaggaa ctgtccctgg ctttcaggct ccaacatcct    420

ccccctgtca agatgtggca cctcaaactt tgtgcagtcc tcatgatctt cctgttgctg    480
```

-continued

```
ttgggccagg taaggaggga aggatactta tgtgtgtgtg tggagtgtgg agatgatagt    540
ggtggtggaa cttgaaagct agattcagtc ctgaggaatg gttcctctgt tctgagtcta    600
cagcatctgc ggaatggaat gatcactctt ccaaggtgtg cagcagggtg tcaacacttt    660
catatctgaa tgtctttgcc cttacagata gatggctccc caataccaga agtgagttcg    720
gcaaagagaa ggccacggag aatgacccca ttttggagag gggtttccct caggcctatt    780
ggagcctcct gccgggatga ttctgagtgt atcacaaggc tatgcagaaa aagacgctgt    840
tccttaagtg tggcccagga atgatgtaca taccagggaa agaaaggaca gcagtcacct    900
ccgacaatgc tccgttctat ggaatattga ttaactgcat tttggctgga gacacccaag    960
tgaagcaatc ttgtattttt aatatttaaa ggcagatgta cgctttaaat tggtctccat   1020
ttcttcttag aatgttgata tatggataag cataactaaa cttgtcaatt tagagtttat   1080
ttttctatgg atactattaa atgtctcaaa ttgaaatttt agcagtctgg aattcaagct   1140
tttgagggaa agaaggattc actttgtata ctaaagaaaa aaacagcatt gcccaataat   1200
gtgttaactt ctcaatctgg aaagtgtagt gagagctaca taatcaatag ctacgtaatc   1260
aacttcagca agttcctaag ctgtggccct ggatcccttc actccatact cttcagggag   1320
gtgtcaaagg tggtcaagct tgggaggctg aggcaggaga atggcgtgaa ccgggagacg   1380
gacttgcagt gagccgagat ccgccactga ctccagcctg ggaaagagc gagactccgt    1440
tcacaaaaaa aaagaatcaa aaaaaaaaag gggagccccc ccttggtatc ggaagaccca   1500
gtcctgtaat tcacacaggt tgagttcaag gcattaagcc ctgtaagggc cacttcggcc   1560
cctcagagtt gctgttctga tccaacggaa gccgcttaca aatttccctt cggaatttgc   1620
ctccggcatt ccctaggggc ggtatttgga agcaaagtcc ttttaacagc cagtgtatnc   1680
naannncggg nnngcccttg cggcnnnngn ccananattg ctccttcttc ncctcttctn   1740
tttnttcccc cccgtgtcga cagggggtgt ggtc                               1774
```

<210> SEQ ID NO 25
<211> LENGTH: 4158
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: 3667-3722
<223> OTHER INFORMATION: a, t, c, g, or other

<400> SEQUENCE: 25

```
ctcccacaac aatttcattg ttgttagcat atctatttct ccatacattg taaaactgta     60
atccttaggt atttctaaaa cataaagagg agaattaagt cagctgcaga acaatggggc    120
tgattcttct gctttttctc tggaaaatct ttcattgctt ttggtggaaa tttacctaga    180
ggttacaacc acaggatgta gcttggtctc ttatttgcct ttttgggaaa ccaattaaga    240
ttaatacagg ataaaggaaa aaagcaatct attcattata taacacagtt gtttgtatta    300
cttgttccct gcaaaggaaa tctgttgaat gcttgcattt tgaattcttt tctaatagaa    360
caaccaaaaa aggcttctta tggtgcagca ggaaaaaaga tcatttttat agctttgcat    420
tcttaacata gcatttaaag agcggcatga attagaggaa agacatggaa cacacaggta    480
gtcggtttga gatcatcggc ttaaaagtat cctaggatgg taatgaccca gaagtatttc    540
```

-continued

```
cagttgtcta gtggtgtggt atgcaggaat gagaagtgtt ttctttccat ttcctgttgg    600
acaggtggca atcttagcag agccactatt tggagttgat aactaaagat gcaaataacg    660
tgactatgcc ttctggtcat cctacgacta tttggagttc tccaaaacct tgtaagaggc    720
atgtcaggca tgcagtaaaa gcatctacaa cttcagctgg gcactggcag cataggtctc    780
atcttggacc atacagtccc actttataga agagagtgga agttctccaa aacaatatcc    840
acaacaaagt ctgacctcac tctgagggag atgggaagtg ggaggaagaa ggactaacca    900
gctccctgga gtaagaggaa tttgctttcc ctgtctgccc accaggggct atatgtgcca    960
cctttcaggt tggggccaag gaagtgatgt cagtgtgaca gaagggagag ttagacctcc   1020
agacgtcagc ctccctccca tggggtacat tttcaatctg agtgttgttg ccttagctgt   1080
gttggtatta gcttgattgg ttggtccgct ggttatgagg tgtagggagg cagtttttgt   1140
ttagttttta ggactttgcc tcttcctttg tccttagcat aatttctagg cagagcatcc   1200
acgaagtcgg ttttcattgc cagctcaaga gcgacaatca tttacgagtt cctatgttat   1260
gttaggtgcc ttatgtatat tatcccaaat ccactgcatg gtttaaatac aggcactgga   1320
atataaatga aaaaggtcat tacagtcact gactttctgc aggaccttaa acatttctct   1380
ttccacaagt ttccccttaa tcatgtgtca aacctctctt cctgacggga atgttgtgct   1440
ataatgaatc tgcataacgc ttgggattct aggaggaagg aaggttccat ggacatgtaa   1500
gtacagcata ttcccctcag tcttctagga gggcagagtg aatcccagaa ctggtaagat   1560
tgggaatctg agcattgcca ctttaatctt agaatattta tcattttgac acatcctgtt   1620
ttttagagag gaaaacaaac acagtttctg cattggtagt gtaaagcata ccttgttagg   1680
aacgtgtttt gtaagacaca tttgggttgt cattctagag catgtcaaac tttgtacttc   1740
aaaatatatt tagtatgatt gttagtggta acatatatca aggctttgaa ttaactgttt   1800
tatttaattt tcacaagaag cacttatttt agccatagga aaaccaatct gagctacaaa   1860
tagttcttta aaataagccc aggttattta gctattctag aaagtgccga cttctttcaa   1920
gaagcaggca ttgtaggaca gctgagaatt atcacatagc ctaaattcta gcctggcagc   1980
aagagtcaca tctgagatgt ccaaaaaaaa aaaaaaaaaa cacctgatct acattgaaag   2040
ggggtagact aacgtatgtg agaccatttt cctatttgca gttacaaggt taaagaactt   2100
tgaaggtcat tcggctgcta agaggcatgt cgaacactct gtgtggctct ttcacagtaa   2160
accctcctaa gagcagaaga cacatggctg ttagtgtctg cgtttagatt taatttctca   2220
aataaaggcc cttggctgcg tatcatttca tccagttata aactagggct cctgcaagca   2280
cccccattct aagggtgaat tattgaaatc agttgctatt tgatgagtca caactggccc   2340
agcaggcagg gcatttgaag tcatggtcat caaaaagaaa tgattgtttt ttgaaaagct   2400
aaatgcttaa aatgcttcta gagggaagtc gtggggcgtg tgctcattct ctttaaaatc   2460
agggttgttg agtttgtttt taaacatttt tataagttca tgagaaaaaa tatataaatt   2520
ctaagaacca acactgtatt cccagaaaca tgaccctcgc tggtcttggg tccacatatc   2580
attggactct gggggacaca aagatgcctg tgacactttg gtgttgccga gttagtcaac   2640
aattattctg ggaaaaagca gaattgaatt cttctctaga tgtcctacca gggttggcca   2700
agggccacaa agcaggctaa taaattccca caggatccag acaccaggca aaattgctct   2760
aagaagccag ttactgtcat ccctctatgg ttctagaaaa aatagtacaa aaatgacagg   2820
```

-continued

```
tcatcctatg agcgtcatgc caatgaaacc ccatcttctg gagaagccct tgaatcagaa   2880

ttatcttttt tcttgatgtc gtcagatgca gccagtttct taattttttt aaaaactgta   2940

tgtttctgtg gtatgtatat ttgtacacct aactacctgg cacttggaaa tcacagcact   3000

actcagaggc aattgaataa agagaaattt aattttaaat atcaagtcct gtcaaacatt   3060

tctcaaactt ctgattttat caaaggtttg ccagccaata aagtgcatcc caagtataca   3120

ggggagaaag ctagactcct acagggtcct agagtttaag taattttttt gttattaata   3180

taggtaataa tttttctaat ttttattttt tggttccaaa tgtaaagctc cttgtgttta   3240

cctctgttta tgtcattctt gacatgttta tctaaattat gtgtgctctg tgacaggtga   3300

aatgtaaatc tgggatccat agtcaagata tcataaggac ctacttccca gcctaccttt   3360

cttcctctac ctgataatga taatactcaa aataacaaca ttcaaaggaa acacaaagaa   3420

atcctgcttt cacatctcct atttcttggg ctccttaata actactgatg gtttgttcat   3480

gaaaaaaaat ttttaaatca aaagattgta cttggccctg agttgaaaaa atttcaaaaa   3540

tcaaaagttt gtacttggcc ctgagttgaa aaaaaaaatt cacattctaa gaataaacag   3600

aaaaatgttc ttcttggaag taaataacaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa   3660

aaaaaannnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   3720

nntcttctat agtgtcacct aaattcaatt cactggccgt cgttttacaa cgtcgtgact   3780

gggaaaaccc tggcgttacc caacttaatc gccttgcagc acatccccct ttcgcaagct   3840

ggcgtaatac gcgaagaggc ccgaaccgtt ggcccttccc aacagttgcg cagcctgaat   3900

ggcgaatggg acgcgccctg tagcggcgca ttaagcgcgg cgggtgtggt ggttagtaga   3960

ggtgtgccgt aaaaaataga ataatttttt ttcaagagat gagcagaatt gagtaggaat   4020

gattacgggg aggaaaagat ctagaagata gacaatagag aggagagaaa aagagggacg   4080

aggaggctga gaggaaaaga gtagaagcgt gatatgaata tatacagaaa cagaaaaagg   4140

agagagggta agacataa                                                 4158


<210> SEQ ID NO 26
<211> LENGTH: 366
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 26

Met Ala Leu Arg Phe Leu Leu Gly Phe Leu Leu Ala Gly Val Asp
  1               5                  10                  15

Leu Gly Val Tyr Leu Met Arg Leu Glu Leu Cys Asp Pro Thr Gln
                 20                  25                  30

Arg Leu Arg Val Ala Leu Ala Gly Glu Leu Val Gly Val Gly Gly
                 35                  40                  45

His Phe Leu Phe Leu Gly Leu Ala Leu Val Ser Lys Asp Trp Arg
                 50                  55                  60

Phe Leu Gln Arg Met Ile Thr Ala Pro Cys Ile Leu Phe Leu Phe
                 65                  70                  75

Tyr Gly Trp Pro Gly Leu Phe Leu Glu Ser Ala Arg Trp Leu Ile
                 80                  85                  90

Val Lys Arg Gln Ile Glu Glu Ala Gln Ser Val Leu Arg Ile Leu
```

-continued

```
                95                  100                 105

Ala Glu Arg Asn Arg Pro His Gly Gln Met Leu Gly Glu Glu Ala
                110                 115                 120

Gln Glu Ala Leu Gln Asp Leu Glu Asn Thr Cys Pro Leu Pro Ala
                125                 130                 135

Thr Ser Ser Phe Ser Phe Ala Ser Leu Leu Asn Tyr Arg Asn Ile
                140                 145                 150

Trp Lys Asn Leu Leu Ile Leu Gly Phe Thr Asn Phe Ile Ala His
                155                 160                 165

Ala Ile Arg His Cys Tyr Gln Pro Val Gly Gly Gly Gly Ser Pro
                170                 175                 180

Ser Asp Phe Tyr Leu Cys Ser Leu Leu Ala Ser Gly Thr Ala Ala
                185                 190                 195

Leu Ala Cys Val Phe Leu Gly Val Thr Val Asp Arg Phe Gly Arg
                200                 205                 210

Arg Gly Ile Leu Leu Leu Ser Met Thr Leu Thr Gly Ile Ala Ser
                215                 220                 225

Leu Val Leu Leu Gly Leu Trp Asp Tyr Leu Asn Glu Ala Ala Ile
                230                 235                 240

Thr Thr Phe Ser Val Leu Gly Leu Phe Ser Ser Gln Ala Ala Ala
                245                 250                 255

Ile Leu Ser Thr Leu Leu Ala Ala Glu Val Ile Pro Thr Thr Val
                260                 265                 270

Arg Gly Arg Gly Leu Gly Leu Ile Met Ala Leu Gly Ala Leu Gly
                275                 280                 285

Gly Leu Ser Gly Pro Ala Gln Arg Leu His Met Gly His Gly Ala
                290                 295                 300

Phe Leu Gln His Val Val Leu Ala Ala Cys Ala Leu Leu Cys Ile
                305                 310                 315

Leu Ser Ile Met Leu Leu Pro Glu Thr Lys Arg Lys Leu Leu Pro
                320                 325                 330

Glu Val Leu Arg Asp Gly Glu Leu Cys Arg Arg Pro Ser Leu Leu
                335                 340                 345

Arg Gln Pro Pro Pro Thr Arg Cys Asp His Val Pro Leu Leu Ala
                350                 355                 360

Thr Pro Asn Pro Ala Leu
                365
```

What is claimed is:

1 A combination comprising a plurality of cDNAs wherein the cDNAs are SEQ ID NOs:1-25 and the complements of SEQ ID NOs:1-25.

2. An isolated cDNA comprising a polynucleotide having the nucleic acid sequence of SEQ ID No:8.

3. A method for detecting differential expression of one or more cDNAs in a sample containing nucleic acids, the method comprising:

a) hybridizing the combination of claim 1 with nucleic acids of the sample, thereby forming one or more hybridization complexes;

b) detecting the hybridization complexes; and

c) comparing the hybridization complexes with those of a standard, wherein differences between the standard and sample hybridization complexes indicate differential expression of cDNAs in the sample.

4. The method of claim 3, wherein the nucleic acids of the sample are amplified prior to hybridization.

5. The method of claim 3, wherein the sample is from a subject with a disorder associated with atherosclerosis.

6. The method of claim 3, wherein the sample is from aorta, arteries, arterioles, endothelial cells, plaque, or blood.

7. The method of claim 3 wherein the combination is immobilized on a substrate.

8. The method of claim 3 wherein differential expression is diagnostic of angina pectoris, coronary artery disease, myocardial infarction, hypertension, transient cerebral

ischemia, mesenteric ischemia, peripheral vascular disease, renal artery stenosis, or stroke.

9. The method of claim 6 wherein the substrate is a microarray.

10. A method of using cDNAs to screen a sample to identify a ligand which specifically binds a cDNA, the method comprising:

a) combining the combination of claim 1 with the sample under conditions to allow specific binding; and

b) detecting specific binding between each cDNA and at least ligand, thereby identifying a ligand that specifically binds to each cDNA.

11. The method of claim 10 wherein the ligands are DNA molecules, proteins, RNA molecules, or transcription factors.

12. A vector containing the cDNA of claim 2.

13. A host cell containing the vector of claim 12.

14. A method for producing a protein, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions for expression of protein; and

b) recovering the protein from the host cell culture.

15. A protein comprising a polypeptide having the amino acid of SEQ ID NO:26.

16. A method for using a protein to screen a plurality of molecules or compounds to identify at least one ligand which specifically binds the protein, the method comprising:

a) combining the protein encoded by a cDNA of claim 1 with the plurality of molecules or compounds under conditions to allow specific binding; and

b) detecting specific binding between the protein and a molecule or compound, thereby identifying a ligand which specifically binds the protein.

17. The method of claim 16 wherein the plurality of molecules or compounds is selected from agonists, antagonists, antibodies, DNA molecules, small molecule drugs, immunoglobulins, inhibitors, mimetics, peptide nucleic acids, peptides, pharmaceutical agents, proteins, RNA molecules, and ribozymes.

18. A method of using a protein to produce an antibody, the method comprising:

a) immunizing an animal with the protein encoded by a cDNA of claim 1 under conditions to elicit an antibody response;

b) isolating animal antibodies; and

c) screening the isolated antibodies with the protein, thereby identifying an antibody which specifically binds the protein.

19. A antibody produced by the method of claim 18.

20. A method for using an antibody to detect gene expression in a sample, the method comprising:

a) combining the antibody of claim 19 with a sample under conditions which allow the formation of antibody:protein complexes; and

b) detecting complex formation, wherein complex formation indicates expression of the protein in the sample.

* * * * *

专利名称(译)	动脉粥样硬化相关基因		
公开(公告)号	US20030129176A1	公开(公告)日	2003-07-10
申请号	US10/219664	申请日	2002-08-14
[标]申请(专利权)人(译)	Incyte公司基因组学		
申请(专利权)人(译)	Incyte公司基因组学，INC.		
当前申请(专利权)人(译)	Incyte公司基因组学，INC.		
[标]发明人	JONES KAREN ANNE VOLKMUTH WAYNE WALKER MICHAEL G MURRY LYNN E		
发明人	JONES, KAREN ANNE VOLKMUTH, WAYNE WALKER, MICHAEL G. MURRY, LYNN E.		
IPC分类号	G01N33/50 A61K38/00 A61P9/10 A61P43/00 C07K14/47 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12P21/02 C12Q1/68 G01N33/15 G01N33/53 G01N33/566 G01N37/00 A61K38/43 C07H21/04 C12N9/00 C12N5/06		
CPC分类号	C07K14/47 A61K38/00		
外部链接	Espacenet USPTO		

摘要(译)

本发明涉及包含多种在心血管疾病中差异表达的cDNA的组合。所述组合和组合物可以完全或部分用于诊断，分期，治疗或监测与动脉粥样硬化相关的病症的进展或治疗。