



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2634146 A1 2007/06/28

(21) **2 634 146**

(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2006/12/19
(87) Date publication PCT/PCT Publication Date: 2007/06/28
(85) Entrée phase nationale/National Entry: 2008/06/19
(86) N° demande PCT/PCT Application No.: US 2006/048246
(87) N° publication PCT/PCT Publication No.: 2007/073478
(30) Priorité/Priority: 2005/12/19 (US60/751,420)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01),
G06F 19/00 (2006.01)

(71) Demandeur/Applicant:
GENIZON BIOSCIENCES INC., CA

(72) Inventeurs/Inventors:
BELOUCHI, ABDELMAJID, CA;
RAELSON, JOHN VERNER, CA;
BRADLEY, WALTER EDWARD, CA;
PAQUIN, BRUNO, CA;
FOURNIER, HELENE, CA;
NGUYEN-HUU, QUYNH, CA;
CROTEAU, PASCAL, CA;

...

(74) Agent: OGILVY RENAULT LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : CARTE GENETIQUE DU GENE HUMAIN ASSOCIE A LA MALADIE DE CROHN

(54) Title: GENEMAP OF THE HUMAN GENES ASSOCIATED WITH CROHN'S DISEASE

(57) Abrégé/Abstract:

Published without an Abstract



(72) **Inventeurs(suite)/Inventors(continued)**: ALLARD, RENE, CA; BRIAND, SANDIE, CA; VAN EERDEWEGH, PAUL, CA;
LITTLE, RANDALL DAVID, CA; SEGAL, JONATHAN, CA; KEITH, TIM, CA

GENEMAP OF THE HUMAN GENES ASSOCIATED WITH CROHN'S DISEASE

INVENTORS: Abdelmajid Belouchi, John Verner Raelson, Walter Edward Bradley, Bruno Paquin, Helene Fournier, Quynh Nguyen-Huu, Pascal Croteau, Rene Allard, Sandie Briand, Paul Van Eerdewegh, Randall David Little, Jonathan Segal and Tim
5 Keith.

FIELD OF THE INVENTION

The invention relates to the field of genomics and genetics, including genome analysis and the study of DNA variations. In particular, the invention relates to the fields of
10 pharmacogenomics, diagnostics, patient therapy and the use of genetic haplotype information to predict an individual's susceptibility to inflammatory bowel disease (IBD), e.g. Crohn's disease and/or their response to a particular drug or drugs, so that drugs tailored to genetic differences of population groups may be developed and/or administered to the appropriate population.

15 The invention also relates to a GeneMap for IBD (e.g. Crohn's disease), which links variations in DNA (including both genic and non-genic regions) to an individual's susceptibility to IBD and/or response to a particular drug or drugs. The invention further relates to the genes disclosed in the GeneMap (see Tables 2-4), which is related to methods and reagents for detection of an individual's increased or decreased risk for
20 Crohn's disease by identifying at least one polymorphism in one or a combination of the genes from the GeneMap. Also related are the candidate regions identified in Table 1, which are associated with IBD. In addition, the invention further relates to nucleotide sequences of those genes including genomic DNA sequences, cDNA sequences, single nucleotide polymorphisms (SNPs), other types of polymorphisms (insertions, deletions,
25 microsatellites), alleles and haplotypes (see Sequence Listing and Tables 5-36).

The invention further relates to isolated nucleic acids comprising these nucleotide sequences and isolated polypeptides or peptides encoded thereby. Also related are expression vectors and host cells comprising the disclosed nucleic acids or fragments thereof, as well as antibodies that bind to the encoded polypeptides or peptides.

30 The present invention further relates to ligands that modulate the activity of the disclosed genes or gene products. In addition, the invention relates to diagnostics and therapeutics

for IBD (in particular UC and Crohn's disease), utilizing the disclosed nucleic acids, polymorphisms, chromosomal regions, gene maps, polypeptides or peptides, antibodies and/or ligands and small molecules that activate or repress relevant signaling events.

5 BACKGROUND OF THE INVENTION

Inflammatory bowel disease (IBD) is a collective term used to describe two intestinal disorders whose etiology is not completely understood: ex: Crohn's disease (CD) and ulcerative colitis (UC). The course and prognosis of IBD, which occurs worldwide and afflicts several million people, varies widely. Onset of IBD is predominant in young adulthood. Symptoms of IBD include abdominal cramps and pain, diarrhea, weight loss and intestinal bleeding. Anemia and weight loss are also common signs of IBD. Between 10% and 15% of people with IBD require surgery over a ten-year period. Patients with IBD are also at increased risk for the development of intestinal cancer. These diseases are accompanied by a high frequency of psychological symptoms, including anxiety and depression.

There are common features in many of the later stages of IBD. Inflammation at the disease site/target organ is typically present, caused by the release of inflammatory (also termed "proinflammatory") cytokines by T cells and by other cells that contribute to the activation steps and effector pathways of immune/inflammatory processes. The current consensus opinion regarding the pathogenesis of IBD centers on the role of genetically determined dysregulation in the host immune response toward the resident bacterial flora.

UC involves the rectum and spreads proximally to contiguous portions or to the entire colon. Disease activity is usually intermittent, with relapses and periods of quiescence. The sigmoidoscopic or colonoscopic picture is characteristic. In mild disease, the colonic mucosa appears hyperemic and granular. In more severe disease, tiny punctuate ulcers are present and the mucosa is characteristically friable and may bleed spontaneously. Histologically, the inflammatory cell infiltrate in active disease usually includes neutrophils, often invading crypts as well as being associated with epithelial damage and crypt distortion. An increased number of lymphocytes in the lamina propria and basal plasmacytosis are usually present. Between 500,000 and 700,000 patients suffer from UC in the United States. Extra-colonic manifestations of UC include arthritis, uveitis,

aphthous stomatitis, pyoderma gangrenosum, and erythema nodosum. Initial therapy for patients with mild to moderate disease is usually an aminosalicylate. In controlled trials, disease improvement by various criteria occurred in up to 30% of subjects in the placebo groups; thus, no specific treatment may be an option for patients with very mild disease.

5 In patients with active UC who do not respond to standard 5-ASA treatment and in those with more severe disease, oral corticosteroids have been the mainstay of acute symptomatic therapy. However, corticosteroids are not effective in long-term maintenance of remission in patients with UC given that their use is associated with significant toxicity over time. Although the pathogenesis of UC is not fully understood,

10 there is increasing evidence that UC may be an autoimmune disorder, with B cells playing a role in disease pathophysiology. B cells, as well as T cells, are present in basal lymphoid aggregates, a histopathologic feature considered indicative of UC and seen in histologic sections from patients with active UC. Whereas mucosal inflammation in UC is thought to be driven by activated T cells, these patients have a T-helper-2 (Th2) cytokine

15 expression pattern profile. As Th2 cytokines classically drive B-cell immune responses and antibody production, a central role for B cell may be postulated in UC.

Crohn's disease is an Inflammatory Bowel Disease (IBD) in which inflammation extends beyond the inner gut lining and penetrates deeper layers of the intestinal wall of any part of the digestive system (esophagus, stomach, small intestine, large intestine, and/or

20 anus). Crohn's disease is a chronic, lifelong disease which can cause painful, often life altering symptoms including diarrhea, cramping and rectal bleeding. Crohn's disease occurs most frequently in the industrialized world and the typical age of onset falls into two distinct ranges, 15 to 30 years of age and 60 to 80 years of age. The highest mortality is during the first years of disease, and in cases where the disease symptoms

25 are long lasting, an increased risk of colon cancer is observed. Crohn's disease presently accounts for approximately two thirds of IBD-related physician visits and hospitalizations, and 50 to 80% of Crohn's disease patients eventually require surgical treatment. Development of Crohn's disease is influenced by environmental and host specific factors, together with "exogenous biological factors" such as constituents of the

30 intestinal flora (the naturally occurring bacteria found in the intestine). It is believed that in genetically predisposed individuals, exogenous factors such as infectious agents, and host-specific characteristics such as intestinal barrier function and/or blood supply, combine with specific environmental factors to cause a chronic state of improperly regulated immune system function. In this hypothetical model, microorganisms trigger an

35 immune response in the intestine, and in susceptible individuals, this immune response

is not turned off when the microorganism is cleared from the body. The chronically "turned on" immune response causes damage to the intestine resulting in the symptoms of Crohn's disease.

Current treatments for Crohn's disease are primarily aimed at reducing symptoms by suppressing inflammation and do not address the root cause of the disease. Despite a preponderance of evidence showing inheritance of a risk for Crohn's disease through epidemiological studies and genome wide linkage analyses, the genes affecting Crohn's disease have yet to be discovered (Hugot JP, and Thomas G., 1998). There is a need in the art for identifying specific genes related to Crohn's disease to enable the development of therapeutics that address the causes of the disease rather than relieving its symptoms. The failure in past studies to identify causative genes in complex diseases, such as Crohn's disease, has been due to the lack of appropriate methods to detect a sufficient number of variations in genomic DNA samples (markers), the insufficient quantity of necessary markers available, and the number of needed individuals to enable such a study.

Unfortunately, new therapies for IBD are few, and both diagnosis and treatment have been hampered by a lack of detailed knowledge of the etiology. Despite the progress noted above, there remains a need in the art for new and improved methods for treating this debilitating group of diseases, and the present inventors have made a significant step forward with the invention disclosed herein. The invention includes a method for treating IBD and inflammatory diseases in a subject, comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition that comprises (a) a compound that inhibits inflammation; and (b) a pharmaceutically acceptable carrier, thereby treating the disease.

The DNA sequences between two human genomes are 99.9% identical. The variations in DNA sequence between individuals can be, as an example, deletions of small or large stretches of DNA, insertions of stretches of DNA, variations in the number of repetitive DNA elements, and changes in single base positions in the genome called "single nucleotide polymorphisms" (SNPs). Human DNA sequence variation accounts for a large fraction of observed differences between individuals, including susceptibility to disease.

Many common diseases, like IBD, are complex genetic traits and are believed to involve several disease-genes rather than single genes, as is observed for rare diseases. This makes detection of any particular gene substantially more difficult than in a rare disease,

where a single gene mutation that segregates according to a Mendelian inheritance pattern is the causative mutation. Any one of the multiple interacting gene mutations involved in the etiology of a complex disease will impart a lower relative risk for the disease than will the single gene mutation involved in a simple genetic disease. Low
5 relative risk alleles are more difficult to detect and, as a result, the success of positional cloning using linkage mapping that was achieved for simple genetic disease genes has not been repeated for complex diseases.

Several approaches have been proposed to discover and characterize multiple genes in complex genetic traits. These gene discovery methods can be subdivided into
10 hypothesis-free disorder association studies and hypothesis-driven candidate gene or region studies. The candidate gene approach relies on the analysis of a gene in patients who have a disorder in which the gene is thought to play a role. This approach is limited in utility because it only provides for the investigation of genes with known functions. Although variant sequences of candidate genes may be identified using this approach, it
15 is inherently limited by the fact that variant sequences in other genes that contribute to the phenotype will be necessarily missed when the technique is employed. Genome-wide scans (GWS) have been shown to be efficient in identifying disease genes, such as Crohn's disease susceptibility genes (NOD2/CARD15 and OCTN). In contrast to the candidate gene approach, a GWS searches throughout the genome without any a priori
20 hypothesis and consequently can identify genes that are not obvious candidates for the disease as well as genes that are relevant candidates for the disease it can also identify chromosomal regions that are structurally important where mutations can influence gene function of specific genes.

Family-based linkage mapping methods were initially used for disorder locus
25 identification. This technique locates genes based on the relatively limited number of genetic recombination events within the families used in the study, and results in large chromosomal regions containing hundreds of genes, any one of which could be the disorder-causing gene. Population-based, or linkage disequilibrium (LD) mapping is based on the premise that regions adjacent to a gene of interest are co-transmitted
30 through the generations along with the gene. As a result, LD extends over shorter genetic regions than does linkage (Hewett *et al.*, 2002), and can facilitate detection of genes with lower relative risk than family linkage mapping approaches. LD-based mapping also defines much smaller candidate regions, which may contain only a few genes, making the identification of the actual disorder gene much easier.

It has been estimated that a GWS that uses a general population and case/control association (LD) analysis would require approximately 700,000 SNP markers (Carlson *et al.*, 2003). The cost of a GWS at this marker density for a sufficient sample size for statistical power is economically prohibitive. The use of a special founder population (genetic isolate), such as the French Canadian population of Quebec, is one solution to the problem with LD analysis. The French Canadian population in Quebec (Quebec Founder Population - QFP) provides one of the best resources in the world for gene discovery based on its high levels of genetic sharing and genetic homogeneity. By combining DNA collected from the QFP, high throughput genotyping capabilities and proprietary algorithms for genetic analysis, a comprehensive genome-wide association study was facilitated. The present invention relates specifically to a set of IBD (E.g., Crohn's disease)-causing genes (GeneMap) and targets, which present attractive points of therapeutic intervention.

In view of the foregoing, identifying susceptibility genes associated with IBD (e.g. Crohn's disease) and their respective biochemical pathways will facilitate the identification of diagnostic markers as well as novel targets for improved therapeutics. It will also improve the quality of life for those afflicted by this disease and will reduce the economic costs of these afflictions at the individual and societal level. The identification of those genetic markers would provide the basis for novel genetic tests and eliminate or reduce the therapeutic methods currently used. The identification of those genetic markers will also provide the development of effective therapeutic intervention for the battery of laboratory, radiological, and endoscopic evaluations typically required to diagnose IBD. The present invention satisfies this need and provides related advantages as well.

25

DESCRIPTION OF THE FILES CONTAINED ON THE CD-R

The contents of the submission on compact discs submitted herewith are incorporated herein by reference in their entirety: A compact disc copy of the Sequence Listing (COPY 1) (filename: GENI 011 01WO SeqList.txt, date recorded: December 19, 2006, file size 39,614,000 bytes); a duplicate compact disc copy of the Sequence Listing (COPY 2) (filename: GENI 011 01WO

SeqList.txt, date recorded: December 19, 2006, file size 39,614,000 bytes); a duplicate compact disc copy of the Sequence Listing (COPY 3) (filename: GENI 011 01WO SeqList.txt, date recorded: December 19, 2006, file size 39,614,000 bytes); a computer readable format copy of the Sequence Listing (CRF COPY)
5 (filename: GENI 011 01WO SeqList.txt, date recorded: December 19, 2006, file size 39,614,000 bytes).

DEFINITIONS

10 Throughout the description of the present invention, several terms are used that are specific to the science of this field. For the sake of clarity and to avoid any misunderstanding, these definitions are provided to aid in the understanding of the specification and claims:

Allele: One of a pair, or series, of forms of a gene or non-genic region that occur at a given locus in a chromosome. Alleles are symbolized with the same basic symbol (e.g.,
15 B for dominant and b for recessive; B1, B2, Bn for n additive alleles at a locus). In a normal diploid cell there are two alleles of any one gene (one from each parent), which occupy the same relative position (locus) on homologous chromosomes. Within a population there may be more than two alleles of a gene. See multiple alleles. SNPs also have alleles, *i.e.*, the two (or more) nucleotides that characterize the SNP.

20 Amplification of nucleic acids: refers to methods such as polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known in the art and are described, for example, in U.S. Patent Nos. 4,683,195 and 4,683,202. Reagents and hardware for conducting PCR are commercially available. Primers useful for amplifying
25 sequences from the disorder region are preferably complementary to, and preferably hybridize specifically to, sequences in the disorder region or in regions that flank a target region therein. Genes from Tables 2-4 generated by amplification may be sequenced directly. Alternatively, the amplified sequence(s) may be cloned prior to sequence analysis.

30 Antigenic component: is a moiety that binds to its specific antibody with sufficiently high affinity to form a detectable antigen-antibody complex.

Antibodies: refer to polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, that can bind to proteins and fragments thereof or to nucleic acid sequences from the disorder region, particularly from the disorder gene products or a portion thereof. The term antibody is used both to refer to a homogeneous
5 molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Proteins may be prepared synthetically in a protein synthesizer and coupled to a carrier molecule and injected over several months into rabbits. Rabbit sera are tested for immunoreactivity to the protein or fragment. Monoclonal antibodies may be made by injecting mice with the proteins, or fragments thereof. Monoclonal antibodies
10 can be screened by ELISA and tested for specific immunoreactivity with protein or fragments thereof (Harlow *et al.* 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). These antibodies will be useful in developing assays as well as therapeutics.

Associated allele: refers to an allele at a polymorphic locus that is associated with a
15 particular phenotype of interest, *e.g.*, a predisposition to a disorder or a particular drug response.

cDNA: refers to complementary or copy DNA produced from an RNA template by the action of RNA-dependent DNA polymerase (reverse transcriptase). Thus, a cDNA clone means a duplex DNA sequence complementary to an RNA molecule of interest, included
20 in a cloning vector or PCR amplified. This term includes genes from which the intervening sequences have been removed.

cDNA library: refers to a collection of recombinant DNA molecules containing cDNA inserts that together comprise essentially all of the expressed genes of an organism or tissue. A cDNA library can be prepared by methods known to one skilled in the art (see,
25 *e.g.*, Cowell and Austin, 1997, "DNA Library Protocols," *Methods in Molecular Biology*). Generally, RNA is first isolated from the cells of the desired organism, and the RNA is used to prepare cDNA molecules.

Cloning: refers to the use of recombinant DNA techniques to insert a particular gene or other DNA sequence into a vector molecule. In order to successfully clone a desired
30 gene, it is necessary to use methods for generating DNA fragments, for joining the fragments to vector molecules, for introducing the composite DNA molecule into a host cell in which it can replicate, and for selecting the clone having the target gene from amongst the recipient host cells.

Cloning vector: refers to a plasmid or phage DNA or other DNA molecule that is able to replicate in a host cell. The cloning vector is typically characterized by one or more endonuclease recognition sites at which such DNA sequences may be cleaved in a determinable fashion without loss of an essential biological function of the DNA, and
5 which may contain a selectable marker suitable for use in the identification of cells containing the vector.

Coding sequence or a protein-coding sequence: is a polynucleotide sequence capable of being transcribed into mRNA and/or capable of being translated into a polypeptide or peptide. The boundaries of the coding sequence are typically determined by a translation
10 start codon at the 5'-terminus and a translation stop codon at the 3'-terminus.

Complement of a nucleic acid sequence: refers to the antisense sequence that participates in Watson-Crick base-pairing with the original sequence.

Disorder region: refers to the portions of the human chromosomes displayed in Table 1 bounded by the markers from Tables 2-36.

15 Disorder-associated nucleic acid or polypeptide sequence: refers to a nucleic acid sequence that maps to region of Table 1 or the polypeptides encoded therein (Tables 2-4, nucleic acids, and polypeptides). For nucleic acids, this encompasses sequences that are identical or complementary to the gene sequences from Tables 2-4, as well as sequence-conservative, function-conservative, and non-conservative variants thereof.
20 For polypeptides, this encompasses sequences that are identical to the polypeptide, as well as function-conservative and non-conservative variants thereof. Included are the alleles of naturally-occurring polymorphisms causative of IBD such as, but not limited to, alleles that cause altered expression of genes of Tables 2-4 and alleles that cause altered protein levels or stability (e.g., decreased levels, increased levels, expression in
25 an inappropriate tissue type, increased stability, and decreased stability).

Expression vector: refers to a vehicle or plasmid that is capable of expressing a gene that has been cloned into it, after transformation or integration in a host cell. The cloned gene is usually placed under the control of (*i.e.*, operably linked to) a regulatory sequence.

30 Function-conservative variants: are those in which a change in one or more nucleotides in a given codon position results in a polypeptide sequence in which a given amino acid residue in the polypeptide has been replaced by a conservative amino acid substitution.

Function-conservative variants also include analogs of a given polypeptide and any polypeptides that have the ability to elicit antibodies specific to a designated polypeptide.

5 Founder population: Also called a population isolate, this is a large number of people who have mostly descended, in genetic isolation from other populations, from a much smaller number of people who lived many generations ago.

Gene: Refers to a DNA sequence that encodes through its template or messenger RNA a sequence of amino acids characteristic of a specific peptide, polypeptide, or protein. The term "gene" also refers to a DNA sequence that encodes an RNA product. The term gene as used herein with reference to genomic DNA includes intervening, non-coding
10 regions, as well as regulatory regions, and can include 5' and 3' ends. A gene sequence is wild-type if such sequence is usually found in individuals unaffected by the disorder or condition of interest. However, environmental factors and other genes can also play an important role in the ultimate determination of the disorder. In the context of complex disorders involving multiple genes (oligogenic disorder), the wild type, or normal
15 sequence can also be associated with a measurable risk or susceptibility, receiving its reference status based on its frequency in the general population.

GeneMaps: are defined as groups of gene(s) that are directly or indirectly involved in at least one phenotype of a disorder (some non-limiting example of GeneMaps comprises various combinations of genes from Tables 2-4). As such, GeneMaps enable the
20 development of synergistic diagnostic products, creating "theranostics".

Genotype: Set of alleles at a specified locus or loci.

Haplotype: The allelic pattern of a group of (usually contiguous) DNA markers or other polymorphic loci along an individual chromosome or double helical DNA segment. Haplotypes identify individual chromosomes or chromosome segments. The presence of
25 shared haplotype patterns among a group of individuals implies that the locus defined by the haplotype has been inherited, identical by descent, from a common ancestor. Detection of identical by descent haplotypes is the basis of linkage disequilibrium (LD) mapping. Haplotypes are broken down through the generations by recombination and mutation. In some instances, a specific allele or haplotype may be associated with
30 susceptibility to a disorder or condition of interest, e.g., Crohn's disease. In other instances, an allele or haplotype may be associated with a decrease in susceptibility to a disorder or condition of interest, i.e., a protective sequence.

Host: includes prokaryotes and eukaryotes. The term includes an organism or cell that is the recipient of an expression vector (e.g., autonomously replicating or integrating vector).

Hybridizable: nucleic acids are hybridizable to each other when at least one strand of the nucleic acid can anneal to another nucleic acid strand under defined stringency conditions. In some embodiments, hybridization requires that the two nucleic acids contain at least 10 substantially complementary nucleotides; depending on the stringency of hybridization, however, mismatches may be tolerated. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementarity, and can be determined in accordance with the methods described herein.

Identity by descent: Identity among DNA sequences for different individuals that is due to the fact that they have all been inherited from a common ancestor. LD mapping identifies Identity by descent haplotypes as the likely location of disorder genes shared by a group of patients.

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 the case may be, 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 A.M. Lesk (ed), 1988, Computational Molecular Biology, Oxford University Press, NY; D.W. Smith (ed), 1993, Biocomputing. Informatics and Genome Projects, Academic Press, NY; A.M. Griffin and H.G. Griffin, H. G (eds), 1994, Computer Analysis of Sequence Data, Part 1, Humana Press, NJ; G. von Heinje, 1987, Sequence Analysis in Molecular Biology, Academic Press; and M. Gribskov and J. Devereux (eds), 1991, Sequence Analysis Primer, M Stockton Press, NY; H. Carillo and D. Lipman, 1988, SIAM J. Applied Math., 48:1073.

Immunogenic component: is a moiety that is capable of eliciting a humoral and/or cellular immune response in a host animal.

Isolated nucleic acids: are nucleic acids separated away from other components (e.g., DNA, RNA, and protein) with which they are associated (e.g., as obtained from cells,

chemical synthesis systems, or phage or nucleic acid libraries). Isolated nucleic acids are at least 60% free, preferably 75% free, and most preferably 90% free from other associated components. In accordance with the present invention, isolated nucleic acids can be obtained by methods described herein, or other established methods, including
5 isolation from natural sources (e.g., cells, tissues, or organs), chemical synthesis, recombinant methods, combinations of recombinant and chemical methods, and library screening methods.

Isolated polypeptides or peptides: are those that are separated from other components (e.g., DNA, RNA, and other polypeptides or peptides) with which they are associated
10 (e.g., as obtained from cells, translation systems, or chemical synthesis systems). In a preferred embodiment, isolated polypeptides or peptides are at least 10% pure; more preferably, 80% or 90% pure. Isolated polypeptides and peptides include those obtained by methods described herein, or other established methods, including isolation from natural sources (e.g., cells, tissues, or organs), chemical synthesis, recombinant
15 methods, or combinations of recombinant and chemical methods. Proteins or polypeptides referred to herein as recombinant are proteins or polypeptides produced by the expression of recombinant nucleic acids. A portion as used herein with regard to a protein or polypeptide, refers to fragments of that protein or polypeptide. The fragments can range in size from 5 amino acid residues to all but one residue of the entire protein
20 sequence. Thus, a portion or fragment can be at least 5, 5-50, 50-100, 100-200, 200-400, 400-800, or more consecutive amino acid residues of a protein or polypeptide.

Linkage disequilibrium (LD): the situation in which the alleles for two or more loci do not occur together in individuals sampled from a population at frequencies predicted by the product of their individual allele frequencies. In other words, markers that are in LD do
25 not follow Mendel's second law of independent random segregation. LD can be caused by any of several demographic or population artifacts as well as by the presence of genetic linkage between markers. However, when these artifacts are controlled and eliminated as sources of LD, then LD results directly from the fact that the loci involved are located close to each other on the same chromosome so that specific combinations
30 of alleles for different markers (haplotypes) are inherited together. Markers that are in high LD can be assumed to be located near each other and a marker or haplotype that is in high LD with a genetic trait can be assumed to be located near the gene that affects that trait. The physical proximity of markers can be measured in family studies where it is called linkage or in population studies where it is called linkage disequilibrium.

LD mapping: population based gene mapping, which locates disorder genes by identifying regions of the genome where haplotypes or marker variation patterns are shared statistically more frequently among disorder patients compared to healthy controls. This method is based upon the assumption that many of the patients will have
5 inherited an allele associated with the disorder from a common ancestor (IBD), and that this allele will be in LD with the disorder gene.

Locus: a specific position along a chromosome or DNA sequence. Depending upon context, a locus could be a gene, a marker, a chromosomal band or a specific sequence of one or more nucleotides.

10 Minor allele frequency (MAF): the population frequency of one of the alleles for a given polymorphism, which is equal or less than 50%. The sum of the MAF and the Major allele frequency equals one.

Markers: an identifiable DNA sequence that is variable (polymorphic) for different individuals within a population. These sequences facilitate the study of inheritance of a
15 trait or a gene. Such markers are used in mapping the order of genes along chromosomes and in following the inheritance of particular genes; genes closely linked to the marker or in LD with the marker will generally be inherited with it. Two types of markers are commonly used in genetic analysis, microsatellites and SNPs.

20 Microsatellite: DNA of eukaryotic cells comprising a repetitive, short sequence of DNA that is present as tandem repeats and in highly variable copy number, flanked by sequences unique to that locus.

Mutant sequence: if it differs from one or more wild-type sequences. For example, a nucleic acid from a gene listed in Tables 2-4 containing a particular allele of a single nucleotide polymorphism may be a mutant sequence. In some cases, the individual
25 carrying this allele has increased susceptibility toward the disorder or condition of interest. In other cases, the mutant sequence might also refer to an allele that decreases the susceptibility toward a disorder or condition of interest and thus acts in a protective manner. The term mutation may also be used to describe a specific allele of a polymorphic locus.

30 Non-conservative variants: are those in which a change in one or more nucleotides in a given codon position results in a polypeptide sequence in which a given amino acid residue in a polypeptide has been replaced by a non-conservative amino acid

substitution. Non-conservative variants also include polypeptides comprising non-conservative amino acid substitutions.

5 Nucleic acid or polynucleotide: purine- and pyrimidine-containing polymers of any length, either polyribonucleotides or polydeoxyribonucleotide or mixed polyribo polydeoxyribonucleotides. This includes single-and double-stranded molecules, *i.e.*, DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as protein nucleic acids (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases.

10 Nucleotide: a nucleotide, the unit of a DNA molecule, is composed of a base, a 2'-deoxyribose and phosphate ester(s) attached at the 5' carbon of the deoxyribose. For its incorporation in DNA, the nucleotide needs to possess three phosphate esters but it is converted into a monoester in the process.

15 Operably linked: means that the promoter controls the initiation of expression of the gene. A promoter is operably linked to a sequence of proximal DNA if upon introduction into a host cell the promoter determines the transcription of the proximal DNA sequence(s) into one or more species of RNA. A promoter is operably linked to a DNA sequence if the promoter is capable of initiating transcription of that DNA sequence.

Ortholog: denotes a gene or polypeptide obtained from one species that has homology to an analogous gene or polypeptide from a different species.

20 Paralog: denotes a gene or polypeptide obtained from a given species that has homology to a distinct gene or polypeptide from that same species.

Phenotype: any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to, a disorder.

25 Polymorphism: occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals at a single locus. A polymorphic site thus refers specifically to the locus at which the variation occurs. In some cases, an individual carrying a particular allele of a polymorphism has an increased or decreased susceptibility toward a disorder or condition of interest.

30 Portion and fragment: are synonymous. A portion as used with regard to a nucleic acid or polynucleotide refers to fragments of that nucleic acid or polynucleotide. The

fragments can range in size from 8 nucleotides to all but one nucleotide of the entire gene sequence. Preferably, the fragments are at least about 8 to about 10 nucleotides in length; at least about 12 nucleotides in length; at least about 15 to about 20 nucleotides in length; at least about 25 nucleotides in length; or at least about 35 to about 55
5 nucleotides in length.

Probe or primer: refers to a nucleic acid or oligonucleotide that forms a hybrid structure with a sequence in a target region of a nucleic acid due to complementarity of the probe or primer sequence to at least one portion of the target region sequence.

Protein and polypeptide: are synonymous. Peptides are defined as fragments or portions
10 of polypeptides, preferably fragments or portions having at least one functional activity (e.g., proteolysis, adhesion, fusion, antigenic, or intracellular activity) as the complete polypeptide sequence.

Recombinant nucleic acids: nucleic acids which have been produced by recombinant DNA methodology, including those nucleic acids that are generated by procedures which
15 rely upon a method of artificial replication, such as the polymerase chain reaction (PCR) and/or cloning into a vector using restriction enzymes. Portions of recombinant nucleic acids which code for polypeptides can be identified and isolated by, for example, the method of M. Jasin *et al.*, U.S. Patent No. 4,952,501.

Regulatory sequence: refers to a nucleic acid sequence that controls or regulates
20 expression of structural genes when operably linked to those genes. These include, for example, the lac systems, the trp system, major operator and promoter regions of the phage lambda, the control region of fd coat protein and other sequences known to control the expression of genes in prokaryotic or eukaryotic cells. Regulatory sequences will vary depending on whether the vector is designed to express the operably linked
25 gene in a prokaryotic or eukaryotic host, and may contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements and/or translational initiation and termination sites.

Sample: as used herein refers to a biological sample, such as, for example, tissue or fluid isolated from an individual or animal (including, without limitation, plasma, serum,
30 cerebrospinal fluid, lymph, tears, nails, hair, saliva, milk, pus, and tissue exudates and secretions) or from *in vitro* cell culture-constituents, as well as samples obtained from, for example, a laboratory procedure.

Single nucleotide polymorphism (SNP): variation of a single nucleotide. This includes the replacement of one nucleotide by another and deletion or insertion of a single nucleotide. Typically, SNPs are biallelic markers although tri- and tetra-allelic markers also exist. For example, SNP A\C may comprise allele C or allele A (Tables 5-36). Thus, a nucleic acid molecule comprising SNP A\C may include a C or A at the polymorphic position. For clarity purposes, an ambiguity code is used in Tables 5-36 and the sequence listing, to represent the variations. For a combination of SNPs, the term "haplotype" is used, *e.g.* the genotype of the SNPs in a single DNA strand that are linked to one another. In certain embodiments, the term "haplotype" is used to describe a combination of SNP alleles, *e.g.*, the alleles of the SNPs found together on a single DNA molecule. In specific embodiments, the SNPs in a haplotype are in linkage disequilibrium with one another.

Sequence-conservative: variants are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position (*i.e.*, silent mutation).

Substantially homologous: a nucleic acid or fragment thereof is substantially homologous to another if, when optimally aligned (with appropriate nucleotide insertions and/or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least 60% of the nucleotide bases, usually at least 70%, more usually at least 80%, preferably at least 90%, and more preferably at least 95-98% of the nucleotide bases. Alternatively, substantial homology exists when a nucleic acid or fragment thereof will hybridize, under selective hybridization conditions, to another nucleic acid (or a complementary strand thereof). Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% sequence identity over a stretch of at least about nine or more nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% (M. Kanehisa, 1984, Nucl Acids Res. 11:203-213). The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least 14 nucleotides, usually at least 20 nucleotides, more usually at least 24 nucleotides, typically at least 28 nucleotides, more typically at least 32 nucleotides, and preferably at least 36 or more nucleotides.

Wild-type gene from Tables 2-4: refers to the reference sequence. The wild-type gene sequences from Tables 2-36 used to identify the variants (polymorphisms, alleles, and haplotypes) described in detail herein.

Technical and scientific terms used herein have the meanings commonly understood by one of ordinary skill in the art to which the present invention pertains, unless otherwise defined. Reference is made herein to various methodologies known to those of skill in the art. Publications and other materials setting forth such known methodologies to which reference is made are incorporated herein by reference in their entireties as though set forth in full. Standard reference works setting forth the general principles of recombinant DNA technology include J. Sambrook *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; P.B. Kaufman *et al.*, (eds), 1995, *Handbook of Molecular and Cellular Methods in Biology and Medicine*, CRC Press, Boca Raton; M.J. McPherson (ed), 1991, *Directed Mutagenesis: A Practical Approach*, IRL Press, Oxford; J. Jones, 1992, *Amino Acid and Peptide Synthesis*, Oxford Science Publications, Oxford; B.M. Austen and O.M.R. Westwood, 1991, *Protein Targeting and Secretion*, IRL Press, Oxford; D.N Glover (ed), 1985, *DNA Cloning*, Volumes I and II; M.J. Gait (ed), 1984, *Oligonucleotide Synthesis*; B.D. Hames and S.J. Higgins (eds), 1984, *Nucleic Acid Hybridization*; Quirke and Taylor (eds), 1991, *PCR-A Practical Approach*; Harries and Higgins (eds), 1984, *Transcription and Translation*; R.I. Freshney (ed), 1986, *Animal Cell Culture; Immobilized Cells and Enzymes*, 1986, IRL Press; Perbal, 1984, *A Practical Guide to Molecular Cloning*, J. H. Miller and M. P. Calos (eds), 1987, *Gene Transfer Vectors for Mammalian Cells*, Cold Spring Harbor Laboratory Press; M.J. Bishop (ed), 1998, *Guide to Human Genome Computing*, 2d Ed., Academic Press, San Diego, CA; L.F. Peruski and A.H. Peruski, 1997, *The Internet and the New Biology. Tools for Genomic and Molecular Research*, American Society for Microbiology, Washington, D.C. Standard reference works setting forth the general principles of immunology include S. Sell, 1996, *Immunology, Immunopathology & Immunity*, 5th Ed., Appleton & Lange, Publ., Stamford, CT; D. Male *et al.*, 1996, *Advanced Immunology*, 3d Ed., Times Mirror Int'l Publishers Ltd., Publ., London; D.P. Stites and A.L. Terr, 1991, *Basic and Clinical Immunology*, 7th Ed., Appleton & Lange, Publ., Norwalk, CT; and A.K. Abbas *et al.*, 1991, *Cellular and Molecular Immunology*, W. B. Saunders Co., Publ., Philadelphia, PA. Any suitable materials and/or methods known to those of skill can be utilized in carrying out the present invention; however, preferred materials and/or methods are described. Materials, reagents, and the like to which reference is made in the following description and examples are generally obtainable from commercial sources, and specific vendors are cited herein.

DETAILED DESCRIPTION OF THE INVENTION

Inflammatory Bowel Disease (IBD) is characterized by excessive and chronic inflammation at various sites in the gastro-intestinal tract. IBD describes two clinical conditions called Crohn's disease (CD) and ulcerative colitis (UC). CD and UC share many clinical and pathological characteristics but they also have some markedly different features. There is strong scientific support suggesting that the main pathological processes in these two diseases are distinct. This patent application will focus primarily on IBD (ex: UC and CD), but is not limited to IBD and encompasses inflammatory diseases in general for therapeutic targets, since IBD and inflammatory diseases can share common therapeutic targets.

Previously identified genes and loci

Genetic studies have previously indicated the presence of several loci predisposing to Inflammatory Bowel Diseases. Nine IBD loci have been identified: IBD1 (CARD15/NOD2, Caspase recruitment domain family, member 15 (NOD2 protein)), IBD2 (Inflammatory bowel disease-2), IBD3 (Inflammatory bowel disease-3), IBD4 (Inflammatory bowel disease-4), IBD5 (Inflammatory bowel disease-5), IBD6 (Inflammatory bowel disease-6), IBD7 (Inflammatory bowel disease-7), IBD8 (Inflammatory bowel disease-8), and IBD9 (Inflammatory bowel disease 9).

Several loci have been identified and replicated to date; they are located on chromosomes 16q12, 12q13.2-q24.1, 6p21, 14q11-q12, 5q31-q33, 19p13, 1p36, 16p, and 3p26 respectively (Wild and Rioux 2004; Duerr *et al.*, 2002). Results from linkage studies have suggested that CD and UC share some loci but do not share others, such as locus 16q12 which is unique to CD (The IBD International Genetics Consortium 2001). Most of the genes determining susceptibility in each of these chromosomal regions remain to be identified. The most widely replicated loci are on chromosomes 16 (caused by mutations in the NOD2/CARD15 gene), 12, 6, 14 and 5 (OCTN2 genes at IBD5 locus), which predispose to early-onset Crohn's disease. Thus, there is a continuing need in the medical arts for genetic markers of IBD and guidance for the use of such markers.

Several years ago, two different groups reported the association of CARD15 (NOD2) variants with CD (Hugot 2001; Ogura 2001). CARD15 is located at the IBD1 locus. The gene codes for an intracellular receptor involved in the innate immune detection of bacterial products. This detection induces the activation of the NF- κ B pathway which is of particular importance in immune and inflammatory responses (Philpott and Viala .
5 2004). Three major mutations represent 82% of the total CARD15 mutations: R702W, G908R, and 1007fsinsC. However they do not explain over 20% of the genetic predisposition to the disease, and altogether, they are carried by 30-50% of CD patients and 15-20% of healthy controls in Caucasian populations. These values are much lower
10 in Japanese or Africans (reviewed in Girardin 2003).

Recently, the OCTN1 and OCTN2 genes at IBD5 locus have been associated with CD (Peltekova 2004). Variants in these genes are in strong linkage disequilibrium and create a two-allele risk haplotype enriched in patients with CD. Both proteins are trans-membrane sodium-dependent carnitine transporters and sodium-independent organic
15 cation transporters. The variants may cause disease by impairing OCTN activity or expression, reducing carnitine transport in a cell-type and disease-specific manner (Peltekova 2004). The IBD5 locus contains multiple candidate genes including the genes for organic cation transporters (OCTN2/SLC22A4 and SLC22A5), the gene for a LIN4-domain-containing protein (RILIPDLIM3), the gene for the oc2 subunit of proline
20 hydroxylase (P4HA2) and a gene of unknown function (NCBI UniGene identifier Hs.70932). Because of extensive linkage disequilibrium (LD) in this region, it has not been possible to further refine the SNP map and unambiguously identify a single susceptibility gene. With respect to OCTN2, the basal transcription is downregulated by the C allele of G-207C, and this allele disrupts the ability of SLC22A5 to be upregulated
25 in response to heat shock or arachidonic acid as a result of impaired HSF I binding and subsequent transcriptional activation.

The mechanism by which these organic cation transporters contribute to the pathology of Crohn's disease may relate to the in vivo metabolic importance of carnitine. Carnitine facilitates transport of long chain fatty acids across the mitochondrial inner membrane for
30 subsequent β -oxidation, and is also important in the maintenance of cellular CoenzymeA levels. Carnitine uptake into lymphocytes, along with a corresponding decrease in plasma levels, is a physiological response to inflammation. Symptoms of the related condition Ulcerative Colitis may be due to an energy deficiency in colonic epithelium secondary to poor mitochondrial function due to decreased long-chain fatty acid

transport to the mitochondria (Roediger WE 1980). The haplotype of mutations in Crohn's disease patients provided previously might affect cellular metabolic energy levels in inflamed tissue by combining impaired OCTN1 transporter function with downregulation and inability to respond to heat or inflammatory stress by the OCTN2 gene. Heat shock proteins and arachidonic acids are involved in response to a variety of cellular stresses, including sepsis, metabolic stress and ischaemia. The heat shock response modulates inflammation through modulation of NF-kB activation. OCTN2 is a heat stress inducible protein. Thus, the OCTN2 gene is normally upregulated in response to inflammation through binding of HSF I protein to its promoter. This in turn mobilizes carnitine and bolsters metabolism in the inflamed tissue. Impaired OCTN1 carnitine transporter activity and lowered OCTN2 expression level results in reduced metabolism and either triggers or worsens cellular stress in areas of inflammation. The two OCTN transporters may therefore function in the inflammatory pathology of Crohn's disease. Additionally, OCTN1 is a polyspecific cation transporter, and might have a role in the uptake of drugs used to treat CD from the gut. As described above, OCTN2 is a transporter protein with the ability to transport carnitine in a sodium dependent manner. Missense mutations and nonsense mutations in the organic cation transporter OCTN2 had previously been identified in patients with primary Systemic Carnitine Deficiency (SCD), an autosomal recessive disorder characterized by progressive cardiomyopathy, skeletal myopathy, hypoglycemia and hyperammonemia.

Genes encoding proteins involved in the immune system, epithelial functions, and host response to micro-organisms represent good potential candidates for IBD and have been examined in numerous case-control studies. However, many of the published associations of genetic variants with IBD have not been replicated in follow-up studies.

The genetic variants that have been identified so far in IBD explain only a fraction of the genetic predisposition to this disorder. It is clear that multiple components contribute to disease risk, each component having a modest effect on disease susceptibility. Thus the development of GeneMaps for IBD may lead to a better understanding of pathogenesis and to the identification of new pathways involved in the disease, ultimately leading to better treatments for the patients. GeneMaps may also lead to molecular diagnostic tools that will identify subjects at risk for CD or for serious complications of the disease.

Genome wide association study to construct a GeneMap for IBD (ex: Crohn's disease)

The present invention is based on the discovery of genes associated with IBD and inflammatory diseases in general. In the preferred embodiment, disease-associated loci (candidate regions; Table 1) are identified by the statistically significant differences in allele or haplotype frequencies between the cases and the controls. For the purpose of
5 the present invention, 654 candidate regions (Table 1) are identified.

The invention provides a method for the discovery of genes associated with IBD (E.g. Crohn's disease) and the construction of a GeneMap for Crohn's disease in a human population, comprising the following steps (see Example section herein):

Step 1: Recruit patients (cases) and controls

- 10 In the preferred embodiment, 500 patients diagnosed for Crohn's disease along with two family members are recruited from the Quebec Founder Population (QFP). The preferred trios recruited are parent-parent-child (PPC) trios. Trios can also be recruited as parent-child-child (PCC) trios. In another preferred embodiment, more or less than 500 trios are recruited. In another embodiment, independent case and control samples are recruited.
- 15 In another embodiment, the present invention is performed as a whole or partially with DNA samples from individuals of another founder population than the Quebec population or from the general population.

Step 2: DNA extraction and quantitation

- Any sample comprising cells or nucleic acids from patients or controls may be used.
- 20 Preferred samples are those easily obtained from the patient or control. Such samples include, but are not limited to blood, peripheral lymphocytes, buccal swabs, epithelial cell swabs, nails, hair, bronchoalveolar lavage fluid, sputum, or other body fluid or tissue obtained from an individual.

- In one embodiment, DNA is extracted from such samples in the quantity and quality
25 necessary to perform the invention using conventional DNA extraction and quantitation techniques. The present invention is not linked to any DNA extraction or quantitation platform in particular.

Step 3: Genotype the recruited individuals

- In one embodiment, assay-specific and/or locus-specific and/or allele-specific
30 oligonucleotides for every SNP marker of the present invention (Tables 5-36) are

organized onto one or more arrays. The genotype at each SNP locus is revealed by hybridizing short PCR fragments comprising each SNP locus onto these arrays. The arrays permit a high-throughput genome wide association study using DNA samples from individuals of the Quebec founder population. Such assay-specific and/or locus-specific and/or allele-specific oligonucleotides necessary for scoring each SNP of the present invention are preferably organized onto a solid support. Such supports can be arrayed on wafers, glass slides, beads or any other type of solid support.

In another embodiment, the assay-specific and/or locus-specific and/or allele-specific oligonucleotides are not organized onto a solid support but are still used as a whole, in panels or one by one. The present invention is therefore not linked to any genotyping platform in particular.

In another embodiment, one or more portions of the SNP maps (publicly available maps, proprietary maps from Perlegen Sciences, Inc. (Mountain View, CA, USA), and our own proprietary QLDM map) are used to screen the whole genome, a subset of chromosomes, a chromosome, a subset of genomic regions or a single genomic region.

The 1,500 individuals composing the 500 trios are preferably individually genotyped with at least 80,000 markers, generating at least a few million genotypes; more preferably, at least a hundred million.

Step 4: Exclude the markers that did not pass the quality control of the assay.

Preferably, the quality controls consist of, but are not limited to, the following criteria: eliminate SNPs that had a high rate of Mendelian errors (cut-off at 1% Mendelian error rate), that deviate from the Hardy-Weinberg equilibrium, that are non-polymorphic in the Quebec founder population or have too many missing data (cut-off at 1% missing values or higher), or simply because they are non-polymorphic in the Quebec founder population (cut-off at $1\% \leq 10\%$ minor allele frequency (MAF)).

Step 5: Perform the genetic analysis on the results obtained using haplotype information as well as single-marker association.

In the preferred embodiment, genetic analysis is performed on all the genotypes from step 3.

In another embodiment, genetic analysis is performed on a total of 248,535 SNPs.

In one embodiment, the genetic analysis consists of, but is not limited to features corresponding to Phase information and haplotype structures. Phase information and haplotype structures are preferably deduced from trio genotypes using Phasefinder. Since chromosomal assignment (phase) cannot be estimated when all trio members are heterozygous, an Expectation-Maximization (EM) algorithm may be used to resolve chromosomal assignment ambiguities after Phasefinder.

In yet another embodiment, the PL-EM algorithm (Partition-Ligation EM; Niu *et al.*, Am. J. Hum. Genet. 70:157 (2002)) can be used to estimate haplotypes from the "genotype" data as a measured estimate of the reference allele frequency of a SNP in 15-marker windows that advance in increments of one marker across the data set. The results from such algorithms are converted into 15-marker haplotype files. Subsequently, the individual 15-marker block files are assembled into one continuous block of haplotypes for the entire chromosome. These extended haplotypes can then be used for further analysis. Such haplotype assembly algorithms take the consensus estimate of the allele call at each marker over all separate estimations (most markers are estimated 15 different times as the 15 marker blocks pass over their position).

In the preferred embodiment, the haplotypes for both the controls and the patients are derived in this manner. The preferred control of a trio structure is the spouse if the patient is one of the parents or the non-transmitted chromosomes (chromosomes found in parents but not in affected child) if the patient is the child.

In another embodiment, the haplotype frequencies among patients are compared to those among the controls using LDSTATS, a program that assesses the association of haplotypes with the disease. Such program defines haplotypes using multi-marker windows that advance across the marker map in one-marker increments. Such windows can be 1, 3, 5, 7 or 9 markers wide, and all these window sizes are tested concurrently. Larger multi-marker haplotype windows can also be used. At each position the frequency of haplotypes in cases is compared to the frequency of haplotypes in controls. Such allele frequency differences for single marker windows can be tested using Pearson's Chi-square with any degree of freedom. Multi-allelic haplotype association can be tested using Smith's normalization of the square root of Pearson's Chi-square. Such significance of association can be reported in two ways:

The significance of association within any one haplotype window is plotted against the marker that is central to that window.

P-values of association for each specific marker are calculated as a pooled P-value across all haplotype windows in which they occur. The pooled P-value is calculated using an expected value and variance calculated using a permutation test that considers covariance between individual windows. Such pooled P-values can yield narrower
5 regions of gene location than the window data (see Example 3 for details on analysis methods, such as LDSTATS v2.0 and v4.0).

In another embodiment, conditional haplotype analyses can be performed on subsets of the original set of cases and controls using the program LDSTATS. The selection of a subset of cases and their matched controls can be based on the carrier status of cases
10 at a gene or locus of interest. Various conditional haplotypes can be derived, such as protective haplotypes and risk haplotypes.

Step 6: Fine Mapping

In this step, the candidate regions that were identified by step 4 are further mapped for the purpose of refinement and validation.

15 In the preferred embodiment, this fine mapping is performed with a density of genetic markers higher than in the genome wide scan (step 3) using any genotyping platform available in the art. Such fine mapping can be, but is not limited to, typing the allele via an allele-specific elongation assay that is then ligated to a locus-specific oligonucleotide. Such assays can be performed directly on the genomic DNA at a highly multiplex level
20 and the products can be amplified using universal oligonucleotides. For each candidate region, the density of genetic markers can be, but is not limited to, a set of SNP markers with an average inter-marker distance of 1-4 Kb distributed over about 400 Kb to 1 Mb, roughly centered at the highest point of the GWS association. The preferred samples are those obtained from Crohn's disease PPC trios including the ones used for the GWS.
25 Other preferred samples are trios or case control samples from another population, such as a General population.

In the preferred embodiment, the genetic analysis of the results obtained using haplotype information as well as single-marker association (as performed as in step 5, described herein) is performed as described herein (step 5 and Example section). The candidate
30 regions that are validated and confirmed after this analysis proceed to a gene mining step described in Example 5, herein, to characterize their marker and genetic content.

Step 7: SNP and DNA polymorphism discovery

In the preferred embodiment, all the candidate genes and regions identified in step 6 are sequenced for polymorphism identification.

In another embodiment, the entire region, including all introns, is sequenced to identify all polymorphisms.

- 5 In yet another embodiment, the candidate genes are prioritized for sequencing, and only functional gene elements (promoters, conserved noncoding sequences, exons and splice sites) are sequenced.

10 In yet another embodiment, previously identified polymorphisms in the candidate regions can also be used. For example, SNPs from dbSNP, Perlegen Sciences, Inc., or others can also be used rather than resequencing the candidate regions to identify polymorphisms.

15 The discovery of SNPs and DNA polymorphisms generally comprises a step consisting of determining the major haplotypes in the region to be sequenced. The preferred samples are selected according to which haplotypes contribute to the association signal observed in the region to be sequenced. The purpose is to select a set of samples that covers all the major haplotypes in the given region. Each major haplotype is preferably analyzed in at least a few individuals.

20 Any analytical procedure may be used to detect the presence or absence of variant nucleotides at one or more polymorphic positions of the invention. In general, the detection of allelic variation requires a mutation discrimination technique, optionally an amplification reaction and optionally a signal generation system. Any means of mutation detection or discrimination may be used. For instance, DNA sequencing, scanning methods, hybridization, extension based methods, incorporation based methods, restriction enzyme-based methods and ligation-based methods may be used in the
25 methods of the invention.

Sequencing methods include, but are not limited to, direct sequencing, and sequencing by hybridization. Scanning methods include, but are not limited to, protein truncation test (PTT), single-strand conformation polymorphism analysis (SSCP), denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE),
30 cleavage, heteroduplex analysis, chemical mismatch cleavage (CMC), and enzymatic mismatch cleavage. Hybridization-based methods of detection include, but are not limited to, solid phase hybridization such as dot blots, multiple allele specific diagnostic

assay (MASDA), reverse dot blots, and oligonucleotide arrays (DNA Chips). Solution phase hybridization amplification methods may also be used, such as Taqman. Extension based methods include, but are not limited to, amplification refracton mutation systems (ARMS), amplification refractory mutation systems (ALEX), and competitive
5 oligonucleotide priming systems (COPS). Incorporation based methods include, but are not limited to, mini-sequencing and arrayed primer extension (APEX). Restriction enzyme-based detection systems include, but are not limited to, restriction site generating PCR. Lastly, ligation based detection methods include, but are not limited to, oligonucleotide ligation assays (OLA). Signal generation or detection systems that may
10 be used in the methods of the invention include, but are not limited to, fluorescence methods such as fluorescence resonance energy transfer (FRET), fluorescence quenching, fluorescence polarization as well as other chemiluminescence, electrochemiluminescence, Raman, radioactivity, colometric methods, hybridization protection assays and mass spectrometry methods. Further amplification methods
15 include, but are not limited to self sustained replication (SSR), nucleic acid sequence based amplification (NASBA), ligase chain reaction (LCR), strand displacement amplification (SDA) and branched DNA (B-DNA).

Step 8: Ultrafine Mapping

This step further maps the candidate regions and genes confirmed in the previous step
20 to identify and validate the responsible polymorphisms associated with Crohn's disease in the human population.

In a preferred embodiment, the discovered SNPs and polymorphisms of step 7 are ultrafine mapped at a higher density of markers than the fine mapping described herein using the same technology described in step 6.

25 Step 9: GeneMap construction

The confirmed variations in DNA (including both genic and non-genic regions) are used to build a GeneMap for IBD (ex: Crohn's disease). The gene content of this GeneMap is described in more detail below. Such GeneMap can be used for other methods of the invention comprising the diagnostic methods described herein, the susceptibility to
30 Crohn's disease, the response to a particular drug, the efficacy of a particular drug, the screening methods described herein and the treatment methods described herein.

As is evident to one of ordinary skill in the art, all of the above steps or the steps do not need to be performed, or performed in a given order to practice or use the SNPs, genomic regions, genes, proteins, etc. in the methods of the invention.

5 Genes from the GeneMap

In one embodiment the GeneMap consists of genes and targets, in a variety of combinations, identified from the candidate regions listed in Table 1. In another embodiment, all genes from Tables 2-4 are present in the GeneMap. In another preferred embodiment, the GeneMap consists of a selection of genes from Tables 2-4.

10 The genes of the invention (Tables 2-4) are arranged by candidate regions and by their chromosomal location. Such order is for the purpose of clarity and does not reflect any other criteria of selection in the association of the genes with IBD (ex: Crohn's disease).

In the preferred embodiment, genes identified in the WGAS and subsequent fine mapping studies for IBD (ex: Crohn's disease) are evaluated using the Ingenuity

15 Pathway Analysis application (IPA, Ingenuity systems) in order to identify direct biological interactions between these genes, and also to identify molecular regulators acting on those genes (indirect interactions) that could be also involved in IBD (ex: Crohn's disease). The purpose of this effort is to decipher the molecules involved in contributing to IBD (ex: Crohn's disease). These gene interaction networks are very

20 valuable tools in the sense that they facilitate extension of the map of gene products that could represent potential drug targets for IBD (ex: Crohn's disease).

Nucleic acid sequences

The nucleic acid sequences of the present invention may be derived from a variety of

25 sources including DNA, cDNA, synthetic DNA, synthetic RNA, derivatives, mimetics or combinations thereof. Such sequences may comprise genomic DNA, which may or may not include naturally occurring introns, genic regions, nongenic regions, and regulatory regions. Moreover, such genomic DNA may be obtained in association with promoter regions or poly (A) sequences. The sequences, genomic DNA, or cDNA may be

30 obtained in any of several ways. Genomic DNA can be extracted and purified from suitable cells by means well known in the art. Alternatively, mRNA can be isolated from a

cell and used to produce cDNA by reverse transcription or other means. The nucleic acids described herein are used in certain embodiments of the methods of the present invention for production of RNA, proteins or polypeptides, through incorporation into cells, tissues, or organisms. In one embodiment, DNA containing all or part of the coding
5 sequence for the genes described in Tables 2-4, or the SNP markers described in Tables 5-36, is incorporated into a vector for expression of the encoded polypeptide in suitable host cells. The invention also comprises the use of the nucleotide sequence of the nucleic acids of this invention to identify DNA probes for the genes described in Tables 2-4 or the SNP markers described in Tables 5-36, PCR primers to amplify the
10 genes described in Tables 2-4 or the SNP markers described in Tables 5-36, nucleotide polymorphisms in the genes described in Tables 2-4, and regulatory elements of the genes described in Tables 2-4. The nucleic acids of the present invention find use as primers and templates for the recombinant production of IBD-associated peptides or polypeptides, for chromosome and gene mapping, to provide antisense sequences, for
15 tissue distribution studies, to locate and obtain full length genes, to identify and obtain homologous sequences (wild-type and mutants), and in diagnostic applications.

Antisense oligonucleotides

In a particular embodiment of the invention, an antisense nucleic acid or oligonucleotide
20 is wholly or partially complementary to, and can hybridize with, a target nucleic acid (either DNA or RNA) having the sequence of SEQ ID NO:1, NO:3 or any SEQ ID from any Tables of the invention. For example, an antisense nucleic acid or oligonucleotide comprising 16 nucleotides can be sufficient to inhibit expression of at least one gene from Tables 2-4. Alternatively, an antisense nucleic acid or oligonucleotide can be
25 complementary to 5' or 3' untranslated regions, or can overlap the translation initiation codon (5' untranslated and translated regions) of at least one gene from Tables 2-4, or its functional equivalent. In another embodiment, the antisense nucleic acid is wholly or partially complementary to, and can hybridize with, a target nucleic acid that encodes a polypeptide from a gene described in Tables 2-4.

30 In addition, oligonucleotides can be constructed which will bind to duplex nucleic acid (*i.e.*, DNA:DNA or DNA:RNA), to form a stable triple helix containing or triplex nucleic acid. Such triplex oligonucleotides can inhibit transcription and/or expression of a gene from Tables 2-4, or its functional equivalent (M.D. Frank-Kamenetskii *et al.*, 1995).

Triplex oligonucleotides are constructed using the basepairing rules of triple helix formation and the nucleotide sequence of the genes described in Tables 2-4.

The present invention encompasses methods of using oligonucleotides in antisense inhibition of the function of the genes from Tables 2-4. In the context of this invention, the term "oligonucleotide" refers to naturally-occurring species or synthetic species formed from naturally-occurring subunits or their close homologs. The term may also refer to moieties that function similarly to oligonucleotides, but have non-naturally-occurring portions. Thus, oligonucleotides may have altered sugar moieties or inter-sugar linkages. Exemplary among these are phosphorothioate and other sulfur containing species, which are known in the art. In preferred embodiments, at least one of the phosphodiester bonds of the oligonucleotide has been substituted with a structure that functions to enhance the ability of the compositions to penetrate into the region of cells where the RNA whose activity is to be modulated is located. It is preferred that such substitutions comprise phosphorothioate bonds, methyl phosphonate bonds, or short chain alkyl or cycloalkyl structures. In accordance with other preferred embodiments, the phosphodiester bonds are substituted with structures which are, at once, substantially non-ionic and non-chiral, or with structures which are chiral and enantiomerically specific. Persons of ordinary skill in the art will be able to select other linkages for use in the practice of the invention. Oligonucleotides may also include species that include at least some modified base forms. Thus, purines and pyrimidines other than those normally found in nature may be so employed. Similarly, modifications on the furanosyl portions of the nucleotide subunits may also be effected, as long as the essential tenets of this invention are adhered to. Examples of such modifications are 2'-O-alkyl- and 2'-halogen-substituted nucleotides. Some non-limiting examples of modifications at the 2' position of sugar moieties which are useful in the present invention include OH, SH, SCH₃, F, OCH₃, OCN, O(CH₂), NH₂ and O(CH₂)_n CH₃, where n is from 1 to about 10. Such oligonucleotides are functionally interchangeable with natural oligonucleotides or synthesized oligonucleotides, which have one or more differences from the natural structure. All such analogs are comprehended by this invention so long as they function effectively to hybridize with at least one gene from Tables 2-4 DNA or RNA to inhibit the function thereof.

The oligonucleotides in accordance with this invention preferably comprise from about 3 to about 50 subunits. It is more preferred that such oligonucleotides and analogs comprise from about 8 to about 25 subunits and still more preferred to have from about

12 to about 20 subunits. As defined herein, a "subunit" is a base and sugar combination suitably bound to adjacent subunits through phosphodiester or other bonds. Antisense nucleic acids or oligonucleotides can be produced by standard techniques (see, *e.g.*, Shewmaker *et al.*, U.S. Patent No. 6,107,065). The oligonucleotides used in accordance
5 with this invention may be conveniently and routinely made through the well-known technique of solid phase synthesis. Any other means for such synthesis may also be employed; however, the actual synthesis of the oligonucleotides is well within the abilities of the practitioner. It is also well known to prepare other oligonucleotides such as phosphorothioates and alkylated derivatives.

10 The oligonucleotides of this invention are designed to be hybridizable with RNA (*e.g.*, mRNA) or DNA from genes described in Tables 2-4. For example, an oligonucleotide (*e.g.*, DNA oligonucleotide) that hybridizes to mRNA from a gene described in Tables 2-4 can be used to target the mRNA for RnaseH digestion. Alternatively an oligonucleotide that can hybridize to the translation initiation site of the mRNA of a gene described in
15 Tables 2-4 can be used to prevent translation of the mRNA. In another approach, oligonucleotides that bind to the double-stranded DNA of a gene from Tables 2-4 can be administered. Such oligonucleotides can form a triplex construct and inhibit the transcription of the DNA encoding polypeptides of the genes described in Tables 2-4. Triple helix pairing prevents the double helix from opening sufficiently to allow the
20 binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described (see, *e.g.*, J.E. Gee *et al.*, 1994, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY).

As non-limiting examples, antisense oligonucleotides may be targeted to hybridize to the following regions: mRNA cap region; translation initiation site; translational termination
25 site; transcription initiation site; transcription termination site; polyadenylation signal; 3' untranslated region; 5' untranslated region; 5' coding region; mid coding region; and 3' coding region. Preferably, the complementary oligonucleotide is designed to hybridize to the most unique 5' sequence of a gene described in Tables 2-4, including any of about 15-35 nucleotides spanning the 5' coding sequence. In accordance with the present
30 invention, the antisense oligonucleotide can be synthesized, formulated as a pharmaceutical composition, and administered to a subject. The synthesis and utilization of antisense and triplex oligonucleotides have been previously described (*e.g.*, Simon *et al.*, 1999; Barre *et al.*, 2000; Elez *et al.*, 2000; Sauter *et al.*, 2000).

Alternatively, expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express nucleic acid sequence that is complementary to the nucleic acid sequence encoding a polypeptide from the genes described in Tables 2-4. These techniques are described both in Sambrook *et al.*, 1989 and in Ausubel *et al.*, 1992. For example, expression of at least one gene from Tables 2-4 can be inhibited by transforming a cell or tissue with an expression vector that expresses high levels of untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a nonreplicating vector, and even longer if appropriate replication elements are included in the vector system. Various assays may be used to test the ability of gene-specific antisense oligonucleotides to inhibit the expression of at least one gene from Tables 2-4. For example, mRNA levels of the genes described in Tables 2-4 can be assessed by Northern blot analysis (Sambrook *et al.*, 1989; Ausubel *et al.*, 1992; J.C. Alwine *et al.* 1977; I.M. Bird, 1998), quantitative or semi-quantitative RT-PCR analysis (see, *e.g.*, W.M. Freeman *et al.*, 1999; Ren *et al.*, 1998; J.M. Cale *et al.*, 1998), or in situ hybridization (reviewed by A.K. Raap, 1998). Alternatively, antisense oligonucleotides may be assessed by measuring levels of the polypeptide from the genes described in Tables 2-4, *e.g.*, by western blot analysis, indirect immunofluorescence and immunoprecipitation techniques (see, *e.g.*, J.M. Walker, 1998, Protein Protocols on CD-ROM, Humana Press, Totowa, NJ). Any other means for such detection may also be employed, and is well within the abilities of the practitioner.

Mapping Technologies

The present invention includes various methods, which employ mapping technologies to map SNPs and polymorphisms. For purpose of clarity, this section comprises, but is not limited to, the description of mapping technologies that can be utilized to achieve the embodiments described herein. Mapping technologies may be based on amplification methods, restriction enzyme cleavage methods, hybridization methods, sequencing methods, and cleavage methods using agents.

Amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990), transcriptional amplification system (Kwoh *et al.*, 1989), Q-Beta Replicase (Lizardi *et al.*, 1988), isothermal amplification (e.g. Dean *et al.*, 2002; and Hafner *et al.*, 2001), or any other nucleic acid amplification method, followed by the detection of the amplified
5 molecules using techniques well known to those of ordinary skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low number.

Restriction enzyme cleavage methods include: isolating sample and control DNA, amplification (optional), digestion with one or more restriction endonucleases,
10 determination of fragment length sizes by gel electrophoresis and comparing samples and controls. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, sequence specific ribozymes (see, e.g., U.S. Pat. No. 5,498,531) or DNAzyme (e.g. U.S. Pat. No. 5,807,718) can be used to score for the presence of specific mutations by development or loss of a ribozyme or
15 DNAzyme cleavage site.

Hybridization methods include any measurement of the hybridization or gene expression levels, of sample nucleic acids to probes corresponding to about 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 50, 75, 100, 200, 500, 1000 or more genes, or ranges of these numbers, such as about 2-10, about 10-20, about 20-50, about 50-100, about 100-200, about 200-
20 500 or about 500-1000 genes of Tables 2-4.

SNPs and SNP maps of the invention can be identified or generated by hybridizing sample nucleic acids, e.g., DNA or RNA, to high density arrays or bead arrays containing oligonucleotide probes corresponding to the polymorphisms of Tables 5-36 (see the Affymetrix arrays and Illumina bead sets at www.affymetrix.com and www.illumina.com
25 and see Cronin *et al.*, 1996; or Kozal *et al.*, 1996).

Methods of forming high density arrays of oligonucleotides with a minimal number of synthetic steps are known. The oligonucleotide analogue array can be synthesized on a single or on multiple solid substrates by a variety of methods, including, but not limited to, light-directed chemical coupling, and mechanically directed coupling (see Pirrung,
30 U.S. Patent No. 5,143,854).

In brief, the light-directed combinatorial synthesis of oligonucleotide arrays on a glass surface proceeds using automated phosphoramidite chemistry and chip masking

techniques. In one specific implementation, a glass surface is derivatized with a silane reagent containing a functional group, *e.g.*, a hydroxyl or amine group blocked by a photolabile protecting group. Photolysis through a photolithographic mask is used selectively to expose functional groups which are then ready to react with incoming 5' photoprotected nucleoside phosphoramidites. The phosphoramidites react only with those sites which are illuminated (and thus exposed by removal of the photolabile blocking group). Thus, the phosphoramidites only add to those areas selectively exposed from the preceding step. These steps are repeated until the desired array of sequences have been synthesized on the solid surface. Combinatorial synthesis of different oligonucleotide analogues at different locations on the array is determined by the pattern of illumination during synthesis and the order of addition of coupling reagents.

In addition to the foregoing, additional methods which can be used to generate an array of oligonucleotides on a single substrate are described in PCT Publication Nos. WO 93/09668 and WO 01/23614. High density nucleic acid arrays can also be fabricated by depositing pre-made or natural nucleic acids in predetermined positions. Synthesized or natural nucleic acids are deposited on specific locations of a substrate by light directed targeting and oligonucleotide directed targeting. Another embodiment uses a dispenser that moves from region to region to deposit nucleic acids in specific spots.

Nucleic acid hybridization simply involves contacting a probe and target nucleic acid under conditions where the probe and its complementary target can form stable hybrid duplexes through complementary base pairing. See WO 99/32660. The nucleic acids that do not form hybrid duplexes are then washed away leaving the hybridized nucleic acids to be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (*e.g.*, low temperature and/or high salt) hybrid duplexes (*e.g.*, DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus, specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (*e.g.*, higher temperature or lower salt) successful hybridization tolerates fewer mismatches. One of skill in the art will appreciate that hybridization conditions may be selected to provide any degree of stringency.

In a preferred embodiment, hybridization is performed at low stringency, in this case in 6x SSPET at 37°C (0.005% Triton X-100), to ensure hybridization and then subsequent

washes are performed at higher stringency (e.g., 1x SSPET at 37°C) to eliminate mismatched hybrid duplexes. Successive washes may be performed at increasingly higher stringency (e.g., down to as low as 0.25x SSPET at 37°C to 50°C) until a desired level of hybridization specificity is obtained. Stringency can also be increased by
5 addition of agents such as formamide. Hybridization specificity may be evaluated by comparison of hybridization to the test probes with hybridization to the various controls that can be present (e.g., expression level control, normalization control, mismatch controls, *etc.*).

In general, there is a tradeoff between hybridization specificity (stringency) and signal
10 intensity. Thus, in a preferred embodiment, the wash is performed at the highest stringency that produces consistent results and that provides a signal intensity greater than approximately 10% of the background intensity. Thus, in a preferred embodiment, the hybridized array may be washed at successively higher stringency solutions and read between each wash. Analysis of the data sets thus produced will reveal a wash
15 stringency above which the hybridization pattern is not appreciably altered and which provides adequate signal for the particular oligonucleotide probes of interest.

Probes based on the sequences of the genes described above may be prepared by any commonly available method. Oligonucleotide probes for screening or assaying a tissue or cell sample are preferably of sufficient length to specifically hybridize only to
20 appropriate, complementary genes or transcripts. Typically the oligonucleotide probes will be at least about 10, 12, 14, 16, 18, 20 or 25 nucleotides in length. In some cases, longer probes of at least 30, 40, or 50 nucleotides will be desirable.

As used herein, oligonucleotide sequences that are complementary to one or more of the genes or gene fragments described in Tables 2-4 refer to oligonucleotides that are
25 capable of hybridizing under stringent conditions to at least part of the nucleotide sequences of said genes. Such hybridizable oligonucleotides will typically exhibit at least about 75% sequence identity at the nucleotide level to said genes, preferably about 80% or 85% sequence identity or more preferably about 90% or 95% or more sequence identity to said genes (see GeneChip® Expression Analysis Manual, Affymetrix, Rev. 3,
30 which is herein incorporated by reference in its entirety).

The phrase "hybridizing specifically to" or "specifically hybridizes" refers to the binding, duplexing, or hybridizing of a molecule substantially to or only to a particular nucleotide

sequence or sequences under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA.

As used herein a "probe" is defined as a nucleic acid, capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (*i.e.*, A, G, U, C, or T) or modified bases (7-deazaguanosine, inosine, *etc.*). In addition, the bases in probes may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages.

A variety of sequencing reactions known in the art can be used to directly sequence nucleic acids for the presence or the absence of one or more polymorphisms of Tables 5-36. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert (1977) or Sanger (1977). It is also contemplated that any of a variety of automated sequencing procedures can be utilized, including sequencing by mass spectrometry (see, *e.g.* PCT International Publication No. WO 94/16101; Cohen *et al.*, 1996; and Griffin *et al.*, 1993), real-time pyrophosphate sequencing method (Ronaghi *et al.*, 1998; and Permutt *et al.*, 2001) and sequencing by hybridization (see *e.g.* Drmanac *et al.*, 2002).

Other methods of detecting polymorphisms include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA, DNA/DNA or RNA/DNA heteroduplexes (Myers *et al.*, 1985). In general, the technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing a wild-type sequence with potentially mutant RNA or DNA obtained from a sample. The double-stranded duplexes are treated with an agent who cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of a mutation or SNP (see, for example, Cotton *et al.*, 1988; and

Saleeba *et al.*, 1992). In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping polymorphisms. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches (Hsu *et al.*, 1994). Other examples include, but are not limited to, the MthLS enzyme complex of *E. coli* (Smith and Modrich *Proc.* 1996) and Cel 1 from the celery (Kulinski *et al.*, 2000) both cleave the DNA at various mismatches. According to an exemplary embodiment, a probe based on a polymorphic site corresponding to a polymorphism of Tables 5-6 is hybridized to a cDNA or other DNA product from a test cell or cells. The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039. Alternatively, the screen can be performed *in vivo* following the insertion of the heteroduplexes in an appropriate vector. The whole procedure is known to those ordinary skilled in the art and is referred to as mismatch repair detection (see *e.g.* Fakhrai-Rad *et al.*, 2004).

In other embodiments, alterations in electrophoretic mobility can be used to identify polymorphisms in a sample. For example, single strand conformation polymorphism (SSCP) analysis can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.*, 1989; Cotton *et al.*, 1993; and Hayashi 1992). Single-stranded DNA fragments of case and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence. The resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Kee *et al.*, 1991).

In yet another embodiment, the movement of mutant or wild-type fragments in a polyacrylamide gel containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.*, 1985). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature,

for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum *et al.*, 1987). In another embodiment, the mutant fragment is detected
5 using denaturing HPLC (see e.g. Hoogendoorn *et al.*, 2000).

Examples of other techniques for detecting polymorphisms include, but are not limited to, selective oligonucleotide hybridization, selective amplification, selective primer extension, selective ligation, single-base extension, selective termination of extension or invasive cleavage assay. For example, oligonucleotide primers may be prepared in
10 which the polymorphism is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.*, 1986; Saiki *et al.*, 1989). Such oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA. Alternatively, the amplification, the
15 allele-specific hybridization and the detection can be done in a single assay following the principle of the 5' nuclease assay (e.g. see Livak *et al.*, 1995). For example, the associated allele, a particular allele of a polymorphic locus, or the like is amplified by PCR in the presence of both allele-specific oligonucleotides, each specific for one or the other allele. Each probe has a different fluorescent dye at the 5' end and a quencher at
20 the 3' end. During PCR, if one or the other or both allele-specific oligonucleotides are hybridized to the template, the Taq polymerase via its 5' exonuclease activity will release the corresponding dyes. The latter will thus reveal the genotype of the amplified product.

Hybridization assays may also be carried out with a temperature gradient following the principle of dynamic allele-specific hybridization or like e.g. Jobs *et al.*, (2003); and
25 Bourgeois and Labuda, (2004). For example, the hybridization is done using one of the two allele-specific oligonucleotides labeled with a fluorescent dye, and an intercalating quencher under a gradually increasing temperature. At low temperature, the probe is hybridized to both the mismatched and full-matched template. The probe melts at a lower temperature when hybridized to the template with a mismatch. The release of the
30 probe is captured by an emission of the fluorescent dye, away from the quencher. The probe melts at a higher temperature when hybridized to the template with no mismatch. The temperature-dependent fluorescence signals therefore indicate the absence or presence of an associated allele, a particular allele of a polymorphic locus, or the like (e.g. Jobs *et al.*, 2003). Alternatively, the hybridization is done under a gradually

decreasing temperature. In this case, both allele-specific oligonucleotides are hybridized to the template competitively. At high temperature none of the two probes are hybridized. Once the optimal temperature of the full-matched probe is reached, it hybridizes and leaves no target for the mismatched probe (e.g. Bourgeois and Labuda, 2004). In the
5 latter case, if the allele-specific probes are differently labeled, then they are hybridized to a single PCR-amplified target. If the probes are labeled with the same dye, then the probe cocktail is hybridized twice to identical templates with only one labeled probe, different in the two cocktails, in the presence of the unlabeled competitive probe.

Alternatively, allele specific amplification technology that depends on selective PCR
10 amplification may be used in conjunction with the present invention. Oligonucleotides used as primers for specific amplification may carry the associated allele, a particular allele of a polymorphic locus, or the like, also referred to as "mutation" of interest in the center of the molecule, so that amplification depends on differential hybridization (Gibbs
15 *et al.*, 1989) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner, 1993). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.*, 1992). It is anticipated that in certain
20 embodiments, amplification may also be performed using Taq ligase for amplification (Barany, 1991). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known associated allele, a particular allele of a polymorphic locus, or the like at a specific site by looking for the presence or absence of amplification. The products of such an oligonucleotide
25 ligation assay can also be detected by means of gel electrophoresis. Furthermore, the oligonucleotides may contain universal tags used in PCR amplification and zip code tags that are different for each allele. The zip code tags are used to isolate a specific, labeled oligonucleotide that may contain a mobility modifier (e.g. Grossman *et al.*, 1994).

In yet another alternative, allele-specific elongation followed by ligation will form a template for PCR amplification. In such cases, elongation will occur only if there is a perfect match at the 3' end of the allele-specific oligonucleotide using a DNA
30 polymerase. This reaction is performed directly on the genomic DNA and the extension/ligation products are amplified by PCR. To this end, the oligonucleotides contain universal tags allowing amplification at a high multiplex level and a zip code for SNP identification. The PCR tags are designed in such a way that the two alleles of a SNP are amplified by different forward primers, each having a different dye. The zip code

tags are the same for both alleles of a given SNPs and they are used for hybridization of the PCR-amplified products to oligonucleotides bound to a solid support, chip, bead array or like. For an example of the procedure, see Fan *et al.* (Cold Spring Harbor Symposia on Quantitative Biology, Vol. LXVIII, pp. 69-78 2003).

- 5 Another alternative includes the single-base extension/ligation assay using a molecular inversion probe, consisting of a single, long oligonucleotide (see *e.g.* Hardenbol *et al.*, 2003). In such an embodiment, the oligonucleotide hybridizes on both side of the SNP locus directly on the genomic DNA, leaving a one-base gap at the SNP locus. The gap-filling, one-base extension/ligation is performed in four tubes, each having a different
10 dNTP. Following this reaction, the oligonucleotide is circularized whereas unreactive, linear oligonucleotides are degraded using an exonuclease such as exonuclease I of *E. coli*. The circular oligonucleotides are then linearized and the products are amplified and labeled using universal tags on the oligonucleotides. The original oligonucleotide also contains a SNP-specific zip code allowing hybridization to oligonucleotides bound to a
15 solid support, chip, and bead array or like. This reaction can be performed at a high multiplexed level.

In another alternative, the associated allele, a particular allele of a polymorphic locus, or the like is scored by single-base extension (see *e.g.* U.S. Pat. No. 5,888,819). The template is first amplified by PCR. The extension oligonucleotide is then hybridized next
20 to the SNP locus and the extension reaction is performed using a thermostable polymerase such as ThermoSequenase (GE Healthcare) in the presence of labeled ddNTPs. This reaction can therefore be cycled several times. The identity of the labeled ddNTP incorporated will reveal the genotype at the SNP locus. The labeled products can be detected by means of gel electrophoresis, fluorescence polarization (*e.g.* Chen *et al.*,
25 1999) or by hybridization to oligonucleotides bound to a solid support, chip, and bead array or like. In the latter case, the extension oligonucleotide will contain a SNP-specific zip code tag.

In yet another alternative, a SNP is scored by selective termination of extension. The template is first amplified by PCR and the extension oligonucleotide hybridizes in the
30 vicinity of the SNP locus, close to but not necessarily adjacent to it. The extension reaction is carried out using a thermostable polymerase such as ThermoSequenase (GE Healthcare) in the presence of a mix of dNTPs and at least one ddNTP. The latter has to terminate the extension at one of the allele of the interrogated SNP, but not both such that the two alleles will generate extension products of different sizes. The extension

product can then be detected by means of gel electrophoresis, in which case the extension products need to be labeled, or by mass spectrometry (see e.g. Storm *et al.*, 2003).

5 In another alternative, SNPs are detected using an invasive cleavage assay (see U.S. Pat. No. 6,090,543). There are five oligonucleotides per SNP to interrogate but these are used in a two step-reaction. During the primary reaction, three of the designed oligonucleotides are first hybridized directly to the genomic DNA. One of them is locus-specific and hybridizes up to the SNP locus (the pairing of the 3' base at the SNP locus is not necessary). There are two allele-specific oligonucleotides that hybridize in tandem
10 to the locus-specific probe but also contain a 5' flap that is specific for each allele of the SNP. Depending upon hybridization of the allele-specific oligonucleotides at the base of the SNP locus, this creates a structure that is recognized by a cleavase enzyme (U.S. Pat. No. 6,090,606) and the allele-specific flap is released. During the secondary reaction, the flap fragments hybridize to a specific cassette to recreate the same
15 structure as above except that the cleavage will release a small DNA fragment labeled with a fluorescent dye that can be detected using regular fluorescence detector. In the cassette, the emission of the dye is inhibited by a quencher.

20 *Methods to identify agents that modulate the expression of a nucleic acid encoding a gene involved in IBD (ex: Crohn's disease).*

The present invention provides methods for identifying agents that modulate the expression of a nucleic acid encoding a gene from Tables 2-4. Such methods may utilize any available means of monitoring for changes in the expression level of the nucleic acids of the invention. As used herein, an agent is said to modulate the expression of a
25 nucleic acid of the invention if it is capable of up- or down- regulating expression of the nucleic acid in a cell. Such cells can be obtained from any parts of the body such as the GI track, colon, esophagus, stomach, rectum, jujenum, ileum, mucosa, submucosa, cecum, rectum, scalp, blood, dermis, epidermis, skin cells, cutaneous surfaces, intertrigious areas, genitalia, vessels and endothelium. Some non-limiting examples of
30 cells that can be used are: muscle cells, nervous cells, blood and vessels cells, dermis, epidermis and other skin cells, T cell, mast cell, CD4+ lymphocyte, monocyte, macrophage, synovial cell, glial cell, villous intestinal cell, neutrophilic granulocyte,

eosinophilic granulocyte, keratinocyte, lamina propria lymphocyte, intraepithelial lymphocyte, epithelial cells and lymphocytes.

In one assay format, the expression of a nucleic acid encoding a gene of the invention (see Tables 2-4) in a cell or tissue sample is monitored directly by hybridization to the nucleic acids of the invention. Cell lines or tissues are exposed to the agent to be tested under appropriate conditions and time and total RNA or mRNA is isolated by standard procedures such as those disclosed in Sambrook *et al.*, (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press).

Probes to detect differences in RNA expression levels between cells exposed to the agent and control cells may be prepared as described above. Hybridization conditions are modified using known methods, such as those described by Sambrook *et al.*, and Ausubel *et al.*, as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA RNA can be accomplished in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support and the solid support exposed to at least one probe comprising at least one, or part of one of the sequences of the invention under conditions in which the probe will specifically hybridize. Alternatively, nucleic acid fragments comprising at least one, or part of one of the sequences of the invention can be affixed to a solid support, such as a silicon chip or a porous glass wafer. The chip or wafer can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences will specifically hybridize to the RNA. By examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up or down regulate expression are identified.

Methods to identify agents that modulate the activity of a protein encoded by a gene involved in IBD (ex: Crohn's disease).

The present invention provides methods for identifying agents that modulate at least one activity of the proteins described in Tables 2-4. Such methods may utilize any means of monitoring or detecting the desired activity. As used herein, an agent is said to modulate the expression of a protein of the invention if it is capable of up- or down- regulating expression of the protein in a cell. Such cells can be obtained from any parts of the body such as the GI track, colon, esophagus, stomach, rectum, jejunum, ileum, mucosa,

submucosa, cecum, rectum, scalp, blood, dermis, epidermis, skin cells, cutaneous surfaces, intertrigous areas, genitalia, vessels and endothelium. Some non-limiting examples of cells that can be used are: muscle cells, nervous cells, blood and vessels cells, dermis, epidermis and other skin cells, T cell, mast cell, CD4+ lymphocyte, 5 monocyte, macrophage, synovial cell, glial cell, villous intestinal cell, neutrophilic granulocyte, eosinophilic granulocyte, keratinocyte, lamina propria lymphocyte, intraepithelial lymphocyte, epithelial cells and lymphocytes.

In one format, the specific activity of a protein of the invention, normalized to a standard unit, may be assayed in a cell population that has been exposed to the agent to be 10 tested and compared to an unexposed control cell population may be assayed. Cell lines or populations are exposed to the agent to be tested under appropriate conditions and times. Cellular lysates may be prepared from the exposed cell line or population and a control, unexposed cell line or population. The cellular lysates are then analyzed with the probe.

15 Antibody probes can be prepared by immunizing suitable mammalian hosts utilizing appropriate immunization protocols using the proteins of the invention or antigen-containing fragments thereof. To enhance immunogenicity, these proteins or fragments can be conjugated to suitable carriers. Methods for preparing immunogenic conjugates with carriers such as BSA, KLH or other carrier proteins are well known in the art. In 20 some circumstances, direct conjugation using, for example, carbodiimide reagents may be effective; in other instances linking reagents such as those supplied by Pierce Chemical Co. (Rockford, IL) may be desirable to provide accessibility to the hapten. The hapten peptides can be extended at either the amino or carboxy terminus with a cysteine residue or interspersed with cysteine residues, for example, to facilitate linking to a 25 carrier. Administration of the immunogens is conducted generally by injection over a suitable time period and with use of suitable adjuvants, as is generally understood in the art. During the immunization schedule, titers of antibodies are taken to determine adequacy of antibody formation. While the polyclonal antisera produced in this way may be satisfactory for some applications, for pharmaceutical compositions, use of 30 monoclonal preparations is preferred. Immortalized cell lines which secrete the desired monoclonal antibodies may be prepared using standard methods, see e.g., Kohler & Milstein (1992) or modifications which affect immortalization of lymphocytes or spleen cells, as is generally known. The immortalized cell lines secreting the desired antibodies can be screened by immunoassay in which the antigen is the peptide hapten,

polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production in ascites fluid. The desired monoclonal antibodies may be recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonal antibodies or the polyclonal antisera which contain the immunologically significant portion(s) can be used as antagonists, as well as the intact antibodies. Use of immunologically reactive fragments, such as Fab or Fab' fragments, is often preferable, especially in a therapeutic context, as these fragments are generally less immunogenic than the whole immunoglobulin. The antibodies or fragments may also be produced, using current technology, by recombinant means. Antibody regions that bind specifically to the desired regions of the protein can also be produced in the context of chimeras derived from multiple species. Antibody regions that bind specifically to the desired regions of the protein can also be produced in the context of chimeras from multiple species, for instance, humanized antibodies. The antibody can therefore be a humanized antibody or a human antibody, as described in U.S. Patent 5,585,089 or Riechmann *et al.* (1988).

Agents that are assayed in the above method can be randomly selected or rationally selected or designed. As used herein, an agent is said to be randomly selected when the agent is chosen randomly without considering the specific sequences involved in the association of the protein of the invention alone or with its associated substrates, binding partners, etc. An example of randomly selected agents is the use of a chemical library or a peptide combinatorial library, or a growth broth of an organism. As used herein, an agent is said to be rationally selected or designed when the agent is chosen on a non-random basis which takes into account the sequence of the target site or its conformation in connection with the agent's action. Agents can be rationally selected or rationally designed by utilizing the peptide sequences that make up these sites. For example, a rationally selected peptide agent can be a peptide whose amino acid sequence is identical to or a derivative of any functional consensus site. The agents of the present invention can be, as examples, oligonucleotides, antisense polynucleotides, interfering RNA, peptides, peptide mimetics, antibodies, antibody fragments, small molecules, vitamin derivatives, as well as carbohydrates. Peptide agents of the invention can be prepared using standard solid phase (or solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides may be synthesized using commercially available oligonucleotide synthesis instrumentation and produced recombinantly using standard recombinant production systems. The

production using solid phase peptide synthesis is necessitated if non-gene-encoded amino acids are to be included.

Another class of agents of the present invention includes antibodies or fragments thereof that bind to a protein encoded by a gene in Tables 2-4. Antibody agents can be obtained
5 by immunization of suitable mammalian subjects with peptides, containing as antigenic regions, those portions of the protein intended to be targeted by the antibodies (see section above of antibodies as probes for standard antibody preparation methodologies).

In yet another class of agents, the present invention includes peptide mimetics that mimic the three-dimensional structure of the protein encoded by a gene from Tables 2-4.
10 Such peptide mimetics may have significant advantages over naturally occurring peptides, including, for example: more economical production, greater chemical stability, enhanced pharmacological properties (half-life, absorption, potency, efficacy, etc.), altered specificity (e.g., a broad-spectrum of biological activities), reduced antigenicity and others. In one form, mimetics are peptide-containing molecules that mimic elements
15 of protein secondary structure. The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is expected to permit molecular interactions similar to the natural molecule. In another form, peptide analogs are commonly used in the
20 pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compounds are also referred to as peptide mimetics or peptidomimetics (Fauchere, 1986; Veber & Freidinger, 1985; Evans *et al.*, 1987) which are usually developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to therapeutically useful peptides may be
25 used to produce an equivalent therapeutic or prophylactic effect. Generally, peptide mimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage using methods known in the art. Labeling of peptide mimetics usually involves covalent attachment of one or more labels, directly or through
30 a spacer (e.g., an amide group), to non-interfering position(s) on the peptide mimetic that are predicted by quantitative structure-activity data and molecular modeling. Such non-interfering positions generally are positions that do not form direct contacts with the macromolecule(s) to which the peptide mimetic binds to produce the therapeutic effect. Derivatization (e.g., labeling) of peptide mimetics should not substantially interfere with

the desired biological or pharmacological activity of the peptide mimetic. The use of peptide mimetics can be enhanced through the use of combinatorial chemistry to create drug libraries. The design of peptide mimetics can be aided by identifying amino acid mutations that increase or decrease binding of the protein to its binding partners.

5 Approaches that can be used include the yeast two hybrid method (see Chien *et al.*, 1991) and the phage display method. The two hybrid method detects protein-protein interactions in yeast (Fields *et al.*, 1989). The phage display method detects the interaction between an immobilized protein and a protein that is expressed on the surface of phages such as lambda and M13 (Amberg *et al.*, 1993; Hogrefe *et al.*, 1993).

10 These methods allow positive and negative selection for protein-protein interactions and the identification of the sequences that determine these interactions.

Method to diagnose Crohn's disease

The present invention also relates to methods for diagnosing inflammatory bowel

15 disease or a related disease, preferably Crohn's disease (CD), a disposition to such disease, predisposition to such a disease and/or disease progression. In some methods, the steps comprise contacting a target sample with (a) nucleic acid molecule(s) or fragments thereof and comparing the concentration of individual mRNA(s) with the concentration of the corresponding mRNA(s) from at least one healthy donor. An

20 aberrant (increased or decreased) mRNA level of at least one gene from Tables 2-4, at least 5 or 10 genes from Tables 2-4, at least 50 genes from Tables 2-4, at least 100 genes from Tables 2-4 or at least 200 genes from Tables 2-4 determined in the sample in comparison to the control sample is an indication of Crohn's disease or a related disease or a disposition to such kinds of diseases. For diagnosis, samples are,

25 preferably, obtained from inflamed colon tissue. Samples can also be obtained from any parts of the body such as the GI track, colon, esophagus, stomach, rectum, jejunum, ileum, mucosa, submucosa, cecum, rectum, scalp, blood, dermis, epidermis, skin cells, cutaneous surfaces, intertrigous areas, genitalia, vessels and endothelium. Some non-limiting examples of cells that can be used are: muscle cells, nervous cells, blood and

30 vessels cells, dermis, epidermis and other skin cells, T cell, mast cell, CD4+ lymphocyte, monocyte, macrophage, synovial cell, glial cell, villous intestinal cell, neutrophilic granulocyte, eosinophilic granulocyte, keratinocyte, lamina propria lymphocyte, intraepithelial lymphocyte, epithelial cells and lymphocytes.

For analysis of gene expression, total RNA is obtained from cells according to standard procedures and, preferably, reverse-transcribed. Preferably, a DNase treatment (in order to get rid of contaminating genomic DNA) is performed. Some non-limiting examples of cells that can be used are: muscle cells, nervous cells, blood and vessels cells, dermis, epidermis and other skin cells, T cell, mast cell, CD4+ lymphocyte, monocyte, macrophage, synovial cell, glial cell, villous intestinal cell, neutrophilic granulocyte, eosinophilic granulocyte, keratinocyte, lamina propria lymphocyte, intraepithelial lymphocyte, epithelial cells and lymphocytes.

The nucleic acid molecule or fragment is typically a nucleic acid probe for hybridization or a primer for PCR. The person skilled in the art is in a position to design suitable nucleic acids probes based on the information provided in the Tables of the present invention. The target cellular component, *i.e.* mRNA, *e.g.*, in colon tissue, may be detected directly in situ, *e.g.* by in situ hybridization or it may be isolated from other cell components by common methods known to those skilled in the art before contacting with a probe. Detection methods include Northern blot analysis, RNase protection, in situ methods, *e.g.* in situ hybridization, *in vitro* amplification methods (PCR, LCR, QRNA replicase or RNA-transcription/amplification (TAS, 3SR), reverse dot blot disclosed in EP-B10237362) and other detection assays that are known to those skilled in the art. Products obtained by *in vitro* amplification can be detected according to established methods, *e.g.* by separating the products on agarose or polyacrylamide gels and by subsequent staining with ethidium bromide. Alternatively, the amplified products can be detected by using labeled primers for amplification or labeled dNTPs. Preferably, detection is based on a microarray.

The probes (or primers) (or, alternatively, the reverse-transcribed sample mRNAs) can be detectably labeled, for example, with a radioisotope, a bioluminescent compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.

The present invention also relates to the use of the nucleic acid molecules or fragments described above for the preparation of a diagnostic composition for the diagnosis of Crohn's disease or a disposition to such a disease.

The present invention also relates to the use of the nucleic acid molecules of the present invention for the isolation or development of a compound which is useful for therapy of Crohn's disease. For example, the nucleic acid molecules of the invention and the data obtained using said nucleic acid molecules for diagnosis of Crohn's disease might allow

for the identification of further genes which are specifically dysregulated, and thus may be considered as potential targets for therapeutic interventions.

The invention further provides prognostic assays that can be used to identify subjects having or at risk of developing Crohn's disease. In such method, a test sample is
5 obtained from a subject and the amount and/or concentration of the nucleic acid described in Tables 2-4 is determined; wherein the presence of an associated allele, a particular allele of a polymorphic locus, or the likes in the nucleic acids sequences of this invention (see SEQ ID from Tables 2-36) can be diagnostic for a subject having or at risk of developing Crohn's. As used herein, a "test sample" refers to a biological sample
10 obtained from a subject of interest. For example, a test sample can be a biological fluid, a cell sample, or tissue. A biological fluid can be, but is not limited to saliva, serum, mucus, urine, stools, spermatozoids, vaginal secretions, lymph, amniotic liquid, pleural liquid and tears. Cells can be, but are not limited to: muscle cells, nervous cells, blood and vessels cells, dermis, epidermis and other skin cells, T cell, mast cell, CD4+
15 lymphocyte, monocyte, macrophage, synovial cell, glial cell, villous intestinal cell, neutrophilic granulocyte, eosinophilic granulocyte, keratinocyte, lamina propria lymphocyte, intraepithelial lymphocyte, epithelial cells and lymphocytes.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic,
20 polypeptide, nucleic acid such as antisense DNA or interfering RNA (RNAi), small molecule or other drug candidate) to treat Crohn's disease. Specifically, these assays can be used to predict whether an individual will have an efficacious response or will experience adverse events in response to such an agent. For example, such methods can be used to determine whether a subject can be effectively treated with an agent that
25 modulates the expression and/or activity of a gene from Tables 2-4 or the nucleic acids described herein. In another example, an association study may be performed to identify polymorphisms from Tables 5-36 that are associated with a given response to the agent, e.g., an efficacious response or the likelihood of one or more adverse events. Thus, one embodiment of the present invention provides methods for determining whether a
30 subject can be effectively treated with an agent for a disease associated with aberrant expression or activity of a gene from Tables 2-4 in which a test sample is obtained and nucleic acids or polypeptides from Tables 2-4 are detected (e.g., wherein the presence of a particular level of expression of a gene from Tables 2-4 or a particular allelic variant of such gene, such as polymorphisms from Tables 5-36 is diagnostic for a subject that

can be administered an agent to treat a disorder such as Crohn's disease). In one embodiment, the method includes obtaining a sample from a subject suspected of having Crohn's disease or an affected individual and exposing such sample to an agent. The expression and/or activity of the nucleic acids and/or genes of the invention are monitored before and after treatment with such agent to assess the effect of such agent. After analysis of the expression values, one skilled in the art can determine whether such agent can effectively treat such subject. In another embodiment, the method includes obtaining a sample from a subject having or susceptible to developing Crohn's disease and determining the allelic constitution of polymorphisms from Tables 5-36 that is associated with a particular response to an agent. After analysis of the allelic constitution of the individual at the associated polymorphisms, one skilled in the art can determine whether such agent can effectively treat such subject.

The methods of the invention can also be used to detect genetic alterations in a gene from Tables 2-4, thereby determining if a subject with the lesioned gene is at risk for a disease associated with Crohn's disease. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one alteration linked to or affecting the integrity of a gene from Tables 2-4 encoding a polypeptide or the misexpression of such gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of: (1) a deletion of one or more nucleotides from a gene from Tables 2-4; (2) an addition of one or more nucleotides to a gene from Tables 2-4; (3) a substitution of one or more nucleotides of a gene from Tables 2-4; (4) a chromosomal rearrangement of a gene from Tables 2-4; (5) an alteration in the level of a messenger RNA transcript of a gene from Tables 2-4; (6) aberrant modification of a gene from Tables 2-4, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a gene from Tables 2-4; (8) inappropriate post-translational modification of a polypeptide encoded by a gene from Tables 2-4; and (9) alternative promoter use. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a gene from Tables 2-4. A preferred biological sample is a peripheral blood sample obtained by conventional means from a subject. Another preferred biological sample is a buccal swab. Other biological samples can be, but are not limited to, urine, stools, spermatozooids, vaginal secretions, lymph, amniotic liquid, pleural liquid and tears.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, 1988; and Nakazawa *et al.*, 1994), the latter of which can be particularly useful for detecting point mutations in a gene from Tables 2-4 (see Abavaya *et al.*, 1995). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic DNA, mRNA, or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene from Tables 2-4 under conditions such that hybridization and amplification of the nucleic acid from Tables 2-4 (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with some of the techniques used for detecting a mutation, an associated allele, a particular allele of a polymorphic locus, or the like described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990), transcriptional amplification system (Kwoh *et al.*, 1989), Q-Beta Replicase (Lizardi *et al.*, 1988), isothermal amplification (e.g. Dean *et al.*, 2002); and Hafner *et al.*, 2001), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of ordinary skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low number.

In an alternative embodiment, alterations in a gene from Tables 2-4, from a sample cell can be identified by identifying changes in a restriction enzyme cleavage pattern. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicate a mutation(s), an associated allele, a particular allele of a polymorphic locus, or the like in the sample DNA. Moreover, sequence specific ribozymes (see, e.g., U.S. Pat. No. 5,498,531 or DNAzyme e.g. U.S. Pat. No. 5,807,718) can be used to score for the presence of specific associated allele, a particular allele of a polymorphic locus, or the likes by development or loss of a ribozyme or DNAzyme cleavage site.

The present invention also relates to further methods for diagnosing IBD (ex: Crohn's disease) Crohn's disease or a related disorder, preferably Inflammation, a disposition to such disorder, predisposition to such a disorder and/or disorder progression. In some methods, the steps comprise contacting a target sample with (a) nucleic molecule(s) or fragments thereof and determining the presence or absence of a particular allele of a polymorphism that confers a disorder-related phenotype (e.g., predisposition to such a disorder and/or disorder progression). The presence of at least one allele from Tables 5-36 that is associated with Crohn's disease ("associated allele"), at least 5 or 10 associated alleles from Tables 5-36, at least 50 associated alleles from Tables 5-36 at least 100 associated alleles from Tables 5-36, or at least 200 associated alleles from Tables 5-36 determined in the sample is an indication of Crohn's disease or a related disorder, a disposition or predisposition to such kinds of disorders, or a prognosis for such disorder progression. Samples may be obtained from any parts of the body such as the GI track, colon, esophagus, stomach, rectum, jejunum, ileum, mucosa, submucosa, cecum, rectum, scalp, blood, dermis, epidermis, skin cells, cutaneous surfaces, intertrigous areas, genitalia, vessels and endothelium. Some non-limiting examples of cells that can be used are: muscle cells, nervous cells, blood and vessels cells, dermis, epidermis and other skin cells, T cell, mast cell, CD4+ lymphocyte, monocyte, macrophage, synovial cell, glial cell, villous intestinal cell, neutrophilic granulocyte, eosinophilic granulocyte, keratinocyte, lamina propria lymphocyte, intraepithelial lymphocyte, epithelial cells and lymphocytes.

In other embodiments, alterations in a gene from Tables 2-4 can be identified by hybridizing sample and control nucleic acids, e.g., DNA or RNA, to high density arrays or bead arrays containing tens to thousands of oligonucleotide probes (Cronin *et al.*, 1996; Kozal *et al.*, 1996). For example, alterations in a gene from Tables 2-4 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al.*, (1996). Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations, associated alleles, particular alleles of a polymorphic locus, or the like. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants, mutations, alleles detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence a gene from Tables 2-4 and detect an associated allele, a particular allele of a polymorphic locus, or the like by comparing the sequence of the sample gene from Tables 2-4 with the corresponding wild-type (control) sequence.

5 Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert (1977) or Sanger (1977). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Bio/Techniques 19:448, 1995) including sequencing by mass spectrometry (see, e.g. PCT International Publication No. WO 94/16101; Cohen *et al.*, 1996; and Griffin *et al.* 1993), real-time pyrophosphate sequencing method (Ronaghi *et al.*, 1998; and Permutt *et al.*, 2001) and sequencing by hybridization (see e.g. Drmanac *et al.*, 2002).

Other methods of detecting an associated allele, a particular allele of a polymorphic locus, or the likes in a gene from Tables 2-4 include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA, DNA/DNA or RNA/DNA heteroduplexes (Myers *et al.*, 1985). In general, the technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type gene sequence from Tables 2-4 with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of an associated allele, a particular allele of a polymorphic locus, or the like (see, for example, Cotton *et al.*, 1988; Saleeba *et al.*, 1992). In a preferred embodiment, the control DNA or RNA can be labeled for detection, as described herein.

30 In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point an associated allele, a particular allele of a polymorphic locus, or the likes in a gene from Tables 2-4 cDNAs obtained from samples of cells. For example, the mutY enzyme of E.

coli cleaves A at G/A mismatches (Hsu *et al.*, 1994). Other examples include, but are not limited to, the MutHLS enzyme complex of *E. coli* (Smith and Modrich., 1996) and Cel 1 from the celery (Kulinski *et al.*, 2000) both cleave the DNA at various mismatches. According to an exemplary embodiment, a probe based on a gene sequence from
5 Tables 2-4 is hybridized to a cDNA or other DNA product from a test cell or cells. The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected using electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039. Alternatively, the screen can be performed *in vivo* following the insertion of the heteroduplexes in an appropriate vector. The whole procedure is known to those
10 ordinary skilled in the art and is referred to as mismatch repair detection (see e.g. Fakhrai-Rad *et al.*, 2004).

In other embodiments, alterations in electrophoretic mobility can be used to identify an associated allele, a particular allele of a polymorphic locus, or the likes in genes from Tables 2-4. For example, single strand conformation polymorphism (SSCP) analysis can
15 be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.*, 1993; see also Cotton, 1993; and Hayashi *et al.*, 1992). Single-stranded DNA fragments of sample and control nucleic acids from Tables 2-4 will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence; the resulting alteration in electrophoretic mobility
20 enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in
25 electrophoretic mobility (Kee *et al.*, 1991).

In yet another embodiment, the movement of mutant or wild-type fragments in a polyacrylamide gel containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.*, 1985). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature,
30 for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum *et al.*, 1987). In another embodiment, the mutant fragment is detected using denaturing HPLC (see e.g. Hoogendoorn *et al.*, 2000).

Examples of other techniques for detecting point mutations, an associated allele, a particular allele of a polymorphic locus, or the like include, but are not limited to, selective oligonucleotide hybridization, selective amplification, selective primer extension, selective ligation, single-base extension, selective termination of extension or invasive cleavage assay. For example, oligonucleotide primers may be prepared in which the known associated allele, particular allele of a polymorphic locus, or the like is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.*, 1986; Saiki *et al.*, 1989). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA of a number of different associated alleles, a particular allele of a polymorphic locus, or the likes where the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA. Alternatively, the amplification, the allele-specific hybridization and the detection can be done in a single assay following the principle of the 5' nuclease assay (e.g. see Livak *et al.*, 1995). For example, the associated allele, a particular allele of a polymorphic locus, or the like locus is amplified by PCR in the presence of both allele-specific oligonucleotides, each specific for one or the other allele. Each probe has a different fluorescent dye at the 5' end and a quencher at the 3' end. During PCR, if one or the other or both allele-specific oligonucleotides are hybridized to the template, the Taq polymerase via its 5' exonuclease activity will release the corresponding dyes. The latter will thus reveal the genotype of the amplified product.

The hybridization may also be carried out with a temperature gradient following the principle of dynamic allele-specific hybridization or like (e.g. Jobs *et al.*, 2003; and Bourgeois and Labuda, 2004). For example, the hybridization is done using one of the two allele-specific oligonucleotides labeled with a fluorescent dye, an intercalating quencher under a gradually increasing temperature. At low temperature, the probe is hybridized to both the mismatched and full-matched template. The probe melts at a lower temperature when hybridized to the template with a mismatch. The release of the probe is captured by an emission of the fluorescent dye, away from the quencher. The probe melts at a higher temperature when hybridized to the template with no mismatch. The temperature-dependent fluorescence signals therefore indicate the absence or presence of the associated allele, particular allele of a polymorphic locus, or the like (e.g. Jobs *et al.* *supra*). Alternatively, the hybridization is done under a gradually decreasing temperature. In this case, both allele-specific oligonucleotides are hybridized to the template competitively. At high temperature none of the two probes is hybridized. Once the optimal temperature of the full-matched probe is reached, it hybridizes and leaves no

target for the mismatched probe. In the latter case, if the allele-specific probes are differently labeled, then they are hybridized to a single PCR-amplified target. If the probes are labeled with the same dye, then the probe cocktail is hybridized twice to identical templates with only one labeled probe, different in the two cocktails, in the presence of the unlabeled competitive probe.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the present invention. Oligonucleotides used as primers for specific amplification may carry the associated allele, particular allele of a polymorphic locus, or the like of interest in the center of the molecule, so that amplification depends on differential hybridization (Gibbs *et al.*, 1989) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner, 1993). In addition it may be desirable to introduce a novel restriction site in the region of the associated allele, particular allele of a polymorphic locus, or the like to create cleavage-based detection (Gasparini *et al.*, 1992). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany, 1991). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known associated allele, a particular allele of a polymorphic locus, or the like at a specific site by looking for the presence or absence of amplification. The products of such an oligonucleotide ligation assay can also be detected by means of gel electrophoresis. Furthermore, the oligonucleotides may contain universal tags used in PCR amplification and zip code tags that are different for each allele. The zip code tags are used to isolate a specific, labeled oligonucleotide that may contain a mobility modifier (e.g. Grossman *et al.*, 1994).

In yet another alternative, allele-specific elongation followed by ligation will form a template for PCR amplification. In such cases, elongation will occur only if there is a perfect match at the 3' end of the allele-specific oligonucleotide using a DNA polymerase. This reaction is performed directly on the genomic DNA and the extension/ligation products are amplified by PCR. To this end, the oligonucleotides contain universal tags allowing amplification at a high multiplex level and a zip code for SNP identification. The PCR tags are designed in such a way that the two alleles of a SNP are amplified by different forward primers, each having a different dye. The zip code tags are the same for both alleles of a given SNP and they are used for hybridization of the PCR-amplified products to oligonucleotides bound to a solid support, chip, bead

array or like. For an example of the procedure, see Fan *et al.* (Cold Spring Harbor Symposia on Quantitative Biology, Vol. LXVIII, pp. 69-78, 2003).

Another alternative includes the single-base extension/ligation assay using a molecular inversion probe, consisting of a single, long oligonucleotide (see *e.g.* Hardenbol *et al.*,
5 2003). In such an embodiment, the oligonucleotide hybridizes on both sides of the SNP locus directly on the genomic DNA, leaving a one-base gap at the SNP locus. The gap-filling, one-base extension/ligation is performed in four tubes, each having a different dNTP. Following this reaction, the oligonucleotide is circularized whereas unreactive, linear oligonucleotides are degraded using an exonuclease such as exonuclease I of *E.*
10 *coli*. The circular oligonucleotides are then linearized and the products are amplified and labeled using universal tags on the oligonucleotides. The original oligonucleotide also contains a SNP-specific zip code allowing hybridization to oligonucleotides bound to a solid support, chip, bead array or the like. This reaction can be performed at a highly multiplexed level.

15 In another alternative, the associated allele, particular allele of a polymorphic locus, or the like is scored by single-base extension (see *e.g.* U.S. Pat. No. 5,888,819). The template is first amplified by PCR. The extension oligonucleotide is then hybridized next to the SNP locus and the extension reaction is performed using a thermostable polymerase such as ThermoSequenase (GE Healthcare) in the presence of labeled
20 ddNTPs. This reaction can therefore be cycled several times. The identity of the labeled ddNTP incorporated will reveal the genotype at the SNP locus. The labeled products can be detected by means of gel electrophoresis, fluorescence polarization (*e.g.* Chen *et al.*, 1999) or by hybridization to oligonucleotides bound to a solid support, chip, bead array or the like. In the latter case, the extension oligonucleotide will contain a SNP-specific zip
25 code tag.

In yet another alternative, the variant is scored by selective termination of extension. The template is first amplified by PCR and the extension oligonucleotide hybridizes in vicinity to the SNP locus, close to but not necessarily adjacent to it. The extension reaction is carried out using a thermostable polymerase such as ThermoSequenase (GE
30 Healthcare) in the presence of a mix of dNTPs and at least one ddNTP. The latter has to terminate the extension at one of the alleles of the interrogated SNP, but not both such that the two alleles will generate extension products of different sizes. The extension product can then be detected by means of gel electrophoresis, in which case the

extension products need to be labeled, or by mass spectrometry (see *e.g.* Storm *et al.*, 2003).

In another alternative, the associated allele, particular allele of a polymorphic locus, or the like is detected using an invasive cleavage assay (see U.S. Pat. No. 6,090,543).

5 There are five oligonucleotides per SNP to interrogate but these are used in a two step-reaction. During the primary reaction, three of the designed oligonucleotides are first hybridized directly to the genomic DNA. One of them is locus-specific and hybridizes up to the SNP locus (the pairing of the 3' base at the SNP locus is not necessary). There are two allele-specific oligonucleotides that hybridize in tandem to the locus-specific
10 probe but also contain a 5' flap that is specific for each allele of the SNP. Depending upon hybridization of the allele-specific oligonucleotides at the base of the SNP locus, this creates a structure that is recognized by a cleavase enzyme (U.S. Pat. No. 6,090,606) and the allele-specific flap is released. During the secondary reaction, the flap fragments hybridize to a specific cassette to recreate the same structure as above
15 except that the cleavage will release a small DNA fragment labeled with a fluorescent dye that can be detected using regular fluorescence detector. In the cassette, the emission of the dye is inhibited by a quencher.

Other types of markers can also be used for diagnostic purposes. For example, microsatellites can also be useful to detect the genetic predisposition of an individual to a
20 given disorder. Microsatellites consist of short sequence motifs of one or a few nucleotides repeated in tandem. The most common motifs are polynucleotide runs, dinucleotide repeats (particularly the CA repeats) and trinucleotide repeats. However, other types of repeats can also be used. The microsatellites are very useful for genetic mapping because they are highly polymorphic in their length. Microsatellite markers can
25 be typed by various means, including but not limited to DNA fragment sizing, oligonucleotide ligation assay and mass spectrometry. For example, the locus of the microsatellite is amplified by PCR and the size of the PCR fragment will be directly correlated to the length of the microsatellite repeat. The size of the PCR fragment can be detected by regular means of gel electrophoresis. The fragment can be labeled internally
30 during PCR or by using end-labeled oligonucleotides in the PCR reaction (*e.g.* Mansfield *et al.*, 1996). Alternatively, the size of the PCR fragment is determined by mass spectrometry. In such a case, however, the flanking sequences need to be eliminated. This can be achieved by ribozyme cleavage of an RNA transcript of the microsatellite repeat (Krebs *et al.*, 2001). For example, the microsatellite locus is amplified using

oligonucleotides that include a T7 promoter on one end and a ribozyme motif on the other end. Transcription of the amplified fragments will yield an RNA substrate for the ribozyme, releasing small RNA fragments that contain the repeated region. The size of the latter is determined by mass spectrometry. Alternatively, the flanking sequences are specifically degraded. This is achieved by replacing the dTTP in the PCR reaction by dUTP. The dUTP nucleosides are then removed by uracyl DNA glycosylases and the resulting abasic sites are cleaved by either abasic endonucleases such as human AP endonuclease or chemical agents such as piperidine. Bases can also be modified post-PCR by chemical agents such as dimethyl sulfate and then cleaved by other chemical agents such as piperidine (see *e.g.* Maxam and Gilbert, 1977; U.S. Pat. No. 5,869,242; and U.S. Patent pending serial No. 60/335,068).

In another alternative, an oligonucleotide ligation assay can be performed. The microsatellite locus is first amplified by PCR. Then, different oligonucleotides can be submitted to ligation at the center of the repeat with a set of oligonucleotides covering all the possible lengths of the marker at a given locus (Zirvi *et al.*, 1999). Another example of design of an oligonucleotide assay comprises the ligation of three oligonucleotides; a 5' oligonucleotide hybridizing to the 5' flanking sequence, a repeat oligonucleotide of the length of the shortest allele of the marker hybridizing to the repeated region and a set of 3' oligonucleotides covering all the existing alleles hybridizing to the 3' flanking sequence and a portion of the repeated region for all the alleles longer than the shortest one. For the shortest allele, the 3' oligonucleotide exclusively hybridizes to the 3' flanking sequence (U.S. Pat. No. 6,479,244).

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid selected from the SEQ ID of Tables 2-36, or antibody reagent described herein, which may be conveniently used, for example, in a clinical setting to diagnose patient exhibiting symptoms or a family history of a disorder or disorder involving abnormal activity of genes from Tables 2-4.

Method to treat an animal suspected of having IBD (ex: Crohn's disease)

The present invention provides methods of treating a disease associated with IBD (ex: Crohn's disease) by expressing in vivo the nucleic acids of at least one gene from Tables 2-4. These nucleic acids can be inserted into any of a number of well-known vectors for

the transfection of target cells and organisms as described below. The nucleic acids are transfected into cells, *ex vivo* or *in vivo*, through the interaction of the vector and the target cell. The nucleic acids encoding a gene from Tables 2-4, under the control of a promoter, then express the encoded protein, thereby mitigating the effects of absent, partial inactivation, or abnormal expression of a gene from Tables 2-4.

Such gene therapy procedures have been used to correct acquired and inherited genetic defects, cancer, and viral infection in a number of contexts. The ability to express artificial genes in humans facilitates the prevention and/or cure of many important human disorders, including many disorders which are not amenable to treatment by other therapies (for a review of gene therapy procedures, see Anderson, 1992; Nabel & Felgner, 1993; Mitani & Caskey, 1993; Mulligan, 1993; Dillon, 1993; Miller, 1992; Van Brunt, 1998; Vigne, 1995; Kremer & Perricaudet 1995; Doerfler & Bohm 1995; and Yu *et al.*, 1994).

Delivery of the gene or genetic material into the cell is the first critical step in gene therapy treatment of a disorder. A large number of delivery methods are well known to those of skill in the art. Preferably, the nucleic acids are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see the references included in the above section.

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

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher *et al.*, 1992; Johann *et al.*, 1992; Sommerfelt *et al.*, 1990; Wilson *et al.*, 1989; Miller *et al.*, 1999; and PCT/US94/05700).

In applications where transient expression of the nucleic acid is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (see, e.g., West *et al.*, 1987; U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, 1994; Muzyczka, 1994). Construction of recombinant AAV vectors is described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, 1985; Tratschin, *et al.*, 1984; Hermonat & Muzyczka, 1984; and Samulski *et al.*, 1989.

In particular, numerous viral vector approaches are currently available for gene transfer in clinical trials, with retroviral vectors by far the most frequently used system. All of these viral vectors utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent. pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar *et al.*, 1995; Kohn *et al.*, 1995; Malech *et al.*, 1997). PA317/pLASN was the first therapeutic vector used in a gene therapy trial (Blaese *et al.*, 1995). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors (Ellem *et al.*, 1997; and Dranoff *et al.*, 1997).

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

Replication-deficient recombinant adenoviral vectors (Ad) are predominantly used in transient expression gene therapy; because they can be produced at high titer and they
10 readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply the deleted gene function in trans. Ad vectors can transduce multiple types of tissues *in vivo*, including
15 nondividing, differentiated cells such as those found in the liver, kidney and muscle tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Stermann *et al.*, 1998). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al.*,
1996; Stermann *et al.*, 1998; Welsh *et al.*, 1995; Alvarez *et al.*, 1997; Topf *et al.*, 1998.

20 Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ 2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent
25 integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a
30 helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR

sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

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

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

Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with a nucleic acid (gene or cDNA), and re-infused back into the subject organism (e.g., patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (see, e.g., Freshney *et al.*, 1994; and the references cited therein for a discussion of how to isolate and culture cells from patients).

In one embodiment, stem cells are used in *ex vivo* procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into

other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells *in vitro* into clinically important immune cell types using cytokines such as GM-CSF, IFN- γ and TNF- α are known (see Inaba *et al.*, 1992).

- 5 Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells).
- 10 Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic nucleic acids can be also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered.

Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells, as described above. The nucleic acids from
15 Tables 2-4 are administered in any suitable manner, preferably with the pharmaceutically acceptable carriers described above. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route (see Samulski
20 *et al.*, 1989). The present invention is not limited to any method of administering such nucleic acids, but preferentially uses the methods described herein.

The present invention further provides other methods of treating IBD (ex: Crohn's disease or UC) such as administering to an individual having IBD (ex: Crohn's disease or UC) an effective amount of an agent that regulates the expression, activity or physical
25 state of at least one gene from Tables 2-4. An "effective amount" of an agent is an amount that modulates a level of expression or activity of a gene from Tables 2-4, in a cell in the individual at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80% or more, compared to a level of the respective gene from Tables 2-4 in a cell in the
30 individual in the absence of the compound. The preventive or therapeutic agents of the present invention may be administered, either orally or parenterally, systemically or locally. For example, intravenous injection such as drip infusion, intramuscular injection, intraperitoneal injection, subcutaneous injection, suppositories, intestinal lavage, oral

GI

enteric coated tablets, and the like can be selected, and the method of administration may be chosen, as appropriate, depending on the age and the conditions of the patient. The effective dosage is chosen from the range of 0.01 mg to 100 mg per kg of body weight per administration. Alternatively, the dosage in the range of 1 to 1000 mg, preferably 5 to 50 mg per patient may be chosen. The therapeutic efficacy of the treatment may be monitored by observing various parts of the GI tract, by endoscopy, barium, colonoscopy, or any other monitoring methods known in the art. Other ways of monitoring efficacy can be, but are not limited to monitoring inflammatory conditions involving the upper gastrointestinal tract such as monitoring the amelioration on the esophageal discomfort, decrease in pain, improved swallowing, reduced chest pain, decreased heartburn, decreased regurgitation of solids or liquids after swallowing or eating, decrease in vomiting, or improvement in weight gain or improvement in vitality.

The present invention further provides a method of treating an individual clinically diagnosed with IBD (ex: Crohn's disease or UC). The methods generally comprises analyzing a biological sample that includes a cell, in some cases, a GI track cell, from an individual clinically diagnosed with IBD (ex: Crohn's disease or UC) for the presence of modified levels of expression of at least 1 gene, at least 10 genes, at least 50 genes, at least 100 genes, or at least 200 genes from Tables 2-4. A treatment plan that is most effective for individuals clinically diagnosed as having a condition associated with Crohn's disease is then selected on the basis of the detected expression of such genes in a cell. Treatment may include administering a composition that includes an agent that modulates the expression or activity of a protein from Tables 2-4 in the cell. Information obtained as described in the methods above can also be used to predict the response of the individual to a particular agent. Thus, the invention further provides a method for predicting a patient's likelihood to respond to a drug treatment for a condition associated with IBD (ex: Crohn's disease or UC), comprising determining whether modified levels of a gene from Tables 2-4 is present in a cell, wherein the presence of protein is predictive of the patient's likelihood to respond to a drug treatment for the condition. Examples of the prevention or improvement of symptoms accompanied by IBD (ex: Crohn's disease or UC) that can monitored for effectiveness include prevention or improvement of diarrhea, prevention or improvement of weight loss, inhibition of bowel tissue edema, inhibition of cell infiltration, inhibition of surviving period shortening, and the like, and as a result, a preventing or improving agent for diarrhea, a preventing or improving agent for weight loss, an inhibitor for bowel tissues edema, an inhibitor for cell infiltration, an inhibitor for surviving period shortening, and the like can be identified.

The invention also provides a method of predicting a response to therapy in a subject having IBD (ex: Crohn's disease or UC) by determining the presence or absence in the subject of one or more markers associated with IBD (ex: Crohn's disease or UC) described in Tables 5-36, diagnosing the subject in which the one or more markers are present as having IBD (ex: Crohn's disease or UC), and predicting a response to a therapy based on the diagnosis e.g., response to therapy may include an efficacious response and/or one or more adverse events. The invention also provides a method of optimizing therapy in a subject having IBD (ex: Crohn's disease or UC) by determining the presence or absence in the subject of one or more markers associated with a clinical subtype of IBD (ex: Crohn's disease or UC), diagnosing the subject in which the one or more markers are present as having a particular clinical subtype of IBD (ex: Crohn's disease or UC), and treating the subject having a particular clinical subtype of IBD (ex: Crohn's disease or UC) based on the diagnosis. As an example, treatment for the fibrostenotic subtype of IBD (ex: Crohn's disease or UC) currently includes surgical removal of the affected, strictured part of the bowel.

Thus, while there are a number of treatments for IBD (ex: Crohn's disease or UC) currently available, they all are accompanied by various side effects, high costs, and long complicated treatment protocols, which are often not available and effective in a large number of individuals. Accordingly, there remains a need in the art for more effective and otherwise improved methods for treating and preventing IBD (ex: Crohn's disease or UC). Thus, there is a continuing need in the medical arts for genetic markers of IBD (ex: Crohn's disease or UC) and guidance for the use of such markers. The present invention fulfills this need and provides further related advantages.

EXAMPLES

Example 1: Identification of cases and controls

All individuals were sampled from the Quebec founder population (QFP). Membership in the founder population was defined as having four grandparents with French Canadian family names who were born in the Province of Quebec, Canada or in adjacent areas of the Provinces of New Brunswick and Ontario or in New England or New York State. The Quebec population is characterized both by extended LD and by decreased genetic heterogeneity. The increased extent of LD allows the detection of disease associated genes using a reasonable marker density, while still allowing the increased meiotic resolution of population-based mapping. The specific combination of age in generations, optimal number of founders and large present population size makes the QFP optimal for LD-based gene mapping.

Patient inclusion criteria for the study include diagnosis for Crohn's disease by any one of the following: a colonoscopy, a radiological examination with barium, an abdominal surgical operation or a biopsy or a surgical specimen. The colonoscopy diagnosis consists of observing linear, deep or serpentigenous ulcers, pseudopolyps, or skip areas. The barium radiological examination consists of the detection of strictures, ulcerations and string signs by observing the barium enema and the small bowel followed through an NMRI series.

Patients that were diagnosed with ulcerative colitis, infectious colitis or other intestinal diseases were excluded from the study. All human sampling was subject to ethical review procedures.

All enrolled QFP subjects (patients and controls) provided a 30 ml blood sample (3 barcoded tubes of 10 ml). Samples were processed immediately upon arrival at Genizon's laboratory. All samples were scanned and logged into a LabVantage Laboratory Information Management System (LIMS), which served as a hub between the clinical data management system and the genetic analysis system. Following centrifugation, the buffy coat containing the white blood cells was isolated from each tube. Genomic DNA was extracted from the buffy coat from one of the tubes, and stored at 4°C until required for genotyping. DNA extraction was performed with a commercial kit using a guanidine hydrochloride based method (FlexiGene, Qiagen) according to the manufacturer's instructions. The extraction method yielded high molecular weight DNA,

and the quality of every DNA sample was verified by agarose gel electrophoresis. Genomic DNA appeared on the gel as a large band of very high molecular weight. The remaining two buffy coats were stored at -80°C as backups.

The QFP samples were collected as family trios consisting of Crohn's disease subjects and two first degree relatives. Of the 500 trios, 477 were Parent, Parent, Child (PPC) trios; the remainders were Parent, Child, Child (PCC) trios. Only the PPC trios were used for the analysis reported here because they produced equal numbers of more accurately estimated case and control haplotypes than the PCC trios. 382 trios were used in the genome wide scan. One member of each trio was affected with Crohn's disease. For the 382 trios used in the genome wide scan, these included 189 daughters, 90 sons, 54 mothers and 49 fathers. When a child was the affected member of the trio, the two non-transmitted parental chromosomes (one from each parent) were used as controls, when one of the parents was affected, that person's spouse provided the control chromosomes. The recruitment of trios allowed a more precise determination of long extended haplotypes.

Example 2: Genome Wide Association

Genotyping was performed using Perlegen's ultra-high-throughput platform. Marker loci were amplified by PCR and hybridized to wafers containing arrays of oligonucleotides. Allele discrimination was performed through allele-specific hybridization. In total, 248,535 SNPs, distributed as evenly as possible throughout the genome, were genotyped on the 382 QFP trios for a total of 372,802,500 genotypes. These markers were mostly selected from various databases including the ~1.6 million SNP database of Perlegen Life Sciences (Patil, 2001); several thousand were obtained from the HapMap consortium database and/or dbSNP at NCBI. The SNPs were chosen to maximize uniformity of genetic coverage and to cover a distribution of allele frequencies. All SNPs that did not pass the quality controls for the assay, that is, that had a minor allele frequency of less than 1%, a Mendelian error rate within trios greater than 1%, that deviated significantly from the Hardy-Weinberg equilibrium, or that had excessive missing data (cut-off at 1% missing values or higher) were removed from the analysis. Genetic analysis was performed on a total of 165,803 SNPs (158,811 autosomal, 6851 X chromosome and 141 Y chromosome). The average gap size was approximately 17 kb.

Of the 165,803 markers, ~140,000 had a minor allele frequency (MAF) greater than 10% for the QFP.

5 The genotyping information was entered into a Unified Genotype Database (a proprietary database under development) from which it was accessed using custom-built programs for export to the genetic analysis pipeline. Analyses of these genotypes were performed with the statistical tools described in Example 3. The GWS permitted the identification of 371 candidate chromosomal regions linked to Crohn's disease (Table 1).

Example 3: Genetic Analysis

10 1. *Dataset quality assessment*

Prior to performing any analysis, the dataset from the GWS was verified for completeness of the trios. The programs Famcheck and Fampull removed any trios with abnormal family structure or missing individuals (e.g. trios without a proband, duos, singletons, etc.), and calculated the total number of complete trios in the dataset. The
15 trios were also tested to make sure that no subjects within the cohort were related more closely than second cousins (6 meiotic steps).

Subsequently, the program DataStats was used to calculate the following statistics per marker and per family:

20 Minor allele frequency (MAF) for each marker; Missing values for each marker and family; Hardy Weinberg Equilibrium for each marker; and Mendelian segregation error rate.

The following acceptance criteria were applied for internal analysis purposes:

25 MAF > 1%;
Missing values < 1%;
Observed non-Mendelian segregation < 0.33%;
Non significant deviation in allele frequencies from Hardy Weinberg equilibrium.

Markers or families not meeting these criteria were removed from the dataset in the following step. Analyses of variance were performed using the algorithm GenAnova, to
30 assess whether families or markers have a greater effect on missing values and/or non-

Mendelian segregation. This was used to determine the smallest number of data points to remove from the dataset in order to meet the requirements for missing values and non-Mendelian segregation. The families and/or markers were removed from the dataset using the program DataPull, which generates an output file that is used for subsequent analysis of the genotype data.

2. Phase Determination

The program PhaseFinderSNP2.0 was used to determine phase from trio data on a marker-by-marker, trio-by-trio basis. The output file contains haplotype data for all trio members, with ambiguities present when all trio members are heterozygous or where data is missing. The program AllHaps2PatCtrl was then used to determine case and control haplotypes and to prepare the data in the proper input format for the next stage of analysis, using the expectation maximization algorithm, PL-EM, to call phase on the remaining ambiguities. This stage consists of several modules for resolution of the remaining phase ambiguities. PLEMinOut1 was first used to recode the haplotypes for input into the PL-EM algorithm in 15-marker blocks for the genome wide scan data and for 11 marker blocks for fine and ultra-fine mapping data sets. The haplotype information was encoded as genotypes, allowing for the entry of known phase into the algorithm; this method limits the possible number of estimated haplotypes conditioned on already known phase assignments. The PL-EM algorithm was used to estimate haplotypes from the "pseudo-genotype" data in 11 or 15-marker windows, advancing in increments of one marker across the chromosome. The results were then converted into multiple haplotype files using the program PLEMinOut2. Subsequently PLEMBlockGroup was used to convert the individual 11 or 15-marker block files into one continuous block of haplotypes for the entire chromosome and to generate files for further analysis by LDSTATS and SINGLETYPE. PLEMBlockGroup takes the consensus estimation of the allele call at each marker over all separate estimations (most markers are estimated 11-15 different times as the 11 or 15 marker blocks pass over their position).

3. Haplotype association analysis

Haplotype association analysis was performed using the program LDSTATS. LDSTATS tests for association of haplotypes with the disease phenotype. The algorithms LDSTATS (v2.0) and LDSTATS (v4.0) define haplotypes using multi-marker windows

that advance across the marker map in one-marker increments. Windows can contain any odd number of markers specified as a parameter of the algorithm. Other marker windows can also be used. At each position the frequency of haplotypes in cases and controls was calculated and a chi-square statistic was calculated from case control frequency tables. For LDSTATS v2.0, the significance of the chi-square for single marker and 3-marker windows was calculated as Pearson's chi-square with degrees of freedom. Larger windows of multi-allelic haplotype association were tested using Smith's normalization of the square root of Pearson's Chi-square. In addition, LDSTATS v2.0 calculates Chi-square values for the transmission disequilibrium test (TDT) for single markers in situations where the trios consisted of parents and an affected child.

LDSTATS v4.0 calculates significance of chi-square values using a permutation test in which case-control status is randomly permuted until 350 permuted chi-square values are observed that are greater than or equal to chi-square value of the actual data. The P value is then calculated as 350 / the number of permutations required.

The best signal at a given location was determined by comparing the significance (p-value) of the association with Crohn disease for window sizes of 1, 3, 5, 7, and 9 SNPs, and selecting the most significant window. For a given window size at a given location, the association with Crohn disease was evaluated by comparing the overall distribution of haplotypes in the cases with the overall distribution of haplotypes in the controls. Haplotypes with a relative risk greater than one increase the risk of developing Crohn disease while haplotypes with a relative risk less than one are protective and decrease the risk.

4. Singletype analysis

The SINGLETYPE algorithm assesses the significance of case-control association for single markers using the genotype data from the laboratory as input in contrast to LDSTATS single marker window analyses, in which case-control alleles for single markers from estimated haplotypes in file, hapatctr.txt, as input. SINGLETYPE calculates P values for association for both alleles, 1 and 2, as well as for genotypes, 11, 12, and 22, and plots these as $-\log_{10}$ P values for significance of association against marker position.

5. Conditional Haplotype Analyses

Conditional haplotype analyses were performed on subsets of the original set of 382 cases and 382 controls using the program LDSTATS (v2.0). The selection of a subset of cases and their matched controls was based on the carrier status of cases at a gene or locus of interest. We selected the gene NOD2 (alias CARD15) on chromosome 16 and the gene IL23R on chromosome 1 based on our association findings using LDSTAT (v2.0) in our fine mapping of these two loci with 477 trios (see below). Additionally, we selected the gene IRAK2 on chromosome 3 and identified haplotypes based on SNPs in the genome wide scan on 382 trios. The most significant association signal in NOD2, using build 35, was obtained with a haplotype window of size 7 containing SNPs corresponding to SEQ IDs 21467, 21468, 21469, 21470, 21471, 21472, 21473 (see Table below for conversion to the specific DNA alleles used). A reduced haplotype diversity was observed and we selected two sets of risk haplotypes for conditional analyses. The first set consisted of haplotypes 2121222 and 1121211 and the second set contained the above two haplotypes and haplotype 2121211. Using the first set, we partitioned the cases into two groups; the first group consisting of those cases that were carrier of a risk haplotype and the second group consisting of the remaining cases, the non-carriers. The resulting sample sizes were respectively 125 and 227. The sample sizes were 200 and 152 for the two groups when we partitioned the cases into two groups using the second set containing three risk haplotypes. LDSTAT (v2.0) was run in each group and regions showing association with Crohn's disease are reported in Tables 5-20. Regions associated with Crohn's disease in the group of carriers (B35#has_NOD2_cr_3haps) indicate the presence of an epistatic interaction between risk factors in those regions and risk factors in NOD2 (Table 5). Regions associated with Crohn's disease in the group of non carriers (B35#not_NOD2_cr_2haps) indicate the existence of risk factors acting independently of NOD2 (Table 6). The most significant association signal in NOD2, using build 36, was obtained with a haplotype window of size 7 containing SNPs corresponding to SEQ IDs 21474, 21467, 21468, 21469, 21470, 21471, 21472 (see Table below for conversion to the specific DNA alleles used). A reduced haplotype diversity was observed and we selected a set of risk haplotypes and a set of protective haplotypes for conditional analyses. The risk set consisted of haplotypes 1112121 and 1212122, while excluding genotype 1212122/1221121 due to dominance effects. The protective set consisted of allele 1222211, excluding heterozygote carriers of either allele 1112121 or 1212122, and of homozygote individuals with genotype 1221121/1221121. The resulting sample sizes were 132 and 238 for the subsetting based on the risk set and 158 and 212 for the protective set, respectively for

the carriers and the non-carriers. LDSTAT (v2.0) was run in each group and regions showing association with Crohn's disease are reported in Tables 5-20. Regions associated with Crohn's disease in the group of non-carriers of a risk haplotype (B36#not_NOD2_cr) and in the group of carriers of a protective haplotype (B36#has_NOD2_cp) indicate the existence of risk factors acting independently of NOD2 (Tables 20 and 17).

A second conditional analysis was performed using the gene IL23R on chromosome 1. The most significant association in IL23R, using build 35, was obtained with a haplotype window of size 9 containing SNPs corresponding to SEQ IDs 21481, 21482, 21483, 21484, 21485, 21486, 21487, 21488, 21489 (see Table below for conversion to the specific DNA alleles used). A reduced haplotype diversity was observed and we selected one set of protective haplotypes and a risk haplotype for conditional analyses. The protective set consisted of haplotypes 212111122 and 212111121 and the risk haplotype was 221211121. However, due to dominance effects involving the risk haplotype, we also considered the risk haplotype 221211121 while excluding heterozygotes involving haplotypes 122122212, 222122211 or 212111122 and the risk haplotype. Using the set of protective haplotypes, we partitioned the cases into two groups; the first group consisting of those cases that were carrier of a protective haplotype and the second group consisting of the remaining cases, the non-carriers. The resulting sample sizes were respectively 204 and 162. LDSTATS (v2.0) was run in each group and regions showing association with Crohn's disease are reported in Tables 5-20. Regions associated with Crohn's disease in the group of carriers (B35#has_IL23R_cp2hap), indicate the existence of risk factors acting independently of IL23R. Regions associated with Crohn's disease in the group of non-carriers (B35#not_IL23R_cp2hap) indicate the presence of an epistatic interaction between risk factors in those regions and risk factors in IL23R (Tables 7 and 10). We repeated the process of partitioning the cases into two groups using the risk haplotype in IL23R. The carriers were cases with a haplotype in the risk set, the non-carriers were the remaining set of cases. The sample sizes for the two groups were 184 and 182 respectively. Regions associated with Crohn's disease in the group of carriers (B35#has_IL23R_cr) and (B35#has_IL23R_cr_not3) indicate the presence of an epistatic interaction between risk factors in those regions and risk factors in IL23R (Tables 8 and 9). Regions associated with Crohn's disease in the group of non-carriers (B35#not_IL23R_cr) and (B35#not_IL23R_cr_not3) indicate the presence of risk factors acting independently of IL23R (Tables 11 and 12). The most significant association in IL23R, using build 36, was

obtained with a haplotype window of size 5 containing SNPs corresponding to SEQ IDs 21488, 21489, 21492, 21493, 21494 (see Table below for conversion to the specific DNA alleles used). A reduced haplotype diversity was observed and we selected one set of risk haplotypes for conditional analyses. The risk set consisted of haplotype 21222, while
5 excluding heterozygote genotypes with haplotypes 12122 or 22122 and haplotype 21222 due to dominance effects. We partitioned the cases into two groups; the first group consisting of those cases that were carrier of a haplotype in the risk set and the second group consisting of the remaining cases, the non-carriers. The resulting sample sizes were respectively 194 and 174. LDSTATS (v2.0) was run in each group and regions
10 showing association with Crohn's disease are reported in Tables 5-20. Regions associated with Crohn's disease in the group of carriers (B36#has_IL23R_cr) indicate the presence of an epistatic interaction between risk factors in those regions and risk factors in IL23R (Table 15). Regions associated with Crohn's disease in the group of non-carriers of a risk haplotype (B36#not_IL23R_cr) indicate the existence of risk factors
15 acting independently of IL23R (Table 16). When considering the program LDSTATS (v4.0), the most significant association in IL23R, using build 36, was obtained with a haplotype window of size 3 containing SNPs corresponding to SEQ IDs 21482, 21483, 21484 (see Table below for conversion to the specific DNA alleles used). A reduced haplotype diversity was observed and we selected one set of risk and a set of protective
20 haplotypes for conditional analyses. The risk set consisted of genotypes 212/221 and 212/212. We partitioned the cases into two groups; the first group consisting of those cases that were carrier of a genotype in the risk set and the second group consisting of the remaining cases, the non-carriers. The resulting sample sizes were respectively 112 and 256. The protective set consisted of genotypes 121/221 and 121/121. As was done
25 with the risk set, we partitioned the cases into two groups; the first group consisting of those cases that were carrier of a genotype in the protective set and the second group consisting of the remaining cases, the non-carriers. The resulting sample sizes were respectively 111 and 257. LDSTATS (v2.0) was run in each group and regions showing association with Crohn's disease are reported in Tables 5-20. Regions associated with
30 Crohn's disease in the group of carriers (B36#has_IL23R -1_cr) indicate the presence of an epistatic interaction between risk factors in those regions and risk factors in IL23R (Table 14). Regions, associated with Crohn's disease in the group of non-carriers of a risk haplotype (B36#not_IL23R -1cr) and in the group of carriers of a protective haplotype (B36#has_IL23R -1_cp) indicate the existence of risk factors acting
35 independently of IL23R (Tables 18 and 13).

A third conditional analysis was performed using the gene IRAK2 on chromosome 3. The most significant association in IRAK2, using build 36, was obtained with a single SNP corresponding to SEQ ID 21498 (see Table below for conversion to the specific DNA alleles used). A reduced haplotype diversity was observed and we selected a risk
5 genotype for conditional analyses. The risk set consisted of genotype 2/2. We partitioned the cases into two groups; the first group consisting of those cases that were carrier of the risk genotype and the second group consisting of the remaining cases, the non-carriers. The resulting sample sizes were respectively 111 and 259. LDSTATS (v2.0) was run in each group and regions showing association with Crohn's disease are
10 reported in Table 5-20. Regions, associated with Crohn's disease in the group of non carriers of a risk haplotype (B36#not_IRAK2_cr) indicate the existence of risk factors acting independently of IRAK2 (Table 19).

For each region that was associated with Crohn's disease in the conditional analyses, we report in Tables 21-36 the allele frequencies and the relative risk (RR) for
15 the haplotypes contributing to the best signal at each SNP in the region. The best signal at a given location was determined by comparing the significance (p-value) of the association with Crohn's disease for window sizes of 1, 3, 5, 7, and 9 SNPs, and selecting the most significant window. For a given window size at a given location, the association with Crohn's disease was evaluated by comparing the overall distribution of
20 haplotypes in the cases with the overall distribution of haplotypes in the controls. Haplotypes with a relative risk greater than one increase the risk of developing Crohn's disease while haplotypes with a relative risk less than one are protective and decrease the risk.

DNA alleles used in haplotypes (NOD2/CARD15 Build 35)

SeqID	21475	21476	21477	21478	21479	21480	21480
Position B35	49303427	49305205	49308676	49308899	49310925	49314041	49314275
Allele	T/C	T/C	T/G	A/G	A/G	G/C	T/C
2121222	C	T	G	A	G	C	C
1121211	T	T	G	A	G	G	T
2121211	C	T	G	A	G	G	T

DNA alleles used in haplotypes (IL23R Build 35)

SeqID	21481	21482	21483	21484	21485	21486	21487	21490	21491
Position B35	67381537	67382937	67384786	67387749	67388943	67390943	67392949	67395507	67397137
Alleles	T/G	A/G	A/G	T/G	T/C	A/C	T/C	T/C	T/C
221211121	G	G	A	G	T	A	T	C	T
122122212	T	G	G	T	C	C	C	T	C
222122211	G	G	G	T	C	C	C	T	T
212111122	G	A	G	T	T	A	T	C	C
212111121	G	A	G	T	T	A	T	C	T

DNA alleles used in haplotypes (NOD2/CARD15 Build 36)

SeqID	21474	21475	21476	21477	21478	21479	21480
Position B36	49302189	49303427	49305205	49308676	49308899	49310925	49314041
Alleles	A/G	T/C	T/C	T/G	A/G	A/G	G/C
1112121	A	T	T	G	A	G	G
1212122	A	C	T	G	A	G	C
1221121	A	C	C	T	A	G	G
1222211	A	C	C	G	G	A	G

5

DNA alleles used in haplotypes (IL23R Build 36)

SeqID	21490	21491	21492	21493	21494
Position B36	67456074	67457704	67457975	67458031	67459652
Alleles	T/C	T/C	T/C	A/G	T/C
12122	T	C	T	G	C
21222	C	T	C	G	C
22122	C	C	T	G	C

DNA alleles used in haplotypes (IL23R Build 36 LDV4)

SeqID	21495	21496	21497
Position B36	67443504	67445353	67448316
Alleles	A G	A G	T G
121	A	G	T
212	G	A	G
221	G	G	T

DNA alleles used in haplotypes (IRAK2 Build 36)

SeqID	21498
Position B36	10253980
Alleles	A C
1	A
2	C

Example 4: Gene identification and characterization

A series of gene characterization was performed for each candidate region described in Table 1. Any gene or EST mapping to the interval based on public map data or proprietary map data was considered as a candidate Crohn's disease gene. The approach used to identify all genes located in the critical regions is described below.

Public gene mining

Once regions were identified using the analyses described above, a series of public data mining efforts were undertaken, with the aim of identifying all genes located within the critical intervals as well as their respective structural elements (*i.e.*, promoters and other regulatory elements, UTRs, exons and splice sites). The initial analysis relied on annotation information stored in public databases (e.g. NCBI, UCSC Genome Bioinformatics, Entrez Human Genome Browser, OMIM – see below for database URL information). Tables 2-4 lists the genes that have been mapped to the 654 candidate regions.

Database URLs	
Name	URL
Biocarta	http://www.biocarta.com/
BioCyc	http://www.biocyc.org/
Biomolecular Interaction Network Database (BIND)	http://bind.ca/
Database of Interacting Proteins	http://dip.doe-mbi.ucla.edu/
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo/
Human Genome Browser	http://www.ensembl.org/Homo_sapiens/
Interdom	http://interdom.lit.org.sg/help/term.php
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/

Molecular Interactions Database (MINT)	http://mint.bio.uniroma2.it/mint/
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
OmniViz	http://www.omniviz.com/applications/omni_viz.htm
Pathway Enterprise	http://www.omniviz.com/applications/pathways.htm
Reactome	http://www.reactome.org/
Transpath	http://www.biobase.de/pages/products/transpath.html
UCSC Genome Bioinformatics	http://genome.ucsc.edu/index.html?org=Human
UniGene	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene

For some genes the available public annotation was extensive, whereas for others very little was known about a gene's function. Customized analysis was therefore performed to characterize genes that corresponded to this latter class. Importantly, the presence of rare splice variants and artifactual ESTs was carefully evaluated. Subsequent cluster analysis of novel ESTs provided an indication of additional gene content in some cases. The resulting clusters were graphically displayed against the genomic sequence, providing indications of separate clusters that may contribute to the same gene, thereby facilitating development of confirmatory experiments in the laboratory. While much of this information was available in the public domain, the customized analysis performed revealed additional information not immediately apparent from the public genome browsers.

A unique consensus sequence was constructed for each splice variant and a trained reviewer assessed each alignment. This assessment included examination of all putative splice junctions for consensus splice donor/acceptor sequences, putative start codons,

consensus Kozak sequences and upstream in-frame stops, and the location of polyadenylation signals. In addition, conserved noncoding sequences (CNSs) that could potentially be involved in regulatory functions were included as important information for each gene. The genomic reference and exon sequences were then archived for future reference. A master assembly that included all splice variants, exons and the genomic structure was used in subsequent analyses (*i.e.*, analysis of polymorphisms). Table 3 lists gene clusters based on the publicly available EST and cDNA clustering algorithm, ECGene.

An important component of these efforts was the ability to visualize and store the results of the data mining efforts. A customized version of the highly versatile genome browser GBrowse (<http://www.gmod.org/>) was implemented in order to permit the visualization of several types of information against the corresponding genomic sequence. In addition, the results of the statistical analyses were plotted against the genomic interval, thereby greatly facilitating focused analysis of gene content.

Computational Analysis of Genes and GeneMaps

In order to assist in the prioritization of candidate genes for which minimal annotation existed, a series of computational analyses were performed that included basic BLAST searches and alignments to identify related genes. In some cases this provided an indication of potential function. In addition, protein domains and motifs were identified that further assisted in the understanding of potential function, as well as predicted cellular localization.

A comprehensive review of the public literature was also performed in order to facilitate identification of information regarding the potential role of candidate genes in the pathophysiology of Crohn's disease. In addition to the standard review of the literature, public resources (Medline and other online databases) were also mined for information regarding the involvement of candidate genes in specific signaling pathways. A variety of pathway and yeast two hybrid databases were mined for information regarding protein-protein interactions. These included BIND, MINT, DIP, Interdom, and Reactome, among others. By identifying homologues of genes in the Crohn's candidate regions and exploring whether interacting proteins had been identified already, knowledge regarding the GeneMaps for Crohn's disease was advanced. The pathway information gained from the use of these resources was also integrated with the literature review efforts, as described above.

Genes identified in the WGAS were evaluated using the Ingenuity Pathway Analysis application (IPA, Ingenuity systems) in order to identify direct biological interactions between these genes, and also to identify molecular regulators acting on those genes (indirect interactions) that could be also involved in CD. The purpose of this effort was to decipher the molecules involved in contributing to CD. These gene interaction networks are very valuable tools in the sense that they facilitate extension of the map of gene products that could represent potential drug targets for CD.

Expression Studies

In order to determine the expression patterns for genes, relevant information was first extracted from public databases. The UniGene database, for example, contains information regarding the tissue source for ESTs and cDNAs contributing to individual clusters. This information was extracted and summarized to provide an indication in which tissues the gene was expressed. Particular emphasis was placed on annotating the tissue source for bona fide ESTs, since many ESTs mapped to Unigene clusters are artifactual. In addition, SAGE and microarray data, also curated at NCBI (Gene Expression Omnibus), provided information on expression profiles for individual genes. Particular emphasis was placed on identifying genes that were expressed in tissues known to be involved in the pathophysiology of Crohn's disease, e.g. intestinal and immune system tissues.

Polymorphism analysis

Polymorphisms identified in candidate genes, including those from the public domain as well as those identified by sequencing candidate genes and regions, are evaluated for potential function. Initially, polymorphisms are examined for potential impact upon encoded proteins. If the protein is a member of a gene family with reported 3-dimensional structural information, this information is used to predict the location of the polymorphism with respect to protein structure. This information provided insight into the potential role of polymorphisms in altering protein or ligand interactions, as well as suitability as a drug target. In a second phase of analysis we evaluate the potential role of polymorphisms in other biological phenomena, including regulation of transcription, splicing and mRNA stability, etc. There are many examples of the functional involvement of naturally occurring polymorphisms in these processes. As part of this analysis, polymorphisms located in promoter or other regulatory elements, canonical splice sites,

exonic and intronic splice enhancers and repressors, conserved noncoding sequences and UTRs are localized.

Example 5: *SNP and Polymorphism Discovery (SNPD)*

Candidate regions are selected for sequencing in order to identify all polymorphisms. In cases where the critical interval, identified by fine mapping, is relatively small (~50 kb), the entire region, including all introns, is sequenced to identify polymorphisms. In situations where the region is large (>50 kb), candidate genes are prioritized for sequencing, and/or only functional gene elements (promoters, exons and splice sites) are sequenced.

The samples sequenced are selected according to which haplotypes contribute to the association signal observed in the region. The purpose is to select a set of samples that covered all the major haplotypes in the given region. Each major haplotype must be present in a few copies. The first step therefore consisted of determining the major haplotypes in the region to be sequenced.

Once a region is defined with the two boundary markers, all the markers used in fine mapping that are located within the region are used to determine the major haplotypes. Long haplotypes covering the whole region are thus inferred using the middle marker as an anchor. The results included two series of haplotype themes that define the major haplotypes, comparing the cases and the controls. This exercise is repeated using an anchor in the peripheral regions to ensure that major haplotype subsets that are not anchored at the original middle marker are not missed.

Once the major haplotypes are determined as described above, appropriate genomic DNA samples are selected such that each major haplotype and haplotype subset were represented in at least two to four copies.

The sequencing protocol included the following steps, once a region was delimited:

1. Primer design

The design of the primers is performed using a proprietary primer design tool. A primer quality control step was included in the primer design process. Primers that successfully passed the control quality process are synthesized by Integrated DNA Technologies

(IDT). The sense and anti-sense oligos are separated such that the sense oligos were placed on one plate in the same position as their anti-sense counterparts on another plate. Two additional plates are created from each storage plate, one for use in PCR and the other for sequencing. For PCR, the sense and anti-sense oligos of the same pair are combined in the same well to achieve a final concentration of 1.5 μ M for each oligonucleotide.

2. PCR optimization

PCR conditions are optimized by testing a variety of conditions that included varying salt concentrations and temperatures, as well as including various additives. PCR products are checked for robust amplification and minimal background by agarose gel electrophoresis.

3. PCR on selected samples

PCR products used for sequencing are amplified using the conditions chosen during optimization. The PCR products are purified free of salts, dNTPs and unincorporated primers by use of a MultiScreen PCR384 filter plate manufactured by Millipore. Following PCR, the amplicons are quantified by use of a lambda/Hind III standard curve. This is done to ensure that the quantity of PCR product required for sequencing had been generated. The raw data is measured against the standard curve data in Excel by use of a macro.

4. Sequencing

Sequencing of PCR products is performed by DNA Landmarks using ABI 3730 capillary sequencing instruments.

5. Sequence analysis

The ABI Prism SeqScape software (Applied Biosystems) is used for SNP identification. The chromatogram trace files are imported into a SeqScape sequencing project and the base calling is automatically performed. Sequences are then aligned and compared to each other using the SeqScape program. The base calling is checked manually, base by base; editing was performed if needed.

Example 6: *Ultra Fine Mapping (UFM)*

Once polymorphisms are identified by sequencing efforts as described in Example 5, additional genotyping of all newly found polymorphisms is performed on the samples used in the fine mapping study. Various types of genotyping assays may need to be utilized based on the type of polymorphism identified (*i.e.*, SNP, indel, microsatellite). The assay type can be, but is not restricted to, Sentrix Assay Matrix on Illumina BeadStations, microsatellite on MegaBACE, SNP on ABI or Orchid. The frequencies of genotypes and haplotypes in cases and controls are analyzed in a similar manner as the GWS data. By examining all SNPs in a region, polymorphisms are identified that increase an individual's susceptibility to Crohn's disease. The result of ultra-fine mapping is to identify the polymorphism that is most associated with disease phenotype as part of the search for the actual DNA polymorphism that confers susceptibility to disease.

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

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and H (D. N. Glover ed., 4); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); *Mullis et al.* U.S. Patent No. 4,683,195; *Nucleic Acid Hybridization* (B.D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Haines & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. 1. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes*

(IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J.H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu *et al.* eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

REFERENCES

Aaltonen, J., M. P. Laitinen, et al. (1999). "Human growth differentiation factor 9 (GDF-9) and its novel homolog GDF-9B are expressed in oocytes during early folliculogenesis." *J Clin Endocrinol Metab.* 84(8): 2744-50.

Abbott, D. W., A. Wilkins, et al. (2004). "The Crohn's Disease Protein, NOD2, Requires RIP2 in Order to Induce Ubiquitylation of a Novel Site on NEMO." *Curr Biol* 14(24): 2217-27.

Adcock, I. M., M. J. Peters, et al. (1993). "Transcription factor interactions in human lung." *Biochem Soc Trans* 21(Pt 3)(3): 277S.

Agnholt et al (2004). Increased production of granulocyte-macrophage colony-stimulating factor in Crohn's disease--a possible target for infliximab treatment. *Eur J Gastroenterol Hepatol.* 16(7):649-55.

Aithal, G. P., C. P. Day, et al. (2001). "Association of single nucleotide polymorphisms in the interleukin-4 gene and interleukin-4 receptor gene with Crohn's disease in a British population." *Genes Immun* 2(1): 44-7.

Alvarado-Kristensson, M. and T. Andersson (2005). "Protein phosphatase 2A regulates apoptosis in neutrophils by dephosphorylating both p38 MAPK and its substrate caspase 3." *J Biol Chem.* 280(7): 6238-44.

Alwine, J. C., D. J. Kemp, et al. (1977). "Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes." *Proc Natl Acad Sci U S A* 74(12): 5350-4.

Anderson, W. F. (1992). "Human gene therapy." *Science* 256(5058): 808-13.

Andres, P. G. and L. S. Friedman (1999). "Epidemiology and the natural course of inflammatory bowel disease." *Gastroenterol Clin North Am* 28(2): 255-81, vii.

Angeloni, D., F. M. Duh, et al. (2003). "C to A single nucleotide polymorphism in intron 18 of the human MST1R (RON) gene that maps at 3p21.3." *Mol Cell Probes.* 17(2-3): 55-7.

Aoki et al (1995). A novel gene, Translin, encodes a recombination hotspot binding protein associated with chromosomal translocations. *Nat Genet.* 10(2):167-74.

- Bai, R. Y., C. Koester, et al. (2002). "SMIF, a Smad4-interacting protein that functions as a co-activator in TGFbeta signalling." *Nat Cell Biol.* 4(3): 181-90.
- Barany, F. (1991). "Genetic disease detection and DNA amplification using cloned thermostable ligase." *Proc Natl Acad Sci U S A* 88(1): 189-93.
- Barre, F. X., S. Ait-Si-Ali, et al. (2000). "Unambiguous demonstration of triple-helix-directed gene modification." *Proc Natl Acad Sci U S A* 97(7): 3084-8.
- Becker, C., S. Wirtz, et al. (2003). "Constitutive p40 promoter activation and IL-23 production in the terminal ileum mediated by dendritic cells." *J Clin Invest* 112(5): 693-706.
- Bednarek, A. K., C. L. Keck-Waggoner, et al. (2001). "WWOX, the FRA16D gene, behaves as a suppressor of tumor growth." *Cancer Res* 61(22): 8068-73.
- Behl, C. (1997). "Amyloid beta-protein toxicity and oxidative stress in Alzheimer's disease." *Cell Tissue Res* 290(3): 471-80.
- Belkhiri, A., A. Zaika, et al. (2005). "Darpp-32: a novel antiapoptotic gene in upper gastrointestinal carcinomas." *Cancer Res.* 65(15): 6583-92.
- Bender, C. F., M. L. Sikes, et al. (2002). "Cancer predisposition and hematopoietic failure in Rad50(S/S) mice." *Genes Dev.* 16(17): 2237-51.
- Bespalova, I. N., G. Van Camp, et al. (2001). "Mutations in the Wolfram syndrome 1 gene (WFS1) are a common cause of low frequency sensorineural hearing loss." *Hum Mol Genet* 10(22): 2501-8.
- Bielenberg, D. R., Y. Hida, et al. (2004). "Semaphorin 3F, a chemorepellent for endothelial cells, induces a poorly vascularized, encapsulated, nonmetastatic tumor phenotype." *J Clin Invest.* 114(9): 1260-71.
- Bjursten, M., P. W. Bland, et al. (2005). "Long-term treatment with anti-alpha 4 integrin antibodies aggravates colitis in G alpha i2-deficient mice." *Eur J Immunol.* 35(8): 2274-83.
- Blanchard, C., S. Durual, et al. (2004). "IL-4 and IL-13 up-regulate intestinal trefoil factor expression: requirement for STAT6 and de novo protein synthesis." *J Immunol* 172(6): 3775-83.

- Blaser, S., J. Horn, et al. (2004). "The novel human platelet septin SEPT8 is an interaction partner of SEPT4." *Thromb Haemost.* 91(5): 959-66.
- Bloch, K. D., J. R. Wolfram, et al. (1995). "Three members of the nitric oxide synthase II gene family (NOS2A, NOS2B, and NOS2C) colocalize to human chromosome 17." *Genomics* 27(3): 526-30.
- Boengler, K., F. Pipp, et al. (2003). "The ankyrin repeat containing SOCS box protein 5: a novel protein associated with arteriogenesis." *Biochem Biophys Res Commun* 302(1): 17-22.
- Boren, J., A. Ramos-Montoya, et al. (2006). "In situ localization of transketolase activity in epithelial cells of different rat tissues and subcellularly in liver parenchymal cells." *J Histochem Cytochem.* 54(2): 191-9.
- Bouma, G. and W. Strober (2003). "The immunological and genetic basis of inflammatory bowel disease." *Nat Rev Immunol* 3(7): 521-33.
- Bourgeois, S. and D. Labuda (2004). "Dynamic allele-specific oligonucleotide hybridization on solid support." *Anal Biochem* 324(2): 309-11.
- Brightbill, H. D., D. H. Libraty, et al. (1999). "Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors." *Science* 285(5428): 732-6.
- Brodbeck, J., A. Davies, et al. (2002). "The ducky mutation in *Cacna2d2* results in altered Purkinje cell morphology and is associated with the expression of a truncated alpha 2 delta-2 protein with abnormal function." *J Biol Chem.* 277(10): 7684-93.
- Brown, J. L., L. Stowers, et al. (1996). "Human Ste20 homologue hPAK1 links GTPases to the JNK MAP kinase pathway." *Curr Biol* 6(5): 598-605.
- Browning, C. M., M. J. Smith, et al. (2001). "Human GFI-2 is a tat activation response element-independent Tat cofactor." *J Virol* 75(5): 2314-23.
- Bulavin, D. V., C. Phillips, et al. (2004). "Inactivation of the Wip1 phosphatase inhibits mammary tumorigenesis through p38 MAPK-mediated activation of the p16(Ink4a)-p19(Arf) pathway." *Nat Genet* 36(4): 343-50.

Cai, D., L. K. Clayton, et al. (1999). "AND-34, a novel p130Cas-binding thymic stromal cell protein regulated by adhesion and inflammatory cytokines." *J Immunol.* 163(4): 2104-12.

Cai, D., K. N. Felekis, et al. (2003). "The GDP exchange factor AND-34 is expressed in B cells, associates with HEF1, and activates Cdc42." *J Immunol.* 170(2): 969-78.

Cale, J. M., C. E. Shaw, et al. (1998). "Optimization of a reverse transcription-polymerase chain reaction (RT-PCR) mass assay for low-abundance mRNA." *Methods Mol Biol* 105: 351-71.

Chamouard, P., L. Grunebaum, et al. (1995). "Prothrombin fragment 1 + 2 and thrombin-antithrombin III complex as markers of activation of blood coagulation in inflammatory bowel diseases." *Eur J Gastroenterol Hepatol.* 7(12): 1183-8.

Chang, N. S., N. Pratt, et al. (2001). "Hyaluronidase induction of a WW domain-containing oxidoreductase that enhances tumor necrosis factor cytotoxicity." *J Biol Chem* 276(5): 3361-70.

Chavany, C., C. Vicario-Abejon, et al. (1998). "Transgenic mice for interleukin 3 develop motor neuron degeneration associated with autoimmune reaction against spinal cord motor neurons." *Proc Natl Acad Sci U S A* 95(19): 11354-9.

Chen, X., L. Levine, et al. (1999). "Fluorescence polarization in homogeneous nucleic acid analysis." *Genome Res* 9(5): 492-8.

Chen, S., X. Yin, et al. (2003). "The C-terminal kinase domain of the p34cdc2-related PITSLRE protein kinase (p110C) associates with p21-activated kinase 1 and inhibits its activity during anoikis." *J Biol Chem* 278(22): 20029-36. Epub 2003 Mar 6.

Chen, Q., L. Rabach, et al. (2005). "IL-11 receptor alpha in the pathogenesis of IL-13-induced inflammation and remodeling." *J Immunol* 174(4): 2305-13.

Chien, C. T., P. L. Bartel, et al. (1991). "The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest." *Proc Natl Acad Sci U S A* 88(21): 9578-82.

Choi, J., B. Nannenga, et al. (2002). "Mice deficient for the wild-type p53-induced phosphatase gene (Wip1) exhibit defects in reproductive organs, immune function, and cell cycle control." *Mol Cell Biol* 22(4): 1094-105.

Clavell, M., H. Correa-Gracian, et al. (2000). "Detection of interferon regulatory factor-1 in lamina propria mononuclear cells in Crohn's disease." *J Pediatr Gastroenterol Nutr*. 30(1): 43-7.

Cobrin, G. M. and M. T. Abreu (2005). "Defects in mucosal immunity leading to Crohn's disease." *Immunol Rev*. 206: 277-95.

Cohen, A. S., D. L. Smisek, et al. (1996). "Emerging technologies for sequencing antisense oligonucleotides: capillary electrophoresis and mass spectrometry." *Adv Chromatogr* 36: 127-62.

Cohn, R. D., M. D. Henry, et al. (2002). "Disruption of DAG1 in differentiated skeletal muscle reveals a role for dystroglycan in muscle regeneration." *Cell* 110(5): 639-48.

Combaret, L., O. A. Adegoke, et al. (2005). "USP19 is a ubiquitin-specific protease regulated in rat skeletal muscle during catabolic states." *Am J Physiol Endocrinol Metab*. 288(4): E693-700.

Costello et al (2005). Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Med*. Aug;2(8):e199.

Cotton, R. G., N. R. Rodrigues, et al. (1988). "Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations." *Proc Natl Acad Sci U S A* 85(12): 4397-401.

Couve, A., S. Restituto, et al. (2004). "Marlin-1, a novel RNA-binding protein associates with GABA receptors." *J Biol Chem* 279(14): 13934-43. Epub 2004 Jan 12.

Couzens, M., M. Liu, et al. (2000). "Peptide YY-2 (PYY2) and pancreatic polypeptide-2 (PPY2): species-specific evolution of novel members of the neuropeptide Y gene family." *Genomics* 64(3): 318-23.

Cronin, M. T., R. V. Fucini, et al. (1996). "Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays." *Hum Mutat* 7(3): 244-55.

Croucher, P. J., S. Mascheretti, *et al.* (2003). "Lack of association between the C3435T MDR1 gene polymorphism and inflammatory bowel disease in two independent Northern European populations." *Gastroenterology* 125(6): 1919-20; author reply 1920-1.

Cuppen, E., M. van Ham, *et al.* (2000). "The zyxin-related protein TRIP6 interacts with PDZ motifs in the adaptor protein RIL and the protein tyrosine phosphatase PTP-BL." *Eur J Cell Biol.* 79(4): 283-93.

Dalwadi, H., B. Wei, *et al.* (2003). "B cell developmental requirement for the G alpha i2 gene." *J Immunol.* 170(4): 1707-15.

Dean, F. B., S. Hosono, *et al.* (2002). "Comprehensive human genome amplification using multiple displacement amplification." *Proc Natl Acad Sci U S A* 99(8): 5261-6.

del Peso, L., R. Hernandez-Alcoceba, *et al.* (1997). "Rho proteins induce metastatic properties in vivo." *Oncogene* 15(25): 3047-57.

Dentelli, P., A. Rosso, *et al.* (2004). "IL-3 affects endothelial cell-mediated smooth muscle cell recruitment by increasing TGF beta activity: potential role in tumor vessel stabilization." *Oncogene* 23(9): 1681-92.

DeSalle, L. M., E. Latres, *et al.* (2001). "The de-ubiquitinating enzyme Unp interacts with the retinoblastoma protein." *Oncogene.* 20(39): 5538-42.

Deschenes, C., L. Alvarez, *et al.* (2004). "The nucleocytoplasmic shuttling of E2F4 is involved in the regulation of human intestinal epithelial cell proliferation and differentiation." *J Cell Physiol.* 199(2): 262-73.

Diefenbach, A., H. Schindler, *et al.* (1999). "Requirement for type 2 NO synthase for IL-12 signaling in innate immunity." *Science* 284(5416): 951-5.

Dijkstra, G., A. J. Zandvoort, *et al.* (2002). "Increased expression of inducible nitric oxide synthase in circulating monocytes from patients with active inflammatory bowel disease." *Scand J Gastroenterol.* 37(5): 546-54.

Dillon, N. (1993). "Regulating gene expression in gene therapy." *Trends Biotechnol* 11(5): 167-73.

- Ding, Z., L. L. Gillespie, et al. (2003). "Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain." *Mol Cell Biol*. 23(1): 250-8.
- Donato, J. L., J. Ko, et al. (2002). "Human HTm4 is a hematopoietic cell cycle regulator." *J Clin Invest*. 109(1): 51-8.
- Donovan, F. M., C. J. Pike, et al. (1997). "Thrombin induces apoptosis in cultured neurons and astrocytes via a pathway requiring tyrosine kinase and RhoA activities." *J Neurosci* 17(14): 5316-26.
- Dranoff, G., R. Soiffer, et al. (1997). "A phase I study of vaccination with autologous, irradiated melanoma cells engineered to secrete human granulocyte-macrophage colony stimulating factor." *Hum Gene Ther* 8(1): 111-23.
- Drmanac, R., S. Drmanac, et al. (2002). "Sequencing by hybridization (SBH): advantages, achievements, and opportunities." *Adv Biochem Eng Biotechnol* 77: 75-101.
- Dryja, T. P., L. B. Hahn, et al. (1996). "Missense mutation in the gene encoding the alpha subunit of rod transducin in the Nougaret form of congenital stationary night blindness." *Nat Genet*. 13(3): 358-60.
- Ducale, A. E., S. I. Ward, et al. (2005). "Regulation of hyaluronan synthase-2 expression in human intestinal mesenchymal cells: mechanisms of interleukin-1beta-mediated induction." *Am J Physiol Gastrointest Liver Physiol*. 289(3): G462-70.
- Duerr, R. H. (2002). "The genetics of inflammatory bowel disease." *Gastroenterol Clin North Am* 31(1): 63-76.
- Dunbar, C. E., M. Cottler-Fox, et al. (1995). "Retrovirally marked CD34-enriched peripheral blood and bone marrow cells contribute to long-term engraftment after autologous transplantation." *Blood* 85(11): 3048-57.
- Duseti, N. J., Y. Jiang, et al. (2002). "Cloning and expression of the rat vacuole membrane protein 1 (VMP1), a new gene activated in pancreas with acute pancreatitis, which promotes vacuole formation." *Biochem Biophys Res Commun* 290(2): 641-9.
- Elez, R., A. Piiper, et al. (2000). "Polo-like kinase1, a new target for antisense tumor therapy." *Biochem Biophys Res Commun* 269(2): 352-6.

Ellem, K. A., M. G. O'Rourke, et al. (1997). "A case report: immune responses and clinical course of the first human use of granulocyte/macrophage-colony-stimulating-factor-transduced autologous melanoma cells for immunotherapy." *Cancer Immunol Immunother* 44(1): 10-20.

Esaki, M., M. Furuse, et al. (1999). "Polymorphism of heat-shock protein gene HSP70-2 in Crohn disease: possible genetic marker for two forms of Crohn disease." *Scand J Gastroenterol* 34(7): 703-7.

Esworthy R.S., R. Aranda, et. (2001). "Mice with combined disruption of Gpx1 and Gpx2 genes have colitis." *Am J Physiol Gastrointest Liver Physiol*. 281(3):G848-55. PMID: 11518697.

Evans, B. E., K. E. Rittle, *et al.* (1987). "Design of nonpeptidal ligands for a peptide receptor: cholecystikinin antagonists." *J Med Chem* 30(7): 1229-39.

Fackler, O. T. and B. M. Peterlin (2000). "Endocytic entry of HIV-1." *Curr Biol* 10(16): 1005-8.

Fakhrai-Rad, H., J. Zheng, *et al.* (2004). "SNP discovery in pooled samples with mismatch repair detection." *Genome Res* 14(7): 1404-12.

Fan, J. and A. B. Malik (2003). "Toll-like receptor-4 (TLR4) signaling augments chemokine-induced neutrophil migration by modulating cell surface expression of chemokine receptors." *Nat Med* 9(3): 315-21.

Farmer, R. G., G. Whelan, et al. (1985). "Long-term follow-up of patients with Crohn's disease. Relationship between the clinical pattern and prognosis." *Gastroenterology* 88(6): 1818-25.

Fechir, M., K. Linker, et al. (2005). "The RNA binding protein TIAR is involved in the regulation of human iNOS expression." *Cell Mol Biol (Noisy-le-grand)*. 51(3): 299-305.

Fidder, H. H., S. Olschwang, et al. (2003). "Association between mutations in the CARD15 (NOD2) gene and Crohn's disease in Israeli Jewish patients." *Am J Med Genet* 121A(3): 240-4.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." *Nature* 340(6230): 245-6.

- Frank-Kamenetskii, M. D. and S. M. Mirkin (1995). "Triplex DNA structures." *Annu Rev Biochem* 64: 65-95.
- Freeman, W. M., S. J. Walker, et al. (1999). "Quantitative RT-PCR: pitfalls and potential." *Biotechniques* 26(1): 112-22, 124-5.
- Frolova, E. I., G. M. Dolganov, et al. (1991). "Linkage mapping of the human CSF2 and IL3 genes." *Proc Natl Acad Sci U S A* 88(11): 4821-4.
- Fujita, Y., Y. Ezura, et al. (2004). "Hypercholesterolemia associated with splice-junction variation of inter-alpha-trypsin inhibitor heavy chain 4 (ITIH4) gene." *J Hum Genet* 49(1): 24-8.
- Fukushige, S., K. Matsubara, et al. (1986). "Localization of a novel v-erbB-related gene, c-erbB-2, on human chromosome 17 and its amplification in a gastric cancer cell line." *Mol Cell Biol* 6(3): 955-8.
- Fukushima, T., J. M. Zapata, et al. (2006). "Critical function for SIP, a ubiquitin E3 ligase component of the beta-catenin degradation pathway, for thymocyte development and G1 checkpoint." *Immunity*. 24(1): 29-39.
- Gainetdinov, R. R., L. M. Bohn, et al. (1999). "Muscarinic supersensitivity and impaired receptor desensitization in G protein-coupled receptor kinase 5-deficient mice." *Neuron*. 24(4): 1029-36.
- Gan, B., Z. K. Melkounian, et al. (2005). "Identification of FIP200 interaction with the TSC1-TSC2 complex and its role in regulation of cell size control." *J Cell Biol*. 170(3): 379-89.
- Gary-Gouy, H., J. Harriague, et al. (2002). "CD5-negative regulation of B cell receptor signaling pathways originates from tyrosine residue Y429 outside an immunoreceptor tyrosine-based inhibitory motif." *J Immunol*. 168(1): 232-9.
- Gasparini, P., A. Bonizzato, et al. (1992). "Restriction site generating-polymerase chain reaction (RG-PCR) for the probeless detection of hidden genetic variation: application to the study of some common cystic fibrosis mutations." *Mol Cell Probes* 6(1): 1-7.

Geyer, M., H. Yu, et al. (2002). "Subunit H of the V-ATPase binds to the medium chain of adaptor protein complex 2 and connects Nef to the endocytic machinery." *J Biol Chem* 277(32): 28521-9. Epub 2002 May 24.

Giachino, D., M. M. van Duist, et al. (2004). "Analysis of the CARD15 variants R702W, G908R and L1007fs in Italian IBD patients." *Eur J Hum Genet* 12(3): 206-12.

Gibbs, R. A., P. N. Nguyen, et al. (1989). "Detection of single DNA base differences by competitive oligonucleotide priming." *Nucleic Acids Res* 17(7): 2437-48.

Girardin, S. E., I. G. Boneca, et al. (2003). "Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection." *J Biol Chem* 278(11): 8869-72.

Girardin, S. E., J. P. Hugot, et al. (2003). "Lessons from Nod2 studies: towards a link between Crohn's disease and bacterial sensing." *Trends Immunol* 24(12): 652-8.

Griffin, H. G. and A. M. Griffin (1993). "DNA sequencing. Recent innovations and future trends." *Appl Biochem Biotechnol* 38(1-2): 147-59.

Grossman, P. D., W. Bloch, et al. (1994). "High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation." *Nucleic Acids Res* 22(21): 4527-34.

Guatelli, J. C., K. M. Whitfield, et al. (1990). "Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication." *Proc Natl Acad Sci U S A* 87(5): 1874-8.

Guatelli, J. C., K. M. Whitfield, et al. (1990). "Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication." *Proc Natl Acad Sci U S A* 87(19): 7797.

Hafner, G. J., I. C. Yang, et al. (2001). "Isothermal amplification and multimerization of DNA by Bst DNA polymerase." *Biotechniques* 30(4): 852-6, 858, 860 passim.

Hampe, J., J. Grebe, et al. (2002). "Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: a cohort study." *Lancet* 359(9318): 1661-5.

Harada, M., Y. Qin, et al. (2005). "G-CSF prevents cardiac remodeling after myocardial infarction by activating the Jak-Stat pathway in cardiomyocytes." *Nat Med*. 11(3): 305-11.

Hardenbol, P., J. Baner, et al. (2003). "Multiplexed genotyping with sequence-tagged molecular inversion probes." *Nat Biotechnol* 21(6): 673-8. Epub 2003 May 5.

Hardwick, J. C., G. R. Van Den Brink, et al. (2004). "Bone morphogenetic protein 2 is expressed by, and acts upon, mature epithelial cells in the colon." *Gastroenterology*. 126(1): 111-21.

Hartman, M. E., J. C. O'Connor, et al. (2004). "Insulin receptor substrate-2-dependent interleukin-4 signaling in macrophages is impaired in two models of type 2 diabetes mellitus." *J Biol Chem* 279(27): 28045-50.

Hasegawa, T., A. Yagi, et al. (2000). "Interaction between GADD34 and kinesin superfamily, KIF3A." *Biochem Biophys Res Commun*. 267(2): 593-6.

Hayashi, K. (1992). "PCR-SSCP: a method for detection of mutations." *Genet Anal Tech Appl* 9(3): 73-9.

Heintz, K., K. Palme, et al. (1992). "The Ncyp1 gene from *Neurospora crassa* is located on chromosome 2: molecular cloning and structural analysis." *Mol Gen Genet* 235(2-3): 413-21.

Hellevuo, K., M. Yoshimura, et al. (1993). "A novel adenylyl cyclase sequence cloned from the human erythroleukemia cell line." *Biochem Biophys Res Commun*. 192(1): 311-8.

Helms, C., L. Cao, et al. (2003). "A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis." *Nat Genet* 35(4): 349-56.

Henry, M. D. and K. P. Campbell (1998). "A role for dystroglycan in basement membrane assembly." *Cell*. 95(6): 859-70.

Hermonat, P. L. and N. Muzyczka (1984). "Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells." *Proc Natl Acad Sci U S A* 81(20): 6466-70.

Hirai, H., K. Tanaka, et al. (2002). "Cutting edge: agonistic effect of indomethacin on a prostaglandin D2 receptor, CRTH2." *J Immunol*. 168(3): 981-5.

- Hisamatsu, T., M. Suzuki, et al. (2003). "Interferon-gamma augments CARD4/NOD1 gene and protein expression through interferon regulatory factor-1 in intestinal epithelial cells." *J Biol Chem.* 278(35): 32962-8.
- Hobbs, M. R., V. Udhayakumar, et al. (2002). "A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children." *Lancet* 360(9344): 1468-75.
- Hoffman, J., F. Kuhnert, et al. (2004). "Wnts as essential growth factors for the adult small intestine and colon." *Cell Cycle.* 3(5): 554-7.
- Hogrefe, H. H., R. L. Mullinax, et al. (1993). "A bacteriophage lambda vector for the cloning and expression of immunoglobulin Fab fragments on the surface of filamentous phage." *Gene* 128(1): 119-26.
- Hoogendoorn, B., N. Norton, et al. (2000). "Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools." *Hum Genet* 107(5): 488-93.
- Hsu, I. C., Q. Yang, et al. (1994). "Detection of DNA point mutations with DNA mismatch repair enzymes." *Carcinogenesis* 15(8): 1657-62.
- Hu, W., F. Yin, et al. (2002). "[Effect of human calcyclin binding protein encoding gene on development of multiple drug resistance in gastric cancer]." *Zhonghua Zhong Liu Za Zhi.* 24(5): 426-9.
- Hugot, J. P., M. Chamaillard, et al. (2001). "Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease." *Nature* 411(6837): 599-603.
- Hugot, J. P. and G. Thomas (1998). "Genome-wide scanning in inflammatory bowel diseases." *Dig Dis* 16(6): 364-9.
- Hunt, T. W., T. A. Fields, et al. (1996). "RGS10 is a selective activator of G alpha i GTPase activity." *Nature.* 383(6596): 175-7.
- Hutchings, N. J., N. Clarkson, et al. (2003). "Linking the T cell surface protein CD2 to the actin-capping protein CAPZ via CMS and CIN85." *J Biol Chem.* 278(25): 22396-403.
- Ince-Dunn, G., B. J. Hall, et al. (2006). "Regulation of thalamocortical patterning and synaptic maturation by NeuroD2." *Neuron.* 49(5): 683-95.

Inostroza, J. A., F. H. Mermelstein, et al. (1992). "Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription." *Cell*. 70(3): 477-89.

Irvine, E. J. (2004). "Patients' fears and unmet needs in inflammatory bowel disease." *Aliment Pharmacol Ther* 20 Suppl 4: 54-9.

Ishibashi, K., M. Suzuki, et al. (2001). "Identification of a new multigene four-transmembrane family (MS4A) related to CD20, HTm4 and beta subunit of the high-affinity IgE receptor." *Gene* 264(1): 87-93.

Ishigaki, S., J. Niwa, et al. (2000). "Two novel genes, human neugrin and mouse m-neugrin, are upregulated with neuronal differentiation in neuroblastoma cells." *Biochem Biophys Res Commun* 279(2): 526-33.

Ishii, M., H. Fei, et al. (2003). "Targeted disruption of GPR7, the endogenous receptor for neuropeptides B and W, leads to metabolic defects and adult-onset obesity." *Proc Natl Acad Sci U S A* 100(18): 10540-5. Epub 2003 Aug 18.

Ishitani, T., S. Kishida, et al. (2003). "The TAK1-NLK mitogen-activated protein kinase cascade functions in the Wnt-5a/Ca(2+) pathway to antagonize Wnt/beta-catenin signaling." *Mol Cell Biol* 23(1): 131-9.

Itoh et al (2001). "Decreased Bax expression by mucosal T cells favours resistance to apoptosis in Crohn's disease." *Gut*. 49(1):35-41.

Ivanov, A. I., A. Nusrat, et al. (2004). "The epithelium in inflammatory bowel disease: potential role of endocytosis of junctional proteins in barrier disruption." *Novartis Found Symp* 263: 115-24.

Jan, Y., M. Matter, et al. (2004). "A mitochondrial protein, Bit1, mediates apoptosis regulated by integrins and Groucho/TLE corepressors." *Cell* 116(5): 751-62.

Jandrig et al (2004). ST18 is a breast cancer tumor suppressor gene at human chromosome 8q11.2. *Oncogene*. 23(57):9295-302.

Jiang, P. H., Y. Motoo, et al. (2004). "Expression of vacuole membrane protein 1 (VMP1) in spontaneous chronic pancreatitis in the WBN/Kob rat." *Pancreas* 29(3): 225-30.

Jobs, M., W. M. Howell, et al. (2003). "DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays." *Genome Res* 13(5): 916-24.

Johann, S. V., J. J. Gibbons, et al. (1992). "GLVR1, a receptor for gibbon ape leukemia virus, is homologous to a phosphate permease of *Neurospora crassa* and is expressed at high levels in the brain and thymus." *J Virol* 66(3): 1635-40.

Johannesen, J., A. Pie, et al. (2001). "Linkage of the human inducible nitric oxide synthase gene to type 1 diabetes." *J Clin Endocrinol Metab* 86(6): 2792-6.

Johnson, E. N. and K. M. Druey (2002). "Heterotrimeric G protein signaling: role in asthma and allergic inflammation." *J Allergy Clin Immunol* 109(4): 592-602.

Kaarniranta, K., T. Ryhanen, et al. (2005). "Geldanamycin activates Hsp70 response and attenuates okadaic acid-induced cytotoxicity in human retinal pigment epithelial cells." *Brain Res Mol Brain Res* 137(1-2): 126-31.

Kajita, M., Y. Ezura, et al. (2003). "Association of the -381T/C promoter variation of the brain natriuretic peptide gene with low bone-mineral density and rapid postmenopausal bone loss." *J Hum Genet* 48(2): 77-81.

Kaltschmidt, C., B. Kaltschmidt, et al. (1994). "Constitutive NF-kappa B activity in neurons." *Mol Cell Biol* 14(6): 3981-92.

Kanazawa, N., M. Kurosaki, et al. (2004). "Regulation of hepatitis C virus replication by interferon regulatory factor 1." *J Virol* 78(18): 9713-20.

Kanehisa, M. (1984). "Use of statistical criteria for screening potential homologies in nucleic acid sequences." *Nucleic Acids Res* 12(1 Pt 1): 203-13.

Kanei-Ishii, C., J. Ninomiya-Tsuji, et al. (2004). "Wnt-1 signal induces phosphorylation and degradation of c-Myb protein via TAK1, HIPK2, and NLK." *Genes Dev* 18(7): 816-29.

Karason, A., J. E. Gudjonsson, et al. (2003). "A susceptibility gene for psoriatic arthritis maps to chromosome 16q: evidence for imprinting." *Am J Hum Genet* 72(1): 125-31.

Kato, A., Y. Nagata, et al. (2004). "Delta-tubulin is a component of intercellular bridges and both the early and mature perinuclear rings during spermatogenesis." *Dev Biol* 269(1): 196-205.

Kashio, N., W. Matsumoto, et al. (1998). "The second domain of the CD45 protein tyrosine phosphatase is critical for interleukin-2 secretion and substrate recruitment of TCR-zeta in vivo." *J Biol Chem.* 273(50): 33856-63.

Kashio et al (2003). Galectin-9 induces apoptosis through the calcium-calpain-caspase-1 pathway. *J Immunol.*;170(7):3631-6.

Kayed et al (2004). Expression analysis of MAC30 in human pancreatic cancer and tumors of the gastrointestinal tract. *Histol Histopathol.* 19(4):1021-31.

Kim, S., J. Lee, et al. (2003). "BP75, bromodomain-containing M(r) 75,000 protein, binds dishevelled-1 and enhances Wnt signaling by inactivating glycogen synthase kinase-3 beta." *Cancer Res.* 63(16): 4792-5.

Kimmel, S. G., R. Mo, et al. (2000). "New mouse models of congenital anorectal malformations." *J Pediatr Surg.* 35(2): 227-30; discussion 230-1.

Klausz, G., T. Molnar, et al. (2005). "Polymorphism of the heat-shock protein gene Hsp70-2, but not polymorphisms of the IL-10 and CD14 genes, is associated with the outcome of Crohn's disease." *Scand J Gastroenterol.* 40(10): 1197-204.

Klein, B. K., J. J. Shieh, et al. (2001). "Receptor binding kinetics of human IL-3 variants with altered proliferative activity." *Biochem Biophys Res Commun* 288(5): 1244-9.

Klein, W., A. Tromm, et al. (2001). "Interleukin-4 and interleukin-4 receptor gene polymorphisms in inflammatory bowel diseases." *Genes Immun* 2(5): 287-9.

Kobayashi, K. S., M. Chamaillard, et al. (2005). "Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract." *Science* 307(5710): 731-4.

Koh, K. P., Y. Wang, et al. (2004). "T cell-mediated vascular dysfunction of human allografts results from IFN-gamma dysregulation of NO synthase." *J Clin Invest* 114(6): 846-56.

Kohler, G. and C. Milstein (1992). "Continuous cultures of fused cells secreting antibody of predefined specificity. 1975." *Biotechnology* 24: 524-6.

Kohn, D. B., K. I. Weinberg, et al. (1995). "Engraftment of gene-modified umbilical cord blood cells in neonates with adenosine deaminase deficiency." *Nat Med* 1(10): 1017-23.

- Kojima et al (2005). STAT3 regulates Nemo-like kinase by mediating its interaction with IL-6-stimulated TGFbeta-activated kinase 1 for STAT3 Ser-727 phosphorylation. *Proc Natl Acad Sci U S A*. 102(12):4524-9.
- Kolesnick, R. and H. R. Xing (2004). "Inflammatory bowel disease reveals the kinase activity of KSR1." *J Clin Invest* 114(9): 1233-7.
- Kolodziejwski, P. J., A. Musial, et al. (2002). "Ubiquitination of inducible nitric oxide synthase is required for its degradation." *Proc Natl Acad Sci U S A* 99(19): 12315-20.
- Kong et al (2005). ELL-associated factors 1 and 2 are positive regulators of RNA polymerase II elongation factor ELL. *Proc Natl Acad Sci U S A*. 102(29):10094-8.
- Kontani, K., T. Chano, et al. (2003). "RB1CC1 suppresses cell cycle progression through RB1 expression in human neoplastic cells." *Int J Mol Med*. 12(5): 767-9.
- Kopp, E. and S. Ghosh (1994). "Inhibition of NF-kappa B by sodium salicylate and aspirin." *Science* 265(5174): 956-9.
- Kotin, R. M. (1994). "Prospects for the use of adeno-associated virus as a vector for human gene therapy." *Hum Gene Ther* 5(7): 793-801.
- Kovalenko et al (2003). "The tumour suppressor CYLD negatively regulates NF-kappaB signalling by deubiquitination." *Nature*. 424(6950):801-5.
- Kozal, M. J., N. Shah, et al. (1996). "Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays." *Nat Med* 2(7): 753-9.
- Kray, A. E., R. S. Carter, et al. (2005). "Positive regulation of IkappaB kinase signaling by protein serine/threonine phosphatase 2A." *J Biol Chem*. 280(43): 35974-82.
- Krebs, S., D. Seichter, et al. (2001). "Genotyping of dinucleotide tandem repeats by MALDI mass spectrometry of ribozyme-cleaved RNA transcripts." *Nat Biotechnol* 19(9): 877-80.
- Kremer, E. J. and M. Perricaudet (1995). "Adenovirus and adeno-associated virus mediated gene transfer." *Br Med Bull* 51(1): 31-44.
- Kulinski, J., D. Besack, et al. (2000). "CEL I enzymatic mutation detection assay." *Biotechniques* 29(1): 44-6, 48.

- Kunapuli, P. and J. L. Benovic (1993). "Cloning and expression of GRK5: a member of the G protein-coupled receptor kinase family." *Proc Natl Acad Sci U S A* 90(12): 5588-92.
- Kure et al (1998). "A missense mutation (His42Arg) in the T-protein gene from a large Israeli-Arab kindred with nonketotic hyperglycinemia." *Hum Genet.* 102(4):430-4.
- Kure, S., T. Shinka, et al. (1998). "A one-base deletion (183delC) and a missense mutation (D276H) in the T-protein gene from a Japanese family with nonketotic hyperglycinemia." *J Hum Genet* 43(2): 135-7.
- Kuroki, T., F. Trapasso, et al. (2002). "Genetic alterations of the tumor suppressor gene WWOX in esophageal squamous cell carcinoma." *Cancer Res* 62(8): 2258-60.
- Kusugami, K., A. Fukatsu, et al. (1995). "Elevation of interleukin-6 in inflammatory bowel disease is macrophage- and epithelial cell-dependent." *Dig Dis Sci.* 40(5): 949-59.
- Kwoh, D. Y., G. R. Davis, et al. (1989). "Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format." *Proc Natl Acad Sci U S A* 86(4): 1173-7.
- Langmann et al. (2004). "Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes." *Gastroenterology.* 127(1):26-40.
- Larkin, D., D. Murphy, et al. (2004). "ICln, a novel integrin alphaIIb beta3-associated protein, functionally regulates platelet activation." *J Biol Chem* 279(26): 27286-93.
- Lesage, S., H. Zouali, et al. (2002). "CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease." *Am J Hum Genet* 70(4): 845-57.
- Leveille, C., A. L.-D. R, et al. (1999). "CD20 is physically and functionally coupled to MHC class II and CD40 on human B cell lines." *Eur J Immunol.* 29(1): 65-74.
- Li, J., Y. Yang, et al. (2002). "Oncogenic properties of PPM1D located within a breast cancer amplification epicenter at 17q23." *Nat Genet* 31(2): 133-4.
- Li, Z., X. Dong, et al. (2005). "Regulation of PTEN by Rho small GTPases." *Nat Cell Biol* 7(4): 399-404. Epub 2005 Mar 27.

- Ligumsky et al (1997). Analysis of cytokine profile in human colonic mucosal Fc epsilonRI-positive cells by single cell PCR: inhibition of IL-3 expression in steroid-treated IBD patients. *FEBS Lett.* 413(3):436-40.
- Lin, Y., A. Khokhlatchev, et al. (2002). "Death-associated protein 4 binds MST1 and augments MST1-induced apoptosis." *J Biol Chem* 277(50): 47991-8001.
- Lin, C. H., S. Hansen, et al. (2005). "The dosage of the neuroD2 transcription factor regulates amygdala development and emotional learning." *Proc Natl Acad Sci U S A.* 102(41): 14877-82.
- Linden, D. R., K. F. Foley, et al. (2005). "Serotonin transporter function and expression are reduced in mice with TNBS-induced colitis." *Neurogastroenterol Motil.* 17(4): 565-74.
- Linstedt et al (1993). Giantin, a novel conserved Golgi membrane protein containing a cytoplasmic domain of at least 350 kDa. *Mol Biol Cell.* 4(7):679-93.
- Liu, J., X. Guan, et al. (2004). "Synergistic activation of interleukin-12 p35 gene transcription by interferon regulatory factor-1 and interferon consensus sequence-binding protein." *J Biol Chem* 279(53): 55609-17.
- Liu, C., C. Kuei, et al. (2005). "INSL5 is a high affinity specific agonist for GPCR142 (GPR100)." *J Biol Chem.* 280(1): 292-300.
- Livak, K. J., J. Marmaro, et al. (1995). "Towards fully automated genome-wide polymorphism screening." *Nat Genet* 9(4): 341-2.
- Lizardi, P. M., X. Huang, et al. (1998). "Mutation detection and single-molecule counting using isothermal rolling-circle amplification." *Nat Genet* 19(3): 225-32.
- Loftus, E. V., Jr. (2004). "Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences." *Gastroenterology* 126(6): 1504-17.
- Lohi, H., S. Makela, et al. (2002). "Upregulation of CFTR expression but not SLC26A3 and SLC9A3 in ulcerative colitis." *Am J Physiol Gastrointest Liver Physiol.* 283(3): G567-75.

Lombardi, M. S., A. Kavelaars, et al. (1999). "Decreased expression and activity of G-protein-coupled receptor kinases in peripheral blood mononuclear cells of patients with rheumatoid arthritis." *Faseb J.* 13(6): 715-25.

Lowenstein, C. J., C. S. Glatt, et al. (1992). "Cloned and expressed macrophage nitric oxide synthase contrasts with the brain enzyme." *Proc Natl Acad Sci U S A* 89(15): 6711-5.

Machida et al (2005). Association of polymorphic alleles of CTLA4 with inflammatory bowel disease in the Japanese. *World J Gastroenterol.* 11(27):4188-93.

Mady et al (2002). Expression of E2F-4 gene in colorectal adenocarcinoma and corresponding covering mucosa: an immunohistochemistry, image analysis, and immunoblot study. *Appl Immunohistochem Mol Morphol.* 10(3):225-30.

Magro, F., M. A. Vieira-Coelho, et al. (2002). "Impaired synthesis or cellular storage of norepinephrine, dopamine, and 5-hydroxytryptamine in human inflammatory bowel disease." *Dig Dis Sci.* 47(1): 216-24.

Malech, H. L., P. B. Maples, et al. (1997). "Prolonged production of NADPH oxidase-corrected granulocytes after gene therapy of chronic granulomatous disease." *Proc Natl Acad Sci U S A* 94(22): 12133-8.

Malo, M. S., W. Zhang, et al. (2004). "Thyroid hormone positively regulates the enterocyte differentiation marker intestinal alkaline phosphatase gene via an atypical response element." *Mol Endocrinol.* 18(8): 1941-62.

Mannon, P. J., I. J. Fuss, et al. (2004). "Anti-interleukin-12 antibody for active Crohn's disease." *N Engl J Med* 351(20): 2069-79.

Mansfield, E. S., M. Vainer, et al. (1996). "Sensitivity, reproducibility, and accuracy in short tandem repeat genotyping using capillary array electrophoresis." *Genome Res* 6(9): 893-903.

Matsumoto et al (1998). Human ecalectin, a variant of human galectin-9, is a novel eosinophil chemoattractant produced by T lymphocytes. *J Biol Chem.* 273(27):16976-84.

Maxam, A. M. and W. Gilbert (1977). "A new method for sequencing DNA." *Proc Natl Acad Sci U S A* 74(2): 560-4.

- Mazzocchi, G., P. Rebuffat, et al. (2005). "G Protein Receptors (GPR) 7 and 8 are Expressed in Human Adrenocortical Cells, and their Endogenous Ligands Neuropeptides B and W Enhance Cortisol Secretion by Activating Adenylate Cyclase- and Phospholipase C-dependent Signaling Cascades." *J Clin Endocrinol Metab* 29: 29.
- Mazzocco, M., M. Maffei, et al. (2002). "The identification of a novel human homologue of the SH3 binding glutamic acid-rich (SH3BGR) gene establishes a new family of highly conserved small proteins related to Thioredoxin Superfamily." *Gene* 291(1-2): 233-9.
- Mecklenbrauker, I., S. L. Kalled, et al. (2004). "Regulation of B-cell survival by BAFF-dependent PKCdelta-mediated nuclear signalling." *Nature* 431(7007): 456-61.
- Meili, R., A. T. Sasaki, et al. (2005). "Rho Rocks PTEN." *Nat Cell Biol* 7(4): 334-335.
- Melkounian, Z. K., X. Peng, et al. (2005). "Mechanism of cell cycle regulation by FIP200 in human breast cancer cells." *Cancer Res.* 65(15): 6676-84.
- Metcalf, D. (1985). "The granulocyte-macrophage colony stimulating factors." *Cell* 43(1): 5-6.
- Miller, A. D. (1992). "Human gene therapy comes of age." *Nature* 357(6378): 455-60.
- Miller, A. D., J. V. Garcia, et al. (1991). "Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus." *J Virol* 65(5): 2220-4.
- Milojevic, T., V. Reiterer, et al. (2006). "The ubiquitin-specific protease Usp4 regulates the cell surface level of the A2A receptor." *Mol Pharmacol.* 69(4): 1083-94.
- Mischak, H., A. Bodenteich, et al. (1991). "Mouse protein kinase C-delta, the major isoform expressed in mouse hemopoietic cells: sequence of the cDNA, expression patterns, and characterization of the protein." *Biochemistry* 30(32): 7925-31.
- Mitani, K. and C. T. Caskey (1993). "Delivering therapeutic genes--matching approach and application." *Trends Biotechnol* 11(5): 162-6.
- Miyamoto, A., K. Nakayama, et al. (2002). "Increased proliferation of B cells and autoimmunity in mice lacking protein kinase Cdelta." *Nature* 416(6883): 865-9.
- Monteleone, G., A. Kumberova, et al. (2001). "Blocking Smad7 restores TGF-beta1 signaling in chronic inflammatory bowel disease." *J Clin Invest.* 108(4): 601-9.

Monticelli, S., D. C. Solymar, et al. (2004). "Role of NFAT proteins in IL13 gene transcription in mast cells." *J Biol Chem* 279(35): 36210-8.

Moreira, E. S., T. J. Wiltshire, et al. (2000). "Limb-girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin." *Nat Genet.* 24(2): 163-6.

Mosialos, G. (1997). "The role of Rel/NF-kappa B proteins in viral oncogenesis and the regulation of viral transcription." *Semin Cancer Biol* 8(2): 121-9.

Muzyczka, N. (1994). "Adeno-associated virus (AAV) vectors: will they work?" *J Clin Invest* 94(4): 1351.

Myers, R. M., Z. Larin, et al. (1985). "Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes." *Science* 230(4731): 1242-6.

Myers, R. M., N. Lumelsky, et al. (1985). "Detection of single base substitutions in total genomic DNA." *Nature* 313(6002): 495-8.

Mykkanen, O. M., M. Gronholm, et al. (2001). "Characterization of human palladin, a microfilament-associated protein." *Mol Biol Cell* 12(10): 3060-73.

Nagashima et al (2003). "A novel PHD-finger motif protein, p47ING3, modulates p53-mediated transcription, cell cycle control, and apoptosis." *Oncogene.* 22(3):343-50.

Nabel, G. J. and P. L. Felgner (1993). "Direct gene transfer for immunotherapy and immunization." *Trends Biotechnol* 11(5): 211-5.

Nagata, K., A. Puls, et al. (1998). "The MAP kinase kinase kinase MLK2 co-localizes with activated JNK along microtubules and associates with kinesin superfamily motor KIF3." *Embo J.* 17(1): 149-58.

Nagley, P. and Y. H. Wei (1998). "Ageing and mammalian mitochondrial genetics." *Trends Genet* 14(12): 513-7.

Nanao et al (1994). "Identification of the mutations in the T-protein gene causing typical and atypical nonketotic hyperglycinemia." *Hum Genet.* 93(6):655-8.

Nanao, K., G. Takada, et al. (1994). "Structure and chromosomal localization of the aminomethyltransferase gene (AMT)." *Genomics*. 19(1): 27-30.

Napolitano, M., F. Miceli, et al. (2000). "Expression and relationship between endothelin-1 messenger ribonucleic acid (mRNA) and inducible/endothelial nitric oxide synthase mRNA isoforms from normal and preeclamptic placentas." *J Clin Endocrinol Metab* 85(6): 2318-23.

Nathan, C. and M. U. Shiloh (2000). "Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens." *Proc Natl Acad Sci U S A* 97(16): 8841-8.

Neil, G. A., R. W. Summers, et al. (1992). "CD5+ B cells are decreased in peripheral blood of patients with Crohn's disease." *Dig Dis Sci*. 37(9): 1390-5.

Netzer, C., L. Rieger, et al. (2001). "SALL1, the gene mutated in Townes-Brocks syndrome, encodes a transcriptional repressor which interacts with TRF1/PIN2 and localizes to pericentromeric heterochromatin." *Hum Mol Genet*. 10(26): 3017-24.

Neurath, M. F., B. Weigmann, et al. (2002). "The transcription factor T-bet regulates mucosal T cell activation in experimental colitis and Crohn's disease." *J Exp Med*. 195(9): 1129-43.

Newman, B., X. Gu, et al. (2005). "A risk haplotype in the Solute Carrier Family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease." *Gastroenterology* 128(2): 260-9.

Ngsee, J. K., L. A. Elferink, et al. (1991). "A family of ras-like GTP-binding proteins expressed in electromotor neurons." *J Biol Chem* 266(4): 2675-80.

Niu, T., Z. S. Qin, et al. (2002). "Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms." *Am J Hum Genet* 70(1): 157-69.

Odashima, M., G. Bamias, et al. (2005). "Activation of A2A adenosine receptor attenuates intestinal inflammation in animal models of inflammatory bowel disease." *Gastroenterology*. 129(1): 26-33.

Ogura, Y., D. K. Bonen, et al. (2001). "A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease." *Nature* 411(6837): 603-6.

Ogura, Y., N. Inohara, et al. (2001). "Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB." *J Biol Chem* 276(7): 4812-8. Epub 2000 Nov 21.

Oppmann, B., R. Lesley, et al. (2000). "Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12." *Immunity* 13(5): 715-25.

Orita, M., H. Iwahana, et al. (1989). "Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms." *Proc Natl Acad Sci U S A* 86(8): 2766-70.

Ouburg, S., R. Mallant-Hent, et al. (2005). "The toll-like receptor 4 (TLR4) Asp299Gly polymorphism is associated with colonic localisation of Crohn's disease without a major role for the *Saccharomyces cerevisiae* mannan-LBP-CD14-TLR4 pathway." *Gut* 54(3): 439-40.

Pahl, H. L., B. Krauss, et al. (1996). "The immunosuppressive fungal metabolite gliotoxin specifically inhibits transcription factor NF-kappaB." *J Exp Med* 183(4): 1829-40.

Parham, C., M. Chirica, et al. (2002). "A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rbeta1 and a novel cytokine receptor subunit, IL-23R." *J Immunol* 168(11): 5699-708.

Paris, S., R. Sesboue, et al. (2002). "Inhibition of tumor growth and metastatic spreading by overexpression of inter-alpha-trypsin inhibitor family chains." *Int J Cancer*. 97(5): 615-20.

Pelteková, V. D., R. F. Wintle, et al. (2004). "Functional variants of OCTN cation transporter genes are associated with Crohn disease." *Nat Genet* 36(5): 471-5. Epub 2004 Apr 11.

Pfeffer, S. R. (2005). "Structural clues to Rab GTPase functional diversity." *J Biol Chem* 3: 3.

Pirone, D. M., S. Fukuhara, et al. (2000). "SPECs, small binding proteins for Cdc42." *J Biol Chem* 275(30): 22650-6.

Podolsky, D. K. (1991). "Inflammatory bowel disease (1)." *N Engl J Med* 325(13): 928-37.

- Podolsky, D. K. (1991). "Inflammatory bowel disease (2)." *N Engl J Med* 325(14): 1008-16.
- Pol, O., J. R. Palacio, et al. (2003). "The expression of delta- and kappa-opioid receptor is enhanced during intestinal inflammation in mice." *J Pharmacol Exp Ther*. 306(2): 455-62.
- Praskova, M., A. Khoklatchev, et al. (2004). "Regulation of the MST1 kinase by autophosphorylation, by the growth inhibitory proteins, RASSF1 and NORE1, and by Ras." *Biochem J* 381(Pt 2): 453-62.
- Qiu, Y., L. Ravi, et al. (1998). "Requirement of ErbB2 for signalling by interleukin-6 in prostate carcinoma cells." *Nature* 393(6680): 83-5.
- Raap, A. K. (1998). "Advances in fluorescence in situ hybridization." *Mutat Res* 400(1-2): 287-98.
- Rebollo et al (2001). "The association of Aiolos transcription factor and Bcl-xL is involved in the control of apoptosis." *J Immunol*. 167(11):6366-73.
- Reiley et al (2004). "Negative regulation of JNK signaling by the tumor suppressor CYLD." *J Biol Chem*. 279(53):55161-7.
- Remick, D. G. (1995). "Applied molecular biology of sepsis." *J Crit Care* 10(4): 198-212.
- Ren et al (2002). "E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints." *Genes Dev*. 16(2):245-56.
- Ren, L. Q., N. Gourmala, et al. (1998). "Lipopolysaccharide-induced expression of IP-10 mRNA in rat brain and in cultured rat astrocytes and microglia." *Brain Res Mol Brain Res* 59(2): 256-63.
- Reuter, U., A. Chiarugi, et al. (2002). "Nuclear factor-kappaB as a molecular target for migraine therapy." *Ann Neurol* 51(4): 507-16.
- Ridley, A. J. (1997). "The GTP-binding protein Rho." *Int J Biochem Cell Biol* 29(11): 1225-9.
- Roediger, W. E. (1980). "The colonic epithelium in ulcerative colitis: an energy-deficiency disease?" *Lancet* 2(8197): 712-5.

Roessler, E., Y. Z. Du, et al. (2003). "Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features." *Proc Natl Acad Sci U S A* 100(23): 13424-9.

Romero et al (1999). "Aiolos transcription factor controls cell death in T cells by regulating Bcl-2 expression and its cellular localization." *EMBO J.* 18(12):3419-30.

Ronaghi, M., M. Uhlen, et al. (1998). "A sequencing method based on real-time pyrophosphate." *Science* 281(5375): 363, 365.

Rosenecker, J., K. H. Harms, et al. (1996). "Adenovirus infection in cystic fibrosis patients: implications for the use of adenoviral vectors for gene transfer." *Infection* 24(1): 5-8.

Sahl et al (1998). "Granulocyte-macrophage colony-stimulating factor and interleukin-3 potentiate interferon-gamma-mediated endothelin production by human monocytes: role of protein kinase C. *Immunology.*" 95(3):473-9.

Saiki, R. K., T. L. Bugawan, et al. (1986). "Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes." *Nature* 324(6093): 163-6.

Saiki, R. K., P. S. Walsh, et al. (1989). "Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes." *Proc Natl Acad Sci U S A* 86(16): 6230-4.

Saleeba, J. A. and R. G. Cotton (1993). "Chemical cleavage of mismatch to detect mutations." *Methods Enzymol* 217: 286-95.

Samulski, R. J., L. S. Chang, et al. (1989). "Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression." *J Virol* 63(9): 3822-8.

Sandborn, W. J. and W. A. Faubion (2004). "Biologics in inflammatory bowel disease: how much progress have we made?" *Gut* 53(9): 1366-73.

Sanders, L. C., F. Matsumura, et al. (1999). "Inhibition of myosin light chain kinase by p21-activated kinase." *Science* 283(5410): 2083-5.

- Sandford, A. J., T. Shirakawa, et al. (1993). "Localisation of atopy and beta subunit of high-affinity IgE receptor (Fc epsilon RI) on chromosome 11q." *Lancet* 341(8841): 332-4.
- Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." *Proc Natl Acad Sci U S A* 74(12): 5463-7.
- Sans, M., D. Tassies, et al. (2003). "The 4G/4G genotype of the 4G/5G polymorphism of the type-1 plasminogen activator inhibitor (PAI-1) gene is a determinant of penetrating behaviour in patients with Crohn's disease." *Aliment Pharmacol Ther* 17(8): 1039-47.
- Santelli, E., M. Leone, et al. (2005). "Structural analysis of Siah1-Siah-interacting protein interactions and insights into the assembly of an E3 ligase multiprotein complex." *J Biol Chem*. 280(40): 34278-87.
- Sarraf et al (1998). "Differentiation and reversal of malignant changes in colon cancer through PPARgamma." *Nat Med* 4(9):1046-52.
- Sato, A., S. Kishida, et al. (2004). "Sall1, a causative gene for Townes-Brocks syndrome, enhances the canonical Wnt signaling by localizing to heterochromatin." *Biochem Biophys Res Commun*. 319(1): 103-13.
- Satoh, A. K., F. Tokunaga, et al. (1997). "Rab proteins of *Drosophila melanogaster*: novel members of the Rab-protein family." *FEBS Lett* 404(1): 65-9.
- Sauter, E. R., M. Herlyn, et al. (2000). "Prolonged response to antisense cyclin D1 in a human squamous cancer xenograft model." *Clin Cancer Res* 6(2): 654-60.
- Schaffer, C. J. and L. B. Nanney (1996). "Cell biology of wound healing." *Int Rev Cytol* 169: 151-81.
- Schreck, R., P. Rieber, et al. (1991). "Reactive oxygen intermediates as apparently widely used messengers in the activation of the NF-kappa B transcription factor and HIV-1." *Embo J* 10(8): 2247-58.
- Segain, J. P., D. Raingeard de la Bletiere, et al. (2003). "Rho kinase blockade prevents inflammation via nuclear factor kappa B inhibition: evidence in Crohn's disease and experimental colitis." *Gastroenterology* 124(5): 1180-7.
- Sells, M. A., U. G. Knaus, et al. (1997). "Human p21-activated kinase (Pak1) regulates actin organization in mammalian cells." *Curr Biol* 7(3): 202-10.

Shekarabi, M., S. W. Moore, et al. (2005). "Deleted in colorectal cancer binding netrin-1 mediates cell substrate adhesion and recruits Cdc42, Rac1, Pak1, and N-WASP into an intracellular signaling complex that promotes growth cone expansion." *J Neurosci* 25(12): 3132-41.

Shi, G. X., K. Harrison, et al. (2004). "Toll-like receptor signaling alters the expression of regulator of G protein signaling proteins in dendritic cells: implications for G protein-coupled receptor signaling." *J Immunol.* 172(9): 5175-84.

Shi, Y., W. Hu, et al. (2004). "Regulation of drug sensitivity of gastric cancer cells by human calcyclin-binding protein (CacyBP)." *Gastric Cancer* 7(3): 160-6.

Silkov et al (2002). "Enhanced apoptosis of B and T lymphocytes in TAFII105 dominant-negative transgenic mice is linked to nuclear factor-kappa B." *J Biol Chem.* 277(20):17821-9.

Shivakumar, L., J. Minna, et al. (2002). "The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation." *Mol Cell Biol.* 22(12): 4309-18.

Simon, H., I. Fortsch, et al. (1999). "Triple helix formation inhibits DNA gyrase activity." *Antisense Nucleic Acid Drug Dev* 9(6): 527-31.

Smith, C. A., T. Farrah, et al. (1994). "The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death." *Cell* 76(6): 959-62.

Smith, J. and P. Modrich (1996). "Mutation detection with MutH, MutL, and MutS mismatch repair proteins." *Proc Natl Acad Sci U S A* 93(9): 4374-9.

Smith, M. J., S. D. Gitlin, et al. (2001). "GLI-2 modulates retroviral gene expression." *J Virol* 75(5): 2301-13.

Sommerfelt, M. A. and R. A. Weiss (1990). "Receptor interference groups of 20 retroviruses plating on human cells." *Virology* 176(1): 58-69.

Steindler, C., Z. Li, et al. (2004). "Jamip1 (marlin-1) defines a family of proteins interacting with janus kinases and microtubules." *J Biol Chem* 279(41): 43168-77.

Steinhauer, D. A., S. A. Wharton, et al. (1991). "Amantadine selection of a mutant influenza virus containing an acid-stable hemagglutinin glycoprotein: evidence for virus-

specific regulation of the pH of glycoprotein transport vesicles." *Proc Natl Acad Sci U S A* 88(24): 11525-9.

Sterman, D. H., J. Treat, et al. (1998). "Adenovirus-mediated herpes simplex virus thymidine kinase/ganciclovir gene therapy in patients with localized malignancy: results of a phase I clinical trial in malignant mesothelioma." *Hum Gene Ther* 9(7): 1083-92.

Stevens, T. H. and M. Forgac (1997). "Structure, function and regulation of the vacuolar (H⁺)-ATPase." *Annu Rev Cell Dev Biol* 13: 779-808.

Stohr, H., N. Mohr, et al. (2002). "Cloning and characterization of WDR17, a novel WD repeat-containing gene on chromosome 4q34." *Biochim Biophys Acta* 1579(1): 18-25.

Storm, N., B. Darnhofer-Patel, et al. (2003). "MALDI-TOF mass spectrometry-based SNP genotyping." *Methods Mol Biol* 212: 241-62.

Straub, R. H., K. Stebner, et al. (2005). "Key role of the sympathetic microenvironment for the interplay of tumour necrosis factor and interleukin 6 in normal but not in inflamed mouse colon mucosa." *Gut*. 54(8): 1098-106.

Strom, T. M., K. Horthagel, et al. (1998). "Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein." *Hum Mol Genet*. 7(13): 2021-8.

Sugawara et al (2005). "Linkage to peroxisome proliferator-activated receptor-gamma in SAMP1/YitFc mice and in human Crohn's disease." *Gastroenterology*. 128(2):351-60.

Sugimura, K., K. D. Taylor, et al. (2003). "A novel NOD2/CARD15 haplotype conferring risk for Crohn disease in Ashkenazi Jews." *Am J Hum Genet* 72(3): 509-18.

Suzuki, S., L. F. Chuang, et al. (2002). "Interactions of opioid and chemokine receptors: oligomerization of mu, kappa, and delta with CCR5 on immune cells." *Exp Cell Res*. 280(2): 192-200.

Tanaka, S., M. Mori, et al. (1997). "Coexpression of Grb7 with epidermal growth factor receptor or Her2/erbB2 in human advanced esophageal carcinoma." *Cancer Res*. 57(1): 28-31.

Tanaka, S., M. Mori, et al. (1998). "A novel variant of human Grb7 is associated with invasive esophageal carcinoma." *J Clin Invest* 102(4): 821-7.

- Teller, I. C. and J. F. Beaulieu (2001). "Interactions between laminin and epithelial cells in intestinal health and disease." *Expert Rev Mol Med*. 2001: 1-18.
- Tratschin, J. D., M. H. West, et al. (1984). "A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase." *Mol Cell Biol* 4(10): 2072-81.
- Tremaine, W. J. (2003). Clinical features and complications of Crohn's disease. Inflammatory bowel disease, from bench to bedside. S. R. Targan, F. Shanahan and L. C. Karp. Dordrecht, Kluwer Acad. Pub.: 291-304.
- Trinchieri, G., S. Pflanz, et al. (2003). "The IL-12 family of heterodimeric cytokines: new players in the regulation of T cell responses." *Immunity* 19(5): 641-4.
- Ustun, S., N. Turgay, et al. (2004). "Interleukin (IL) 5 levels and eosinophilia in patients with intestinal parasitic diseases." *World J Gastroenterol* 10(24): 3643-6.
- Van Molle, W., B. Wielockx, et al. (2002). "HSP70 protects against TNF-induced lethal inflammatory shock." *Immunity*. 16(5): 685-95.
- Vermeire, S., G. Wild, et al. (2002). "CARD15 Genetic Variation in a Quebec Population: Prevalence, Genotype- Phenotype Relationship, and Haplotype Structure." *Am J Hum Genet* 71(1): 74-83.
- Vitt, U. A., S. Mazerbourg, et al. (2002). "Bone morphogenetic protein receptor type II is a receptor for growth differentiation factor-9." *Biol Reprod*. 67(2): 473-80.
- Vojtek, A. B. and C. J. Der (1998). "Increasing complexity of the Ras signaling pathway." *J Biol Chem* 273(32): 19925-8.
- Wang, P., P. Wu, et al. (1995). "Interleukin (IL)-10 inhibits nuclear factor kappa B (NF kappa B) activation in human monocytes. IL-10 and IL-4 suppress cytokine synthesis by different mechanisms." *J Biol Chem* 270(16): 9558-63.
- Warabi, K., M. D. Richardson, et al. (2002). "Human substance P receptor undergoes agonist-dependent phosphorylation by G protein-coupled receptor kinase 5 in vitro." *FEBS Lett* 521(1-3): 140-4.
- Weichert, D., J. Gobom, et al. (2006). "Analysis of NOD2-mediated proteome response to muramyl dipeptide in HEK293 cells." *J Biol Chem*. 281(4): 2380-9.

- West, M. H., J. P. Trempe, et al. (1987). "Gene expression in adeno-associated virus vectors: the effects of chimeric mRNA structure, helper virus, and adenovirus VA1 RNA." *Virology* 160(1): 38-47.
- Wild, G. E. and J. D. Rioux (2004). "Genome scan analyses and positional cloning strategy in IBD: successes and limitations." *Best Pract Res Clin Gastroenterol* 18(3): 541-553.
- Wilson, C., M. S. Reitz, et al. (1989). "Formation of infectious hybrid virions with gibbon ape leukemia virus and human T-cell leukemia virus retroviral envelope glycoproteins and the gag and pol proteins of Moloney murine leukemia virus." *J Virol* 63(5): 2374-8.
- Wolf, F. W., R. M. Marks, et al. (1992). "Characterization of a novel tumor necrosis factor-alpha-induced endothelial primary response gene." *J Biol Chem* 267(2): 1317-26.
- Wonsey, D. R., K. I. Zeller, et al. (2002). "The c-Myc target gene PRDX3 is required for mitochondrial homeostasis and neoplastic transformation." *Proc Natl Acad Sci U S A*. 99(10): 6649-54.
- Wu, J. Y., Y. Jin, et al. (2005). "Impaired TGF-beta responses in peripheral T cells of G alpha i2-/- mice." *J Immunol*. 174(10): 6122-8.
- Yamada, R., T. Tanaka, et al. (2001). "Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinatorial effect that two genetic loci have on susceptibility to the disease." *Am J Hum Genet* 68(3): 674-85.
- Yamamoto, S., G. Yang, et al. (2003). "Activation of Mst1 causes dilated cardiomyopathy by stimulating apoptosis without compensatory ventricular myocyte hypertrophy." *J Clin Invest* 111(10): 1463-74.
- Yamazaki, K., M. Takazoe, et al. (2002). "Absence of mutation in the NOD2/CARD15 gene among 483 Japanese patients with Crohn's disease." *J Hum Genet* 47(9): 469-72.
- Yamit-Hezi, A. and R. Dikstein (1998). "TAFII105 mediates activation of anti-apoptotic genes by NF-kappaB." *Embo J*. 17(17): 5161-9.

- Yan, F., S. K. John, et al. (2004). "Kinase suppressor of Ras-1 protects intestinal epithelium from cytokine-mediated apoptosis during inflammation." *J Clin Invest* 114(9): 1272-80.
- Yin et al (2003). "Cloning and characterization of the human IFT20 gene." *Mol Biol Rep.* 30(4):255-60.
- Yu, M., E. Poeschla, et al. (1994). "Progress towards gene therapy for HIV infection." *Gene Ther* 1(1): 13-26.
- Yu, Q., S. J. Cok, et al. (2003). "Translational repression of human matrix metalloproteinases-13 by an alternatively spliced form of T-cell-restricted intracellular antigen-related protein (TIAR)." *J Biol Chem.* 278(3): 1579-84.
- Zaratin, P. F., A. Quattrini, et al. (2005). "Schwann cell overexpression of the GPR7 receptor in inflammatory and painful neuropathies." *Mol Cell Neurosci* 28(1): 55-63.
- Zhang, W. J., W. A. Koltun, et al. (2000). "Absence of GNAI2 codon 179 oncogene mutations in inflammatory bowel disease." *Inflamm Bowel Dis.* 6(2): 103-6.
- Zhang, M., P. Liu, et al. (2002). "MLN64 mediates mobilization of lysosomal cholesterol to steroidogenic mitochondria." *J Biol Chem.* 277(36): 33300-10.
- Zhang, S. X., E. G. Gras, et al. (2005). "Identification of direct serum response factor gene targets during DMSO induced P19 cardiac cell differentiation." *J Biol Chem* 28: 28.
- Zhang, Z., A. Andoh, et al. (2005). "Interleukin-1beta and tumor necrosis factor-alpha upregulate interleukin-23 subunit p19 gene expression in human colonic subepithelial myofibroblasts." *Int J Mol Med* 15(1): 79-83.
- Zhou, J., J. Ma, et al. (2004). "BRD7, a novel bromodomain gene, inhibits G1-S progression by transcriptionally regulating some important molecules involved in ras/MEK/ERK and Rb/E2F pathways." *J Cell Physiol.* 200(1): 89-98.

BOOKS:

Abbas AK, Litchman AH. Cellular and Molecular Immunology. Philadelphia: Saunders; 1994. 417 p.

Austen BM and. Westwood OMR. Protein Targeting and Secretion. Oxford: IRL Press; 1991. 85 p.

Bishop MJ, editor. Guide to Human Genome Computing, 2d ed. San Diego: Academic Press; 1998. 306 p.

Cowell IG, Austin CA, editors. DNA Library Protocols. Methods in Molecular Biology. Vol. 69 Totowa, N.J.:Humana Press; 1997. 321p.

Freshney RI, editor. Animal Cell Culture: A Practical Approach. Oxford: IRL Press; 1986.

Freshney RI. Culture Of Animal Cells: A Manual of Basic Technique. New York: AR Liss; 1987. 397 p.

Glover DM, editor. DNA Cloning: A Pratical Approach. Vols 1 & 2. Oxford; Washington: IRL Press; 1985.

Gribskov M, Devereux J, editors. Sequence Analysis Primer. Oxford University Press; 1994. 296 p.

Griffin AM, Griffin HG, editors. ComputerAnalysis of Sequence Data, Part 1. Totowa, N.J.Humana Press; 1994. 392 p.

Hames BD, Higgins SJ, editors. Nucleic Acid Hybridization: A Practical Approach. Oxford: IRL Press; 1985. 245 p.

Hames BD, Higgins SJ, editors. Transcription and Translation: A Practical Approach. Oxford: IRL Press; 1984. 328 p.

Harlow Ed, Lane D. Antibodies: A Laboratory Manual. New York: Cold Spring Harbor Laboratory;1988. 726 p.

Heinje G. von. Sequence Analysis in Molecular Biology. San Diego: Academic Press; 1987. 188 p.

Hogan B, Costantini F, Lacy E, editors. Manipulating the Mouse Embryo: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press; 1986. 332 p.

Huber BE, Carr BI. Molecular and Immunologic Approaches. Mt. Kisco, NY: Futura Publishing Co; 1994.

Jones J. Amino Acid and Peptide Synthesis. Oxford; New York: Oxford Science Publications; 1992. 86 p.

Kaufman PB, William W, Donghern K, editors. Handbook of Molecular and Cellular Methods in Biology and Medicine. Boca Raton: CRC Press; 1995. 484 p.

Lesk AM, editor. Computational Molecular Biology: sources and methods for sequence analysis. New York: Oxford University Press; 1988. 254p.

Male D, Cooke A, Owen M, Trowsdale J, Champion B, editors. Advanced Immunology. 3rd ed. London; Baltimore: Mosby; 1996. 273 p.

McPherson MJ, editor. Directed Mutagenesis: A Practical Approach. New York: IRL Press; 1991. 257 p.

McPherson MJ, Quirke P, Taylor JR, editors. PCR: A Practical Approach. Oxford; New York: IRL Press; 1991. 253 p.

Miller JH, Calos MP, editors. Gene Transfer Vectors for Mammalian Cells. New York: Cold Spring Harbor Laboratory Press; 1987. 169 p.

Miller JH, Calos MP, editors. Gene Transfer Vectors For Mammalian Cells. New York: Cold Spring Harbor Laboratory; 1987. 169p.

Pawlowitzki IH, Edwards JH, Thompson EA, editors. Genetic Mapping of Disease Genes. Academic Press London; 1997. 288 p.

Perbal BV. A Practical Guide to Molecular Cloning. 1st ed. New York: Wiley Interscience Publication; 1984. 554 p.

Perbal BV. A Practical Guide To Molecular Cloning. New York: Wiley; 1984. 554 p.

Peruski LF, Peruski AH. The Internet and the New Biology. Tools for Genomic and Molecular Research. Washington, D.C.: American Society for Microbiology Press; 1997.

Sambrook J. Molecular Cloning: A Laboratory Manual. 2nd ed. 3 vols. New York: Cold Spring Harbor Laboratory Press; 1989.

Sell S. Immunology, Immunopathology & Immunity. 5th ed. Stamford, CT: Appleton & Lange; 1996. 1014 p.

Smith DW, editor. Biocomputing. Informatics and Genome Projects, New York: Academic Press; 1993. 336p.

Stites DP, Terr AT, editors. Basic and Clinical Immunology. 7th ed. Norwalk, CT: Appleton & Lange; 1991. 870 p.

Walker JM. Protein Protocols on CD-ROM, Humana Press, Totowa, NJ.

Weir DM, Herzenberg LA, Blackwell C, editors. Handbook Of Experimental Immunology. 4 vols. Oxford: Blackwell; 1986.

Woodward J. Immobilized Cells And Enzymes: A Practical Approach. Oxford: IRL Press; 1986.

Wu R, Grossman L, editors. Methods in Enzymology: Rexcombinant DNA Part E. Vol. 154. Amsterdam: Elsevier Science; 1987. 576 p.

Wu R, Grossman L, editors. Methods in Enzymology: Rexcombinant DNA Part F. Vol. 155. Amsterdam: Elsevier Science; 1987. 628 p.

Patents

U.S. 4,683,202.

U.S. 4,952,501.

WO03042661A2

US 20040009479A1

U.S. 5,315,000

WO1997US0005216

U.S. 5,498,531

U.S. 5,807,718

U.S. 5,888,819

U.S. 6,090,543

WO 2007/073478

PCT/US2006/048246

U.S. 6,090,606

U.S. 5,585,089

U.S. 4,683,195

U.S. 4,683,202

U.S. 5,459,039.

U.S. 6,090,543).

U.S. 6,090,606

U.S. 5,869,242

U.S. 60/335,068

U.S. 6,479,244

PCT/US94/05700

U.S. 4,797,368

WO 93/24641

U.S. 5,173,414

WE CLAIM:

1. A method of constructing a GeneMap for Crohn's disease comprising identifying at least two chromosomal loci associated with Crohn's disease, wherein said at least two chromosomal loci are selected from the genomic regions listed in Table 1.
2. The method of claim 1, wherein said population is a general population.
3. The method of claim 1, wherein said population is a founder population.
4. The method of claim 3, wherein said founder population is the population of Quebec.
5. The method of claim 1, wherein said at least two chromosomal regions are selected from the genes in Table 2, 3, or 4.
6. The method of claim 5, wherein said genes are used to construct gene networks based on the functional relationship of gene products interactions.
7. The method of claim 6, wherein the interactions are direct, indirect, or a combination thereof.
8. The method of claim 1, wherein the identifying comprises screening for the presence or absence of at least one single nucleotide polymorphisms (SNPs) from Tables 5 and 6, in at least one sample.
9. The method of claim 8, wherein the screening comprises the steps of: (a) obtaining biological samples from at least one disease patient; (b) screening for the presence or absence of at least one SNP or a group of SNPs from Tables 5-36 within each biological sample; and (c) evaluating whether said SNP or a group of SNPs shows a statistically significant skewed genotype distribution between a group of patients compared to a control.
10. The method of claim 9, wherein said biological samples are fluid, serum, tissue or buccal swabs, saliva, mucus, urine, stools, spermatozoids, vaginal secretions, lymph, amniotic liquid, pleural liquid or tears.

11. The method of claim 9, wherein said patients and controls are from a human population.
12. The method of claim 11, wherein said patients and controls are recruited independently according to specific phenotypic criteria.
13. The method of claim 11, wherein said patients and controls are recruited in the form of trios comprising two parents and one child or one parent and two children.
14. The method of claim 8, wherein said screening is performed by a method selected from the group consisting of an allele-specific hybridization assay, an oligonucleotide ligation assay, an allele-specific elongation/ligation assay, an allele-specific amplification assay, a single-base extension assay, a molecular inversion probe assay, an invasive cleavage assay, a selective termination assay, RFLP, a sequencing assay, SSCP, a mismatch-cleaving assay, and denaturing gradient gel electrophoresis.
15. The method of claim 8, wherein said screening is carried out on each individual of a cohort at each of at least one SNP or a group of SNPs from Tables 5-36.
16. The method of claim 8, wherein said screening is carried out on pools of patients and pools of controls.
17. The method of claim 8, wherein the genotype distribution is compared one SNP at a time.
18. The method of claim 8, wherein the genotype distribution is compared with a group of markers from Tables 5-36 forming a haplotype.
19. The method of claim 17, wherein the genotype distribution is compared using the allelic frequencies between the patient pools and control pools.
20. The method of claim 1, wherein the GeneMap comprises all of the genes of Tables 2, 3 and 4.
21. A method of diagnosing Crohn's disease, the predisposition to Crohn's disease, or the progression or prognostication of Crohn's disease, comprising

determining the amount and/or concentration of at least one polypeptide from Table 2, 3 or 4 and/or at least one nucleic acid encoding the polypeptide present in said biological sample

22. The method of claim 21, wherein the diagnosing comprises the steps of: (a) obtaining a biological sample of mammalian body fluid or tissue to be diagnosed; (b) comparing the amount and/or concentration of said polypeptide and/or nucleic acid encoding the polypeptide determine in said biological sample with the amount and/or concentration of said polypeptide and/or nucleic acid encoding the polypeptide as determined in a control sample, wherein the difference in the amount of said polypeptide and/or nucleic acid encoding the polypeptide is indicative of Crohn's disease or the stage of Crohn's disease.
23. The method of claim 21, wherein a nucleic acid probe is used for determining the amount and/or concentration of at least one nucleic acid sequence from Table 2, 3 or 4 encoding the polypeptide.
24. The method of claim 23, wherein said nucleic acid probe is selected from the nucleic acid sequences designated as SEQ ID NO: 1 to 21466.
25. The method of claim 23, wherein said nucleic acid probe comprises nucleic acids hybridizing to the nucleic acid sequences designated as SEQ ID NO: 1 to 21466, and/or fragments thereof.
26. The method of claim 23, wherein said nucleic acid probe comprises nucleic acids hybridizing to at least five nucleic acid sequences from Table 2, 3 or 4.
27. The method of claim 23, wherein said nucleic acid probe specifically hybridizes to at least 10 nucleic acid sequences from Table 2, 3 or 4.
28. The method of claim 23, wherein said nucleic acid probe specifically hybridizes to at least 20 nucleic acid sequences from Table 2, 3 or 4.
29. The method of claim 23, wherein said nucleic acid probe specifically hybridizes to at least 50 nucleic acid sequences from Table 2, 3 or 4.
30. The method of claim 23, wherein said nucleic acid probe specifically hybridizes to at least 100 nucleic acid sequences from Table 2, 3 or 4.

31. The method of claim 23, wherein said nucleic acid probe specifically hybridizes to at least 100 nucleic acid sequences from Table 2, 3 or 4.
32. The method of claim 23, wherein said nucleic acid probe is at least about 10 nucleotides in length.
33. The method of claim 23, wherein said nucleic acid probe is at least about 30 nucleotides in length.
34. The method of claim 23, wherein said nucleic acid probe is at least about 50 nucleotides in length.
35. The method of claim 23, wherein a PCR technique is used for determining the amount and/or concentration of at least one nucleic acid from Table 2, 3 or 4.
36. The method of claim 21, wherein a specific antibody is used for determining the amount and/or concentration of at least one polypeptide from Table 2, 3 or 4.
37. The method of claim 36 wherein said antibody is selected from the group comprising polyclonal antiserum, polyclonal antibody, monoclonal antibody, antibody fragments, single chain antibodies and diabodies.
38. The method of claim 21, wherein the amounts and/or concentrations of at least five polypeptides or nucleic acids are determined.
39. A method of detecting susceptibility to Crohn's disease comprising detecting at least one mutation or polymorphism in the nucleic acid molecule selected from Table 2, 3 or 4 in a patient.
40. The method of claim 39, wherein said method comprises hybridizing a probe to said patient's sample of DNA or RNA under stringent conditions which allow hybridization of said probe to nucleic acid comprising said mutation or polymorphism, wherein the presence of a hybridization signal indicates the presence of said mutation or polymorphism in at least one gene from Table 2, 3 or 4.
41. The method of claim 39, wherein the patient's DNA or RNA has been amplified and said amplified DNA or RNA is hybridized.

42. The method of claim 39, wherein said method comprises using a single-stranded conformation polymorphism technique to assay for said mutation.
43. The method of claim 39, wherein said method comprises sequencing at least one gene from Table 2, 3 or 4 in a sample of RNA or DNA from a patient.
44. The method of claim 39, wherein said method comprises determining the sequence of at least one gene from Table 2, 3 or 4 by preparing cDNA from RNA taken from said patient and sequencing said cDNA to determine the presence or absence of a mutation.
45. The method of claim 39, wherein said method comprises performing an RNase assay.
46. The method of claim 39, wherein said probe is attached to a microarray or a bead.
47. The method of claim 39, wherein said probes are oligonucleotides.
48. The method of claim 40, wherein said sample is selected from the group consisting of blood, normal tissue and tumor tissue.
49. The method of claim 39, wherein the mutation is selected from the group consisting of at least one of the SNPs from Tables 5-366, alone or in combination.
50. The method of claim 21, further comprising comparing the level of expression or activity of a polypeptide of Table 2, 3 or 4 in a test sample from a patient with the level of expression or activity of the same polypeptide in a control sample wherein a difference in the level of expression or activity between the test sample and control sample is indicative of Crohn's disease.
51. A method of treatment of IBD in a mammal in need thereof, comprising the steps of: performing steps a) to c) according to claim 22; and treating the mammal in need of said treatment; wherein said medical treatment is based on the stage of the disease.
52. A method of diagnosing susceptibility to Crohn's disease in an individual, comprising screening for an at-risk haplotype of at least one gene or gene

region from Table 2, 3 or 4, that is more frequently present in an individual susceptible to Crohn's disease compared to a control individual, wherein the presence of the at-risk haplotype is indicative of a susceptibility to Crohn's disease.

53. The method of claim 52 wherein the at-risk haplotype is indicative of increased risk for Crohn's disease.
54. The method of claim 53, wherein the risk is increased at least about 20%.
55. The method of claim 52, wherein the at-risk haplotype is characterized by the presence of at least one single nucleotide polymorphism from Tables 5 and 6.
56. The method of claim 52, wherein screening for the presence of an at-risk haplotype in at least one gene from Table 2, 3 or 4, comprises enzymatic amplification of nucleic acid from said individual or amplification using universal oligos on elongation/ligation products.
57. The method of claim 56, wherein the nucleic acid is DNA.
58. The method of claim 57, wherein the DNA is human DNA.
59. The method of claim 52, wherein screening for the presence of an at-risk haplotype in at least one gene from Table 2, 3 or 4 comprises: (a) obtaining material containing nucleic acid from the individual; (b) amplifying said nucleic acid; and (c) determining the presence or absence of an at-risk haplotype in said amplified nucleic acid.
60. The method of claim 59, wherein determining the presence of an at-risk haplotype is performed by electrophoretic analysis.
61. The method of claim 59, wherein determining the presence of an at-risk haplotype is performed by restriction length polymorphism analysis.
62. The method of claim 59, wherein determining the presence of an at-risk haplotype is performed by sequence analysis.
63. The method of claim 59, wherein determining the presence of an at-risk haplotype is performed by hybridization analysis.

64. A method of diagnosing a susceptibility to Crohn's disease, comprising detecting an alteration in the expression or composition of a polypeptide encoded by at least one gene from Table 2, 3 or 4 in a test sample, in comparison with the expression or composition of a polypeptide encoded by said gene in a control sample, wherein the presence of an alteration in expression or composition of the polypeptide in the test sample is indicative of a susceptibility to Crohn's Disease.
65. The method of claim 64, wherein the alteration in the expression or composition of a polypeptide encoded by said gene comprises expression of a splicing variant polypeptide in a test sample that differs from a splicing variant polypeptide expressed in a control sample.
66. A drug screening assay comprising: a) administering a test compound to an animal having IBD, or a cell population isolated therefrom; and (b) comparing the level of gene expression of at least one gene from Table 2, 3 or 4 in the presence of the test compound with the level of said gene expression in normal cells; wherein test compounds which provide the level of expression of one or more genes from Table 2, 3 or 4 similar to that of the normal cells are candidates for drugs to treat IBD.
67. A pharmaceutical preparation for treating an animal having IBD comprising a compound identified by the assay of claim 66 and a pharmaceutically acceptable excipient.
68. A method for treating an animal having IBD comprising administering a compound identified by the assay of claim 66.
69. A method for predicting the efficacy of a drug for treating IBD in a human patient, comprising: (a) obtaining a sample of cells from the patient; (b) obtaining a gene expression profile from the sample in the absence and presence of the drug ; the gene expression profile comprising one or more genes from Table 2, 3 or 4; and (c) comparing the gene expression profile of the sample with a reference gene expression profile, wherein similarity between the sample expression profile and the reference expression profile predicts the efficacy of the drug for treating IBD in the patient.

70. The method of claim 69, further comprising exposing the sample to the drug for treating IBD prior to obtaining the gene expression profile of the sample.
71. The method of claim 69, wherein the sample of cells is derived from a tissue selected from the group consisting of: the jejunum, ileum, mucosa, submucosa, cecum, inner and outer intestinal coatings, muscle and nervous tissue.
72. The method of claim 71, wherein the cells are selected from the group consisting of: smooth muscle cell, neutrophil, dendritic cell, T cell, mast cell, CD4+ lymphocyte, monocyte, macrophage, dendritic cell, synovial cell, glial cell, villous intestinal cell, neutrophilic granulocyte, eosinophilic granulocyte, keratinocyte, lamina propria lymphocyte, intraepithelial lymphocyte and epithelial cell.
73. The method of claim 69, wherein the sample is obtained via small bowel or colon biopsy.
74. The method of claim 69, wherein the gene expression profile comprises expression values for all of the genes listed in Table 2, 3 or 4.
75. The method of claim 74, wherein the gene expression profile of the sample is obtained by detecting the protein products of said genes.
76. The method of claim 69, wherein the gene expression profile of the sample is obtained using a hybridization assay to oligonucleotides contained in a microarray.
77. The method of claim 76, wherein the oligonucleotides comprises nucleic acid molecules at least 95% identical to the gene sequences from Tables 2, 3 and 4.
78. The method of claim 69, wherein the reference expression profile is that of cells derived from patients that do not have IBD.
79. The method of claim 69, wherein the drug is selected from the group consisting of symptom relievers and anti-inflammatory drugs for an inflammatory disease condition.

80. The method of claim 69, wherein said patient's sample of DNA has been amplified or cloned.
81. A method for predicting the efficacy of a drug for treating IBD in a human patient, comprising: a) obtaining a sample of cells from the patient; b) obtaining a set of genotypes from the sample, wherein the set of genotypes comprises genotypes of one or more polymorphic loci from Tables 2, 3, 4, 5, and 6; and c) comparing the set of genotypes of the sample with a set of genotypes associated with efficacy of the drug, wherein similarity between the set of genotypes of the sample and the set of genotypes associated with efficacy of the drug predicts the efficacy of the drug for treating IBD in the patient.
82. The method of claim 81, wherein the sample of cells is derived from a tissue selected from the group consisting of: the scalp, GI track, muscle, sebaceous gland, nerve, blood, dermis, epidermis and other skin cells, cutaneous surfaces, intertrigous areas, genitalia, vessels and endothelium.
83. The method of claim 82, wherein the cells are selected from the group consisting of: melanocytes, hair follicle cells, muscle cells, nerve cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells and lymphocytes.
84. The method of claim 81, wherein the sample is obtained via biopsy.
85. The method of claim 81, wherein the set of genotypes from the sample comprises genotypes of at least two of the polymorphic loci listed in Tables 2, 3, 4, 5-36.
86. The method of claim 81 wherein the set of genotypes from the sample is obtained by hybridization to allele-specific oligonucleotides complementary to the polymorphic loci from Tables 2, 3, 4, 5-36, wherein said allele-specific oligonucleotides are contained on a microarray.
87. The method of claim 86, wherein the oligonucleotides comprise nucleic acid molecules at least 95% identical to SEQ ID from Tables 2, 3, 4, 5-36.

88. The method of claim 81 wherein the set of genotypes from the sample is obtained by sequencing said polymorphic loci in said sample.
89. The method of claim 81, wherein the drug is selected from the group consisting of symptom relievers and drugs for IBD.
90. A method of treating IBD in a patient in need thereof, comprising expressing *in vivo* at least one gene from Tables 2, 3, or 4 in an amount sufficient to treat the disease.
91. The method of claim 90, comprising: (a) administering to a patient a vector comprising a gene selected from Table 2, 3 or 4 that encodes the protein; and (b) allowing said protein to be expressed from said gene in said patient in an amount sufficient to treat the disease.
92. The method of claim 91, wherein said vector is selected from the group consisting of an adenoviral vector, and a lentiviral vector.
93. The method of claim 91, wherein said vector is administered by a route selected from the group consisting of: topical administration, intraocular administration, parenteral administration, intranasal administration, intratracheal administration, intrabronchial administration and subcutaneous administration.
94. The method of claim 91, wherein said vector is a replication-defective viral vector.
95. The method of claim 91, wherein said gene encodes a human protein.
96. A method of treating IBD in a patient in need thereof, comprising administering an agent that regulates the expression, activity or physical state of at least one gene or its encoding RNA from Table 2, 3 or 4 in the patient.
97. The method of claim 96, wherein the encoded protein from said gene comprises an alteration.
98. The method of claim 96, wherein said gene comprises a mutation that modulates the expression of the encoded protein.

99. The method of claim 96, wherein said agent is selected from the group consisting of chemical compounds, oligonucleotides, peptides and antibodies.
100. The method of claim 99, wherein said agent is an antisense molecule or interfering RNA.
101. The method of claim 99, wherein said agent is an expression modulator.
102. The method of claim 101, wherein said modulator is an activator.
103. The method of claim 101, wherein said modulator is a repressor.
104. The method of claim 96, wherein said gene comprises a mutation that modifies at least one property or function of the encoded protein.
105. The method of claim 96, wherein the agent modulates at least one property or function of said gene.
106. A method of treating IBD in a patient in need thereof, comprising administering an agent that regulates the expression, activity or physical state of at least one polypeptide encoded by a gene from Table 2, 3 or 4 in the patient.
107. The method of claim 106, wherein the encoded protein from said gene comprises an alteration, wherein said alteration is encoded by a polymorphic locus in said gene.
108. The method of claim 106, wherein said gene comprises an associated allele, a particular allele of a polymorphic locus, or the like that modulates the expression of the encoded protein.
109. The method of claim 106, wherein said agent is selected from the group consisting of chemical compounds, oligonucleotides, peptides and antibodies.
110. The method of claim 106, wherein said agent is an antisense molecule or interfering RNA.
111. The method of claim 106, wherein said agent is an expression modulator.
112. The method of claim 111, wherein said modulator is an activator.

113. The method of claim 111, wherein said modulator is a repressor.
114. The method of claim 106, wherein said gene comprises an associated allele, a particular allele of a polymorphic locus, or the like that modifies at least one property or function of the encoded protein.
115. A method for preventing the occurrence of IBD in an individual in need thereof, comprising regulating the level of at least one gene from Table 2, 3 or 4 compared to a control.
116. The method of claim 115, wherein said level is regulated by regulating expression of at least one gene from Table 2, 3 or 4 by a binding agent, a receptor to said gene, a peptidomimetic, a fusion protein, a prodrug, an antibody or a ribozyme.
117. The method of claim 115, wherein said level is controlled by genetically altering the expression level of at least one gene from Table 2, 3 or 4, whereby the regulated level of said gene mimics the level in a healthy individual.
118. A method for identifying a gene that regulates drug response in IBD, comprising: (a) obtaining a gene expression profile for at least one gene from Table 2, 3 or 4 in a resident tissue cell induced for a pro-inflammatory like state in the presence of the candidate drug; and (b) comparing the expression profile of said gene to a reference expression profile for said gene in a cell induced for the pro-inflammatory like state in the absence of the candidate drug, wherein genes whose expression relative to the reference expression profile is altered by the drug may identifies the gene as a gene that regulates drug response in IBD.
119. A method for identifying an agent that alters the level of activity or expression of a polypeptide of Table 2, 3 or 4 for use in diagnostics, prognostics, prevention, treatment, or study of IBD, comprising: (a) contacting a cell, cell lysate, or the polypeptide, with an agent to be tested; (b) assessing a level of activity or expression of the polypeptide; and (c) comparing the level of activity or expression of the polypeptide with a control sample in an absence of the agent, wherein if the level of activity or expression of the polypeptide in the presence of the agent differs by an amount that is statistically significant

from the level in the absence of the agent then the agent alters the activity or expression of the polypeptide.

120. A kit for diagnosing susceptibility to Crohn's disease in an individual comprising: primers for nucleic acid amplification of a region of at least one gene from Table 2, 3 or 4.
121. The kit of claim 120, wherein the primers comprise a segment of nucleic acids of length suitable for nucleic acid amplification of a target sequence, selected from the group consisting of: single nucleotide polymorphism from Tables 5-36, and combinations thereof.
122. A kit for assessing a patient's risk of having or developing Crohn's disease, comprising: (a) detection means for detecting the differential expression, relative to a normal cell, of at least one gene shown in Table 2-4 or the gene product thereof; and (b) instructions for correlating the differential expression of said gene or gene product with a patient's risk of having or developing Crohn's disease.
123. The kit of claim 122, wherein the detection means includes nucleic acid probes for detecting the level of mRNA of said genes.
124. A kit for assessing a patients risk of having or developing Crohn's disease, comprising: (a) at least one means for amplifying or detecting a sequence of at least one gene in Table 2, 3 or 4, wherein the detection means includes nucleic acid probes or primers for detecting the presence or absence of mutations or changes to at least one sequence of Table 2, 3 or 4.
125. The kit of claim 124, wherein the detection means includes an immunoassay for detecting the level of at least one gene product from Table 2, 3 or 4.
126. A kit for assessing a patient's risk of having or developing Crohn's disease, comprising: a) a detection means for detecting the genotype of at least one polymorphic locus shown in Tables 2, 3, 4, or 5-36; and b) instructions for correlating the genotype of said at least one polymorphic locus with a patient's risk of having or developing Crohn's disease.

127. The kit of claim 126, wherein the detection means includes nucleic acid probes for detecting the genotype of said at least one polymorphic locus.
128. A diagnostic composition for diagnosing or detecting susceptibility to Crohn's disease comprising a set of oligonucleotide probes that specifically hybridizes to at least two geneonic regions listed in Table 1.
129. The composition of claim 128, wherein said set of oligonucleotide probes specifically hybridize to sequences of at least two genes selected from the genes in Table 2, 3 or 4.
130. The composition of claim 128, wherein the oligonucleotide probes are detectably labeled with an agent selected from the group consisting of a fluorescent dye, a radioisotope, a bioluminescent compound, a chemiluminescent compound, a fluorescent compound, a metal chelate and an enzyme.
131. The composition of claim 130, wherein the oligonucleotide probes are labeled with different fluorescent compounds.
132. The composition of claim 128, wherein the set of oligonucleotide probes hybridizes in situ.
133. The composition of claim 128, wherein the set of oligonucleotide probes hybridizes at a gradually changing temperature.
134. The composition of claim 128, wherein the oligonucleotide probes are between 2 to 100 bases.
135. The composition of claim 128, wherein the oligonucleotide probes are between 3 to 50 bases.
136. The composition of claim 128, wherein the oligonucleotide probes are between 8 to 25 bases.
137. A method of assessing a patient's risk of having or developing Crohn's disease, comprising: (a) determining the level of expression of at least one gene from Table 2, 3 or 4 or gene products thereof, and comparing the level of expression to a normal cell; and (b) assessing a patient's risk of having or

developing Crohn's disease by determining the correlation between the differential expression of said genes or gene products with known changes in expression of said genes measured in at least one patient suffering from Crohn's disease.

138. A method of assessing a patient's risk of having or developing Crohn's disease, comprising (a) determining a genotype for at least one polymorphic locus from Tables 2, 3, 4, or 5-36 in a patient; (b) comparing said genotype of (a) to a genotype for at least one polymorphic locus from Tables 2, 3, 4, or 5-36 that is associated with Crohn's disease; and (c) assessing the patient's risk of having or developing Crohn's disease, wherein said patient has a higher risk of having or developing Crohn's disease if the genotype for at least one polymorphic locus from Tables 2, 3, 4, or 5-36 in said patient is the same as said genotype for at least one polymorphic locus from Tables 2, 3, 4, 5, or 6 that is associated with Crohn's disease.
139. A method for assaying the presence of a nucleic acid associated with resistance or susceptibility to Crohn's disease in a sample, comprising: contacting said sample with a nucleic acid recited in claim 5 under stringent hybridization conditions; and detecting a presence of a hybridization complex.
140. A method for assaying the presence or amount of a polypeptide encoded by a gene of Tables 2, 3 or 4 for use in diagnostics, prognostics, prevention, treatment, or study of Crohn's disease, comprising: contacting a sample with an antibody that specifically binds to a gene of Tables 2, 3 or 4 under conditions appropriate for binding; and assessing the sample for the presence or amount of binding of the antibody to the polypeptide.