



US 20080293582A1

(19) **United States**

(12) **Patent Application Publication**
LI et al.

(10) **Pub. No.: US 2008/0293582 A1**

(43) **Pub. Date: Nov. 27, 2008**

(54) **MARKERS AND METHODS FOR ASSESSING AND TREATING ULCERATIVE COLITIS AND RELATED DISORDERS USING A 43 GENE PANEL**

(76) Inventors: **XILIN LI**, WALLINGFORD, PA (US); **XIAO-YU SONG**, BRIDGEWATER, NJ (US)

Correspondence Address:
PHILIP S. JOHNSON
JOHNSON & JOHNSON
ONE JOHNSON & JOHNSON PLAZA
NEW BRUNSWICK, NJ 08933-7003 (US)

(21) Appl. No.: **11/847,737**

(22) Filed: **Aug. 30, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/823,976, filed on Aug. 30, 2006.

Publication Classification

(51) Int. Cl.	
<i>C40B 30/04</i>	(2006.01)
<i>C40B 30/00</i>	(2006.01)
<i>C12Q 1/68</i>	(2006.01)
<i>G01N 33/53</i>	(2006.01)

(52) **U.S. Cl. 506/9; 506/7; 435/6; 435/7.1**

(57) **ABSTRACT**

A method for prognostic or diagnostic assessment of a gastrointestinal-related disorder, such as ulcerative colitis, in a subject correlates the presence, absence, and/or magnitude of a gene in a sample with a reference standard to determine the presence and/or severity of the disorder, and/or the response to treatment for the disorder. The method enables identification of the effectiveness of candidate therapies.

**MARKERS AND METHODS FOR ASSESSING
AND TREATING ULCERATIVE COLITIS AND
RELATED DISORDERS USING A 43 GENE
PANEL**

CLAIM TO PRIORITY

[0001] This application claims the benefit of U.S. Provisional Application Ser. No. 60/823,976, filed 30 Aug. 2006, the entire contents of which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0002] The invention relates to the identification of expression profiles and the nucleic acids indicative of gastrointestinal-related disorders, such as ulcerative colitis, and to the use of such expression profiles and nucleic acids in diagnosis of ulcerative colitis and related diseases. The invention further relates to methods for identifying, using, and testing candidate agents and/or targets which modulate ulcerative colitis.

BACKGROUND OF THE INVENTION

[0003] Ulcerative colitis (UC) is a multifactorial autoimmune disease with a complex pathogenesis involving unidentified genetic, microbial, and environmental factors. Recent studies using microarray analysis of inflamed colonoscopic tissue biopsy vs. non-inflamed biopsy samples from UC patients revealed dysregulation of a few inflammatory cytokines (1), however, the etiology, pathogenesis, and role of tumor necrosis factor-alpha (TNF α) in UC is still poorly understood. TNF α is a critical proinflammatory cytokine in Crohn's disease as demonstrated by the therapeutic effect of infliximab on the induction and maintenance of clinical remission, closure of enterocutaneous, perianal, and rectovaginal fistulas, maintenance of fistula closure, and steroid tapering in Crohn's disease patients (2-5). However, the evidence to support a role of TNF α in the pathogenesis of UC has been controversial (6-10) despite the fact that it is also found at increased levels in the blood, colonic tissue, and stools of UC patients (11-13). A recent clinical study (ACT-1) by Rutgeerts et al. showed that infliximab is effective when administered at weeks 0, 2, 6 and every 8 weeks thereafter in achieving clinical response and remission in patients with moderate-to-severe active UC despite the use of conventional therapy supporting a critical pathogenic role of TNF α in UC (14).

[0004] Microarray technology is a powerful tool since it enables analysis of the expression of thousands of genes simultaneously and can also be automated allowing for a high-throughput format. In diseases associated with complex host functions, such as those known as immune mediated inflammatory diseases, such as UC, microarray results can provide a gene expression profile that can be of utility in designing new approaches to disease diagnosis and management. These approaches also serve to identify novel genes and annotating genes of unknown function heretofore unassociated with the disease or condition. Accordingly, there is a need to identify and characterize new gene markers useful in developing methods for diagnosing and treating autoimmune disorders, such as UC and Crohn's disease, as well as other diseases and conditions and how a patient would respond to a therapeutic intervention.

[0005] Gene expression can be modulated in several different ways, including by the use of siRNAs, shRNAs, antisense

molecules and DNazymes. siRNAs and shRNAs both work via the RNAi pathway and have been successfully used to suppress the expression of genes. RNAi was first discovered in worms and the phenomenon of gene silencing related to dsRNA was first reported in plants by Fire and Mello and is thought to be a way for plant cells to combat infection with RNA viruses. In this pathway, the long dsRNA viral product is processed into smaller fragments of 21-25 bp in length by a DICER-like enzyme and then the double-stranded molecule is unwound and loaded into the RNA induced silencing complex (RISC). A similar pathway has been identified in mammalian cells with the notable difference that the dsRNA molecules must be smaller than 30 bp in length in order to avoid the induction of the so-called interferon response, which is not gene specific and leads to the global shut down of protein synthesis in the cell.

[0006] Synthetic siRNAs have been successfully designed to selectively target a single gene and can be delivered to cells in vitro or in vivo. ShRNAs are the DNA equivalents of siRNA molecules and have the advantage of being incorporated into a cells' genome where they are replicated during every mitotic cycle.

[0007] DNazymes have also been used to modulate gene expression. DNazymes are catalytic DNA molecules that cleave single-stranded RNA. They are highly selective for the target RNA sequence and as such can be used to down-regulate specific genes through targeting of the messenger RNA.

[0008] Accordingly, there is a need to identify and characterize new gene markers useful in developing methods for diagnosing and treating autoimmune disorders, such as UC and Crohn's disease, as well as other diseases and conditions.

SUMMARY OF THE INVENTION

[0009] The present invention relates to a method of diagnosing and/or treating UC and/or related diseases or disorders by identifying and using candidate agents and/or targets which modulate such diseases or disorders. The present invention includes the discovery of panels of genes, one of 43 genes, that have modified expression levels in patients with UC and/or treated with an agent effective in reducing the symptoms of UC (and modified levels in patients whose UC treatment has not been effective). The modified expression levels constitute a profile that can serve as a biomarker profile indicative of UC and/or the response of a subject to treatment.

[0010] In a particular embodiment, the present invention comprises a method of determining the efficacy of the treatment for UC based on the pattern of gene expression of one or more of the 43 genes which constitute the profile. One or more of these genes may be from a category of genes, for example, an innate or adaptive immune response-related gene, a cell-cell interaction, cell-matrix interaction or matrix regulation-related gene, a cell-cell, intracellular signaling pathway-related gene, a cell growth and apoptosis-related gene, a protein regulation-related gene, a metabolic regulation-related gene, a cytoskeleton organization-related gene, a developmental regulation-related gene, and a transcriptional regulation-related gene, and the like. This can be done for a subject, for example, prior to the manifestation of other gross measurements of clinical response. In one embodiment, the method of screening drug candidates includes comparing the level of expression in the absence of the drug candidate to the level of expression in the presence of the drug candidate, wherein the concentration of the drug candidate can vary when present,

and wherein the comparison can occur during treatment or after treatment with the drug candidate. In a typical embodiment, the cell specimen expresses at least two expression profile genes. The profile genes may show an increase or decrease.

[0011] In one embodiment, the UC-related gene profile is used to create an array-based method for prognostic or diagnostic purposes, the method comprising:

[0012] (a) preparing a representative mixture of nucleic acids from a specimen obtained from a patient and causing said sample nucleic acids in the mixture to be labeled with a detectable marker;

[0013] (b) contacting a sample with an array comprising a plurality of nucleic acid segments, wherein each nucleic acid segment is immobilized to a discrete and known address on a substrate surface wherein the panel of UC-related biomarkers is identified as a feature of the array by address, the array further comprises at least one calibration nucleic acid at a known address on the substrate, and contacting is performed under conditions in which a sample nucleic acid specifically may bind to the nucleic acid segment immobilized on the arrays;

[0014] (c) performing a statistical comparison of all test samples from treated patients and a reference standard; and

[0015] (d) comparing the pattern of intensity changes in features for the test sample to the pattern of intensity changes for those features which are members of the UC-related gene profile with historical patterns for samples taken from patients responsive to treatment with an anti-TNF antibody.

[0016] Optionally, statistical analysis is performed on the changes in levels of members of the gene panel to evaluate the significance of these changes and to identify which members are meaningful members of the panel.

[0017] In an alternative embodiment, the present invention comprises a kit for diagnosing UC and/or related diseases or disorders by identifying and using candidate agents and/or targets which modulate such diseases or disorders and for determining the efficacy of the treatment for UC and/or related diseases or disorders based on the pattern of gene expression.

[0018] Another embodiment of the present invention relates to agonists and/or antagonists of the transcription of the genes or of the gene products of the UC-related gene panel and a method of using UC-related gene panel antagonists, including antibodies directed toward UC-related gene panel products, to treat UC or related disorders.

[0019] In one aspect, the UC-related gene panel antagonist is an antibody that specifically binds UC-related gene panel product. A particular advantage of such antibodies is that they are capable of binding UC-related gene panel product in a manner that prevents its action. The method of the present invention thus employs antibodies having the desirable neutralizing property which makes them ideally suited for therapeutic and preventative treatment of disease states associated with various UC-related disorders in human or nonhuman patients. Accordingly, the present invention is directed to a method of treating UC or a related disease or condition in a patient in need of such treatment which comprises administering to the patient an amount of a neutralizing UC-related gene panel product antibody to inhibit the UC-related disease or condition.

[0020] In another aspect, the invention provides methods for modulating activity of a member of a UC-related gene panel comprising contacting a cell with an agent (e.g., antagonist or agonist) that modulates (inhibits or enhances) the activity or expression of the member of the UC-related gene panel such that activity or expression in the cell is modulated. In a preferred embodiment, the agent is an antibody that specifically binds to the UC-related gene panel. In other embodiments, the modulator is a peptide, peptidomimetic, or other small molecule.

[0021] The present invention also provides methods of treating a subject having UC or related disorder wherein the disorder can be ameliorated by modulating the amount or activity of the UC-related gene panel. The present invention also provides methods of treating a subject having a disorder characterized by aberrant activity of the UC-related gene panel product or one of their encoding polynucleotide by administering to the subject an agent that is a modulator of the activity of the UC-related gene panel product or a modulator of the expression of a UC-related gene panel.

[0022] In one embodiment, the modulator is a polypeptide or small molecule compound. In another embodiment, the modulator is a polynucleotide. In a particular embodiment, the UC-related gene panel antagonist is an siRNA molecule, an shRNA molecule, an antisense molecule, a ribozyme, or a DNzyme capable of preventing the production of UC-related gene panel by cells.

[0023] The present invention further provides any invention described herein.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0024] The following definitions are set forth to illustrate and define the meaning and scope of various terms used to describe the invention herein.

[0025] An "activity," a biological activity, and a functional activity of a polypeptide refers to an activity exerted by a gene of the UC-related gene panel in response to its specific interaction with another protein or molecule as determined in vivo, in situ, or in vitro, according to standard techniques. Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein, or an indirect activity, such as a cellular process mediated by interaction of the protein with a second protein or a series of interactions as in intracellular signaling or the coagulation cascade.

[0026] An "antibody" includes any polypeptide or peptide containing molecule that comprises at least a portion of an immunoglobulin molecule, such as but not limited to, at least one complementarity determining region (CDR) of a heavy or light chain or a ligand binding portion thereof, a heavy chain or light chain variable region, a heavy chain or light chain constant region, a framework region, or any portion, fragment or variant thereof. The term "antibody" is further intended to encompass antibodies, digestion fragments, specified portions and variants thereof, including antibody mimetics or comprising portions of antibodies that mimic the structure and/or function of an antibody or specified fragment or portion thereof, including single chain antibodies and fragments thereof. For example, antibody fragments include, but are not limited to, Fab (e.g., by papain digestion), Fab' (e.g., by pepsin digestion and partial reduction) and F(ab')₂ (e.g., by pepsin digestion), fabc (e.g., by plasmin digestion), pFc' (e.g., by pepsin or plasmin digestion), Fd (e.g., by pepsin digestion),

partial reduction and reaggregation), Fv or scFv (e.g., by molecular biology techniques) fragments, and single domain antibodies (e.g., V_H or V_L), are encompassed by the invention (see, e.g., Colligan, et al., eds., *Current Protocols in Immunology*, John Wiley & Sons, Inc., NY (1994-2001); Colligan et al., *Current Protocols in Polypeptide Science*, John Wiley & Sons, NY (1997-2001)).

[0027] The terms “array” or “microarray” or “biochip” or “chip” as used herein refer to articles of manufacture or devices comprising a plurality of immobilized target elements, each target element comprising a “clone,” “feature,” “spot” or defined area comprising a particular composition, such as a biological molecule, e.g., a nucleic acid molecule or polypeptide, immobilized to a solid surface, as discussed in further detail, below.

[0028] “Complement of” or “complementary to” a nucleic acid sequence of the invention refers to a polynucleotide molecule having a complementary base sequence and reverse orientation as compared to a first polynucleotide.

[0029] A “gene” is a set of segments of nucleic acid that contains the information necessary to produce a functional RNA product in a controlled manner. By “gene” is meant a DNA sequence capable of being transcribed to produce a unique gene product, which product will usually be a protein synthesized from the transcribed, properly processed, and translated gene sequence. Some genes encode gene products that are transcribed but not translated, such as rRNA genes and tRNA genes. Gene expression, or simply “expression”, is the process by which the inheritable information which comprises a gene, such as the DNA sequence, is made manifest as a biologically functional gene product, such as protein or RNA. The genes of eukaryotic organisms can contain non-coding regions called introns that are removed from the messenger RNA in a process known as splicing. Exons are the regions that encode the gene product. One single gene can lead to the synthesis of multiple proteins through the different arrangements of exons produced by alternative splicings. Several steps in the gene expression process may be modulated, including the transcription step and mRNA processing step(s). The level of gene expression can have a profound effect on the functions (actions) of the gene and therefore of the gene product in the organism. A gene may exist in one of multiple alternative forms, each of which is a viable DNA sequence occupying a given position, or locus on a chromosome known as alleles with nucleic acid variations which may produce changes in the encoded protein gene product or, by virtue of the redundancy in the genetic code, be silent. Thus, DNA fragments representative of a single gene may comprise variations in length of the segment or variations in sequence.

[0030] “Identity,” as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, “identity” also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. “Identity” and “similarity” can be readily calculated by known methods, including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A. M., ed., Oxford University Press, New York, 1988; *Bio-computing: Informatics and Genome Projects*, Smith, D. W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A. M., and Griffin, H. G., eds., Humana Press, N.J., 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and

Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *Siam J. Applied Math.*, 48:1073 (1988). In addition, values for percentage identity can be obtained from amino acid and nucleotide sequence alignments generated using the default settings for the AlignX component of Vector NTI Suite 8.0 (Informax, Frederick, Md.).

[0031] The terms “specifically hybridize to,” “hybridizing specifically to,” “specific hybridization” and “selectively hybridize to,” as used herein refer to the binding, duplexing, or hybridizing of a nucleic acid molecule preferentially to a particular nucleotide sequence under stringent conditions. The term “stringent conditions” refers to conditions under which a probe will hybridize preferentially to its target sub-sequence; and to a lesser extent to, or not at all to, other sequences. A “stringent hybridization” and “stringent hybridization wash conditions” in the context of nucleic acid hybridization (e.g., as in array, Southern or Northern hybridizations) are sequence dependent, and are different under different environmental parameters. Alternative hybridization conditions that can be used to practice the invention are described in detail, below. In alternative aspects, the hybridization and/or wash conditions are carried out under moderate conditions, stringent conditions and very stringent conditions, as described in further detail, below. Alternative wash conditions are also used in different aspects, as described in further detail, herein.

[0032] The phrases “labeled biological molecule” or “labeled with a detectable composition” or “labeled with a detectable moiety” as used herein refer to a biological molecule, e.g., a nucleic acid, comprising a detectable composition, i.e., a label, as described in detail, below. The label can also be another biological molecule, as a nucleic acid, e.g., a nucleic acid in the form of a stem-loop structure as a “molecular beacon,” as described below. This includes incorporation of labeled bases (or, bases which can bind to a detectable label) into the nucleic acid by, e.g., nick translation, random primer extension, amplification with degenerate primers, and the like. Any label can be used, e.g., chemiluminescent labels, radiolabels, enzymatic labels and the like. The label can be detectable by any means, e.g., visual, spectroscopic, photochemical, biochemical, immunochemical, physical, chemical and/or chemiluminescent detection. The invention can use arrays comprising immobilized nucleic acids comprising detectable labels.

[0033] The term “nucleic acid” as used herein refers to a deoxyribonucleotide (DNA) or ribonucleotide (RNA) in either single- or double-stranded form. The term encompasses nucleic acids containing known analogues of natural nucleotides. The term nucleic acid is used interchangeably with gene, DNA, RNA, cDNA, mRNA, oligonucleotide primer, probe and amplification product. The term also encompasses DNA backbone analogues, such as phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene (methylimino), 3'-N-carbamate, morpholino carbamate, and peptide nucleic acids (PNAs).

[0034] The terms “sample” or “sample of nucleic acids” as used herein refer to a sample comprising a DNA or RNA, or nucleic acid representative of DNA or RNA isolated from a natural source. A “sample of nucleic acids” is in a form suitable for hybridization (e.g., as a soluble aqueous solution) to another nucleic acid (e.g., immobilized probes). The sample nucleic acid may be isolated, cloned, or extracted

from particular cells or tissues. The cell or tissue sample from which the nucleic acid sample is prepared is typically taken from a patient having or suspected of having UC or a related disease or condition. Methods of isolating cell and tissue samples are well known to those of skill in the art and include, but are not limited to, aspirations, tissue sections, needle biopsies, and the like. Frequently the sample will be a "clinical sample" which is a sample derived from a patient, including sections of tissues such as frozen sections or paraffin sections taken for histological purposes. The sample can also be derived from supernatants (of cells) or the cells themselves taken from patients or from cell cultures, cells from tissue culture and other media in which it may be desirable to detect the response to drug candidates. In some cases, the nucleic acids may be amplified using standard techniques such as PCR, prior to the hybridization. The probe can be produced from and collectively can be representative of a source of nucleic acids from one or more particular (pre-selected) portions of, e.g., a collection of polymerase chain reaction (PCR) amplification products, substantially an entire chromosome or a chromosome fragment, or substantially an entire genome, e.g., as a collection of clones, e.g., BACs, PACs, YACs, and the like (see below).

[0035] "Nucleic acids" are polymers of nucleotides, wherein a nucleotide comprises a base linked to a sugar which sugars are in turn linked one to another by an interceding at least bivalent molecule, such as phosphoric acid. In naturally occurring nucleic acids, the sugar is either 2'-deoxyribose (DNA) or ribose (RNA). Unnatural poly- or oligonucleotides contain modified bases, sugars, or linking molecules, but are generally understood to mimic the complementary nature of the naturally occurring nucleic acids after which they are designed. An example of an unnatural oligonucleotide is an antisense molecule composition that has a phosphorothioate backbone. An "oligonucleotide" generally refers to a nucleic acid molecule having less than 30 nucleotides.

[0036] The term "profile" means a pattern and relates to the magnitude and direction of change of a number of features. The profile may be interpreted stringently, i.e., where the variation in the magnitude and/or number of features within the profile displaying the characteristic is substantially similar to a reference profile or it may be interpreted less stringently, for example, by requiring a trend rather than an absolute match of all or a subset of feature characteristics.

[0037] The terms "protein," "polypeptide," and "peptide" include "analogs," or "conservative variants" and "mimetics" or "peptidomimetics" with structures and activity that substantially correspond to the polypeptide from which the variant was derived, as discussed in detail above.

[0038] A "polypeptide" is a polymer of amino acid residues joined by peptide bonds, and a peptide generally refers to amino acid polymers of 12 or less residues. Peptide bonds can be produced naturally as directed by the nucleic acid template or synthetically by methods well known in the art.

[0039] A "protein" is a macromolecule comprising one or more polypeptide chains. A protein may further comprise substituent groups attached to the side groups of the amino acids not involved in formation of the peptide bonds. Typically, proteins formed by eukaryotic cell expression also contain carbohydrates. Proteins are defined herein in terms of their amino acid sequence or backbone and substituents are not specified, whether known or not.

[0040] The term "receptor" denotes a molecule having the ability to affect biological activity, in e.g., a cell, as a result of interaction with a specific ligand or binding partner. Cell membrane bound receptors are characterized by an extracellular ligand-binding domain, one or more membrane spanning or transmembrane domains, and an intracellular effector domain that is typically involved in signal transduction. Ligand binding to cell membrane receptors causes changes in the extracellular domain that are communicated across the cell membrane, direct or indirect interaction with one or more intracellular proteins, and alters cellular properties, such as enzyme activity, cell shape, or gene expression profile. Receptors may also be untethered to the cell surface and may be cytosolic, nuclear, or released from the cell altogether. Non-cell associated receptors are termed soluble receptors or ligands.

[0041] All publications or patents cited herein are entirely incorporated herein by reference, whether or not specifically designated accordingly, as they show the state of the art at the time of the present invention and/or provide description and enablement of the present invention. Publications refer to any scientific or patent publications, or any other information available in any media format, including all recorded, electronic or printed formats. The following references are entirely incorporated herein by reference: Ausubel, et al., ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., NY (1987-2001); Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor, N.Y. (1989); Harlow and Lane, *antibodies, a Laboratory Manual*, Cold Spring Harbor, N.Y. (1989); Colligan, et al., eds., *Current Protocols in Immunology*, John Wiley & Sons, Inc., NY (1994-2001); Colligan et al., *Current Protocols in Protein Science*, John Wiley & Sons, NY (1997-2001).

Gene Panel Identification and Validation

[0042] The present invention provides novel methods for diagnosis of disorders associated with UC, as well as methods for screening for compositions which modulate the symptoms of UC, particularly the mucosal layer of the rectum and all or part of the colon. By "UC" or grammatical equivalents as used herein, is meant a disease state or condition which is marked by diarrhea, rectal bleeding, tenesmus, passage of mucus, and crampy abdominal pain.

[0043] In one aspect, the expression levels of genes are determined in different patient samples for which diagnosis information is desired, to provide expression profiles. An expression profile of a particular sample is essentially a "fingerprint" of the state of the sample; while two states may have any particular gene similarly expressed, the evaluation of a number of genes simultaneously allows the generation of a gene expression profile that is unique to the state of the patient sample. That is, normal tissue may be distinguished from lesion tissue and tissue from a treated patient may be distinguished from an untreated patient. By comparing expression profiles of tissue in different disease states that are known, information regarding which genes are important (including both up- and down-regulation of genes) in each of these states is obtained.

[0044] The identification of sequences (genes) that are differentially expressed in disease tissue allows the use of this information in a number of ways. For example, the evaluation of a particular treatment regime may be evaluated. Similarly, diagnosis may be done or confirmed by comparing patient samples with the known expression profiles. Furthermore,

these gene expression profiles (or individual genes) allow screening of drug candidates with an eye to mimicking or altering a particular expression profile; for example, screening can be done for drugs that suppress the angiogenic expression profile.

[0045] This may be done by making biochips comprising sets of the important disease genes, which can then be used in these screens. These methods can also be performed on the protein basis; that is, protein expression levels of the UC-related gene product proteins can be evaluated for diagnostic purposes or to screen candidate agents. In addition, the nucleic acid sequences comprising the UC-related gene profile can be used to design a therapeutic including the administration of antisense nucleic acids, or the protein coded for by the gene sequence can be administered as a component of a vaccine.

[0046] Thus, the present invention provides information on nucleic acid and protein sequences that are differentially expressed in UC, herein termed "UC-related gene sequences." As outlined below, UC-related gene sequences include those that are upregulated (i.e., expressed at a higher level) in disorders associated with UC, as well as those that are down-regulated (i.e., expressed at a lower level). In a preferred embodiment, the UC-related gene sequences are from humans; however, as will be appreciated by those in the art, UC-related gene sequences from other organisms may be useful in animal models of disease and drug evaluation; thus, other UC-related gene sequences are provided, from vertebrates, including mammals, including rodents (rats, mice, hamsters, guinea pigs, etc.), primates, farm animals (including sheep, goats, pigs, cows, horses, etc). UC-related gene sequences from other organisms may be obtained using the techniques known in the art.

[0047] UC-related gene sequences can include both nucleic acid and amino acid sequences. In a preferred embodiment, the UC-related gene sequences are recombinant nucleic acids. By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed *in vitro*, in general, by the manipulation of nucleic acid by polymerases and endonucleases, in a form not normally found in nature. Thus, an isolated nucleic acid, in a linear form, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e., using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention.

Method of Practicing the Invention

[0048] The invention provides *in silico*, array-based methods relying on the relative amount of a binding molecule (e.g., nucleic acid sequence) in two or more samples. Also provided are computer-implemented methods for determining the relative amount of a binding molecule (e.g., nucleic acid sequence) in two or more samples and using the determined relative binding amount to diagnose and stage disease, predict responsiveness to a particular therapy, and monitor and enhance therapeutic treatment.

[0049] In practicing the methods of the invention, two or more samples of labeled biological molecules (e.g., nucleic

acid) are applied to two or more arrays, where the arrays have substantially the same complement of immobilized binding molecule (e.g., immobilized nucleic acid capable of hybridizing to labeled sample nucleic acid). The two or more arrays are typically multiple copies of the same array. However, because each "spot," "clone" or "feature" on the array has similar biological molecules (e.g., nucleic acids of the same sequence) and the biological molecules (e.g., nucleic acid) in each spot is known, as is typical of nucleic acid and other arrays, it is not necessary that the multiple arrays used in the invention be identical in configuration it is only necessary that the position of each feature on the substrate be known, that is, have an address. Thus, in one aspect, multiple biological molecules (e.g., nucleic acid) in samples are comparatively bound to the array (e.g., hybridized simultaneously) and the information gathered is coded so that the results are based on the inherent properties of the feature (e.g., the nucleic acid sequence) and not its position on the substrate.

[0050] Amplification of Nucleic Acids

[0051] Amplification using oligonucleotide primers can be used to generate nucleic acids used in the compositions and methods of the invention, to detect or measure levels of test or control samples hybridized to an array, and the like. The skilled artisan can select and design suitable oligonucleotide amplification primers. Amplification methods are also well known in the art, and include, e.g., polymerase chain reaction, PCR (PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., ligase chain reaction (LCR) (see, e.g., Wu (1989) *Genomics* 4:560; Landegren (1988) *Science* 241:1077; Barringer (1990) *Gene* 89:117); transcription amplification (see, e.g., Kwoh (1989) *Proc. Natl. Acad. Sci. USA* 86:1173); and, self-sustained sequence replication (see, e.g., Guatelli (1990) *Proc. Natl. Acad. Sci. USA* 87:1874); Q Beta replicase amplification (see, e.g., Smith (1997) *J. Clin. Microbiol.* 35:1477-1491), automated Q-beta replicase amplification assay (see, e.g., Burg (1996) *Mol. Cell. Probes* 10:257-271) and other RNA polymerase mediated techniques (e.g., NASBA, Canguene, Mississauga, Ontario); see also Berger (1987) *Methods Enzymol.* 152:307-316; Sambrook; Ausubel; U.S. Pat. Nos. 4,683,195 and 4,683,202; Sookninan (1995) *Biotechnology* 13:563-564.

[0052] Hybridizing Nucleic Acids

[0053] In practicing the methods of the invention, test and control samples of nucleic acid are hybridized to immobilized probe nucleic acid, e.g., on arrays. In alternative aspects, the hybridization and/or wash conditions are carried out under moderate conditions, stringent conditions and very stringent conditions. An extensive guide to the hybridization of nucleic acids is found in, e.g., Sambrook Ausubel, Tijssen. Generally, highly stringent hybridization and wash conditions are selected to be about 5° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on an array or a filter in a Southern or northern blot is 42° C. using standard hybridization solutions (see, e.g., Sambrook), with the hybridization being carried out overnight. An example of highly stringent wash

conditions is 0.15 M NaCl at 72° C. for about 15 minutes. An example of stringent wash conditions is a 0.2×SSC wash at 65° C. for 15 minutes (see, e.g., Sambrook). Often, a high stringency wash is preceded by a medium or low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1×SSC at 45° C. for 15 minutes. An example of a low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4× to 6×SSC at 40° C. for 15 minutes.

[0054] In alternative aspects of the compositions and methods of the invention, e.g., in practicing comparative nucleic acid hybridization, such as comparative genomic hybridization (CGH) with arrays, the fluorescent dyes Cy3® and Cy5® are used to differentially label nucleic acid fragments from two samples, e.g., the array-immobilized nucleic acid versus the sample nucleic acid, or, nucleic acid generated from a control versus a test cell or tissue. Many commercial instruments are designed to accommodate the detection of these two dyes. To increase the stability of Cy5®, or fluors or other oxidation-sensitive compounds, antioxidants and free radical scavengers can be used in hybridization mixes, the hybridization and/or the wash solutions. Thus, Cy5® signals are dramatically increased and longer hybridization times are possible. See WO 0194630 A2 and U.S. Patent Application No. 20020006622.

[0055] To further increase the hybridization sensitivity, hybridization can be carried out in a controlled, unsaturated humidity environment; thus, hybridization efficiency is significantly improved if the humidity is not saturated. See WO 0194630 A2 and U.S. Patent Application No. 20020006622. The hybridization efficiency can be improved if the humidity is dynamically controlled, i.e., if the humidity changes during hybridization. Mass transfer will be facilitated in a dynamically balanced humidity environment. The humidity in the hybridization environment can be adjusted stepwise or continuously. Array devices comprising housings and controls that allow the operator to control the humidity during pre-hybridization, hybridization, wash and/or detection stages can be used. The device can have detection, control and memory components to allow pre-programming of the humidity and temperature controls (which are constant and precise or which fluctuate), and other parameters during the entire procedural cycle, including pre-hybridization, hybridization, wash and detection steps. See WO 0194630 A2 and U.S. Patent Application No. 20020006622.

[0056] The methods of the invention can comprise hybridization conditions comprising osmotic fluctuation. Hybridization efficiency (i.e., time to equilibrium) can also be enhanced by a hybridization environment that comprises changing hyper-/hypo-tonicity, e.g., a solute gradient. A solute gradient is created in the device. For example, a low salt hybridization solution is placed on one side of the array hybridization chamber and a higher salt buffer is placed on the other side to generate a solute gradient in the chamber. See WO 0194630 A2 and U.S. Patent Application No. 20020006622.

[0057] Blocking the Ability of Repetitive Nucleic Acid Sequences to Hybridize

[0058] The methods of the invention can comprise a step of blocking the ability of repetitive nucleic acid sequences to hybridize (i.e., blocking “hybridization capacity”) in the immobilized nucleic acid segments. The hybridization capacity of repetitive nucleic acid sequences in the sample nucleic acid sequences can be blocked by mixing sample nucleic acid

sequences with unlabeled or alternatively labeled repetitive nucleic acid sequences. Sample nucleic acid sequences can be mixed with repetitive nucleic acid sequences before the step of contacting with the array-immobilized nucleic acid segments. Blocking sequences are for example, Cot-1 DNA, salmon sperm DNA, or specific repetitive genomic sequences. The repetitive nucleic acid sequences can be unlabeled. A number of methods for removing and/or disabling the hybridization capacity of repetitive sequences using, e.g., Cot-1 are known; see, e.g., Craig (1997) *Hum. Genet.* 100: 472-476; WO 93/18186. Repetitive DNA sequences can be removed from library probes by means of magnetic purification and affinity PCR, see, e.g., Rauch (2000) *J. Biochem. Biophys. Methods* 44:59-72.

[0059] Arrays are generically a plurality of target elements immobilized onto the surface of the plate as defined “spots” or “clusters,” or “features,” with each target element comprising one or more biological molecules (e.g., nucleic acids or polypeptides) immobilized to a solid surface for specific binding (e.g., hybridization) to a molecule in a sample. The immobilized nucleic acids can contain sequences from specific messages (e.g., as cDNA libraries) or genes (e.g., genomic libraries), including a human genome. Other target elements can contain reference sequences and the like. The biological molecules of the arrays may be arranged on the solid surface at different sizes and different densities. The densities of the biological molecules in a cluster and the number of clusters on the array will depend upon a number of factors, such as the nature of the label, the solid support, the degree of hydrophobicity of the substrate surface, and the like. Each feature may comprise substantially the same biological molecule (e.g., nucleic acid), or, a mixture of biological molecules (e.g., nucleic acids of different lengths and/or sequences). Thus, for example, a feature may contain more than one copy of a cloned piece of DNA, and each copy may be broken into fragments of different lengths.

[0060] Array substrate surfaces onto which biological molecules (e.g., nucleic acids) are immobilized can include nitrocellulose, glass, quartz, fused silica, plastics and the like, as discussed further, below. The compositions and methods of the invention can incorporate in whole or in part designs of arrays, and associated components and methods, as described, e.g., in U.S. Pat. Nos. 6,344,316; 6,197,503; 6,174,684; 6,159,685; 6,156,501; 6,093,370; 6,087,112; 6,087,103; 6,087,102; 6,083,697; 6,080,585; 6,054,270; 6,048,695; 6,045,996; 6,022,963; 6,013,440; 5,959,098; 5,856,174; 5,843,655; 5,837,832; 5,770,456; 5,723,320; 5,700,637; 5,695,940; 5,556,752; 5,143,854; see also, e.g., WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; WO 89/10977; see also, e.g., Johnston (1998) *Curr. Biol.* 8:R171-174; Schummer (1997) *Biotechniques* 23:1087-1092; Kern (1997) *Biotechniques* 23:120-124; Solinas-Toldo (1997) *Genes, Chromosomes & Cancer* 20:399-407; Bowtell (1999) *Nature Genetics Supp.* 21:25-32; Epstein (2000) *Current Opinion in Biotech.* 11:36-41; Mendoza (1999) *Biotechniques* 27: 778-788; Lueking (1999) *Anal. Biochem.* 270:103-111; Davies (1999) *Biotechniques* 27:1258-1261.

[0061] Substrate Surfaces

[0062] Substrate surfaces that can be used in the compositions and methods of the invention include, for example, glass (see, e.g., U.S. Pat. No. 5,843,767), ceramics, and quartz. The arrays can have substrate surfaces of a rigid, semi-rigid or flexible material. The substrate surface can be flat or planar, be shaped as wells, raised regions, etched trenches, pores,

beads, filaments, or the like. Substrate surfaces can also comprise various materials such as nitrocellulose, paper, crystalline substrates (e.g., gallium arsenide), metals, metalloids, polycryloylmorpholide, various plastics and plastic copolymers, Nylon®, Teflon®, polyethylene, polypropylene, latex, polymethacrylate, poly (ethylene terephthalate), rayon, nylon, poly(vinyl butyrate), and cellulose acetate. The substrates may be coated and the substrate and the coating may be functionalized to, e.g., enable conjugation to an amine.

[0063] Arrays Comprising Sequences Representative of Human Genes

[0064] As genomic DNA comprises nucleic acid sequences that do not code for gene products, e.g. sequences involved in gene regulation and intervening sequences (introns), arrays comprising discreet probes or DNA fragments representative of exons of a gene which are expressed and form functional gene products may be used rather than arrays created e.g. from random fragmentation of a genome or chromosome.

[0065] In one embodiment, a DNA chip comprising DNA fragments which representative of coding sequences of specified genetic loci, preferably specific named genes, are used to detect the expression patterns of genes from samples of UC patients. One example of such a commercially available DNA chip is the Human Genome U133 (HG-U133) Set, consisting of two GeneChip® arrays, available from Affymetrix (Sunnyvale, Calif.). The Human Genome U133 contains almost 45,000 probe sets representing more than 39,000 transcripts derived from approximately 33,000 well-substantiated human genes. According to the documentation available from Affymetrix, the Human Genome U133 set design uses sequences selected from GenBank®, dbEST, and RefSeq. The sequence clusters were created from the UniGene database (Build 133, Apr. 20, 2001). They were then refined by analysis and comparison with a number of other publicly available databases including the Washington University EST trace repository and the University of California, Santa Cruz Golden Path human genome database (April 2001 release). While some commercially available gene chips are useful for research purposes, similar arrays using probe sets of oligonucleotides or DNA fragments representative of the UC-gene product panels of the present invention for detecting gene expression related to the treatment, prediction, or diagnosis of UC can be manufactured based on the techniques described in U.S. Pat. Nos. 7,135,285, 6,610,482, 5,800,992, and 6,054,270.

[0066] Arrays Comprising Calibration Sequences

[0067] The invention contemplates the use of arrays comprising immobilized calibration sequences for normalizing the results of array-based hybridization reactions, and methods for using these calibration sequences, e.g., to determine the copy number of a calibration sequence to “normalize” or “calibrate” ratio profiles. The calibration sequences can be substantially the same as a unique sequence in an immobilized nucleic acid sequence on an array. For example, a “marker” sequence from each “spot” or “biosite” on an array (which is present only on that spot, making it a “marker” for that spot) is represented by a corresponding sequence on one or more “control” or “calibration” spot(s).

[0068] The “control spots” or “calibration spots” are used for “normalization” to provide information that is reliable and repeatable. Control spots can provide a consistent result independent of the labeled sample hybridized to the array (or a labeled binding molecule from a sample). The control spots can be used to generate a “normalization” or “calibration”

curve to offset possible intensity errors between the two arrays (or more) used in the in silico, array-based methods of the invention.

[0069] One method of generating a control on the array would be to use an equimolar mixture of all the biological molecules (e.g., nucleic acid sequences) spotted on the array and generating a single spot. This single spot would have equal amounts of the biological molecules (e.g., nucleic acid sequences) from all the other spots on the array. Multiple control spots can be generated by varying the concentration of the equimolar mixture.

[0070] Samples and Specimens

[0071] The sample nucleic acid may be isolated, cloned, or extracted from particular cells, tissues, or other specimens. The cell or tissue sample from which the nucleic acid sample is prepared is typically taken from a patient having or suspected of having UC or a related condition. Methods of isolating cell and tissue samples are well known to those of skill in the art and include, but are not limited to, aspirations, tissue sections, needle biopsies, and the like. Frequently, the sample will be a “clinical sample” which is a sample derived from a patient, including whole blood, or sections of tissues, such as frozen sections or paraffin sections taken for histological purposes. The sample can also be derived from supernatants (of cells) or the cells themselves taken from patients or from cell cultures, cells from tissue culture and other media in which it may be desirable to detect the response to drug candidates. In some cases, the nucleic acids may be amplified using standard techniques such as PCR, prior to the hybridization.

[0072] In one embodiment, the present invention is a post-treatment method of monitoring disease resolution. The method includes (1) taking a colon biopsy or other specimen from an individual diagnosed with UC or a related disease or disorder, (2) measuring the expression levels of the profile genes of the panel, (3) comparing the post-treatment expression level of the genes with a pre-treatment reference profile for the individual, and (4) determining the prognosis for resolution of the UC condition by monitoring at least one constituent of the UC-related gene profile.

[0073] In another embodiment, the present invention is a diagnostic method for UC and the reference standard (sample) is taken from an uninvolved site and the test sample from a suspect biopsy.

[0074] Methods of Assessing Biomarker Utility

[0075] The diagnostic and prognostic utility of the present biomarker gene panel for assessing a patient’s response to treatment, prognosis, or presence, extent, severity or stage of disease can be validated by using other means for assessing a patient’s state of health or disease. For example, gross measurement of disease may be assessed and recorded by certain imaging methods, such as but not limited to: physician evaluation, imaging by photographic, radiometric, or magnetic resonance technology. General indices of health or disease further include serum or blood composition (protein, liver enzymes, pH, electrolytes, red cell volume, hematocrit, hemoglobin, or specific protein). However, in some diseases, the etiology is still poorly understood. UC is an example of one such disease.

Patient Assessment and Monitoring

[0076] Some of the genes in the panel have been reported to be aberrantly expressed in UC patients previously, such as IL-1b, IL-1ra, IL-6, superoxide dismutase, selecting, inte-

grins, and various MMPs etc., the expression patterns of the genes over the course of treatment have not been studied in the treatment of UC, and none has been identified as having predictive value. The panel of gene expression biomarkers disclosed herein permits the generation of methods for rapid and reliable prediction, diagnostic tools that predict the clinical outcome of a UC trial, or prognostic tools for tracking the efficacy of UC therapy. Diagnostic and prognostic methods based on detecting these genes in a sample are provided. These compositions may be used, for example, for the diagnosis, prevention and treatment of a range of immune-mediated inflammatory diseases.

[0077] Therapeutic Agents

[0078] Antagonists

[0079] As used herein, the term “antagonists” refer to substances which inhibit or neutralize the biologic activity of the gene product of the UC-related gene panel of the invention. Such antagonists accomplish this effect in a variety of ways. One class of antagonists will bind to the gene product protein with sufficient affinity and specificity to neutralize the biologic effects of the protein. Included in this class of molecules are antibodies and antibody fragments (such as, for example, F(ab) or F(ab')₂ molecules). Another class of antagonists comprises fragments of the gene product protein, muteins or small organic molecules, i.e., peptidomimetics, that will bind to the cognate binding partners or ligands of the gene product, thereby inhibiting the biologic activity of the specific interaction of the gene product with its cognate ligand or receptor. The UC-related gene antagonist may be of any of these classes as long as it is a substance that inhibits at least one biological activity of the gene product.

[0080] Antagonists include antibodies directed to one or more regions of the gene product protein or fragments thereof, antibodies directed to the cognate ligand or receptor, and partial peptides of the gene product or its cognate ligand which inhibit at least one biological activity of the gene product. Another class of antagonists includes siRNAs, shRNAs, antisense molecules and DNazymes targeting the gene sequence as known in the art are disclosed herein.

[0081] Suitable antibodies include those that compete for binding to UC-related gene products with monoclonal antibodies that block UC-related gene product activation or prevent UC-related gene product binding to its cognate ligand, or prevent UC-related gene product signalling.

[0082] A therapeutic targeting the inducer of the psoriasis-related gene product may provide better chances of success. Gene expression can be modulated in several different ways including by the use of siRNAs, shRNAs, antisense molecules and DNazymes. Synthetic siRNAs, shRNAs, and DNazymes can be designed to specifically target one or more genes and they can easily be delivered to cells in vitro or in vivo.

[0083] The present invention encompasses antisense nucleic acid molecules, i.e., molecules that are complementary to a sense nucleic acid encoding a UC-related gene product polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a

UC-related gene product polypeptide. The non-coding regions (“5' and 3' untranslated regions”) are the 5' and 3' sequences that flank the coding region and are not translated into amino acids.

[0084] The invention also provides chimeric or fusion proteins. As used herein, a “chimeric protein” or “fusion protein” comprises all or part (preferably biologically active) of a UC-related gene product polypeptide operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same UC-related gene product polypeptide). Within the fusion protein, the term “operably linked” is intended to indicate that the UC-related gene product polypeptide and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the amino-terminus or the carboxyl-terminus of the UC-related gene product polypeptide. In another embodiment, a UC-related gene product polypeptide or a domain or active fragment thereof can be fused with a heterologous protein sequence or fragment thereof to form a chimeric protein, where the polypeptides, domains or fragments are not fused end to end but are interposed within the heterologous protein framework.

[0085] In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a UC-related gene product polypeptide is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction in vivo. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a UC-related gene product polypeptide. Inhibition of ligand/receptor interaction can be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (e.g., promoting or inhibiting) cell survival. A preferred embodiment of an immunoglobulin chimeric protein is a C_H1 domain-deleted immunoglobulin or “mimetic body” having an active polypeptide fragment interposed within a modified framework region as taught in co-pending application PCT WO/04002417. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a UC-related gene product polypeptide in a subject, to purify ligands and in screening assays to identify molecules that inhibit the interaction of receptors with ligands.

[0086] Compositions and Their Uses

[0087] In accordance with the invention, the neutralizing anti-UC-related gene product antagonists, such as monoclonal antibodies, described herein can be used to inhibit UC-related gene product activity. Additionally, such antagonists can be used to inhibit the pathogenesis of UC and -related inflammatory diseases amenable to such treatment, which may include, but are not limited to, rheumatic diseases. The individual to be treated may be any mammal and is preferably a primate, a companion animal which is a mammal and most preferably a human patient. The amount of antagonist administered will vary according to the purpose it is being used for and the method of administration.

[0088] The UC-related gene antagonists may be administered by any number of methods that result in an effect in tissue in which pathological activity is desired to be prevented or halted. Further, the anti-UC-related gene product antago-

nists need not be present locally to impart an effect on the UC-related gene product activity, therefore, they may be administered wherever access to body compartments or fluids containing UC-related gene product is achieved. In the case of inflamed, malignant, or otherwise compromised tissues, these methods may include direct application of a formulation containing the antagonists. Such methods include intravenous administration of a liquid composition, transdermal administration of a liquid or solid formulation, oral, topical administration, or interstitial or inter-operative administration. Administration may be affected by the implantation of a device whose primary function may not be as a drug delivery vehicle.

[0089] For antibodies, the preferred dosage is about 0.1 mg/kg to 100 mg/kg of body weight (generally about 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of about 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, the use of lower dosages and less frequent administration is often possible. Modifications, such as lipidation, can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

[0090] The UC-related gene product antagonist nucleic acid molecules can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Pat. No. 5,328,470), or by stereotactic injection (see, e.g., Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[0091] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0092] Pharmacogenomics

[0093] Agents, or modulators that have a stimulatory or inhibitory effect on activity or expression of a UC-related gene product polypeptide as identified by a screening assay described herein, can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a UC-related gene product polypeptide, expression of a UC-related gene product nucleic acid, or mutation con-

tent of a UC-related gene product gene in an individual can be determined to thereby select an appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

[0094] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism." These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[0095] As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

[0096] Thus, the activity of a UC-related gene product polypeptide, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

[0097] Methods of Treatment

[0098] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with

aberrant expression or activity of a UC-related gene product polypeptide and/or in which the UC-related gene product polypeptide is involved.

[0099] The present invention provides a method for modulating or treating at least one UC-related gene product related disease or condition, in a cell, tissue, organ, animal, or patient, as known in the art or as described herein, using at least one UC-related gene product antagonist.

[0100] Compositions of UC-related gene product antagonist may find therapeutic use in the treatment of UC or related conditions, such as Crohn's disease or other gastrointestinal disorders.

[0101] The present invention also provides a method for modulating or treating at least one gastrointestinal, immune related disease, in a cell, tissue, organ, animal, or patient including, but not limited to, at least one of gastric ulcer, inflammatory bowel disease, ulcerative colitis, Crohn's pathology, and the like. See, e.g., the Merck Manual, 12th-17th Editions, Merck & Company, Rahway, N.J. (1972, 1977, 1982, 1987, 1992, 1999), Pharmacotherapy Handbook, Wells et al., eds., Second Edition, Appleton and Lange, Stamford, Conn. (1998, 2000), each entirely incorporated by reference.

[0102] Disorders characterized by aberrant expression or activity of the UC-related gene product polypeptides are further described elsewhere in this disclosure.

1. Prophylactic Methods

[0103] In one aspect, the invention provides a method for at least substantially preventing in a subject, a disease or condition associated with an aberrant expression or activity of a UC-related gene product polypeptide, by administering to the subject an agent that modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease that is caused or contributed to by aberrant expression or activity of a UC-related gene product can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

[0104] Another aspect of the invention pertains to methods of modulating expression or activity of UC-related gene or gene product for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. In another embodiment, the agent inhibits one or more of the biological activities of the UC-related gene or gene product polypeptide. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies and other methods described herein. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g.,

by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a UC-related gene product polypeptide. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulate (e.g., up-regulate or down-regulate) expression or activity. Inhibition of activity is desirable in situations in which activity or expression is abnormally high or up-regulated and/or in which decreased activity is likely to have a beneficial effect. **[0105]** While having described the invention in general terms, the embodiments of the invention will be further disclosed in the following examples which should not be construed as limiting the scope of the claims.

EXAMPLE 1

Sample Analysis by Using Nucleic Acid Microarrays

[0106] Colon Biopsies from Infliximab Treated Ulcerative Colitis Patients

[0107] Sample Collection and RNA Isolation

[0108] Patients with moderate to severe active UC were randomly assigned 1:1:1 to intravenous placebo or infliximab (anti-TNF antibody) at a dose of 5 or 10 mg/kg at 0, 2, 6 and every 8 weeks thereafter. Colonoscopic punch biopsies were obtained from disease tissues at weeks 0 (prior to therapy), 8, and 30 and kept frozen until RNA preparation. RNA isolated from the biopsy samples was subsequently used for Affymetrix (oligonucleotide) microarray analysis. One hundred and twenty-three colon biopsy samples were collected from 49 subjects in this study. Gene expression profiles from 36 infliximab treatment responder samples in both 5 and 10 mg/kg treatment group at both weeks 8 and 30 were compared to that of 13 non-responder samples across both dose groups at both time points as described herein. Treatment responders showed a marked clinical improvement following therapy defined by a decrease from baseline Mayo score by at least 3 points and at least 30% with an accompanying decrease in rectal bleeding sub-score of at least 1 point or an absolute rectal bleeding sub-score of 0 or 1.

[0109] Total RNA was isolated with an RNeasy mini kit according to the manufacturer's instructions (Qiagen Inc., Valencia, Calif.). The colon biopsy samples were lysed and homogenized in the presence of 600 μ L of GITC (guanidine isothiocyanate)-containing buffer, which immediately inactivates RNase to ensure isolation of intact RNA. 600 μ L of 70% ethanol was added to provide appropriate binding conditions and the sample was then applied to an RNeasy mini spin column where the total RNA binds to the membrane and contaminants were efficiently washed away. High-quality RNA was then eluted in 30 μ L of water. RNA quality and quantity was analyzed with 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, Calif.).

[0110] Microarray Data Analysis

[0111] Microarray analysis was performed on GeneChip Human Genome U133 Plus 2.0 arrays that allow the analysis of the expression level of more than 47,000 transcripts and variants, including 38,500 well-characterized human genes. RNA amplification, target synthesis and labeling, chip hybridization, washing and staining were performed in accordance with the manufacture's protocol (Affymetrix, Santa Clara, Calif.). The GeneChips were scanned using the GeneChip Scanner 3000. The data were analyzed with GCOS 1.4

(GeneChip Operating System) using Affymetrix default analysis settings and global scaling as normalization method. The trimmed mean target intensity of each array was arbitrarily set to 500.

[0112] Data quality was assessed by hybridization intensity distribution and Pearson's correlation in Partek Pro software version 6.1 (Partek Inc., St. Charles, Mo.), and was deemed good except for two samples, E36507_P43_5 mg/kg_W30 & E36498_P39_placebo_W8. These samples were regarded as outliers and removed from data analysis.

[0113] Using GeneSpring™ software version 7.2 (Agilent Technologies, Palo Alto, Calif.), the intensity for probe set was normalized across all samples. Each measurement was divided by the median of all measurements in that sample. The intensity of a probe set was then normalized to the median intensity of that probe set in the control group. The control groups in this study were all 45 week 0 samples. Normalized intensity of probe set A in sample X was calculated as follows:

$$\frac{(\text{Signal intensity of probe set A in sample X})}{(\text{Median intensity of all measurements in sample X}) \times (\text{Median intensity of probe set A across all week-0 samples})}$$

[0114] Using Partek Pro 6.1, statistical analysis was done to identify significant treatment effects, and the differences between responders and non-responders, using log 2 transformed normalized intensities. Standard ANOVA was conducted between responders at each treatment condition (5 mg/kg week 8, 5 mg/kg week 30, 10 mg/kg week 8, and 10 mg/kg week 30) vs. the corresponding baseline, and between responders and non-responders under each treatment condition. Subject effect was tested in the mix-model of ANOVA as a random factor. Differences were considered statistically significant at p-value <0.05. Using linear scaled data, genes showing more than 2× significant differential expression for a specific comparison were identified. Only the genes designated Present or Marginally Present at least once among the samples representing the condition with a higher expression level in a comparison were documented.

[0115] Class Prediction Analysis. Classification of infliximab responsiveness for each patient sample was generated with the 'K-Nearest Neighbors' algorithm (Cover TM HP. Nearest neighbor pattern classification. IEEE Transactions on Information Theory 1967; 13:21-27). Week-8 samples comprised the training set and week-30 samples the test set. Fisher's Exact Test was used to select a smaller set of transcripts from the training set yielding the treatment-response-specific class prediction at week 30. Transcripts are scored based on the best prediction for a class. The predictive strength is the negative natural logarithm of the p-value for a hypergeometric test of predicted versus actual class membership for this class versus others. The class prediction analysis led to the 43-gene panel.

[0116] Gene expression signatures between responder and nonresponder samples were compared at week 8. Classification of infliximab responsiveness for each patient sample was generated by the 'K-Nearest Neighbors' algorithm (Cover TM HP. Nearest neighbor pattern classification. IEEE Transactions on Information Theory 1967; 13:21-27), using 27 week-8 samples as the training set (20 responders and 7 nonresponders) to predict infliximab responsiveness of the 22

week-30 samples in the test set (16 responders and 6 nonresponders). A common set of 143 transcripts was identified that passed ANOVA and 2-fold change cut-off in both the 5- and 10-mg/kg dose groups between responders and nonresponders at week 8. Upon subsequent Fisher's Exact Test, the top 50 predictive transcripts (43 genes) were selected to achieve an acceptable predictive accuracy with a minimal number of transcripts (Table 1). Transcripts are scored based on the best prediction for a class. The predictive strength is the negative natural logarithm of the p-value for a hypergeometric test of predicted versus actual class membership for this class versus others. This 43-gene classifier correctly identified 21 patients as determined by clinical outcome measurement and misclassified one nonresponder indicating that this set of transcripts provides 100% sensitivity and 83% specificity for prediction of treatment responsiveness at week 30.

[0117] Differences in gene expression profiles between weeks 8 and 30 were also noted when infliximab 5 and 10 mg/kg treatment responder vs. nonresponder samples were compared. Distinct transcripts were associated with the maintenance therapy up to week 30 that were different from those affected by the induction regimen up to week 8. Among the transcripts unique to week 30, immune response genes, such as IL-17A, were downregulated. IL-17A has been shown to play a key role in autoimmune diseases and animal models of inflammatory diseases, and increased expression has been associated with UC and CD. Also, chemokines that can be induced by IL-17A, e.g., CXCL2, 6, and 8 (IL-8), and chemokines important for neutrophil migration, innate immunity, acute inflammation, and T cell migration/adaptive immunity, including CXCL3, 5, 9, 10, and 11, respectively, were all downregulated in responder samples. Downregulation of matrix remodeling genes, such as matrix metalloproteinases (MMPs) 7, 9, 10, 12, and 19, and tissue inhibitor of metalloproteinase (TIMP1) was also observed.

[0118] To explore differential gene expression profiles for infliximab non-responders in UC at various follow-up time points, gene expression changes were examined in the infliximab nonresponder samples for both dose groups (n=6) at week 30 relative to baseline samples (n=13). The differential expression profiles were then compared with those in the infliximab responder samples (n=10 in the 10 mg/kg group) at week 30 relative to baseline samples (n=17). Among the genes showing unique expression changes in the nonresponder expression profiles, IL-23p19, CCR1, and serum amyloid protein A (SAA) were significantly upregulated by 2.3-, 2.0-, and 2.3-fold, respectively. Conversely, these genes were consistently and significantly downregulated by infliximab in responder samples. Additionally, a parathyroid hormone-like hormone (PTHrP), G-protein coupled receptor 86 (GPR86), and a Ral-GDS-related protein (Rgr) were also significantly upregulated in the nonresponder samples. Expression of other genes that were significantly downregulated by infliximab treatment in the responder samples was not changed significantly in nonresponder samples at weeks 8 and 30 relative to baseline. The combination of the significant and nonsignificant gene expression changes in nonresponder vs. responder samples suggests a unique molecular signature for the infliximab treatment nonresponders.

[0119] Microarray Results

[0120] Biopsies taken from infliximab treatment responders and non-responders at weeks 8 and 30 allowed an understanding of the potential mechanism underlying treatment response and non-response in UC. The post-treatment

responder samples analyzed were taken from patients who showed a marked clinical improvement following infliximab therapy as defined above. The non-responder samples were taken from patients who did not achieve the treatment response as defined above.

[0121] Genes that were expressed at lower levels in the infliximab treatment responders in the response signature can be grouped into 7 main categories based on their functions. The first category consists of genes reported to be involved in immune and inflammatory responses as represented by IL-1 β , IL-1ra, IL-6, IL-8R β , IL-11, IL-13R α 2, IL-23A, IL-24, oncostatin M (OSM), TNF α -inducible protein 6 (TNFAIP6), superoxide dismutase 2, selectin E, selectin L, T-cell activation GTPase (TAGAP), TLR2, and TREM1. The second class consists of genes reported to be involved in cell growth, proliferation, maintenance, apoptosis, cell-cell signaling, and cell adhesion, such as TNFR superfamily member 10c (TNFRSF10c), BCL2A1, BCL6, integrin alpha X (ITGAX), and protocadherin 17. The third class consists of genes reported to be involved in signal transduction, such as WNT5A and prokineticin 2. The fourth class consists of genes reported to be involved in matrix turnover, such as MMP3 and MMP25. The fifth class consists of genes that have been reported to be important for various metabolisms and the transporter genes. The sixth class is composed of genes reported to be involved in cytoskeleton organizations, such as myosin 1F and Kelch-like 5 gene, and the last class consists of genes reported to be involved in hormonal regulations, such as PTH (parathyroid hormone) like hormone. In the response signature, the two genes that were expressed at higher levels in the infliximab treatment responder samples were thyroid hormone receptor beta (THRB) and carboxypeptidase A6 (CPA6).

[0122] The genes disclosed above, not identified in SEQ ID NOS: 1-43, and those identified in SEQ ID NOS: 1-43, individually or in combination, are useful as biomarkers to assess the presence or severity of UC-related diseases or disorders, the response to treatment with a particular therapy (e.g., an anti-TNF antibody, such as infliximab), such as a treatment responder or non-responder, and as therapeutic targets for UC-related diseases or disorders.

[0123] Utility of the Response Signature

[0124] The response signature for infliximab treatment in UC described herein can be assessed and used as described below.

[0125] 1) Archived RNA samples from treatment non-responder samples (5-10) as early as 8 weeks post-treatment are used for subsequent comparison analysis.

[0126] 2) Colonoscopic biopsy samples are obtained from lesional sites of patients with active UC as early as 8 weeks post-treatment. RNA will then be isolated from the biopsy samples and subjected to real time RT-PCR analysis. One microgram of total RNA in the volume of 50 μ l was converted to cDNA in the presence of MultiScribe Reverse Transcriptase. The reaction was carried out by incubating for 10 minutes at 25° C. followed by 30 minutes at 48° C. Reverse Transcriptase was inactivated at 95° C. for 5 minutes. Twenty-five nanograms of cDNA per reaction was used in real time PCR with ABI 7900 system (Foster City, Calif.). In the presence of AmpliTaq Gold DNA polymerase (ABI biosystem, Foster City, Calif.), the reaction was incubated for 2 minutes at 50° C. followed by 10 minutes at 95° C. Then the reaction was run for 40 cycles at 15 seconds, at 95° C. and 1 minute, 60° C. per cycle using primer/probe sets specific for the genes in the response signature. House keeping genes, such as GAPDH or actin, will be used as internal calibrators. The relative change in gene expression is calculated using the delta-delta Ct method described by Applied Biosystems using values in the non-responder samples as the calibrator or comparator.

[0127] 3) If a similar gene expression profile meets the parameters of the gene profile signature, i.e., 43 of the same signature genes showed lower expression with at least 2 fold change in the responder samples as compared with that in the non-responder samples and two genes (THRB and CPA6) showed elevated expression with at least 2 fold change in the responder vs. non-responder samples, then the patient is defined as a treatment responder. In which case, the patient will be kept on therapy.

[0128] 4) If the gene expression profile does not meet the parameters of the gene profile signature, based on the direction of the change in expression level or magnitude of the changes, then the patient is defined as a treatment non-responder. In which case, the patient should discontinue the therapy. This enables a patient to avoid therapy earlier after being deemed a non-responder. This can allow the patient to receive a different type of therapy.

TABLE 1

43 genes (50 transcripts) as predictors of infliximab responsiveness in UC				
GeneBank Accession Number	Name (SEQ ID NO)	Name	Functional categories	Predictive Strength*
NM_006850	IL24 (1)	Interleukin 24	Immune response	11.62
NM_014459	PCDH17 (2)	protocadherin 17	Cell adhesion	10.65
NM_020361	CPA6 (3)	Carboxypeptidase A6	Proteolysis and peptidolysis	10.65
AF010316	PTGES (4)	Prostaglandin E synthase	Signal transduction	10.65
AW469523	DGAT2 (5)	diacylglycerol O-acyltransferase homolog 2 (mouse)	Lipid metabolism	10.65
N39230	LOC389865 (6)	Unknown	Unknown	10.65
BG437034	OSM (7)	Oncostatin M	Immune/	10.11
AI079327			inflammatory response	8.254

TABLE 1-continued

43 genes (50 transcripts) as predictors of infliximab responsiveness in UC				
GeneBank Accession Number	Name (SEQ ID NO)	Name	Functional categories	Predictive Strength*
NM_005795	CALCRL (8)	calcitonin receptor-like	G-protein signaling	10.11
NM_006334	OLFM1 (9)	olfactomedin 1	Development	10.11
R38389				8.254
M83248	SPP1 (10)	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	Immune/inflammatory response	8.909
NM_000759	CSF3 (11)	Colony stimulating factor 3	Defense response	8.909
BF433902	TNFRSF11B (12)	Tumor necrosis factor receptor superfamily, member 11b	Inflammatory response	8.909
AV756141	CSF2RB (13)	Colony stimulating factor 2 receptor, beta, low-affinity	Defense response	8.909
BG494007	THRB (14)	Thyroid hormone receptor, beta	Hormone regulation	8.909
NM_001557	IL8RB (15)	Interleukin 8 receptor beta	Immune/inflammatory response	8.909
W46388	SOD2 (16)	Superoxide dismutase 2, mitochondrial	Inflammatory response	8.909
X15132				8.254
NM_000641	IL11 (17)	Interleukin 11	Immune response	8.909
U90939	FCGR2A (18)	Fc fragment of IgG, low affinity IIa receptor (CD32)	Immune response	8.909
NM_004904	CREB5 (19)	cAMP responsive element binding protein 5	Transcription regulation	8.748
NM_022977	ACSL4 (20)	acyl-CoA synthetase long-chain family member 4	metabolism	8.748
NM_018643	TREM1 (21)	Triggering receptor expressed on myeloid cells 1	Innate immune response	8.254
BC020691	PBEF1 (22)	Pre-B-cell colony enhancing factor 1	Cell-cell signaling	8.254
NM_005746				7.898
BF575514				7.898
AF288391	C1orf24 (23)	unknown	unknown	8.254
D87291	KCNJ15 (24)	potassium inwardly-rectifying channel, subfamily J, member 15	ion transport	8.254
NM_001706	BCL6 (25)	B-cell CLL/lymphoma 6 (zinc finger protein 51)	regulation of cell growth	8.254
AW264036				8.254
AI968085	WNT5A (26)	wingless-type MMTV integration site family, member 5A	signal transduction	8.254
NM_003392				7.898
NM_170776	GPR97 (27)	G protein-coupled receptor 97	G-protein signaling	8.254
J03223	PRG1 (28)	proteoglycan 1, secretory granule	matrix	8.254
NM_000167	GK (29)	glycerol kinase	Carbohydrate metabolism	8.254
NM_006317	BASP1 (30)	brain abundant, membrane attached signal protein 1	Signal transduction	8.254
AA650281	FLJ23153 (31)	unknown	unknown	8.254
AL359062	COL8A1 (32)	collagen, type VIII, alpha 1	Collagen metabolism	8.254

TABLE 1-continued

43 genes (50 transcripts) as predictors of infliximab responsiveness in UC				
GeneBank Accession Number	Name (SEQ ID NO)	Name	Functional categories	Predictive Strength*
AW576600	TAGAP (33)	T-cell activation GTPase activating protein	immune response	7.898
AK002174	KLHL5 (34)	kelch-like 5 (<i>Drosophila</i>)	cytoskeleton organization and biogenesis	7.898
NM_000450	SELE (35)	selectin E (endothelial adhesion molecule 1)	inflammatory response	7.898
NM_002029	FPR1 (36)	formyl peptide receptor-like 1	G-protein signaling	7.898
NM_003841	TNFRSF10C (37)	tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain	apoptosis	7.898
AW665748	Transcribed sequences (38)	unknown	unknown	7.898
X90579	CYP3A5 (39)	cytochrome P450, family 3, subfamily A, polypeptide 5	Enzymes	7.898
AK055340	clone FEBRA2000809 (40)	unknown	unknown	7.898
AL524520	GPR49 (41)	G protein-coupled receptor 49	G-protein signaling	7.898
H16258	FLJ37034 (42)	unknown	unknown	7.898
AF493929	RGS5 (43)	regulator of G- protein signaling 5	G-protein signaling	7.405

*Transcripts are scored based on the best prediction for a class.

[0129] These results are novel findings in that clinical response outcome to infliximab treatment in moderate to severe UC can also be detected at the gene expression levels of a panel of selective genes. Furthermore, the panel of genes encompasses a multitude of pathogenic pathways underlying UC that are impacted by infliximab treatment. These include both innate and adaptive immune response genes, such as CSF receptors, NCF2, TLR2, TREM1 and IL-23A, IL-8R β , IL-11, IL-13R α 2, and IL-24. Various pro-inflammatory cytokines, such as IL-1 β , IL-6, a number of TNFL-inducible genes and TNFRSF members were all significantly down regulated in infliximab responders when compared with non-responder samples. In addition, genes important for regulation of cell growth, proliferation, death and cell-cell signaling and those that affect matrix remodeling also showed differential expression in responder samples vs. non-responders samples. Therefore, a constellation of the expression changes in a panel of genes as represented in Table 1 can constitute a profile that can serve as a biomarker profile indicative of the response of a subject to treatment.

[0130] Real Time PCR (TaqMan) Confirmation:

[0131] In order to confirm the microarray finding by an independent means, Real Time PCR technology was employed. One microgram of total RNA in the volume of 50 μ l was converted to cDNA in the presence of MultiScribe Reverse Transcriptase. The reaction was carried out by incubating for 10 minutes at 25° C. followed by 30 minutes at 48° C. Reverse Transcriptase was inactivated at 95° C. for 5 minutes. Twenty-five nanograms of cDNA per reaction were used in real time PCR with ABI 7900 system (Foster City,

Calif.). In the presence of AmpliTaq Gold DNA polymerase (ABI biosystem, Foster City, Calif.), the reaction was incubated for 2 minutes at 50° C. followed by 10 minutes at 95° C. Then the reaction was run for 40 cycles at 15 seconds, at 95° C. and 1 minute, 60° C. per cycle. The housekeeping gene GAPDH (glyceraldehydes-3-phosphate dehydrogenase) was used to normalize gene expression. The Taqman results on a selected number of genes are consistent with the observation from the microarray analysis.

[0132] The present invention discloses the discovery of a panel of potential molecular biomarkers that is indicative of favorable outcome for the treatment of UC. The panel of identified genes represents a UC-related gene panel, which can be used as a tool to monitor the efficacy of any UC therapeutic, such as infliximab, and provide valuable information that guides dosing regimens.

[0133] A panel of genes identified as UC-related genes herein have demonstrated relevance to UC, IBD, and inflammation. As demonstrated by the present analysis, the panel as a whole provides a fingerprint for gauging the efficacy of a treatment of UC that leads to an improvement in the involvement and severity of disease lesions.

[0134] In summary, a panel of potential molecular biomarkers that is indicative of favorable outcome for the treatment of UC has been identified along with the direction in which they are modulated. This panel of biomarkers is particularly useful in guiding clinical development, as the change in expression of genes in this panel can appear prior to improvement of clinically measurable parameters, such as improvement in microscopic changes of the lesions, can be achieved

and/or detected. Thus, the 43 identified genes represent a UC-related gene panel which can be used as a tool to monitor the efficacy of any UC therapeutic, such as anti-TNF antibody, and provide valuable information that guides dosing regimens.

[0135] A panel of genes identified as UC-related genes herein have demonstrated relevance to UC and Crohn's disease. As demonstrated by the present analysis, the panel as a whole provides a fingerprint for gauging the efficacy of a treatment of UC that leads to an improvement in the involvement and severity of UC in patients. A number of the genes, which are members of the UC-related gene panel, have been previously shown to be aberrantly expressed in UC patient samples. For example, increased levels of IL-11, TREM1, superoxide dismutase, selectins, integrins, and various MMP-shave been associated with UC. Thus, together, monitoring genes in this panel provides a method for evaluating drug candidates and in so far as the modulation of the expression of these genes predicts the clinical outcome of a UC therapy.

[0136] Although illustrated and described above with reference to certain specific embodiments, the present invention is nevertheless not intended to be limited to the details shown. Rather, the present invention is directed to the UC-related genes and gene products. Polynucleotides, antibodies, apparatus, and kits disclosed herein and uses thereof, and methods for controlling the levels of the UC-related biomarker genes, and various modifications may be made in the details within the scope and range of equivalents of the claims and without departing from the spirit of the invention.

REFERENCES

- [0137] 1. Okahara, S., Arimura, Y., Yabana, T., Kobayashi, K., Gotoh, A., Motoya, S., Imamura, A., Endo, T., and Imai, K. 2005. Inflammatory gene signature in ulcerative colitis with cDNA microarray analysis. *Aliment Pharmacol Ther* 21:1091-1097.
- [0138] 2. Hanauer, S. B., Feagan, B. G., Lichtenstein, G. R., Mayer, L. F., Schreiber, S., Colombel, J. F., Rachmilewitz, D., Wolf, D. C., Olson, A., Bao, W., et al. 2002. Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet* 359:1541-1549.
- [0139] 3. Rutgeerts, P., Feagan, B. G., Lichtenstein, G. R., Mayer, L. F., Schreiber, S., Colombel, J. F., Rachmilewitz, D., Wolf, D. C., Olson, A., Bao, W., et al. 2004. Comparison of scheduled and episodic treatment strategies of infliximab in Crohn's disease. *Gastroenterology* 126:402-413.
- [0140] 4. Sands, B. E., Anderson, F. H., Bernstein, C. N., Chey, W. Y., Feagan, B. G., Fedorak, R. N., Kamm, M. A., Korzenik, J. R., Lashner, B. A., Onken, J. E., et al. 2004. Infliximab maintenance therapy for fistulizing Crohn's disease. *N Engl J Med* 350:876-885.
- [0141] 5. Sands, B. E., Blank, M. A., Patel, K., and van Deventer, S. J. 2004. Long-term treatment of rectovaginal fistulas in Crohn's disease: response to infliximab in the ACCENT II Study. *Clin Gastroenterol Hepatol* 2:912-920.
- [0142] 6. Mizoguchi, E., Mizoguchi, A., Takedatsu, H., Cario, E., de Jong, Y. P., Ooi, C. J., Xavier, R. J., Terhorst, C., Podolsky, D. K., and Bhan, A. K. 2002. Role of tumor necrosis factor receptor 2 (TNFR2) in colonic epithelial hyperplasia and chronic intestinal inflammation in mice. *Gastroenterology* 122:134-144.
- [0143] 7. Melgar, S., Yeung, M. M., Bas, A., Forsberg, G., Suhr, O., Oberg, A., Hammarstrom, S., Danielsson, A., and Hammarstrom, M. L. 2003. Over-expression of interleukin 10 in mucosal T cells of patients with active ulcerative colitis. *Clin Exp Immunol* 134:127-137.
- [0144] 8. Leeb, S. N., Vogl, D., Gunckel, M., Kiessling, S., Falk, W., Goke, M., Scholmerich, J., Gelbmann, C. M., and Rogler, G. 2003. Reduced migration of fibroblasts in inflammatory bowel disease: role of inflammatory mediators and focal adhesion kinase. *Gastroenterology* 125:1341-1354.
- [0145] 9. Ten Hove, T., The Olle, F., Berkhout, M., Bruggeman, J. P., Vyth-Dreese, F. A., Slors, J. F., Van Deventer, S. J., and Te Velde, A. A. 2004. Expression of CD45RB functionally distinguishes intestinal T lymphocytes in inflammatory bowel disease. *J Leukoc Biol* 75:1010-1015.
- [0146] 10. Amasheh, S., Barmeyer, C., Koch, C. S., Tavalali, S., Mankertz, J., Epple, H. J., Gehring, M. M., Florian, P., Kroesen, A. J., Zeitz, M., et al. 2004. Cytokine-dependent transcriptional down-regulation of epithelial sodium channel in ulcerative colitis. *Gastroenterology* 126:1711-1720.
- [0147] 11. Murch, S. H., Lamkin, V. A., Savage, M. O., Walker-Smith, J. A., and MacDonald, T. T. 1991. Serum concentrations of tumour necrosis factor alpha in childhood chronic inflammatory bowel disease. *Gut* 32:913-917.
- [0148] 12. Murch, S. H., Braegger, C. P., Walker-Smith, J. A., and MacDonald, T. T. 1993. Location of tumour necrosis factor alpha by immunohistochemistry in chronic inflammatory bowel disease. *Gut* 34:1705-1709.
- [0149] 13. Braegger, C. P., Nicholls, S., Murch, S. H., Stephens, S., and MacDonald, T. T. 1992. Tumour necrosis factor alpha in stool as a marker of intestinal inflammation. *Lancet* 339:89-91.
- [0150] 14. Rutgeerts, P., Sandbom, W. J., Feagan, B. G., Reinisch, W., Olson, A., Johanns, J., Travers, S., Rachmilewitz, D., Hanauer, S. B., Lichtenstein, G. R., et al. 2005. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 353:2462-2476.
- [0151] 15. Pender, S. L., and MacDonald, T. T. 2004. Matrix metalloproteinases and the gut—new roles for old enzymes. *Curr Opin Pharmacol* 4:546-550.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 43

<210> SEQ ID NO 1

<211> LENGTH: 1975

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 1

```

cttgectgca aacctttact totgaaatga cttccacggc tgggacggga accttccacc   60
cacagctatg cctctgattg gtgaatgggtg aagggtgcctg tctaactttt ctgtaaaaag   120
aaccagctgc ctccaggcag ccagccctca agcatcactt acaggaccag agggacaaga   180
catgactgtg atgaggagct gctttcgcca atttaacacc aagaagaatt gaggctgctt   240
gggaggaagg ccaggaggaa cacgagactg agagatgaat tttcaacaga ggctgcaaag   300
cctgtggact ttagccagac ccttctgccc tcctttgctg gcgacagcct ctcaaatgca   360
gatggttgtg ctccttggc tgggttttac cctgcttctc tggagccagg tatcaggggc   420
ccagggccaa gaattccact ttgggccctg ccaagtgaag ggggttgttc cccagaaact   480
gtgggaagcc tctgggctg tgaaagacac tatgcaagct caggataaca tcacgagtgc   540
ccggctgtg cagcaggagg ttctgcagaa cgtctcggat gctgagagct gttaccttgt   600
ccacacctg ctggagtctt acttgaaaac tgttttcaaa aactaccaca atagaacagt   660
tgaagtcagg actctgaagt cattctctac tctggccaac aactttgttc tcatcgtgtc   720
acaactgcaa cccagtcagg aaaatgagat gttttccatc agagacagtg cacacaggcg   780
gtttctgcta ttcgggagag cattcaaaac gttggacgta gaagcagctc tgaccaaagc   840
ccttggggaa gtggacatct ttctgacctg gatgcagaaa ttctacaagc tctgaatgtc   900
tagaccagga cctccctccc cctggcactg gtttgttccc tgtgtcattt caaacagtct   960
cccttctat gctgttcaact ggacacttca cgccttggc catgggtccc attcttggcc  1020
caggattatt gtcaaagaag tcattcttta agcagcgcca gtgacagtca gggaaagggtc  1080
ctctggatgc tgtgaagagt ctacagagaa gattcttcta tttattacaa ctctatttaa  1140
ttaatgtcag tatttcaact gaagttctat ttatttgtga gactgtaagt tacatgaagg  1200
cagcagaata ttgtgcecca tgcttcttta ccctcaciaa tccttgccac agtgtggggc  1260
agtggatggg tgcttagtaa gtacttaata aactgtgggt ctttttttgg cctgtctttg  1320
gattgttaaa aaacagagag ggatgcttgg atgtaaaact gaacttcaga gcatgaaaat  1380
cacactgtct tctgatatct gcaggacag agcattgggg tgggggtaag gtgcactctgt  1440
ttgaaaagta aacgataaaa tgtggattaa agtgcccagc acaaagcaga tcctcaataa  1500
acatttcatt tcccaccac actcgcagc tcacccatc atcccttcc cttggtgccc  1560
tcctttttt tttatcctag tcattcttcc ctaatcttcc acttgagtgt caagctgacc  1620
ttgctgatgg tgacattgca cctggatgta ctatccaatc tgtgatgaca ttcctgcta  1680
ataaaagaca acataactca agtctggcag actttcttct ctatttctgg atgaatgccc  1740
agtgagactg tgtgtacag ctagaaaagg ccttcttccc aatagcaagg ctgtgcatct  1800
agcctcaagc tctggctgaa ctttgtggtc gacatcaatc taaagataca gtgtctgact  1860
ataaccttgt tccaaaaacc taggcaaaga gtatatgtag gaggtgggat atcacttcca  1920
tgacataagt gctattgcag agccgtggcc acccaggaac tctgactgc tttcc  1975

```

<210> SEQ ID NO 2

<211> LENGTH: 4966

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

-continued

gtagatgcag tccgcccgcg ccgctgcctc agccagcaat gcaagattag atctctaaat	60
gcagcaaac actgcctgaa aacagaccgg cccgcgcagc aagcagacat ttcacgggtc	120
gctggggaag cttcaaaaata tatctgtgac tctgtcttcg ttgctcttca tccccatcaa	180
tttcatcacg ggaggcgcgc agcaagtaag aatttcaact tcggatctgc ctagagacac	240
acctccctgc tccctccccc actcgatgtg aagagtatto cggagtctcc gggcgggagt	300
agatttgacg caccctagcg ggagcgcgga aaacctactg attcttttagc tcattatcat	360
ctctcccaga cgagatttcc ttcttatcgc ctgcctcctc gctcaagttt gacccctccc	420
aagtccgggc gggagagacg aaaccccctg ctcaacccca gccgcaggaa gccaccgect	480
tgtcccgaag ccttcgcagct ctgctgcacc gcagcttctc acccagtgcg gatgctgtag	540
atcaacaggt tcagggaact tgagcagaat aaggagagac caccgggtgc cgcagctcgg	600
gtgcagaggg aaaaaaggac ccatagactt gtggctcgcg tcgcgcgcgc acgctgcgcc	660
agggcccag gctggcgcgc actccctctc tggctcctcc agtcagattg ctectgcccc	720
caccttacag gtctgggatg tacctttcca tctgttgctg cttttcttca tgggcccctg	780
ccctcactct caagaacctc aactactcgc tgcgcgagga gcaaggggcc ggcacgggtga	840
tcgggaacat cggcagggat gctcgcactg agcctgggct tccgctgca gagcgcggcg	900
gcgaggggcg cagcaagtgc ggtagctacc ggggtgctgga gaactccgca ccgcacctgc	960
tggacgtgga cgcagacagc gggctcctct acaccaagca gcgcctcgc cgcagatccc	1020
tgtgcccca caatgccaaag tgcagctgt cctcgcaggt gttcgccaac gacaaggaga	1080
tctgatgat caaggtagag atccaggaca tcaacgacaa cgcgccctcc ttctcctcgg	1140
accagatcga aatggacatc tcggagaacg ctgctccggg caccoccttc cccctacca	1200
gcgcacatga ccccgacgcc ggcgagaatg ggctccgcac ctacctgctc acgcgcgacg	1260
atcacggcct ctttgactg gacgttaagt cccgcccgcg cggcaccagg tcccagaac	1320
tggtcatcca gaaggctctg gaccgcgagc aacagaatca ccatcgcctc gtgctgactg	1380
ccctggacgg tggcgcgact ccacgttccg ccaccgtaca gatcaacgtg aagggtattg	1440
actccaacga caacagcccc gtcttcgagg cgcctccta cttgggtgaa ctgcccgaga	1500
acgctccgct ggttacagtg gtcctcgcac tgaacgccac cgacgcgat gaaggtccca	1560
atggtgaagt gctctactct ttcagcagct acgtgcctga ccgcgtgcgg gagctcttct	1620
ccatcgacc caagaccggc ctaatccgtg tgaaggcaca tctggactat gaggaaaacg	1680
ggatgctgga gattgacgtg caggcccag acctggggcc taacctatc ccagcccact	1740
gcaaaagtcac ggtcaagctc atcgaccgca acgacaatgc gccgtccatc ggtttcgtct	1800
ccgtgcgcca gggggcgcgt agcgcggcgc cccctcccgg caccgctatc gccctgggtc	1860
gggtcactga ccgggactct ggcaagaacg gacagctgca gtgtcgggtc ctaggcggag	1920
gagggacggg cggcggcggg ggcctgggcg ggcccggggg ttcgctccc tcaagcttg	1980
aggagaacta cgacaacttc tacacgggtg tgaactgacc cccgctggac cgcgagacac	2040
aagacgagta caacgtgacc atcgtggcgc gggacggggg ctctcctccc ctcaactcca	2100
ccaagtctgt cgcgatcaag attctagacg agaacgacaa cccgctcgg tcaaccaaag	2160
ggctctactg gcttcagggt cagcagaaca acatcccggg agagtacctg ggctctgtgc	2220
tcgcccagga tcccgcctg ggcagaacg gcaccgtatc ctactctatc ctgcccctgc	2280

-continued

acatcggcga	cgtgtctatc	tacaacctatg	tgtctgtgaa	tcccacgaac	ggggccatct	2340
acgccctgcg	ctcctttaac	ttcagacaga	ccaaggcttt	tgagttcaag	gtgcttgcta	2400
aggactcggg	ggcgcccgcg	cacttgagga	gcaacgccac	ggtgaggggtg	acagtgctag	2460
acgtgaatga	caacgcgcca	gtgatcgtgc	tccccacgct	gcagaacgac	accgcggagc	2520
tgcaggtgcc	gcgcaacgct	ggcctgggct	atctgggtgag	cactgtgcgc	gccctagaca	2580
gcgacttcgg	cgagagcggg	cgtctcacct	acgagatcgt	ggacggcaac	gacgaccacc	2640
tgtttgagat	cgaccctgcc	agcggcgaga	tccgcacgct	gcacccttc	tgggaggacg	2700
tgacgccctg	ggtggagctg	tggtggaagg	tgaccgacca	cggcaagcct	accctgtccg	2760
cagtggccaa	gctcatcatc	cgctcgggtg	gcggatccct	tcccaggggg	gtaccacggg	2820
tgaatggcga	gcagcaccac	tgggacatgt	cgctgccgct	catcgtgact	ctgagcacta	2880
tctccatcat	cctcctagcg	gccatgatca	ccatcgccgt	caagtgcaag	cgcgagaaca	2940
aggagatccg	cacttaaca	tgccgcacgc	ccgagtacag	ccaccgcag	ctgggtgggg	3000
gcaagggcaa	gaagaagaag	atcaacaaaa	atgatatcat	gctggtgcag	agcgaagtgg	3060
aggagaggaa	cgccatgaac	gtcatgaacg	tggtgagcag	cccctccctg	gccacctccc	3120
ccatgtactt	cgactaccag	accgcctgc	ccctcagctc	gccccggctg	gaggtgatgt	3180
atctcaaacc	ggcctccaac	aacctgactg	tcctcaggg	gcacgcgggc	tgccacacca	3240
gcttcaccgg	acaagggact	aatgcaagcg	agaccctgc	cactcggatg	tcataattc	3300
agacagacaa	ttttcccgca	gagcccaatt	acatgggcag	caggcagcag	ttgttcaaa	3360
gtatttcagt	agctccacgt	ttaaggaccc	agaaagagcc	agcctgagag	acagtgggca	3420
cggggacagt	gatcaggctg	acagtgaacca	agacactaac	aaaggctcct	gctgtgacat	3480
gtctgttagg	gaggcactca	agatgaaaa	tacttcaact	aaaagccaac	cacttgaaca	3540
agaaccagaa	gagtgtgtta	attgcacaga	tgaatgccga	gtgcttggtc	attctgacag	3600
gtgctggatg	ccacagttcc	ctgcagccaa	tcaggctgaa	aatgcagatt	accgcacaaa	3660
tctctttgta	cctacagttg	aagctaattg	tgagactgag	acttacgaaa	ctgtgaaatcc	3720
cactgggaaa	aagacttttt	gtacatttgg	aaaagacaag	cgagagcaca	ctattctcat	3780
tgccaacggt	aaaccttatt	taaaagccaa	acgtgccctg	agcctctctc	tccaagaggt	3840
cccctcagca	tcaagcagcc	caaccaaggc	gtgcatcgag	ccttgacact	caacaaaagg	3900
ctccctggat	ggctgtgaag	caaaaccagg	agcctggct	gaagcaagca	gtcagtactt	3960
gcccactgac	agtcaatata	tgtcacctag	taagcaacca	agagaccctc	ccttcattggc	4020
ttccgatcag	atggcaaggg	tctttgcaga	tgtgcattcc	agagccagcc	gggattccag	4080
tgagatgggt	gctgttcttg	agcagcttga	ccaccccaac	agggatctgg	gcagagagtc	4140
tgtggatgca	gaggaagttg	tgagagaaat	tgataagctt	ttgcaagact	gccggggaaa	4200
cgaccctgtg	gctgtgagaa	agtgaaaaaa	gaaaaaaaa	aaggcattgg	cattttcttg	4260
tctcttctgt	tgatttaaaa	atgatccctc	ctgggtgataa	cccattttac	agggatgaag	4320
aaagaccaat	gctgctttaa	ggcttttagt	gaacatctga	agtgccca	agtatgttct	4380
ttccactgct	gattttcttt	tcagagataa	caatggtttc	gttttgacca	aacttgtatt	4440
aggacagaat	taatgatgct	taaagagaaa	agaaaaaaag	agagaagaaa	aaggagagat	4500
gaaaaaggag	gatgaggaga	agaattacct	tttgacaatc	tgtaggaag	gtatgcagtg	4560

-continued

```

tgagaactga agtatttctg atcactctca gactgtcctc cgtgatttat gctgacttaa 4620
ctgtttacct ataaacccca tacaagcag ggtcataatt tgtgatctgt ggtggatttc 4680
tagcagtcac cacaggtctc tactgaaagt cctgaaaaga ccttgagta gtccaagcta 4740
caccaaacat taacacatat ttgtggtaaa cttttctgta taaagttacc tgacacacat 4800
ataaacacaa ggaacattcc atatcattag tcgaaaacaa aaacaaaaaa aaaacctttg 4860
gtcatttcta agacatctca tgtcatataa aagttaaatg taaaaagata cagtccattt 4920
tgtcctgcac acacgtagac taattcacgt caaaaaaaaa aaaaaa 4966

```

<210> SEQ ID NO 3

<211> LENGTH: 1906

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

```

cccaagacca agtcgtaata gcaacttccc ttcctcagct gcctgaaactt ttttttccc 60
ttgtagctgg agagaagtgt cacattttgc tcaactctcaa ccttctctgc ccaccccctt 120
cccggagaac ctgtgcgggtg tgtagagggt gctgtgagcc acctccagcc tcgggtggct 180
gcttaagtaa ctttcaactc ctctcttctt aacactatga agtgtctcgg gaagcgcagg 240
ggccaggcag ctgctttcct gcctctttgc tggctctttt tgaagattct gcaaccgggg 300
cacagccacc tttataacaa ccgctatgct ggtgataaag tgataagatt tattcccaaa 360
acagaagagg aagcatatgc actgaagaaa atactctatc aacttaaggt ggacctgtgg 420
cagcccagca gtatctccta tgatcagag ggaacagtta ctgatgtcca tatccccca 480
aatggttccc gagccctgtt agccttctta caggaagcca acatocagta caaggtcctc 540
atagaagatc ttcagaaaac actggagaag ggaagcagct tgcacacca gagaaaccga 600
agatccctct ctggatataa ttatgaagtt tatcactcct tagaagaaat tcaaaattgg 660
atgcatcctc tgaataaaa tcaactcaggc ctcatcaca tgttctctat tggaaagatc 720
tatgagggaa gatgtctttt tattttaaag ctgggcagac gatcagcact caaaagagct 780
gtttggatag actgtgggat tcatgcaaga gaatggattg gtctgcctt ttgtcagtgg 840
tttgtaaaag aagctcttct aacatataag agtgaccag ccatgagaaa aatggtgaat 900
catctatatt tctatatcat gcctgtgttt aacgtcagtg gataccattt tagttggacc 960
aatgatcgat tttggagaaa aacaaggtca aggaactcaa ggtttcctg ccgtggagtg 1020
gatgccaata gaaactggaa agtgaagtgg tgtgatgaag gagcttctat gcacccttgt 1080
gatgacacat actgtggccc ttttccagaa tctgagccgg aagtgaaggc tgtagctaac 1140
ttccttcgaa aacacagaaa gcacattagg gcttatctct cctttcatgc atagctcag 1200
atgttactgt atcccattc ttacaaatat gcaacaattc ccaattttag atgtgtggaa 1260
tctgcagctt ataaagctgt gaatgcactt cagtcagtat acggggtagc atacagatat 1320
ggaccagcct ccacaacgtt gtatgtgagc tctggtagct caatggattg ggctacaaa 1380
aatggaatac cttatgcatt tgctttcgaa ctacgtgaca ctggatattt tggattttta 1440
ctcccagaga tgctcatcaa acccacctgt acagaaacta tgctggctgt gaaaaatctc 1500
acaatgcacc tgctaagaa atgtccctga gacagcccaa ggctcaggtc aactgccata 1560
ggattctgag caaggcctac ttggccctgg atagaaattg ttttcaaaga gaagggcagc 1620

```

-continued

```

tgcttagagt gaacatgtct atggacttta aaaagacccc acgcaatttg actttgtggc 1680
aatagaaaac agtaaaaaac agggcatagc ctagtttgtt ataagaaaa gcatccattt 1740
tctatccttt tagagtctta tttgattatg gtgggagggg atgttttcaa atttccatt 1800
tctcaagaaa tgttcatatt aattgaggat ttcccttcaa taaatctcat gtcctcaatt 1860
ataaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa aaaaaa 1906

```

```

<210> SEQ ID NO 4
<211> LENGTH: 291
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 4

```

```

ggctttgggt tctgtatgg tggaaactgg gtgagccaag gacagggctg gtcctctctg 60
ccccgtgac gtttcccttg ccgttggtt tggatgtctt tgctgcagtc ttctctctgg 120
ctcaggtgtg ggtgggaggg gccacagga agctcagcct tctctctcca aggtttgagt 180
ccctccaaag ggcagtgggt ggaggaccgg gagctttggg tgaccagcca ctcaaaggaa 240
ctttctggtc cttcagtat cttcaaggtt tggaaactgc aaatgctccc t 291

```

```

<210> SEQ ID NO 5
<211> LENGTH: 515
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (33) (34) (41) (431) (444)
<223> OTHER INFORMATION: Wherein n can be either a, c, t, or g

```

```

<400> SEQUENCE: 5

```

```

gcctggcaag aatgcagtca ccctgcggaa ccnaagggc ntttgtgaaa ctggcctctg 60
gtcatggagc tgacctggtt cccatctact cttttggaga gaatgaagtg tacaagcagg 120
tgatcttoga ggagggtctc tggggccgat ggggtccagaa gaagtccag aaatacattg 180
gtttgcctcc atgcatcttc catggtcgag gcctcttctc ctccgacacc tgggggctgg 240
tgcctactc caagccctc accactgttg tgggagagcc catcaccatc cccaagctgg 300
agcaccacac ccagcaagac atcgacctgt accacaccat gtacatggag gcctgtgtga 360
agctcttoga caagcacaag accaagttcg gcctcccgga gactgaggtc ctggaggtga 420
actgagccag ncttcggggc caantccctg gaggaaccag ctgcaaatca cttttttgct 480
ctgtaaatth ggaagtgtca tgggtgtctg tgggt 515

```

```

<210> SEQ ID NO 6
<211> LENGTH: 566
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (42) (67) (70) (102) (106) (110) (146) (149) (165) (178) (194)
(197) (206) (239) (241) (248) (262) (268) (272) (273) (274)
<223> OTHER INFORMATION: Wherein n can be either a, c, t, or g

```

```

<400> SEQUENCE: 6

```

```

gatgggggcg caatagtctt gaacattgta taaagtgtcc angaatggaa gtgctctttg 60
attcatnatn atttttcttc ttcattatcc cctcccagag tntconatcn taggacatca 120
gcattctcac acaagcctaa tggctnatnc tgagtaagca gggentagaa attcactntt 180

```

-continued

```

cttgatactc agtnctngcc ttctanacac tccttgatct tgcctacctc tcccctttnc 240
nacatgtntt ttccctgtagg ancacttntc cnnnttattc ctgcctatcc aattcttccc 300
tatatttccct ggaccagcta aagtccagtg ttccagaga cttttgaaag tcaacttaca 360
ctttttcctt ctccattcac aaagctcttc tccctgggc cctgggatgt atgcctttct 420
ctcctactgt ctaatagcac ctcgtaaatt gtcaatgaac ttttctaagg ggtattcttg 480
aattcccaac tagattgtga gcttctggaa gacaaggcta tgtctttgat tgttgtctcc 540
cctaccacag cccagtactt tagtta 566

```

```

<210> SEQ ID NO 7
<211> LENGTH: 601
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (354)
<223> OTHER INFORMATION: Wherein n can be a, c, t, or g

```

```

<400> SEQUENCE: 7
ccccttgccg gtgaacggat ggggcagggt ctctgtggat gagaggaacc atcgcaggat 60
gacagctccc gggtccecaa agctgttccc ctcttgctac tagccactga gaagtgcact 120
ttaagagggt ggagctgggc agaccctctc acctctcca ggctgggaga cagagtcagg 180
ctgttgccgc tcccactca gccccaagtt ccccaggccc agtgagggtg gccgggcatg 240
ggccacgagg gaaccgactt tccattgatt cagggcgtct gatgacacag gctgagttag 300
aggctgtaca aggccccac tgctgtcgg ttgcttggat tccctgacgt aagntggata 360
ttaaaaatct gtaaatcagg acaggtgggt caaatggcgc tgggaggtgt acacggaggt 420
ctctgtaaaa gcagaccac ctcccagcgc cgggaagccc gtcttgggtc ctgctgctg 480
gctgctcccc ctgggtgggt atcttgaat tttctcacgc aggagccatt gctctctag 540
aggggggtctc agaaactgcg agggcagttc cttggaggga catgactaat ttatcgattt 600
t 601

```

```

<210> SEQ ID NO 8
<211> LENGTH: 5006
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 8
gctgctgac acttacaatc tgacaacact tacaatctac tcagaacaac ctctctctct 60
ccagcagaga gtgtcacctc ctgctttagg accatcaagc tctgctaact gaatctcatc 120
ctaattgcag gatcacattg caaagcttcc actcttccc accttgcttg tgggtaaatc 180
tcttctcggg aatctcagaa agtaaagttc catcctgaga atatttcaca aagaatttcc 240
ttaagagctg gactgggtct tgaccctga atttaagaaa ttcttaaga caatgtcaaa 300
tatgatccaa gagaaaatgt gatttgagac tggagacaat tgtgcatatc gtctaataat 360
aaaaacccat actagcctat agaaaacaat atttgaaaga ttgctaccac taaaagaaa 420
actactacaa cttgacaaga ctgctgcaaa cttcaatttg tcaaccacaa cttgacaagg 480
ttgctataaa acaagattgc tacaacttct agtttatgtt atacagcata tttcattttg 540
gcttaatgat ggagaaaaag tgtaccctgt atttctgggt tctcttgect tttttatgat 600

```

-continued

ttcttgttac agcagaatta gaagagagtc ctgaggactc aattcagttg ggagttacta	660
gaaataaaat catgacagct caatatgaat gttaccacaaa gattatgcaa gaccccatc	720
aacaagcaga aggcgtttac tgcaacagaa cctgggatgg atggctctgc tggaaacgatg	780
ttgcagcagg aactgaatca atgcagctct gccctgatta ctttcaggac tttgatccat	840
cagaaaaagt tacaagatc tgtgaccaag atggaaactg gtttagacat ccagcaagca	900
acagaacatg gacaaattat acccagtgta atgttaacac ccacgagaaa gtgaagactg	960
cactaaatth gttttacctg accataattg gacacggatt gtctattgca tcaactgctta	1020
tctcgcttgg catattcttt tatttcaaga gcctaagttg ccaaaggatt accttacaca	1080
aaaatctgtt cttctcattt gtttgtaact ctgttgtaac aatcattcac ctcaactgcag	1140
tggccaacaa ccaggcctta gtagccacaa atcctgttag ttgcaaagtg tcccagttca	1200
ttcatcttta cctgatgggc tgtaattact tttggatgct ctgtgaaggc atttacctac	1260
acacactcat tgtggtggcc gtgtttgcag agaagcaaca tttaatgtgg tattattttc	1320
ttggctgggg atttccactg attcctgctt gtatacatgc cattgctaga agcttatatt	1380
acaatgacaa ttgctggatc agttctgata cccatctcct ctacattatc catggcccaa	1440
tttgtgctgc tttactgggtg aatctttttt tcttgttaaa tattgtacgc gttctcatca	1500
ccaagttaaa agttacacac caagcggaaat ccaatctgta catgaaagct gtgagagcta	1560
ctcttatctt ggtgccattg cttggcattg aatttgtgct gattccatgg cgacctgaag	1620
gaaagattgc agaggaggta tatgactaca tcatgcacat ccttatgcac ttccagggtc	1680
ttttggtctc taccattttc tgettcttta atggagaggt tcaagcaatt ctgagaagaa	1740
actggaatca atacaaaatc caatttggaa acagcttttc caactcagaa gctcttcgta	1800
gtgcgtctta cacagtgtca acaatcagtg atggctcagg ttatagtcac gactgtccta	1860
gtgaacactt aaatggaaaa agcatccatg atattgaaaa tgttctctta aaaccagaaa	1920
atztatataa ttgaaaatag aaggatgggt gtctcactgt tttgtgcttc tccaaactca	1980
aggacttggg cccatgactc tgtagccaga agacttcaat attaaatgac tttttgaatg	2040
tcataaagaa gagccttcac atgaaattag tagtgtgttg ataagagtgt aacatccagc	2100
tctatgtggg aaaaaagaaa tccctggttg taatgtttgt cagtaaatac tcccactatg	2160
cctgatgtga cgtactaac ctgacatcac caagtgtgga attggagaaa agcacaatca	2220
acttttctga gctggtgtaa gccagttcca gcacaccatt gcatgaattc acaacaaat	2280
ggctgtaaaa ctaaacatac atgttgggca tgattctacc cttattgcc caagagacct	2340
agctaaggtc tataaacatg aagggaaaaat tagcttttag ttttaaaact ctttatccca	2400
tcttgattgg ggcagttgac ttttttttg cccagagtgc cgtagtcctt tttgtaacta	2460
ccctctcaaa tggacaatac cagaagtgaa ttatccctgc tggtttctt ttctctatga	2520
aaagcaactg agtacaattg ttatgatcta ctcatctgct gacacatcag ttatatcttg	2580
tggcatatcc attgtggaat ctggatgaac aggatgtata atatgcaatc ctacttctat	2640
atcattagga aaacatctta gttgatgcta caaaacacct tgtcaacctc ttctgtctt	2700
accaaacagt gggagggaat tccatagctg aaatataaat tttgtccctt ccatttctac	2760
tgtataaaca aattagcaat cattttatat aaagaaaatc aatgaaggat ttcttatttt	2820
cttgaatth tgtaaaaaga aattgtgaaa aatgagcttg taaatactcc attattttat	2880

-continued

```

tttatagtct caaatcaaat acatacaacc tatgtaattt ttaaagcaaa tatataatgc 2940
aacaatgtgt gtatgttaat atctgatact gtatctgggc tgatttttta aataaaatag 3000
agtctggaat gctatatttg gtaaatattt taaagacaac cagatgccag catcagaagt 3060
ctgtttgaga actaagagaa cagaacatc tatcataaga tatatttatt ttaaaaacac 3120
aaggcacta ttttattgaa tatatttggt ttgataactc ataccttaat aatagggtgtg 3180
tttgacatat tcttttttgc attttgacaa tgaactcaca ttctaacca gaaattttaa 3240
acaactactg tgataaatac caatctgcta cttttataga ttttaccaca ttaaaatatt 3300
actttactga cttttactat gtgaagatat atagctttgg aaatgtccca ggctattcaa 3360
gaaatataaa aaactagaag gatactatat ataccatata caatgcttta atattttaat 3420
agagctactg tatataatac aaattaggga aatacttgaa tatatcattg agaaaaaatt 3480
attgtcagat cttactgaat tattgtcaga ctttattaaa taaagataga agaaaacctt 3540
gctaatgaat taaagtgaat tttgcatggg attcagtttc tctaagtta tttccgctg 3600
aaatctctaa agaacaagaa tgacttcaat tagtaaaagt caattttggg aaaagtcattg 3660
ggtatctggt ttttaagtgt gtcaatctga ttaaaatgga tgaacaaat tactcatcat 3720
aagttgttgc ttaagctgtc aatagtcaa tagatgggta gttcagaact tatttcaaat 3780
tgctaagaca aattatctaa atctgtaaga attaacatat agaatggctc ggtcagtaca 3840
tttataattt atctatgcat gaaaagtat tgttttggtt gaaacatgaa tttcatagca 3900
agctgccata gaaaggaacg caggctgttc tagaccttca actgocataa ttatacaaaa 3960
attcatttta ataaactcaa ttattagcta tttattatc aaagacctat atttaaatcc 4020
tttgctgacc atgttgacat atatcagcct tcttctagac aaactgtcaa ctctcaacca 4080
tcttgacagt agaagtgaca gtaaaaaatg ttgaatgatc agagattata ttaaaataaa 4140
catgtaattt tcaagtattt ttgttggtct tttataatat taattctaga tcagatttat 4200
tttatagcca gggtttgtct gttgttaggt cttgaggcgt agcagtcatt catgattaat 4260
cactgttagt tttgtaccca tatattttta gaatagtttt aaatgttaga tttctcaaaa 4320
gctaaatgct acttaatatc tttgtatcat actcataaag caaagtaaat ctgacacttt 4380
ttttaagca aacttctttg ctgtcaaaaa aataaatttg gggaaatttc tagcttttaa 4440
aatgtagatc tgcattttac tgtgattact tgtgaaagtc atattttaat tttctaaatt 4500
ctaattgtc attttatctc ctaaagttaa tttccaatgc atttattcat aaaatattca 4560
ttctggaatg cagtgtttgt ttaaagttaa tccaatgat atagaattag tgggtggctgt 4620
agtgtgtat ttattgctta taattttttt taaatgtgaa cttactttta attttctctt 4680
ggttttaatc tgctagtaga aaccactagt tatctgtaaa aatatttca agatattctg 4740
atcaattata acaatttatg ttatgcctag agtatatctc tattttttga ttgtatgaaa 4800
atattaaagt tatgagttaa agtttatttt cactgatatt tactacagtg ccaataatc 4860
taatttataa acataattct tacagtaatc aatgggatac ttctcaaaa taacaaatct 4920
cttaacaaaa tatatctttt gccctcttta aagtcttcag taaaccagta aatgaattca 4980
ataaaccaat taagaaaaaa aaaaaa 5006

```

<210> SEQ ID NO 9

<211> LENGTH: 1212

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 9

tgcttccccg ccggcccgcc ccctcctccc cgggagagag cgaggcgcgc gggtcctct 60
gcgccacccc cgccccgcc ccttcgagc aaacttttg caccacccgc agcccagcgc 120
gcgttcgtgc tccgcagggc gcgcctctct ccgccaatgc caggcgcgcg ggggagccat 180
taggaggcga ggagagagga gggcgcagct ccgcccagc ccagccctgc ccagccctgc 240
ccggaggcag acgcgcggga accgggacgc gataaatatg cagagcggag gcttcgcgca 300
gcagagcccc cgcgccgccc gctccgggtg ctgaatccag gcgtggggac acgagccagg 360
cgccgccccc ggagccagcg gagccggggc cagagccgga gcgcgtccgc gtcccagcag 420
ccgcggccgc gccagcacc agggccctgc atgccaggtc gttggaggtg gcagcgcgac 480
atgcaccgag ccgggaagct cctcagcctc ctcttctca tcctgatggg cactgaactc 540
actcaagtgc tgcccaccaa cctgaggag agctggcagg tgtacagctc tgcccaggac 600
agcgagggca ggtgtatctg cacagtggc gctccacagc agaccatgtg ttcacgggat 660
gccccacaaa aacagctgag gcagctactg gagaagggtc agaacatgc tcaatccata 720
gaggctcttg acaggcggac ccagagagac ttgcagtacg tggagaagat ggagaaccaa 780
atgaaaggac tggagtccaa gttcaaacag gtggaggaga gtcataagca acacctggcc 840
aggcagttta agggctaact taaaagagtt ttttcaatgc tgcagtgact gaagaagcag 900
tccactccca tgtaaccatg aaagagagcc agagagcttt ttgcaccatg catttttact 960
attattttcc aatacttagc accatttcc taaggaacct tgaatacaac caggatcctc 1020
ctttgcatgc gactgtagct gcatttcatg aatagtttga acccttgta atgcattttt 1080
tgaaaaagaa agaaaaaaaa aacttcgtgt atgtgactca aagcatgtaa ccttaagatg 1140
ttgcattcta aactgacaat aaagacctt cccaaataaa aaaaaaaaaa aaaaaaaaaa 1200
aaaaaaaaaa aa 1212

```

```

<210> SEQ ID NO 10
<211> LENGTH: 1278
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10

gcaggaggag gcagagcaca gcatcgctcg gaccagactc gtctcaggcc agttgcagcc 60
ttctcagcca aacgccgacc aaggaaaact cactaccatg agaattgcag tgatttgctt 120
ttgcctccta ggcacacact gtgccatacc agttaaacag gctgattctg gaagttctga 180
ggaaaagcag ctttacaaca aatacccaga tgctgtggcc acatggctaa accctgacct 240
atctcagaag cagaatctcc tagcccaca gaatgctgtg tcctotgaag aaaccaatga 300
ctttaacaaa gagaccctcc caagtaagtc caacgaaagc catgaccaca tggatgatata 360
ggatgatgaa gatgatgacg accatgtgga cagccaggac tccattgact cgaacgactc 420
tgatgatgta gatgacactg atgattctca ccagtctgat gagtctcacc attctgatga 480
atctgatgaa ctggtcactg attttccac ggacctgcca gcaaccgaag ttttactctc 540
agttgtcccc acagtagaca catatgatgg ccgaggtgat agtgtggttt atggactgag 600
gtcaaatctc aagaagtctc gcagacctga catccagtac cctgatgcta cagacgagga 660

```

-continued

```

catcacctca cacatggaaa gcgaggagtt gaatggtgca tacaaggcca tccccgttgc 720
ccaggacctg aacgcgcctt ctgattggga cagccgtggg aaggacagtt atgaaacgag 780
tcagctggat gaccagagtg ctgaaaccca cagccacaag cagtccagat tatataagcg 840
gaaagctaata gatgagagca atgagcattc cgatgtgatt gatagtcagg aactttccaa 900
agtcagccgt gaattccaca gccatgaatt tcacagccat gaagatatgc tggttgtaga 960
ccccaaaagt aaggaagaag ataaacacct gaaatttcgt atttctcatg aattagatag 1020
tgcattctct gaggtcaatt aaaaggagaa aaaatacaat ttctcacttt gcatttagtc 1080
aaaagaaaa atgctttata gcaaaatgaa agagaacatg aaatgcttct ttctcagttt 1140
attggttgaa tgtgtatcta tttgagctcg gaaataactg atgtgtttga taattagttt 1200
agtttggtgc ttcattgaaa ctccctgtaa actaaaagct tcagggttat gtctatgttc 1260
attctataga agaaatgc 1278

```

```

<210> SEQ ID NO 11
<211> LENGTH: 1518
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 11

```

```

aaaaacagccc ggagcctgca gccagcccc acccagacc atggctggac ctgccacca 60
gagccccatg aagctgatgg ccctgcagct gctgctgtgg cacagtgcac tctggacagt 120
gcaggaagcc acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg 180
cttagagcaa gtgaggaaga tccagggcga tggcgcagcg ctccaggaga agctggtgag 240
tgagtgtgcc acctacaagc tgtgcccccc cgaggagctg gtgctgctcg gacactctct 300
gggcattccc tgggctcccc tgagcagctg ccccagccag gccctgcagc tggcaggtg 360
cttgagccaa ctccatagcg gccctttcct ctaccagggg ctccctgcagg ccctggaagg 420
gatctcccc gagttgggtc ccaccttggc cacactgcag ctggacgtcg ccgactttgc 480
caccaccatc tggcagcaga tggagaact gggaaatggcc cctgccctgc agcccacca 540
gggtgccatg ccggccttgc cctctgcttt ccagcgcagg gcaggagggg tctggttgc 600
ctcccatctg cagagcttcc tggaggtgct gtaccgcgtt ctacgccacc ttgccagcc 660
ctgagccaag cctccccat cccatgtatt tatctctatt taatatttat gtctatttaa 720
gcctcatatt taaagacagg gaagagcaga acggagcccc aggcctctgt gtccttcct 780
gcattttga gtttcattct cctgcctgta gcagtgagaa aaagctctg tcctccatc 840
ccctggactg ggaggtagat aggtaaatac caagtattta ttactatgac tgcctcccag 900
ccctggctct gcaatgggca ctgggatgag ccgctgtgag cccctggtcc tgaggggtccc 960
cacctgggac ccttgagagt atcaggtctc ccacgtggga gacaagaaat cctgtttaa 1020
tatttaaaca gcagtgttcc ccatctgggt ccttgcaccc ctcaactctgg cctcagccga 1080
ctgcacagcg gccctgcat ccccttggct gtgaggcccc tggacaagca gaggtggcca 1140
gagctgggag gcattgccct ggggtcccac gaatttctg gggaaatctg tttttctct 1200
taagactttt gggacatggt ttgactcccc aacatcaccg acgtgtctcc tgtttttctg 1260
ggtggcctcg ggacacctgc cctgccccca cgagggtcag gactgtgact ctttttaggg 1320
ccaggcaggt gcctggacat ttgccttctg ggacggggac tggggatgtg ggagggagca 1380

```

-continued

```

gacaggagga atcatgtcag gcctgtgtgt gaaaggaagc tccactgtca ccccccacct 1440
cttcaccccc cactcaccag tgtcccctcc actgtcacat tgtaactgaa cttcaggata 1500
ataaagtgtt tgctccca 1518

```

```

<210> SEQ ID NO 12
<211> LENGTH: 380
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 12
tactgcttgc agtaattcaa ctggaaatta aaaaaaaaaa actagactcc attgtgcctt 60
actaaatag ggaatgtcta acttaaataag ctttgagatt tcagctatgc tagaggcttt 120
tattagaaag ccatatTTTT ttctgtaaaa gttactaata tatctgtaac actattacag 180
tattgctatt tatattcatt cagatataag atttgtacat attatcatcc tataaagaaa 240
cggtatgact taattttaga aagaaaatta tattctgttt attatgacaa atgaaagaga 300
aaatatatat ttttaatgga aagttttag catttttcta ataggtagctg ccatatTTTT 360
ctgtgtggag tatttttata 380

```

```

<210> SEQ ID NO 13
<211> LENGTH: 468
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 13
gatttacaaa ggtcctccca ttgcaaagca gtgtttgtcc taatttatat attgtttttc 60
tagttcattt tgtgtttcca acttttcatg taaaatttta attatttttg aatgtgtgga 120
tgtgagactg aggtgccttt tggtagctgaa attctttttc catgtacctg aagtgttact 180
tttvtgatat aggaaatcct tgtatatata ctttatttgg ccttaggett cctattttgt 240
taccttgcct tctctatggc atccaccatt ttgattgttc tacttttatg atatgttttc 300
ataagtgggt aagcaagtat tctcgttact tttgctctta aatccctatt cattacagca 360
atgttvtgtg tcaaagaaaa tgataaaca cttgaatgtt caatgttctt gaaatacata 420
acaacatttt agtacattgt aaagtagaat cctctgttca taatgaac 468

```

```

<210> SEQ ID NO 14
<211> LENGTH: 398
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (200) (245)
<223> OTHER INFORMATION: Wherein n can be a, c, t, or g

```

```

<400> SEQUENCE: 14
ggaactctaa cctattcgtg tcatattgac cttttgctgc atgagtcata aattatgaaa 60
tcagtcttac agtttttgaa atgtagccag cttttgtaag gctaaacctt tttcatgaac 120
tgaatttaag tgaataacca agccacagtt cctcctcaaa tggagagtga tgatcgacat 180
ttgaatctct ttgccctttn ccaacggcta tggcatcagg ttctaaaata agctcgtaat 240
ttttncctgt tattttaata atatggaaat attagcatag tgtttctttt gatagtgata 300
gactataatc catatttaaa ttttatagag aagaaatTTT attgtactgt gatgtagata 360

```

-continued

```

tttattatcc aggtaaggat ttgcccgggtg tgtatttt      398

<210> SEQ ID NO 15
<211> LENGTH: 2859
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 15
cattcagaga cagaagggtg atagacaaat ctccaccttc agactggtag gctcctccag      60
aagccatcag acaggaagat gtgaaaatcc ccagcactca tcccagaatc actaagtggc      120
acctgtcctg ggccaaagtc ccaggacaga cctcattggt cctctgtggg aatacctccc      180
caggagggca tcttggattt cccccttgca acccaggtea gaagtttcat cgtcaagggt      240
gtttcatctt ttttttctg tctaacagct ctgactacca cccaaccttg aggcacagtg      300
aagacatcgg tggccactcc aataacagca ggtcacagct gctcttctgg aggtgtccta      360
caggtgaaaa gccacagcac ccagtcagga ttaagtta cctcaaaaat ggaagatttt      420
aacatggaga gtgacagctt tgaagatttc tggaaagggt aagatcttag taattacagt      480
tacagctcta cctgccccct ttttctacta gatgccgccc catgtgaacc agaatccctg      540
gaaatcaaca agtatcttgg ggtcattatc tatgcctcgg tattcctgct gaggcctgctg      600
ggaaactccc tctgtgatgt ggtcatctta tacagcaggg tcggccgctc cgtcactgat      660
gtctacctgc tgaacctagc cttggccgac ctactctttg cctgacctt gcccatctgg      720
gccgcctoca agtgtaatgg ctggattttt ggcaattccc tgtgcaagggt ggtctcactc      780
ctgaaggaag tcaacttcta tagtggcatc ctgctactgg cctgcatcag tgtggaccgt      840
tacctggcca ttgtccatgc cacacgcaca ctgaccaga agcgctactt ggtcaaatc      900
atatgtctca gcactctggg tctgtccttg ctctctggcc tgcctgtctt acttttccga      960
aggaccgtct actcatccaa tggtagccca gcctgctatg aggacatggg caacaataca     1020
gcaaaactggc ggatgctggt acggatcctg ccccagtcct ttggcttcat cgtgccactg     1080
ctgatcatgc tgttctgcta cggattcacc ctgcgtacgc tgtttaaggc ccacatgggg     1140
cagaagcacc gggccatgag ggteatcttt gctgtctgct tcatcttctt gctctgctgg     1200
ctgccctaca acctggctct gctggcagac acctcatga ggaccaggt gatccaggag     1260
acctgtgagc gccgcaatca catcgaccgg gctctggatg ccaccgagat tctgggcatc     1320
cttcacagct gcctcaaccc cctcatctac gccttcattg gccagaagtt tcgccatgga     1380
ctcctcaaga ttctagctat acatggcttg atcagcaagg actccctgcc caaagacagc     1440
aggccttctt ttgttggctc ttcttcaggc cacacttcca ctactctcta agacctcctg     1500
cctaagtgca gccccgtggg gttcctccct tctcttcaca gtcacattcc aagcctcatg     1560
tccactgggt cttcttggtc tcagtgtaa tgcagcccc attgtggtea caggaagtag     1620
aggaggccac gttcttacta gtttcccttg catggtttag aaagcttgcc ctggtgcctc     1680
acctcttgcc ataattacta tgtcatttgc tggagctctg cccatcctgc cctgagcccc     1740
atggcactct atgttctaag aagtgaaaa ctacactcca gtgagacagc tctgcatact     1800
cattaggatg gctagtatca aaagaaagaa aatcaggctg gccaacgggg tgaaccctg     1860
tctctactaa aaatacaaaa aaaaaaaaaa attagccggg cgtgggtggg agtgctgta     1920
atcacageta cttgggaggc tgagatggga gaatcacttg aaccggggag gcagagggtg     1980

```

-continued

```

cagtgagccg agattgtgcc cctgcaactcc agcctgagcg acagtgagac tctgtctcag 2040
tccatgaaga tgtagaggag aaactggaac tctcgagcgt tgctgggggg gattgtaaaa 2100
tggtgtgacc actgcagaag acagtatggc agctttcctc aaaacttcag acatagaatt 2160
aacacatgat cctgcaatc cacttatagg aattgaccca caagaaatga aagcagggac 2220
ttgaacccat atttgtacac caatattcat agcagcttat tcacaagacc caaaaggcag 2280
aagcaacca aatgttcac aatgaatgaa tgaatggcta agcaaatgt gatatgtacc 2340
taacgaagta tccttcagcc tgaagagga atgaagtact catacatggt acaacacgga 2400
cgaacctga aaactttatg ctaagtgaat taagccagac atcaacagat aaatagttta 2460
tgattccacc tacatgaggt actgagagtg aacaaattha cagagacaga aagcagaaca 2520
gtgattacca gggactgagc ggaggggagc atgggaagt acgggttaat gggcacaggc 2580
tttatgttta gtagttgaa aaagtctgc agataaacag tagtgatagt tgtaccgcaa 2640
tgtgacttaa tgccactaaa ttgacactta aaaatggttt aaatggtaaa ttttggttatg 2700
tatatthtat atcaatthta aaaaaaacct gagccccaaa aggtatthta atcaccaagg 2760
ctgattaaac caaggctaga accactgcc tatatthttt gttaatgat ttcattcaat 2820
atctthtttt taataaacca tthttacttg ggtgtttat 2859

```

```

<210> SEQ ID NO 16
<211> LENGTH: 796
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (147) (154) (155) (159) (172) (228)
<223> OTHER INFORMATION: Wherein n can be a, c, t, or g

```

```

<400> SEQUENCE: 16
acctcagccc taacggtggt ggagaaccca aaggggagtt gctggaagcc atcaaacgtg 60
actttggttc ctttgacaag ttttaaggaga agctgacggc tgcactctgt ggtgtccaag 120
gctcaggttg gggttggcct ggtttcnaat aagnnaacng gggacactta cnaaattgct 180
gcttgtccaa atcaggatcc actgcaagga acaacaggcc ttattccnac tgetggggat 240
tgatgtgtgg gagcacgctt actacctca gtataaaat gtcaggcctg attatctaaa 300
agctatthtg aatgtaatca actgggagaa tgtaactgaa agatacatgg cttgcaaaaa 360
gtaaacaccg atcgttatgc tgagttggct tggtttcaat aaggaacggg gacacttaca 420
aattgctgct tgtccaaatc aggatccact gcaaggaaca acaggcetta tteactgct 480
ggggattgat gttgaggagc acgcttacta ccttcagtat aaaaatgtca ggctgatta 540
tctaaaagct atthggaatg taatcaactg ggagaatgta actgaaagat acatggcttg 600
caaaaagtaa accacgatcg ttatgctgag tatgttaagc tctttatgac tgtttttgta 660
gtggtataga gtactgcaga atacagtaag ctgctctatt gtagcatttc ttgatgttgc 720
ttagtcaact atttcataaa caacttaatg ttctgaataa tttcttacta aacatthtgt 780
tattgggcaa gtgatt 796

```

```

<210> SEQ ID NO 17
<211> LENGTH: 2354
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

-continued

<400> SEQUENCE: 17

```

gctcagggca catgcctccc ctccccaggc cgcggcccag ctgaccctcg gggtccccc 60
ggcagcggac agggaagggt taaaggcccc cggtccctcg cccctgccc tggggaaccc 120
ctggccctgt ggggacatga actgtgtttg ccgcctggtc ctggctgtgc tgagcctgtg 180
gccagatata gctgtcgcgc ctggggcacc acctggcccc cctcgagttt ccccagacce 240
tcgggcccag ctggacagca ccgtgtcctc gaccgcctct ctctggcgg acacgcggca 300
gctggctgca cagctgaggg acaaatcccc agctgacggg gaccacaacc tggattccct 360
gcccaccctg gccatgagtg cgggggcact gggagctcta cagctcccag gtgtgctgac 420
aaggctcga gcggacctac tgtcctacct gcggcacgtg cagtggctgc gccgggcagg 480
tggtcttcc ctgaagacct tggagcccga gctgggcacc ctgcaggccc gactggaccg 540
gctgctgcgc cggctgcagc tcctgatgtc ccgcctggcc ctgcccagc cacccccgga 600
cccgccggcg cccccctgg cgcccccctc ctacgctgg gggggcatca gggcccacca 660
cgccatcctg ggggggctgc acctgacct tgactgggcc gtgaggggac tctgctgct 720
gaagactcgg ctgtgaccgg gggcccaaa ccaccaccgt ccttccaaag ccagatctta 780
tttatttatt tatttcagta ctgggggcga aacagccagg tgatccccc gccattatct 840
ccccctagtt agagacagtc cttccgtgag gcctgggggg catctgtgcc ttatttatac 900
ttatttattt caggagcagg ggtgggaggg aggtggactc ctgggtcccc gaggagggag 960
ggactggggg cccggattct tgggtctcca agaagtctgt ccacagactt ctgcctggc 1020
tcttcccct ctaggcctgg gcaggaaact atattattta ttaagcaat tacttttcat 1080
gttggggtgg ggacggaggg gaaagggaa cctgggtttt tgtacaaaa tgtgagaaac 1140
ctttgtgaga cagagaacag ggaattaaat gtgtcataca tatccacttg agggcgattt 1200
gtctgagagc tggggctgga tgcttgggta actggggcag ggcaggtgga ggggagacct 1260
ccattcaggt ggaggtcccc agtgggcggg gcagcgactg ggagatgggt cggtcacca 1320
gacagctctg tggaggcagg gtctgagcct tgctggggc cccgcaactg atagggcctt 1380
ttgtttgttt tttgagatgg agtctcctc tgttgcttag gctggagtgc agtgaggcaa 1440
tctgaggtca ctgcaacctc cacctcccg gttcaagcaa ttctcctgcc tcagcctccc 1500
gattagctgg gatcacaggt gtgcaccacc atgcccagct aattatttat ttcttttcta 1560
tttttagtag agacaggggt tcaccatggt ggccaggctg gtttcgaact cctgacctca 1620
ggtgatcctc ctgcctcggc ctcccaaagt gctgggatta caggtgtgag ccaccacacc 1680
tgaccatag gtcttcaata aatatttaat ggaaggttcc acaagtcacc ctgtgatcaa 1740
cagtaccctg atgggacaaa gctgcaaggt caagatggtt cattatggct gtgttcacca 1800
tagcaaacct gaaacaatct agatatccaa cagtgagggg taagcaaat ggtgcatctg 1860
tggatagaac gccaccacgc cgcocggagc agggactgtc attcaggag gctaaggaga 1920
gaggcttgct tgggatatag aaagataacc tgacattggc caggcatggt ggtcacgcc 1980
tgtaatcctg gcactttggg aggacgaagc gagtggatca ctgaagtcca agagttcgag 2040
accggcctgc gagacatggc aaaaccctgt ctcaaaaaag aaagaatgat gtectgacat 2100
gaaacagcag gctacaaaa cactgcatgc tgtgatccca attttgtgtt tttctttcta 2160
tatatggatt aaaacaaaa tcctaaaggg aaatacgcca aaatgttgac aatgactgtc 2220

```

-continued

```

tccagggtcaa aggagagagg tgggattgtg ggtgactttt aatgtgtatg attgtctgta 2280
ttttacagaa tttctgccat gactgtgtat tttgcatgac acattttaaa aataataaac 2340
actattttta gaat 2354

```

```

<210> SEQ ID NO 18
<211> LENGTH: 1019
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 18
tgatgggaat cctgtcatto ttacctgtcc ttgccactga gagtgactgg gctgactgca 60
agtcccccca gccttgggggt catatgcttc tgtggacagc tgtgctattc ctggtcctcg 120
ttgctggggac acctgcagct cccccaaagg ctgtgctgaa actcgagccc cagtggatca 180
acgtgctcca agaggactct gtgactctga catgcccggg gactcacagc cctgagagcg 240
actccattca gtggttccac aatgggaatc tcattccac ccacacgcag cccagctaca 300
ggttcaaggc caacaacaat gacagcgggg agtacacgtg ccagactggc cagaccagcc 360
tcagcgaccc tgtgcatctg actgtgcttt ctgagtggct ggtgctccag acccctcacc 420
tggagttcca ggaggagaaa accatcgtgc tgaggtgcca cagctggaag gacaagcctc 480
tggtaagggt cacattcttc cagaatggaa aatccaagaa attttcccg tccgatccca 540
acttctccat cccacaagca aaccacagtc acagtgggta ttaccactgc acaggaaaaca 600
taggctacac gctgtactca tocaagcctg tgaccatcac tgtccaagct cccagctctt 660
caccgatggg gatcattgtg gctgtgggta ctgggattgc tgtagcggcc attggtgctg 720
ctgtagtggc cttgatctac tgcaggaaaa agcggatttc agccaattcc actgatcctg 780
tgaaggctgc ccaatttgag atgctttcct gcagccacct ggacgtcaaa tgattgccat 840
cagaaagaga caacctgaag aaaccaacaa tgactatgaa acagctgacg gcggctacat 900
gactctgaac cccagggcac ctactgacga tgataaaaac atctacctga ctcttctctc 960
caacgacccat gtcaacagta ataactaaag agtaacgtta tgccatgtgg tcatctaga 1019

```

```

<210> SEQ ID NO 19
<211> LENGTH: 8213
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 19
ttttagtggg ggagtcaatt tatttctgag acgatctcat ttacctgaat gaggagctca 60
tatttatttt caggatttat gaggaatcca agatgaattt ggagcaggag aggcctttg 120
tctgcagtgc cccaggctgc tcccagcgtc tcccacaga ggaccatctg atgattcata 180
ggcacaaca tgaaatgact ttgaagtctc cttcaataaa aacagacaat atgttatcag 240
atcaaaactcc gacccaacg agattcctga agaactgcga ggaggtgggc ctcttcagcg 300
agctggactg ctccctggag cacgagttca ggaaggctca ggaagaggag agcagcaagc 360
ggaatatctc gatgcataat gcagttggg gggccatgac ggggcccgga actcaccagc 420
ttagcagcgc tcggctgccc aaccatgaca ccaacgttgt gattcagcaa gccatgccgt 480
cgctcagtc cagctctgtc atcactcagg caccttccac caaccgccag atcgggctcg 540
tcccagctc tctatctct ctgtacatc tccacaacag acagagacag cccatgccag 600

```

-continued

cctccatgcc tgggaccctg cccaacccta caatgccagg atcttccgcc gtcttgatgc	660
caatggagcg acaaatgtca gtgaactcca gcatcatggg gatgcaaggt ccaaatctca	720
gcaacccctg tgtcttccc caggtccagc caatgcattc agaagccaaa atgaggttga	780
aggctgcatt gactcaccac cctgctgcca tgtcaaatgg gaacatgaac accatgggac	840
acatgatgga gatgatgggc tcccggcagg accagacgcc acaccatcac atgcactcgc	900
accgcatca gcaccagaca ctgccacccc atcaccccta cccacaccag caccagcacc	960
cagcacacca tctcaccct caaccccatc accagcagaa ccatccacat caccactccc	1020
attcccact tcatgcacac ccagcacatc accagacctc gccacatccg cccctgcaca	1080
ccggcaacca agcacaggtt tcaccagcaa cacaacagat gcagccaacc cagacaatac	1140
agccacccca gccacaggg gggcgccggc gaagggtggt agacgaggat cgggacgaga	1200
ggcgccgaa atttctggaa cggaaaccggg cagctgccac cggctgcaga cagaagagga	1260
aggctgggt gatgtcattg gaaaagaaag cagaagaact caccagaca aacatgcagc	1320
ttcagaatga agtgtctatg ttgaaaaatg agtgggcca gctgaaacag ttgttgttaa	1380
cacataaaga ctgccaata acagccatgc agaaagaatc acaaggatat ctaagtccag	1440
agagtagccc tctgctagt cctgtcccag cttgtccca gcaacaagtc atccagcata	1500
ataccatcac tacttctca tcggtcagcg agtggttagg aagctccacc ctccagcagc	1560
tcaccactca cagaacagac ctgaatccga ttcttataaa tgcaccatca gacctggcct	1620
ccaagaagag ctgtagcgtt ccatgctgcc tttcttttaa gggcattttt agaattaact	1680
cagacctgga agactcctca gttcttcaaa gactggcttt catttttata gttattatgg	1740
aatgttgtc ttttatactt agttatataa gaaaaaggg agttatgcaa ttaatatcta	1800
tcagctggg aaacgcttg gtgctttct ccagtttct ggtaccagtt acttgtttat	1860
aaactgaacc tttctgtat atagccatgg ttctattctt atcagtccaa ccctttgcct	1920
gaaacattga atcttgtaa accacagctt ttagctaaaa tgaggatata ctagatgtca	1980
agtaagacag atccaaggta actgggtagg aaatcttttg acatctaac tcatgttgag	2040
tttgtctgt ggtgtccaca gaattccaga taaacacaca gcctttccca taccttttt	2100
tttcttacta taaaatatta taagatccat tgatgtccaa ataataccac caagcatctc	2160
ttcacctctc ctctcttgg tccacttgc atgcccagt tttcttctcc atttccactt	2220
tttcttaggc tccctattta ctattcattt tgacttctt ctgttttatt tttttccctt	2280
tagcattgca tgtgaataag aaaataatgt ttaaagaaaa aaaaaaaaaa gcaaacctcc	2340
aaaacgtgga cctaaccatt gcttccacta cacttcaccc acagctggag ttcatccaac	2400
tcttgctttt cacaaaatag taaccaggag atgtttaatg tgcctgattt aatgttttta	2460
ataatcacag caaatgaaag gtggtttagt tataagtgaa gcatggttga ataccagctg	2520
gggagacact agggaagggg gctttgtaag ccttgattgc gaaagtccaa attttgatgt	2580
ggggctataa catgacaccc ttggattgcg actggtttta tacggcctgc ctataacgctt	2640
gaaaaatccat gtactacata ataattcaga agggctctat tcaactacaca gattacattg	2700
ttcaatcatc agctgctaag agcetaagat ttattttttt ttttttetta agcctatgga	2760
accggctttg ctgttctggg ggggtgaaat agactaacta ctggagaaac aaagagagaa	2820
agaaaaccca gtgtttccat aggggcactt ttagccttcc cacaacagtt aagcactctt	2880

-continued

tgactgctga aggaacccca tggatgaggt gcaggctact tcaactcttt tttttctttt	2940
ttgagacaga gtctcaccta ttgcccagac tgaagtgcag tgggtgcgac atggatcaact	3000
gcagcagcat cctccgagtt caagctatcc ttccacctcc gcctcctgag tagctgggac	3060
cacaggttca cataaccatg cctggctaata ttatttttac ttttatttta aaataaaaaga	3120
tgaggtctgt cttatgttgc ccaggctggt ctcaactat cctacttctt cctcccaaag	3180
tgttgggatt ataggtgtga gccactgcac ccagcctact tcaactctct gaattattct	3240
gatttatttt caacaacttt tgtgaacttg cccgtgatac aaagcagata gtccctgaac	3300
cacagtcgtg cctccttgaa acaagccatt ctactgtgct aatgttttaa taccacatct	3360
cacaaataac aggggtgaat gtttctctct agcaatctag gcagggtctg gtgtttcatc	3420
tccatttgaa tctctgacct cttaatgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgttc	3480
atgggtttta aaagaacagt attttcaaaa aggtgtagct tttataagag tgcagaaaag	3540
ggaaggatgt gttttttct ctcaactatg tataagaatc tattttggag aaaaaaagaa	3600
aatatgaggg tctcgaagca tgatttttat ataactagtt tcagttttat ctaataactt	3660
actttttaaa tcaatattta tcaacaatct ttccttgat gcagtgcttt caaaagatgg	3720
ttttgagtgt ccagtgaaac ttatgacttg gatatatggt tgaagaatca aaacaaaagc	3780
aaaaaaaaaa agcaaaaaaa gaaaagagaa aaaaaaagaa aatgcaaatg gaataatttt	3840
ctattatatt ttagacaaaac atatcatttt cgagtatttt aaactactgaa tccatagttg	3900
ttgtttttta aattccaaca gtaacagctg aatggtttaa tctgactggc ttcctaagaa	3960
atgtttaaga ctacagctta aaaagaagtt aacattcata tctctgtttt gaaatcaaaa	4020
atcatatttc aaaattcttt cctaggacca tctatgtgtc tcccctcccc tccacaaaaa	4080
ggagaaaagag tgcattaaaa tgtttagtgt ggttttttaa ttttaattt ttagttatg	4140
ttttgctttg ttttaagtaa acaaaaattt ttctttcttt actgcatgca tagcacttaa	4200
taaaaatggat ttttaaaaaa tccactagta atacagaat gtccagggag tgactgtcac	4260
tacaatgatg gtttagttta cttctgttcc accttttgat tgaatatatt agttgttagg	4320
ctgaaagcct cggcagttaa gaacttgctt gagttttctt cgttcagcaa cttgacagtt	4380
tgactgatgt gcattatata tagctcaatt atgtctgttt tttatgctaa gtaggaaaac	4440
caaccacaca cattagcaaa ccggcctcaa catataatta gaataaactg tcttctgttt	4500
ctactcaggg cctttaggtg tgttcattca cggtatggaa atacagtaaa tgaagattc	4560
caactagttg tcagtgcttc ttgaaattcc aaacagaaaag atacattggt caaatccaac	4620
acttggttta tcaatattaa gtcttttacc taaaggccca gccgtcacca gacaacagaa	4680
taatcaatct gcctgaaaaa cctcctcctt tgcctacac tttttgctg tttgggagaa	4740
tatctttgta ctccattctc ctccctcagc cagttactgg gtcacccatc catgtgttca	4800
tgaatcaatc atcacggcct gcagagcacc tgccttaagg agggaaaatc ctgtcacact	4860
gcctcctccc attcgtgtgt ggttttcttg atcgggtgaga tctgtctctg aagtcactgc	4920
cagcctcctt gggaaactct atagtgcctc cctgcctta tgtgatggga gtttaacaact	4980
cagataagta caectgagag catttctatc aggtaaaact tcaactaaat ggaggtgtcc	5040
acatcttaat tgtttctctt tgacacattt ctcaatccac gaagccagga gaggtagagt	5100
gaaaaatccca gccatggatg aatgtactaa tttgaaagcc aagtgttaag tccgatgttt	5160

-continued

tcccgttaca ctactactca gccctctcct gcggccacat caacggatgc aagtcacagt	5220
cttaacacag cctgtgggag acaagcagtt tgtgtgctca cagtataat tatagtaatt	5280
agggtgactt agagcaata ctcttcagat cctatgtagt cagtgaaca aatggagag	5340
cgtattctga tagaaggacg tcgacgggta atgttctggt gggtgtgccc tgttaagtaa	5400
actttagtgt gtaagttgag tttgtcatta aaatcataaa ccagctgcgg taacagacaa	5460
gcctttggct ggggagtttt aagcctcggg aactgctata aaactagcca tccagttagg	5520
atagaatgtg tttctttctg gtaaaaaaa ggaaaaacca tctaagaaaa tatatatgta	5580
tgtatgtgtg tatacagtgg aattcaaagg accaaagcaa aatttgaaca ggaatctatt	5640
aatttagaat ttataagat atttattaat aaatgttatt tttaaacatt ccatttgaac	5700
agtattctgt aggatctact tgtttttaaa gtgttagtcc ataataaact actatagtta	5760
tgtgtatttt catttttcag ggtttcaaat ggctattctc catcatttgg tggaaatgtt	5820
tgcttagatc tctgtgcata gacatttcaa ggatttttat tgcctgtgta gttatttttt	5880
aatcaacatt ctgaacagtt ttttttaaac atttatttct gtgtgttcat ttttaaagta	5940
agctctttca tttaggaagc agagtccagc taaagggat cagtaactct aactggaaca	6000
gctttcttgt agaagtgtaa aaacagcttc atctctgcct ctctcacc caccccaatt	6060
tcctagaaag ccttgcaacta ttcagctccc ttagtgcttt ttgtcccttc ccgaacaata	6120
tgcagtactt ttaagccatt caagctccat tatgcagtat atctgagaag gaaaggaaa	6180
caaccattt aaatttgaat aaaaccgtgc ctatgcgaac agtagcaatt tagaatctct	6240
tttctgcttt taaaataatt tataattaaa aattgcactt tagctttttg atccctttgt	6300
atttctctta ttctctttct aacctcttct ctgtcctcaa acttgccttt gctctccttt	6360
acaatacccc ccaccctcc tccaaggctc tgagcggcat catttaaat actttacaga	6420
tatttgcacc aggtacattt atgtgcgtcc attggtagca cagctgagac ctgtgtctca	6480
catcagccta ggtgaagcct actacaagaa tgccaaggag aagagccagt acactatag	6540
gtttatactc tttatccctt tatteatagc atgtttttta aaaatgttat attatgcaac	6600
agatgtgagg cagcagctaa gctatactta agaattttct ctcaoctcc aaaccaaggt	6660
gtcctgaata agccaggaga cttattcttt tgtgcaccct ggtgcacatc tgactgttgt	6720
cctagccata gactctctga ggccactgaa agaacagtgg ccctatcgat ttcattccta	6780
ggtctcaaaa atacaatgtt gccttgtaac ataattaggg acagcacctc tatttcacaa	6840
ttataatcta aggtaggata agacgacaca gcagcaataa acttacaagt aaaattcaat	6900
acaaaaaca acacaagaa atttaaaaaa caaaaaacct agctcatcat gttgtgaaaa	6960
tgaaaaagtg aatgtccatt caaaatattt tactatttct tgtggagttt ttcagtgatg	7020
taatgcttgt agccaaattg cttaaagagt gtttatatat tttttcctt ataaattgtc	7080
tattttttaa aaaagctatt taaccacagc tgaagtgggg ggtaaggcca aattgccaac	7140
acttgtaaaa agattaatac tcttaagtgg cactctgata cctttccaac ttgtcatcag	7200
aaaggaatca ataattacca actgttgtat ttagaccaac ttacaatata tagctcatta	7260
gaagccagga tctagaagac tccctctaag ccatttaaga tattottaca ttgagcttca	7320
tattatagaa ctttatagga ttggatattt tacaatagaa taatttagcc tcaggactga	7380
gaatgtggaa gctgaataaa ttagctttaa atacatcatt aaaatcttat gcacaataag	7440

-continued

```

ctcattagat tctagtttcc tcccttagaa taccaatgcc acagacacta caggagataa 7500
tgaaaggat cagttgtgtt gagtggagg agttaaagag aaaggacct tcccaaccag 7560
cagccagtag aaaatacaac ctactcacct tttcccttc taagttctgc taaatcacat 7620
ctgcctcata gagaaaggaa tgttgctttt gagaactgtc ttggagaaca gataagcttg 7680
aatgttttc tctagagagg acatagggtt tgggatcctc tgaaaaggcc cagaaaaata 7740
gctcagttca aatacaatgt tctaggacaa ttggaatata aatattgtcc aaaaatataa 7800
ttaaagaaa aaagttagc actgtgtaaa gtaagtgtta actgaggaag tcccaaaaag 7860
gtgtgtcac ttaagtctt ggacttgggg ttctttgtat ttgtaaacag caaagcattt 7920
gtgtttgtt gtctattttt aaagcaacca ccttccttat tggaggaga aaaaaagggg 7980
tacatacatg taaatacttg ctgcagcatt taatatgttt aattttgtgt taagcttttt 8040
gttgcacgt gaacacattt attgttacca atggacaatg agttcattaa gactgttcaa 8100
ctaggtcaga tttttacatc tctttctagc aagaagagac aagattttgt gcatttgtac 8160
aatgttaat atcactgcaa ttccaatata ataaagcact caaatgcaaa taa 8213

```

```

<210> SEQ ID NO 20
<211> LENGTH: 5356
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 20

```

```

cgggattcgg ctggctctgc cacaccaccg cgcgccccgc ctccgccgc cctccggggc 60
gogtcttttc cgggctcgcg ctgagtcacc cctccgccgg ctgtccgggt gcgcgcgcgc 120
cgctgcggct tttctctcgg cctccgccgc gcgctcctcc tcgtcccagc gctagcgggc 180
acgcggttcc tttttcggag ctttcagagt gccagcgccc ggcggctgc gaagacgcgg 240
tgggcccgcc ctccgattga aatcacagaa gatattcgtg ttcttcttaa gagaaaaaga 300
ggacatttta gctttctcag ttgaaggcgt actttattgt cggcttccaa agattactaa 360
cttttatctg taccactaag attgaaactgc cttggctgta ctgctattct tactgtctgt 420
tctattattg ccttcttcag cacaataagg ctttcaaaag ccaagaata acaagaataa 480
agcaccattt tagaagcctt tccactatga aacttaagct aaatgtgctc accattattt 540
tgctgcctgt ccacttgta ataacaatat acagtgcctt tatatttatt ccatggtatt 600
ttcttaccaa tgccaagaag aaaaacgcta tggcaagag aataaaagct aagcccactt 660
cagacaaaacc tggaaagtcca tctcgtctg tcacacactt cgaactcacta gctgtaatag 720
acatccctgg agcagatact ctggataaat tatttgacca tgctgtatcc aagtttggga 780
agaaggacag ccttgggacc agggaaatcc taagtgaaga aaatgaaatg cagccaaatg 840
gaaaagtttt taagaagtta attcttggga attataaatg gatgaactat cttgaagtga 900
atcgcagagt gaataacttt ggtagtggac tcaactgcact gggactaaaa ccaagaaca 960
ccattgcat cttctgtgag accagggccc aatggatgat tgcagcacag acctgcttta 1020
agtacaactt tctcttgtg actttatag ccacacttgg caaagaagca gtagttcatg 1080
ggctaaatga atctgaggct tctatctgta ttaccagtgt tgaactctg gaaagtaaac 1140
ttaagactgc attgttagat atcagttgtg ttaaacatat ctttatgtg gacaataagg 1200
ctatcaataa agcagagtac cctgaaggat ttgagattca cagcatgcaa tcagtagaag 1260

```

-continued

agttgggatc taaccagaa aacttggca ttctccaag tagaccaacg ccttcagaca	1320
tggccattgt tatgtatact agtggttcta ctggccgacc taaggagtg atgatgcac	1380
atagcaatth gatagctgga atgacagccc agtgtgaaag aatacctgga ctgggaccga	1440
aggacacata tattggctac ttgcctttgg ctcatgtgct agaactgaca gcagagatat	1500
cttgctttac ctatggctgc aggattggat attcttctcc gcttacctc tctgaccagt	1560
ccagcaaat taaaaagga agcaaggag actgtactgt actgaagccc acacttatgg	1620
ctgctgttcc ggaaatcatg gatagaatth ataagaatgt tatgagcaaa gtccaagaga	1680
tgaattatat tcagaaaa ctgttcaaga tagggatga ttacaaattg gaacagatca	1740
aaaagggata tgatgcacct ctttgcaatc tgttactgtt taaaaaggtc aaggccctgc	1800
tgggaggaa tgtccgatg atgctgtctg gaggggcccc gctatctcct cagacacacc	1860
gattcatgaa tgtctgcttc tgctgcccc ttggccaggg ttatggactg acagaatcat	1920
gtggtgctgg gacagttact gaagtaactg actatactac tggcagagtt ggagcacctc	1980
ttatttgctg tgaattaag ctaaaagact ggcaagaagg cggttatata attaatgaca	2040
agccaaacc cagagtgaa atcgtaattg gtggacagaa catctccatg ggatattth	2100
aaaatgaaga gaaaacagca gaagattatt ctgtggatga aaatggaca agtggtttt	2160
gcactggtga tattggagaa ttccatccc atggatgth acagattata gatcgtaaga	2220
aagatctagt gaagttaca gcaggagagt atgtatctct tgggaaagta gaagctgcac	2280
tgaagaattg tccactatt gacaacatct gtgcttttgc caaaagtgat cagtctatg	2340
tgatcagtht tgtggttctc aaccagaaaa ggttgacact tttggcaca cagaaaggg	2400
tagaaggaac ttgggttat atctgcaata atcctgctat ggaagctgaa atactgaaag	2460
aaatcgaga agctgcaaat gccatgaaat tggagcgatt tgaattcca atcaaggtc	2520
gattaagccc agagccatgg acccctgaaa ctggtttgg aactgatgct ttcaaactga	2580
aaaggaagga gctgaggaac cttacctca aagacattga acgaatgat gggggcaaat	2640
aaaatgtht tgtcttattg acagttgtgc aggagtagc ctggtgtht tcaacctca	2700
gaattthtaag cttttgttga actgtagaa tgaaggat atcattctaa agatagagta	2760
aaaagaaac aaaacaaaa gttatthaaa ttgttgcctg gttacttht acttagtht	2820
gcatagtht agtgcagctg aaattgaaaa gttatttccc tttagctgth ttattataga	2880
gcagaaatc tgtttthaaa aattagccta agatatactt gttttgtht agaaaaat	2940
ttaatgtht acaaaataaa ttggagttg agtagaatgt agttgagga aattgacgc	3000
ttccaatgcc tcttgcctc ctatttcaga agththaaata ttaagcatga cagaaat	3060
gtatthaacac tactcaagc aaaagtgcg cagggttht aaattctctt ccaaccatt	3120
atcttgaagg aaaaattcaa tagtaata atacacaaaa tcaataata cttagaagg	3180
tattaagatt ataattgtht cataggttag atatagagtc attgthtgt tgtgaataat	3240
tacagthcct aaaataagaa tagaacaaca tataaacac caaaaaat ctagtht	3300
atthaaagg aaattgagct gcttttttg aaacttgag atcthaaat aactgtht	3360
atthaatga ctaaggagaa agtacattt ttgaaatgct gaaaattgcc tttctgtht	3420
tattcaaac gaaaagctga gaccaagagc aaggaagtht aaaagtht acaggcaacat	3480
ttctcttag aaaagtht gaaatcataa gtatttgtht ttagaacct tgcacagc	3540

-continued

```

tgaacctggg aaagagattt aaactctgaa tttatctttg ataacagggg ttgattttaa 3600
aatgtacatg tattaaatta catttgtaat ttaaggtctg tttgctgttg ctgattttat 3660
tcttgatcag tagtttgcac ttcagaaagc ctttcatttt gctttaagtt tagcaaagcg 3720
gggttataat gaatgacttc cccaatatct tgcttgaact tacagtgatt aacttggatg 3780
agttttggga agttaaaggg aagaaaacac tgttatcatt ttttctgtt tgggaagagc 3840
ttagaaactg gaaatactag atttgggaga agggcagagt tacttgataa gggacttgat 3900
gtttgtgcag taacttggga gtgtggttc tttttgaatc ttaattaaa acctgggatt 3960
atatatccct gataaatatt cacacttgaa ccatagttac tgtaaaatgc aaaaaatctt 4020
aatactgtta ttctttgcac tttttcttaa tcatttttta tatatatgca tatatatatg 4080
tgtgtgtgtg tgttgcttat gttgtttgt acagatgtgg gccaccattg caacaaaata 4140
cattcttttt gctctaaaat atttatgaag aaaatactta aatgttatgt ataggtggt 4200
aataagggaa aaatcaagta ttataacaa gaatgaaggt ttttgtaaag atttctgttc 4260
agcgttttgc aaggtaaaat tttaggcaag ttttccctga agttatgtgt atgtgagtat 4320
tctcattctt cccaacttgc ctttgaagag tgaatatca ttattatcaa gtagactact 4380
gttcagcttt tattccttcc ctggttgttt atcccttagg aatgagtttc ttagactttc 4440
ccaatatgtg attttttttc ccatttagaa tgggtatttt aaatgtgtga gtgcatgtac 4500
tatcttatct cagataattg caccocaaat ctgccccaa ctcccaaaag ctagaacact 4560
gccaactgat ctgttatagg tccttttaga acacataatt aaccttaag gttgggtgct 4620
gctaattctt tgcaaaaatc caaatattgt taagggacca gggagatgcc actaccctt 4680
gattttccat ctaaaaatat acatgtttat gtaaacaaat ctttccatat ccatagtgac 4740
ttttcaagta ttttaagcct aagattttga tctcacattt ttatacctgt ttaaattgct 4800
cacagttatt acatacacac cagccatcaa ctaaagtgtt actttaaaaa tttactacaa 4860
tatgtacatt tctaagtcaa acacttgtga cttttgcttt aattocatga atgttctctg 4920
ctccttgata tttgtattta ttcttttttt ctctagagta gaggtataat tgtgtgatat 4980
ttcagaaata cagataaatg attcaaaaag tcacagttaa ggagaatcat gtttctttga 5040
tcatgaataa ctgattagta agtcttgctt atattttcct gatagcatat gacaaatggt 5100
tctaaggtaa caagatgaga acagataaag attgtgtggt gttttggatt tggagagaaa 5160
tattttaatt tttaaatgca gttacaaatt ataatgtatt catatttga ctttctgtta 5220
aaatgcatga ttgcagaatt gtttagattt tgtgtttatt cttgatgaaa agctttgttt 5280
gttcttgttt ttaagtttgc actcaaatct taagaaataa atccaccat gttatcaaaa 5340
aaaaaaaaa aaaaaa 5356

```

<210> SEQ ID NO 21

<211> LENGTH: 948

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 21

```

attgtgtgct cttgtagctg tccccggagc cctcagcagc agttggagct ggtgcacagg 60
aaggatgagg aagaccaggc tctgggggct gctgtggatg ctctttgtct cagaactccg 120
agctgcaact aaattaactg aggaaaagta tgaactgaaa gaggggcaga ccttgatgt 180

```

-continued

gaaatgtgac tacacgctag agaagtttgc cagcagccag aaagcttggc agataataag	240
ggacggagag atgcccaga ccttggcatg cacagagagg ccttcaaaga attcccatcc	300
agtccaagtg gggaggatca tactagaaga ctacatgat catggtttac tgcgcgtccg	360
aatggtaaac cttcaagtgg aagattctgg actgtatcag tgtgtgatct accagcctcc	420
caaggagcct cacatgctgt tcgatcgcct ccgcttggtg gtgaccaagg gtttttcagg	480
gacccctggc tccaatgaga attctacca gaatgtgat aagattcttc ctaccaccac	540
taaggccttg tgcccactct ataccagccc cagaactgtg acccaagctc cacccaagtc	600
aactgccgat gtctccactc ctgactctga aatcaacctt acaaatgtga cagatatcat	660
cagggttccg gtgttcaaca ttgtcattct cctggctggg ggattcctga gtaagagcct	720
ggctttctct gtctgttttg ctgtcaecgt gaggtcattt gtaccctagg cccacgaacc	780
cacgagaatg tctctgact tccagccaca tccatctggc agttgtgcca agggaggagg	840
gaggagtaa aaggcaggga gtaataaca tgaattaaat ctgtaatcac cagctatttc	900
taaagtcagc gtctcacctt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	948

<210> SEQ ID NO 22

<211> LENGTH: 1240

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22

gagctcgag cgcgcggccc ctgtctccg gcccgagatg aatcctcggg cagaagccc	60
gttcaacatc ctctggcca ccgactccta caaggttact cactataaac aatatccacc	120
caacacaagc aaagtttatt cctactttga atgccgtgaa aagaagacag aaaactccaa	180
attaaggaag gtgaaatatg aggaaacagt attttatggg ttgcagtaca ttcttaataa	240
gtacttaaaa ggtaaagtga taaccacaaga gaaaatccag gaagccaaag atgtctacaa	300
agaacatttc caagatgatg tctttaatga aaagggatgg aactacattc ttgagaagta	360
tgatgggcat cttccaatg aaataaaagc tgttccctgag ggctttgtca tcccagagg	420
aaatgttctc ttcacgggtg aaaaacacaga tccagagtgt tactggctta caaattggat	480
tgagactatt cttgttcagt cctggtatcc aatcacagtg gccacaaatt ctagagagca	540
gaagaaaaata ttggccaaat atttgttaga aacttctggt aacttagatg gtctggaata	600
caagttacat gattttggct acagaggagt ctcttcccaa gagactgctg gcataggagc	660
atctgctcac ttggttaact tcaaaggaac agatacagta gcaggacttg ctetaattaa	720
aaaatattat ggaacgaaag atcctgttcc aggctattct gttccagcag cagaacacag	780
taccataaca gcttggggga aagaccatga aaaagatgct tttgaacata ttgtaacaca	840
gttttcatca gtgcctgtat ctgtggctcag cgatagctat gacatttata atgcgtgtga	900
gaaaatatgg ggtgaagatc taagacattt aatagtatcg agaagtacac aggcaccact	960
aataatcaga cctgattctg gaaaccctct tgacactgtg ttaaaggttt tggagathtt	1020
aggtaagaag tttcctgtta ctgagaactc aaagggttac aagttgctgc caccttatct	1080
tagagttatt caaggggatg gagtagatat taatacctta caagaggat gtgttttata	1140
ttaaaagttt caataaggca tttcttataa ttaagtttgt ttatgtttga taaagaacac	1200
aatataaata caaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	1240

-continued

<210> SEQ ID NO 23

<211> LENGTH: 6919

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 23

```
cccacggatc atcagaaggc gcgacctgg aggaggcgc ccagaaggcg acgcctcttg      60
cctcctgtct ctgcctcttc gaaggaagtt tgctcttaat ttcagagccg ggttcgccgt      120
cggatcaacc tccaggagct agcagcgggc gcggaccggg cagtttccgc gctcagcaca      180
ggcagctcgc ggtcatgggc ggctcagcct ccagccagct ggacgagggc aagtgcgctt      240
acatccgagg gaaaaactgag gctgccatca aaaacttcag tccctactac agtcgtcagt      300
actctgtggc tttctgcaat cacgtgcgca ctgaagtaga acagcaaaga gatttaacgt      360
cacagttttt gaagaccaag ccaccattgg cgcttgaac tattttgtat gaagcagagc      420
tatcacaatt tctgaagac ataaagaagt ggaaggagag atacgttgta gttaaaaatg      480
attatgctgt ggagagctat gagaataaag aggcctatca gagaggagct gtcctctaaat      540
gtcgaattct tccagccggg ggcaaggtgt taacctcaga agatgaatat aatctgttgt      600
ctgacaggca tttcccagac cctcttgctt ccagtggaga ggagaacact cagccctttg      660
tggctcctgc caaggaattc ccagtgtacc tgtggcagcc cttcttcaga cacggctact      720
tctgcttcca cgaggctgct gaccagaaga ggtttagtgc cctcctgagt gactgcgctca      780
ggcatctcaa tcatgattac atgaagcaga tgacatttga agcccaagcc tttttagaag      840
ctgtgcaatt cttccgacag gagaagggtc actatggttc ctgggaaatg atcactgggg      900
atgaaatcca gatcctgagt aacctggtga tggaggagct cctgccact cttcagacag      960
acctgctgcc taagatgaag ggaagaaga atgacagaaa gaggacgtgg cttggtctcc     1020
tcgaggaggg ctacaccctg gttcagcacc aagtttcaga aggattaagt gccttgaagg     1080
aggaatgcag agctctgaca aagggcctgg aaggaacgat ccgttctgac atggatcaga     1140
ttgtgaaetc aaagaactac ttaattggaa agatcaaagc gatggtggcc cagccggcgg     1200
agaaaagctg cttggagagt gtgcagccat tcttggcacc catcctggag gagctcatgg     1260
gaccagtgag ctccggatcc agtgaagtac gtgtactctt tgagaaagag gtgaatgaag     1320
tcagccagaa cttccagacc accaaagaca gtgtccagct aaaggagcat ctgaccggc     1380
ttatgaatct tccgtgcat tccgtgaaga tggaaacctg ttatactaaa gtcaacctgc     1440
ttcacgagcg cctgcaggat ctcaagagcc gcttcagatt ccccccacatt gatctggtgg     1500
ttcagaggac acagaactac atgcaggagc taatggagaa tgcagtgttc acttttgagc     1560
agttgctttc cccacatctc caaggagagg cctccaaaac tgcagttgcc attgagaagg     1620
ttaaactocg agtctttaaag caatatgatt atgacagcag caccatccga aagaagatat     1680
ttcaagaggc actagttaa atcacacttc ccaactgtgca gaaggcactg gcgtccacat     1740
gcaaacccaga gcttcagaaa tacgagcagt tcatctttgc agatcatacc aatgatgatc     1800
acgttgaaaa tgtctatgag gagattttac atcagatcct gcttgatgaa actctgaaag     1860
tgataaagga agctgctatc ttgaagaaac acaacttatt tgaagataac atggccttgc     1920
ccagtgaag tgtgtccagc ttaacagatc taaagcccc cacagggtca aaccaggcca     1980
gccctgccag gagagcttct gccattctgc caggagtctt gggtagtggag accctcagta     2040
```

-continued

acgaagtatt ccaggagtca gaggaagaga agcagcctga ggccctagc tcgttggcca	2100
aaggagaaag cctttctctc cctgggcca gcccaccccc agatgggact gagcaggtga	2160
ttatttcaag agtggatgac cccgtggtga atcctgtggc aacagaggac acagcaggac	2220
tcccgggcac atgctcatca gagctggagt ttggaggac ccttgaggat gaagaacccg	2280
cccaggaaga gccagaacct atcactgect cgggttcttt gaaggcctc agaaagtgc	2340
tgacagcgtc cgtggaagta ccagtggact ctgctccagt gatggaagaa gatacgaatg	2400
gggagagcca cgttcccaaa gaaaatgaag aagaagagga aaaagagccc agtcaggcag	2460
ctgccatcca ccccgacaac tgtgaagaaa gtgaagtcag cgagaggag gcccacctc	2520
cctgtcccga ggcccattgg gaggagttgg ggggatttcc agaggtaggc agcccagcct	2580
ctcccagc cagtggaggg ctaccagg agcccctggg gccatggag ggggagctcc	2640
caggagaggc ctgcacactc actgccatg aaggaagagg gggcaagtgt accgaggaag	2700
gggatgcctc acagcaagag ggctgcacct taggttctga cccatctgc ctcaagtga	2760
gccaggttcc tgaggaacaa gaagagatgg gagggcaaaag cagcgggccc caggccacgg	2820
ccagtgtgaa tgcagaggag atcaaggtag cccgtattca tgagtgtcag tgggtggtgg	2880
aggatgctcc aaacccggat gtctctgtgt cacacaaaga tgactggaag gagggagaag	2940
gtggtcagga gagtttcca gagctgacct cagaggagt aaagggaca tttggtgaa	3000
gtcttctctc gaaaaaagc aaacgcttat aggggtacac ttaggggttg catgcaagct	3060
gttaccaaaa aatttttaag tattttctta atttgaataa taaaaccaga ggaatgcat	3120
acagggcatg agcaactgag gcaaaccttt gtggacatga attgttctac gatgaattt	3180
tgctttagta ttttaataag aattacaaag acaatggcat actgggggtg agaggagct	3240
gaggatgtct gagggaggaa tagtattgca gggaagactg agaaaacagt aggatgacag	3300
ttttgagtat actctgcact tttcaattgt gcaatcttct tgtgcacttt aaggctttt	3360
aattttgttt gagaatgcaa atgtatactg taagtctacc tttactatct actatgcta	3420
cttaccatc tcttaaggac tcggcatttg tccacagtca gactgcaaga gaggttaggt	3480
catgaacagt caccctgtct ggctgtagcc cccacagagg caatcatgcc caatagatc	3540
aagagaagct aagcggaaat ggaggggtga aggtgtgatc tgtgggactg tctggcctg	3600
ttactcatcc tgctatcaat ttcttattaa ttaatcttga tgattcttat taattaatca	3660
catttgcagg aattcagat gaggcaagaa aattttattg gcctgggtaa gactgaaagc	3720
attccaaatt aggccttagc tgtgcaaaag gcttagctaa gttatcgagc ttaaaacccg	3780
tcaattaaac aaacattatt tgaacagtta ctgcatgcca cgcactgtgt tgggcttagt	3840
aataaaaaaa agaaaagata agtgcttgtt ctagcataaa ttaaaaggtc caaggaatt	3900
taatctggaa gagaacatat gccaatttt aaactatgac agctttttt tctctttcca	3960
ttcaaatagt cctggttcat tcccagaagg gcacaaaatg aatgaataaa taataaatg	4020
aataaagaca aaagccaagg tgatgctct caagttccaa agatgttatc aaamctgaa	4080
atcatttgtt tggtcattca gcaagctaat tgagtctctg ttatatacca agcactggg	4140
ataccatggc gaaaaacaac tttgttctt cctcctagaa cttacatttt aatggaata	4200
gacaaaaac atcttcttaa cggatggtga cctataacca ttaatgttga aatggaaga	4260
gacttgcttc caaaagatta aaaggagttg ttcttttctc cttcagaaaa ataccagatc	4320

-continued

atttccataa atctccagtc ccaagtatta catcgtggtt tccctccccg acttttttatt	4380
ttatttttatt ctattttttt gagatggagt ctctgctctgt cgccaaggct ggagtgcagt	4440
gggtgatct cggctcactg caacctccgc ctctctgggt caggagattc tcctctgtca	4500
gcgtcccaag tagctggaat tacagccatg cggcaccatt cccggctaatt tttgtattt	4560
ttggtggaga tgggttttca ccatgttggc caggctggtc tcractcct gatttcaggc	4620
gatccacca cctcgacttc ccaaagtgtc gggattacag cgtgagctac ctggccccctc	4680
cccaaatttt aacatcagat ctcaacatat ttgttgagaa cttgtttaga tcatcactat	4740
cgagtaaatt tatgtcagtt ctttagagca ttgtcttcc acacagtca accatttaac	4800
cagaaattaa aatatcccc ttttcagccc tcccatgaag aatttctagg tcataaacat	4860
gaataaact cagtatatgt ataatatcat atttaattgt tgcaacttt cttggcgtgt	4920
ttaaatctc atgtctttga gggatcccac cacagtgaa ggaagtaaag accacttctc	4980
cccagtcctc tgctttctgc agtctcact catatcctgg catttagcat ggtgtgttga	5040
aattaggttt acttcttctc tctccaagtg gctcctctct taaaagacag aaactagggt	5100
ttagtcatca tttgtcttc ctgctagaaa cccacagcct tgaataatgg cttcctgcct	5160
ccttgagtca ctttaatat cctgggtgat aggacctggc atgctcagg ctgctcggga	5220
aatgtgggaa ctggacacc agaacactgc tgtgctgggg ctatttgggg cctgctgtca	5280
ggcagaaaga cgttttgaat tgggctttct gcctctgttg agttttctct taagtaaagt	5340
ccaaagtcca aggggcagat ggccagatgc actgccagc aaggcaggaa gccaaagagg	5400
cagtgccagc cccacaaaag ctgcccgaact ccctgggaca gtagtgtgga gtcccagccc	5460
aggctgacct cacaccggag cttcctagct tcctttcttt gctcaatgca gggcttcttg	5520
caccccttg aaagctaaga gattttttc aaccctaaaa gagagtacct ttcactgcat	5580
tggatggatg aacatcagtg ccctaacttt atccatcatg taggtcaggg aggactgggc	5640
actatttggc aggatgtact ccagaatata atcaaagaat tttctgtaca tatttcacta	5700
aagacaagtt tttgggctgg atgtggtagc ttacacctat aatcccagca ctttgggagg	5760
ctgaggtgag agggctcact gagcccagga gttctagaca agcctgggaa acatagcaag	5820
atctcactc tacaaaaaat aataatggtg tgtgcctgca gtcctagcta cttgggtggt	5880
tgaggcagga ggattgtttg agcgagggaa gttgaggctg cagtgaacta tgattttacc	5940
actgcactcc agcttgggag acggagtgag accctgtctc aaaaaaaaaa aaagttttct	6000
agaataagca ggatgattgt ttaatttgaa gatggaacag gaaactagag tgcatttaaa	6060
atactctgtc ttcattttaa catgttgaat ggaataactg catatcacca tgagtttgtt	6120
ttgcttttca tacagacttg tatgtgcat ttgagtgggt tcagattgg agcgagggtta	6180
ttctgatcta aatgaacagc atttttttcc ttagcctctg tttgocactc tgggtatcgt	6240
ctcctatggc aaagccatta gaaatgcata aaacctcgag acatggtttt tggcaaaaac	6300
tccatgactt taaactagct cttttactac tgacctttca cagagaaaaa atatttccct	6360
tgaaaaaaac tgggcttgtc attttttccc ttgtagcttt aagcagagac ataagtgcct	6420
tgcattacac atagtaaac ttcttttaaa aaaaaaaaaa aagattttgg agactaccag	6480
ggtaagattc caacttgtcc aaaagctttc tggccttaca tattttatta taaaaattct	6540
caagtctggt aatcttctat gtcagageta gtgatttcaa aaggtttcac aattcccaca	6600

-continued

```

gacaaaagtg attttcgttc attataataa ggtaagtga tatgtgattc ataacaattt 6660
tgatgtgaag aagggaaagga catcattgac ttaataatag tatcagtcgg tgcaacagtt 6720
ggcaacatgt gccttcacac tttaccataa agagacgggt ttgagggttt gccttctaaa 6780
gtctgcaact tcaagaaaaa aaatcgacac tgtggattga ctttccgggt cactatataa 6840
agcaataaaa cttaaaacac tttgtaacca tgtatttact ctgccagggt cctatattcc 6900
aataaaatgt tcatccttg 6919

```

```

<210> SEQ ID NO 24
<211> LENGTH: 1489
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 24

```

```

ctcctctgcc caatgtctcc caatctcttt ctttctctc ttcagttcct ccaggtaatt 60
cttactcaaa cttgtaccaa cttgtttttg actgacagtg aacagtgaga gagttttctt 120
cattttgagg aaccctaaac acctatcttt cccaaggcaa cctgtctgga ctgagcattt 180
ctctgacttg acataacttc ccatccagcc aggagtctgc actcttcagt ctttgcaggc 240
agtagcagaa tcccatggta gccagggtgg tgaaggggag cgaggacggt ctacctgcct 300
tgaagaagac acctgacctg cggagtgagt gaccagtgtt tccagagcct ggcaatggat 360
gccattcaca tcggcatgtc cagcaccctc ctggtgaagc aactgctgg ggctgggctc 420
aaggccaaca gaccccgctg catgtccaag agtgggcaca gcaacgtgag aattgacaaa 480
gtggatggca tatacctact ctacctgcaa gacctgtgga ccacagttat cgacatgaag 540
tggagataca aactcaccct gttcgtctgc acttttgtga tgacctgggt cctttttgga 600
gtcatctact atgccatcgc gtttattcat ggggacttag aaccogatga gccatttca 660
aatcataccc cctgcatcat gaaagtgagc tctctcactg gggcgtttct cttttccctg 720
gaatcccaga caaccattgg ctatggagtc cgttccatca cagaggaatg tctctatgcc 780
atcttctctg tggttgctca gttggtcatc acgaccttga ttgagatctt catcaccgga 840
accttctctg ccaaaaatgc cagaccctaa aagcgggctg agaccatcaa gttcagccac 900
tgtgcagtca tcaccaagca gaatgggaag ctgtgcttgg tgattcaggt agccaatatg 960
aggaagagcc tcttgattca gtgccagctc tctggcaagc tctgcagac ccacgtcacc 1020
aaggaggggg agcggattct cctcaaccaa gccactgtca aattccacgt ggactcctcc 1080
tctgagggcc ccttctcat tctgccatg acattctacc atgtgctgga tgagacgagc 1140
cccctgagag acctcaccac ccaaaacctc aaggagaagg agtttgagct tgtggtcctc 1200
ctcaatgcca ctgtggaatc caccagcctg gtctgccaga gccgaacatc ttatatccca 1260
gaggaaatct actgggggtt tgagtttgtg cctgtgggat ctctctcaa aaatggaaaa 1320
tatgtggctg atttcagtca gtttgaacag attcggaaaa gccagattg cacattttac 1380
tgtgcagatt ctgagaaaaa gcaactcgag gagaagtaca ggcaggagga tcagagggaa 1440
agagaactga ggacactttt attacaacag agcaatgtct gatcacagg 1489

```

```

<210> SEQ ID NO 25
<211> LENGTH: 3537
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

-continued

<400> SEQUENCE: 25

```

ggccccctcga gcctcgaacc ggaacctcca aatccgagac gctctgctta tgaggacctc   60
gaaatatgcc ggccagttaa aaaatcttgt ggctttgagg gcttttggtt ggccaggggc   120
agtaaaaatc tcggagagct gacaccaagt cctcccctgc cacgtagcag tggtaaagtc   180
cgaagctcaa attccgagaa ttgagctctg ttgattctta gaactggggt tcttagaagt   240
ggtgatgcaa gaagtttcta ggaaggccg gacaccagg tttgagcaaa attttggact   300
gtgaagcaag gcattggtga agacaaaatg gcctcgcgg ctgacagctg tatccagttc   360
accgcctatg ccagtgatgt tcttctcaac cttaatcgtc tccggagtcg agacatcttg   420
actgatgttg tcattgttgt gagccgtgag cagtttagag cccataaaac ggtcctcatg   480
gcctgcagtg gcctgttcta tagcatcttt acagaccagt tgaatgcaa ccttagtgtg   540
atcaatctag atcctgagat caaccctgag ggattctgca tccctcggga cttcatgtac   600
acatctcggc tcaatttgcg ggagggcaac atcatggctg tgatggccac ggctatgtac   660
ctgcagatgg agcatgttgt ggacacttgc cggaaagtta ttaaggccag tgaagcagag   720
atggtttctg ccatcaagcc tccctgtgaa gagttcctca acagccggat gctgatgcc   780
caagacatca tggcctatcg gggctgtgag gtggtggaga acaacctgcc actgaggagc   840
gcccctgggt gtgagagcag agcctttgcc cccagcctgt acagtggcct gtccacaccg   900
ccagcctctt attcctatga cagccacctc cctgtcagca gcctcctctt ctccgatgag   960
gagtttcggg atgtccggat gcctgtggcc aacctctcc ccaaggagcg ggcactccca  1020
tgtgatagtg ccaggccagt cctggtgag tacagccggc cgactttgga ggtgtcccc   1080
aatgtgtgcc acagcaatat ctattcacc aaggaacaa tcccagaaga ggcacgaagt   1140
gatatgcact acagtgtggt tgagggcctc aaacctgctg cccctcagc ccgaaatgcc   1200
ccctacttcc cttgtgacaa ggccagcaaa gaagaagaga gacctcctc ggaagatgag   1260
attgccttgc atttcgagcc cccaatgca cccctgaacc ggaaggtctt ggttagtcca   1320
cagagcccc agaaatctga ctgccagccc aactcgccea cagagtcttg cagcagtaag   1380
aatgcctgca tctccaggtc ttctggtctc cctccagcca agagccccac tgaccccaaa   1440
gcctgcaact ggaagaaata caagttcatc gtgctcaaca gcctcaacca gaatgcaaaa   1500
ccagaggggc ctgagcaggc tgagctgggc cgctttccc cacgagccta cacggcccca   1560
cctgcctgcc agccaccat ggagcctgag aaccttgacc tccagtcccc aaccaagctg   1620
agtgccagcg gggaggactc caccatccca caagccagcc ggctcaataa catcgttaac   1680
aggtoctatg cgggctctcc ccgcagcagc agcgagagcc actcaccact ctacatgcac   1740
ccccgaagt gcacgtcctg eggctctcag tccccacagc atgcagagat gtgcctccac   1800
accgctggcc ccagcttccc tgaggagatg ggagagaccc agtctgagta ctacagattct   1860
agctgtgaga acggggcctt cttctgcaat gagtgtgact gccgcttctc tgaggaggcc   1920
tcaactcaaga ggcacacgct gcagacccac agtgacaaac cctacaagtg tgaccgctgc   1980
caggcctcct tccgctacaa gggcaacctc gccagccaca agaccgtcca taccggtgag   2040
aaacctatc gttgcaacat ctgtggggcc cagttcaacc ggccagccaa cctgaaaacc   2100
cacactcgaa tcaactctgg agagaagccc tacaaatgag aaacctgagg agccagattt   2160
gtacaggtgg cccacctccg tgcccatgtg cttatccaca ctggtgagaa gccctatccc   2220

```

-continued

```

tgtgaaatct gtggcaccgc ttcccgccac cttcagactc tgaagagcca cctgcgaatc 2280
cacacaggag agaaacctta ccattgtgag aagtgttaacc tgcatttccg tcacaaaagc 2340
cagctgcgac ttcacttgcg ccagaagcat ggcgccatca ccaacaccaa ggtgcaatac 2400
cgcggtgcag ccactgaect gcctccggag ctccccaaag cctgctgaag catggagtgt 2460
tgatgctttc gtctccagcc ccttctcaga atctacccaa aggatactgt aacactttac 2520
aatgttcac ccatgatgta gtgcctcttt catccactag tgcaaatcat agctgggggt 2580
tgggggtggg gggggctggg gcctggggga ctgggagccg cagcagctcc cctcccccca 2640
ctgccataaa acattaagaa aatcatattg cttcttctcc tatgtgtaag gtgaaccatg 2700
tcagcaaaaa gcaaaatcat tttatatgtc aaagcagggg agtatgcaaa agttctgact 2760
tgactttagt ctgcaaaatg aggaatgtat atgttttgtg ggaacagatg tttcttttgt 2820
atgtaaatgt gcattctttt aaaagacaag acttcagtat gttgtcaaag agagggcttt 2880
aattttttta accaaagggtg aaggaatata tggcagagtt gtaaataatata aaatataatata 2940
atatataaaa taaataatata taaacctaac aaagataatata taaaaatata aaactgcgctt 3000
aaaggctcga ttttgtatct gcagggcagac acggatctga gaatctttat tgagaaagag 3060
cacttaagag aatattttta gtattgcatc tgtataagta agaaaatatt ttgtctaaaa 3120
tgcctcagtg tatttgtatt tttttgcaag tgaaggttta caatttaca agtgtgtatt 3180
aaaaaaaaaca aaaagaacaa aaaaatctgc agaaggaaaa atgtgtaatt ttgttctagt 3240
tttcagtttg tatatacccg tacaacgtgt cctcacggtg ccttttttca cggaggtttt 3300
caatgatggg cgagcgtgca ccatcccttt ttgaagtgta ggcagacaca gggacttgaa 3360
gttgttacta actaaactct ctttgggaat gtttgtctca tcccattctg cgtcatgctt 3420
gtgttataac tactccggag acagggtttg gctgtgtctc aactgcatta ccgctgtgta 3480
aaatatagct gtacaaatat aagaataaaa tgttgaaaag tcaaactgga aaaaaaa 3537

```

<210> SEQ ID NO 26

<211> LENGTH: 5855

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

```

agttgcctgc gcgcccctgc cggaccggcg gctccctagt tgcgccccga ccaggccctg 60
cccttctctc cggtctgcgc gcgtccgcgc cccctccatt cctgggcgca tcccagctct 120
gccccaaactc gggagtcagc gcccgggcgc cagtgcctgc ttcagctccg gttcactgcg 180
cccgcgggac gcgcccggga ggactccgca gccctgctcc tgaccgtccc cccaggctta 240
accggctcgc tccgctcgga ttctctggct gcgctcctgc ggggtggcgc ttctctcccg 300
cgccccctcc cctctgccc atgaagaatc attggaatat taagcccagg agttgctttg 360
gggatggctg gaagtgcaat gtcttccaag ttcttctag tggctttggc catatttttc 420
tccttcgccc aggttgtaat tgaagccaat tcttggtggt cgctaggtat gaataaccct 480
gttcagatgt cagaagtata tattatagga gcacagcctc tctgcagcca actggcagga 540
ctttctcaag gacagaagaa actgtgccac ttgtatcagg accacatgca gtacatcgga 600
gaaggcgcga agacaggcat caaagaatgc cagtatcaat tccgacatcg aaggtggaac 660
tgcagcactg tggataaac ctctgttttt ggcaggtgta tgcagatagg cagccgcgag 720

```

-continued

acggccttca catacgcggt gagcgcagca ggggtggtga acgccatgag ccgggctgtc	780
cgcgagggcg agctgtccac ctgctgctgc agccgcgccc cgcgcccaca ggacctgccc	840
cgggactggc tctggggcgg ctgctggcgc aacatcgact atggctaccg ctttgccaag	900
gagttcgtgg acgcccgcga gcgggagcgc atccacgcca agggctccta cgagagtgt	960
cgcacctca tgaacctgca caacaacgag gccggccgca ggacggtgta caacctggct	1020
gatgtggcct gcaagtgcc tgggggtgccc ggctcatgta gcctgaagac atgctggctg	1080
cagctggcag acttccgcaa ggtgggtgat gccctgaagg agaagtacga cagcgcggcg	1140
gccatgcggc tcaacagccc gggcaagtgt gtacaggcca acagccgctt caactcggcc	1200
accacacaag acctggtcta catcgacccc agccctgact actgctgctg caatgagagc	1260
accggctcgc tgggcacgca gggccgcctg tgcaacaaga cgtcggaggg catggatggc	1320
tgcgagctca tgtgctgccc ccgtggctac gaccagtcca agaccgtgca gacggagcgc	1380
tgccactgca agtccactg gtgctgctac gtcaagtgca agaagtgcac ggagatcgtg	1440
gaccagtgtg tgtgcaagta gtgggtgcca cccagcactc agccccgctc ccaggacccc	1500
cttatttata gaaagtacag tgattctggt ttttggttt tagaaatatt ttttatttt	1560
ccccagaat tgcaaccgga accatttttt ttctgttac catetaagaa ctctgtggtt	1620
tattattaat attataatta ttatttgcca ataatggggg tgggaaccaa gaaaaatatt	1680
tattttgtgg atctttgaaa aggtaataca agacttcttt tgatagtata gaatgaaggg	1740
gaaataacac atacctaac ttagctgtgt ggacatggt cacaaccaga aggtaagaa	1800
atacatttc tttttctcaa atatgccatc atatgggatg ggtaggttcc agttgaaaga	1860
gggtggtaga aatctattca caattcagct tctatgacca aatgagttg taaattctct	1920
ggtgcaagat aaaagttctt gggaaaaaca acaaaaaca acaaacctc ccttcccag	1980
cagggtgct agcttgcttt ctgcattttc aaaatgataa tttacaatgg aaggacaaga	2040
atgtcatatt ctcaaggaaa aaaggatat cacatgtctc attctctca aatattccat	2100
ttgcagacag accgtcatal tctaatagct catgaaattt gggcagcagg gaggaagtc	2160
cccagaaatt aaaaaattta aaactcttat gtcaagatgt tgatttgaag ctgttataag	2220
aattaggatt ccagattgta aaaagatccc caaatgattc tggacactag attttttgt	2280
ttggggaggt tggcttgaac ataaatgaaa atactctgtt attttcttag ggatacttgg	2340
ttagtaaat ataatagtaa aaataataca tgaatccat tcacaggttc tcagcccag	2400
caacaaggta attgcgtgcc attcagcact gcaccagagc agacaaccta tttgaggaaa	2460
aacagtgaaa tccaccttcc tcttcacact gagccctctc tgattctctc gtgtttgat	2520
gtgatgctgg ccacgtttcc aaacggcagc tccactgggt cccctttggt tgtaggacag	2580
gaaatgaaac attaggagct ctgcttggaa aacagttcac tacttaggga tttttgttc	2640
ctaaaacttt tattttgagg agcagtagtt ttctatgttt taatgacaga acttgctaa	2700
tggaaatcac agaggtgttg cagcgtatca ctgttatgat cctgtgttta gattatccc	2760
tcatgcttct cctattgtac tgcaggtgta ccttaaaact gttcccagtg tacttgaaca	2820
gttgattta taagggggga aatgtggttt aatgggtgct gatattctca agtctttgt	2880
acataacata tatatatata tacatatata taaatataaa tataaatata tctcattgca	2940
gccagtgatt tagatttaca gtttactctg gggttatttc tctgtctaga gcattgtgt	3000

-continued

ccttcactgc agtccagttg ggattattcc aaaagttttt tgagtcttga gcttgggctg	3060
tggccctgct gtgatcatac cttgagcacg acgaagcaac cttgtttctg aggaagcttg	3120
agttctgact cactgaaatg cgtgttgggt tgaagatato ttttttctt tctgcctcac	3180
ccctttgtct ccaaccteca tttctgttca ctttgtggag agggcattac ttgttcgtta	3240
tagacatgga cgtaagaga tattcaaac tcagaagcat cagcaatggt tctcttttct	3300
tagttcattc tgacagaatg aaacccatgc ctattagaaa tgacagtaact tattaattga	3360
gtccctaagg aatattcagc ccactacata gatagctttt ttttttttt ttttaataag	3420
gacacctctt tccaaacagt gccatcaaat atgttcttat ctcagactta cgttgtttta	3480
aaagtttggg aagatacaca tctttcatac ccccttagg caggttggct ttcatatcac	3540
ctcagccaac tgtggctctt aatttattgc ataatgatat tcacatcccc tcagttgcag	3600
tgaattgtga gcaaaagatc ttgaaagcaa aaagcactaa ttagttaaaa atgtcacttt	3660
tttggttttt attatacaaa aaccatgaag tactttttt atttgctaaa tcagattggt	3720
cctttttagt gactcatggt tatgaagaga gttgagtta acaatcctag cttttaaaag	3780
aaactattta atgtaaaaa ttctacatgt cattcagata ttatgtatat cttctagcct	3840
ttattctgta cttttaatgt acatatttct gtcttgcgtg atttgtatat ttcactgggt	3900
taaaaaaca acatcgaaag gcttatgcca aatggaagat agaataaaa ataaaaagtt	3960
acttgatat tggtaagtgg ttccaattgt ccttcagata attcatgtgg agattttttg	4020
agaaacctg acggatagtt taggatgact acatgtcaaa gtaataaaa agtgggtgaat	4080
tttaccaaaa ccaagctatt tggaaagctc aaaaggttct tatatgtaat ggaacaaaag	4140
gggaattctc ttttctata tatgttctct acaaaaaaaaa aaaaaaaga aatcaagcag	4200
atggcttaaa gctgggtata ggattgctca cattctttta gcattatgca tgtaacttaa	4260
ttgttttaga gcgtgttgct gttgtaacat cccagagaag aatgaaaagg cacatgcttt	4320
tatccgtgac cagattttta gtccaaaaa atgtattttt ttgtgtgttt accactgcaa	4380
ctattgcacc tctctatttg aatttactgt ggacctgtg tgggtgtctct atgcccttg	4440
aaagcagttt ttataaaaag aaagcccggg tctgcagaga atgaaaactg gttgaaaact	4500
aaaggttcat tgtgttaagt gcaattaata caagttattg tgcttttcaa aaatgtacac	4560
ggaaatctgg acagtgctgc acagattgat acattagcct ttgctttttc tctttccgga	4620
taaccttgta acatattgaa accttttaag gatgccaaga atgcattatt ccacaaaaa	4680
acagcagacc aacatataga gtgtttaaaa tagcatttct gggcaaatc aaactcttgt	4740
ggttctagga ctcacatctg tttcagtttt tctcagttg tatattgacc agtgttcttt	4800
attgcaaaaa catatacccg atttagcagt gtcagcgtat tttttctct catcctggag	4860
cgtattcaag atcttcccaa tacaagaaaa ttaataaaaa atttatatat aggcagcagc	4920
aaaagagcca tgttcaaat agtcattatg ggctcaaata gaaagaagac ttttaagttt	4980
taatccagtt tatctgttga gttctgtgag ctactgacct cctgagactg gcaactgtga	5040
agttttagtt gctacccta gctctttct cgtacaattt tgccaatacc aagtttcaat	5100
ttgtttttac aaaacattat tcaagccact agaattatca aatatgacgc tatagcagag	5160
taaatactct gaataagaga ccggtactag ctaactccaa gagatcgtta gcagcatcag	5220
tccacaaaca cttagtggcc cacaatatat agagagatag aaaaggtagt tataacttga	5280

-continued

```

agcatgtatt taatgcaaat aggcacgaag gcacaggctc aaaatactac attgtcactg 5340
taagctatac ttttaaaata tttatTTTTT ttaaagtatt ttctagtctt ttctctctct 5400
gtggaatggg gaaagagaga tgccgtgTtt tGaaagtaag atgatgaaat gaatttttaa 5460
ttcaagaaac attcagaaac ataggaatta aaacttagag aaatgatcta atttccctgt 5520
tcacacaaac tttacacttt aatctgatga ttggatattt tattttagtG aaacatcatc 5580
ttgttagcta actttaaaaa atggatgtag aatgattaaa ggttggtatg attttttttt 5640
aatgtatcag tttgaaccta gaatattgaa ttaaaatgct gtctcagtat tttaaaagca 5700
aaaaaggaat ggaggaaaat tgcactctag accattttta tatgcagtgt acaatttgct 5760
gggctagaaa tgagataaag attatTTatt tttgttcata tcttgactt ttctattaaa 5820
atcattttat gaaatccaaa aaaaaaaaaa aaaaa 5855

```

<210> SEQ ID NO 27

<211> LENGTH: 2681

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

```

gcagggtggg ggcaggccag ctcagcagag cctggggcca gagggccaga cagccacaga 60
gctcctggcg tgggcaaggc tggccaagga tggcgacgcc caggggcctg ggggcccctg 120
tctgctect cctgctccc acctcaggtc aggaaaagcc caccgaaggg ccaagaaaca 180
cctgctggg gagcaacaac atgtacgaca tctcaactt gaatgacaag gctttgtgct 240
tcaccaagtG caggcagtcg ggcagcgact cctgcaatgt ggaaaacttg cagagatact 300
ggctaaaacta cgaggcccct ctgatgaagg aaggtttgac gcagaaggtg aacacgcctt 360
tctgaaggc tttggtccag aacctcagca ccaacactgc agaagacttc tatttctctc 420
tggagccctc tcaggttccg aggcaggTga tgaaggacga ggacaagccc cctgacagag 480
tgcgacttcc caagagcctt tttcgatccc tgccaggcaa caggctctgG gtcgcttgg 540
ccgtcaccat tctggacatt ggtccaggga ctctcttcaa gggcccccg ctcggcctgg 600
gagatggcag cggcgtgTtg aacaatcgcc tggTgggttT gagTgTggga caaatgcAtg 660
tcaccaagct ggctgagcct ctggagatcg tcttctctca ccagcgaccg ccccctaaca 720
tgaccctcac ctgtgtatcc tgggatgtga ctaaaggGac cactggagac tggctctctg 780
agggtgctc caccgaggtc agacctgagg ggacctgtg ctgctgtgac cacctgacct 840
tttgcgccc gctcctgaga cccaccttg accagtcac ggtgcatatc ctcacacgca 900
tctcccaggc gggctgtggg gtctccatga tcttctggc cttcaccatt attctttatg 960
cctttctgag gctttcccgG gagaggTtca agtcagaaga tgccccaaag atccacgtgg 1020
ccctgggtgG cagcctgttc ctctgaatc tggccttctt ggTcaatgtg gggagtggct 1080
caaaggggtc tgatgctgcc tgetgggccc ggggggctgt cttccactac ttctgctct 1140
gtgcttccac ctggatgggc cttgaagcct tccacctcta cctgctcgtc gtcagggtct 1200
tcaaacaccta cttegggcac tacttctga agctgagcct ggtgggctgG ggcctgcccG 1260
ccctgatggT catcgcaact gggagtGcca acagctacgg cctetacacc atecgtgata 1320
gggagaaccG cacctctctg gagctatgct ggttccgtga agggacaacc atgtacgccc 1380
tctatatac cgtccacggc tacttctca tcaccttct ctttggcatg gtggctcctg 1440

```

-continued

```

ccctgggtgt ctggaagatc ttcaccctgt cccgtgctac agtgggtcaag gagcggggga 1500
agaaccggaa gaaggtgctc accctgctgg gcctctcgag cctgggtgggt gtgacatggg 1560
ggttggccat cttcaccctc ttgggcctct ccaccgtcta catctttgca cttttcaact 1620
ccttgcaagg tgtcttcato tgctgctggg tcaccatect ttacctccca agtcagagca 1680
ccacagtctc ctctctact gcaagattgg accaggccca ctccgcatct caagaatagg 1740
aaggcacggc cctgcaatat ggactcagct ctggctctct gtgtgacctt gggcagctcc 1800
gtgcctctct ctgtactccc tcagtttctt tctctgtaca atgtggctgg ggaggagag 1860
gatgggacca ggttggaaca cgtggcatca gaggtcccat ccagatccaa ctataggctc 1920
aagagtccac gtaagcaggt ttgcaaggct ctaaagtctc tatagtctct agaccctctg 1980
ccagcaaaga gtgacagtca cctccatgcc ctgccctcat tgcaaagccc tcactcacct 2040
tctgggtctc gcaagggagg agagtctggt gctggcatag cctggaagg agccccagc 2100
ctctcccctc ctctccttg tcactggcct cccacaactc ccttctggc tgctgtaac 2160
cttgaggggc attcaggagg ccagcgttcc ctccagcact gggggtttgt tttgggggt 2220
gggagttgat cctcccacc agtctgcccc tggctctctc ccatccaatc agagcccacc 2280
ctctggaag agaccctctg gttcagagtg ctggcagccc tgcacgtgtc cagggacact 2340
gcatttcaaa gaaccactga gtgggtgagc taccttgggc aaacccccca ctctgactc 2400
tgactgccac gtgggtggcc gcacctctga cctgctgtca tcatagaggt agaaagcaaa 2460
caatctgggg ctccagcac acgtgggtgc tccactcat tcagtgtgtg gggccctga 2520
gcagaggctg ggcattgcca ctagaacctg agctcctaga gagcaaggac ctgggtggcc 2580
tcgcttactg ttccagccca ggccaagcac agggcctggc tcgtggcaaa ccttgaataa 2640
atatttgttg gctgaaaaaa aaaaaaaaaa aaaaaaaaaa a 2681

```

```

<210> SEQ ID NO 28
<211> LENGTH: 659
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 28

```

```

gtgcagctgg gagagctaga ctaagttggt catgatgcag aagctactca aatgcagctg 60
gcttgctctg gctcttgccc tcatcctggg tctggaatcc tcagttcaag gttatcctac 120
gcagagagcc aggtaccaat ggggtgcctg caatccagac agtaattctg caaactgcct 180
tgaagaaaaa ggaccaatgt tcgaactact tccaggtgaa tccaacaaga tccccctct 240
gaggactgac ctttttccaa agacgagaat ccaggacttg aatcgtatct tcccactttc 300
tgaggactac tctggatcag gcttcggctc cggctccggc tctggatcag gatctgggag 360
tggtttccta acggaatgg aacaggatta ccaactagta gacgaaagtg atgctttcca 420
tgacaacctt aggtctcttg acaggaatct gccctcagac agccaggact tgggtcaaca 480
tggattagaa gaggatttta tgttataaaa gaggattttc ccaccttgac accaggcaat 540
gtagttagca tattttatgt accatggtta tatgattaat cttgggacaa agaattttat 600
agaaattttt aaacatctga aaaagaagct taagttttat catccttttt ttttctcat 659

```

```

<210> SEQ ID NO 29
<211> LENGTH: 3573

```

-continued

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

```

gcgttcagcg gacgcgcgcg gcctcgatct ctggactcgt cacetgcccc tccccctccc   60
gccgccgtca cccaggaaac cggcgcgaat cgccggccga cctgaagctg gtttcatggc   120
agcctcaaag aaggcagttt tggggccatt ggtgggggcg gtggaccagg gcaccagttc   180
gacgcgcttt ttggttttca attcaaaaac agctgaacta cttagtcata atcaagtaga   240
aataaaacaa gagttcccaa gagaaggatg ggtggaacag gaccctaagg aaattctaca   300
ttctgtctat gagtgtatag agaaaacatg tgagaaactt ggacagctca atattgatat   360
ttccaacata aaagctattg gtgtcagcaa ccagagggaa accactgtag tctgggacaa   420
gataactgga gacccctctc acaatgtgtt ggtgtggctt gatctaagaa cccagctctac   480
cgttgagagt cttagtaaaa gaattccagg aaataataac tttgtcaagt ccaagacagg   540
ccttcacttt agcacttact tcagtcagat gaaacttcgt tggctccttg acaatgtgag   600
aaaagtcaa aaggccgttg aagaaaaacg agctcttttt gggactattg attcatggct   660
tatttgagtg ttgacaggag gagtcaatgg aggtgtccac tgtacagatg taacaaatgc   720
aagtaggact atgcttttca acattcattc tttggaatgg gataaacaac tctgcgaatt   780
ttttggaatt ccaatggaaa ttcttccaaa tgtccggagt tcttctgaga tctatggcct   840
aatgaaagct ggggccttgg aaggtgtgcc aatatctggg tgtttagggg accagctctgc   900
tgcattggtg ggacaaatgt ccttcagatg tggacaagcc aaaaatacgt atggaacagg   960
atgtttctta ctatgtaata caggccataa gtgtgtattt tctgatcatg gccttctcac  1020
cacagtggtc taaaaacttg gcagagacaa accagtatat tatgctttgg aaggttctgt  1080
agctatagct ggtgctgtta ttcgctggct aagagacaat cttggaatta taagacctc  1140
agaagaaatt gaaaaacttg ctaaagaagt aggtacttct tatggctgct acttcgtccc  1200
agcattttcg gggttatatg caccttattg ggagcccagc gcaagagggg taatctgtgg  1260
actcactcag ttcaccaata aatgccatat tgcttttgct gcattagaag ctgtttgttt  1320
ccaaactcga gagatttttg atgccatgaa tcgagactgt ggaattccac tcagtcattt  1380
gcaggtagat ggaggaatga ccagcaacaa aattcttatg cagctacaag cagacattct  1440
gtatatacca gtagtgaagc cctcaatgcc cgaaccact gcaactgggtg cggctatggc  1500
ggcaggggct gcagaaggag tcggcgtatg gagtctcgaa cccgaggatt tgtctgccgt  1560
cacgatggag cggtttgaac ctcagattaa tgcggaggaa agtgaaattc gttattctac  1620
atggaagaaa gctgtgatga agtcaatggg ttgggttaca actcaatctc cagaaagtgg  1680
tattccataa aaacctacaa ctcatggatt cccaagatgt gagcttttta cataatgaaa  1740
gaaccagca attctgtctc ttaatgcaat gacactatc atagactttg attttattta  1800
taagccactt gctgcatgac cctccaagta gacctgtggc ttaaaataaa gaaaatgcag  1860
caaaaagaat gctatagaaa tatttgggtg tttttttttt ttttaaacat ccacagttaa  1920
ggttgggcca gctacctttg gggctgacct cctccattgc cataacatcc tgcctcattc  1980
cctctaagat gtaggaagaa ttcggatcct taccattgga atcttccatc gaacatactc  2040
aaacactttt ggaccaggat ttgagtctct gcatgacata tacttgatta aaaggttatt  2100
actaacctgt taaaaatcag cagctctttg cttttaacag acaccctaaa agtcttcttt  2160

```

-continued

```

tctacatagt tgaagacagc aacatcttca ctgaatgttt gaatagaaac ctctactaaa 2220
ttattaaaat agacatttag tgttctcaca gcttggatat ttttctgaaa agttatttgc 2280
caaaactgaa atccttcaga tgttttccat ggtcccacta attataatga ctttctgtct 2340
ggatcttata ggaaaagata ctttcttttt tcttccatct ttccttttta tttttttac 2400
tttgatgta taacatacat gcttatatat tttatacact gagggtagcc catttataaa 2460
ttaaagacac atttatattc gaaggttcta acagggtctg tcttaagtga accactgtgt 2520
atataaatat gttgaaaaac agctgtatac atttttgggc aacggttatg cataatattt 2580
accaggagaa ttttttctt aacaagccaa cttttaaata ttatgtttta tgtcaataaa 2640
agaaaatata ctttattgtg acttcaacta ttttcttat cccttacatt tttatttaat 2700
tgtcttagct taaaaaaga agaaactgtg gaatactaca gtaaatattg ttttcaaca 2760
caagcaataa tcaaatagat ttttttctt ttgaattaat tttagacata ttttggatcc 2820
tattgagggg ataagaggat gtcaaaaaag ttaaatacct aagtagaaaa aaatatagaa 2880
ataaagccaa gaactctctt cagttcaaat gttatcaatt gttaataaga aattgctatc 2940
tgggatgaca gaattacctc tcttagtat ctcattataa ctgaaagaag gtttatcatt 3000
acaaatacct tccaatgaaa ccaagaatct ctcaaaatat ttaatgtcac atattataag 3060
aagttaccta atcctgcttc ttaacatcaa tttttaaaaa tatcttaaaa ttactttgtt 3120
ttgtagtaa cagtgaagaa aagattgect cctaattatt ttttcaatg agtgcgtgaat 3180
gggaaacat ttatatctta ctataaaagg ttctgttttg tttggaatca atggtagctt 3240
tattgactgt tctgattgtg ctgtttctaa tttattgaat ctgctaggtt ttattgatgc 3300
agccaccact taagtacat aaatattata gaaaggtact gtgaaatgat cactttgtgg 3360
caggggtact ttaaacata aatgtttcta caaaagtagg ttgagttcat tgtaaataat 3420
tgtgaaagcc actggtcaca taattttaag attacattaa ttttctata aattggaaga 3480
tttataaatg tttgaaattg tacacattga ttttaatga caaatttact taaaataaat 3540
tgacccttg ttcttaaaaa aaaaaaaaaa aaa 3573

```

<210> SEQ ID NO 30

<211> LENGTH: 1820

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30

```

gcgcgacga cgcgcgcggc agcgctccaa ctggctcctc gctcgggct cgcctcga 60
gccgggagag agcctccgcc agcggccagg caccagccag acgacgccag cgacccggc 120
ctctcgcgcg caccgcgcta actcaggggc tgcataggca cccagagccg aactccaaga 180
tgggaggcaa gctcagcaag aagaagaagg gctacaatgt gaacgacgag aaagccaagg 240
agaaagacaa gaaggccgag ggcgcggcga cggaagagga ggggacccc aaggagagtg 300
agccccaggc ggcgcgagag cccgcccagg ccaaggaggg caaggagaag cccgaccagg 360
acgccgaggg caaggccgag gagaaggagg gcgagaagga cgcggcggt gccaaaggag 420
aggccccgaa ggcggagccc gagaagacgg agggcgcggc agaggccaag gctgagcccc 480
cgaaggcgc cgagcaggag caggcggccc ccggccccgc tgcggcggc gagggcccca 540
aagctgctga ggcgcgcgag gcccccggcc agagcgcggc ccctgcccgc ggggaggagc 600

```

-continued

```

ccagcaagga ggaaggggaa cccaaaaaga ctgaggcgcc cgcagctect gccgcccagg 660
agacccaaag tgacggggcc ccagcttcag actcaaaacc cggcagctcg gaggetgccc 720
cctcttccaa ggagaccccc gcagccacgg aagcgcttag ttccacaccc aaggcccagg 780
gccccgcagc ctctgcagaa gagcccaagc cggtgaggc cccggcagct aattccgacc 840
aaaccgtaac cgtgaaagag tgacaaggac agcctatagg aaaaaaata ccacttaaaa 900
caatctctctc tctctctctc tctctctctc tctatctctc tctctatctc ctctctctct 960
ctctctctct atctctctct tctctctctc ctatactaac ttgtttcaaa ttggaagtaa 1020
tgatatgtat tgcccaagga aaaatacagg atgttgctcc atcaaggag ggaggggggtg 1080
ggagaatcca aatagtattt ttgtggggaa atatctaata taccttcagt caactttacc 1140
aagaagtctt ggatttccaa gatcgcgctc tgaagtgca gtacatcgtt tgtacctgaa 1200
actgccgcca catgcactcc tccaccgctg agagttgaa agcttttctt ctgcaatggg 1260
agttgggagt gatgcgtttg attctgccc cagggcctgt gccaaagcaa tcagatcttt 1320
atgagagcag tttttctgt gttttctttt taatttacag cctttcttat ttgatattt 1380
ttttaatggt gtggatgaat gccagcttcc agacagagcc cacttagctt gtccacatgg 1440
atctcaatgc caatctccca ttctctctct ccagatattt ttgggagtga caaacattct 1500
ctcatctctc ttagcctacc tagatttctc atgacgagtt aatgcctgct cgtggttggg 1560
tgcacctgta gttctgttta ttggtcagtg gaaatgaaaa aaaaaaaaaa aaaaagtctg 1620
cgttcattgc agttccagtt tctcttccat tctgtgtcac agacaccaac acaccactca 1680
ttgaaaaatg gaaaaaaaaa acaaaaaaaaa acaaaaaaaaa tgtacaatgg atgcattgaa 1740
attatatgta attgtataaa tgggtgcaaca gtaataaagt taaacaatta aaaagaaaaa 1800
aaaaaaaaaa aaaaaaaaaa 1820

```

```

<210> SEQ ID NO 31
<211> LENGTH: 533
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (154)(305)
<223> OTHER INFORMATION: Wherein n can be a, c, t, or g
<400> SEQUENCE: 31

```

```

gctcatagtc cgtcaccgaa aatagaaaaat gccatccata ggtaaaatgc tgacctatag 60
aaaaaaaaatga actctacttt tatagcctag taaaaatgct ctacctgagt agttaaagc 120
aattcatgaa gcttgaagct aaagagcact ctgntgggtt tggcataata gctgcatttc 180
cagacctgac ctttggcccc aaccacaagt gctccaagcc ccaccagctg accaaagaaa 240
gcccagttc tctttctgct cttcccacaa cctccctgct cccaaaaacta tgaattaat 300
ttganccata ttaacacagc tgactcctcc agtttactta aggtagaag aatgagtta 360
caacagatga aaataagtgc tttggggcga ctgtattcct ttaacagat ccaactatt 420
ttacatttaa aaaaaaagtt aaactaaact tctttactgc tgatatgttt cctgtattct 480
agaaaaattt ttactcttcc acattatttt tgtactcttt ccccatgtta agg 533

```

```

<210> SEQ ID NO 32
<211> LENGTH: 427

```

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 32
gaagaggagc aacatctatg ccaataactg tgcattctac aatggtgcta atctcagacc    60
taaatgatac tccatttaat ttaaaaaaga gttttaaata attatctatg tgectgtatt    120
tcccttttga gtgctgcaca acatgttaac atattagtgt aaaagcagat gaacaacca    180
cgtgttctaa agtctagggg ttgtgctata atccctatgt agttcaaaat taaccagaat    240
tcttccatgt gaaatggacc aaactcatat tattgttatg taaatacaga gttttaatgc    300
agtatgacat cccacagggg aaaagaatgt ctgtagtggg tgactgttat caaatatgtt    360
atagaataca atgaacgggtg aacagactgg taacttggtt gagttcccat gacagatttg    420
agacttg                                           427

<210> SEQ ID NO 33
<211> LENGTH: 424
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (122)(191)
<223> OTHER INFORMATION: Wherein n can be a, c, t, or g

<400> SEQUENCE: 33
agaaaatcat tcacatattg gttcactcaa caagcattta ttaaatatat attcactatt    60
ctagactaat agcaagaccg gggatcttgt ttaggaagaa gcagtccttg ttctcagaa    120
tntgcaaac cattaaaaaa gcacctactt taagccattt ttttcagca aggagtcatt    180
ctgccagaaa natgtagtac acaaatacag gataatataa caaatgtaa atttctcatt    240
tctagtgaat taaactttcc agtaatttta catctcacct cattcttatg atgtcagtt    300
tgcttaatta ttggcaaaa taatggtaaa atgtttgtac tgtattagag ctttactgtt    360
cgattattaa gatatttacc cagtatctta gagatgctgg acttcaattt tcttatttt    420
atct                                           424

<210> SEQ ID NO 34
<211> LENGTH: 488
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34
ttaaatattc attccattac atctagactc accaagaact acatgttatg atgttaagtt    60
gaagttgaaa catgatgttt tgcattaaat ttaagatag caaatattatg tagagaaaat    120
aaatgttata taccctataa tctttcacct aattagtatt taattatag gatttgtttt    180
atattataaa agatgttttg attttgcctt ttgatattga caaaattggt tggatatect    240
tatgttctca agtctgtatc tgectccctc gccttatttc ttatgttttg ccacagttaa    300
cccattgtgc ttctttgtaa tcaaacagtt tgtgggagaa tgggcttatt gaatgtctaa    360
aaaataagtt taaagtgttt gttaccctaa gttttttaca tttttaaact ctaattacat    420
atgtgaatgt tattactctc agtgaattgt tattgtttgc aaaaatgcac tgggcagtaa    480
cattttgt                                           488

```

-continued

```

<210> SEQ ID NO 35
<211> LENGTH: 3834
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 35

cctgagacag aggcagcagt gatacccacc tgagagatcc tgtgttttaa caactgcttc      60
ccaaaacgga aagtatttca agcctaaacc tttgggtgaa aagaactctt gaagtcatga      120
ttgcttcaca gtttctctca gctctcactt tgggtcttct cattaaagag agtggagcct      180
ggctttacaa cacctccacg gaagctatga cttatgatga ggccagtgct tattgtcagc      240
aaaggtacac acacctgggt gcaattcaaa acaagaaga gattgagtac ctaaaactcca      300
tattgagcta ttcaccaagt tattactgga ttggaatcag aaaagtcaac aatgtgtggg      360
tctgggtagg aaccagaaaa cctctgacag aagaagccaa gaactgggct ccaggtgaac      420
ccaacaatag gcaaaaagat gaggactgcg tggagatcta catcaagaga gaaaaagatg      480
tgggcatgtg gaatgatgag aggtgcagca agaagaagct tgcctatgc tacacagctg      540
cctgtaccaa tacatctctc agtggccacg gtgaatgtgt agagaccatc aataattaca      600
cttgcaagtg tgacctggc ttcagtggac tcaagtgtga gaaaattgtg aactgtacag      660
ccctggaatc ccctgagcat ggaagcctgg tttgcagtca cccactggga aacttcagct      720
acaattcttc ctgctctatc agctgtgata ggggttacct gccaaagcagc atggagacca      780
tgcaagtgtat gtctcttgga gaatggagtg ctctatttc agcctgcaat gtggttgagt      840
gtgatgtgtg gacaaaatca gccaatgggt tctggaatg tttccaaaac cctggaagct      900
tcccattgaa cacaaactgt acatttgact gtgaagaagg atttgaacta atgggagccc      960
agagccttca gtgtacctca tctgggaatt gggacaacga gaagccaacg tgtaaagctg     1020
tgacatgcag ggcctgccc cagcctcaga atggctctgt gaggtgcagc cattcccctg     1080
ctggagagtt caccttcaaa tcatctcgca acttcacctg tgaggaaggc ttcattgttc     1140
agggaccagc ccaggtttaa tgcaccactc aagggcagtg gacacagcaa atcccagttt     1200
gtgaagcttt ccagtgcaaa gccttgtcca accccgagcg aggctacatg aattgtcttc     1260
ctagtctctc tggcagtttc cgttatgggt ccagctgtga gttctctctg gacaggggtt     1320
ttgtgtttaa gggatccaaa aggcctcaat gtggccccc aggggagtggt gacaacgaga     1380
agcccacatg tgaagctgtg agatgcgatg ctgtccacca gccccgaag ggtttggtga     1440
gggtgtctca tccccctatt ggagaattca cctacaagtc ctcttgtgcc ttcagctgtg     1500
aggagggatt tgaattatat ggatcaactc aacttgagtg cacatctcag ggacaatgga     1560
cagaagaggt tcttctctgc caagtggtaa aatgttcaag cctggcagtt ccgggaaaga     1620
tcaacatgag ctgcagtggg gagcccgtgt ttggcactgt gtgcaagttc gcctgtcctg     1680
aaggatggac gctcaatggt tctgcagctc ggacatgtgg agccacagga cactggtctg     1740
gcctgtacc tacctgtgaa gctcccactg agtccaacat tcccttggtg gctggacttt     1800
ctgctgtctg actctcccct ctgacattag caccatttct cctctggctt cggaaatgct     1860
tacggaaagc aaagaaattt gttcctgcca gcagctgcca aagcctttaa tcagacggaa     1920
gctacaaaaa gccttcttac atcctttaag ttcaaaaaga tcagaaaacag gtgcatcttg     1980
ggaactagag ggatacactg aagttaacag agacagataa ctctctctgg gtctctggcc     2040
cttcttgect actatgccag atgcctttat ggctgaaacc gcaacaccca tcaccacttc     2100

```

-continued

```

aatagatcaa agtccagcag gcaaggacgg ccttcaactg aaaagactca gtgttccctt 2160
tctactctc aggatcaaga aagtgttggc taatgaaggg aaaggatatt ttcttccaag 2220
caaaggtaa gagaccaaga ctctgaaatc tcagaattcc ttttotaact ctcccttgct 2280
cgctgtaaaa tcttggcaca gaaacacaat attttgtggc tttcttctt ttgcocttca 2340
cagtgtttcg acagctgatt acacagttgc tgcataaga atgaataata attatccaga 2400
gtttagagga aaaaaatgc taaaaatatt ataacttaaa aaaatgacag atgttgaatg 2460
cccacaggca aatgcatgga ggggtgttaa tggtgcaaat cctactgaat gctctgtgcg 2520
agggttacta tgcacaatth aatcactttc atccctatgg gattcagtgc ttcttaaga 2580
gttcttaagg attgtgatat ttttacttgc attgaatata ttataatctt ccatacttct 2640
tcattcaata caagtgtggt agggacttaa aaaacttcta aatgctgtca actatgatat 2700
ggtaaaagtt acttattctc gattaccccc tcattgttta ttaacaaatt atgttacatc 2760
tgttttaat ttatttcaaa aagggaaact attgtcccct agcaaggcat gatgtaacc 2820
agaataaagt tctgagtgtt tttactacag ttgttttttg aaaacatggg agaattggag 2880
agtaaaaact gaatggaagg tttgtatatt gtcagatatt tttcagaaa tatgtggttt 2940
ccacgatgaa aaacttccat gaggccaaac gttttgaact aataaaagca taaatgcaaa 3000
cacacaaagg tataatttta tgaatgtctt tgttgaaaa gaatacagaa agatggatgt 3060
gctttgcatt cctacaaga tgtttgtcag atgtgatatg taaacataat tcttgtatat 3120
tatggaagat tttaaattca caatagaaac tcaccatgta aaagagtcac ctggtagatt 3180
tttaacgaat gaagatgtct aatagttatt ccctatttgt tttcttctgt atgttagggt 3240
gctctggaag agaggaatgc ctgtgtgagc aagcatttat gtttatttat aagcagattt 3300
aacaattcca aaggaatctc cagttttcag ttgatcactg gcaatgaaaa attctcagtc 3360
agtaattgcc aaagctgctc tagccttgag gagtgtgaga atcaaaactc tcctacactt 3420
ccattaactt agcatgtgtt gaaaaaaaa gtttcagaga agttctggct gaacactggc 3480
aacgacaaag ccaacagtc aaacagagat gtgataagga tcagaacagc agaggttctt 3540
ttaaaggggc agaaaaactc tgggaaataa gagagaacaa ctactgtgat caggctatgt 3600
atggaatata gtgttattht ctttgaaatt gtttaagtgt tgtaaatatt tatgtaact 3660
gcattagaaa tttagctgtgt gaaataccag tgtggtttgt gtttgagttt tattgagaat 3720
tttaattat aacttaaaat attttataat ttttaagta tatatttatt taagcttatg 3780
tcagacctat ttgacataac actataaagg ttgacaataa atgtgcttat gttt 3834

```

<210> SEQ ID NO 36

<211> LENGTH: 1334

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 36

```

aatcattaga gectgagtca ctctccccag gagaccaga cctagaacta cccagagcaa 60
gaccacagct ggtgaacagt ccaggagcag acaagatgga gacaaattcc tctctccca 120
cgaacatctc tggagggaca cctgctgtat ctgctggcta tctctctctg gatataatca 180
cttatctggt atttgcagtc acctttgtcc tcgggtctct gggcaacggg cttgtgatct 240
gggtggctgg attccggatg acacacacag tcaccacat cagttacctg aacctggccg 300

```

-continued

```

tggctgactt ctgtttcacc tccactttgc cattcttcat ggtcaggaag gccatgggag   360
gacattggcc ttctggctgg ttctctgtgca aattcgtctt taccatagtg gacatcaact   420
tgttcggaaag tgtcttcctg atcgcocctca ttgctctgga ccgctgtgtt tgcgtcctgc   480
atccagtctg gaccagaac caccgcaccg tgagcctggc caagaagggtg atcattgggc   540
cctgggtgat ggctctgctc ctcaattgc cagttatcat tcgtgtgact acagtacctg   600
gtaaacggg gacagtagcc tgcactttta acttttcgcc ctggaccaac gaccctaaag   660
agaggataaa tgtggccggt gccatgttga cggtgagagg catcatccgg ttcattcattg   720
gcttcagcgc acccatgtcc atcgttctg tcaagttatgg gcttattgcc accaagatcc   780
acaagcaagg cttgattaag tccagtcgct ccttacgggt cctctccttt gtcgcagcag   840
ccttttttct ctgctggctc ccatatcagg tggtgccct tatagccaca gtcagaatcc   900
gtgagttatt gcaaggcatg tacaagaaa ttggtattgc agtggatgtg acaagtgcc   960
tggccttctt caacagctgc ctcaacccca tgctctatgt ctccatgggc caggacttcc 1020
gggagaggct gatccacgcc ctcccgcca gtctggagag ggccctgacc gaggactcaa 1080
cccaaaccag tgacacagct accaattcta ctttaccttc tgcagagggtg gagttacagg 1140
caaagtgagg agggagctgg gggacacttt cgagctccca gctccagctt cgtctcacct 1200
tgagttaggc tgagccacag gcatttcctg cttattttag gattaccac tcatcagaaa 1260
aaaaaaaaaa agcctttgtg tcccctgatt tggggagaat aaacagatat gagtttaaaa 1320
aaaaaaaaaa aaaa                                     1334

```

```

<210> SEQ ID NO 37
<211> LENGTH: 1404
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 37

```

```

ggcagtgacg ctgtgggaac ctctccacgc gcacgaactc agccaacgat ttctgataga   60
tttttgggag tttgaccaga gatgcaagggt gtgaaggagc gcttccacc gttagggaac   120
tctggggaca gagcgcgccg gccgcctgat ggccgaggca ggggtgcgacc caggaccag   180
gacggcgtcg ggaaccatac catggcccg atcccaaga ccctaaagt cgtcgtcgtc   240
atcgtcgcgg tctgctgcc agtccctagct tactctgcca cactgcccgc gaggaggaa   300
gttccccagc agacagtggc cccacagcaa cagaggcaca gctcaagggt ggaggagtgt   360
ccagcaggat ctcatagatc agaacatact ggagcctgta acccgtgcac agagggtgtg   420
gattacacca acgcttccaa caatgaacct tcttcttcc catgtacagt ttgtaaatca   480
gatcaaaaac ataaaagttc ctgcaccatg accagagaca cagtgtgtca gtgtaagaa   540
ggcaccttcc ggaatgaaaa ctccccagag atgtgccgga agtgtagcag gtgccctagt   600
ggggaagtcc aagtcagtaa ttgtacgtcc tgggatgata tccagtgtgt tgaagaattt   660
ggtgccaatg cactgtgga aaccccagct gctgaagaga caatgaacac cagcccgggg   720
actcctgccc cagctgctga agagacaatg aacaccagcc cagggactcc tgccccagct   780
gctgaagaga caatgaccac cagcccgggg actcctgccc cagctgctga agagacaatg   840
accaccagcc cggggactcc tgccccagct gctgaagaga caatgaccac cagcccgggg   900
actcctgccc cttctcatta cctctcatgc accatcgtag ggatcatagt tctaattgtg   960

```

-continued

```

cttctgattg tgtttgtttg aaagacttca ctgtggaaga aattccttcc ttacctgaaa 1020
ggttcaggta ggcgctggct gagggcgggg ggcgctggac actctctgcc ctgectccct 1080
ctgctgtgtt cccacagaca gaaacgcctg ccctgcccc aagtoctggg gtctccagcc 1140
tggtctatc ttectcttg tgatcgtccc atccccacat cccgtgcacc cccaggacc 1200
ctggtctcat cagtcctct cctggagctg ggggtccaca catctcccag ccaagtccaa 1260
gagggcaggg ccagttctc ccatcttcag gccagccag gcagggggca gtgggtcct 1320
caactgggtg acaagggtga ggatgagaag tggtcacggg atttattcag ccttggtcag 1380
agcagaaaaa aaaaaaaaaa aaaa 1404

```

```

<210> SEQ ID NO 38
<211> LENGTH: 454
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (410)
<223> OTHER INFORMATION: Wherein n can be a, c, t, or g

```

<400> SEQUENCE: 38

```

gtggcctggc ccttgaata gtagtgttta ggtagatgct tgtgtaggat tcctgataag 60
agcaactgaa aagaaggaga ggggaagtag taaagggaca agaacaatt ttttttttga 120
ggaaccataa gcaattata gtttgacaag acaagattgg gggacatata tggttaccag 180
ggaattaact cttatgtgtt atatctttat attatttato tctggaaaag agtaccctgc 240
aaaattccct acagctgcaa gcagatgtca cttgatggac agagggggaa ttctgcccct 300
ccggtatcgg gaaatacata ctaaagacat tgcgaaacgc tgaacctctt cccataaata 360
aaaggtttgt ttgtaaaatg gaaatccac ccataataaa tgaacaatan gcaactgccag 420
tttaggcctg ttcattgaatg gatctgcaag acag 454

```

```

<210> SEQ ID NO 39
<211> LENGTH: 1525
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

<400> SEQUENCE: 39

```

ggaattccca gccagcaaaa cagcagcact cagctaaaag gaagactcac agaacacagc 60
tgaagaagga aagtggcgat ggacctcacc ccaaatttgg cggtggaac ctggcttctc 120
ctggctgtca gcctgtgtct cctctatcta tatgggaccc gtacacatgg actttttaag 180
agactgggaa ttccagggcc cacacctctg cctttgttgg gaaatgtttt gtcctatcgt 240
cagggctctc gaaaatttga cacagagtgc tataaaaagt atggaaaaat gtggggtatc 300
tcttccctgt ttggaccaca ttaccttca tcatatgaag ccttgggtgg ctctgtgtg 360
agactcttgc tgtgtgtcac accctaatga actagaacct aaggttgctg tgtgtcgtac 420
aactagggaa cgtatgaagg tcaactccct gtgctggcca tcacagatcc cgacgtgatc 480
agaacagtgc tagtgaaga atgttattct gtcttcaca atcgaaggtc tttaggccca 540
gtgggattta tgaaaagtgc catctcttta gctgaggatg aagaatggaa gagaatacgg 600
tcattgtgtg ctccaacctt caccagcggg aaactcaagg agaaaagaca tcacaaaatt 660
cattacaaaa tgtcacttac tgctccatgc tggagaaagc catatccttc tgggacttga 720

```

-continued

```

gtctgcacat ttaactacag catctttggg gcctacagca tggatgtgat tactggcaca    780
tcatttggag tgaacatcga ctctctcaac aatccacaag acccctttgt ggagagcact    840
aagaagtcc taaatattgg tttcttagat ccattatttc tctcaataat actctttcca    900
ttccttacc cagtttttga agcattaaat gtctctctgt ttccaaaaga taccataaat    960
ttttaagta aatctgtaaa cagaatgaag aaaagtcgcc tcaacgacaa acaaaagcac   1020
cgactagatt tccttcagct gatgattgac tcccagaatt cgaagaaac tgagtcccac   1080
aaagctctgt ctgatctgga gctcgcagcc cagtcaataa tcttcatttt tgctggctat   1140
gaaaccacca gcagtgttct ttccttcaat ttatatgaac tggccactca ccctgatgtc   1200
cagcagaaac tgcaaaagga gattgatgca gttttgcccataaaggtgag gggatgacct   1260
ctggagatga agggaagagg tgaagcctta gcaaaaatgc ctcctacca ctcccagga   1320
gaatttttat aaaaagcata atcaactgatt ccttcaactga cataatgtag gaagcctctg   1380
aggagaaaaa caaagggaga aacatagaga acggttgcta ctggcagaag cataagatct   1440
ttgtacaata ttgctggccc tggttcacct gtttactgtt atcacaataa tgctaagtaa   1500
aaaaaaaaaa aaaaaaaaaa aaaaaa                                         1525

```

```

<210> SEQ ID NO 40
<211> LENGTH: 684
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 40

```

```

cggcctcac ttatcattat taaactctaa catgtgttga tcaactaaaa ctctcgagtt    60
gtactgtgtg acatgtgata gctaccctat ttcattactt catttatccc ctcgaggctt   120
ggagactgat ggtgagggga ggagtctgag ggctcctggg ccctcgccca tgggagctgg   180
gtgcgtggct gtcctgtttg taggacaagc tacagggaga gactgtgttt ccttggccat   240
gtcccgtggg gccagcatg atgtcctgag tagcagaaa atgggtgagg gatgggctgc   300
ctggatgatg actggctgat gcattttctg ccctcagttt tataagtgag actgaacgag   360
ctagaaggaa tgtgaagtcc agagttagaca atgatgacaa ccattttatt aaaaaaaaaat   420
agtggtctact atgtgttagc tgctgtgcta ctaacaggca ctggtttttg ccagtgatct   480
atattacaag geatggccat cctctcctgg gcttccagcc tagtggggag ggctagctgg   540
tcaacagaca gttgtgtgag agcaggtgca gggatgatc acagcacaga gaagcaaccc   600
atccagcat ggggaccaga gaggaacatt actgatggaa gtctaaaacta ggacctaaaa   660
gatcaatggg aaccagctgg ccaa                                         684

```

```

<210> SEQ ID NO 41
<211> LENGTH: 465
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 41

```

```

caatatctga caccactttg gactcaagag actcagtaac gtattatcct gtttatttag    60
cttggtttta gctgtgttct ctctggataa cccacttgat gtttagaaca ttacttctct   120
gcttattcca tattaatact gtgttaggta ttttaagaag caagttatta aataagaaaa   180
gtcaaagtat taattcttac cttctattat cctatattag cttcaatata tccaaaccaa   240

```

-continued

```

atggctgtta ggtagattta tttttatata agcatgttta ttttgatcag atgttttaac    300
ttggatttga aaaaatacat ttatgagatg ttttataaga tgtgtaaata tagaactgta    360
tttattacta tagtaaaggt tcagtaacat taaggacat gataatgata ataaccttg    420
tacagtggca tattctttga tttatattgt gtttctctgc ccatt                    465

```

```

<210> SEQ ID NO 42
<211> LENGTH: 371
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: unsure
<222> LOCATION: (142) (143) (144) (145) (146) (147) (148) (150) (151) (152) (153)
(154) (155) (156) (157) (158) (159) (160) (161)
<223> OTHER INFORMATION: Wherein n can be a, c, t, or g

```

```

<400> SEQUENCE: 42
tcaccttcac cttctaacta actagcctcc ggatgagggtg gctgccacca ggcccgaatg    60
atccccagga gccccagcttc caaaccccaa catcgaatca aacatctcca tcccgaagtg    120
cagtaacaca caaaaaccaa annnnnnnan nnnnnnnnnn nccctgggaa aggcctggtg    180
cgattctcag taggactcac acccacccta cctagaagta ctgggctggc ctgggtactg    240
catcctgtgt ttttgataag ggggtgatgt ggccacgccc ttatctagat ttcactttgt    300
atccactggg cacagatatt ctagagaact tatctttcac tcttgtaaaa gccacatata    360
cacatctctt t                                                    371

```

```

<210> SEQ ID NO 43
<211> LENGTH: 546
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 43
atgtgcaaag gacttgcagc tttgcccac tcatgcctgg aaagggccaa ggagattaag    60
atcaagttgg gaattctcct ccagaagcca gactcagttg gtgacctgt cattccgtac    120
aatgagaagc cagagaaacc agccaagacc cagaaaacct cgctggacga ggccctgcag    180
tggcgtgatt cctctggaaa actcctgcag aacaactatg gacttgccag tttcaaaagt    240
ttcctgaagt ctgaattcag tgaggaaaac cttgagttct ggattgcctg tgaggattac    300
aagaagatca agtcccctgc caagatggct gagaaggcaa agcaaattta tgaagaattc    360
attcaaacgg aggcctctaa agaggtgaat attgaccact tcaactaagga catcacaatg    420
aagaacctgg tggaaccttc cctgagcagc tttgacatgg ccagaaaag aatccatgcc    480
ctgatggaaa aggattctct gcctcgtctt gtgcgctctg agttttatca ggagttaatc    540
aagtag                                                    546

```

What is claimed:

1. A method for prognostic or diagnostic assessment of a gastrointestinal-related disorder in a subject, comprising:

- a) preparing a sample of nucleic acids from a specimen obtained from the subject;
- b) contacting the sample with a panel of nucleic acid segments consisting of at least 2 genes represented by nucleic acids from the group consisting of SEQ ID NOS: 1-43 to detect the levels of the panel segments;

c) evaluating the sample against a reference standard to determine the magnitude of change in the amounts of at least 2 members present in the sample; and

d) correlating the magnitude of change with the presence or resolution of the gastrointestinal-related disorder.

2. The method of claim 1, wherein the subject is a patient having a gastrointestinal-related disorder and steps a) through d) are performed before, during, and/or after treatment of the patient with a therapy for the gastrointestinal-related disorder.

3. The method of claim 2, wherein steps a) through d) are performed during treatment of the patient with a therapy for the gastrointestinal-related disorder and about 30 weeks after commencement of treatment.

4. The method of claim 2, wherein the gastrointestinal-related disorder is ulcerative colitis.

5. The method of claim 2, wherein the reference standard is from the group consisting of colon biopsy from a normal patient, colon biopsy from an untreated ulcerative colitis patient, and colon biopsy from a treated ulcerative colitis patient.

6. The method of claim 2, wherein the reference standard is from the subject prior to treatment with a therapy, the sample of nucleic acids is from the subject after treatment with a therapy, and the correlating step evaluates the effectiveness of treatment with the therapy.

7. The method of claim 2, wherein the therapy is an anti-TNF α antibody.

8. The method of claim 1, wherein the collection is an array of nucleic acid segments.

9. The method of claim 2, wherein the sample is from a colon biopsy of a patient selected from the group consisting of patients suspected of having ulcerative colitis, patients diagnosed with ulcerative colitis undergoing treatment with an approved agent, and patients diagnosed with ulcerative colitis undergoing treatment with an experimental agent.

10. The method of claim 2, wherein the sample is from a source selected from the group consisting of a patient providing the sample prior to administration of a therapy, a placebo treated patient having a gastrointestinal-related disorder, and a sample from a biobank.

11. The method of claim 1, wherein the at least one gene from the collection is selected from the group consisting of cytokines, chemokines, transcription factors, proteases, protease inhibitors, structural and adhesion molecules, and genes for proteins involved in lipid metabolism.

12. The method of claim 1, wherein the sample comprises a colon biopsy sample.

13. The method of claim 1, wherein the sample comprises peripheral blood cells.

14. The method of claim 1, wherein the sample is contacted with a panel of nucleic acid segments comprising at least 4 members from the group consisting of SEQ ID NOS: 1-43.

15. The method of claim 13, wherein the at least four nucleic acid segments are representative of or selected from an innate or adaptive immune response-related gene selected from the group consisting of SEQ ID NOS: 1, 7, 10-13, 15-18, 21, 33, and 35; a cell-cell interaction, cell-matrix interaction or matrix regulation-related gene selected from the group consisting of SEQ ID NOS: 2, 28, and 32; a cell-cell, intracellular signaling pathway-related gene selected from the group consisting of SEQ ID NOS: 4, 8, 22, 26, 27, 30, 36, 41, and 43; a cell growth and apoptosis-related gene selected from the group consisting of SEQ ID NOS: 25 and 37; a protein regulation-related gene selected from the group consisting of SEQ ID NOS: 3 and 39; a metabolic regulation-related gene selected from the group consisting of SEQ ID NOS: 5, 14, 20, 24, and 29; and a cytoskeleton organization-related gene of SEQ ID NO: 34; a developmental regulation-related gene of SEQ ID NO: 9; and a transcriptional regulation-related gene of SEQ ID NO: 19.

16. The method of claim 1, wherein at least one of the at least two nucleic acid segments is representative of or selected from the group consisting of SEQ ID NOS: 1, 7, 10-13, 15-18, 21, 33, and 35.

17. The method of claim 1, wherein the at least two gene segments are representative of or selected from the group consisting of SEQ ID NOS: SEQ ID NOS: 1, 7, 10-13, 15-18, 21, 33, and 35 and SEQ ID NOS: 25 and 37.

18. A method for prognostic or diagnostic assessment of a gastrointestinal-related disorder in a subject, comprising:

- a) preparing a sample of nucleic acids from a sample obtained from a patient;
- b) contacting the sample with a panel of nucleic acid segments consisting of at least one member represented by nucleic acids from the group consisting of SEQ ID NOS: 1, 7, 10-13, 15-18, 21, 33, and 35 to detect the presence of the panel segments;
- c) evaluating the sample against a reference standard to determine the change and/or magnitude of change-in the expression level of the amounts of the at least one member present in the sample; and
- d) correlating the change and/or magnitude of expression level with the presence or resolution of the gastrointestinal-related disorder.

19. An array-based testing method for prognostic or diagnostic assessment of a gastrointestinal-related disorder in a patient, comprising:

- a) preparing a mixture of nucleic acids from a specimen obtained from a patient;
- b) labeling said specimen nucleic acids with a detectable marker to form a sample;
- c) contacting the sample with an array comprising a plurality of nucleic acid segments, wherein each nucleic acid segment is immobilized to a discrete and known address on a substrate surface of the array, wherein at least two members of a gastrointestinal-related gene panel represented by nucleic acids consisting of SEQ ID NOS: 1-43 are identified as features of the array by address, and wherein said array further comprises at least one calibration nucleic acid at a known address on the substrate;
- d) determining the degree of binding of the specimen nucleic acids to the nucleic acid segments; and
- e) comparing the degree of binding to a reference standard to enable a prognostic or diagnostic assessment.

20. The method of claim 18, further comprising the step of performing a statistical comparison of the specimen nucleic acids from gastrointestinal-related disorder patients treated with a therapy to a reference standard to evaluate the effect of treatment with the therapy.

21. The method of claim 19, wherein the gastrointestinal-related disorder is ulcerative colitis and the gastrointestinal-related gene panel is an ulcerative colitis-related gene panel.

22. The method of claim 19, wherein the therapy is an anti-TNF α antibody.

23. The method of claim 18, wherein the specimen is from a colon biopsy of a patient selected from the group consisting of patients suspected of having ulcerative colitis, patients diagnosed with ulcerative colitis not undergoing treatment, and patients diagnosed with ulcerative colitis undergoing treatment with a therapy.

24. The method of claim 18, wherein the specimen is from a source selected from the group consisting of a patient providing the specimen prior to administration of a therapy, a

patient having a similar disease or condition treated with a placebo, and a sample from a biobank.

25. The method of claim **18**, wherein the members of the gene panel are selected from the group consisting of cytokines, chemokines, transcription factors, proteases, protease inhibitors, structural and adhesion molecules, and genes for proteins involved in lipid metabolism.

26. The method of claim **18**, wherein the specimen comprises a colon biopsy sample.

27. The method of claim **18**, wherein the specimen comprises peripheral blood cells.

28. The method of claim **20**, wherein the comparing the degree of binding step further comprises a stringent test of the similarity of feature intensity changes of the array of the ulcerative colitis-related gene panel.

29. A reagent for testing the responsiveness of a cell or subject to a therapy for a gastrointestinal-related disorder, comprising at least one member selected from the group consisting of an oligonucleotide comprising at least 15 nucleotides complementary to a nucleotide sequence of one of SEQ ID NOS: 1-43, a polypeptide encoded by at least a portion of one of SEQ ID NOS: 1-43, and a ligand for the polypeptide encoded by at least a portion of one of SEQ ID NOS: 1-43.

30. The reagent of claim **28**, wherein the gastrointestinal-related disorder is ulcerative colitis.

31. A method of testing for responsiveness to a therapy for a gastrointestinal-related disorder in a patient sample comprising contacting the patient sample with the reagent of claim **28** and comparing the levels of at least a portion of one of the genes or proteins of SEQ ID NOS: 1-43 to a reference standard.

32. The method of claim **30**, wherein the testing is done by RT-PCR.

33. The method of claim **30**, wherein the testing is done by ELISA.

34. A method of testing the effectiveness of a therapy for a gastrointestinal-related disorder, comprising:

- a. contacting a sample from a patient being treated for the gastrointestinal-related disorder with the reagent of claim **28**;
- b. measuring levels of the at least one member; and
- c. correlating the levels of the at least one member with the effectiveness of the therapy.

35. The method of claim **33**, wherein the correlating step comprises comparing the levels with levels of the at least one member of a sample from the patient prior to treatment with the therapy and wherein a decrease of at least about 2-fold in the level of the at least one member from the patient being treated versus the patient prior to treatment indicates a responder to the therapy.

36. The method of claim **34**, wherein the gastrointestinal-related disorder is ulcerative colitis.

37. The method of claim **34**, wherein the therapy comprises an antagonist of TNF α .

38. The method of claim **36**, wherein the antagonist is an antibody to TNF α .

39. The method of claim **37**, wherein the antibody to TNF α is infliximab.

40. A kit for prognostic or diagnostic use, comprising an oligonucleotide comprising at least 15 nucleotides complementary to a polynucleotide comprising the nucleotide sequence of a marker gene or the complementary strand thereof and cells expressing the marker gene, wherein the marker gene is represented by nucleic acids selected from the group consisting of SEQ ID NOS: 1-43.

41. A kit for screening for a therapeutic agent for UC, the kit comprising an antibody which recognizes a peptide comprising an amino acid sequence encoded by a marker gene and cells expressing the marker gene, wherein the marker gene is represented by nucleic acids selected from the group consisting of SEQ ID NOS: 1-43.

42. A method of testing the effectiveness of a therapy for ulcerative colitis, comprising:

- a) contacting a sample from a patient being treated for ulcerative colitis with at least two members of the reagent of claim **28**;
- b) measuring levels of the at least two members; and
- c) correlating the levels of the at least two members with the effectiveness of the therapy.

43. The method of claim **41**, wherein the correlating step comprises comparing the levels with levels of at least two members of a sample from the patient prior to treatment with the therapy and wherein a decrease of at least about 2-fold in the level of the at least two members from the patient being treated versus the patient prior to treatment indicates a responder to the therapy.

44. The method of claim **42**, wherein the therapy comprises an antagonist of TNF α .

45. The method of claim **43**, wherein the antagonist is an antibody to TNF α .

46. The method of claim **44**, wherein the antibody to TNF α is infliximab.

47. A method for prognostic or diagnostic assessment of a gastrointestinal-related disorder in a subject, comprising:

- a) preparing a sample of nucleic acids from a specimen obtained from the subject;
- b) contacting the sample with a panel of nucleic acid segments consisting of at least 2 members from the group of genes represented by nucleic acids selected from the group consisting of SEQ ID NOS: 1-43 to detect the levels of the panel segments;
- c) evaluating the sample against a reference standard to determine the magnitude of change in the amounts of the at least 2 members present in the sample; and
- d) correlating the magnitude of change with the presence or resolution of the gastrointestinal-related disorder.

48. The method of claim **47**, wherein the subject is a patient having a gastrointestinal-related disorder and steps a) through d) are performed before, during, and/or after treatment of the patient with a therapy for the gastrointestinal-related disorder.

49. Any invention described herein.

* * * * *

专利名称(译)	使用43 Gene Panel评估和治疗溃疡性结肠炎和相关疾病的标记和方法		
公开(公告)号	US20080293582A1	公开(公告)日	2008-11-27
申请号	US11/847737	申请日	2007-08-30
[标]申请(专利权)人(译)	李希林 宋肖宇		
申请(专利权)人(译)	李希林 宋小瑜		
当前申请(专利权)人(译)	李希林 宋小瑜		
[标]发明人	LI XILIN SONG XIAO YU		
发明人	LI, XILIN SONG, XIAO-YU		
IPC分类号	C40B30/04 C40B30/00 C12Q1/68 G01N33/53		
CPC分类号	C12Q1/6883 C12Q2600/106 C12Q2600/136 C12Q2600/158		
优先权	60/823976 2006-08-30 US		
外部链接	Espacenet USPTO		

摘要(译)

用于对受试者中的胃肠相关病症（例如溃疡性结肠炎）进行预后或诊断评估的方法将样品中基因的存在，不存在和/或量级与参考标准相关联以确定存在和/或严重性。疾病的治疗方法，和/或对疾病治疗的反应。该方法能够识别候选疗法的有效性。

$$(\text{Signal intensity of probe set } A \text{ in sample } X)$$

$$(\text{Median intensity of all measurements in sample } X) \times$$

$$(\text{Median intensity of probe set } A \text{ across all week-0 samples})$$