US 20030180764A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2003/0180764 A1

Shang et al. (43) Pub. Date: Sep. 25, 2003

(54) GENES AFFECTED BY CHOLESTEROL TREATMENT AND DURING ADIPOGENESIS

(75) Inventors: **Jin Shang**, Fremont, CA (US); **Benjamin Bowen**, Berkeley, CA (US)

Correspondence Address:
QUINE INTELLECTUAL PROPERTY LAW GROUP, P.C.
P O BOX 458
ALAMEDA, CA 94501 (US)

(73) Assignee: **Lynx Therapeutics, Inc.**

(21) Appl. No.: **10/339,793**

(22) Filed: **Jan. 8, 2003**

Related U.S. Application Data

(60) Provisional application No. 60/347,286, filed on Jan. 9, 2002.

Publication Classification

(51) **Int. Cl.**[7] **C12Q 1/68**; G01N 33/53; G06F 19/00; G01N 33/48; G01N 33/50; C07H 21/04

(52) **U.S. Cl.** **435/6**; 435/7.1; 702/19; 702/20; 536/23.2

(57) **ABSTRACT**

Genes, nucleic acids, proteins, antibodies, marker sets, and arrays are provided. Methods of detecting conditions associated with elevated cholesterol and lipid, as well as during adipogenesis, are also provided.

GENES AFFECTED BY CHOLESTEROL TREATMENT AND DURING ADIPOGENESIS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/347,286 filed Jan. 9, 2002, entitled "GENES AFFECTED BY CHOLESTEROL TREATMENT AND DURING ADIPOGENESIS" and naming Jin Shang et al. as the inventors. This prior application is hereby incorporated by reference in its entirety.

COPYRIGHT NOTIFICATION

[0002] Pursuant to 37 C.F.R. 1.71(e), Applicants note that a portion of this disclosure contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0003] Not Applicable.

FIELD OF THE INVENTION

[0004] The invention relates to new candidate target genes for human diseases related to high cholesterol and high fat, such as atherosclerosis, diabetes mellitus, and obesity. More specifically, it relates to the identification of new genes that exhibit significant changes in expression regulated by cholesterol and during adipogenesis.

BACKGROUND

[0005] Diets high in fat and cholesterol are associated with increased morbidity and mortality due to a variety of interrelated human diseases, including obesity, atherosclerosis, coronary artery heart disease (CAHD), Non-insulin dependent diabetes mellitus (NIDDM) as well as numerous associated pathophysiologic conditions including arthritis, cancers, hypertension, vascular disorders, and liver and gall bladder disease.

[0006] Cholesterol is a component of eukaryotic plasma membranes. In higher organisms, cholesterol is needed for the growth and viability of the cell; but, high levels of cholesterol in the serum can cause disease and death. As a result, organisms have evolved a variety of mechanisms to regulate cholesterol homeostasis. The type of regulation used to maintain cholesterol homeostasis depends on the source of the cholesterol. In an organism, the sources of cholesterol are diet and de novo synthesis. In cells that synthesize cholesterol de novo, there is a feedback regulation of cholesterol synthesis in response to dietary intake of cholesterol, e.g., when dietary cholesterol is high, the gene for 3-hydroxy-3-methylglutaryl CoA reductase is suppressed thereby blocking de novo synthesis of cholesterol. In cells that do not synthesize cholesterol, the uptake of cholesterol from the serum is regulated, e.g., when serum cholesterol is high, additional uptake of cholesterol from the serum is blocked by suppressing the synthesis of new low-density lipoprotein (LDL) receptors. A family of transcription factors, sterol regulatory element binding proteins (SREBPs), regulate numerous genes involved in cholesterol biosynthesis, endocytosis of LDL as well as fatty acid biosynthesis and glucose metabolism.

[0007] When cholesterol homeostasis is disrupted, disease and death can occur. For example, atherosclerosis is the primary cause of heart disease and stroke. Among the many genetic and environmental risk factors that have been identified by epidemiological studies, elevated levels of cholesterol are probably unique in being sufficient to drive the development of atherosclerosis in humans and animal models. Epidemiological studies have shown that the genetic contribution to atherosclerosis is high, frequently exceeding 50%. Although studies on rare Mendelian forms of atherosclerosis have revealed several aberrant single genes underlying disorders that either elevate plasma LDL or decrease plasma HDL (e.g., LDLR, apoB-100, ARH, ABCG5/ABCG8, ABCA1), genes contributing to common multigenic forms of atherosclerosis remain to be identified.

[0008] Furthermore, a potent class of cholesterol lowering drugs, "statins", have been shown to significantly reduce cardiovascular mortality in hypercholesterolemic patients; however, they are not sufficient to fully prevent the progression of atherosclerosis in many susceptible patients. An understanding of genome-wide responses of cells to cholesterol level changes or alterations in cholesterol homeostasis is needed to identify other key players in cholesterol homeostasis and in the development of atherosclerosis.

[0009] Adipocytes play a critical role in energy homeostasis. They synthesize and store lipids when nutrients are plentiful, and release fatty acids into the circulation when nutrients are required. Numerous adipogenic genes are expressed in functional adipocytes, whereas they are not expressed in preadipocytes in which lipid are not accumulated either. Adipocyte development has been extensively studied in cell culture as well as in animal models. There are several lines of evidence supporting that adipose tissue dysfunction plays an important role in the pathogenesis of type II diabetes mellitus, i.e. failure of adipocyte differentiation is a predisposition to developing diabetes, see, e.g., Danforth (2000) *Failure of adipocyte diferentiation causes type II diabetes mellitus? Nature Genetics* 26: 13.

[0010] Adipogenesis in vivo and in vitro is subject to hormonal and transcriptional control, in part mediated by a cascade of transcription factors including members of the CCAAT/enhancer binding protein family, basic helix-loop-helix leucine zipper (bBLH-LZ) family, e.g., ADD1/SREBP1 and peroxisome proliferator activated receptor gamma (PPARgamma) (See, e.g., Wu et al. (1999) *Transcriptional activation of adipogenesis Current Opin. Cell Biol* 11:689-694, Rosen and Spiegelman (2000) *Molecular regulation of adipogenesis Annu Rev Cell Dev Biol* 16:145-171, for recent reviews, as well as Kim and Spiegelman (1996) *ADD1/SREBP1 promotes adipocyte differentiation and gene expression linkedfatty acid metabolism Genes Devel* 10:1096-1107). However details regarding cellular targets of such transcription factors remain largely undetermined, as do the mechanisms underlying their action in physiological and pathological processes.

[0011] Efforts aimed at understanding the molecular mechanisms underlying the cholesterol and lipid homeostasis and metabolism have recently turned to large-scale

analysis of gene expression in either cholesterol loaded cells or in cell culture models of adipogenesis. Most commonly, these studies rely on microarray technology, in which the choice of genes examined is predetermined by the selection among available ESTs and gene annotation accompanying sequence databases.

[0012] For example, following cholesterol loading in a cell culture model of human macrophages, gene expression was evaluated by probing a microarray of 9808 human cDNA products with cellular RNA products. Changes in gene expression were analyzed over a four day period revealing numerous expression products that were either induced or suppressed in response to cholesterol, Shiffman et al. (2000) *Large Scale Gene Expression Analysis of Cholesterol-loaded Macrophages J. Biol. Chem.* 275: 37324-37332. Similarly, analysis of gene expression following induction of adipocyte development using microarray (Soukas et al. (2001) *Distinct transcriptional profiles of adipogenesis in vivo and in vitro. J Biol Chem* 276: 34167-34174) and SAGE technologies (Ji et al. (2000) *Patterns of gene expression associated with BMP*-2-induced osteoblast and adipocyte differentiation of mesenchymal progenitor cell 3*T3-F442A. J Bone Miner Metab* 18:132-9) has demonstrated that a variety of known target sequences are differentially regulated during development of adipocytes in vivo and from 3T3 cells in vitro.

[0013] To date, none of these studies has simultaneously examined the effects of cholesterol and adipogenesis on the regulation of gene expression. More importantly, these studies are biased by the selection of genes present on the microarray. The present invention is based on the discovery of nucleic acid sequences that are regulated by both cholesterol and by adipogenesis. Furthermore, by using Massively Parallel Signature Sequencing (MPSS) Technology, the sequences are not limited to previously characterized EST and cDNA sequences.

SUMMARY OF THE INVENTION

[0014] The present invention relates to a set of polynucleotide sequences that are differentially regulated in response to cholesterol and adipogenesis, exemplified by SEQ ID NO: 1 through SEQ ID NO:443.

[0015] In a first aspect, the invention relates to compositions including one or more nucleic acid expression vectors including the polynucleotide sequences of the invention. For example, such expression vectors include nucleic acids including at least one polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443. Similarly, sequences that hybridize under stringent hybridization conditions, or that are at least about 70%, (or at least about 75%, about 80%, about 85%, about 90%, about 95%, about 97%, about 98%, or at least about 99%) identical to one or more of SEQ ID NOs:1-443 can be included in the expression vectors of the invention. Polynucleotides encoding polypeptides or peptides having a subsequence encoded by such sequences, e.g., SEQ ID NO:1 to SEQ ID NO:443, as well as polypeptides or peptides that are conservative variations thereof are also polynucleotides of the invention. Likewise, expression vectors incorporating nucleic acids with subsequences of at least about 10 contiguous nucleotides of SEQ ID NOs:1-443 (or at least about 12, about 14, about 16, or about 17 contiguous nucleotides of one of the designated sequences) are included among the compositions of the invention. Polynucleotide sequences that correspond to sequences that are physically linked in the human genome to a nucleic acid comprising one of the above polynucleotide sequences are also polynucleotides of the invention. The expression vectors of the invention also include polynucleotide sequences complementary to any one of the above polynucleotide sequences. In some embodiments, the expression vector includes a promoter operably linked to one or more of the nucleic acids described above. Such expression vectors can encode expression products such as sense or antisense RNAs, or polypeptides.

[0016] Recombinant or isolated polypeptides including a sequence or subsequence encoded by a polynucleotide of the invention, such as a sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, and conservatively modified variants thereof, are also a feature of the invention. Similarly, polypeptides encoded by polynucleotides that hybridize under stringent conditions to one of SEQ ID NO:1 through SEQ ID NO:443, or which are at least about 70% identical to one of SEQ ID NO:1 through SEQ ID NO:443, are polypeptides of the invention. Polypeptides (and oligopeptides and peptides) including amino acid subsequences encoded by SEQ ID NO:1 through SEQ ID NO:443 are also a feature of the invention. For example, fusion proteins including a polypeptide subsequence, e.g., an antigenic subsequence, encoded by any of SEQ ID NO:1 through SEQ ID NO:443, are included in the polypeptides of the invention. Likewise, proteins having a subsequence encoded by SEQ ID NO:1 to SEQ ID NO:443 and homologous or variant polypeptides and a peptide or polypeptide tag, such as a reporter peptide or polypeptide, localization signal or sequence, or antigenic epitope, are included among the polypeptides of the invention.

[0017] Cells including an expression vector, and/or expressing a polypeptide as described above, are also a feature of the invention. In certain embodiments, the expressed polypeptide is encoded by an exogenous polynucleotide, i.e., an expression vector. Such expression vectors typically include a polynucleotide sequence encoding the polypeptide of interest operably linked to, and under the transcriptional regulation of, a constitutive or inducible promoter. In other embodiments, the polypeptide is encoded by an endogenous polynucleotide sequence activated by an exogenous promoter and/or enhancer.

[0018] Antibodies specific for a polypeptide having an amino acid sequence or subsequence encoded by a polynucleotide sequence of the invention are also a feature of the invention. Such specific antibodies can be either derived from a polyclonal antiserum or can be monoclonal antibodies. For example, such antibodies are specific for an epitope including or derived from a subsequence encoded by one of SEQ ID NO:1-SEQ ID NO:443.

[0019] Compositions comprising any of the above nucleic acids, polypeptides, peptides, antibodies or cells optionally also include an excipient to facilitate administration, e.g., in an experimental model such as a cell, tissue or non-human mammal or in a non-human or human subject. Where administration to a human subject is contemplated, the excipient is a pharmaceutically acceptable excipient.

[0020] Another aspect of the invention provides labeled nucleic acid or polypeptide (or peptide) probes. For

example, nucleic acid probes of the invention include DNA or RNA molecules incorporating a polynucleotide sequence of the invention, e.g., selected from SEQ ID NO:1 to SEQ ID NO:443, sequences that hybridize under stringent conditions to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are at least about 70% identical to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that encode a polypeptide or peptide comprising a subsequence encoded by any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are physically linked in the human genome to any one of SEQ ID NO:1-SEQ ID NO:443, sequences complementary to any such sequences, or subsequences thereof including at least about 10 contiguous nucleotides. Optionally, the subsequences include at least about 12 contiguous nucleotides of any one of SEQ ID NOs:1-443. Often such subsequences include at least about 14 contiguous nucleotides, typically at least about 16 contiguous nucleotides, and usually at least about 17 contiguous nucleotides of SEQ ID NO:1 to SEQ ID NO:443. These nucleic acid probes can be, e.g., synthetic oligonucleotides and probes, cDNA molecules, amplification products (e.g., produced by PCR or LCR), transcripts, or restriction fragments.

[**0021**] In other embodiments, the labeled probes are polypeptides, i.e., polypeptides or peptides with an amino acid subsequence encoded by a polynucleotide of the invention, e.g., SEQ ID NOs:1-443. Antibodies specific for such polypeptides or peptides are also a feature of the invention (as are polypeptides which bind to such antibodies). For example, a polypeptide probe can be a fusion protein, or a polypeptide with an epitope tag. In one embodiment, the peptide probe includes an antigenic peptide encoded by one of SEQ ID NO:1 through SEQ ID NO:443.

[**0022**] The label of the nucleic acid, polypeptide or antibody probe can be any of a variety of detectable moieties including isotopic, fluorescent, fluorogenic, or colorimetric labels.

[**0023**] In another aspect, the invention relates to a marker set, e.g., for predicting one or more conditions or characteristics related to cholesterol exposure and/or adipogenesis. Such marker sets can include a plurality of nucleic acids including one or more polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443, sequences that hybridize under stringent conditions to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are at least about 70% (or at least about 75%, about 80%, about 85%, about 90%, about 95%, about 97%, about 98%, or at least about 99%) identical to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that encode a polypeptide or peptide comprising a subsequence encoded by any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are physically linked in the human genome to any one of SEQ ID NO:1-SEQ ID NO:443, sequences complementary to any such sequences, or subsequences thereof including at least about 10 contiguous nucleotides of SEQ ID NOs:1-443 (or at least about 12, about 14, about 16, or about 17 contiguous nucleotides of one of the designated sequences).

[**0024**] In one embodiment, the marker set includes a plurality of oligonucleotides, such as synthetic oligonucleotides. In other embodiments, the marker set includes expression products, amplification products, nucleic acid probes, or the like. The marker set of the invention can also include multiple nucleic acids selected from among different molecular classifications, e.g., oligonucleotides, expression products (such as cDNAs), amplification products, restriction fragments, etc. In one embodiment, the marker set is made up of nucleic acids including polynucleotide sequences corresponding to each of SEQ ID NO:1 through SEQ ID NO:443.

[**0025**] Markers of the invention can also be polypeptides, e.g., polypeptides with a subsequence encoded by SEQ ID NO:1-SEQ ID NO:443, or polypeptide or peptide subsequences thereof. Typically a peptide subsequence comprises at least about 5 contiguous amino acids.

[**0026**] Markers of the invention can also be antibodies, e.g., monoclonal or polyclonal antibodies or anti-sera specific for an epitope encoded by one of SEQ ID NO:1 through SEQ ID NO:443.

[**0027**] In certain useful embodiments, the marker set is logically or physically arrayed. For example, the members of the marker set, whether nucleic acid, polypeptide, peptide or antibody, or a combination thereof, can be physically arrayed in a solid phase or liquid phase array, such as a bead (or microbead) array. Arrays, including a plurality of the polynucleotides of the invention, e.g., SEQ ID NO:1 to SEQ ID NO:443, polypeptides including subsequences encoded thereby, or antibodies specific therefor, are also a feature of the invention. In some embodiments, the arrays include polynucleotides corresponding to majority of SEQ ID NO:1 to SEQ ID NO:443, polypeptides including subsequences encoded thereby or antibodies specific therefor. In one embodiment, the array includes polynucleotides corresponding to each of SEQ ID NO:1 to SEQ ID NO:443, polypeptides or peptides encoded by each of SEQ ID NO:1 to SEQ ID NO:443, or antibodies specific therefor. In an embodiment, the marker set is a mixed marker set including members that are selected from nucleic acids, polypeptides or peptides, and antibodies.

[**0028**] In one embodiment, the marker set of the invention is used for evaluating a condition or characteristic associated with alterations in cholesterol and lipid homeostasis and metabolism and/or adipogenesis, by hybridizing one or more nucleic acids of the marker set to a DNA or RNA sample from a cell or tissue (e.g., from a patient), and detecting at least one polymorphic polynucleotide or differentially expressed expression product in the sample. For example, the marker sets are favorably used for evaluating adverse effects of elevated cholesterol. In another related embodiment, differentially expressed expression products are detected using an antibody array.

[**0029**] Another aspect of the invention provides methods for modulating a condition or characteristic associated with alterations in cholesterol or lipid homeostasis and metabolism and/or adipogenesis in a cell, tissue or organism, such as a cell line or tissue of a human or non-human mammal, e.g., a human, a mouse, a rat, a rabbit, a dog, a pig, a sheep or a non-human primate. For example, a physiologic or pathologic response to cholesterol and/or adipogenesis, e.g., associated with the adverse effects of elevated cholesterol, is modulated in one or more cell-types such as liver, adipose tissue, gall bladder, pancreas, monocytes, macrophages, foam cells, T cells, endothelia and smooth muscle derived from blood vessels and gut, fibroblasts, and/or glia and nerve cells. The methods of the invention for regulating a response to cholesterol (or lipids) and/or adipogenesis in a cell or

tissue optionally include modulating expression or activity of at least one polypeptide encoded by a polynucleotide of the invention, such as a nucleic acid with a polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443, sequences that hybridize under stringent conditions to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are at least about about 70% (or at least about 75%, about 80%, about 85%, about 90%, about 95%, about 97%, about 98%, or at least about 99%) identical to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that encode a polypeptide or peptide comprising a subsequence encoded by any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are physically linked in the human genome to any one of SEQ ID NO:1-SEQ ID NO:443, sequences complementary to any such sequences, or subsequences thereof including at least about 10 contiguous nucleotides of SEQ ID NOs:1-443 (or at least about 12, about 14, about 16, or about 17 contiguous nucleotides of one of the designated sequences).

[0030] In one preferred embodiment, a physiologic or pathologic response to cholesterol and/or adipogenesis is regulated by modulating expression or activity of at least one polypeptide contributing to a condition such as obesity, atherosclerosis, diabetes mellitus and/or coronary artery heart disease. In an embodiment, expression is modulated by expressing an exogenous nucleic acid including a polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443. In other embodiments, expression of an endogenous nucleic acid including a subsequence corresponding to one of SEQ ID NO:1 to SEQ ID NO:443 is induced or suppressed, for example, by integrating an exogenous nucleic acid including at least one promoter that regulates expression of the endogenous nucleic acid. In other embodiments, expression or activity is modulated in response to cholesterol and/or lipid.

[0031] In some embodiments, the methods involve detecting altered expression or activity of an expression product, such as an RNA or polypeptide, encoded by a nucleic acid including a polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443. In some cases, altered expression or activity in response to a pharmaceutical agent is detected. In other cases, altered expression or activity in response to diet is detected. In certain embodiments, a plurality of expression products are detected, e.g., in a high-throughput assay. For example, a plurality of expression products can be detected in an array, such as a bead array.

[0032] In an embodiment, a data record related to the altered expression or activity is recorded in a database. For example, a data record can be a character string recorded in a database made up of a plurality of character strings recorded in a computer or on a computer readable medium.

[0033] In another aspect, the invention provides methods evaluating a condition or characteristic associated with alterations is cholesterol or lipid homeostasis or metabolism, and/or adipogenesis in a subject, such as a human subject. For example, the methods of the invention are useful for evaluating conditions and characteristics associated with elevated cholesterol and/or adipogenesis (such as obesity, atherosclerosis, diabetes mellitus (type II) and coronary artery heart disease). The methods of the invention for detecting such a condition or characteristic involve providing a subject cell or tissue sample of nucleic acids and detecting at least one polymorphic polynucleotide sequence or expression product corresponding to a polynucleotide sequence of the invention, such as: a polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443, sequences that hybridize under stringent conditions to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are at least about 70% (or at least about 75%, about 80%, about 85%, about 90%, about 95%, about 97%, about 98%, or at least about 99% identical to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that encode a polypeptide or peptide comprising a subsequence encoded by any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are physically linked in the human genome to any one of SEQ ID NO:1-SEQ ID NO:443, sequences complementary to any such sequences, or subsequences) thereof including at least about 10 contiguous nucleotides of SEQ ID NOs:1-443 (or at least about 12, about 14, about 16, or about 17 contiguous nucleotides of one of the designated sequences).

[0034] Detection of expression products is performed either qualitatively (presence or absence of one or more product of interest) or quantitatively (by monitoring the level of expression of one or more product of interest). In one embodiment, the expression product is an RNA expression product, such as differentially expressed RNA. The present invention optionally includes monitoring an expression level of a nucleic acid or polypeptide as noted herein for detection of a condition or characteristic associated with a physiologic or pathologic response to cholesterol or lipid and/or adipogenesis in an individual, such as a human, or in a population such as a human population.

[0035] Kits which incorporate one or more of the nucleic acids, polypeptides, antibodies, or arrays noted above are also a feature of the invention. Such kits can include any of the above noted components and further include, e.g., instructions for use of the components in any of the methods noted herein, packaging materials, containers for holding the components, and/or the like.

[0036] Digital systems which incorporate one or more representation (e.g., character string, data table, or the like) of one or more of the nucleic acids or polypeptides herein are also a feature of the invention.

DETAILED DISCUSSION

[0037] Lipid and sterol metabolism are integrated in cells, and high fat and high cholesterol diets, typically defined as diets having in excess of 30% of total calories from fat and in excess of 300 mg cholesterol, are a risk factor for a number of human diseases, such as atherosclerosis, obesity, coronary artery heart disease and diabetes mellitus (Type II). The present invention is based on a genome-wide determination of cellular genetic and metabolic responses to both cholesterol and fat loading.

[0038] In recent years, large-scale gene expression analysis of either cholesterol-loaded cells or during adipogenesis have been reported. These studies relied on microarray technology, in which choice of the interrogated genes is defined by ESTs and gene annotation.

[0039] However, to date both processes have not been investigated in concert. In the present invention, MPSS technology has been applied to gene expression profiling of fat cell development and cholesterol loading in cultured cells leading to the identification of numerous genes exhibiting

significant expression changes common to these two interrelated processes. MPSS is a sequence-based, open system with no a priori assumptions, allowing the discovery of novel genes. Furthermore, the sensitivity, dynamic range and quantitative discrimination of MPSS are determined by the number of clones sequenced, which is superior to microarray technology. By sequencing between a large number (e.g., 0.6 million to 1.8 million, as described herein in the Examples), genes expressed at low copy numbers can be readily identified, facilitating the development of novel therapeutic approaches to controlling conditions and diseases associated with excessive dietary cholesterol and fat, as well as genetic conditions exacerbating the effects of cholesterol and fat consumption.

[0040] Definitions

[0041] Before describing the present invention in detail, it is to be understood that this invention is not limited to particular devices or biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "an excipient" includes a combination of two or more excipients; reference to "bacteria" includes mixtures of bacteria, and the like.

[0042] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. In describing and claiming the present invention, the following terminology will be used in accordance with the definitions set out below.

[0043] The term "correlatable," when used relative to, e.g., a condition associated with cholesterol homeostasis and/or adipogenesis, indicates that the designated subject, e.g., a polymorphic nucleic acid or the expression or activity of an expression product, is statistically associated with that condition.

[0044] The term "nucleic acid" is generally used in its art-recognized meaning to refer to a ribose nucleic acid (RNA) or deoxyribose nucleic acid (DNA) polymer, or analog thereof, e.g., a nucleotide polymer comprising modifications of the nucleotides, a peptide nucleic acid, or the like. In certain applications, the nucleic acid can be a polymer that includes both RNA and DNA subunits. A nucleic acid can be, e.g., a chromosome or chromosomal segment, a vector (e.g., an expression vector), a naked DNA or RNA polymer, the product of a polymerase chain reaction (PCR), an oligonucleotide, a probe, etc.

[0045] The term "polynucleotide sequence" refers to a contiguous sequence of nucleotides in a nucleic acid or to a representation, e.g., a character string, thereof. "Polymorphic polynucleotides" are polynucleotide sequences corresponding to a single locus, i.e., alleles at a locus, characterized by at least one variant (or alternative) nucleotide subunit. Thus, a polymorphic polynucleotide is a polynucleotide that differs, e.g., from another allele at the same locus, or between an otherwise homologous or similar polynucleotide, at one or more nucleotide positions.

[0046] The term "unique nucleotides" refers to a polynucleotide sequence corresponding to a unique locus, e.g., a non-repetitive, or unduplicated, locus in the human genome.

[0047] An "expression vector" is a vector, e.g., a plasmid, capable of producing transcripts and, potentially, polypeptides encoded by a polynucleotide sequence included therein. Typically, an expression vector is capable of producing transcripts in an exogenous cell, e.g., a bacterial cell, or a mammalian cultured cell. Expression of a product can be either constitutive or inducible depending, e.g., on the promoter selected.

[0048] In the context of an expression vector, a promoter is said to be "operably linked" to a polynucleotide sequence if it is capable of regulating expression of the associated polynucleotide sequence. The term also applies to alternative exogenous gene constructs, such as expressed or integrated transgenes. Similarly, the term operably linked applies equally to alternative or additional transcriptional regulatory sequences such as enhancers, associated with a polynucleotide sequence.

[0049] An "expression product" is a transcribed sense or antisense RNA, or a translated polypeptide corresponding to a polynucleotide sequence. Depending on context, the term also can be used to refer to an amplification product (amplicon) or cDNA corresponding to the RNA expression product transcribed from the polynucleotide sequence.

[0050] A polynucleotide sequence is said to "encode" a sense or antisense RNA molecule, or a polypeptide, if the polynucleotide sequence can be transcribed (in spliced or unspliced form) or translated into the RNA or polypeptide, or a fragment of thereof.

[0051] A probe and a gene (or expression product) are said to "correspond" when they share substantial structural identity, or complementarity, depending on context. For example, a probe or an expression product, e.g., a messenger RNA, corresponds to a gene when it is derived from a genetic element with substantial sequence identity.

[0052] An "antibody" refers to a protein made up of one or more polypeptides substantially or partially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The term "antibody," as used herein also includes antibody fragments either produced by the modification of whole antibodies or synthesized de novo using molecular biology techniques. Antibodies include single chain antibodies, including single chain Fv (sFv) antibodies in which a variable heavy and a variable light chain are joined together (directly or through a peptide linker) to form a contiguous polypeptide.

[0053] The term "pharmaceutical composition" means a composition suitable for pharmaceutical use in a subject, including an animal or human. A pharmaceutical composition generally comprises an effective amount of an active agent and a pharmaceutically acceptable excipient or carrier.

[0054] The term "effective amount" means a dosage or amount sufficient to produce a desired result. The desired result can comprise an objective or subjective improvement in the recipient of the dosage or amount.

[0055] A "prophylactic treatment" is a treatment administered to a subject who does not display signs or symptoms of a disease, pathology, or medical disorder, or displays only

early signs or symptoms of a disease, pathology, or disorder, such that treatment is administered for the purpose of diminishing, preventing, or decreasing the risk of developing the disease, pathology, or medical disorder. A prophylactic treatment functions as a preventative treatment against a disease or disorder. A "prophylactic activity" is an activity of an agent, such as a nucleic acid, vector, gene, polypeptide, protein, substance, or composition thereof that, when administered to a subject who does not display signs or symptoms of pathology, disease or disorder, or who displays only early signs or symptoms of pathology, disease, or disorder, diminishes, prevents, or decreases the risk of the subject developing a pathology, disease, or disorder. A "prophylactically useful" agent or compound (e.g., nucleic acid or polypeptide) refers to an agent or compound that is useful in diminishing, preventing, treating, or decreasing development of pathology, disease or disorder.

[0056] A "therapeutic treatment" is a treatment administered to a subject who displays symptoms or signs of pathology, disease, or disorder, in which treatment is administered to the subject for the purpose of diminishing or eliminating those signs or symptoms of pathology, disease, or disorder. A "therapeutic activity" is an activity of an agent, such as a nucleic acid, vector, gene, polypeptide, protein, substance, or composition thereof, that eliminates or diminishes signs or symptoms of pathology, disease or disorder, when administered to a subject suffering from such signs or symptoms. A "therapeutically useful" agent or compound (e.g., nucleic acid or polypeptide) indicates that an agent or compound is useful in diminishing, treating, or eliminating such signs or symptoms of a pathology, disease or disorder.

[0057] Polynucleotides of the Invention

[0058] The present invention is based on the identification and isolation of a set of genes regulated by cholesterol and adipogenesis. The specified sequences are implicated in the regulation and metabolism of cholesterol, and in adipogenesis by their differential regulation in response to experimental conditions indicative of cellular metabolic processes either induced by or suppressed by cholesterol and by their regulation during adipogenesis. Unlike the vast majority of polynucleotide sequences present in the human genome, e.g., randomly selected unique or repetitive polynucleotide sequences, this defined and limited group of polynucleotide sequences, possess an extraordinarily high probability of association with loci involved in the genetic and metabolic programs regulating cholesterol and lipid homeostasis and metabolism and adipogenesis.

[0059] Accordingly, in one aspect, the polynucleotide sequences of the invention are useful for identifying chromosomal segments and corresponding cDNAs associated with cholesterol and lipid homeostasis and adipogenesis, and related conditions and disorders, e.g., conditions associated with a physiologic or pathologic response to elevated cholesterol. More generally, the polynucleotide sequences of the invention and corresponding polypeptides are useful, individually and/or collectively, as probes (e.g., probes labeled with a detectable moiety) and markers. Such probes and markers are useful not only for identifying genes encoding products that are candidates for development of therapeutic and prophylactic interventions, but also for evaluating metabolic and genetic responses to cholesterol and lipid (e.g., for diagnostic or prognostic assays for evaluating presence of or susceptibility to a condition related to excess cholesterol and lipid and/or adipose tissue dysfunction in a subject, such as a human subject, or patient). In addition, the polynucleotide sequences of the invention are useful for the production of animal and cell culture models useful for the evaluation of monitoring of therapeutic agents and protocols aimed at reducing risk of morbidity and mortality due to conditions such as obesity, atherosclerosis, diabetes mellitus (type II), and coronary artery heart disease related to excess cholesterol and lipid and/or adipose tissue dysfunction.

[0060] Polynucleotides of the invention include the polynucleotide sequences including the nucleotide sequences represented by SEQ ID NO:1 through SEQ ID NO:443. In addition to the sequences expressly provided in the accompanying sequence listing, polynucleotide sequences that are highly related both structurally and functionally are polynucleotides of the invention. Thus, polynucleotide sequences of the invention include polynucleotide sequences that hybridize to a polynucleotide sequence comprising any of SEQ ID NO:1-SEQ ID NO:443.

[0061] In addition to the polynucleotide sequences of the invention, e.g., enumerated in SEQ ID NO:1 to SEQ ID NO:443, polynucleotide sequences that are substantially identical to a polynucleotide of the invention can be used in the compositions and methods of the invention. Substantially identical, or substantially similar polynucleotide (or polypeptide) sequences are defined as polynucleotide (or polypeptide) sequences that are identical, on a nucleotide by nucleotide basis, with at least a subsequence of a reference polynucleotide (or polypeptide) e.g., selected from SEQ ID NO:1-443. Such polynucleotides can include, e.g., insertions, deletions, and substitutions relative to any of SEQ ID NO:1-443. For example, such polynucleotides are typically at least about 70% identical to a reference polynucleotide (or polypeptide) selected from among SEQ ID NO:1 through SEQ ID NO:443. That is, at least 7 out of 10 nucleotides (or amino acids) within a window of comparison are identical to the reference sequence selected SEQ ID NO:1-443. Frequently, such sequences are at least about 80%, e.g., at least about 90%, and often at least about 95%, or even at least about 98%, or about 99%, identical to the reference sequence, e.g., at least one of SEQ ID NO:1 to SEQ ID NO:443.

[0062] Additionally, the polynucleotide sequences of the invention include polynucleotide sequences that are proximally linked in the human genome to any one of SEQ ID NO:1 through SEQ ID NO:443. In the context of the invention, the term "proximally linked" or "linked" is used to indicate that the sequences reside on the same physical nucleic acid. Most typically, the nucleic acid is an expression product, such as a full length cDNA, or chromosomal segment including the coding domain of an expression product. Using well-known procedures such as genome or chromosome walking (using molecular or bioinformatic approaches), it is a routine matter to identify and isolate such linked nucleic acids. Chromosome walking (and jumping procedures) are well known in the art and are further described, e.g., in Poustka et al. (1987) *Construction and use of human chromosome jumping libraries from NotI-digested DNA Nature* 325:353-5; Jones et al. (1993) *Genome walking with 2- to 4-kb steps using panhandle PCR PCR Methods*

Appl 2:197-203; Shyamala and Ames (1989) *Genome walking by single-specific primer polymerase chain reaction: SSP-PCR Gene* 84:1-8; Kere et al. (1992) *Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts Genomics* 14:241-8; Sandford and Elgar (1992) *A novel method for rapid genomic walking using lambda vectors Nucleic Acids Res* 20:4665-6; and, Cross and Little (1986) *A cosmid vector for systematic chromosome walking Gene* 49: 9-22.

[**0063**] For example, as described in further detail below, labeled probes corresponding to any one or more of SEQ ID NOs:1-443 can be used to screen expression (i.e., cDNA) or genomic (i.e., chromosomal) libraries to identify expression products or genomic segments that include adjacent polynucleotide sequences along with the polynucleotide sequence hybridizing to the probe selected from SEQ ID NO:1 to SEQ ID NO:443. Such linked polynucleotide sequences are also a feature of the invention and are useful in the methods and compositions described herein.

[**0064**] Polynucleotides encoding polypeptides having amino acids sequences or subsequences encoded by SEQ ID NOs:1-443 are also an embodiment of the invention. Subsequences of SEQ ID NO:1-443 including at least about 10 contiguous nucleotides or complementary subsequences are also a feature of the invention. More commonly a subsequence includes, e.g., at least about 12 contiguous nucleotides of one or more of SEQ ID NO: 1 through SEQ ID NO:443. Typically, the subsequence includes at least about 14, frequently at least about 16, and usually at least about 17 contiguous nucleotides of one of the specified polynucleotide sequences. Such subsequences are typically oligonucleotides, such as synthetic oligonucleotides.

[**0065**] In addition, polynucleotide sequences complementary to any of the above described sequences are included among the polynucleotide sequences of the invention.

[**0066**] Where polynucleotide sequences are translated to form a polypeptide or subsequence of a polypeptide, nucleotide changes can result in either conservative or non-conservative amino acid variations. Conservative amino acid variations result from the substitution of residues having functionally similar side chains, i.e., conservative substitutions. Conservative substitution tables providing functionally similar amino acids are well known in the art. Table 1 sets forth six groups which contain amino acids that are "conservative substitutions" for one another. Alternative conservative substitution charts are available in the art and can be used in a similar manner.

TABLE 1

Conservative Substitution Groups				
1	Alanine (A)	Serine (S)	Threonine (T)	
2	Aspartic acid (D)	Glutamic acid (E)		
3	Asparagine (N)	Glutamine (Q)		
4	Arginine (R)	Lysine (K)		
5	Isoleucine (I)	Leucine (L)	Methionine (M)	Valine (V)
6	Phenylalanine (F)	Tyrosine (Y)	Tryptophan (W)	

[**0067**] One of skill will appreciate that many conservative variations of the nucleic acid constructs which are disclosed yield a functionally identical construct. For example, as discussed above, owing to the degeneracy of the genetic code, "silent substitutions" (i.e., substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid. Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence (e.g., about 1%, about 2%, about 3%, about 4%, about 5%, about 6%, about 7%, about 8%, about 9%, about 10% or more) are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

[**0068**] Methods for obtaining conservative variants, as well as more divergent versions of the nucleic acids and polypeptides of the invention are widely known in the art. In addition to naturally occurring homologues which can be obtained, e.g., by screening genomic or expression libraries according to any of a variety of well-established protocols, see, e.g., Ausubel, Sambrook, Berger, additional variants can be produced by a variety of mutagenesis procedures. Many such procedures are known in the art, including site directed mutagenesis, oligonucleotide-directed mutagenesis, and many others. For example, site directed mutagenesis is described, e.g., in Smith (1985) *In vitro mutagenesis Ann. Rev. Genet.* 19:423-462, and references therein, Botstein & Shortle (1985) *Strategies and applications of in vitro mutagenesis Science* 229:1193-1201; and Carter (1986) *Site-directed mutagenesis Biochem. J.* 237:1-7. Oligonucleotide-directed mutagenesis is described, e.g., in Zoller & Smith (1982) *Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment Nucleic Acids Res.* 10:6487-6500). Mutagenesis using modified bases is described e.g., in Kunkel (1985) *Rapid and efficient site-specific mutagenesis without phenotypic selection Proc. Natl. Acad. Sci. USA* 82:488-492, and Taylor et al. (1985) *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA Nucl. Acids Res.* 13: 8765-8787. Mutagenesis using gapped duplex DNA is described, e.g., in Kramer et al. (1984) *The gapped duplex DNA approach to oligonucleotide-directed mutation construction Nucl. Acids Res.* 12: 9441-9456). Point mismatch repair is described, e.g., by Kramer et al. (1984) *Point Mismatch Repair Cell* 38:879-887). Double-strand break repair is described, e.g., in Mandecki (1986) *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis Proc. Natl. Acad. Sci. USA,* 83:7177-7181, and in Arnold (1993) *Protein engineering for unusual environments Current Opinion in Biotechnology* 4:450-455). Mutagenesis using repair-deficient host strains is described, e.g., in Carter et al. (1985) *Improved oligonucleotide site-directed mutagenesis using M13 vectors Nucl. Acids Res.* 13: 4431-4443. Mutagenesis by total gene synthesis is described e.g., by Nambiar et al. (1984) *Total synthesis and cloning of a gene coding for the ribonuclease S protein Science* 223: 1299-1301. DNA shuffling is described, e.g., by Stemmer (1994) *Rapid evolution of a protein in vitro by DNA shuffling Nature* 370:389-391, and Stemmer (1994) *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution Proc. Natl. Acad. Sci. USA* 91:10747-10751.

[0069] Many of the above methods are further described in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods. Kits for mutagenesis, library construction and other diversity generation methods are also commercially available. For example, kits are available from, e.g., Amersham International plc (e.g., using the Eckstein method above), Anglian Biotechnology Ltd (e.g., using the Carter/Winter method above), Bio/Can Scientific, Bio-Rad (e.g., using the Kunkel method described above), Boehringer Mannheim Corp., Clonetech Laboratories, DNA Technologies, Epicentre Technologies (e.g., the 5 prime 3 prime kit); Genpak Inc, Lemargo Inc, Life Technologies (Gibco BRL), New England Biolabs, Pharmacia Biotech, Promega Corp., Quantum Biotechnologies, Stratagene (e.g., QuickChange™ site-directed mutagenesis kit; and Chameleon™ double-stranded, site-directed mutagenesis kit).

[0070] Determining Sequence Relationships

[0071] A variety of methods for determining relationships between two or more sequences (e.g., identity, similarity and/or homology) are available, and well known in the art. The methods include manual alignment, computer assisted sequence alignment and combinations thereof. A number of algorithms (which are generally computer implemented) for performing sequence alignment are widely available, or can be produced by one of skill. These methods include, e.g., the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2:482; the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443; the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. (USA)* 85:2444; and/or by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.).

[0072] For example, software for performing sequence identity (and sequence similarity) analysis using the BLAST algorithm is described in Altschul et al. (1990) *J. Mol. Biol.* 215:403-410. This software is publicly available, e.g., through the National Center for Biotechnology Information on the World Wide Web at ncbi.nlm.nih.gov. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always>0) and N (penalty score for mismatching residues; always<0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=−4, and a comparison of both strands. For amino acid sequences, the BLASTP (BLAST Protein) program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see, Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

[0073] Additionally, the BLAST algorithm performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001.

[0074] Another example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle (1987) *J. Mol. Evol.* 35:351-360. The method used is similar to the method described by Higgins & Sharp (1989) *CABIOS* 5:151-153. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster can then be aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences can be aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program can also be used to plot a dendogram or tree representation of clustering relationships. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison.

[0075] An additional example of an algorithm that is suitable for multiple DNA, or amino acid, sequence alignments is the CLUSTALW program (Thompson, J. D. et al. (1994) *Nucl. Acids. Res.* 22: 4673-4680). CLUSTALW performs multiple pairwise comparisons between groups of sequences and assembles them into a multiple alignment based on homology. Gap open and Gap extension penalties can be, e.g., 10 and 0.05 respectively. For amino acid alignments, the BLOSUM algorithm can be used as a protein weight matrix. See, e.g., Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919.

[0076] Nucleic Acid Hybridization

[0077] Similarity between nucleic acids can also be evaluated by "hybridization" between single stranded (or single stranded regions of) nucleic acids with complementary or partially complementary polynucleotide sequences. Hybridization is a measure of the physical association between nucleic acids, typically, in solution, or with one of the nucleic acid strands immobilized on a solid support, e.g., a

membrane, a bead, a chip, a filter, etc. Nucleic acid hybridization occurs based on a variety of well characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. Numerous protocols for nucleic acid hybridization are well known in the art. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), as well as in Ausubel et al. *Current Protocols in Molecular Biology* (supplemented through 2001) John Wiley & Sons, New York ("Ausubel"); Sambrook et al. *Molecular Cloning—A Laboratory Manual* (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989 ("Sambrook"), and Berger and Kimmel *Guide to Molecular Cloning Techniques Methods in Enzymology volume* 152 Academic Press, Inc., San Diego, Calif. ("Berger"). Hames and Higgins (1995) *Gene Probes* 1, IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 1) and Hames and Higgins (1995) *Gene Probes* 2, IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides.

[0078] Conditions suitable for obtaining hybridization, including differential hybridization, are selected according to the theoretical melting temperature (T_m) between complementary and partially complementary nucleic acids. Under a given set of conditions, e.g., solvent composition, ionic strength, etc., the T_m is the temperature at which the duplex between the hybridizing nucleic acid strands is 50% denatured. That is, the T_m corresponds to the temperature corresponding to the midpoint in transition from helix to random coil; it depends on length, nucleotide composition, and ionic strength for long stretches of nucleotides.

[0079] After hybridization, unhybridized nucleic acids can be removed by a series of washes, the stringency of which can be adjusted depending upon the desired results. Low stringency washing conditions (e.g., using higher salt and lower temperature) increase sensitivity, but can produce nonspecific hybridization signals and high background signals. Higher stringency conditions (e.g., using lower salt and higher temperature that is closer to the T_m) lower the background signal, typically with primarily the specific signal remaining. See, also, Rapley, R. and Walker, J. M. eds., *Molecular Biomethods Handbook* (Humana Press, Inc. 1998).

[0080] "Stringent hybridization wash conditions" or "stringent conditions" in the context of nucleic acid hybridization experiments, such as Southern and northern hybridizations, are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), supra, and in Hames and Higgins 1 and Hames and Higgins 2, supra.

[0081] An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 2×SSC, 50% formamide at 42° C., with the hybridization being carried out overnight (e.g., for approximately 20 hours). An example of stringent wash conditions is a 0.2×SSC wash at 65° C. for 15 minutes (see Sambrook, supra for a description of SSC buffer). Often, the wash determining the stringency is preceded by a low stringency wash to remove signal due to residual unhybridized probe. An example low stringency wash is 2×SSC at room temperature (e.g., 20° C. for 15 minutes).

[0082] In general, a signal to noise ratio of at least 2.5×-5× (and typically higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Detection of at least stringent hybridization between two sequences in the context of the present invention indicates relatively strong structural similarity to, e.g., the nucleic acids of the present invention provided in the sequence listings herein.

[0083] For purposes of the present invention, generally, "highly stringent" hybridization and wash conditions are selected to be about 5° C. or less lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH (as noted below, highly stringent conditions can also be referred to in comparative terms). Target sequences that are closely related or identical to the nucleotide sequence of interest (e.g., "probe") can be identified under stringent or highly stringent conditions. Lower stringency conditions are appropriate for sequences that are less complementary.

[0084] For example, in determining stringent or highly stringent hybridization (or even more stringent hybridization) and wash conditions, the hybridization and wash conditions are gradually increased (e.g., by increasing temperature, decreasing salt concentration, increasing detergent concentration and/or increasing the concentration of organic solvents, such as formamide, in the hybridization or wash), until a selected set of criteria are met. For example, the hybridization and wash conditions are gradually increased until a probe comprising one or more polynucleotide sequences of the invention, e.g., selected from SEQ ID NO:1-443, and/or complementary polynucleotide sequences binds to a perfectly matched complementary target (again, a nucleic acid comprising one or more nucleic acid sequences or subsequences selected from SEQ ID NO:1 to SEQ ID NO:443, and/or complementary polynucleotide sequences thereof), with a signal to noise ratio that is at least 2.5×, and optionally 5×, or 10×, or 100×or more as high as that observed for hybridization of the probe to an unmatched target, as desired.

[0085] Using the polynucleotides of the invention, or subsequences thereof, novel target nucleic acids can be obtained, such target nucleic acids are also a feature of the invention. For example, such target nucleic acids include sequences that hybridize under stringent conditions to a unique oligonucleotide probe corresponding to any of the polypeptides of the invention, e.g., SEQ ID NOs:1-443.

[0086] For example, hybridization conditions are chosen under which a target polynucleotide or oligonucleotide that is perfectly complementary to the oligonucleotide probe hybridizes to the probe with at least about a 5-10× higher signal to noise ratio than for hybridization of the target polynucleotide (oligonucleotide) to a control nucleic acid, e.g., a nucleic acid that is not a polynucleotide sequence of the invention (e.g., sequences unrelated to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that hybridize under stringent conditions to any one of SEQ ID NO:1-SEQ ID

NO:443, sequences that are at least about 70% identical to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that encode a polypeptide or peptide comprising a subsequence encoded by any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are physically linked in the human genome to any one of SEQ ID NO:1-SEQ ID NO:443, sequences complementary to any such sequences, or subsequences thereof).

[0087] Higher ratios of signal to noise can be achieved by increasing the stringency of the hybridization conditions such that ratios of about 15×, 20×, 30×, 50× or more are obtained. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a calorimetric label, a radio active label, or the like.

[0088] Probes

[0089] Nucleic acids including one or more polynucleotide sequence of the invention are favorably used as probes for the detection of corresponding or related nucleic acids in a variety of contexts, such as the nucleic hybridization experiments discussed above. The probes can be either DNA or RNA molecules, such as restriction fragments of genomic or cloned DNA, cDNAs, amplification products, transcripts, and oligonucleotides, and can vary in length from oligonucleotides as short as about 10 nucleotides in length to chromosomal fragments or cDNAs in excess of one or more kilobases. For example, in some embodiments, a probe of the invention includes a polynucleotide sequence or subsequence selected from among SEQ ID NO:1 to SEQ ID NO:443, or sequences complementary thereto. Alternatively, polynucleotide sequences that are variants of one of the above designated sequences are used as probes. Most typically, such variants include one or a few nucleotide variations. For example, pairs (or sets) of oligonucleotides can be selected, in which the two (or more) polynucleotide sequences are substantially identical variants of each other, wherein one polynucleotide sequence correspond identically to a first allele or allelic variant and the other(s) correspond identically to additional alleles or allelic variants. Such pairs of oligonucleotide probes are particularly useful, e.g., for allele specific hybridization experiments to detect polymorphic nucleotides. In other applications, probes are selected that are more or less divergent, that is probes that are at least about 70% (or about 80%, about 90%, about 95%, about 98%, or about 99%) identical are selected.

[0090] The probes of the invention, e.g., as exemplified by SEQ ID NO:1 through SEQ ID NO:443, can also be used to identify additional useful polynucleotide sequences according to procedures routine in the art. In one set of preferred embodiments, one or more probes, as described above, are utilized to screen libraries of expression products or chromosomal segments (i.e., expression libraries or genomic libraries) to identify clones that include sequences identical to, or with significant sequence identity to, one or more of SEQ ID NO:1-443, e.g., allelic variants, homologues or orthologues. In turn, each of these identified sequences can be used to make probes, including pairs or sets of variant probes as described above. It will be understood that in addition to such physical methods as library screening, computer assisted bioinformatic approaches, e.g., BLAST and other sequence homology search algorithms,-and the like, can also be used for identifying related or physically linked (e.g., in the human genome) polynucleotide sequences. Polynucleotide sequences identified in this manner are also a feature of the invention.

[0091] For example, oligonucleotide probes, most typically produced by well known synthetic methods, such as the solid phase phosphoramidite triester method described by Beaucage and Caruthers (1981) *Tetrahedron Letts.* 22(20):1859-1862, e.g., using an automated synthesizer, as described in Needham-VanDevanter et al. (1984) *Nucleic Acids Res.,* 12:6159-6168. Oligonucleotides can also be custom made and ordered from a variety of commercial sources known to persons of skill. Purification of oligonucleotides, where necessary, is typically performed by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson and Regnier (1983) *J. Chrom.* 255:137-149. The sequence of the synthetic oligonucleotides can be verified using the chemical degradation method of Maxam and Gilbert (1980) in Grossman and Moldave (eds.) Academic Press, New York, *Methods in Enzymology* 65:499-560. Custom oligos can also easily be ordered from a variety of commercial sources known to persons of skill.

[0092] In addition, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (available on the World Wide Web genco.com), ExpressGen Inc. (available on the World Wide Web at expressgen.com), Operon Technologies Inc. (Alameda, Calif.) and many others. Similarly, peptides and antibodies can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, inc. (available on the World Wide Web htibio.com), BMA Biomedicals Ltd (U.K.), Bio.Synthesis, Inc., and many others.

[0093] As noted, in one embodiment, oligonucleotide probes of the invention include subsequences of SEQ ID NO:1 through SEQ ID NO:443, and/or complementary sequences thereof, e.g., of at least about 10 contiguous nucleotides in length. Commonly, the oligonucleotide probes are at least about 12 contiguous nucleotides in length; usually, the oligonucleotides are at least about 14 contiguous nucleotides in length; frequently, the oligonucleotides are at least about 16 contiguous nucleotides in length, and in many cases the oligonucleotides are at least about 17 contiguous nucleotides of at least one sequence selected from SEQ ID NO:1 to SEQ ID NO:443. In some cases, the oligonucleotide probes consist of a polynucleotide sequence selected from SEQ ID NO:1 through SEQ ID NO:443.

[0094] In other circumstances, e.g., relating to functional attributes of cells or organisms expressing the polynucleotides and polypeptides of the invention, probes that are polypeptides, peptides or antibodies are favorably utilized. For example, polypeptides, polypeptide fragments and peptides encoded by or having subsequences encoded by the polynucleotides of the invention, e.g., SEQ ID NO:1 to SEQ ID NO:443, etc., are favorably used to identify and isolate antibodies or other binding proteins, e.g., from phage display libraries, combinatorial libraries, polyclonal sera, and the like.

[0095] Antibodies specific for a polypeptide subsequence encoded by any of SEQ ID NO:1 to SEQ ID NO:443 are likewise valuable as probes for evaluating expression products, e.g., from cells or tissues. In addition, antibodies are

particularly suitable for evaluating expression of proteins comprising amino acid subsequences encoded by SEQ ID Nos:1-443, in situ, in a cell, tissue or organism, e.g., an organism providing an experimental model of cholesterol homeostasis or adipogenesis. Antibodies can be directly labeled with a detectable reagent as described below, or detected indirectly by labeling of a secondary antibody specific for the heavy chain constant region (i.e., isotype) of the specific antibody. Additional details regarding production of specific antibodies are provided below in the section entitled "Antibodies."

[0096] Labeling and Detecting Probes

[0097] Numerous methods are available for labeling and detection of the nucleic acid and polypeptide (or peptide or antibody) probes of the invention, these include: 1) Fluorescence (using, e.g., fluorescein, Cy-5, rhodamine or other fluorescent tags); 2) Isotopic methods, e.g., using end-labeling, nick translation, random priming, or PCR to incorporate radioactive isotopes into the probe polynucleotide/oligonucleotide; 3) Chemifluorescence using Alkaline Phosphatase and the substrate AttoPhos (Amersham) or other substrates that produce fluorescent products; 4) Chemiluminescence (using either Horseradish Peroxidase and/or Alkaline Phosphatase with substrates that produce photons as breakdown products, kits providing reagents and protocols are available from such commercial sources as Amersham, Boehringer-Mannheim, and Life Technologies/Gibco BRL); and, 5) Colorimetric methods (again using both Horseradish Peroxidase and Alkaline Phosphatase with substrates that produce a colored precipitate, kits are available from Life Technologies/Gibco BRL, and Boehringer-Mannheim). Other methods for labeling and detection will be readily apparent to one skilled in the art.

[0098] More generally, a probe can be labeled with any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical, chemical or other available means. Useful labels in the present invention include spectral labels such as fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, ^{32}P, ^{33}P, etc.), enzymes (e.g., horse-radish peroxidase, alkaline phosphatase, etc.), spectral colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads. The label may be coupled directly or indirectly to a component of the detection assay (e.g., a probe, such as an oligonucleotide, isolated DNA, amplicon, restriction fragment, or the like) according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions. In general, a detector which monitors a probe-target nucleic acid hybridization is adapted to the particular label which is used. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters, cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill. Commonly, an optical image of a substrate comprising a nucleic acid array with particular set of probes bound to the array is digitized for subsequent computer analysis.

[0099] Because incorporation of radiolabeled nucleotides into nucleic acids is straightforward, this detection represents one favorable labeling strategy. Exemplar technologies for incorporating radiolabels include end-labeling with a kinase or phosphatase enzyme, nick translation, incorporation of radio-active nucleotides with a polymerase and many other well known strategies.

[0100] Fluorescent labels are desirable, having the advantage of requiring fewer precautions in handling, and being amenable to high-throughput visualization techniques. Preferred labels are typically characterized by one or more of the following: high sensitivity, high stability, low background, low environmental sensitivity and high specificity in labeling. Fluorescent moieties, which are incorporated into the labels of the invention, are generally are known, including Texas red, fluorescein isothiocyanate, rhodamine, etc. Many fluorescent tags are commercially available from SIGMA chemical company (Saint Louis, Mo.), Molecular Probes (Eugene, Oreg.), R&D systems (Minneapolis, Minn.), Pharmacia LKB Biotechnology (Piscataway, N.J.), CLONTECH Laboratories, Inc. (Palo Alto, Calif.), Chem Genes Corp., Aldrich Chemical Company (Milwaukee, Wis.), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersberg, Md.), Fluka Chemica-Biochemika Analytika (Fluka Chemie AG, Buchs, Switzerland), and Applied Biosystems (Foster City, Calif.) as well as other commercial sources known to one of skill. Similarly, moieties such as digoxygenin and biotin, which are not themselves fluorescent but are readily used in conjunction with secondary reagents, i.e., anti-digoxygenin antibodies, avidin (or streptavidin), that can be labeled, are suitable as labeling reagents in the context of the probes of the invention.

[0101] The label is coupled directly or indirectly to a molecule to be detected (a product, substrate, enzyme, or the like) according to methods well known in the art. As indicated above, a wide variety of labels are used, with the choice of label depending on the sensitivity required, ease of conjugation of the compound, stability requirements, available instrumentation, and disposal provisions. Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to a nucleic acid such as a probe, primer, amplicon, or the like. The ligand then binds to an anti-ligand (e.g., streptavidin) molecule which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. A number of ligands and anti-ligands can be used. Where a ligand has a natural anti-ligand, for example, biotin, thyroxine, and cortisol, it can be used in conjunction with labeled, anti-ligands. Alternatively, any haptenic or antigenic compound can be used in combination with an antibody. Labels can also be conjugated directly to signal generating compounds, e.g., by conjugation with an enzyme or fluorophore or chromophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, etc. Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, e.g., luminol. Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is

optically detectable, typical detectors include microscopes, cameras, phototubes and photodiodes' and many other detection systems which are widely available.

[0102] It will be appreciated that probe design is influenced by the intended application. For example, where several allele-specific probe-target interactions are to be detected in a single assay, e.g., on a single DNA chip, it is desirable to have similar melting temperatures for all of the probes. Accordingly, the length of the probes are adjusted so that the melting temperatures for all of the probes on the array are closely similar (it will be appreciated that different lengths for different probes may be needed to achieve a particular T_m where different probes have different GC contents). Although melting temperature is a primary consideration in probe design, other factors are optionally used to further adjust probe construction, such as selecting against primer self-complementarity and the like.

[0103] Marker Sets

[0104] Sets of probes, including multiple nucleic acids with polynucleotide sequences selected from among the polynucleotide sequences of the invention, e.g., SEQ ID NO:1 through SEQ ID NO:443, are also a feature of the invention. Such sets of probes are useful as marker sets, e.g., for evaluating alterations in cholesterol or lipid homeostasis and adipogenesis, such as conditions or characteristics associated with a physiologic or pathologic response to elevated cholesterol and/or adipogenesis, and the like. For example, the marker sets are useful in monitoring molecular events underlying a metabolic response to excessive dietary cholesterol, prior to the onset of overt symptoms of a disorder associated elevated cholesterol.

[0105] Marker sets of the invention favorably include any of the probe sequences described above, such as polynucleotide sequences that hybridize under stringent conditions to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are at least about 70% identical to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that encode a polypeptide or peptide comprising a subsequence encoded by any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are physically linked in the human genome to any one of SEQ ID NO:1-SEQ ID NO:443, as well as sequences complementary to any such sequences, or subsequences thereof.

[0106] In one embodiment, the marker set of the invention is a plurality of oligonucleotides, e.g., synthetic oligonucleotides produced by the phosporamidite triester synthesis method on an automated synthesizer, as described above. For example, at least two oligonucleotides including a polynucleotide sequence of at least about 10 contiguous nucleotides of a polynucleotide of the invention, e.g., selected from SEQ ID NO:1-SEQ ID NO:443, can be used as a set to evaluate one or more characteristics or conditions associated with responses to cholesterol or adipogenesis. Frequently, the oligonucleotides selected will be longer than 10 contiguous nucleotides in length, for example, oligonucleotides of at least about 12, or about 14, or about 16 or about 17, or more contiguous nucleotides are favorably employed in the marker sets of the invention.

[0107] While as few as one or two probes can constitute a marker set, it is frequently desirable to employ marker sets with more than two members. Typically, a marker set of the invention has at least about 3, often at least about 5 or more members selected from among any of the polynucleotides of the invention. In some embodiments, the marker sets include members corresponding to a majority of SEQ ID NO:1-SEQ ID NO:433. In one favorable embodiment, the marker set includes oligonucleotides corresponding in sequence to at least part of each of SEQ ID NO:1 through SEQ ID NO:443. In another embodiment, the marker sets are made up of expression products such as cDNAs, or amplification products corresponding to cDNA or RNA expression products.

[0108] In some applications, the marker set includes labeled nucleic acid probes as described in the preceding section. In other applications, e.g., certain array applications, a labeled nucleic acid sample is hybridized to a set of unlabeled marker nucleic acids.

[0109] The marker sets of the invention are frequently employed in the context of a polynucleotide sequence array. Any of the polynucleotide sequences of the invention, as described above, can be logically or physically arranged to produce a useful array. For example, nucleic acids, e.g., oligonucleotides, cDNAs, amplicons, or chromosomal segments, can be physically arrayed in a solid phase or liquid phase array. Common solid phase arrays include a variety of solid substrates suitable for attaching nucleic acids in an ordered manner, such as membranes, filters, chips, beads, pins, slides, plates, etc. Common liquid phase arrays include, e.g., arrays of wells (e.g., as in microtiter trays) or containers (e.g., as in arrays of test tubes).

[0110] Nucleic acids of the marker sets are optionally immobilized, for example by direct or indirect cross-linking, to the solid support. Essentially any solid support capable of withstanding the reagents and conditions used in the particular detection assay can be utilized. For example, functionalized glass, silicon, silicon dioxide, modified silicon, any of a variety of polymers, such as (poly)tetrafluoroethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, membranes (e.g., nylon or nitrocellulose), or combinations thereof, can all serve as the substrate for a solid phase array.

[0111] In one embodiment, the array is a "chip" composed, e.g., of one of the above specified materials. Polynucleotide probes, e.g., RNA or DNA, such as cDNA, synthetic oligonucleotides, and the like, as discussed above are adhered to the chip in a logically ordered manner, i.e., in an array. Additional details regarding methods for linking nucleic acids and proteins to a chip substrate, can be found in, e.g., U.S. Pat. No. 5,143,854 "Large Scale Photolithographic Solid Phase Synthesis of Polypeptides and Receptor Binding Screening Thereof" to Pirrung et al., issued Sep. 1, 1992; U.S. Pat. No. 5,837,832 "Arrays of Nucleic Acid Probes on Biological Chips" to Chee et al., issued Nov. 17, 1998; U.S. Pat. No. 6,087,112 "Arrays with Modified Oligonucleotide and Polynucleotide Compositions" to Dale, issued Jul. 11, 2000; U.S. Pat. No. 5,215,882 "Method of Immobilizing Nucleic Acid on a Solid Substrate for Use in Nucleic Acid Hybridization Assays" to Bahl et al., issued Jun. 1, 1993; U.S. Pat. No. 5,707,807 "Molecular Indexing for Expressed Gene Analysis" to Kato, issued Jan. 13, 1998; U.S. Pat. No. 5,807,522 "Methods for Fabricating Microarrays of Biological Samples" to Brown et al., issued Sep. 15, 1998; U.S. Pat. No. 5,958,342 "Jet Droplet Device" to Gamble et al., issued Sep. 28, 1999; U.S. Pat. No. 5,994,076 "Methods of Assaying Differential Expression" to Chenchik et al., issued Nov.

30, 1999; U.S. Pat. No. 6,004,755 "Quantitative Microarray Hybridization Assays" to Wang, issued Dec. 21, 1999; U.S. Pat. No. 6,048,695 "Chemically Modified Nucleic Acids and Method for Coupling Nucleic Acids to Solid Support" to Bradley et al., issued Apr. 11, 2000; U.S. Pat. No. 6,060,240 "Methods for Measuring Relative Amounts of Nucleic Acids in a Complex Mixture and Retrieval of Specific Sequences Therefrom" to Kamb et al., issued May 9, 2000; U.S. Pat. No. 6,090,556 "Method for Quantitatively Determining the Expression of a Gene" to Kato, issued Jul. 18, 2000; and U.S. Pat. No. 6,040,138 "Expression Monitoring by Hybridization to High Density Oligonucleotide Arrays" to Lockhart et al., issued Mar. 21, 2000.

[0112] In addition to being able to design, build and use probe arrays using available techniques, one of skill can simply order custom-made arrays and array-reading devices from manufacturers specializing in array manufacture. For example, Affymetrix Corp., in Santa Clara, Calif. manufactures DNA VLSIP™ arrays.

[0113] In addition to marker sets made up of nucleic acid probes described above, marker sets including polypeptide, peptide, and antibody probes as discussed in the section entitled "Labeled probes" are favorably used in certain applications. As discussed above for individual peptide or polypeptide probes, sets of probes including multiple members encoded by or having subsequences encoded by the polynucleotides of the invention, e.g., SEQ ID NO:1-443, or antibodies specific to such sequences can be used in liquid phase, or immobilized as described above with respect to nucleic acid markers.

[0114] Vectors, Promoters and Expression Systems

[0115] The present invention includes recombinant constructs incorporating one or more of the nucleic acid sequences described above. Such constructs include a vector, for example, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), etc., into which one or more of the polynucleotide sequences of the invention, e.g., comprising any of SEQ ID NO:1-443, or a subsequence thereof, has been inserted, in a forward or reverse orientation. For example, the inserted nucleic acid can include a chromosomal sequence or cDNA including a all or part of at least one of the polynucleotide sequences of the invention. In a preferred embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

[0116] The polynucleotides of the present invention can be included in any one of a variety of vectors suitable for generating sense or antisense RNA, and optionally, polypeptide (or peptide) expression products. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses and many others. Any vector that is capable of introducing genetic material into a cell, and, if replication is desired, which is replicable in the relevant host can be used.

[0117] In an expression vector, the polynucleotide sequence of interest is physically arranged in proximity and orientation to an appropriate transcription control sequence (promoter, and optionally, one or more enhancers) to direct mRNA synthesis. That is, the polynucleotide sequence of interest is operably linked to an appropriate transcription control sequence. Examples of such promoters include: LTR or SV40 promoter, *E. coli* lac or trp promoter, phage lambda PL promoter, and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation, and a transcription terminator. The vector optionally includes appropriate sequences for amplifying expression. In addition, the expression vectors optionally comprise one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells, such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli.*

[0118] Additional Expression Elements

[0119] Where translation of polypeptide encoded by a nucleic acid comprising a polynucleotide sequence of the invention is desired, additional translation specific initiation signals can improve the efficiency of translation. These signals can include, e.g., an ATG initiation codon and adjacent sequences. In some cases, for example, full-length cDNA molecules or chromosomal segments including a coding sequence incorporating, e.g., a polynucleotide sequence of the invention, a translation initiation codon and associated sequence elements are inserted into the appropriate expression vector simultaneously with the polynucleotide sequence of interest. In such cases, additional translational control signals are not required. However, in cases where only a polypeptide coding sequence, or a portion thereof, is inserted, exogenous translational control signals, including an ATG initiation codon is provided for expression of the relevant sequence. The initiation codon is put in the correct reading frame to ensure transcription of the polynucleotide sequence of interest. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf D et al. (1994) *Results Probi Cell Differ* 20:125-62; Bittner et al. (1987) *Methods in Enzymol* 153:516-544).

[0120] Expression Hosts

[0121] The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention by recombinant techniques. Host cells are genetically engineered (i.e., transduced, transformed or transfected) with a vector, such as an expression vector, of this invention. As described above, the vector can be in the form of a plasmid, a viral particle, a phage, etc. Examples of appropriate expression hosts include: bacterial cells, such as *E. coli*, Streptomyces, and *Salmonella typhimurium*; fungal cells, such as *Saccharomyces cerevisiae, Pichia pastoris*, and *Neurospora crassa*; insect cells such as Drosophila and *Spodoptera frugiperda*; mammalian cells such as 3T3, COS, CHO, BHK, HEK 293 or Bowes melanoma; plant cells, etc.

[0122] The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the inserted polynucleotide sequences. The culture conditions,

such as temperature, pH and the like, are typically those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, e.g., Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique, third edition*, Wiley-Liss, New York and the references cited therein. Expression products corresponding to the nucleic acids of the invention can also be produced in non-animal cells such as plants, yeast, fungi, bacteria and the like. In addition to Sambrook, Berger and Ausubel, details regarding cell culture can be found in Payne et al. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla.

[0123] In bacterial systems, a number of expression vectors can be selected depending upon the use intended for the expressed product. For example, when large quantities of a polypeptide or fragments thereof are needed for the production of antibodies, vectors which direct high level expression of fusion proteins that are readily purified are favorably employed. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the coding sequence of interest, e.g., a polynucleotide of the invention as described above, can be ligated into the vector in-frame with sequences for the amino-terminal translation initiating Methionine and the subsequent 7 residues of beta-galactosidase producing a catalytically active beta galactosidase fusion protein; pIN vectors (Van Heeke & Schuster (1989) *J Biol Chem* 264:5503-5509); pET vectors (Novagen, Madison Wis.); and the like.

[0124] Similarly, in the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH can be used for production of the desired expression products. For reviews, see Berger, Ausubel, and, e.g., Grant et al. (1987; *Methods in Enzymology* 153:516-544).

[0125] In mammalian host cells, a number of expression systems, such as viral-based systems, can be utilized. For example, in cases where an adenovirus is used as an expression vector, a coding sequence is optionally ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing the polypeptides of interest in infected host cells (Logan and Shenk (1984) *Proc Natl Acad Sci* 81:3655-3659). In addition, transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, can be used to increase expression in mammalian host cells.

[0126] Transformed or transfected host cells containing the expression vectors described above are also a feature of the invention. The host cell can be a eukaryotic cell, such as a mammalian cell, a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, electroporation, or other common techniques (Davis, L., Dibner, M., and Battey, I. (1986) *Basic Methods in Molecular Biology*).

[0127] A host cell strain is optionally chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a precursor form into a mature form of the protein is sometimes important for correct insertion, folding and/or function. Different host cells such as 3T3, COS, CHO, HeLa, ByK, MDCK, 293, W138, etc. have specific cellular machinery and characteristic mechanisms for such post-translational activities and can be chosen to ensure the correct modification and processing of the introduced, foreign protein.

[0128] For long-term, high-yield production of recombinant proteins encoded by or having subsequences encoded by the polynucleotides of the invention, stable expression systems are typically used. For example, cell lines which stably express a polypeptide of the invention are transfected using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells are allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. For example, resistant groups or colonies of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

[0129] Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell can be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used.

[0130] Polypeptide Production and Recovery

[0131] Following transduction of a suitable host cell line or strain and growth of the host cells to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. The secreted polypeptide product is then recovered from the culture medium. Alternatively, cells can be harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Eukaryotic or microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well know to those skilled in the art.

[0132] Expressed polypeptides can be recovered and purified from recombinant cell cultures by any of a number of methods well known in the art, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography (e.g., using any of the tagging systems noted herein), hydroxylapatite chromatography, and lectin chromatography. Protein refolding steps can be used, as desired, in completing configuration of the mature protein. Finally,

high performance liquid chromatography (HPLC) can be employed in the final purification steps. In addition to the references noted above, a variety of purification methods are well known in the art, including, e.g., those set forth in Sandana (1997) *Bioseparation of Proteins*, Academic Press, Inc.; and Bollag et al. (1996) *Protein Methods, 2nd Edition* Wiley-Liss, NY; Walker (1996) *The Protein Protocols Handbook* Humana Press, NJ, Harris and Angal (1990) *Protein Purification Applications: A Practical Approach* IRL Press at Oxford, Oxford, U.K.; Scopes (1993) *Protein Purification: Principles and Practice 3rd Edition* Springer Verlag, NY; Janson and Ryden (1998) *Protein Purification: Principles, High Resolution Methods and Applications, Second Edition* Wiley-VCH, NY; and Walker (1998) *Protein Protocols* on CD-ROM Humana Press, NJ.

[0133] Alternatively, cell-free transcription/translation systems can be employed to produce polypeptides comprising an amino acid sequence or subsequence encoded by the polynucleotides of the invention. A number of suitable in vitro transcription and translation systems are commercially available. A general guide to in vitro transcription and translation protocols is found in Tymms (1995) *In vitro Transcription and Translation Protocols: Methods in Molecular Biology Volume* 37, Garland Publishing, NY.

[0134] In addition, the polypeptides, or subsequences thereof, e.g., subsequences comprising antigenic peptides, can be produced manually or by using an automated system, by direct peptide synthesis using solid-phase techniques (see, Stewart et al. (1969) *Solid-Phase Peptide Synthesis*, W H Freeman Co, San Francisco; Merrifield J (1963) *J. Am. Chem. Soc.* 85:2149-2154). Exemplary automated systems include the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.). If desired, subsequences can be chemically synthesized separately, and combined using chemical methods to provide full-length polypeptides.

[0135] Conservatively Modified Variations

[0136] The polypeptides of the present invention include conservatively modified variations of the polypeptides comprising subsequences encoded by a polynucleotide of the invention, e.g., SEQ ID NOs: 1-443. Such conservatively modified variations comprise substitutions, additions or deletions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 4%, about 2%, or about 1%) in a polypeptide encoded by a polynucleotide of the invention. Typically, substitutions of amino acids are conservative substitutions according to the six substitution groups set forth in Table 1 (supra).

[0137] Conservative variations also include the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional sequence.

[0138] The polypeptides of the invention, including conservatively substituted sequences, can be present as part of larger polypeptide sequences such as occur upon the addition of one or more domains for purification of the protein (e.g., poly his segments, FLAG tag segments, etc.), e.g., where the additional functional domains have little or no effect on the activity of the protein, or where the additional domains can be removed by post synthesis processing steps such as by treatment with a protease.

[0139] Modified Amino Acids

[0140] Expressed polypeptides of the invention can contain one or more modified amino acid. The presence of modified amino acids can be advantageous in, for example, (a) increasing polypeptide serum half-life, (b) reducing polypeptide antigenicity, (c) increasing polypeptide storage stability. Amino acid(s) are modified, for example, co-translationally or post-translationally during recombinant production (e.g., N-linked glycosylation at N-X-S/T motifs during expression in mammalian cells) or modified by synthetic means (e.g., via PEGylation).

[0141] Non-limiting examples of a modified amino acid include a glycosylated amino acid, a sulfated amino acid, a prenlyated (e.g., farnesylated, geranylgeranylated) amino acid, an acetylated amino acid, an acylated amino acid, a PEG-ylated amino acid, a biotinylated amino acid, a carboxylated amino acid, a phosphorylated amino acid, and the like, as well as amino acids modified by conjugation to, e.g., lipid moieties or other organic derivatizing agents. References adequate to guide one of skill in the modification of amino acids are replete throughout the literature. Example protocols are found in Walker (1998) *Protein Protocols* on CD-ROM Human Press, Towata, N.J.

[0142] Antibodies

[0143] The polypeptides of the invention can be used to produce antibodies specific for the polypeptides comprising amino acid sequences or subsequences encoded by the polynucleotides of the invention. Antibodies specific for antigenic peptides encoded by, e.g., SEQ ID NOs:1-443, and related variant polypeptides are useful, e.g., for diagnostic and therapeutic purposes, e.g., related to the activity, distribution, and expression of target polypeptides. For example, antibodies that block receptor binding, are useful for certain therapeutic applications.

[0144] Antibodies specific for the polypeptides of the invention can be generated by methods well known in the art. Such antibodies can include, but are not limited to, polyclonal, monoclonal, chimeric, humanized, single chain, Fab fragments and fragments produced by an Fab expression library.

[0145] Polypeptides do not require biological activity for antibody production. However, the polypeptide or oligopeptide is antigenic. Peptides used to induce specific antibodies typically have an amino acid sequence of at least about 5 amino acids, and often at least about 10 or about 20 amino acids. Short stretches of a polypeptide, e.g., encoded by a polynucleotide of the invention such a sequence selected from SEQ ID NO:1-SEQ ID NO:443, can be fused with another protein, such as keyhole limpet hemocyanin (KLH), and antibodies produced against the chimeric molecule.

[0146] Numerous methods for producing polyclonal and monoclonal antibodies are known to those of skill in the art, and can be adapted to produce antibodies specific for the polypeptides or peptides of the invention. See, e.g., Coligan (1991) *Current Protocols in Immunology* Wiley/Greene, NY; and Harlow and Lane (1989) *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY; Stites et al. (eds.) *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, Calif., and references cited therein; Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, N.Y.; *Funda-*

mental Immunology, e.g., 4th Edition (or later),W. E. Paul (ed.), Raven Press, N.Y. (1998); and Kohler and Milstein (1975) *Nature* 256: 495-497. Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors. See, Huse et al. (1989) *Science* 246: 1275-1281; and Ward, et al. (1989) *Nature* 341: 544-546. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 μM, preferably at least about 0.01 μM or better, and most typically and preferably, 0.001 μM or better.

[0147] For certain therapeutic applications, humanized antibodies are desirable. Detailed methods for preparation of chimeric (humanized) antibodies can be found in U.S. Pat. No. 5,482,856. Additional details on humanization and other antibody production and engineering techniques can be found in Borrebaeck (ed) (1995) *Antibody Engineering, 2nd Edition* Freeman and Company, NY (Borrebaeck); McCafferty et al. (1996) *Antibody Engineering, A Practical Approach* IRL at Oxford Press, Oxford, England (McCafferty), and Paul (1995) *Antibody Engineering Protocols* Humana Press, Towata, N.J. (Paul). Additional details regarding specific procedures can be found, e.g., in Ostberg et al. (1983), *Hybridoma* 2: 361-367, Ostberg, U.S. Pat. No. 4,634,664, and Engelman et al., U.S. Pat. No. 4,634,666.

[0148] Defining Polypeptides by Immunoreactivity

[0149] The polypeptides of the invention listed in the sequence listing herein, as well as novel variants derived therefrom, which are also encompassed within the present invention, provide a variety of structural features which can be recognized, e.g., in immunological assays. The generation of antisera which specifically binds the polypeptides of the invention, as well as the polypeptides which are bound by such antisera, are a feature of the invention.

[0150] The invention includes polypeptides that specifically bind to or that are specifically immunoreactive with an antibody or antisera generated against an immunogen comprising an amino acid sequence encoded by a polynucleotide of the invention. To eliminate cross-reactivity with non-related polypeptides, the antibody or antisera can be subtracted with unrelated polypeptides or proteins.

[0151] In one typical format, the immunoassay uses a polyclonal antiserum which was raised against one or more polypeptide comprising a sequence or subsequence encoded by one or more of the polynucleotides of the invention, such as SEQ ID NOs:1-443. Such an antigenic peptide or polypeptide is referred to as an "immunogenic polypeptide." The resulting antisera is optionally selected to have low cross-reactivity against unrelated polypeptides, e.g., BSA, and any such cross-reactivity can be removed by immunoabsorbtion with one or more of the unrelated polypeptides, or protein preparations, prior to use of the polyclonal antiserum in the immunoassay.

[0152] In order to produce antisera for use in an immunoassay, one or more of the immunogenic polypeptides is produced and purified as described herein. For example, recombinant protein may be produced in a mammalian cell line. An inbred strain of mice (used in this assay because results are more reproducible due to the virtual genetic identity of the mice) is immunized with the immunogenic protein(s) in combination with a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see, Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a standard description of antibody generation, immunoassay formats and conditions that can be used to determine specific immunoreactivity). Alternatively, one or more synthetic or recombinant polypeptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

[0153] Polyclonal sera are collected and titered against the immunogenic polypeptide in an immunoassay, for example, a solid phase immunoassay with one or more of the immunogenic proteins immobilized on a solid support. Polyclonal antisera with a titer of 10^6 or greater are selected, pooled and subtracted with the control unrelated polypeptides to produce subtracted pooled titered polyclonal antisera.

[0154] If desired, the subtracted pooled titered polyclonal antisera are tested for cross reactivity against any unrelated polypeptides. Discriminatory binding conditions are determined for the subtracted titered polyclonal antisera which result in at least about a 5-10 fold higher signal to noise ratio for binding of the titered polyclonal antisera to the immunogenic polypeptide of interest as compared to binding to the unrelated polypeptide. That is, the stringency of the binding reaction is adjusted by the addition of non-specific competitors such as albumin or non-fat dry milk, or by adjusting salt conditions, temperature, or the like. These binding conditions are used in subsequent assays for determining whether a test polypeptide is specifically bound by the pooled subtracted polyclonal antisera. In particular, test polypeptides which show at least a 2-5× and preferably 10× or higher signal to noise ratio than the control polypeptides under discriminatory binding conditions, and at least about a ½ signal to noise ratio as compared to the immunogenic polypeptide(s) (and typically 90% or more of the signal to noise ratio shown for the immunogenic peptide), shares substantial structural similarity with the immunogenic polypeptide as compared to unrelated polypeptides, and is, therefore, a polypeptide of the invention.

[0155] Such methods are also useful for detecting an unknown test protein or polypeptide, which is also specifically bound by the antisera under conditions as described above. In one format, the immunogenic polypeptide(s) are immobilized to a solid support which is exposed to the subtracted pooled antisera. Test proteins are added to the assay to compete for binding to the pooled subtracted antisera. The ability of the test protein(s) to compete for binding to the pooled subtracted antisera as compared to the immobilized protein(s) is compared to the ability of the immunogenic polypeptide(s) added to the assay to compete for binding (the immunogenic polypeptides compete effectively with the immobilized immunogenic polypeptides for binding to the pooled antisera). The percent cross-reactivity for the test proteins is calculated, using standard calculations.

[0156] In a parallel assay, the ability of the control proteins to compete for binding to the pooled subtracted antisera is determined as compared to the ability of the immunogenic polypeptide(s) to compete for binding to the antisera. Again, the percent cross-reactivity for the control polypeptides is calculated, using standard calculations. Where the percent cross-reactivity is at least 5-10× as high for the test polypeptides, the test polypeptides are said to specifically bind the pooled subtracted antisera.

[0157] In general, the immunoabsorbed and pooled antisera can be used in a competitive binding immunoassay as described herein to compare any test polypeptide to the immunogenic polypeptide(s). In order to make this comparison, the two polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to inhibit 50% of the binding of the subtracted antisera to the immobilized protein is determined using standard techniques. If the amount of the test polypeptide required is less than twice the amount of the immunogenic polypeptide that is required, then the test polypeptide is said to specifically bind to an antibody generated to the immunogenic protein, provided the amount is at least about 5-10× as high as for a control polypeptide.

[0158] As a final determination of specificity, the pooled antisera is optionally fully immunosorbed with the immunogenic polypeptide(s) (rather than the control polypeptides) until little or no binding of the resulting immunogenic polypeptide subtracted pooled antisera to the immunogenic polypeptide(s) used in the immunosorbtion is detectable. This fully immunosorbed antisera is then tested for reactivity with the test polypeptide. If little or no reactivity is observed (i.e., no more than 2× the signal to noise ratio observed for binding of the fully immunosorbed antisera to the immunogenic polypeptide), then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

[0159] Evaluating Metabolic Responses to Cholesterol and Adipogenesis

[0160] The probes and marker sets of the invention are favorably employed in methods for evaluating metabolic and genetic responses to cholesterol (and lipid) and adipogenesis in a subject, such as a patient undergoing medical evaluation for one or more conditions or characteristics associated with elevated cholesterol, e.g., through high dietary cholesterol or fat consumption, and/or adipogenesis, such as obesity, atherosclerosis, diabetes mellitus and coronary artery heart disease. Nucleic acids of a marker set or individual probes including one or more polynucleotides of the invention, as described, e.g., in the section entitled "Probes," are hybridized, e.g., as an array, to a DNA or RNA sample from a subject cell or tissue sample. Upon hybridization of the sample to at least a subset of the probes, a signal is detected corresponding to at least one polymorphic nucleic acid or to expression or activity of an expression product correlatable to the condition or characteristic of interest. When expression is detected, the evaluation can be made on a qualitative basis, that is, detecting whether or not an expression product (or multiple expression products) are expressed in a subject cell or tissue sample. Alternatively, the evaluation can be quantitative, determining whether levels of one or more expression products increase or decrease.

[0161] While a variety of biological samples reflective of cholesterol metabolism and/or adipogenesis can be employed, the subject sample is usually selected for ease of acquisition and to minimize invasiveness of the collection procedure to the subject. Thus, in the context of human subjects, peripheral blood samples, spinal fluid and needle biopsies from liver and adipose tissue are preferred samples, and can be obtained by well-known procedures. In the case of certain experimental applications, e.g., using animal models, alternative samples are preferred, e.g., one or more cell-types selected from the group comprising liver, adipose tissue, gall bladder, pancreas, monocytes, macrophages, foam cells, T cells, endothelia and smooth muscle derived from blood vessels and gut, fibroblasts, glia and nerve cells, etc.

[0162] For example, a marker set including a plurality of the polynucleotides of the invention, can be hybridized individually, or as an array, to an RNA or cDNA sample produced, e.g., by a reverse transcription-polymerase chain reaction (RT-PCR), from a subject RNA sample. Typically, prior to hybridization of the probes or array to a subject or "test" sample, the probe or array is validated and/or calibrated by comparing samples obtained from classes of subjects known to differ in status with respect to the characteristic or condition, e.g., obesity, atherosclerosis, diabetes, coronary artery heart disease. For example, subjects shown, e.g., by metabolic assays or phenotypic evaluation, to be at enhanced risk of one or more of the conditions of interest are compared to subjects that show no increased risk relative to the general population.

[0163] Alternatively, a marker set including a plurality of antibodies, or other binding proteins, specific for a polypeptide or peptide encoded by a polynucleotide of the invention are employed as individual probes or marker sets to evaluate expression of corresponding target proteins in a cell or tissue sample. In this case, rather than, or in addition to, preparing RNA from a sample, proteins are recovered and exposed to the probe or marker set of antibodies, in liquid phase or with either the target of antibody immobilized on a solid substrate, such as a solid phase array.

[0164] Patterns of expression correlatable to one or more of the conditions of interest are detected by hybridization to one or more probes. In some embodiments, a single probe with a high predictive value is favored, e.g., for ease of handling and cost containment. In other embodiments multiple probes, e.g., the entire marker set, are preferred, e.g., to increase sensitivity or diagnostic or prognostic value. Optimal probes and marker sets are readily ascertained on an empirical basis.

[0165] Alternatively, an oligonucleotide or polynucleotide probe that detects sequence polymorphisms rather than expression differences between subjects with different characteristics relative to a condition of interest (e.g., obesity, atherosclerosis, diabetes, coronary artery heart disease) can be used. Polymorphisms at a nucleotide level can correspond either directly or indirectly to the gene of interest underlying the condition of interest, and can be detected in any of several ways, for example, as restriction fragment length polymorphisms, by allele specific hybridization, as amplification length polymorphisms, and the like.

[0166] For example, oligonucleotide probes including variants of a polynucleotide sequences are selected that correspond to polymorphic variations in a target sequence. For example, a probe pair incorporating a single variant nucleotide can be designed to hybridize under allele specific hybridization conditions to allelic target sequences in which one allele is indicative of a condition of interest and the other allele indicates, e.g., an absence of the specified condition or characteristic. For example probe sequences are selected from among SEQ ID NO:1 through SEQ ID NO:443 (or other polynucleotides of the invention) and variants thereof.

In some instances, for example, where the cDNA or chromosomal segment has been sequenced and a particular nucleotide polymorphism is associated with a condition or characteristic of interest, the probes are chosen to detect the nucleotide polymorphism, e.g., by allele specific hybridization.

[0167] Modulating Responses to Cholesterol and Adipogenesis in a Cell or Tissue

[0168] The invention also provides experimental and therapeutic methods for modulating physiologic and pathologic responses to cholesterol and lipid and/or adipogenesis in vitro and in vivo. Tissue culture and animal models useful for elucidating the molecular mechanisms underlying metabolic responses, e.g., to elevated cholesterol and lipid, and adipogenesis (and associated physiological and pathological conditions) as well as for screening and evaluating potential therapeutic targets are produced by modulating expression or activity of polypeptides encoded by the polynucleotides of the invention.

[0169] For example, mammalian cells in culture are transfected with a nucleic acid, e.g., comprising a polynucleotide sequence selected from SEQ ID NO:1 through SEQ ID NO:443, to produce cells that express a polypeptide involved in cholesterol homeostasis and/or adipogenesis. It will be understood, that where exogenous polynucleotide sequences are introduced into cells, tissues or organisms, that the polynucleotide sequences can be selected from among any one of SEQ ID NO:1-SEQ ID NO:443, sequences that hybridize under stringent conditions to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are at least about 70% identical to any one of SEQ ID NO:1-SEQ ID NO:443, sequences that encode a polypeptide or peptide comprising a subsequence encoded by any one of SEQ ID NO:1-SEQ ID NO:443, sequences that are physically linked in the human genome to any one of SEQ ID NO:1-SEQ ID NO:443, sequences complementary to any such sequences, or subsequences thereof.

[0170] In some cases, it is preferable to link the polynucleotide sequence of interest to the regulatory sequences with which it is typically associated in vivo in nature. Alternatively, in cases where constitutive expression at levels that are in excess of those found in nature is desired, exogenous promoters and enhancers can be employed, as described in detail in the section entitled "Vectors, Promoters and Expression Systems."

[0171] Expression and/or activity of the gene or polypeptide can also be modulated in a negative manner, that is, suppressed. For example, knock out mutations can be produced by homologous recombination of an exogenous gene homologue, e.g., bearing stop codon, and/or insertion of, e.g., a selectable marker, that disrupts production of an intact transcript. Alternatively, vectors incorporating the sequence of interest in the antisense orientation can be introduced to suppress translation at a post-transcriptional level.

[0172] Alternatively, cell lines that express polypeptides comprising a subsequence encoded by a polynucleotide of the invention into which vectors have been transduced that randomly activate expression of associated endogenous sequences upon integration can be isolated. Such vectors have been described, e.g., by Harrington et al. (2001) *Creation of genome-wide protein expression libraries using random activation of gene expression Nature Biotechnology* 19: 440-445, which is incorporated herein by reference. Typically, the vector is constructed with a strong exogenous promoter linked to an exon and an unpaired splice donor site. Upon integration into the genome, splicing with a proximal splice-acceptor site occurs activating expression of a chimeric transcript encoding at least a portion of the endogenous gene. Cells expressing a polypeptide of interest can be selected by well known methods, including those based on phenotypic screening methods, antibody or receptor binding, RNA analytical methods, e.g., RT-PCR, northern analysis, MPSS, and the like. By preference, the screening is performed in a high-throughput format.

[0173] In certain embodiments, modulation of expression or activity of the polypeptide encoded by the transfected polynucleotide contributes to a detectable alteration in phenotype indicative of at least one condition associated with elevated cholesterol or lipid and/or adipogenesis. Thus, in one preferred embodiment, modulation of expression or activity of a polypeptide encoded by a polynucleotide of the invention is achieved by inducing or suppressing expression of the polynucleotide or by introducing a mutation that results in an increase or decrease in the activity of the encoded polypeptide.

[0174] The above-described methods for producing cell culture model systems can be adapted for use in the screening or monitoring of therapeutic or dietary interventions, e.g., aimed at regulating cholesterol or lipid levels or adipogenesis (for example, in an experimental model or subject exhibiting or at increased risk (based on genetic or environmental, e.g., dietary factors) for one or more conditions or characteristics such as obesity, atherosclerosis, diabetes and/or coronary artery heart disease. For example, it is desirable to select promoters and enhancers that are modulated in response to cholesterol, e.g. those regulated by the SREBP family of transcription factors. One such promoter is associated with the 3-hydroxy-3methylgutaryl CoA reductase (HMG CoA reductase) gene, which is the target of cholesterol mediated feedback regulation in vivo. Other promoters regulated by SREBP's include the promoters associated with genes encoding LDL receptor, HMG-CoA synthase, farnesyl diphosphate synthase, squalene synthase, acetyl-CoA carboxlyase, fatty acid synthase, stearoyl-CoA desaturase 1, stearoyl-CoA desaturase 2, glycerol-3-phosphate acyltransferase, and ATP-citrate lyase. See e.g. Edwards et al. (2000), *Biochimica et Biophysica Acta* 1529:103-113.

[0175] Following treatment with cholesterol, cholesterol analogues, cholesterol precursors, e.g., mevalonate, or other molecules that regulate cholesterol biosynthesis, e.g., statin drugs altered expression or activity can be detected at the RNA or protein level. Detection of altered levels of RNA is most conveniently accomplished by such methods as RT-PCR, MPSS, or northern analysis. Protein expression is conveniently monitored using, e.g., antibody based detection methods, such as ELISA's, immunoprecipitations, or immunohistochemical methods including Western analysis. In each of these procedures, the sample including the expressed protein of interest is reacted with an antibody (e.g., monoclonal antibody) or antiserum specific for the protein of interest. Methods for generating specific antibodies are well known and further details are provided above in the section entitled "Antibodies."

[0176] The cell culture models can be used to identify pharmaceutical agents capable of favorably regulating the expression or activity of a polypeptide of interest, e.g., a polypeptide comprising an amino acid sequence or subsequence encoded by a polynucleotide of the invention, in a cell culture system as described above. Most typically, this involves exposing the cells to a chemical or biological composition, e.g., a small organic molecule, or biological macromolecule such as a protein, e.g., an antibody, binding protein, or macromolecular cofactor, e.g., an apolipoprotein. Following exposure to the one or more compositions, for example, members of a chemical or biological composition library, such as a combinatorial chemical library, a library of peptide or polypeptide products expressed from a library of nucleic acids, an antibody (or other polypeptide) display library such as a phage display library, etc., modulation of the polypeptide of interest is detected. As discussed above, modulation of the polypeptide can be detected as an alteration in expression at the level of transcription or translation, or as an alteration in the activity of the encoded protein or polypeptide. In some instances, it is desirable to monitor expression or activity of multiple expression products in the same cell, or cell line. The monitored expression products, can be exogenous, i.e., introduced as described above, or endogenous, such as transcripts or polypeptides whose expression or activity is dependent on the amount or activity of the polypeptide of interest.

[0177] In cases where the expression or activity of multiple products are of interest, or where the effect of a plurality of different compounds on the expression or activity of one or more expression products, e.g., screening for pharmaceutical agents as described above, the monitoring assay is conveniently performed in an array. For example, cells can be arrayed by aliquoting into the wells of a multiwell plate, e.g., a 96, 384, 1536, or other convenient format selected according to available equipment. The arrayed cells can exposed to members of a composition library, and the cells sampled and monitored by, e.g., FACS, immunohistochemisty, ELISA, etc. Alternatively, nucleic acids or proteins can be prepared from the arrayed cells, in a manual, semi-automatic or automated procedure, and the products arranged in a liquid or solid phase array for evaluation. Additional details regarding arrays are provided above in the section entitled "Marker Sets." Alternative high throughput processing methods, such as microfluidic devices, are also available, and can favorably be employed in the context of monitoring modulation of expression products.

[0178] Typically, when processing and evaluating large numbers of samples, e.g., in a high throughput assay, data relating to expression or activity is recorded in a database, typically the database includes a character strings representing the data recorded on a computer or in a computer readable medium.

[0179] In addition to tissue culture systems, transgenic animals, most typically non-human mammals, can be produced which have integrated one or more of the polynucleotide sequences of the invention, e.g., comprising a subsequence selected from SEQ ID NO:1 to SEQ ID NO:443. In this context, commonly used experimental animals include, e.g., mouse, rat, rabbit (e.g., New Zealand White), dog, pig, sheep, or a non-human primate. In some cases the animal of choice has a naturally occurring or introduced mutation in a gene which encodes a protein involved in cholesterol homeostasis (e.g., an ApoE deficient mouse).

[0180] Such transgenic animal models are useful, in addition to the cultured cells discussed above, for the evaluation of pharmaceutical agents suitable for the modulation of cholesterol and lipid homeostasis and/or adipogenesis. For example, such transgenic animal models are useful for evaluating the ability of pharmaceutical agents to modulate a physiologic or pathologic response to elevated cholesterol and/or lipid. Transgenic animal models, e.g., expressing a polypeptide comprising a sequence or subsequence encoded by a polynucleotide of the invention are also suitable for evaluating dietary interventions aimed at regulating cholesterol or lipid homeostasis and/or adipogenesis. For example, following administration of a defined diet to a transgenic animal expressing a polypeptide of the invention, cholesterol homeostasis and/or adipogenesis and/or related conditions or characteristics are monitored. Monitoring can involve detecting altered expression or activity of an expression product encoded by one or more polynucleotide of the invention as discussed above. Alternatively, standard clinical laboratory methods for detecting and evaluating cholesterol, triglycerides, and lipoprotein profiles in the serum can be utilized. Such assays can also be adapted to evaluate cholesterol quantity and composition in other tissues and organs, e.g., liver, adipose tissue, etc. In some cases phenotypic measures of, e.g., body weight and composition (e.g., fat/lean body mass ratio), and number and function of adipocytes can be monitored.

[0181] Administration in Patients

[0182] In one aspect, the present invention provides for the administration of one or more of the polynucleotides of the invention, e.g., for gene therapy and/or for the administration of a protein herein as a prophylactic or therapeutic agent to a subject, including, e.g., a mammal, including, e.g., a human, a non-human primate, a mouse, a pig, a cow, a goat, a rabbit, a rat, a guinea pig, a hamster, a horse, and/or a sheep, exhibiting or at risk for a condition or disease associated with excessive cholesterol exposure and/or adipogenesis.

[0183] Whether the therapeutic agent is a nucleic acid, a protein or a modulator of an activity of a nucleic acid or protein, administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering compositions in the context of the present invention to a patient are available, and, although more than one route can be used to administer a particular composition, a particular route can provide a more immediate and more effective reaction than another route.

[0184] The invention also includes compositions comprising any nucleic acid or any isolated or recombinant polypeptide described above and an excipient, e.g., a pharmaceutically acceptable excipient. Transgenic animals, which include any nucleic acid or polypeptide above, e.g., produced by introduction of the vector, are also a feature of the invention. Methods for a remedying or ameliorating a condition associated with elevated cholesterol and/or lipid and/or dysfunctional adipogenesis by administering to a patient an effective amount of at least one expression vector and/or an effective amount of at least one isolated or recombinant polypeptide described above are also included in the present invention.

[0185] Pharmaceutically acceptable excipient or carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention.

[0186] Formulations suitable for parenteral administration, such as, for example, by intraarticular (in the joints), intravenous, intramuscular, intradermal, subdermal, intraperitoneal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. Parenteral administration and intravenous administration are one class of preferred methods of administration. Formulations can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials.

[0187] Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets. Cells transduced by expression vectors or gene therapy vectors (e.g., in the context of ex vivo gene therapy) can also be administered intravenously or parenterally as described above.

[0188] Formulations suitable for oral administration can consist of (a) liquid solutions, such as an effective amount of the packaged nucleic acid suspended in diluents, such as water, saline, buffered saline, ethanol, glycerol, dextrose, PEG 400 and combinations thereof; (b) capsules, sachets or tablets, each containing a predetermined amount of the active ingredient, as liquids, solids, granules or gelatin; (c) suspensions in an appropriate liquid; and (d) suitable emulsions. Tablet forms can include one or more of lactose, sucrose, mannitol, sorbitol, calcium phosphates, corn starch, potato starch, tragacanth, microcrystalline cellulose, acacia, gelatin, colloidal silicon dioxide, croscarmellose sodium, talc, magnesium stearate, stearic acid, and other excipients, colorants, fillers, binders, diluents, buffering agents, moistening agents, preservatives, flavoring agents, dyes, disintegrating agents, and pharmaceutically compatible carriers. Lozenge forms can comprise the active ingredient in a flavor, usually sucrose and acacia or tragacanth, as well as pastilles comprising the active ingredient in an inert base, such as gelatin and glycerin or sucrose and acacia emulsions, gels, and the like containing, in addition to the active ingredient, carriers known in the art.

[0189] The materials, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[0190] Suitable formulations for rectal administration include, for example, suppositories, which consist of the packaged nucleic acid with a suppository base. Suitable suppository bases include natural or synthetic triglycerides or paraffin hydrocarbons. In addition, it is also possible to use gelatin rectal capsules which consist of a combination of materials with a base, including, for example, liquid triglycerides, polyethylene glycols, and paraffin hydrocarbons.

[0191] The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial therapeutic response in the patient over time. The dose will be determined by the efficacy of the particular composition employed and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular composition (e.g., gene therapy vector, transduced cell type, protein or activity modulator) in a particular patient.

[0192] For example, in one aspect, the dose equivalent of a naked nucleic acid encoding a nucleic acid herein is from about 0.1 μg to 1 mg for a typical 70 kilogram patient, and doses of vectors which include a gene therapy or expression vector, such as a retroviral particle, are calculated to yield an approximately equivalent amount of a nucleic acid.

[0193] In the practice of this invention, compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The method of administration will often be local, oral, rectal or intravenous, but materials can also be applied in a suitable vehicle for the topical treatment of related conditions. The agents of this invention can supplement treatment of conditions associated with cholesterol exposure and adipogenesis, e.g., obesity, atherosclerosis, diabetes and coronary artery heart disease, or related conditions by any known conventional therapy, including pain medications, biologic response modifiers and the like.

[0194] For administration, compositions of the present invention can be administered at a rate determined by the LD-50 of composition and the side-effects of the composition at various concentrations, as applied to the mass and overall health of the patient. Administration can be accomplished via single or divided doses.

[0195] For ex-vivo therapy, transduced cells are prepared for reinfusion according to established methods. See, Abrahamsen et al. (1991) *J. Clin. Apheresis* 6:48-53; Carter et al. (1988) *J. Clin. Arpheresis* 4:113-117; Aebersold et al. (1988), *J. Immunol. Methods* 112: 1-7; Muul et al. (1987) *J. Immunol. Methods* 101:171-181 and Carter et al. (1987) *Transfusion* 27:362-365. After a period of about 2-4 weeks in culture, the cells should number between 1×10^8 and 1×10^{12}. In this regard, the growth characteristics of cells vary from patient to patient and from cell type to cell type. About 72 hours prior to reinfusion of the transduced cells, an aliquot is taken for analysis of phenotype, and percentage of cells expressing the therapeutic agent.

[0196] In one embodiment, in ex vivo methods, one or more cells, or a population of the subject's cells of interest are obtained or removed from the subject and contacted with an amount of a molecule of the invention, e.g., nucleic acids or subsequences thereof or isolated or recombinant polypeptides or subsequences thereof or antibodies, that is effective in prophylactically or therapeutically treating the condition in question. The contacted cells are then returned or delivered to the subject to the site from which they were obtained or to another site (e.g., including those defined above) of interest in the subject to be treated. Contacted cells can also be grafted onto a tissue or system site (including all described above) of interest in the subject using standard and well-known grafting techniques or, e.g., delivered to the

blood or lymph system using standard delivery or transfusion techniques. In another embodiment, a construct comprising a molecule, e.g., a nucleic acid comprising a polynucleotide sequence of the invention, that encodes a biologically active peptide that is effective in prophylactically or therapeutically treating the condition in question, is introduced into the one or more cells of interest or a population of cells of interest of the subject. A sufficient amount of the construct and a controlling promoter is used such that uptake of the construct (and promoter) into the cell(s) occurs and sufficient expression of the biologically active peptide produces an amount of the biologically active molecule effective to prophylactically or therapeutically treat the condition in question. Expression of the target nucleic acid can either be induced or occur naturally and a sufficient amount of the molecule is expressed and effective to treat the disease or condition at the site or tissue system.

[0197] In another embodiment, the invention provides in vivo methods in which one or more cells or a population of the subject's cells of interest are contacted directly or indirectly with an amount of a molecule(s) of the invention effective to ameliorate the condition in question. In direct contact/administration formats, the molecule(s) is typically administered or transferred directly to the cells to be treated or to the tissue site of interest by any of a variety of formats, which include injection, e.g., by a needle and/or syringe, vaccine, gene gun delivery, or pushing into the tissue. The molecule(s) can be delivered as described above, or placed within a cavity of the body (including, e.g., during surgery).

[0198] In in vivo indirect contact/administration formats, the molecule(s) is administered or transferred indirectly to the cells to be treated or to the tissue site of interest by contacting or administering the molecule(s) of the invention directly to one or more cells or population of cells from which treatment can be facilitated. For example, cells of interest, e.g., adipose tissue, liver, etc., within the body of the subject can be treated by contacting cells of the blood or lymphatic system with a sufficient amount of the molecule such that delivery of the molecule to the site of interest (e.g., adipose tissue, liver, etc.) occurs and effective prophylactic or therapeutic treatment results. Such contact, administration, or transfer is typically made by using one or more of the routes or modes of administration described above.

[0199] In one embodiment, the invention provides in vivo methods. Typically, one or more cells of interest or a population of subject's cells (e.g., including those cells and cell(s) systems and subjects described above) are transformed in the body of the subject by contacting the cell(s) or population of cells with (or administering or transferring to the cell(s) or population of cells using one or more of the routes or modes of administration described above) a polynucleotide construct comprising a nucleic acid sequence of the invention that encodes a biologically active molecule of interest (e.g., a polynucleotide of the invention) that is effective in prophylactically or therapeutically treating the condition of interest. Expression of the nucleic acid can be induced or occur naturally such that an amount of the encoded polypeptide expressed is sufficient and effective to treat the condition. The polynucleotide construct can include a promoter sequence (e.g., CMV promoter sequence) and optionally, one or more additional nucleotide sequences of the invention, adjuvant, or co-stimulatory molecule, or other polypeptide of interest.

[0200] A variety of viral vectors suitable for in vivo transduction and expression in an organism are known. Such vectors include retroviral vectors (see, Miller (1992) *Curr. Top. Microbiol. Immunol* 158:1-24; Salmons and Gunzburg (1993) *Human Gene Therapy* 4:129-141; Miller et al. (1994) *Methods in Enzymology* 217: 581-599), adeno-associated vectors (reviewed in Carter (1992) *Curr. Opinion Biotech.* 3: 533-539; Muzcyzka (1992) *Curr. Top. Microbiol. Immunol.* 158: 97-129) and other viral vectors (as generally described in, e.g., Jolly (1994) *Cancer Gene Therapy* 1:51-64; Latchman (1994) *Molec. Biotechnol.* 2:179-195; and Johanning et al. (1995) *Nucl. Acids Res.* 23:1495-1501).

[0201] If a patient undergoing infusion of a therapeutic composition develops fevers, chills, or muscle aches, he/she receives the appropriate dose of aspirin, ibuprofen or acetaminophen. Patients who experience reactions to the infusion such as fever, muscle aches, and chills are premedicated 30 minutes prior to the future infusions with either aspirin, acetaminophen, or diphenhydramine. Meperidine is used for more severe chills and muscle aches that do not quickly respond to antipyretics and antihistamines. Cell infusion is slowed or discontinued depending upon the severity of the reaction.

[0202] In general, gene therapy provides methods for combating acquired diseases and some forms of congenital defects such as enzyme deficiencies. Various textbooks describe gene therapy protocols which can be used with the present invention by introducing nucleic acids, e.g., one or more of SEQ ID NO:1 to SEQ ID NO: 443, into patient. One example is Robbins (1996) *Gene Therapy Protocols*, Humana Press, NJ, and Joyner (1993) *Gene Targeting: A Practical Approach*, IRL Press, Oxford, England.

[0203] In addition to the references cited above, several approaches for introducing nucleic acids into cells in vivo, ex vivo and in vitro are also described below along with the references cited within. These include liposome based gene delivery (Debs and Zhu (1993) WO 93/24640 and U.S. Pat. No. 5,641,662; Mannino and Gould-Fogerite (1988) *BioTechniques* 6(7): 682-691; Rose, U.S. Pat. No. 5,279,833; Brigham (1991) WO 91/06309; and Felgner et al. (1987) *Proc. Natl. Acad. Sci. USA* 84: 7413-7414); Brigham et al. (1989) *Am. J. Med. Sci.,* 298:278-281; Nabel et al. (1990) *Science,* 249:1285-1288; Hazinski et al. (1991) *Am. J. Resp. Cell Molec. Biol.,* 4:206-209; and Wang and Huang (1987) *Proc. Natl. Acad. Sci USA,* 84:7851-7855).; adenoviral vector mediated gene delivery, e.g., to treat cancer (see, e.g., Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91: 3054-3057; Tong et al. (1996) *Gynecol. Oncol.* 61: 175-179; Clayman et al. (1995) *Cancer Res.* 5: 1-6; O'Malley et al. (1995) *Cancer Res.* 55: 1080-1085; Hwang et al. (1995) *Am. J. Respir. Cell Mol. Biol.* 13: 7-16; Haddada et al. (1995) *Curr. Top. Microbiol. Immunol.* 199 (Pt. 3): 297-306; Addison et al. (1995) *Proc. Nat'l. Acad. Sci USA* 92: 8522-8526; Colak et al. (1995) *Brain Res* 691: 76-82; Crystal (1995) *Science* 270: 404-410;.Elshami et al. (1996) *Human Gene Ther.* 7: 141-148; Vincent et al. (1996) *J. Neurosurg.* 85: 648-654). Other delivery systems include replication-defective retroviral vectors harboring therapeutic polynucleotide sequence as part of the retroviral genome, particularly with regard to simple MuLV vectors (Miller et al. (1990) *Mol. Cell. Biol.* 10:4239 (1990); Kolberg (1992) *J. NIH Res.* 4:43, and Cornetta et al. (1991) *Hum. Gene Ther.* 2:215), nucleic acid transport coupled to ligand-specific, cation-based trans-

port systems (Wu and Wu (1988) *J. Biol. Chem.,* 263:14621-14624) and naked DNA expression vectors (Nabel et al. (1990), supra); Wolff et al. (1990) *Science,* 247:1465-1468). In general, these approaches can be adapted to the invention by incorporating nucleic acids, i.e., the polynucleotides of the invention, into the appropriate vectors.

[0204] In addition to expression of the nucleic acids of the invention as gene replacement nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, once expression of the nucleic acid is no-longer desired in the cell. Similarly, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can also be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) *Antisense Technology: A Practical Approach* IRL Press at Oxford University, Oxford, England, and in Agrawal (1996) *Antisense Therepeutics* Humana Press, NJ, and the references cited therein.

[0205] Kits and Reagents

[0206] The present invention is optionally provided to a user as a kit. For example, a kit of the invention contains one or more nucleic acid, polypeptide, antibody, or cell line described herein. Most often, the kit contains a diagnostic nucleic acid or polypeptide, e.g., antibody, probe set, e.g., as a cDNA microarray packaged in a suitable container, or other nucleic acid such as one or more expression vector. The kit typically further comprises, one or more additional reagents, e.g., substrates, labels, primers, for labeling expression products, tubes and/or other accessories, reagents for collecting samples, buffers, hybridization chambers, cover slips, etc. The kit optionally further comprises an instruction set or user manual detailing preferred methods of using the kit components for discovery or application of diagnostic gene sets.

[0207] When used according to the instructions, the kit can be used, e.g., for evaluating expression or polymorphisms in a subject sample, i.e., for evaluating a characteristic or condition associated with a physiologic or pathologic response to cholesterol exposure and/or adipogenesis, or for evaluating effects of a pharmaceutical agent or dietary intervention on cholesterol homeostasis and/or adipogenesis in a cell or organism.

[0208] Digital Systems

[0209] The present invention provides digital systems, e.g., computers, computer readable media and integrated systems comprising character strings corresponding to the sequence information herein for the polypeptides and nucleic acids herein, including, e.g., those sequences listed herein and the various silent substitutions and conservative substitutions thereof. Integrated systems can further include, e.g., gene synthesis equipment for making genes corresponding to the character strings.

[0210] Various methods known in the art can be used to detect homology or similarity between different character strings, or can be used to perform other desirable functions such as to control output files, provide the basis for making presentations of information including the sequences and the like. Examples include BLAST, discussed supra. Computer systems of the invention can include such programs, e.g., in conjunction with one or more data file or data base comprising a sequence as noted herein.

[0211] Thus, different types of homology and similarity of various stringency and length can be detected and recognized in the integrated systems herein. For example, many homology determination methods have been designed for comparative analysis of sequences of biopolymers, for spell-checking in word processing, and for data retrieval from various databases. With an understanding of double-helix pair-wise complement interactions among 4 principal nucleobases in natural polynucleotides, models that simulate annealing of complementary homologous polynucleotide strings can also be used as a foundation of sequence alignment or other operations typically performed on the character strings corresponding to the sequences herein (e.g., word-processing manipulations, construction of figures comprising sequence or subsequence character strings, output tables, etc.).

[0212] Thus, standard desktop applications such as word processing software (e.g., Microsoft Word™ or Corel WordPerfect™) and database software (e.g., spreadsheet software such as Microsoft Excel™, Corel Quattro Pro™, or database programs such as Microsoft Access™ or Paradx™) can be adapted to the present invention by inputting a character string corresponding to one or more polynucleotides and polypeptides of the invention (either nucleic acids or proteins, or both). For example, a system of the invention can include the foregoing software having the appropriate character string information, e.g., used in conjunction with a user interface (e.g., a GUI in a standard operating system such as a Windows, Macintosh or LINUX system) to manipulate strings of characters corresponding to the sequences herein. As noted, specialized alignment programs such as BLAST can also be incorporated into the systems of the invention for alignment of nucleic acids or proteins (or corresponding character strings).

[0213] Systems in the present invention typically include a digital computer with data sets entered into the software system comprising any of the sequences herein. The computer can be, e.g., a PC (Intel×86 or Pentium chip-compatible DOS™, OS2™ WINDOWS™ WINDOWS NT™, WINDOWS95™, WINDOWS98™ LINUX based machine, a MACINTOSH™, Power PC, or a UNIX based (e.g., SUN™ work station) machine) or other commercially common computer which is known to one of skill. Software for aligning or otherwise manipulating sequences is available, or can easily be constructed by one of skill using a standard programming language such as Visualbasic, Fortran, Basic, Java, or the like.

[0214] Any controller or computer optionally includes a monitor which is often a cathode ray tube ("CRT") display, a flat panel display (e.g., active matrix liquid crystal display, liquid crystal display), or others. Computer circuitry is often placed in a box which includes numerous integrated circuit chips, such as a microprocessor, memory, interface circuits, and others. The box also optionally includes a hard disk drive, a floppy disk drive, a high capacity removable drive such as a writeable CD-ROM, and other common peripheral elements. Inputting devices such as a keyboard or mouse optionally provide for input from a user and for user selection of sequences to be compared or otherwise manipulated in the relevant computer system.

[0215] The computer typically includes appropriate software for receiving user instructions, either in the form of user input into a set parameter fields, e.g., in a GUI, or in the form of preprogrammed instructions, e.g., preprogrammed for a variety of different specific operations. The software then converts these instructions to appropriate language for instructing the operation of the fluid direction and transport controller to carry out the desired operation.

[0216] The software can also include output elements for controlling nucleic acid synthesis (e.g., based upon a sequence or an alignment of a sequences herein) or other operations.

[0217] In an additional aspect, the present invention provides system kits embodying the methods, composition, systems and apparatus herein. System kits of the invention optionally comprise one or more of the following: (1) an apparatus, system, system component or apparatus component as described herein; (2) instructions for practicing the methods described herein, and/or for operating the apparatus or apparatus components herein and/or for using the compositions herein. In a further aspect, the present invention provides for the use of any apparatus, apparatus component, composition or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit to practice any assay or method herein.

[0218] Molecular Techniques

[0219] In the context of the invention, nucleic acids and/or proteins are manipulated according to well known molecular biology methods. Detailed protocols for numerous such procedures are described in, e.g., in Ausubel et al. *Current Protocols in Molecular Biology* (supplemented through 2001) John Wiley & Sons, New York ("Ausubel"); Sambrook et al. *Molecular Cloning—A Laboratory Manual* (*2nd Ed.*), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989 ("Sambrook"), and Berger and Kimmel *Guide to Molecular Cloning Techniques, Methods in Enzymology volume* 152 Academic Press, Inc., San Diego, Calif. ("Berger").

[0220] In addition to the above references, protocols for in vitro amplification techniques, such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), Qβ-replicase amplification, and other RNA polymerase mediated techniques (e.g., NASBA), useful e.g., for amplifying cDNA probes of the invention, are found in Mullis et al. (1987) U.S. Pat. No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis et al. eds) Academic Press Inc. San Diego, Calif. (1990) ("Innis"); Arnheim and Levinson (1990) *C&EN* 36; *The Journal Of NIH Research* (1991) 3:81; Kwoh et al. (1989) *Proc Natl Acad Sci USA* 86, 1173; Guatelli et al. (1990) *Proc Natl Acad Sci USA* 87:1874; Lomell et al. (1989) *J Clin Chem* 35:1826; Landegren et al. (1988) *Science* 241:1077; Van Brunt (1990) *Biotechnology* 8:291; Wu and Wallace (1989) *Gene* 4: 560; Barringer et al. (1990) *Gene* 89:117, and Sooknanan and Malek (1995) *Biotechnology* 13:563. Additional methods, useful for cloning nucleic acids in the context of the present invention, inlcude Wallace et al. U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) *Nature* 369:684 and the references therein.

[0221] Certain polynucleotides of the invention, e.g., oligonucleotides can be synthesized utilizing various solid-phase strategies involving mononucleotide- and/or trinucleotide-based phosphoramidite coupling chemistry. For example, nucleic acid sequences can be synthesized by the sequential addition of activated monomers and/or trimers to an elongating polynucleotide chain. See e.g., Caruthers, M. H. et al. (1992) *Meth Enzymol* 211:3. In lieu of synthesizing the desired sequences, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (available on the World Wide Web at genco.com), ExpressGen, Inc. (available on the World Wide Web at expressgen-.com), Operon Technologies, Inc. (available on the World Wide Web at operon.com), and many others.

[0222] Similarly, commercial sources for nucleic acid and protein microarrays are available, and include, e.g., Affymetrix, Santa Clara,Calif. (available on the World Wide Web at affymetrix.com); and Agilent, Palo Alto, Calif. (available on the World Wide Web at agilent.com) Zyomyx, Hayward, Calif. (available on the World Wide Web at zyomyx.com); and Ciphergen Biosciences, Fremont, Calif. (available on the World Wide Web at ciphergen.com).

[0223] A variety of techniques can be used to detect differential gene expression and generate the sequence information corresponding to the gene that is differentially expressed. Typically, massively parallel signature sequencing is used; other examples include SAGE data, microarrays and cDNA fragment profiling methods. See, e.g., Brenner et al., (2000), *Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays, Nature Biotech.,* 18:630-634; Tyagi, (2000), *Taking a census of mRNA populations with microbeads, Nature Biotech.,* 18:597-598; Brenner et al., (2000) *In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs, PNAS USA* 97:1665-1670; Okubo et al., (1992), *Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression, Nature Genetics,* 2:173-179; Bachem et al., (1996) *Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development, Plant J.,* 9:745-753; Nelson M, et al., (1993) *Sequencing two DNA templates in five channels by digital compression, PNAS (US),* 90(5):1647-51; and Shimkets et al., (1999) *Gene expression analysis by transcript profiling coupled to database query, Nature Biotechnology,* 17:798-803.

[0224] Massively parallel signature sequencing (MPSS) is designed for large-scale counting of individual mRNA molecules in a sample. MPSS provides data for all genes in a tissue or cell sample, not just those that have been previously identified and characterized. No prior knowledge of a gene's sequence is required for MPSS; thus, gene expression datasets can be generated from any organism. In addition, MPSS has a high sensitivity level. Anywhere from about 100,000 to about ten million molecules are typically counted in any given sample, so that even genes that are expressed at low levels can be quantified with high accuracy. Typically, an MPSS dataset typically involves greater than, e.g., about 100,000 signature sequences, to about 750,000 signature sequences. Two-flow cells with microbeads initiated with either of two different initiating adaptors can be used for each experiment, e.g., a 2-stepper and 4-stepper as described above. Therefore, datasets containing from about 200,000 to

about 1,400,000 signature sequences can be generated for any given sample. The data from multiple MPSS experiments can optionally be combined.

[0225] MPSS is a "digital" gene expression tool that counts all mRNA molecules simultaneously. Counting mRNAs with MPSS is based on the ability to uniquely identify every mRNA in a sample. This is done by generating a sequence of 17 or more bases for each mRNA at a specific site upstream from its poly(A) tail (e.g., the last DpnII site in double stranded cDNA). The sequence of 17 or more bases is then used as an mRNA identification "signature." To measure the level of expression of any given gene in a sample analyzed by MPSS, the total number of signatures for that gene's mRNA are counted.

[0226] MPSS signatures for mRNAs in a sample are generated by sequencing double stranded cDNAs fragments cloned on to microbeads using the Lynx Megaclone technology. A clone refers to a single microbead from which 17 or more bases have been sequenced to create a signature sequence tag from an individual cDNA molecule that has been cloned into the Megaclone library. Fragments from 100,000-10,000,000 individual cDNA molecules from a sample are cloned on to 100,000-10,000,000 separate microbeads using, e.g., the procedure described in Brenner et al., supra, *PNAS*, thereby making a Megaclone library of cloned cDNA fragments.

[0227] MPSS and microbead technology is further described in the following patents and references cited within: U.S. Pat. No. 6,306,597 to Macevicz entitled "DNA sequencing by parallel oligonucleotide extensions" issued Oct. 23, 2001; U.S. Pat. No. 6,280,935 to Macevicz entitled "Method of detecting the presence or absence of a plurality of target sequences using oligonucleotide tags" issued Aug. 28, 2001; U.S. Pat. No. 6,265,163 to Albrecht et al., entitled "Solid phase selection of differentially expressed genes" issued Jul. 24, 2001; U.S. Pat. No. 6,235,475 to Brenner et al., entitled "Oligonucleotide tags for sorting and identification" issued May 22, 2001; U.S. Pat. No. 6,228,589 to Brenner entitled "Measurement of gene expression profiles in toxicity determination" issued May 8, 2001; U.S. Pat. No. 6,175,002 to DuBridge et al., entitled "Adaptor-based sequence analysis" issued Jan. 16, 2001; U.S. Pat. No. 6,172,218 to Brenner entitled "Oligonucleotide tags for sorting and identification" issued Jan. 9, 2001; U.S. Pat. No. 6,172,214 to Brenner entitled "Oligonucleotide tags for sorting and identification" issued Jan. 9, 2001; U.S. Pat. No. 6,150,516 to Brenner et al., entitled "Kits for sorting and identifying polynucleotides" issued Nov. 21, 2000; U.S. Pat. No. 6,140,489 to Brenner entitled "Compositions for sorting polynucleotides" issued Oct. 31, 2000; U.S. Pat. No. 6,138,077 to Brenner entitled "Method, apparatus and computer program product for determining a set of non-hybridizing oligonucleotides" issued on Oct. 24, 2000; U.S. Pat. No. 6,013,445 to Albrecht et al., entitled "Massively parallel signature sequencing by ligation of encoded adaptors" issued Jan. 11, 2000; U.S. Pat. No. 5,962,228 to Brenner entitled "DNA extension and analysis with rolling primers" issued Oct. 5, 1999; U.S. Pat. No. 5,888,737 to DuBridge et al., entitled "Adaptor-based sequence analysis" issued Mar. 30, 1999; U.S. Pat. No. 5,780,231 to Brenner entitled "DNA extension and analysis with rolling primers" issued Jul. 14, 1998; U.S. Pat. No. 5,750,341 to Macevicz entitled "DNA sequencing by parallel oligonucleotide extensions" issued May 12, 1998; U.S. Pat. No. 5,747,255 to Brenner entitled "Polynucleotide detection by isothermal amplification using cleavable oligonucleotides" issued May 5, 1998; U.S. Pat. No. 5,969,119 to Macevicz entitled "DNA sequencing by parallel oligonucleotide extensions" issued Oct. 19, 1999; U.S. Pat. No. 5,863,722 to Brenner entitled "Method of sorting polynucleotides" issued Jan. 26, 1999; U.S. Pat. No. 5,846,719 to Brenner et al. entitled "Oligonucleotide tags for sorting and identification" issued Dec. 8, 1998; U.S. Pat. No. 5,763,175 to Brenner entitled "Simultaneous sequencing of tagged polynucleotides" issued Jun. 9, 1998; U.S. Pat. No. 5,695,934 to Brenner entitled "Massively Parallel sequencing of sorted polynucleotides" issued Dec. 9, 1997; U.S. Pat. No. 5,635,400 to Brenner entitled "Minimally cross-hybridizing sets of oligonucleotide tags" issued Jun. 3, 1997; and, U.S. Pat. No. 5,604,097 to Brenner entitled "Methods for sorting polynucleotides using oligonucleotide tags" issued Feb. 19, 1997.

[0228] In MPSS, DNA is sequenced through an automated series of adaptor ligations and enzymatic steps. Two, e.g., independent sampling, procedures typically used involve either a 4-stepper or 2-stepper, which differ by using two alternative reading-frame adaptors. For example, in a 4 stepper procedure, the process is initiated by ligating an adaptor molecule to the GATC (DpnII) single-stranded overhangs, and then digesting the samples with BbvI, which is a type IIs restriction enzyme that cuts the DNA at a position 9-13 nucleotides away from the recognition sequence. This produces molecules with a 4 base single stranded overhang immediately adjacent to the DpnII recognition sequence. Another set of adaptors, called encoded adaptors, are hybridized and ligated to the 4 base overhangs on each molecule. The encoded adaptors contain a 4 base single stranded overhang with all possible nucleotide combinations at one end, and a single stranded coded sequence at the other end. One member of the encoded adaptor set will find a partner on the DNA molecules attached to the beads in the flow cell. The exact sequence of each encoded adaptor that hybridizes to the DNA on a microbead is decoded through 16 different sequential hybridization reactions with a set of fluorescent decoder probes. This process yields the first 4 nucleotides at the end of each molecule. To collect additional sequence, the encoded adaptor from the first round is removed by digestion with BbvI, and the process is repeated several times. In the end, a 17 or more base signature sequence is generated for each bead in the flowcell. In a 2-stepper, the sequence obtained is in a different reading frame, which is staggered by two bases compared to the 4-stepper.

[0229] Specifically, in a 2-stepper protocol, the recognition site for the type IIS restriction enzyme, e.g., BbvI, used to expose the first four nucleotides to identify the signature sequence, is located 11 nucleotides from the GATC site at the end of the adaptor. In the 4-stepper protocol, the recognition site for the type IIS restriction enzyme, e.g., BbvI, used to expose the first four nucleotides to identify the signature sequence, is located 9 nucleotides from the GATC site at the end of the adaptor. The difference between the 2-stepper protocol and the 4-stepper protocol allows the choice of what overhang will be produced after the first restriction enzyme, e.g., BbvI, digestion. The datasets generated with the two different adaptors are different, because a different set of four base-pair overhangs will be generated for each signature sequence depending on whether a 2-step-

per or 4-stepper protocol is used. Each exposed four base pair can potentially contain a palindromic structure, e.g., 16 of 256 different possible four base-pair overhangs. There can also be additional biases due to the relative efficiency of individual overhangs in the ligation processes involved during the sequencing cycles. The dataset generated and the biases make the 2-stepper and 4-stepper protocols independent sampling methods.

[0230] Ligation-based sequencing is further described in the following patents and references cited within: U.S. Pat. No. 5,714,330 to Brenner et al., entitled "DNA sequencing by stepwise ligation and cleavage" issued Feb. 3, 1998; U.S. Pat. No. 5,599,675 to Brenner entitled "DNA sequencing by stepwise ligation and cleavage" issued Feb. 4, 1997; U.S. Pat. No. 5,831,065 to Brenner entitled "Kits for DNA sequencing by stepwise ligation and cleavage" issued Nov. 3, 1998; U.S. Pat. No. 5,856,093 to Brenner entitled "Method of determining zygosity by ligation and cleavage" issued Jan. 5, 1999; and, U.S. Pat. No. 5,552,278 to Brenner entitled "DNA sequencing by stepwise ligation and cleavage" issued Sep. 3, 1996.

[0231] Another technology that can be used is SAGE technology. SAGE is another transcript counting technique that generates a tag sequence for each mRNA. It also generates a digital gene expression profile. SAGE is based on the principles that a short sequence tag derived from a defined position from a mRNA can uniquely identify the transcript and concatenation of the tags allows for high-throughput sequencing. The length of the SAGE tag is about 10 to about 14 nucleotides. The tag sequence is determined using conventional sequencing technologies. See the following publications and references cited within: Velculescu et al., (1995), *Serial analysis of gene expression, Science,* 270:484-487; and Zhang et al., (1997), *Gene expression profiles in normal and cancer cells, Science,* 276:1268-1272. To determine expression level of a gene from SAGE technique, the frequency of a sequence tag derived from the corresponding mRNA transcript is measured. As with microarray data described below, adjustments to consider bias and normalization are optionally included in the present invention. See, e.g., Marguiles et al., (2001) *Identification and prevention of a GC content bias in SAGE libraries, Nucleic Acid Res.,* 29(12):E60-0.

[0232] Microarrays are also technologies that can be used in the present invention. Typically, a microarray is a solid support that contains a variety of genes. The mRNAs from the sample are then allowed to hybridize to the microarray. Microarrays have the advantage of high throughput analysis of multiple samples. Typically with microarray techniques, some or all of a variety of variables should be considered. These variables include, e.g., that the desired genes are represented on a given array. Second, a microarray exists for the organism of interest. Third, the detection sensitivity is optimized to achieve detection of low expressed genes. Fourth, a sample is compared with a control sample to compensate for several sources of bias and noise in the intensity results. Typically, the experiment is replicated several times to provide a more reliable dataset. Fifth, compensation is made for multiple values for single gene, because multiple values can arise from, e.g., distinct probe sets within different sections within the gene. See Kerr and Churchhill, G. A., (2001), *Statistical design and the analysis of gene expression microarray data, Biostatistics,* 2:183-201; Wodicka et al., (1997), *Genome wide expression monitoring in Saccharomyces cerevisiae, Nature Biotech.,* 15:1359-1367; Lockhart et al., (1996), *Expression monitoring by hybridization to high-density oligonucleotide arrays, Nature Biotech.,* 14:1675-1680; Aach et al., *Systematic management and analysis of yeast gene expression data, Genome Res.,* 10:431-445 and Wittes and Friedman, (1999) *Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data, J. Natl. Cancer Inst.,* 91:400-401.

[0233] More information can be found in the following publications and references cited within: Duggan et al., (1999), *Expression profiling using cDNA microarrays, Nature Genetics,* 21:10-14; Lipshutz et al., *High density synthetic oligonucleotide arrays, Nature Genetics Suppl.* 21:20-24; Evertsz et al., (2000), *Technology and applications of gene expression microarrays, in Microarray Biochip technology*, Schena, M., Ed. BioTechniques Books, Natick, Mass., pp.149-166; Lockhart and Winzeler, (2000), *Genomics, gene expression and DNA arrays, Nature,* 405:827-836; Zhou et al., (2000), *Information processing issues and solutions associated with microarray technology, in Microarray Biochip technology*, Schena, M., Ed., BioTechniques Books, Natick, Mass., pp. 167-200; and Hughes et al., (2001), *Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer, Nature Biotech.,* 19:342-347.

[0234] A comparison between two samples can be made in order to determine, e.g., differential expression. A variety of statistical comparison tests can be used, for example, a two-tailed normal approximation test, a chi-squared test, a Fisher exact test, a generalized linear model, Audic and Clayerie's Bayesian method and the like. Comparison tests are well-known to one of skill in the art; information on statistical tests can be found in variety of places, such as, textbooks, papers and the World Wide Web. For example, see Fisher and van Belle, (1993) *Biostatistics: a Methodology for the Health Science*, John Wiley & Sons, New York; Man et al., (2000) *POWER SAGE: comparing statistical tests for SAGE experiments, Bioinformatics,* 16(11): 953-959; and, Audic and Clayerie, (1997) *The significance of digital gene expression profiles, Genome Research,* 7:986-995. Further details on the use of the two tailed normal approximation test are found in U.S. Patent Application, concurrently filed on Dec. 10, 2002, LOJAQ docket No. 37-000710US, the contents of which are incorporated by reference.

EXAMPLES

[0235] The following examples are offered to illustrate, but not to limit the claimed invention.

[0236] Cholesterol Treatment: Human fibroblast cells were maintained in culture in DMEM with 10% lipoprotein-deficient serum, and then incubated either with 50 μM compactin and 10 μM mevalonate (Induced or "Ncho" condition) or with 1 μg/ml 25-hydroxycholesterol and 10 μg/ml cholesterol (suppressed, or "Ycho" condition). MPSS was performed on cDNA isolated from cells incubated under these two treatment conditions. Sequencing of 629,269 and 807,483 cDNA clones derived from the Ncho and Ycho treated samples, respectively, yielded a total of 24,854 unique signatures.

[0237] Adipogenesis: Human immortalized preadipocytes (PAZ6) differentiated into adipocytes in vitro with the induction of 850 nM insulin, 1 nM triiodothyronine, 100 nM dexamethasone, and 1 μM piaglitazone for 15-21 days (see, Zilberfarb et al. (1997) *Human immortalized adipocytes express functional beta-3-adrenoreceptor coupled to lipolysis Journal of Cell Science* 110:801-807). MPSS was performed on cDNA isolated from pre-adipocytes (bas1) and adipocytes (dffr). A total of 17840 unique 17-base signatures were obtained by sequencing 1,089,207 and 1,825,821 cDNA clones from the preadipocyte and adipocyte samples, respectively.

[0238] MPSS data analysis: Statistical analysis of each of the above two datasets was performed using normal approximation methods, e.g., as described in "Methods for Analysis of Massively Parallel Signature Sequencing" by Jing Zhong Lin et al., filed Dec. 10, 2002 (Attorney Docket No. 37-000710US) incorporated herein by reference, to identify signatures exhibited a significant change in abundance with the treatment. Signatures exhibiting a statistically significant change in abundance in response to either cholesterol or adipocyte differentiation were then corresponded to unique genes using the blast search algorithm against the NCBI NR and EST databases. The two datasets were then compared to identify common signatures. 578 and 256 common signatures exhibited differential expression at the significant level of $p<0.01$ and $p<0.001$, respectively.

	Cholesterol	Adipogenesis	Common
Total Signatures	24,854	17,840	10,675
Expressed Differentially $p < 0.01$	3619	3278	578
Expressed Differentially $p < 0.001$	1526	1625	256

[0239] Using the following criteria: (1) sequences exhibited a significant change in abundance ($p<0.01$) in both fibroblast cholesterol treatment and during in vitro adipogenesis; (2) the abundance change is great than 2 fold with both treatments; (3) signature sequences can be assigned to a previously annotated gene or cDNA, a total of 277 genes/cDNAs were obtained. We further grouped these sequences into 4 categories: (a) induced by cholesterol and adipogenesis; (b) suppressed by cholesterol and adipogenesis; (c) induced by cholesterol but suppressed by adipogenesis; (d) suppressed by cholesterol but induced by adipogenesis. The detailed signature abundance and gene annotation are provided in Appendix I.

	Cholesterol Induced	Cholesterol Suppressed
Adipogenesis Induced	78	65
Adipogenesis Suppressed	82	52

[0240] Similarly, a list of MPSS signatures induced by both cholesterol and adipogenesis, i.e., that are absent from the Ncho and basl samples and present in the cholesterol loaded sample (Ycho) and fat cell sample (dffr). A total of 166 unique signatures exhibited such an expression pattern. Of note, several genes encoding products that are likely to play a role in the synthesis or degradation of HSPG (exostoses 1, nidogen 2, mannose-6-phosphate receptor (cation dependent)) were detected in this category, e.g., heparin sulfate proteoglycan (HSPG). Detailed information is provided in Appendix II.

[0241] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims.

[0242] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

SEQ ID NO	Code	Sequence
1	166-1	GATCAAAACTATTAAAA
2	166-2	GATCAAACTACTCCGAG
3	166-3	GATCAAATGGAACATTT
4	166-4	GATCAAATGTGTGGGCA
5	166-5	GATCAACAGGCTGATTA
6	166-6	GATCAACATCATTAATT
7	166-7	GATCAACGGCACGACAG
8	166-8	GATCAAGGAGACCCGGA
9	166-9	GATCAAGTATCTGCTCA
10	166-10	GATCAAGTATGCTAATT
11	166-11	GATCAAGTTCAATATTC
12	166-12	GATCAAGTTTAGTTATT
13	166-13	GATCAATGAATTGCACA
14	166-14	GATCAATGTTTAGTAAA
15	166-15	GATCACAAACACTGGCA
16	166-16	GATCACACCACTCCAGT
17	166-17	GATCACATGTCAGGACA
18	166-18	GATCACCAAAATTAATT
19	166-19	GATCACCACTCTAACCC
20	166-20	GATCACCAGTGGGCACG
21	166-21	GATCACCGTGACAGCCA

-continued

SEQ ID NO	Code	Sequence
22	166-22	GATCACCGTTGCTTGGT
23	166-23	GATCACGCCACTGAACT
24	166-24	GATCACTCTGTTCATAA
25	166-25	GATCACTTGTGTTGAGG
26	166-26	GATCACTTTGTGGCAGG
27	166-27	GATCAGCCCAGGCAACA
28	166-28	GATCAGCTAATTCAGAG
29	166-29	GATCAGGAACAGACATG
30	166-30	GATCAGGAGGCATGACT
31	166-31	GATCAGTAAGGATTGCA
32	166-32	GATCAGTACAGACATTT
33	166-33	GATCAGTATTTTTTATC
34	166-34	GATCATAATAATTCCAA
35	166-35	GATCATGAAATTCTAAT
36	166-36	GATCATGGGCTCTGGAA
37	166-37	GATCATGTCCTAATTCA
38	166-38	GATCATGTTACCATGCA
39	166-39	GATCATGTTTCATTTCT
40	166-40	GATCATTATTATTTTTT
41	166-41	GATCATTCAAACATTTG
42	166-42	GATCATTTACAAATTTG
43	166-43	GATCATTTGTGATTCCT
44	166-44	GATCCAAACCAAGAGGA
45	166-45	GATCCAAACTAATATTG
46	166-46	GATCCAAAGAGCTCTTA
47	166-47	GATCCAAAGCAGTTGTT
48	166-48	GATCCAAATAAACACAC
49	166-49	GATCCAACAATAAAATA
50	166-50	GATCCAACACCGACTAC
51	166-51	GATCCAATACTATTTAG
52	166-52	GATCCACCCACCTCGGG
53	166-53	GATCCACCCAGTAAATC
54	166-54	GATCCACCCATCTCGCC
55	166-55	GATCCACTTCCTAAAAG
56	166-56	GATCCAGAAATAAGTTT
57	166-57	GATCCAGAGAGCTGACA

-continued

SEQ ID NO	Code	Sequence
58	166-58	GATCCAGCAATACCCCA
59	166-59	GATCCAGCAATCCCACC
60	166-60	GATCCAGCAGTCCCACT
61	166-61	GATCCAGCAGTCTCACT
62	166-62	GATCCAGGGGAAGCAAA
63	166-63	GATCCAGGTCTCCAATT
64	166-64	GATCCAGTGCTTAGTTC
65	166-65	GATCCAGTGTCCACGGA
66	166-66	GATCCAGTTCGTGCAGC
67	166-67	GATCCAGTTTTAGTTTT
68	166-68	GATCCATCAGTGTGCTG
69	166-69	GATCCATCTGCCTTTCT
70	166-70	GATCCATCTTCAAATCA
71	166-71	GATCCATTTGATTCTTA
72	166-72	GATCCCAACTGCTCCTT
73	166-73	GATCCCAAGAGACTTAC
74	166-74	GATCCCACCAATCCTCG
75	166-75	GATCCCACCATTGCACT
76	166-76	GATCCCACCTTATGGTG
77	166-77	GATCCCAGAAAAACAAC
78	166-78	GATCCCAGAGGCTGATT
79	166-79	GATCCCAGTGTTGCCTT
80	166-80	GATCCCATCACCCAGGT
81	166-81	GATCCCCCAACTCTGKA
82	166-82	GATCCCCTTTGGTCTTA
83	166-83	GATCCCTGGTATTGATT
84	166-84	GATCCCTTCCCTCAATT
85	166-85	GATCCGAGTCGTCGGGA
86	166-86	GATCCGATTTCTTTACC
87	166-87	GATCCCCACCCCCCCCC
88	166-88	GATCCCCAGGCAGACAC
89	166-89	GATCCGCAGGCAGACAG
90	166-90	GATCCGTATGTGGTTAA
91	166-91	GATCCGTCCCTAACTAC
92	166-92	GATCCTAACAAAACCTA
93	166-93	GATCCTACAAGTTGATC
94	166-94	GATCCTATTTTAATTTT

-continued

SEQ ID NO	Code	Sequence
95	166-95	GATCCTCAAAATAACCT
96	166-96	GATCCTCCGAGTAATTG
97	166-97	GATCCTCCTACATCTGC
98	166-98	GATCCTCTAGAGCCAGC
99	166-99	GATCCTCTTGCATTCTG
100	166-100	GATCCTGGAGCATCTCC
101	166-101	GATCCTGGCAATTTATT
102	166-102	GATCCTGCCCTGAGAAA
103	166-103	GATCCTGGGAATTATTT
104	166-104	GATCCTGTAGTTTATGT
105	166-105	GATCCTTTTTTTCATCT
106	166-106	GATCGAAGGAAACAACC
107	166-107	GATCGAGTCTGCTTTCC
108	166-108	GATCCCAAATCACAGCT
109	166-109	GATCGCCACTCAGAAAG
110	166-110	GATCGCTTCAGCCCGGG
111	166-111	GATCGCTTTGCTGTGCT
112	166-112	GATCGGAATCTATTAAG
113	166-113	GATCGGCATTCAGATTC
114	166-114	GATCGGGAGGACCTGTC
115	166-115	GATCGTGCCATTCCACT
116	166-116	GATCGTTCTTCAAGTAT
117	166-117	GATCRTGCCTTAAAAAG
118	166-118	GATCTAACTCTCTTCTT
119	166-119	GATCTAAGAGTTCCCTG
120	166-120	GATCTAAGGGTTTTAGT
121	166-121	GATCTAATGTTATTTTC
122	166-122	GATCTACCAAAACAGTT
123	166-123	GATCTACCAACCAACAG
124	166-124	GATCTACTACTTACTTA
125	166-125	GATCTACTCTGTGCCAG
126	166-126	GATCTAGAAAACTTTGG
127	166-127	GATCTAGACACAAAGGA
128	166-128	GATCTCAAATTCTCTAT
129	166-129	GATCTCACAGGCTCAGA
130	166-130	GATCTCACCTCTACGGG
131	166-131	GATCTCCTCCAGGAACA

-continued

SEQ ID NO	Code	Sequence
132	166-132	GATCTCTAGAGCTGTCT
133	166-133	GATCTCTGCGTTCTCAC
134	166-134	GATCTCTGCTTCCTTCC
135	166-135	GATCTCTTCAGAGGTAT
136	166-136	GATCTGAACTTTTTATC
137	166-137	GATCTGAATGGGGCTTT
138	166-138	GATCTGAATTTGTACCA
139	166-139	GATCTGACAGGGGTCTG
140	166-140	GATCTGATGGCTTTATA
141	166-141	GATCTGCACAGCCAGAC
142	166-142	GATCTGCCCACCGCAGC
143	166-143	GATCTGCTTGCCTTGGT
144	166-144	GATCTGGACAAATGGCA
145	166-145	GATCTGGATGGAGTGGT
146	166-146	GATCTGGGCTACCTCCT
147	166-147	GATCTGTGTTTGCCCTh
148	166-148	GATCTGTTACCGCTAAA
149	166-149	GATCTGTTATGAAACGA
150	166-150	GATCTGTTTATTCTTTT
151	166-151	GATCTGTTTCTTTTTTT
152	166-152	GATCTTAATGATGTTTT
153	166-153	GATCTTAGTTATCTGTA
154	166-154	GATCTTATTTCAAAGGA
155	166-155	GATCTTCAAAAATACTA
156	166-156	GATCTTCCTTTTCTCAG
157	166-157	GATCTTCTGGGTTGGCA
158	166-158	GATCTTCTTTATTTTTT
159	166-159	GATCTTGACTTCTTGAT
160	166-160	GATCTTGCAGTTTTCCT
161	166-161	GATCTTGGACGCCAGCC
162	166-162	GATCTTTCAATAGTCTG
163	166-163	GATCTTTCTTTAATACT
164	166-164	GATCTTTGTCGGTTAGA
165	166-165	GATCTTTTGCTTACCCA
166	166-166	GATCTTTTTAACATTGA
167	277-1	GATCACAGGCCTTGCTT
168	277-2	GATCACCATCCCAGTCA

-continued

SEQ ID NO	Code	Sequence
169	277-3	GATCACTGTCCTATCAC
170	277-4	GATCAGAATCATGGTCT
171	277-5	GATCAGATTCCGATTTG
172	277-6	GATCATGAATAGGAGCC
173	277-7	GATCATGATTTTGTAGT
174	277-8	GATCATGTTTTGCTACA
175	277-9	GATCATTCCTTCTCTAG
176	277-10	GATCCACACACGTTGGT
177	277-11	GATCCCAAATTTGTCCA
178	277-12	GATCCCAACGGCCTTAG
179	277-13	GATCCCAGGATTCAGTA
180	277-14	GATCCCATCTCTACGCC
181	277-15	GATCCCCAACAATGTCA
182	277-16	GATCCCCGGCCTCAGTC
183	277-17	GATCCCTGAAGTTGCCC
184	277-18	GATCCTGGAGGATTTCC
185	277-19	GATCCTTCAGCACAGGA
186	277-20	GATCCTCTTGTATGGTG
187	277-21	GATCGTGTATTGAGATT
188	277-22	GATCGTTGACAAGTATG
189	277-23	GATCTATCATTACTCCA
190	277-24	GATCTCTGTGCTGTAAA
191	277-25	GATCTGATTTATTTATT
192	277-26	GATCTGCTTGGAGTTTT
193	277-27	GATCTGGAACCTCAGCC
194	277-28	GATCTGGCATGTTAGCC
195	277-29	GATCTGGTGTGAGTGCA
196	277-30	GATCTGTACACAGTAAA
197	277-31	GATCTGTACAGACAGGA
198	277-32	GATCTGTACCTGAGAGG
199	277-33	GATCTGTCTATGGGACC
200	277-34	GATCTTTCCAACCACAT
201	277-35	GATCAACGCCTCACTGA
202	277-36	GATCCAAAGTCATGTGT
203	277-37	GATCCTGTTTCCATTTG
204	277-38	GATCTGTAAAATGTGAT
205	277-39	GATCATTGGTTCCAGTC

-continued

SEQ ID NO	Code	Sequence
206	277-40	GATCAACTTGAGTCCAA
207	277-41	GATCACCGCTTTCCAAT
208	277-42	GATCAGAGCTCAGTTCC
209	277-43	GATCAGCTGAACAGCAG
210	277-44	GATCATGTGCTACTGGT
211	277-45	GATCCCAGCTGATGTAG
212	277-46	GATCCTAGACAGGGCTC
213	277-47	GATCGAGCTCGCCTATG
214	277-48	GATCGAGGCTTGTGATG
215	277-49	GATCTATACTAGATAAT
216	277-50	GATCTCGAACCCTGTCT
217	277-51	GATCTTAGCTTTCATAA
218	277-52	GATCTTTAATGCTTTGG
219	277-53	GATCAAAAGGGACAAGC
220	277-54	GATCAAACCAAGCCCCA
221	277-55	GATCAACCTGGAGCTCT
222	277-56	CATCAAGAACAATGCCT
223	277-57	GATCACAGGCAAACCCA
224	277-58	GATCACAGGGGTGATGG
225	277-59	GATCACATCTGTGTGAA
226	277-60	GATCACATGAATAGGGG
227	277-61	GATCAGAAAAGCAGAAA
228	277-62	GATCAGAGGTGAAGGGA
229	277-63	GATCATCCTCACTCACT
230	277-64	GATCATGGCAGCATGAA
231	277-65	GATCATTCCTCATTCTG
232	277-66	GATCATTTCTTCTTCTT
233	277-67	GATCCAAATCCCATTAC
234	277-68	GATCCAATGGAGCCTGG
235	277-69	GATCCACATCTCAAAGA
236	277-70	GATCCAGAAGGGGTTTG
237	277-71	GATCCACCTGGAAAGCT
238	277-72	GATCCATCATCCGCAAT
239	277-73	GATCCCACCCATTCTTT
240	277-74	GATCCCACTTCCTGTTT
241	277-75	GATCCCAGGAGAATCAC
242	277-76	GATCCCCAGAGTTGGTC

-continued

SEQ ID NO	Code	Sequence
243	277-77	GATCCCCTGAATGCCTT
244	277-78	GATCCCCTTTGCTGCTA
245	277-79	GATCCCGTTCTGCTGCC
246	277-80	GATCCGACATTTTGGAG
247	277-81	GATCCGCAGGAGGGTGC
248	277-82	GATCCGCTTATTTCTGC
249	277-83	GATCCTATAGGGAGGCC
250	277-84	GATCCTGACTGCTGTCA
251	277-85	GATCCTGGAGGACCCTG
252	277-86	GATCGCACCACTGCACG
253	277-87	GATCGCTTTCTACACTG
254	277-88	GATCGTAATGTTTATCA
255	277-89	GATCTACAACACCTGCC
256	277-90	GATCTACAATGAAGCCC
257	277-91	GATCTATTACTGACCGT
258	277-92	GATCTCCCCGAATCTCA
259	277-93	GATCTCCGATGTGATCA
260	277-94	GATCTGAAAAGGCGTCT
261	277-95	GATCTGAAGCCTGAGTG
262	277-96	GATCTGAGGTAAACTTT
263	277-97	GATCTGCGTGGGGCTGG
264	277-98	GATCTGTGTTGAAAGTC
265	277-99	GATCCTCACCTCTTGGA
266	277-100	GATCTGTAAATAAAACT
267	277-101	GATCTTACCTTTTCAAT
268	277-102	GATCAAAGTGGCTGCAG
269	277-103	GATCAACTGGAACCTCT
270	277-104	GATCAAGCAGTTATTTG
271	277-105	GATCAATAAAATGTGAT
272	277-106	GATCAATTTCTAATTGC
273	277-107	GATCACGGCTCTTTTTA
274	277-108	GATCAGCGCTTTAAAAA
275	277-109	GATCAGTTCTCGTGGTT
276	277-110	GATCATGGCATTTAAAT
277	277-111	GATCATTAAAAATGGCT
278	277-112	GATCCAAATCAAAGTGA
279	277-113	GATCCAGAGGCCATGGA

-continued

SEQ ID NO	Code	Sequence
280	277-114	GATCCCCCAAGTACACC
281	277-115	GATCTAAATAAAATGCT
282	277-116	GATCTAAATCTGAACAG
283	277-117	GATCTATTTTTTAATAA
284	277-118	GATCAAACTCCCCACCC
285	277-119	GATCAACAAGAAATGTT
286	277-120	GATCAACATAATGGACC
287	277-121	GATCAACCATCGCTTTA
288	277-122	GATCAATCCTGAATTTC
289	277-123	GATCACAAGCACAAATC
290	277-124	GATCACTGAGTGTACAG
291	277-125	GATCACTGTTCCAAGCA
292	277-126	GATCAGCAAGCACGAGT
293	277-127	GATCAGCAGGGAGTTTA
294	277-128	GATCAGCAGTTCCAGCC
295	277-129	GATCAGTGTCTCTAGTC
296	277-130	GATCATACCTATTAAAA
297	277-131	GATCATCAAACTGATTA
298	277-132	GATCATCTTGATGTCTA
299	277-133	GATCATGTCTTTTCCAT
300	277-134	GATCATGTGTTCTGGAG
301	277-135	GATCATTGTCAAAAAAT
302	277-136	GATCATTTCAAAACTCA
303	277-137	GATCATTTTATTTTACA
304	277-138	GATCCACAGGGGTGGTG
305	277-139	GATCCACTTCTGTGATT
306	277-140	GATCCAGAACATGGGAA
307	277-141	GATCCAGCTAGGCTGGG
308	277-142	GATCCATCACAAAGCGA
309	277-143	GATCCCAGAAAAGTTCT
310	277-144	GATCCCAGAGAGCAGCT
311	277-145	GATCCCCAAGGAGTTCC
312	277-146	GATCCCCCCAGCCTGAC
313	277-147	GATCCCCGGTGGTTTTG
314	277-148	GATCCCCTCAGAAGGCA
315	277-149	GATCCCGCATGCCTGAA
316	277-150	GATCCCTCTACAGAGCT

-continued

SEQ ID NO	Code	Sequence
317	277-151	GATCCCTCTTTTCCAGA
318	277-152	GATCCCTTCATTTGAAT
319	277-153	GATCCGCCTGGCAGCCA
320	277-154	GATCCTCCCTGCCCGCG
321	277-155	GATCCTGATGCCAATAC
322	277-156	GATCCTGCAGGACTACA
323	277-157	GATCCTTGACGAGGAGA
324	277-158	GATCCTTTCAGCTGCCA
325	277-159	GATCCTTTTTTGTACAT
326	277-160	GATCGTGGAGGAGTGTC
327	277-161	GATCTATCATTTTATTG
328	277-162	GATCTATGTTTGTGTGA
329	277-163	GATCTATTTCTCAGTAA
330	277-164	GATCTATTTGGCCTCTC
331	277-165	GATCTCAGTTCTGCGTT
332	277-166	GATCTGATTATTTACTT
333	277-167	GATCTGCTATTGTTATT
334	277-168	GATCTGGAAGATGAGTC
335	277-169	GATCTGGGATAAAACCA
336	277-170	GATCTGTCTCTGCTGTT
337	277-171	GATCTGTTGGGAAAGAT
338	277-172	GATCTGTTTTATTGATA
339	277-173	GATCTTACACATTCTCT
340	277-174	GATCTTCGACACAGAAA
341	277-175	GATCTTGCAACTCCATT
342	277-176	GATCTTGCCTCTTTCCT
343	277-177	GATCTTTCTTTCCAAAA
344	277-178	GATCTTTGTACGTAATT
345	277-179	GATCCCTACCTGCCTGG
346	277-180	GATCAACATTCGCAATG
347	277-181	GATCATGTCCATATCAT
348	277-182	GATCCCTTACCCCCAGG
349	277-183	GATCCTCCTGACCTCAA
350	277-184	GATCTGTTTTGTACTTT
351	277-185	GATCAAAATTTGTGTAA
352	277-186	GATCAAGGTCCTTTCCG
353	277-187	GATCAAGTAACATGTTG

-continued

SEQ ID NO	Code	Sequence
354	277-188	GATCAATTACCTAACTG
355	277-189	GATCAGCTGCATCTAAA
356	277-190	GATCAGTGTTATATTTT
357	277-191	GATCATAGCTGACTTTA
358	277-192	GATCATGTAGCTGAGAC
359	277-193	GATCCTGTCTGCAGTCA
360	277-194	GATCCATTCAGCCCTGG
361	277-195	GATCTACATTTTGTACA
362	277-196	GATCTGGTCTCTTTGGC
363	277-197	GATCTGTGCAGGGTATT
364	277-198	GATCTGTTTTTCTTAAA
365	277-199	GATCTTACTGCAAAGGA
366	277-200	GATCAACAACCCCTCCC
367	277-201	GATCAAGCGTGCTTTCC
368	277-202	GATCACTTTGAGAAACA
369	277-203	GATCAGACAGAATAATA
370	277-204	GATCAGAGCATTGTGCA
371	277-205	GATCAGCACCTTGTATA
372	277-206	GATCATAATAATTCCAA
373	277-207	GATCATCACATTTTGAT
374	277-208	GATCATTCTTGATTTTG
375	277-209	GATCATTGCTCCTTCTC
376	277-210	GATCATTTCACCTGATG
377	277-211	GATCCAAGTTCCAGTGT
378	277-212	GATCCACATTGTTAGGT
379	277-213	GATCCACCTGCTTATTT
380	277-214	GATCCACTACCGGAAGA
381	277-215	GATCCAGCCATTACTAA
382	277-216	GATCCAGCTCAGAACGA
383	277-217	GATCCAGGCTTCTGCCA
384	277-218	GATCCAGTGTCCATGGA
385	277-219	GATCCCCAAGTGGTGAA
386	277-220	GATCCCCCCTGCCTATC
387	277-221	GATCCCCGTTCTTCAAG
388	277-222	GATCCCCTTTGGTTTTA
389	277-223	GATCCGTTCCGTCGTCG
390	277-224	GATCCTCACCAACCTAA

-continued

SEQ ID NO	Code	Sequence
391	277-225	GATCCTCGAACGGAAAG
392	277-226	GATCCTCTGTCTTCAGT
393	277-227	GATCCTCTTGTACTGGG
394	277-228	GATCCTAACACTAAGGA
395	277-229	GATCCTTACGGAAAAGG
396	277-230	GATCCTTTCCCGAAGCA
397	277-231	GATCGCTCTAACACGAG
398	277-232	GATCGTCTGAGCCCCCC
399	277-233	GATCGTTCTCAGGCCCT
400	277-234	GATCTAACCATTTTCAT
401	277-235	GATCTAAGATGATTATT
402	277-236	GATCTAGATTCTACATG
403	277-237	GATCTAGTAAAGTGTTT
404	277-238	GATCTATCACCTGTCAT
405	277-239	GATCTATGGCCTCTGGT
406	277-240	GATCTCACAGGCTGAGA
407	277-241	GATCTCAGTTGTAAATA
408	277-242	GATCTCCCCTTGGACTG
409	277-243	GATCTGCTAAGACCAGG
410	277-244	GATCTGGGCGAGGAAGT
411	277-245	GATCTGTATGTGTTCTA
412	277-246	GATCTGTCTGTCTGAGC
413	277-247	GATCTGTCTGTTGCTTG
414	277-248	GATCTGTCTTGCATTTC
415	277-249	GATCTGTGTTTGCTCTG
416	277-250	GATCTTACCCGTGACAA
417	277-251	GATCTTCTGTGGTGCTT
418	277-252	GATCTTTCATGTGTTAG

-continued

SEQ ID NO	Code	Sequence
419	277-253	GATCAGTTATGAAGAAG
420	277-254	GATCTGAAGTAATTGTG
421	277-255	GATCCGCATCAGCGACA
422	277-256	GATCTGCTATTGCCAGC
423	277-257	GATCAAAATAAAGCCTC
424	277-258	GATCAAATGGAACATTT
425	277-259	GATCAAGTTTAAATGAC
426	277-260	GATCAATGAATTGCACA
427	277-261	GATCAATGCAACGACGT
428	277-262	GATCAATGCCCTCATTA
429	277-263	GATCAATTGTGCTTTAC
430	277-264	GATCAGAAATGGCTAAT
431	277-265	GATCAGCTGGGTTTTGG
432	277-266	GATCATAAATATTAATG
433	277-267	GATCCAAACTAATATTG
434	277-268	GATCCAACAATAAAATA
435	277-269	GATCCACTACAGAAAGG
436	277-270	GATCCAGTGAATATTCA
437	277-271	GATCCCTGCATTTCCTG
438	277-272	GATCTATTTTTTGCATG
439	277-273	GATCTCTAAAGCAGTAG
440	277-274	GATCTGAATTTGTACCA
441	277-275	GATCTGGTCTAGTTAAC
442	277-276	GATCTTCAGATAAATTC
443	277-277	GATCTTTTTGTAAAAGG

[0243]

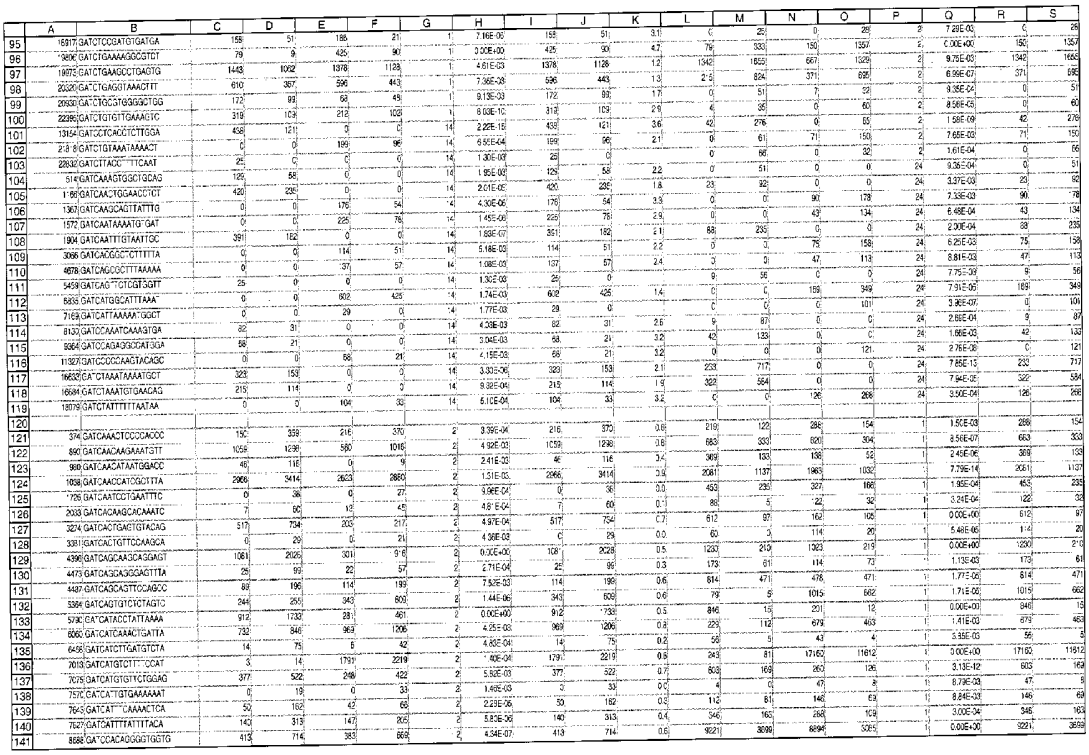
Appendix I

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	ID	signature	Ncho_T	Ycho1_T	Ncho_F	Ycho_F	Ycho_direction	Chi-p	Ncho_taken	Ycho_taken	Ycho_NY_ratio	Bast_T	Diff_T	Bast_F	Diff_F	HLPa_direction	HLPa_p	Bast_taken	Diff_taken
2	2850	GATCACAGGCTTGCTT	53	9	15	12	1	6.11E-04	53	9	5.9	30	9	99	20	1	3.51E-04	99	20
3	2678	GATCACCATCCCAGTCA	1665	1018	3671	1652	1	3.00E-09	3671	652	2.2	93	61	1232	832	1	1.42E-04	1232	832
4	3357	GATCACTGTCTATCAC	208	180	379	258	1	9.95E-03	379	258	1.5	299	60	478	243	1	1.39E-05	478	243
5	3988	GATCAGAATCATGGTGT	4938	3852	6206	5346	1	5.66E-04	6206	5346	1.1	1178	343	2668	1272	1	0.00E+00	2668	1272
6	4329	GATCAGATTCCGATTTG	64	34	34	34	1	2.10E-04	64	34	3.9	102	20	295	126	1	8.51E-04	256	126
7	6533	GATCATGAATAGGASCC	237	156	212	72	1	2.00E-03	237	156	1.7	228	61	135	50	1	1.91E-04	205	91
8	6633	GATCATGATTTTGTAGT	57	21	102	27	1	7.08E-05	108	27	4.0	41	3	122	44	1	2.76E-06	122	44
9	7158	GATCATGTTTTCGTACA	147	72	106	42	1	2.86E-03	147	72	2.0	341	117	327	142	1	1.99E-05	327	142
10	7402	GATCATTCCTGTGTAG	5458	4102	4601	3551	1	4.44E-16	5458	4102	1.3	5543	4315	4591	3304	1	2.59E-06	5543	4319
11	9622	GATCCACACACGTTGGT	2265	1571	2132	1655	1	1.19E-06	2132	1655	1.3	1656	1101	1442	905	1	2.02E-06	1656	1101
12	10506	GATCCCAAATTTGTCCA	438	4	429	9	1	0.00E+00	438	4	109.5	238	122	260	146	1	4.66E-07	260	146
13	10530	GATCCCACAGGCCTTAG	21	0	6	6	1	2.91E-03	21	0		37	0	43	0	1	1.08E-03	43	0
14	10900	GATCCCAGGATTCAGTA	819	654	884	693	1	6.65E-38	884	693	1.3	1052	581	1043	658	1	6.12E-06	1052	681
15	11056	GATCCCATCTCTACGCC	67	9	75	21	1	1.63E-03	75	21	3.6	46	0	57	6	1	1.97E-03	58	6
16	11177	GATCCCCAACAATGTCA	79	51	651	452	1	6.53E-04	651	452	1.4	42	30	1762	903	1	0.00E+00	1762	902
17	11434	GATCCCCGGCCTCAGTC	1592	394	1631	1318	1	1.11E-03	1631	1318	1.2	3087	1563	3493	1914	1	0.00E+00	3493	1914
18	11922	GATCCCTGAAGTTGCCC	244	118	262	144	1	9.75E-34	262	144	1.8	257	66	256	69	1	2.53E-07	256	69
19	14038	GATCCTGGAGGATTTGC	11416	9182	10644	9168	1	0.00E+00	11416	9182	1.2	8554	5329	5953	4259	1	7.49E-07	6554	5329
20	14557	GATCCTCAGCACAGGA	146	31	812	3	1	9.03E+00	1461	31	47.1	706	452	130	81	1	7.33E-04	706	452
21	16134	GATCGTGTTGTATGGTG	215	102	196	97	1	1.48E-04	215	102	2.1	173	56	241	52	1	4.79E-08	241	52
22	16280	GATCGTGTATTGAGATT	79	29	121	53	1	8.61E-03	121	57	2.1	224	117	165	113	1	9.68E-03	224	117
23	16309	GATCGTTGACAAGTATG	43	9	19	12	1	4.76E-03	43	9	4.8	28	10	51	5	1	3.34E-03	51	6
24	17758	GATCTATCATTACTGCA	754	338	638	349	1	4.75E-14	754	338	2.2	711	420	824	321	1	1.32E-04	711	420
25	9523	GATCTCTGTGCTGTAAA	197	121	275	155	1	1.26E-03	275	56	1.8	327	122	501	304	1	5.42E-04	501	304
26	20546	GATCTGATTATTATTATT	99	36	91	27	1	4.34E-03	89	36	2.5	364	215	312	264	1	5.24E-03	364	215
27	21128	GATCTGGTTGGAGTTTT	64	12	65	12	1	5.19E-04	65	12	5.4	86	30	110	29	1	5.27E-04	110	28
28	21208	GATCTGGAACCTCAGCC	502	102	563	247	1	7.75E-10	552	247	2.2	481	159	766	425	1	9.37E-07	766	426
29	21423	GATCTGGCATGTTAGCC	25	12	36	3	1	2.43E-03	36	3	12.0	90	0	311	0	1	5.70E-04	60	0
30	21748	GATCTGGTGTGAGTGCA	229	152	317	144	1	4.89E-06	317	144	2.2	252	107	241	101	1	6.24E-04	252	107
31	21876	GATCTGTACACAGTAAA	32	26	62	18	1	4.58E-03	62	18	3.4	66	9	82	8	1	9.84E-05	82	8
32	21878	GATCTGTACAGACACAGGA	107	55	127	57	1	3.26E-03	127	57	2.2	201	56	252	110	1	2.98E-04	252	110
33	21894	GATCTGTAGCTGAGAGG	65	17	66	18	1	1.40E-11	185	18	10.3	107	66	154	35	1	8.63E-04	154	56
34	22090	GATCTGTCTATGGGACC	203	126	160	114	1	9.27E-03	205	126	1.7	308	143	337	125	1	4.96E-04	308	143
35	24167	GATCTTTCCAACCACAT	125	46	76	45	1	2.52E-04	125	46	2.7	121	42	67	28	1	4.75E-03	121	40
36	1094	GATCAACGCCTCACTGA	25	24	651	392	1	5.21E-06	651	392	1.7	0	0	1296	564	14	0.00E+00	1296	564
37	8098	GATCCAAAGTCATGTGT	3	0	324	78	1	2.64E-12	324	78	4.2	3	0	114	28	14	3.47E-04	114	28
38	14394	GATCCTGTTTCCATTTG	10	0	311	183	1	1.22E-03	311	183	1.7	0	0	76	18	14	5.06E-04	76	12
39	21813	GATCTGTAAAATGTGAT	92	0	232	141	1	7.83E-03	232	141	1.6	0	0	312	186	14	5.20E-03	312	186
40	7553	GATCATTGGTTCCAGTC	118	12	0	0	14	5.29E-09	118	12	9.8	163	15	3	0	1	8.83E-07	163	15
41	1234	GATCAACTTGAGTCCAA	240	136	0	0	14	1.44E-05	240	136	1.8	93	20	0	0	4	2.31E-03	93	20
42	2778	GATCACCGCTTTCCAAT	0	0	55	2	14	2.46E-03	55	12	4.6	0	0	102	36	14	5.27E-03	102	36
43	4159	GATCAGAGCTCAGTTCC	0	0	907	370	14	0.00E+00	907	370	2.5	0	0	260	138	14	2.22E-03	260	138
44	4728	GATCAGCTGAACAGCAG	93	12	0	0	14	7.99E-07	93	12	7.8	145	46	0	0	14	1.40E-03	45	46
45	7052	GATCATGTGCTACTGGT	165	89	0	0	14	5.80E-03	165	89	1.9	163	66	0	0	14	4.19E-03	163	66
46	10880	GATCCCAGCTGATGTAG	0	0	1012	681	14	5.44E-06	1012	681	1.5	0	0	861	602	14	1.06E-03	561	609
47	12938	GATCCTAGACAGGGCTC	38	0	0	0	14	2.67E-04	38	0		46	0	0	0	14	2.51E-03	46	0

Appendix I

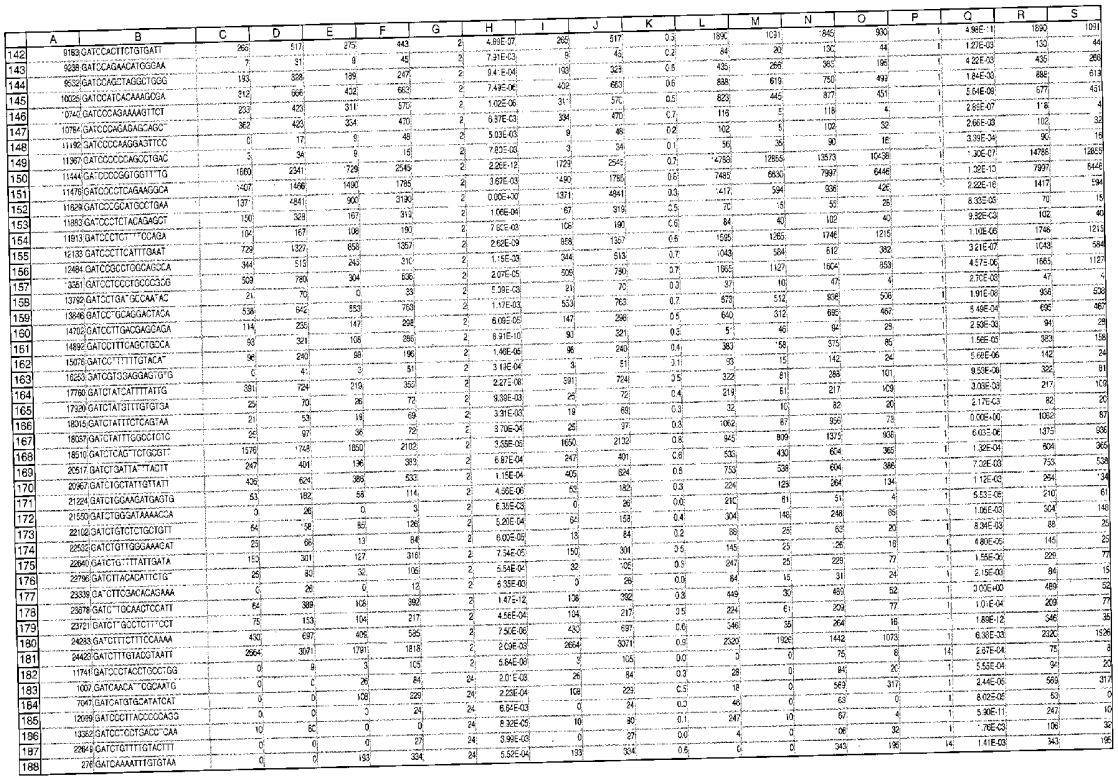
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
48	1526	GATCGAGCTGGCCTATG	0	0	491	156	14	7.00E-14	491	156	3.1	0	0	1315	829	14	1.58E-07	1315	829
49	1526	GATCGAGGCTTGTGATG	0	0	229	6	14	2.22E-16	229	6	25.4	0	0	3	0	14	5.29E-03	31	0
50	17704	GATCTA AC AGATAAT	4474	3871	0	0	14	1.25E-04	4474	3871	1.2	84	0	0	0	14	5.03E-05	94	0
51	19066	GATCTCGAACCCTG CT	801	103	0	0	14	1.02E-06	801	109	2.8	88	0	0	0	14	1.53E-05	93	0
52	22899	GATCT AGCTTTGATAA	0	2	140	72	14	7.79E-03	140	72	1.9	0	0	300	154	14	6.52E-04	300	154
53	24215	GATCTTTAATGCTTTGG	96	58	0	3	14	1.59E-03	96	36	2.7	74	5	0	0	4	5.52E-04	74	5
54																			
55	194	GATCAAAAGGGACAAGC	71	31	85	27	1	1.81E-05	85	27	3.1	0	5	0	5	2	9.35E-04	0	5
56	329	GA GAAACCAAGGCCCA	642	437	1599	1194	1	8.43E-05	1599	1194	1.3	74	636	1502	2117	2	4.31E-07	1509	2117
57	1073	GATCAACCTGCAGGTCT	662	136	1902	705	1	0.00E+00	1902	705	2.7	0	532	162	910	2	0.00E+00	162	910
58	1265	GATCAAGAACAATGCCT	750	685	1568	1191	1	4.76E-05	1568	1191	1.3	37	148	237	751	2	2.22E-16	237	751
59	2344	GATCACAGGCAAGCCA	17	12	153	78	1	4.92E-03	153	75	2.0	0	10	15	97	2	1.16E-04	15	97
60	2307	GATCACAGGGGTGA GG	25	0	16	0	1	1.30E-03	25	0		46	87	59	196	2	1.28E-05	59	196
61	2476	GATCACATCTGTGTGAA	264	26	317	211	1	8.66E-03	317	211	1.5	23	71	59	134	2	7.01E-03	59	134
62	2484	GATCACATGAATAGGGG	125	36	216	66	1	2.75E-07	216	66	3.3	46	245	110	556	2	0.00E+00	110	556
63	3727	GATCAGAAAAGGAGAAA	233	34	252	84	1	1.90E-07	252	84	3.0	130	153	189	378	2	7.55E-05	189	378
64	4158	GATCAGAGGTGAAGGGA	28	0	28	0	1	2.88E-04	28	0		0	51	0	63	2	4.97E-05	0	63
65	6260	GATCATCCTCACTCAGT	66	12	32	21	1	3.29E-06	66	12	7.2	107	225	27	77	2	5.27E-03	107	225
66	6824	GATCATGCGAGCATGAA	287	63	465	51	1	0.00E+00	465	51	9.1	88	225	252	422	2	1.08E-03	252	422
67	7260	GATCATTCCTCATTCTG	25	0	13	0	1	1.90E-03	25	0		0	35	0	20	2	5.63E-03	0	35
68	7725	GATCATTTC TTCTT	17	3	16	0	1	6.58E-03	17	0		14	107	0	42	2	9.58E-06	14	107
69	8125	GATCCAAATCCCATTAC	111	29	66	66	1	2.24E-05	111	29	3.8	135	278	134	215	2	1.57E-03	135	278
70	8523	GATCCAATGGACCCTGC	636	21	759	6	1	0.00E+00	759	6	126.5	4	154	15	196	2	1.98E-09	4	184
71	8728	GATCCACATCTCAAAGA	5537	4796	6638	5220	1	0.00E+00	6638	5220	1.3	1553	4580	397	4572	2	1.12E-03	3971	4572
72	9259	GATCCAGAAGGCGTTTG	140	48	121	57	1	5.60E-05	140	48	2.9	121	242	67	101	2	3.16E-03	121	245
73	9549	GATCCAGCTGGAAAGCT	603	328	666	313	1	8.52E-09	603	328	2.0	388	571	395	666	2	5.64E-10	388	671
74	10049	GATCCATCATCCGCAAT	2388	2084	2624	1348	1	8.20E-03	2388	2084	1.1	897	967	691	1129	2	1.22E-11	837	1167
75	10681	GATCCCACCCATTCTTT	233	60	6	0	1	1.84E-07	233	60	2.9	74	198	0	252	2	5.77E-04	74	198
76	13728	GATCGCACTTCCTGTTT	21	19	42	6	1	7.61E-03	42	9	4.7	100	22	110	211	2	4.49E-03	110	211
77	13880	GATCCCAGGAGAATCAG	193	107	206	114	1	3.66E-03	206	114	1.8	65	158	90	26	2	2.19E-07	65	158
78	11259	GATCCCCAGAGTTGGTC	736	471	808	468	1	4.84E-07	808	468	1.7	411	805	438	768	2	2.19E-07	411	809
79	11513	GATCCCCTGAATGCCTT	206	94	271	162	1	3.06E-05	271	62	4.7	150	245	159	329	2	1.03E-04	159	329
80	11582	GATCCCCTTTGCTGCTA	14	0	29	3	1	7.75E-03	29	3	9.7	0	97	0	69	2	5.06E-06	0	97
81	11636	GATCCCG TCTGCTGCC	57	4	121	39	1	2.55E-04	121	39	3.1	28	97	27	69	2	4.21E-03	27	89
82	12516	GATCCGACATTTTGCAG	136	70	154	57	1	6.40E-03	136	70	1.9	9	6	15	44	2	4.43E-03	9	61
83	12563	GATCCGCAGCAGGGTGG	58	21	42	21	1	4.33E-05	93	21	4.4	4	87	7	44	2	7.24E-05	4	87
84	12515	GATCCGCTTATTTCTGC	964	532	1264	615	1	0.00E+00	1264	615	2.1	1019	1378	944	1142	2	9.03E-04	1019	1378
85	12999	GATCCTATAGGGAGGCC	71	20	81	6	1	2.67E-04	81	15	4.5	4	153	19	97	2	4.58E-08	4	153
86	13715	GATCCTGAC GCTGTCA	693	328	867	352	1	0.00E+00	867	352	2.5	0	40	7	28	2	3.08E-03	0	40
87	4036	GATCCTGGAGGACCCTG	420	182	419	47	1	6.34E-09	490	182	2.3	42	117	47	126	2	2.79E-03	47	126
88	15317	GATCGCACCACTGCACC	32	21	228	30	1	2.96E-13	228	30	8.0	4	5	15	182	2	3.37E-08	15	182
89	15731	GATCGCTTTGTACACTG	413	274	576	352	1	3.23E-05	576	352	1.6	345	906	744	1159	2	5.55E-06	746	1150
90	1607	GATCGTAA GTTTATCA	919	36	918	36	1	0.00E+00	1018	36	28.5	271	461	395	837	2	2.82E-10	395	837
91	17048	GATCTACACACACCTGCC	693	457	717	467	1	3.9HE-05	7 7	467	1.5	290	685	284	559	2	1.05E-03	284	559
92	17066	GATCTACAATGAACCCC	17443	2601	18091	11666	1	0.00E+00	17443	12901	1.4	3312	7204	2916	6096	2	0.00E+00	33 2	7204
93	17331	GATC ATTAGTGACGGT	1269	299	1602	271	1	0.00E+00	1059	299	3.5	168	368	177	300	2	8.12E-05	168	368
94	18360	GATC CCGCGAATCTCA	61	60	114	51	1	5.18E-03	114	51	2.2	4	76	19	69	2	8.67E-03	19	69

Appendix I

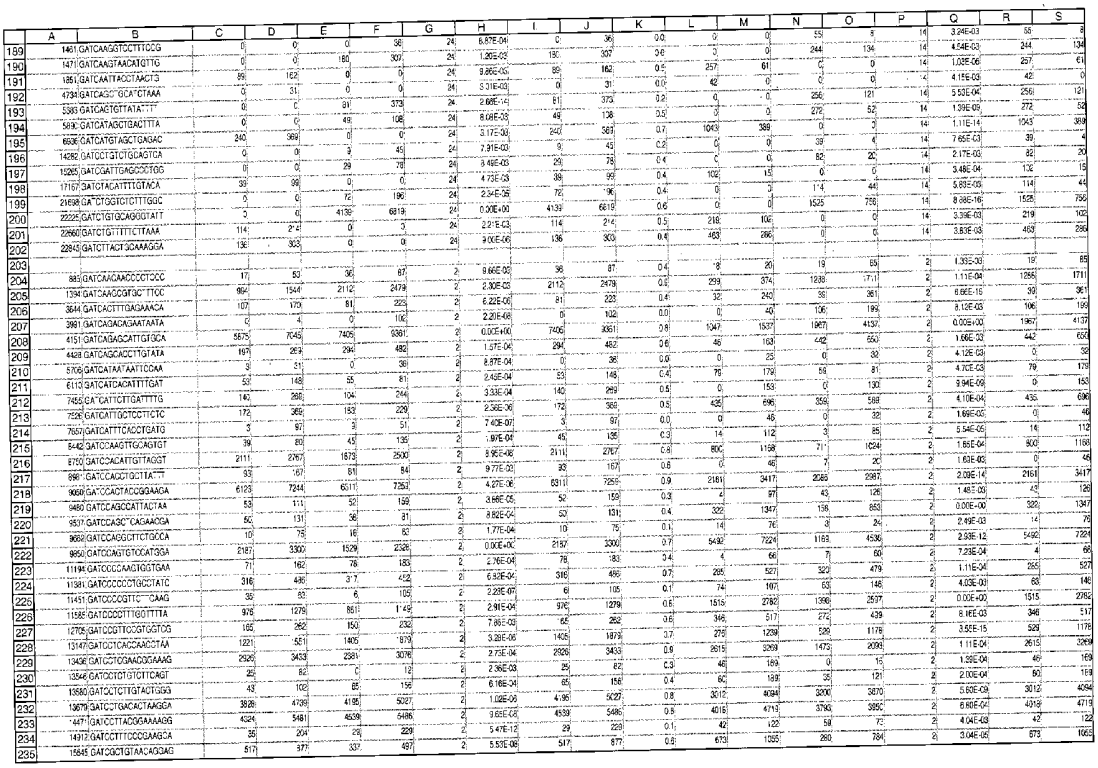


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
95	16917	GATCTCCBATGTGATGA	155	51	186	21	1	7.16E-09	155	51	3.1	0	25	0	25	2	7.29E-02	0	25
96	2456	GATCTGAAAAGGCGTCT	79	9	435	90	1	3.00E+00	425	90	4.7	79	333	150	1357	2	0.00E+00	150	1357
97	19073	GATCTGAAGCCTGAGTG	1443	1092	1978	1128	1	4.61E-03	1978	1128	1.2	1342	1655	667	1529	2	9.75E-03	1342	1655
98	20220	GATCTGAGGTAAACTTT	610	367	596	443	1	7.96E-28	596	443	1.3	215	824	371	695	2	6.99E-07	371	695
99	20930	GATCTGGGTGGGGCTGG	122	99	68	45	1	9.13E-03	122	99	1.2	4	35	0	60	2	8.56E-05	0	60
100	22286	GATCTGTGTTGAAAGTC	319	103	212	103	1	8.60E-10	319	103	2.9	4	35	0	53	2	1.58E-09	42	276
101	13154	GATCCTCACCTCTTGGA	458	121	0	0	14	2.22E-18	458	121	3.6	42	276	0	53	2	7.65E-03	71	150
102	218	GATCTGTAAATAAAACT	0	0	199	99	14	6.92E-04	199	99	2.1	0	61	71	150	2	1.61E-04	0	66
103	22632	GATCTTACC TTTCAAT	25	0	0	0	14	1.30E-07	25	0		0	66	0	32	2	9.36E-04	0	51
104	514	GATCAAAGTGGCTGCAG	125	58	0	0	14	1.95E-02	125	58	2.2	0	511	0	0	24	3.37E-03	25	92
105	1 86	GATCAACTGGAACCTCT	420	235	0	0	14	2.01E-05	420	235	1.8	23	92	0	0	24	7.33E-03	90	78
106	1367	GATCAAGCAGTTATTTG	0	0	175	54	4	4.30E-06	175	54	3.3	0	0	90	178	24	6.48E-04	43	134
107	1572	GATCAATAAAATG GAT	0	0	225	78	14	1.45E-06	225	78	2.9	0	0	43	134	24	2.00E-04	88	235
108	1904	GATCAATTTGTAATTGC	391	182	0	0	14	1.58E-07	391	182	2.1	88	235	0	0	24	6.20E-03	75	156
109	3066	GATCACGGCTCTTTTTA	0	0	114	51	14	5.18E-03	114	51	2.2	0	0	75	156	24	8.81E-03	47	113
110	4678	GATCAGCGCTTTAAAAA	0	0	57	57	14	1.08E-05	137	57	2.4	0	91	47	113	24	7.75E-09	9	56
111	5459	GATCAG TTCTCGTGGTT	25	0	0	0	14	1.30E-09	25	0		9	56	0	0	24	7.91E-05	169	349
112	6838	GATCATGGCATTTAAA	0	0	602	425	14	1.74E-03	602	425	1.4	0	0	169	349	24	3.90E-07	0	101
113	7169	GATCATTAAAAA GGCT	0	0	29	0	4	1.77E-03	29	0		0	0	0	101	24	2.88E-04	9	87
114	8132	GATCCAAATCAAAGTGA	82	31	0	0	14	4.20E-03	82	31	2.6	9	87	0	0	24	1.66E-03	42	132
115	8964	GATCCAGAGGCGATGGA	58	21	0	0	14	3.04E-03	58	21	3.2	42	132	0	0	24	2.76E-06	0	121
116	11207	GATCCCCCAAGTACAGC	0	0	58	21	14	4.15E-06	58	21	3.2	0	0	0	121	24	7.85E-15	230	717
117	16632	GA TCTAAATAAAATGCT	323	153	0	0	14	3.33E-06	323	153	2.1	233	717	0	0	24	7.94E-05	322	584
118	16584	GATCTAAATGTGAACAG	215	114	0	0	14	9.32E-04	215	114	1.9	322	584	0	0	24	3.50E-04	126	268
119	18079	GATCTATTTTTTAATAA	0	0	104	33	14	5.10E-04	104	33	3.2	0	0	126	268	24			
120																			
121	374	GATCAAACTCCCCACCC	190	359	216	370	2	3.39E-04	216	370	0.6	219	122	286	154	1	1.50E-03	286	154
122	890	GATCAACAAGAAATGTT	1055	1236	580	1016	2	4.92E-03	1059	1236	0.8	663	333	820	304	1	8.56E-07	663	333
123	960	GATCAACATAATGGACC	46	116	0	9	2	2.41E-03	46	116	3.4	369	193	108	52	1	2.49E-06	369	193
124	1058	GATCAACCATCGGTTTA	2969	3414	2623	2880	2	1.31E-03	2966	3414	0.9	2081	1137	1963	1032	1	7.79E-14	2081	1137
125	1726	GATCAATCCTGAATTTC	0	36	0	27	2	9.99E-04	0	36	0.0	453	233	327	166	1	1.95E-04	453	233
126	2033	GATCACAAGCACAATC	7	90	12	45	2	4.81E-04	7	60	0.1	88	5	22	32	1	3.24E-04	122	32
127	3274	GATCACTGAGTGTACAG	517	754	203	217	2	4.97E-04	517	754	0.7	612	97	162	105	1	0.00E+00	612	97
128	3311	GATCACTGTTCCAAGCA	0	29	0	21	2	4.36E-03	0	29	0.0	60	3	114	20	1	5.46E-06	114	20
129	4398	GATCAGCAAGCAGGAGT	1061	2025	301	916	2	0.00E+00	108	2028	0.5	1232	210	1023	219	1	0.00E+00	1230	210
130	4473	GATCAGCAGCGGAGTTTA	25	99	22	57	2	2.71E-04	25	99	0.3	173	61	114	73	1	1.13E-03	173	61
131	4497	GATCAGCAGTTCCAGCC	89	196	114	189	2	7.52E-03	114	196	0.6	814	471	478	471	1	1.77E-05	814	471
132	5364	GATCAGTGTCTCTAGTC	244	250	343	800	2	1.44E-06	343	609	0.6	79	9	1015	662	1	1.71E-05	1015	662
133	5790	GA TCATACCTATTAAAA	912	1230	281	461	2	0.00E+00	912	733	0.5	846	15	291	15	1	0.00E+00	846	15
134	6060	GATCATCAAACTGATTA	732	846	969	1206	2	4.25E-03	969	1206	0.8	229	112	679	463	1	1.41E-03	679	463
135	6458	GATCATCTTGATGTCTA	14	75	5	42	2	4.83E-04	14	75	0.2	56	5	43	4	1	3.35E-03	56	5
136	7013	GATCATGTCTT CCAT	3	14	1791	2219	2	40E-04	1791	2219	0.8	243	81	17162	11612	1	0.00E+00	17162	11612
137	7075	GATCATGTGTTCTGGAG	377	522	249	422	2	5.82E-03	377	522	0.7	603	169	260	126	1	3.13E-12	603	169
138	7570	GATCATTGTGAAAAAAT	0	19	0	33	2	1.49E-03	0	33	0.0	4	0	47	9	1	8.79E-03	47	9
139	7643	GATCAT TCAAAACTCA	50	162	42	68	2	2.23E-06	50	162	0.3	112	81	146	63	1	8.84E-03	146	63
140	7627	GATCATTTTATTTTACA	140	313	147	265	2	5.83E-06	140	313	0.4	346	163	268	109	1	3.00E-04	346	163
141	8688	GATCCACACGGGTGGTG	413	714	383	669	2	4.34E-07	413	714	0.6	9221	3699	6894	3065	1	0.00E+00	9221	3699

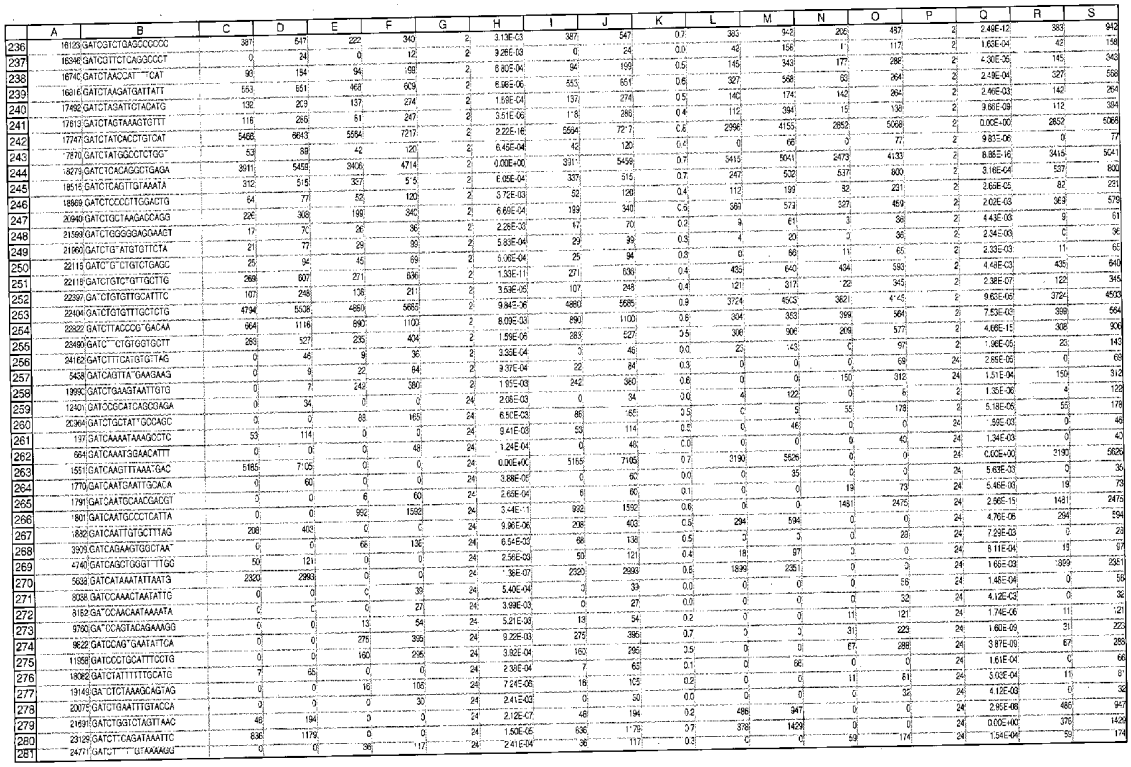
Appendix I



Appendix I



Appendix I



Appendix 1

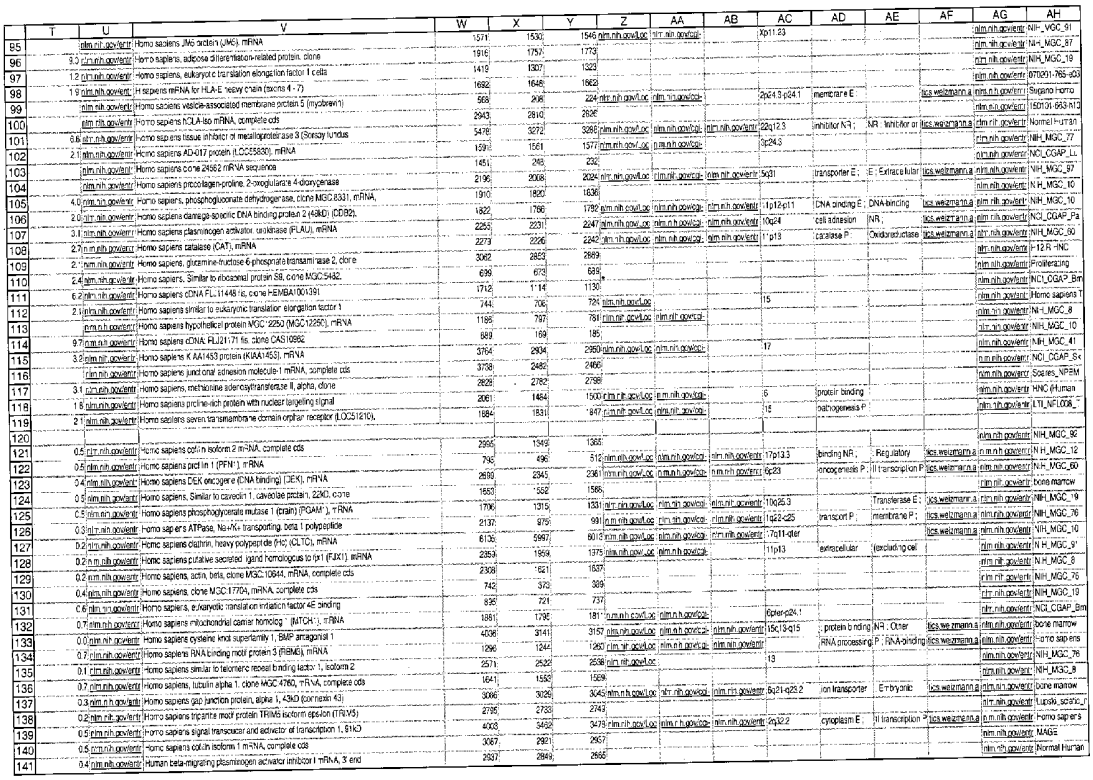
	T	U	V	W	X	Y
1	HuPa_D/B_ratio	LOCUS	DESCRIPTION	LENGTH	MATCH_START	MATCH_END
2	0.2	nlm.nih.gov/entr	Homo sapiens beta-catenin-interacting protein ICAT (LOC56998), mRNA	2334	2686	2702
3	0.7	nlm.nih.gov/entr	Homo sapiens heat shock 27kD protein 1 (HSPB1), mRNA	847	641	657
4	0.5	nlm.nih.gov/entr	Homo sapiens, APEX nuclease (multifunctional DNA repair enzyme).	1446	1137	1153
5	0.4	nlm.nih.gov/entr	Homo sapiens annexin A2 (ANXA2), mRNA	1362	913	929
6	0.5	nlm.nih.gov/entr	Homo sapiens SMT3 (suppressor of mif two 3, yeast) homolog 2	774	261	277
7	0.4	nlm.nih.gov/entr	Homo sapiens acetyl-Coenzyme A acyltransferase 2 (mitochondrial	1584	1397	1413
8	0.4	nlm.nih.gov/entr	Homo sapiens similar to proteasome (prosome, macropain) 26S subunit,	1560	1088	1104
9	0.4	nlm.nih.gov/entr	Homo sapiens HIV-1 Rev binding protein (HRB), mRNA	2584	2506	2524
10	0.8	nlm.nih.gov/entr	Homo sapiens, peptidylprolyl isomerase A (cyclophilin A), clone	753	559	575
11	0.7	nlm.nih.gov/entr	Homo sapiens mRNA; cDNA DKFZp586J021 (from clone DKFZp586J021)	1696	1296	1312
12	0.6	nlm.nih.gov/entr	Homo sapiens, farnesyl-diphosphate farnesyltransferase 1, clone	1709	1331	1347
13	0.0	nlm.nih.gov/entr	Homo sapiens cDNA FLJ14465 fis, clone MAMMA1000388, highly similar to	1917	1714	1730
14	0.6	nlm.nih.gov/entr	Homo sapiens neural precursor cell expressed, developmentally	605	483	499
15	0.1	nlm.nih.gov/entr	Homo sapiens UDP-N-acetyl-alpha-D-galactosamine:polypeptide	2881	2292	2308
16	0.5	nlm.nih.gov/entr	Homo sapiens, tubulin beta 5, clone MGC:4029, mRNA, complete cds	4078	2737	2753
17	0.5	nlm.nih.gov/entr	Homo sapiens adaptor-related protein complex 2, mu 1 subunit (AP2M1),	1827	1725	1741
18	0.3	nlm.nih.gov/entr	Homo sapiens mRNA; cDNA DKFZp564J0323 (from clone DKFZp564J0323)	2199	2073	2089
19	0.8	nlm.nih.gov/entr	Homo sapiens thymosin, beta 10 (TMSB10), mRNA	452	201	217
20	0.6	nlm.nih.gov/entr	Homo sapiens stearoyl-CoA desaturase (SCD) mRNA, complete cds	5221	4832	4848
21	0.2	nlm.nih.gov/entr	Homo sapiens apoptosis-associated nuclear protein (PHLDA1) gene,	4262	3647	3663
22	0.5	nlm.nih.gov/entr	Homo sapiens retinoblastoma 1 (including osteosarcoma) (RB1) mRNA	4741	4589	4605
23	0.2	nlm.nih.gov/entr	Homo sapiens calumenin (CALU), mRNA	3317	929	945
24	0.6	nlm.nih.gov/entr	Homo sapiens, Integrin beta 4 binding protein, clone MGC:1359,	1118	979	995
25	0.6	nlm.nih.gov/entr	Homo sapiens, guanine nucleotide binding protein, beta 1, clone	1673	1545	1561
26	0.6	nlm.nih.gov/entr	Homo sapiens oxoglutarate dehydrogenase (lipoamide) (OGDH), mRNA	4098	4076	4092
27	0.3	nlm.nih.gov/entr	Homo sapiens host cell factor C1 (VP16-accessory protein) (HCFC1),	7711	7320	7336
28	0.6	nlm.nih.gov/entr	Homo sapiens integrin, alpha 5 (fibronectin receptor, alpha	4206	3974	3990
29	0.0	nlm.nih.gov/entr	Homo sapiens a disintegrin and metalloproteinase domain 12 (meltrin	4652	4608	4624
30	0.4	nlm.nih.gov/entr	Homo sapiens cDNA FLJ14929 fis, clone PLACE1009130, weakly similar to	2470	2320	2336
31	0.1	nlm.nih.gov/entr	Human basic transcription factor 3a (BTF3a) gene, complete cds	3953	3142	3158
32	0.4	nlm.nih.gov/entr	Homo sapiens filamin B, beta (actin-binding protein-278) (FLNB), mRNA	9427	8932	8948
33	0.4	nlm.nih.gov/entr	Homo sapiens mevalonate kinase (mevalonic aciduria) (MVK), mRNA	1952	1889	1905
34	0.5	nlm.nih.gov/entr	Homo sapiens PEF protein with a long N-terminal hydrophobic domain	1641	1315	1331
35	0.3	nlm.nih.gov/entr	Homo sapiens vanilloid receptor-like protein 1 (VRL-1), mRNA	2397	2370	2386
36	0.4	nlm.nih.gov/entr	Homo sapiens, pyruvate kinase, muscle, clone MGC:14360, mRNA,	2338	1837	1853
37	0.2	nlm.nih.gov/entr	Homo sapiens mRNA; cDNA DKFZp586P1622 (from clone DKFZp586P1622)	1593	1528	1544
38	0.2	nlm.nih.gov/entr	H.sapiens mRNA for 3'UTR of unknown protein	1420	771	787
39	0.6	nlm.nih.gov/entr	Homo sapiens DKFZp564J157 protein (DKFZP564J157), mRNA	1112	792	808
40	0.1	nlm.nih.gov/entr	Homo sapiens, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase	1873	1226	1242
41	0.2	nlm.nih.gov/entr	Homo sapiens mannosidase, alpha, class 2A, member 1 (MAN2A1), mRNA	4104	3894	3910
42	0.4	nlm.nih.gov/entr	Homo sapiens, clone MGC:10295, mRNA, complete cds	3262	3095	3111
43	0.5	nlm.nih.gov/entr	Homo sapiens meningioma expressed antigen 5 (hyaluronidase) (MGEA5).	5146	5076	5092
44	0.3	nlm.nih.gov/entr	Homo sapiens calpastatin mRNA, complete cds	2290	238	254
45	0.4	nlm.nih.gov/entr	Homo sapiens surfeit 1 (SURF1), mRNA	1019	980	996
46	0.7	nlm.nih.gov/entr	Homo sapiens crystallin, alpha B (CRYAB), mRNA	691	394	410
47	0.0	nlm.nih.gov/entr	Homo sapiens hypothetical protein FLJ11136 (FLJ11136), mRNA	1767	1152	1168

	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	LOCUSLINK	UNIGENE	OMIM	CYTOGENETIC	ENE_ONTOLOG	HER_ONTOLOG	GENECARD	EST_LOCUS	ST_DESCRIPTIO
2	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		1p36.1-p36.32	E ; cell	transcription E ;		nlm.nih.gov/entr	LTI_NFL001_N
3	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	7q	heat shock	Inhibitor or	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_19
4								nlm.nih.gov/entr	NCI_CGAP_Brn
5	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	15q21-q22	membrane E ;	Development	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_42
6	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	6	related protein	conjugation	tics.weizmann.a	nlm.nih.gov/entr	220101-105-h09
7	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	18	NR ; 3-oxoacyl-	generation NR ;	tics.weizmann.a	nlm.nih.gov/entr	290200-017-h09
8	nlm.nih.gov/Loc			14q13.1-q13.3				nlm.nih.gov/entr	Proliferating
9	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	2q36	nuclear pore P ;	Nuclear P ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_89
10								nlm.nih.gov/entr	NIH_MGC_6
11								nlm.nih.gov/entr	Normal Human
12								nlm.nih.gov/entr	NIH_MGC_15
13								nlm.nih.gov/entr	190101-812-g05
14	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	14q12-14q21.3	protein	degradation NR	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_19
15	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	12q13	glycosylation E ;	Transferase E :	tics.weizmann.a	nlm.nih.gov/entr	141299-031-
16								nlm.nih.gov/entr	NIH_MGC_8
17	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	3q28	; vesicle	protein E ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_102
18								nlm.nih.gov/entr	G3.R.R HNC
19	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	2	P ; spermatid		tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_19
20								nlm.nih.gov/entr	NIH_MGC_10
21								nlm.nih.gov/entr	110101-112-b04
22	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr		chromatin P ;	Transcription E	tics.weizmann.a	nlm.nih.gov/entr	Proliferating
23	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr		P ; calcium		tics.weizmann.a	nlm.nih.gov/entr	080101-012-
24								nlm.nih.gov/entr	NIH_MGC_98
25								nlm.nih.gov/entr	Normal Human
26	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	7p14-p13	pathways P ;	generation P ;	tics.weizmann.a	nlm.nih.gov/entr	NCI_CGAP_Brn
27	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	Xq28	factor NR ;	transcription NR	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_12
28	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	12q11-q13	adhesion P ;	membrane NR	tics.weizmann.a	nlm.nih.gov/entr	HOA (Human
29	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	10q26.3	P ; plasma	Protease (other	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_79
30								nlm.nih.gov/entr	NIH_MGC_60
31								nlm.nih.gov/entr	Athersys RAGE
32	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	3p14.3	cytoskeleton P ;	NR :	tics.weizmann.a	nlm.nih.gov/entr	200201-456-
33	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	12q24	kinase E ;		tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_93
34	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		1p34	P ;			nlm.nih.gov/entr	030101-677-g06
35	nlm.nih.gov/Loc	nlm.nih.gov/cgi-			ion transporter	E ; Integral		nlm.nih.gov/entr	NCI_CGAP_Lu
36								nlm.nih.gov/entr	NIH_MGC_12
37								nlm.nih.gov/entr	NIH_MGC_79
38								nlm.nih.gov/entr	LTI_FL013_FB
39	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		19q13.1				nlm.nih.gov/entr	NIH_MGC_42
40								nlm.nih.gov/entr	NIH_MGC_12
41	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	5q21-q22	NR ; membrane	Golgi E ;	tics.weizmann.a	nlm.nih.gov/entr	NT2RP4 Homo
42								nlm.nih.gov/entr	LTI_NFL001_N
43	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	10q24.1-q24.3	degradation E ;	Extracellular	tics.weizmann.a	nlm.nih.gov/entr	Soares_NPBM
44								nlm.nih.gov/entr	NIH_MGC_49
45	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	9q33-q34	transport P ;	generation P ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_77
46	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	11q22.3-q23.1	nucleus NR :	NR ;	tics.weizmann.a	nlm.nih.gov/entr	100401-213-h03
47	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		11cen-q12.1				nlm.nih.gov/entr	NCI_CGAP_Kid

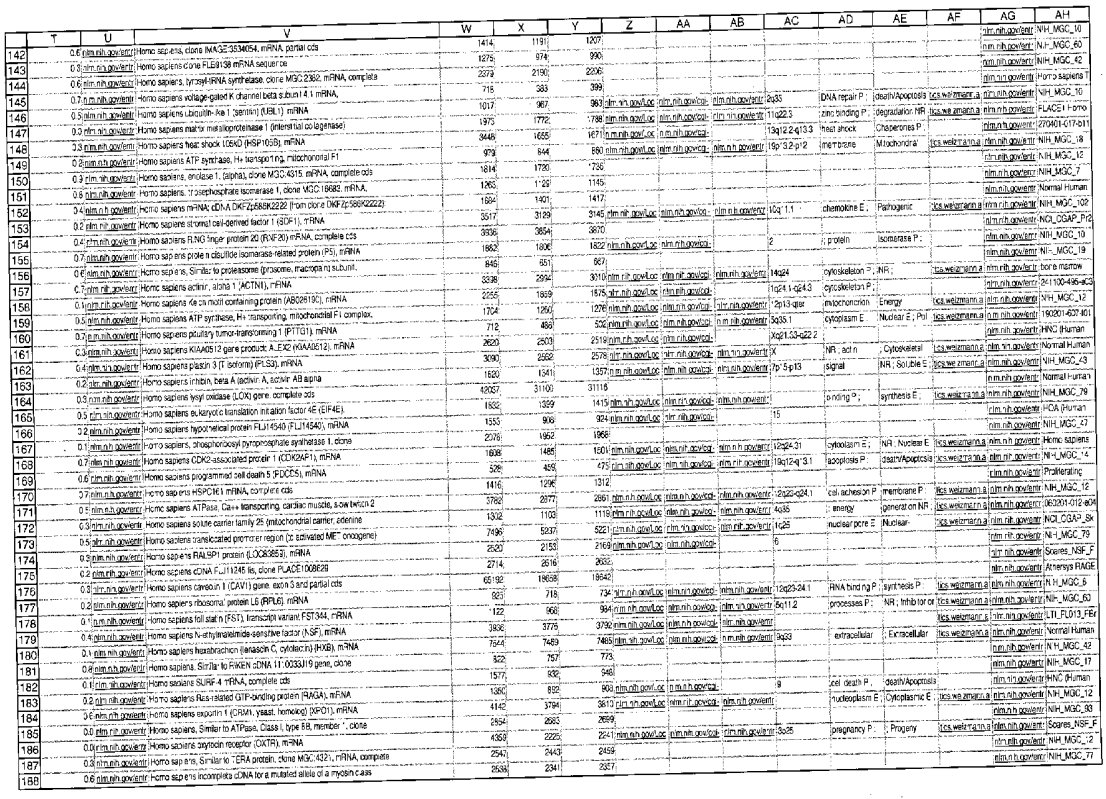
Appendix I

	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
48	0.6	nlm.nih.gov/entr	Homo sapiens chloride intracellular channel 1 (CLIC1), mRNA	1225	920	936	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	6p22.1-p21.2	membrane E ;	Unspecified	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_8
49	0.0	nlm.nih.gov/entr	Homo sapiens Kruppel-like factor 2 (lung) (KLF2), mRNA	1647	1443	1459	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	p13.11			tics.weizmann.a	nlm.nih.gov/entr	90101-006-e10
50	0.0	nlm.nih.gov/entr	Homo sapiens, insulin-like growth factor binding protein 3, clone	2497	2136	2152								nlm.nih.gov/entr	NIH_MGC_102
51	0.0	nlm.nih.gov/entr	Homo sapiens tissue inhibitor of metalloproteinase 3 (Sorsby fundus	5478	5292	5308	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	22q12.3	inhibitor NR ;	NR ; Inhibitor or	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_19
52	0.5	nlm.nih.gov/entr	Homo sapiens, HSPC163 protein, clone MGC:772, mRNA, complete cds	641	425	441								nlm.nih.gov/entr	NIH_MGC_17
53	0.1	nlm.nih.gov/entr	Homo sapiens Friend leukemia virus integration 1 (FLI1), mRNA	2957	2693	2709	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	11q24.1-q24.3	DNA binding P	protein P ; DNA-	tics.weizmann.a	nlm.nih.gov/entr	160101-677-d05
54															
55		nlm.nih.gov/entr	Homo sapiens, clone MGC:12344, mRNA, complete cds	774	113	129								nlm.nih.gov/entr	NIH_MGC_19
56	1.4	nlm.nih.gov/entr	Homo sapiens PRO1181 mRNA, complete cds	1487	1247	1263								nlm.nih.gov/entr	NIH_MGC_8
57	5.6	nlm.nih.gov/entr	Homo sapiens, ferritin, heavy polypeptide 1, clone MGC:1749, mRNA,	909	244	260								nlm.nih.gov/entr	NIH_MGC_8
58	3.2	nlm.nih.gov/entr	Homo sapiens, ribosomal protein L3, clone MGC:14821, mRNA, complete	1322	920	936								nlm.nih.gov/entr	NIH_MGC_12
59	6.5	nlm.nih.gov/entr	Homo sapiens serine (or cysteine) proteinase inhibitor, clade F	1200	860	876	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr		neurogenesis	NR ; Embryonic	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_113
60	3.4	nlm.nih.gov/entr	Homo sapiens adipose differentiation-related protein (ADFP), mRNA	1314	399	415	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	9	lipid particle E ;		tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_42
61	2.3	nlm.nih.gov/entr	Homo sapiens lipase A, lysosomal acid, cholesterol esterase (Wolman	2494	2098	2114	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	10q23.2-q23.3	lipoprotein	Lysosome/vacu	tics.weizmann.a	nlm.nih.gov/entr	HR85 islet
62	5.1	nlm.nih.gov/entr	Homo sapiens mRNA; cDNA DKFZp434D098 (from clone DKFZp434D098);	2473	2374	2390								nlm.nih.gov/entr	NCI_CGAP_GC
63	2.0	nlm.nih.gov/entr	Homo sapiens Mitochondrial Acyl-CoA Thioesterase (MT-ACT48), mRNA	1908	1456	1474	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		X	mitochondrion E ;			nlm.nih.gov/entr	NIH_MGC_15
64		nlm.nih.gov/entr	Homo sapiens decidual protein induced by progesterone (DEPP), mRNA	2120	1744	1760	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		10cen-p11.23				nlm.nih.gov/entr	LTI_NFL006_P
65	2.1	nlm.nih.gov/entr	Homo sapiens mRNA; cDNA DKFZp434E0121 (from clone DKFZp434E0121)	4502	4299	4315								nlm.nih.gov/entr	NCI_CGAP_Pa
66	1.7	nlm.nih.gov/entr	Homo sapiens fatty acid desaturase 1 (FADS1), mRNA	4213	4006	4022	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		11q12.2-q13.1	desaturase E ;	E ; Unspecified	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_17
67		nlm.nih.gov/entr	Homo sapiens catenin (cadherin-associated protein), delta 1 (CTNND1).	6232	2039	2055	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	11q11	NR ;		tics.weizmann.a	nlm.nih.gov/entr	100201-004-b10
68	7.6	nlm.nih.gov/entr	Homo sapiens polymyositis/scleroderma autoantigen 1 (75kD) (PMSCL1).	1543	1366	1352	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		4	nucleolus E ;	Nuclear	tics.weizmann.a	nlm.nih.gov/entr	LTI_NFL003_T
69	2.0	nlm.nih.gov/entr	Homo sapiens mRNA for HMGBCG protein	2746	2666	2684								nlm.nih.gov/entr	NCI_CGAP_GC
70	46.0	nlm.nih.gov/entr	Homo sapiens glucocorticoid-induced GILZ mRNA, complete cds	1946	1766	1782								nlm.nih.gov/entr	Proliferating
71	1.2	nlm.nih.gov/entr	Homo sapiens hypothetical protein FLJ23544 (FLJ23544), mRNA	2454	2226	2242	nlm.nih.gov/Loc	nlm.nih.gov/cgi-						nlm.nih.gov/entr	NIH_MGC_8
72	2.0	nlm.nih.gov/entr	Homo sapiens v-rel avian reticuloendotheliosis viral oncogene homolog	2379	2194	2200	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	11q13	apoptosis E ;	in transcription	tics.weizmann.a	nlm.nih.gov/entr	Normal Human
73	2.2	nlm.nih.gov/entr	Homo sapiens, thioredoxin peroxidase (antioxidant enzyme), clone	954	822	838								nlm.nih.gov/entr	NIH_MGC_60
74	1.9	nlm.nih.gov/entr	Homo sapiens similar to ribosomal protein S28 (H. sapiens)	365	134	150	nlm.nih.gov/Loc			17				nlm.nih.gov/entr	NIH_MGC_8
75	2.7	nlm.nih.gov/entr	Homo sapiens makorin, ring finger protein, 1 (MKRN1), mRNA	3028	2921	2937	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		7q34			tics.weizmann.a	nlm.nih.gov/entr	NCI_CGAP_Brn
76	1.9	nlm.nih.gov/entr	Homo sapiens laminin, gamma 1 (formerly LAMB2) (LAMC1), mRNA	7215	6480	6496	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	1q31	membrane P ;	(excluding cell	tics.weizmann.a	nlm.nih.gov/entr	NCI_CGAP_Sk
77	2.4	nlm.nih.gov/entr	Homo sapiens cDNA FLJ13099 fis, clone NT2RP3002248	2319	2238	2254								nlm.nih.gov/entr	LTI_FL002_PL1
78	2.0	nlm.nih.gov/entr	Homo sapiens platelet-derived growth factor receptor, beta	5576	5397	5413	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	5q31-q32	transduction E ;	membrane NR ;	tics.weizmann.a	nlm.nih.gov/entr	LTI_NFL003_N
79	2.1	nlm.nih.gov/entr	Homo sapiens, clone MGC:1426, mRNA, complete cds	1226	980	996								nlm.nih.gov/entr	NIH_MGC_91
80		nlm.nih.gov/entr	Homo sapiens minichromosome maintenance deficient (S. cerevisiae)	2535	854	838	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	21q22.3	DNA binding P	P ; Hydrolase P	tics.weizmann.a	nlm.nih.gov/entr	160301-001-c08
81	3.3	nlm.nih.gov/entr	Homo sapiens ubiquitin-conjugating enzyme E2L 3 (UBE2L3), mRNA	2845	2803	2819	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	22q11.21	modification E ;	Protein	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_48
82	6.8	nlm.nih.gov/entr	Human Meis1-related protein 2 (MRG2), mRNA, partial cds	1668	981	997								nlm.nih.gov/entr	NCI_CGAP_Brn
83	21.8	nlm.nih.gov/entr	Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 233360	2100	1768	1784								nlm.nih.gov/entr	NIH_MGC_93
84	1.4	nlm.nih.gov/entr	Homo sapiens ribonuclease/angiogenin inhibitor (RNH), mRNA	1921	1860	1876	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	11p15.5	P ; ribonuclease	Inhibitor or	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_10
85	38.3	nlm.nih.gov/entr	Homo sapiens, clone IMAGE:3687782, mRNA, partial cds	1759	1574	1590								nlm.nih.gov/entr	Human Iris
86		nlm.nih.gov/entr	Homo sapiens S100 calcium-binding protein A4 (calcium protein,	512	54	70	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	1q21	E ; invasive	binding protein	tics.weizmann.a	nlm.nih.gov/entr	240101-747-c08
87	2.7	nlm.nih.gov/entr	Homo sapiens G protein-coupled receptor kinase 7 (GPRK7), mRNA	1535	1269	1285	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	19p13.3	phosphorylation	Transferase P ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_93
88	10.8	nlm.nih.gov/entr	Homo sapiens, Similar to DNA replication factor, clone IMAGE:4126642,	2598	2268	2284								nlm.nih.gov/entr	NIH_MGC_106
89	1.5	nlm.nih.gov/entr	Homo sapiens amyloid beta (A4) precursor protein (protease nexin-II,	3573	3280	3296	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	21q21.3	cell death E ;	Soluble E ;	tics.weizmann.a	nlm.nih.gov/entr	HNC (Human
90	2.1	nlm.nih.gov/entr	Homo sapiens fatty acid desaturase 2 (FADS2), mRNA	3149	3080	3096	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		11q12-q13.1	fraction P ; fatty	E ; Integral	tics.weizmann.a	nlm.nih.gov/entr	NCI_CGAP_Brn
91	2.3	nlm.nih.gov/entr	Homo sapiens laminin, beta 2 (laminin S) (LAMB2), mRNA	5815	5674	5690	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	3p21	lamina F ;	transmission	tics.weizmann.a	nlm.nih.gov/entr	230101-001-g08
92	2.2	nlm.nih.gov/entr	Homo sapiens S100 calcium-binding protein A6 gene, complete cds	274	246	262								nlm.nih.gov/entr	NIH_MGC_42
93	2.2	nlm.nih.gov/entr	Homo sapiens core promoter element binding protein (COPEB), mRNA	1472	1275	1291	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr				tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_10
94	3.6	nlm.nih.gov/entr	Homo sapiens gamma-aminobutyric acid (GABA) B receptor, 1 (GABBR1),	4435	4391	4407	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr		receptor E ;	membrane P ;	tics.weizmann.a	nlm.nih.gov/entr	LTI_FL013_FBr

Appendix I



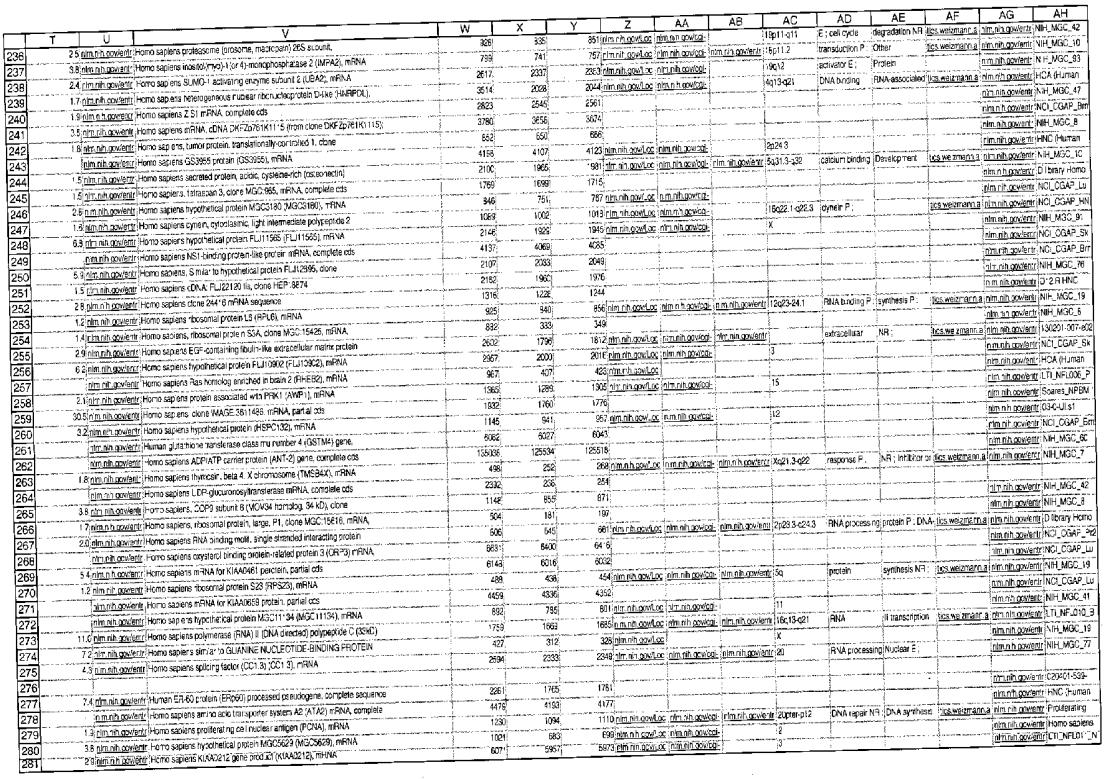
Appendix I



Appendix I

	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
189	0.1	nlm.nih.gov/entr	Homo sapiens tropomyosin 1 (alpha) (TPM1), mRNA	1106	729	745	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	15q22.1	control of heart	NR ; Inhibitor or	tics.weizmann.a	nlm.nih.gov/entr	Homo sapiens
190	0.5	nlm.nih.gov/entr	Homo sapiens, calmodulin 2 (phosphorylase kinase, delta), clone	1164	718	734								nlm.nih.gov/entr	NIH_MGC_12
191	0.2	nlm.nih.gov/entr	Homo sapiens synaptobrevin-like 1 (SYBL1), mRNA	2576	2469	2485	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr				tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_95
192	0.0	nlm.nih.gov/entr	Homo sapiens phospholipase A2-activating protein (PLAA), mRNA	2217	1400	1416	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	9p21	transduction P ;	Activator P ;	tics.weizmann.a	nlm.nih.gov/entr	LTI_NFL006_P
193	0.5	nlm.nih.gov/entr	Homo sapiens nucleosome assembly protein 1-like 1 (NAP1L1), mRNA	1560	1085	1101	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	12q15	P ; nucleosome	P ; Nuclear P ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_12
194	0.2	nlm.nih.gov/entr	Homo sapiens tumor protein p53-binding protein, 2 (TP53BP2), mRNA	4534	4469	4505	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr		cell cycle	Cytoplasmic E ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_60
195	0.4	nlm.nih.gov/entr	Homo sapiens karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	1976	1722	1738	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	17q23.1-q23.3	cytoplasm E ;	translocation	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_10
196	0.1	nlm.nih.gov/entr	Homo sapiens calpain 4, small subunit (30K) (CAPN4), mRNA	1479	1039	1055	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr		positive control	degradation NR	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_42
197	0.2	nlm.nih.gov/entr	Homo sapiens axin 2 (conductin, axil) (AXIN2), mRNA	3072	2612	2628	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	17q23-q24	suppressor P		tics.weizmann.a	nlm.nih.gov/entr	170101-017-g06
198	0.1	nlm.nih.gov/entr	Homo sapiens cDNA FLJ10746 fis, clone NT2RP3001679	2953	2585	2601								nlm.nih.gov/entr	NCI_CGAP_Lu
199	0.4	nlm.nih.gov/entr	Homo sapiens, DKFZP564M082 protein, clone MGC:12527, mRNA, complete	904	725	741								nlm.nih.gov/entr	NIH_MGC_19
200	0.5	nlm.nih.gov/entr	Homo sapiens, actin, gamma 1, clone MGC:4484, mRNA, complete cds	1947	1546	1562								nlm.nih.gov/entr	NIH_MGC_12
201	0.5	nlm.nih.gov/entr	Homo sapiens N-acetylglucosamine-phosphate mutase (AGM1), mRNA	1994	1796	1812	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	6q14.1-q15				nlm.nih.gov/entr	LTI_NFL010_B
202	0.6	nlm.nih.gov/entr	Homo sapiens malate dehydrogenase 1, NAD (soluble) (MDH1), mRNA	1267	995	1011	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	2p16	malate		tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_10
203															
204	4.5	nlm.nih.gov/entr	Homo sapiens mRNA; cDNA DKFZp564B0769 (from clone DKFZp564B0769);	2796	1129	1145								nlm.nih.gov/entr	NCI_CGAP_Sk
205	1.3	nlm.nih.gov/entr	Homo sapiens ribosomal protein L34 (RPL34), mRNA	382	281	297	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		4	protein	synthesis P ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_19
206	9.3	nlm.nih.gov/entr	Homo sapiens stimulated trans-acting factor (50 kDa) (STAF50), mRNA	2809	2629	2645	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		11	immune	pathogen		nlm.nih.gov/entr	Proliferating
207	1.9	nlm.nih.gov/entr	Homo sapiens hypothetical protein FLJ22875 (FLJ22875), mRNA	1483	718	734	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		15				nlm.nih.gov/entr	NIH_MGC_12
208	2.1	nlm.nih.gov/entr	Homo sapiens collagen, type I, alpha 2 (COL1A2), mRNA	5084	4391	4407	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	7q22.1	collagen type I	matrix	tics.weizmann.a	nlm.nih.gov/entr	060101-022-
209	1.5	nlm.nih.gov/entr	Homo sapiens GTP-binding protein SAR1 (SAR1) mRNA, complete cds	3003	2833	2849								nlm.nih.gov/entr	NCI_CGAP_Brn
210		nlm.nih.gov/entr	Homo sapiens cDNA FLJ12687 fis, clone NT2RM4002281	3344	2379	2395								nlm.nih.gov/entr	NIH_MGC_63
211	2.3	nlm.nih.gov/entr	Homo sapiens hypothetical protein FLJ20283 (DPP8), mRNA	2778	2592	2608	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		15q22				nlm.nih.gov/entr	NIH_MGC_43
212		nlm.nih.gov/entr	Homo sapiens, clone IMAGE:2958115, mRNA, partial cds	942	703	719								nlm.nih.gov/entr	NIH_MGC_10
213	1.6	nlm.nih.gov/entr	Homo sapiens clone E8271 liver membrane-bound protein mRNA, complete	1709	1507	1523								nlm.nih.gov/entr	HNC (Human
214		nlm.nih.gov/entr	Homo sapiens full length insert cDNA clone YI62C12	504	137	121								nlm.nih.gov/entr	11-0-UI.s1
215	8.0	nlm.nih.gov/entr	Homo sapiens hypothetical protein MGC16121 (MGC16121), mRNA	674	588	604	nlm.nih.gov/Loc	nlm.nih.gov/cgi-						nlm.nih.gov/entr	111200-021-g05
216	1.5	nlm.nih.gov/entr	Homo sapiens collagen, type I, alpha 2 (COL1A2), mRNA	5084	4644	4660	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	7q22.1	collagen type I	matrix	tics.weizmann.a	nlm.nih.gov/entr	bone marrow
217		nlm.nih.gov/entr	Homo sapiens catenin (cadherin-associated protein), beta 1 (88kD)	3186	2988	2972	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr		oncogenesis P ;	li transcription	tics.weizmann.a	nlm.nih.gov/entr	200301-004-e07
218	1.6	nlm.nih.gov/entr	Homo sapiens, ribosomal protein L13a, clone IMAGE:3545755, mRNA,	1128	500	516								nlm.nih.gov/entr	NIH_MGC_8
219	2.9	nlm.nih.gov/entr	Homo sapiens platelet derived growth factor C (PDGFC), mRNA	3007	2621	2637	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		4			tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_60
220	4.2	nlm.nih.gov/entr	Homo sapiens glycophorin C (Gerbich blood group) (GYPC), mRNA	1071	1025	1041	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	2q14-q21	membrane NR ;	development P ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_87
221	5.4	nlm.nih.gov/entr	Homo sapiens retinoblastoma-binding protein 2 (RBBP2), mRNA	6458	6117	6133	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	12p11	protein binding	transcription NR	tics.weizmann.a	nlm.nih.gov/entr	NCI_CGAP_Kid
222	1.3	nlm.nih.gov/entr	Homo sapiens, follistatin-like 1, clone MGC:1993, mRNA, complete cds	3714	3395	3411								nlm.nih.gov/entr	090801-203-h01
223	16.5	nlm.nih.gov/entr	Homo sapiens ATP-binding cassette, sub-family C (CFTR/MRP), member 3	5155	4825	4841	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	17q22	membrane	Integral	tics.weizmann.a	nlm.nih.gov/entr	160101-694-a02
224	1.8	nlm.nih.gov/entr	Homo sapiens, NADH dehydrogenase (ubiquinone) Fe-S protein 2 (49kD)	2070	1867	1883								nlm.nih.gov/entr	bone marrow
225	2.3	nlm.nih.gov/entr	Homo sapiens cDNA: FLJ21175 fis, clone CAS11071	1666	1506	1522								nlm.nih.gov/entr	LTI_NFL006_P
226	1.8	nlm.nih.gov/entr	Homo sapiens neural precursor cell expressed, developmentally	2249	2070	2086	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	2q37			tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_10
227	1.5	nlm.nih.gov/entr	Homo sapiens, Similar to proteasome (prosome, macropain) 26S	1116	931	947								nlm.nih.gov/entr	NIH_MGC_10
228	2.2	nlm.nih.gov/entr	Homo sapiens procollagen C-endopeptidase enhancer (PCOLCE), mRNA	1508	1371	1387	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	7q22	NR ;	matrix	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_19
229	1.3	nlm.nih.gov/entr	Homo sapiens ribosomal protein L26 (RPL26), mRNA	436	358	374	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	17p	protein	synthesis NR ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_12
230	3.7	nlm.nih.gov/entr	Homo sapiens similar to KIAA0336 gene product (H. sapiens)	6608	2607	2623	nlm.nih.gov/Loc			2				nlm.nih.gov/entr	NIH_MGC_77
231	3.2	nlm.nih.gov/entr	Homo sapiens target of myb1 (chicken) homolog (TOM1), mRNA	2310	2247	2263	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	22q13.1	lysosome P ;		tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_49
232	1.4	nlm.nih.gov/entr	Homo sapiens, ribosomal protein S17, clone MGC:11144, mRNA, complete	1506	1311	1327								nlm.nih.gov/entr	NIH_MGC_8
233	1.2	nlm.nih.gov/entr	Homo sapiens ribosomal protein L17 (RPL17), mRNA	770	582	598	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	nlm.nih.gov/entr	18q21	RNA binding	synthesis NR ;	tics.weizmann.a	nlm.nih.gov/entr	NIH_MGC_8
234	2.9	nlm.nih.gov/entr	Homo sapiens mRNA for KIAA0714 protein, partial cds	5655	1524	1540								nlm.nih.gov/entr	NCI_CGAP_Lu
235	1.6	nlm.nih.gov/entr	Homo sapiens melanoma-associated antigen MG50 mRNA, partial cds	6847	6665	6681								nlm.nih.gov/entr	NCI_CGAP_GC

Appendix I



Appendix I

	AI	AJ	AK	AL
1	EST_LENGTH	T_MATCH_STA	ST_MATCH_END	
2	905	792	808	
3	638	535	551	
4	610	541	557	
5	810	588	604	
6	532	247	263	
7	612	80	64	
8	642	504	488	
9	697	242	258	
10	748	556	572	
11	600	230	214	
12	822	215	231	
13	344	302	318	
14	544	446	462	
15	505	271	287	
16	644	380	396	
17	884	54	38	
18	693	435	451	
19	1112	202	218	
20	633	216	232	
21	343	219	235	
22	506	448	464	
23	380	337	353	
24	755	31	15	
25	629	387	403	
26	509	473	489	
27	744	332	348	
28	365	136	152	
29	752	448	464	
30	747	448	464	
31	449	342	358	
32	575	118	102	
33	585	505	521	
34	381	63	47	
35	418	81	65	
36	516	14	30	
37	815	389	405	
38	918	463	479	
39	766	414	430	
40	773	379	395	
41	577	211	195	
42	367	107	91	
43	539	80	64	
44	744	681	697	
45	510	439	455	
46	501	345	361	
47	312	151	135	

Appendix I

	AI	AJ	AK	AL
48	687	669	685	
49	228	15	31	
50	941	318	302	
51	718	611	627	
52	602	391	407	
53	281	154	170	
54				
55	752	59	75	
56	848	444	460	
57	885	259	275	
58	684	598	614	
59	795	292	276	
60	632	409	425	
61	432	326	342	
62	436	90	74	
63	1070	152	168	
64	1036	205	189	
65	280	192	176	
66	1002	241	257	
67	459	429	445	
68	901	145	161	
69	506	64	48	
70	357	205	189	
71	678	479	495	
72	589	170	154	
73	669	445	461	
74	399	139	155	
75	329	117	101	
76	891	562	578	
77	834	731	747	
78	819	85	69	
79	621	376	392	
80	442	307	323	
81	580	516	532	
82	501	388	404	
83	766	694	710	
84	560	483	499	
85	430	102	86	
86	449	33	49	
87	139	35	51	
88	725	511	527	
89	461	158	174	
90	426	338	354	
91	100	65	49	
92	476	204	220	
93	480	300	316	
94	942	24	8	

Appendix I

	AI	AJ	AK	AL
95	1665	33	49	
96	498	350	366	
97	622	535	551	
98	200	158	174	
99	300	249	265	
100	420	419	403	
101	586	319	303	
102	779	392	408	
103	432	44	28	
104	848	837	821	
105	560	473	489	
106	779	701	717	
107	277	27	11	
108	605	567	583	
109	669	423	439	
110	196	65	81	
111	454	154	170	
112	321	32	16	
113	689	502	486	
114	643	587	603	
115	530	481	497	
116	841	228	212	
117	452	49	33	
118	647	583	599	
119	623	556	572	
120				
121	1016	320	336	
122	734	497	513	
123	650	278	294	
124	704	128	112	
125	632	572	588	
126	988	220	236	
127	814	435	451	
128	935	464	480	
129	996	673	689	
130	608	272	288	
131	654	574	590	
132	739	635	651	
133	685	105	89	
134	338	147	131	
135	922	601	617	
136	885	582	598	
137	685	88	72	
138	357	39	23	
139	481	60	76	
140	675	148	132	
141	430	74	58	

Appendix I

	AI	AJ	AK	AL
142	783	653	669	
143	935	592	608	
144	965	867	883	
145	310	163	147	
146	735	522	538	
147	509	196	180	
148	427	261	277	
149	785	598	614	
150	805	732	748	
151	794	427	443	
152	573	260	244	
153	831	370	354	
154	221	91	75	
155	367	291	307	
156	807	614	630	
157	664	61	45	
158	417	219	203	
159	702	631	647	
160	485	259	275	
161	628	307	323	
162	660	444	428	
163	982	96	80	
164	658	390	374	
165	435	337	353	
166	413	311	327	
167	994	81	65	
168	502	390	406	
169	972	488	504	
170	518	117	101	
171	829	564	548	
172	315	132	148	
173	729	540	556	
174	762	150	166	
175	507	104	88	
176	792	460	476	
177	633	575	591	
178	707	336	352	
179	900	105	89	
180	324	72	56	
181	887	826	842	
182	589	262	278	
183	669	473	489	
184	806	676	692	
185	956	252	268	
186	452	221	205	
187	286	159	175	
188	741	479	495	

Appendix I

	AI	AJ	AK	AL
189	1368	74	90	
190	652	581	597	
191	901	98	82	
192	798	735	719	
193	696	468	484	
194	786	124	108	
195	681	413	429	
196	782	383	399	
197	395	311	327	
198	479	371	387	
199	427	254	270	
200	856	436	452	
201	865	143	127	
202	828	642	658	
203				
204	753	589	605	
205	437	283	299	
206	554	121	105	
207	730	652	668	
208	259	186	202	
209	662	576	592	
210	729	495	511	
211	627	533	549	
212	460	273	289	
213	671	504	520	
214	457	140	124	
215	496	429	445	
216	670	453	437	
217	385	70	86	
218	735	501	517	
219	828	388	404	
220	536	476	492	
221	412	22	6	
222	459	174	190	
223	435	157	141	
224	530	219	203	
225	883	45	29	
226	326	132	148	
227	635	365	381	
228	1009	649	665	
229	517	370	386	
230	780	343	359	
231	988	528	544	
232	746	304	320	
233	626	467	483	
234	329	123	107	
235	486	166	150	

Appendix I

	AI	AJ	AK	AL
236	810	151	167	
237	524	448	464	
238	958	344	360	
239	388	207	223	
240	477	234	218	
241	213	81	97	
242	626	414	430	
243	652	625	641	
244	1031	375	391	
245	89	35	51	
246	470	73	57	
247	425	77	61	
248	675	627	643	
249	407	60	44	
250	406	30	14	
251	703	556	572	
252	589	543	559	
253	944	537	553	
254	853	313	329	
255	411	136	120	
256	857	34	50	
257	592	516	532	
258	1008	36	20	
259	558	148	132	
260	573	208	192	
261	634	24	8	
262	709	590	606	
263	646	298	314	
264				
265	632	409	425	
266	886	234	250	
267	551	60	76	
268	532	209	193	
269	494	477	493	
270	745	447	463	
271	284	127	111	
272	740	719	703	
273	758	650	666	
274	567	333	349	
275	759	523	539	
276				
277	460	285	269	
278	626	379	395	
279	408	163	147	
280	309	309	293	
281	758	633	649	

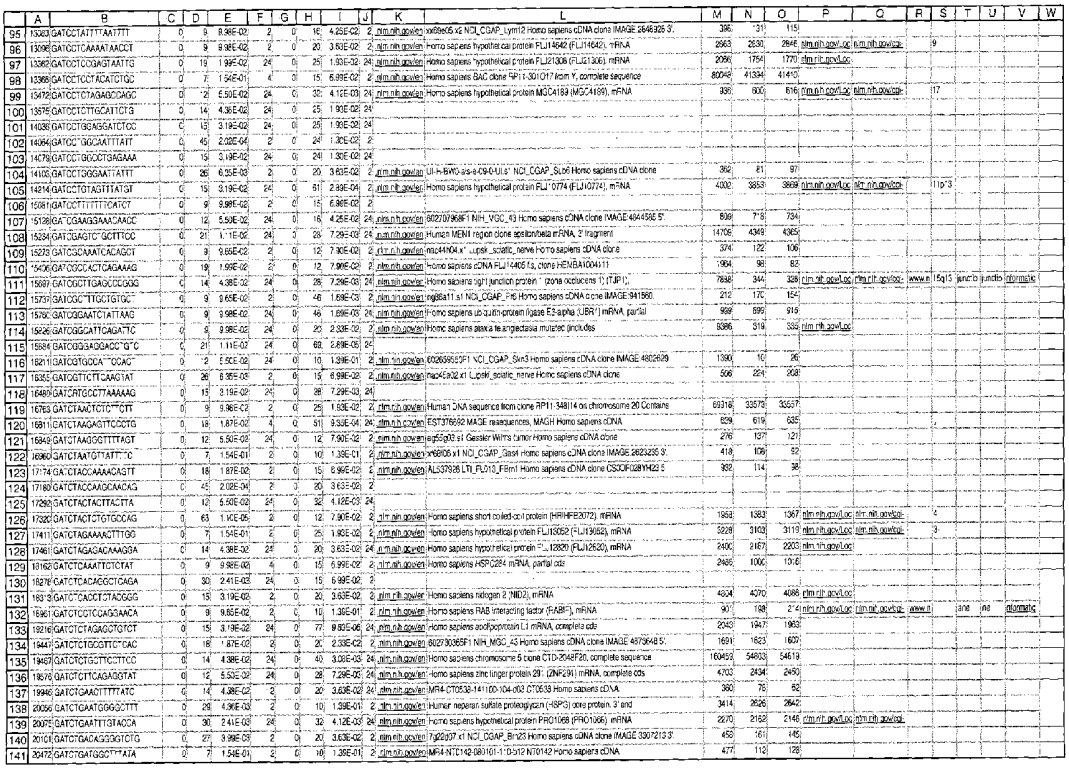
Appendix II

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	ID	sentence	Ncho	Ycho	cho_p	p_direct	Basl	Diff	Pa_probabi	dir	LOCUS	DESCRIPTION	LENGTH	TCH_STA	ATCH_EN	LOCUSLINK	UNIGENE	OMIM	OGEN	ONTO	ONTO	ENECARD	
2	149	GATCAAAACTATTAAAA	0	299	0.00E+00	4	0	97	5.08E-06	24	nlm.nih.gov/en	Homo sapiens, exostoses (multiple) 1, clone MGC:2129 IMAGE:3502232,	2769	2715	2731								
3	361	GATCAAACTACTCCGAG	0	39	5.40E-04	2	0	30	1.04E-02	2	nlm.nih.gov/en	602651454T1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4761732 3'.	813	190	174								
4	664	GATCAAATGGAACATTT	0	48	1.24E-04	24	0	40	1.34E-03	24	nlm.nih.gov/en	Homo sapiens ADP/ATP carrier protein (ANT-2) gene, complete cds	135036	125534	125518								
5	724	GATCAAATGTGTGGGCA	0	21	1.11E-02	24	0	32	4.12E-03	24													
6	946	GATCAACAGGCTGATTA	0	24	6.64E-03	24	0	26	7.29E-03	24	nlm.nih.gov/en	601065264F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3451646 5'.	791	681	697								
7	990	GATCAACATCATTAATT	0	39	5.40E-04	24	0	24	1.30E-02	24	nlm.nih.gov/en	Human DNA sequence from clone RP4-744A17 on chromosome 20 Contains	113639	93214	93198								
8	1101	GATCAACGGCAGGAGAG	0	7	1.54E-01	4	0	16	4.25E-02	2	nlm.nih.gov/en	Homo sapiens c-Cbl-interacting protein (CIN85), mRNA	3348	411	427	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		X				
9	1423	GATCAAGGAGACCCGGA	0	18	1.87E-02	24	0	24	1.30E-02	24	nlm.nih.gov/en	Homo sapiens hypothetical protein FLJ20917 (FLJ20917), mRNA	961	746	762	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		q22			nformatic	
10	1492	GATCAAGTATCTGCTCA	0	72	2.59E-06	24	0	16	4.25E-02	24	nlm.nih.gov/en	602646560F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4768563 5'.	639	27	43								
11	1494	GATCAAGTATGCTAATT	0	21	1.36E-02	24	0	15	6.99E-02	2	nlm.nih.gov/en	Homo sapiens ADP-ribosylation factor binding protein GGA2s (GGA2)	3376	3301	3317								
12	1530	GATCAAGTTCAATATTC	0	24	6.64E-03	2	0	44	7.69E-04	24	nlm.nih.gov/en	602720975F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4837650 5'.	813	342	326								
13	1553	GATCAAGTTTAGTTATT	0	18	1.87E-02	2	0	20	2.33E-02	2	nlm.nih.gov/en	at12c03.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone	514	110	94								
14	1770	GATCAATGAATTGCACA	0	60	3.38E-05	24	0	35	5.63E-03	24	nlm.nih.gov/en	Homo sapiens UDP-glucuronosyltransferase mRNA, complete cds	2392	238	254								
15	1848	GATCAATGTTTAGTAAA	0	14	4.38E-02	24	0	20	3.63E-02	24	nlm.nih.gov/en	Homo sapiens cDNA FLJ14059 fis, clone HEMBB1000573	1601	1369	1385								
16	1960	GATCACAAAGACTGGCA	0	33	1.46E-03	24	0	20	2.33E-02	24	nlm.nih.gov/en	Homo sapiens clone 3932P1, complete sequence	85060	25652	25668								
17	2153	GATCACACCACTGCAGT	0	12	5.50E-02	24	0	16	4.25E-02	24	nlm.nih.gov/en	Homo sapiens hypothetical protein FLJ14950 (FLJ14950), mRNA	3403	2435	2419	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		2				
18	2513	GATCACATGTCAGGACA	0	9	9.65E-02	2	0	24	1.30E-02	24	nlm.nih.gov/en	Homo sapiens mRNA; cDNA DKFZp727M031 (from clone DKFZp727M031)	2637	722	736								
19	2588	GATCACCAAAATTAATT	0	9	9.65E-02	2	0	12	7.90E-02	2	nlm.nih.gov/en	601472414F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE 3875475 5'.	809	119	135								
20	2636	GATCACCACTCTAACCC	0	12	5.50E-02	24	0	24	1.30E-02	24													
21	2666	GATCACCAGTGGGCACG	0	12	6.56E-02	2	0	20	2.33E-02	2													
22	2793	GATCACCGTGACAGCCA	0	36	8.87E-04	24	0	36	2.34E-03	24													
23	2801	GATCACCGTTGCTTGGT	0	13	1.67E-02	2	0	12	7.90E-02	2	nlm.nih.gov/en	Homo sapiens mRNA for KIAA1582 protein, partial cds	4442	1640	1656								
24	3013	GATCACGCCACTGAACT	0	42	3.30E-04	24	0	26	7.29E-03	24	nlm.nih.gov/en	Homo sapiens FKSG83 (FKSG83), mRNA	1164	1115	1131	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		6				
25	3238	GATCACTCTGTTCATAA	0	21	1.11E-02	2	0	12	7.90E-02	2	nlm.nih.gov/en	Homo sapiens protein phosphatase 1, regulatory (inhibitor) subunit 3B	3807	3741	3757	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		8			nformatic	
26	3613	GATCACTTGTGTTGAGG	0	9	9.98E-02	2	0	12	7.90E-02	2	nlm.nih.gov/en	nal59d11.x1 NCI_CGAP_Brn65 Homo sapiens cDNA clone IMAGE:4148397	258	26	10								
27	3664	GATCACTTTGTGGCAGG	0	9	9.98E-02	2	0	12	7.90E-02	2	nlm.nih.gov/en	602464567F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592272 5'.	836	366	382								
28	4573	GATCAGCCCAGGCAACA	0	9	9.65E-02	2	0	24	1.30E-02	2	nlm.nih.gov/en	602812389F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4944286	737	268	284								
29	4696	GATCAGCTAATTCAGAG	0	14	4.38E-02	24	0	35	5.63E-03	24	nlm.nih.gov/en	Homo sapiens mitochondrial translational release factor 1 (MTRF1),	1552	1324	1340	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	www.n	1-14.3	ondrio	n	nformatic	
30	4810	GATCAGGAACAGACATG	0	6	1.75E-01	2	0	32	4.12E-03	2	nlm.nih.gov/en	PM3-HT0907-040900-002-h09 HT0907 Homo sapiens cDNA.	323	60	64								
31	4907	GATCAGGAGGCATGACT	0	15	3.19E-02	24	0	71	9.01E-05	24													
32	5157	GATCAGTAAGGATTGCA	0	9	9.98E-02	24	0	15	6.99E-02	2													
33	5174	GATCAGTACAGACATTT	0	9	9.98E-02	24	0	20	3.63E-02	2	nlm.nih.gov/en	Homo sapiens high-mobility group (nonhistone chromosomal) protein 14	1232	1213	1197	nlm.nih.gov/Loc							
34	5213	GATCAGTATTTTTTATC	0	15	3.19E-02	2	0	25	1.93E-02	4	nlm.nih.gov/en	Human TBP-associated factor (hTAFII130) mRNA, partial cds	4233	4003	4019								
35	5706	GATCATAATAATTCCAA	0	36	8.87E-04	2	0	32	4.12E-03	2	nlm.nih.gov/en	Homo sapiens cDNA FLJ12667 fis, clone NT2RM4002251	3344	2379	2395								
36	6512	GATCATGAAATTCTAAT	0	9	9.98E-02	24	0	25	1.93E-02	24	nlm.nih.gov/en	tq77g01.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2214864 3	381	216	200								
37	6865	GATCATGGGCTCTGGAA	0	30	2.41E-03	2	0	24	1.30E-02	24	nlm.nih.gov/en	7o30c07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3704461 3'	366	252	236								
38	6986	GATCATGTCCTAATTCA	0	55	7.93E-05	2	0	28	7.29E-03	2	nlm.nih.gov/en	CM3-SN0032-060400-140-f08 SN0032 Homo sapiens cDNA.	328	111	95								
39	7082	GATCATGTTACCATGCA	0	12	5.50E-02	24	0	15	4.25E-02	24													
40	7142	GATCATGTTTCATTTCT	0	14	4.38E-02	2	0	16	4.25E-02	2	nlm.nih.gov/en	Human DNA sequence from clone RP11-129K24 on chromosome 6, complete	94153	77811	77827								
41	7298	GATCATTATTATTTTTT	0	14	4.38E-02	24	0	25	1.93E-02	2	nlm.nih.gov/en	EST390040 MAGE resequences, MAGO Homo sapiens cDNA.	616	102	86								
42	7321	GATCATTCAAACATTTG	0	24	9.26E-03	4	0	20	3.63E-02	2	nlm.nih.gov/en	Homo sapiens mRNA; cDNA DKFZp566J2146 (from clone DKFZp566J2146)	1385	681	665								
43	7602	GATCATTTACAAATTTG	0	9	9.98E-02	24	0	10	1.39E-01	2	nlm.nih.gov/en	qu51d01.x1 NCI_CGAP_Lym6 Homo sapiens cDNA clone IMAGE:1968289 3'.	365	241	225								
44	7792	GATCATTTGTGATTCCT	0	12	6.58E-02	2	0	20	3.63E-02	2	nlm.nih.gov/en	Homo sapiens DC42 mRNA, complete cds	1632	404	388								
45	8024	GATCCAAACCAAGAGGA	0	12	6.58E-02	2	0	20	3.63E-02	2	nlm.nih.gov/en	602868218F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:5014612 5'	974	593	609								
46	8038	GATCCAAACTAATATTG	0	39	5.40E-04	24	0	56	1.48E-04	24	nlm.nih.gov/en	Homo sapiens mRNA for KIAA0659 protein, partial cds	4459	4336	4352								
47	8070	GATCCAAAGAGCTCTTA	0	18	1.87E-02	2	0	10	1.39E-01	2	nlm.nih.gov/en	Homo sapiens CRIM1 protein gene, partial cds; and FEZ2 gene, partial	236303	183469	183453								

Appendix II

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
48	8067	GATCCAAAGGAGTTGTT	0	46	1.24E-04	24	0	146	1.17E-09	24													
49	8111	GATCCAAATAAACACAC	0	19	1.99E-02	2	0	15	6.99E-02	2	nlm.nih.gov/en	ab09c09.s1 Stratagene lung (#937210) Homo sapiens cDNA clone	297	219	203								
50	8192	GATCCAACAATAAAATA	0	27	3.99E-03	24	0	32	4.12E-03	24	nlm.nih.gov/en	Homo sapiens hypothetical protein MGC11134 (MGC11134), mRNA	892	785	801	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		.11				
51	8199	GATCCAACAGCGACTAC	0	9	9.65E-02	2	0	16	4.25E-02	2	nlm.nih.gov/en	Homo sapiens similar to heat shock 10kD protein 1 (chaperonin 10)	532	193	177	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	www.n	1-	peroxi	stress	nformatic	
52	8475	GATCCAATACTATTTAG	0	21	1.36E-02	24	0	12	7.90E-02	2	nlm.nih.gov/en	Homo sapiens mRNA; cDNA DKFZp566J2446 (from clone DKFZp566J2446)	1134	714	730								
53	8819	GATCCACCCACCTCGGG	0	15	3.19E-02	2	0	20	2.33E-02	24	nlm.nih.gov/en	CM3-MT0295-170101-628-b12 MT0295 Homo sapiens cDNA.	312	180	196								
54	8841	GATCCACCCAGTAAATC	0	6	1.75E-01	2	0	20	2.33E-02	2	nlm.nih.gov/en	MR4-ET0138-230401-006-b02_1 ET0138 Homo sapiens cDNA.	104	64	80								
55	8845	GATCCACCCATCTCGCC	0	14	4.38E-02	24	0	20	3.63E-02	24	nlm.nih.gov/en	PM0-NT0319-030401-002-b03 NT0319 Homo sapiens cDNA.	250	34	18								
56	9177	GATCCACTTCCTAAAAG	0	9	9.98E-02	2	0	25	1.93E-02	2	nlm.nih.gov/en	yz08e12.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA	648	119	103								
57	9225	GATCCAGAAATAAGTTT	0	21	1.11E-02	2	0	16	4.25E-02	2	nlm.nih.gov/en	602696075F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4828051 5'.	650	516	532								
58	9341	GATCCAGAGAGCTGACA	0	12	6.58E-02	2	0	10	1.39E-01	2	nlm.nih.gov/en	Homo sapiens myosin phosphatase target subunit 2 (MYPT2) gene, exons 11	150000	134214	134230								
59	9417	GATCCAGCAATACCCCA	0	12	6.58E-02	24	0	25	1.93E-02	2	nlm.nih.gov/en	Homo sapiens chromosome 1 open reading frame 12 (C1orf12), mRNA	5163	4445	4461	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		1q42.1	P ;	asmic	nformatic	
60	9418	GATCCAGCAATCCCACC	0	21	1.11E-02	4	0	16	4.25E-02	24	nlm.nih.gov/en	RC6-HT0592-270300-011-A09 HT0592 Homo sapiens cDNA.	283	159	143								
61	9450	GATCCAGCAGTCCCACT	0	21	1.36E-02	2	0	15	6.99E-02	2	nlm.nih.gov/en	CM3-NT0267-110101-588-b11 NT0267 Homo sapiens cDNA.	196	153	137								
62	9451	GATCCAGCAGTCTCACT	0	14	4.38E-02	24	0	20	3.63E-02	24	nlm.nih.gov/en	Homo sapiens T-cell receptor alpha delta locus from bases 1 to 250529	250529	231401	231417								
63	9688	GATCCAGGGGAAGCAAA	0	6	1.75E-01	4	0	15	6.99E-02	2	nlm.nih.gov/en	Homo sapiens chromosome 7 clone RP11-180C16, complete sequence	162634	138848	138832								
64	9711	GATCCAGGTCTCCAATT	0	78	9.94E-07	2	0	30	1.04E-02	2	nlm.nih.gov/en	QV2-NN2003-230401-632-h10 NN2003 Homo sapiens cDNA.	547	124	106								
65	9841	GATCCAGTGCTTAGTTC	0	12	5.50E-02	2	0	24	1.30E-02	2	nlm.nih.gov/en	602299156F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4393597 5'	410	93	109								
66	9849	GATCCAGTGTCCACGGA	0	45	2.02E-04	24	0	30	1.04E-02	2	nlm.nih.gov/en	RC0-NT0013-040400-021-h02 NT0013 Homo sapiens cDNA	385	102	118								
67	9881	GATCCAGTTGGTGCAGC	0	18	1.87E-02	24	0	44	7.69E-04	24	nlm.nih.gov/en	Homo sapiens, clone MGC:17708 IMAGE:3866595, mRNA, complete cds.	839	16	32								
68	9896	GATCCAGTTTTAGTTTT	0	16	1.87E-02	2	0	20	3.63E-02	2	nlm.nih.gov/en	yx89h04.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone	526	259	243								
69	10047	GATCCATCAGTGTGCTG	0	16	1.87E-02	2	0	15	6.99E-02	2	nlm.nih.gov/en	CM1-MT0244-030101-695-h07 MT0244 Homo sapiens cDNA.	366	224	208								
70	10128	GATCCATCTGCCTTTCT	0	12	6.58E-02	24	0	16	4.25E-02	24	nlm.nih.gov/en	oq61d06.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1590827 3'	424	205	189								
71	10148	GATCCATCTTCAAATCA	0	18	1.87E-02	2	0	20	2.33E-02	2	nlm.nih.gov/en	60251446F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4646170 5'	754	415	431								
72	10396	GATCCATTTGATTGTTA	0	15	3.19E-02	2	0	20	3.63E-02	2	nlm.nih.gov/en	DKFZp564F047_s1 564 (synonym: hfbr2) Homo sapiens cDNA clone	510	72	56								
73	10546	GATCCCAACTGCTCCTT	0	12	5.50E-02	24	0	24	1.30E-02	24													
74	10557	GATCCCAAGAGACTTAC	0	7	1.54E-01	2	0	20	2.33E-02	2	nlm.nih.gov/en	Homo sapiens RAB10, member RAS oncogene family (RAB10), mRNA	3127	2970	2954	nlm.nih.gov/Loc							
75	10657	GATCCCACCAATCCTCG	0	17	2.95E-02	2	0	52	2.55E-04	2	nlm.nih.gov/en	MR3-HN0052-080101-005-c05 HN0052 Homo sapiens cDNA.	426	270	286								
76	10678	GATCCCACCATTGCACT	0	14	4.38E-02	2	0	12	7.90E-02	2	nlm.nih.gov/en	Homo sapiens similar to leucine-rich repeat-containing 2 (H. sapiens)	4841	4235	4251	nlm.nih.gov/Loc							
77	10696	GATCCCACCTTATGGTG	0	9	9.65E-02	2	0	24	1.30E-02	2													
78	10739	GATCCCAGAAAAACAAC	0	7	1.54E-01	2	0	10	1.39E-01	2	nlm.nih.gov/en	HNC48-1-ES.R HNC (Human Normal Cartilage) Homo sapiens cDNA.	665	502	518								
79	10798	GATCCCAGAGGCTGATT	0	14	4.38E-02	2	0	12	7.90E-02	2	nlm.nih.gov/en	np77e10.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1132362.	457	266	234								
80	10982	GATCCCAGTGTTGCCTT	0	9	9.65E-02	2	0	20	2.33E-02	24	nlm.nih.gov/en	Homo sapiens hypothetical protein MGC5391 (MGC5391), mRNA	1161	737	753	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		q14 3				
81	1'017	GATCCCATCACCCAGGT	0	7	1.54E-01	2	0	20	2.33E-02	2	nlm.nih.gov/en	Homo sapiens T-cell receptor alpha delta locus from bases 250472 to	251199	135062	135078								
82	11325	GATCCCCCAACTCTGKA	0	33	1.46E-03	24	0	20	2.33E-02	24													
83	11584	GATCCCCTTTGGTCTTA	0	18	1.87E-02	24	0	20	3.63E-02	2													
84	12026	GATCCCTGGTATTGATT	0	12	5.50E-02	2	0	10	1.39E-01	2	nlm.nih.gov/en	Homo sapiens hypothetical protein FLJ22457 (FLJ22457), mRNA	1888	1623	1639	nlm.nih.gov/Loc							
85	12212	GATCCCTTGCCTGAATT	0	19	1.99E-02	24	0	87	1.60E-05	24													
86	12345	GATCCGAGTCGTCGGGA	0	81	6.16E-07	24	0	80	8.56E-05	24													
87	12366	GATCCGATTTCTTTACC	0	9	9.98E-02	2	0	10	1.39E-01	2	nlm.nih.gov/en	602598226F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4707006 5'.	807	135	151								
88	12377	GATCCGCACCCCCCCCC	0	17	2.95E-02	24	0	25	1.93E-02	24													
89	12385	GATCCGCAGGCAGACAC	0	14	4.38E-02	24	0	56	5.19E-04	24													
90	12386	GATCCGCAGGCAGACAG	0	17	2.95E-02	24	0	71	3.01E-05	24													
91	12645	GATCCGTATGTGGTTAA	0	17	2.95E-02	2	0	20	3.63E-02	24	nlm.nih.gov/en	QV0-EN0059-101100-494-b04 EN0059 Homo sapiens cDNA.	565	166	150								
92	12658	GATCCGTCCCTAACTAC	0	46	3.35E-04	24	0	51	9.35E-04	24													
93	12762	GATCCTAACAAAACCTA	0	7	1.54E-01	2	0	10	1.39E-01	2	nlm.nih.gov/en	AV733718 cdA Homo sapiens cDNA clone cdAAZH02 5'.	431	218	234								
94	12647	GATCCTACAAGTTGATG	0	17	2.95E-02	2	0	25	1.93E-02	2	nlm.nih.gov/en	Homo sapiens BAC clone RP11-490I2 from 2, complete sequence	134'74	32153	32137								

Appendix II



Appendix II

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
142	20602	GATCTGCACAGCCAGAC	0	14	4.38E-02	24	0	25	1.93E-02	24													
143	20721	GATCTGCCCACCGCAGC	0	6	1.75E-01	2	0	20	3.63E-02	2	.nlm.nih.gov/en	IL3-NT0280-220101-409-H08 NT0280 Homo sapiens cDNA	545	481	465								
144	21127	GATCTGCTTGCCTTGGT	0	9	9.65E-02	2	0	10	1.39E-01	2	.nlm.nih.gov/en	CM4-HB0026-151200-635-g12 HB0026 Homo sapiens cDNA	358	84	100								
145	21252	GATCTGGACAAATGGCA	0	6	1.75E-01	2	0	15	6.99E-02	2	.nlm.nih.gov/en	Homo sapiens cDNA FLJ11738 fis, clone HEMBA1005474	1651	1265	1281								
146	21361	GATCTGGATGGAGTGGT	0	45	2.02E-04	2	0	10	1.39E-01	2	.nlm.nih.gov/en	Homo sapiens hypothetical protein FLJ21313 (FLJ21313), mRNA	2912	955	971	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		5				
147	21578	GATCTGGGCTACCTCGT	0	7	1.54E-01	2	0	40	1.34E-03	2	.nlm.nih.gov/en	Homo sapiens gene for thymidylate synthase exons 1, 2, 3, 4, 5, 6, 7,	16596	13195	13179								
148	22403	GATCTGTGTTTGCCCTG	0	12	5.50E-02	2	0	30	1.04E-02	24	.nlm.nih.gov/en	zh27b02.s1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone	275	79	63								
149	22426	GATCTGTTACCGCTAAA	0	9	9.98E-02	2	0	20	3.63E-02	2													
150	22449	GATCTGTTATGAAACGA	0	12	5.50E-02	2	0	40	1.34E-03	2	.nlm.nih.gov/en	7q92a01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3705505 3'.	496	407	423								
151	22577	GATCTGTTTATTCTTTT	0	34	2.06E-03	4	0	10	1.39E-01	2	.nlm.nih.gov/en	Homo sapiens hypothetical protein FLJ23518 (FLJ23518), mRNA	1134	635	651	nlm.nih.gov/Loc							
152	22608	GATCTGTTTCTTTTTTT	0	9	9.65E-02	2	0	25	1.93E-02	2													
153	22785	GATCTTAATGATGTTTT	0	19	1.99E-02	24	0	20	3.63E-02	24	.nlm.nih.gov/en	Homo sapiens CRIM1 protein gene, partial cds; and FEZ2 gene, part.a	236303	21368	21364								
154	22916	GATCTTAGTTATCTGTA	0	12	6.58E-02	2	0	12	7.90E-02	2	.nlm.nih.gov/en	nc48c06.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1011370	387	323	339								
155	23000	GATCTTATTTCAAAGGA	0	12	6.58E-02	2	0	10	1.39E-01	2	.nlm.nih.gov/er	Homo sapiens R3P1-like protein (BCAA), transcript variant 2, mRNA	5804	2444	2460	nlm.nih.gov/Loc	nlm.nih.gov/cgi-		q43				
156	23028	GATCTTCAAAAATACTA	0	34	2.08E-03	2	0	20	2.33E-02	2	.nlm.nih.gov/en	602691237F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4823529 5'.	737	251	235								
157	23334	GATCTTCCTTTTCTCAG	0	15	3.19E-02	2	0	25	1.93E-02	2	.nlm.nih.gov/en	nc17c07.r1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:1008396.	242	128	112								
158	23481	GATCTTCTGGGTTGGGA	0	21	1.36E-02	2	0	15	6.99E-02	2	.nlm.nih.gov/en	xw74e11.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2833772 3'	408	242	226								
159	23586	GATCTTCTTTATTTTTT	0	39	5.40E-04	24	0	40	1.34E-03	24	.nlm.nih.gov/en	602675830F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4798231 5'.	761	131	115								
160	23624	GATCTTGACTTCTTGAT	0	24	6.64E-03	24	0	16	4.25E-02	24													
161	23690	GATCTTGCAGTTTTCCT	0	6	1.75E-01	2	0	18	4.25E-02	2	.nlm.nih.gov/en	602503520F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4616684 5'.	835	164	180								
162	23781	GATCTTGGAGGCCAGCC	0	12	6.58E-02	24	0	10	1.39E-01	2	.nlm.nih.gov/en	Homo sapiens polymerase (RNA) mitochondrial (DNA directed) (POLRMT),	3759	3608	3624	nlm.nih.gov/Loc							
163	24125	GATCTTTCAATAGTCTG	0	14	4.38E-02	24	0	51	9.35E-04	2	.nlm.nih.gov/en	Homo sapiens clone MGC:18144 IMAGE:4154047, mRNA, complete cds	3537	3232	3248								
164	24281	GATCTTTCTTTAATAGT	0	9	9.65E-02	4	0	66	1.61E-04	2	.nlm.nih.gov/en	Homo sapiens mannose-6-phosphate receptor (cation dependent) (M6PR),	2454	2145	2161	nlm.nih.gov/Loc	nlm.nih.gov/cgi-	www.n	12p13	embra	n	informatic	
165	24444	GATCTTTGTCGGTTAGA	0	7	1.54E-01	2	0	12	7.90E-02	2	.nlm.nih.gov/en	Homo sapiens staufen (Drosophila, RNA-binding protein) (STAU), mRNA	3345	2712	2696	nlm.nih.gov/Loc							
166	24665	GATCTTTTGCTTACCCA	0	9	9.65E-02	2	0	35	5.63E-03	2	.nlm.nih.gov/en	HHEA90R Atrium cDNA library Human heart Homo sapiens cDNA clone	278	123	107								
167	24700	GATCTTTTAACATTGA	0	9	9.98E-02	2	0	10	1.39E-01	2													

What is claimed is:

1. A composition comprising at least one expression vector, which expression vector comprises a nucleic acid comprising:

(a) at least one polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a polynucleotide sequence complementary thereto;

(b) at least one polynucleotide sequence that hybridizes under stringent conditions to a polynucleotide sequence of (a);

(c) at least one polynucleotide sequence that is at least about 70% identical to a polynucleotide sequence of (a);

(d) at least one polynucleotide that encodes a polypeptide or peptide comprising a subsequence encoded by a polynucleotide sequence of (a);

(e) at least one polynucleotide sequence that hybridizes to a nucleic acid that is physically linked in the human genome to a nucleic acid comprising a polynucleotide sequence of (a), (b), (c) or (d); or,

(f) at least one polynucleotide sequence comprising at least about 10 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a sequence complementary thereto.

2. The vector of claim 1, wherein the vector comprises a promoter operably linked to the nucleic acid comprising the polynucleotide sequence of (a), (b), (c), (d), (e) or (f).

3. The vector of claim 1, wherein the nucleic acid encodes a polypeptide.

4. The vector of claim 1, wherein the nucleic acid encodes a sense or antisense RNA.

5. A composition comprising the at least one expression vector of claim 1 and an excipient.

6. The composition of claim 5, wherein the excipient is a pharmaceutically acceptable excipient.

7. A cell comprising the vector of claim 1.

8. A labeled probe comprising a nucleic acid sequence comprising:

(a) at least one polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a polynucleotide sequence complementary thereto;

(b) at least one polynucleotide sequence that hybridizes under stringent conditions to a polynucleotide sequence of (a);

(c) at least one polynucleotide sequence that is at least about 70% identical to a polynucleotide sequence of (a);

(d) at least one polynucleotide that encodes a polypeptide or peptide comprising a subsequence encoded by a polynucleotide sequence of (a);

(e) at least one polynucleotide sequence that hybridizes to a nucleic acid that is physically linked in the human genome to a nucleic acid comprising a polynucleotide sequence of (a), (b), (c) or (d); or,

(f) at least one polynucleotide sequence comprising at least about 10 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a sequence complementary thereto.

9. The labeled probe of claim 8, the subsequence comprising at least about 12 nucleotides.

10. The labeled probe of claim 8, the subsequence comprising at least about 14 nucleotides.

11. The labeled probe of claim 8, the subsequence comprising at least about 16 nucleotides.

12. The labeled probe of claim 8, the subsequence comprising at least about 17 nucleotides.

13. The labeled probe of claim 8, comprising an isotopic, fluorescent, fluorogenic, or colorimetric label.

14. The labeled probe of claim 8, comprising a DNA or RNA molecule.

15. The labeled probe of claim 8, comprising a cDNA, an amplification product, a transcript, a restriction fragment, or an oligonucleotide.

16. The labeled probe of claim 8, comprising an oligonucleotide consisting of a polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443.

17. The labeled probe of claim 8, wherein the labeled probe is a member of an array of probes comprising a plurality of nucleic acids comprising two or more polynucleotide sequences selected from (a), (b), (c), (d), (e) and/or (f).

18. An array of probes according to claim 17, wherein the plurality of nucleic acids are logically or physically arrayed.

19. A marker set for evaluating a condition or characteristic associated with elevated cholesterol or lipid and/or adipogenesis comprising a plurality of members, which members comprise nucleic acids, polypeptides and/or peptides comprising:

(a) one or more polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a polynucleotide sequence complementary thereto;

(b) one or more polynucleotide sequence that hybridizes under stringent conditions to a polynucleotide sequence of (a);

(c) one or more polynucleotide sequence that is at least about 70% identical to a polynucleotide sequence of (a);

(d) one or more polynucleotide that encodes a polypeptide or peptide comprising a subsequence encoded by a polynucleotide sequence of (a);

(e) one or more polynucleotide sequence that hybridizes to a nucleic acid that is physically linked in the human genome to a nucleic acid comprising a polynucleotide sequence of (a), (b), (c) or (d); and/or,

(f) one or more polynucleotide sequence comprising at least about 10 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a sequence complementary thereto;

(g) one or more polypeptides or peptides comprising an amino acid sequence encoded by a polynucleotide of (a), (b), (c), (d), or (e); and/or,

(h) one or more antibodies specific for a polypeptide or peptide sequence of (g).

20. The marker set of claim 19, comprising a plurality of oligonucleotides.

21. The marker set of claim 20, wherein the oligonucleotides are synthetic oligonucleotides.

22. The marker set of claim 19, comprising a plurality of amplification products or expression products.

23. The marker set of claim 19, comprising a plurality of labeled nucleic acid probes.

24. The marker set of claim 19, comprising a plurality of polypeptides or peptides.

25. The marker set of claim 19, comprising a plurality of antibodies.

26. The marker set of claim 19, comprising a plurality of members, which members include nucleic acids and polypeptides.

27. The marker set of claim 19, wherein the members of the marker set are logically or physically arrayed.

28. The marker set of claim 19, wherein the members of the marker set are physically arrayed in a solid phase or liquid phase array.

29. The marker set of claim 28, wherein the array comprises a bead array.

30. The marker set of claim 19, comprising a majority of sequences or subsequences selected from SEQ ID NO:1-SEQ ID NO:443.

31. The marker set of claim 19, comprising SEQ ID NO:1-SEQ ID NO:443.

32. The marker set of claim 19, wherein the condition or characteristic associated with elevated cholesterol or lipid and/or adipogenesis is predicted by hybridizing the nucleic acids of the marker set to a DNA or RNA sample from a cell or tissue, and detecting at least one polymorphic polynucleotide or differentially expressed expression product.

33. The marker set of claim 19, wherein the condition or characteristic is associated with high cholesterol or fat exposure.

34. The marker set of claim 19, wherein the condition or characteristic is selected from among obesity, atherosclerosis, diabetes mellitus and coronary artery heart disease.

35. An array comprising the marker set of claim 19.

36. A method for modulating a physiologic or pathologic response to elevated cholesterol or lipid and/or adipogenesis in a cell, tissue or organism, the method comprising:

modulating expression or activity of at least one polypeptide encoded by a nucleic acid comprising:

(a) at least one polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a polynucleotide sequence complementary thereto;

(b) at least one polynucleotide sequence that hybridizes under stringent conditions to a polynucleotide sequence of (a);

(c) at least one polynucleotide sequence that is at least about 70% identical to a polynucleotide sequence of (a);

(d) at least one polynucleotide that encodes a polypeptide or peptide comprising a subsequence encoded by a polynucleotide sequence of (a);

(e) at least one polynucleotide sequence that hybridizes to a nucleic acid that is physically linked in the human genome to a nucleic acid comprising a polynucleotide sequence of (a), (b), (c) or (d); or,

(f) at least one polynucleotide sequence comprising at least about 10 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a sequence complementary thereto.

37. The method of claim 36, comprising modulating expression or activity of at least one polypeptide contributing to a condition selected from obesity, atherosclerosis, diabetes mellitus, or coronary artery heart disease.

38. The method of claim 36, comprising modulating a physiologic or pathologic response to elevated cholesterol or lipid and/or adipogenesis in one or more cell-types selected from the group comprising liver, adipose tissue, gall bladder, pancreas, monocytes, macrophages, foam cells, T cells, endothelia and smooth muscle derived from blood vessels and gut, fibroblasts, glia and nerve cells.

39. The method of claim 36, comprising modulating expression by expressing an exogenous nucleic acid comprising a polynucleotide sequence selected from SEQ ID NO:1 to SEQ ID NO:443.

40. The method of claim 36, comprising modulating expression in a cell line or non-human mammal.

41. The method of claim 40, wherein the non-human mammal comprises a mouse, a rat, a dog, a rabbit, a pig, a sheep or a non-human primate

42. The method of claim 39, comprising modulating expression by inducing or suppressing expression of an endogenous nucleic acid.

43. The method of claim 42, wherein the endogenous nucleic acid encodes a polypeptide comprising a subsequence encoded by a sequence selected from among SEQ ID NO:1-SEQ ID NO:443, or homologues thereof.

44. The method of claim 39, comprising introducing an exogenous nucleic acid comprising at least one promoter, which promoter regulates expression of the endogenous nucleic acid modulating cholesterol or lipid homeostasis and metabolism.

45. The method of claim 39, wherein expression is modulated in response to cholesterol and/or lipid.

46. The method of claim 39, further comprising detecting altered expression or activity of an expression product encoded by a nucleic acid comprising a polynucleotide sequence selected from SEQ ID NO:1-SEQ ID NO:443, or conservative variants thereof.

47. The method of claim 46, comprising detecting altered expression or activity in a high throughput assay.

48. The method of claim 45, comprising detecting altered expression or activity in response to administration of a pharmaceutical agent.

49. The method of claim 45, comprising detecting altered expression or activity in response to diet.

50. The method of claim 45, wherein a plurality of expression products are detected.

51. The method of claim 50, wherein the plurality of expression products are detected in an array.

52. The method of claim 51, wherein the array comprises a bead array.

53. The method of claim 45, wherein a data record comprising the altered expression or activity is recorded in a database.

54. The method of claim 52, wherein the database comprises a plurality of character strings recorded on a computer or in a computer readable medium.

55. A method for evaluating a condition or characteristic associated with a physiologic or pathologic response to excessive cholesterol or lipid and/or adipogenesis in a subject, the method comprising:

(i) providing a subject cell or tissue sample of nucleic acids;

(ii) detecting at least one polymorphic nucleic acid or at least one expression product corresponding to a polynucleotide sequence comprising:

(a) at least one polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a polynucleotide sequence complementary thereto;

(b) at least one polynucleotide sequence that hybridizes under stringent conditions to a polynucleotide sequence of (a);

(c) at least one polynucleotide sequence that is at least about 70% identical to a polynucleotide sequence of (a);

(d) at least one polynucleotide that encodes a polypeptide or peptide comprising a subsequence encoded by a polynucleotide sequence of (a);

(e) at least one polynucleotide sequence that hybridizes to a nucleic acid that is physically linked in the human genome to a nucleic acid comprising a polynucleotide sequence of (a), (b), (c) or (d); or,

(f) at least one polynucleotide sequence comprising at least about 10 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a sequence complementary thereto;

wherein the polymorphic nucleic acid or expression or activity of the expression product is correlatable to at least one condition or characteristic associated with a physiologic or pathologic response to elevated cholesterol or lipid and/or adipogenesis.

56. The method of claim 36, wherein the expression product comprises an RNA.

57. The method of claim 36, wherein the expression product comprises a protein or polypeptide.

58. The method of claim 36, wherein the detecting step comprises qualitative detection.

59. The method of claim 36, wherein the detecting step comprises quantitative detection.

60. A method for identifying a gene altering a physiologic or pathologic response to elevated cholesterol or lipid and/or adipogenesis, the method comprising:

(i) providing at least one nucleic acid comprising:

(a) at least one polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a polynucleotide sequence complementary thereto;

(b) at least one polynucleotide sequence that hybridizes under stringent conditions to a polynucleotide sequence of (a);

(c) at least one polynucleotide sequence that is at least about 70% identical to a polynucleotide sequence of (a);

(d) at least one polynucleotide that encodes a polypeptide or peptide comprising a subsequence encoded by a polynucleotide sequence of (a);

(e) at least one polynucleotide sequence that hybridizes to a nucleic acid that is physically linked in the human genome to a nucleic acid comprising a polynucleotide sequence of (a), (b), (c) or (d); or,

(f) at least one polynucleotide sequence comprising at least about 10 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a sequence complementary thereto; and,

(ii) identifying at least one nucleic acid corresponding to a gene capable of altering a physiologic or pathologic response to elevated cholesterol or lipid and/or adipogenesis.

61. The method of claim 60, comprising providing at least one expression vector comprising a polynucleotide sequence selected from among the polynucleotide sequences of (a), (b), (c), (d), (e) or (f).

62. The method of claim 60, comprising providing at least one probe comprising a polynucleotide sequence selected from among the polynucleotide sequences of (a), (b), (c), (d), (e) or (f), and,

hybridizing the at least one probe to an expression product of a gene capable of altering a physiologic or pathologic response to elevated cholesterol or lipid and/or adipogenesis.

63. The method of claim 60, wherein providing the at least one nucleic acid comprises amplifying a target sequence comprising a polynucleotide sequence selected from (a), (b), (c), (d), (e) or (f).

64. The method of claim 63, wherein the amplifying comprises a quantitative reverse transcriptase-polymerase chain reaction (RT-PCR).

65. The method of claim 63, comprising identifying a target sequence that is differentially expressed in response to cholesterol and/or lipid.

66. The method of claim 60, comprising hybridizing the at least one nucleic acid to a target nucleic acid; and, sequencing the target nucleic acid to detect a nucleotide polymorphism, which nucleotide polymorphism is associated with a condition selected from among obesity, atherosclerosis, diabetes mellitus or coronary artery heart disease.

67. An isolated or recombinant polypeptide comprising one or more amino acid sequences or subsequences encoded by a nucleic acid comprising:

(a) at least one polynucleotide sequence selected from the group consisting of SEQ ID NO:1-SEQ ID NO:443, or a polynucleotide sequence complementary thereto;

(b) at least one polynucleotide sequence that hybridizes under stringent conditions to a polynucleotide sequence of (a);

(c) at least one polynucleotide sequence that is at least about 70% identical to a polynucleotide sequence of (a);

(d) at least one polynucleotide sequence that hybridizes to a nucleic acid that is physically linked in the human genome to a nucleic acid comprising a polynucleotide sequence of (a), (b) or (c); or,

(e) at least one polynucleotide sequence comprising at least about 10 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO:443, or a sequence complementary thereto.

68. The isolated or recombinant polypeptide of claim 66, comprising a fusion protein.

69. The isolated or recombinant polypeptide of claim 66, comprising a peptide or polypeptide tag.

70. The isolated or recombinant polypeptide of claim 66, wherein the peptide or polypeptide tag comprises a reporter peptide or polypeptide.

71. The isolated or recombinant polypeptide of claim 66, wherein the peptide or polypeptide tag comprises an epitope.

72. The isolated or recombinant polypeptide of claim 66, wherein the peptide or polypeptide tag comprises a signal sequence.

73. A composition comprising the isolated or recombinant polypeptide of claim 66 and an excipient.

74. The composition of claim 73, wherein the excipient is a pharmaceutically acceptable excipient.

75. An array of polypeptides comprising two or more different polypeptides of claim 66.

76. An antibody specific for an isolated or recombinant polypeptide of claim 66.

77. The antibody of claim 76, wherein the antibody comprises a monoclonal antibody or a polyclonal serum.

78. One or more isolated or recombinant polypeptides that bind to the antibody of claim 66.

* * * * *

专利名称(译)	受胆固醇治疗和脂肪形成过程影响的基因		
公开(公告)号	US20030180764A1	公开(公告)日	2003-09-25
申请号	US10/339793	申请日	2003-01-08
[标]申请(专利权)人(译)	林克斯治疗公司		
申请(专利权)人(译)	LYNX THERAPEUTICS，INC.		
当前申请(专利权)人(译)	LYNX THERAPEUTICS，INC.		
[标]发明人	SHANG JIN BOWEN BENJAMIN		
发明人	SHANG, JIN BOWEN, BENJAMIN		
IPC分类号	C07K14/47 C12N15/10 C12Q1/68 G01N33/92 G06F19/00 G01N33/53 G01N33/48 G01N33/50 C07H21/04		
CPC分类号	C07K14/47 C12Q1/6883 G01N2800/044 G01N33/92 C12Q2600/158		
优先权	60/347286 2002-01-09 US		
外部链接	Espacenet USPTO		

摘要(译)

提供了基因，核酸，蛋白质，抗体，标记组和阵列。还提供了检测与升高的胆固醇和脂质相关的病症以及在脂肪生成过程中的方法。