



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

Table with 4 columns: APPLICATION NUMBER, FILING OR 371(C) DATE, FIRST NAMED APPLICANT, ATTY. DOCKET NO./TITLE. Row 1: 14/957,122, 12/02/2015, Graham Alastair Richard HEAP, 2003144.00142US2

23483
WILMERHALE/BOSTON
60 STATE STREET
BOSTON, MA 02109

CONFIRMATION NO. 1519
PUBLICATION NOTICE



Title: Biomarkers for Increased Risk of Drug-Induced Thiopurine-Induced Pancreatitis
Publication No. US-2016-0153046-A1
Publication Date: 06/02/2016

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Office of Public Records. The Office of Public Records can be reached by telephone at (703) 308-9726 or (800) 972-6382, by facsimile at (703) 305-8759, by mail addressed to the United States Patent and Trademark Office, Office of Public Records, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently http://pair.uspto.gov/. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Management, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101



US 20160153046A1

(19) **United States**

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

(10) **Pub. No.: US 2016/0153046 A1**

(43) **Pub. Date: Jun. 2, 2016**

(54) **BIOMARKERS FOR INCREASED RISK OF  
DRUG-INDUCED THIOPURINE-INDUCED  
PANCREATITIS**

(71) Applicant: **Severe Adverse Event (SAE)  
Consortium**, Chicago, IL (US)

(72) Inventors: **Graham Alastair Richard HEAP**,  
Exeter Devon (GB); **Tariq AHMAD**,  
Exeter Devon (GB)

(21) Appl. No.: **14/957,122**

(22) Filed: **Dec. 2, 2015**

**Related U.S. Application Data**

(60) Provisional application No. 62/086,583, filed on Dec.  
2, 2014.

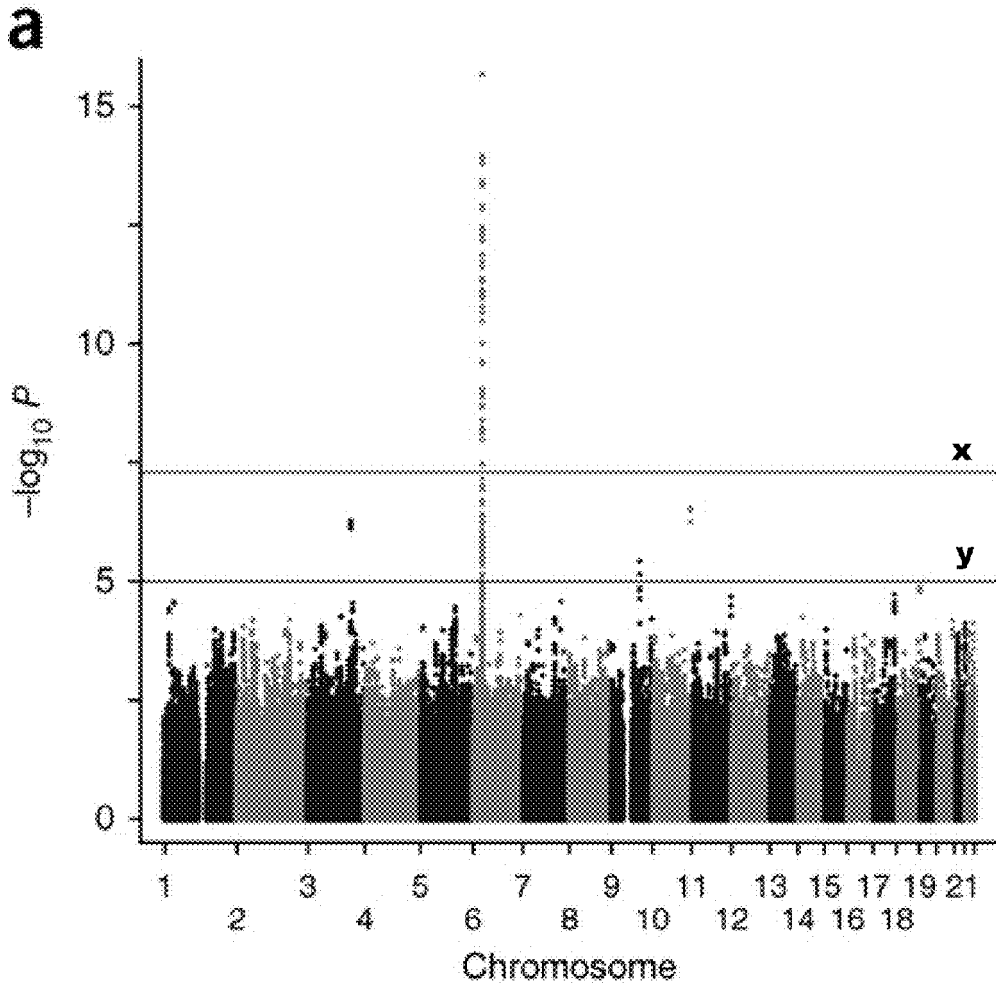
**Publication Classification**

(51) **Int. Cl.**  
*C12Q 1/68* (2006.01)

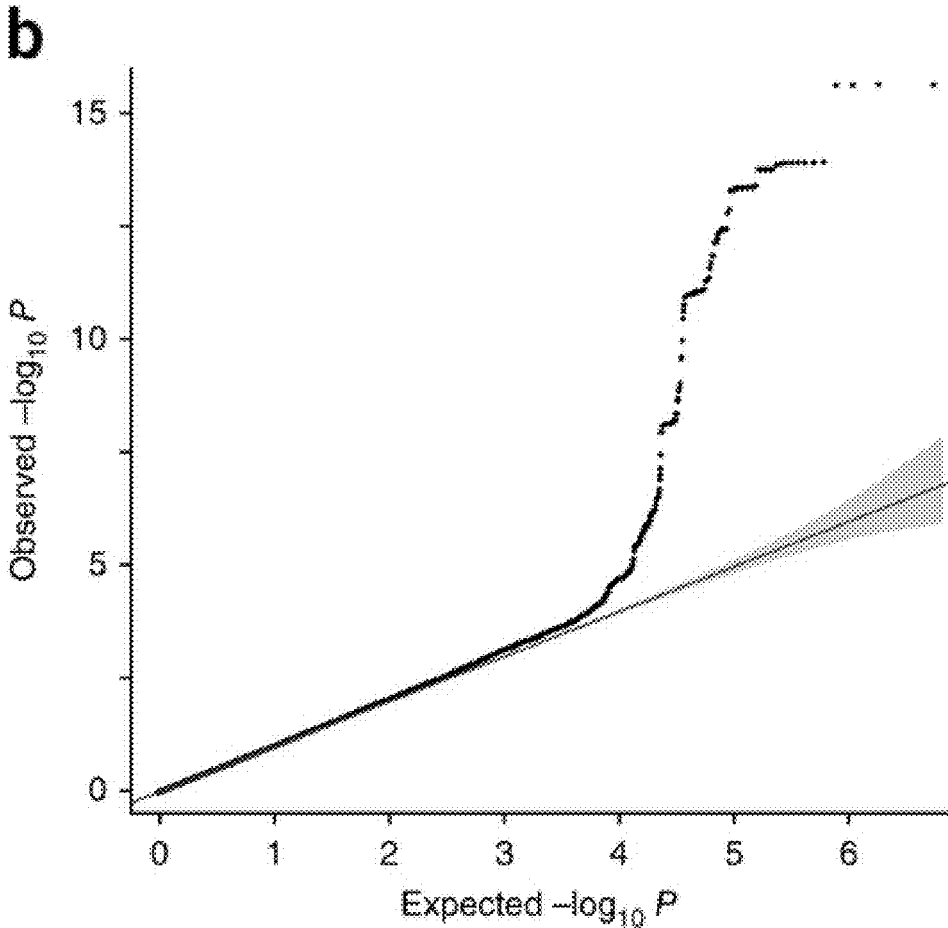
(52) **U.S. Cl.**  
CPC ..... *C12Q 1/6883* (2013.01); *C12Q 2600/156*  
(2013.01)

(57) **ABSTRACT**

The present invention provides a method for predicting the risk of a patient for developing adverse drug reactions, particularly thiopurine induced pancreatitis (TIP). The invention also provides a method of identifying a subject afflicted with, or at risk of, developing TIP. In some aspects, the methods comprise analyzing at least one genetic marker, wherein the presence of the at least one genetic marker indicates that the subject is afflicted with, or at risk of, developing TIP.



**FIGURE 1a**



**FIGURE 1b**

## BIOMARKERS FOR INCREASED RISK OF DRUG-INDUCED THIOPURINE-INDUCED PANCREATITIS

[0001] This application claims the benefit of priority of U.S. Provisional Application Ser. No. 62/086,583, filed Dec. 2, 2014, the contents of which are hereby incorporated by reference in its entirety.

### TECHNICAL FIELD

[0002] The present disclosure relates generally to methods for identifying genetic risk factors for adverse reactions to drugs. More specifically, the present disclosure relates to methods for predicting what drugs will cause thiopurine-induced pancreatitis, and in which patients.

### BACKGROUND

[0003] Adverse reactions to drugs are a major cause of morbidity and death. Frequently occurring adverse drug reactions include thiopurine-induced pancreatitis (TIP). The thiopurine class of drugs (mercaptopurine and azathioprine) are widely used agents to induce and maintain remission in patients with inflammatory bowel disease (Crohn's disease or ulcerative colitis). Four percent of patients treated with these agents develop pancreatitis due to drug administration.

[0004] Common drugs that have been associated with TIP include mercaptopurine and azathioprine and 6-thioguanine.

[0005] There is a need for markers that can predict the existence of or predisposition to TIP. Several studies have identified genetic risk factors for drug-related severe adverse events. However, there is currently no clinically useful method for predicting what drugs will cause TIP and in which patients.

### SUMMARY

[0006] An aspect of the invention provides a method for predicting the risk of a patient for developing adverse drug reactions, particularly thiopurine-induced pancreatitis (TIP).

[0007] TIP may be caused by the thiopurine class of drugs such as mercaptopurine and azathioprine.

[0008] Another aspect of the invention provides a method of identifying a subject afflicted with, or at risk of, developing TIP comprising (a) obtaining a nucleic acid-containing sample from the subject; and (b) analyzing the sample to detect the presence of at least one genetic marker, wherein the presence of the at least one genetic marker indicates that the subject is afflicted with, or at risk of, developing TIP. The method may further comprise treating the subject based on the results of step (b). The method may further comprise taking a clinical history from the subject. Genetic markers that are useful for the invention include, but are not limited to, alleles, microsatellites, SNPs, and haplotypes. The sample may be any sample capable of being obtained from a subject, including but not limited to blood, sputum, saliva, mucosal scraping and tissue biopsy samples.

[0009] In some embodiments of the invention, the genetic markers are SNPs selected from those listed in Table 1. In other embodiments, genetic markers that are linked to each of the SNPs can be used to predict the corresponding TIP risk.

[0010] The presence of the genetic marker can be detected using any method known in the art. Analysis may comprise nucleic acid amplification, such as PCR. Analysis may also comprise primer extension, restriction digestion, sequencing,

hybridization, a DNase protection assay, mass spectrometry, labeling, and separation analysis.

[0011] Other features and advantages of the disclosure will be apparent from the detailed description, drawings and from the claims.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0012] FIG. 1a shows a Manhattan plot depicting association p values at each tested SNP for development of pancreatitis upon administration of the thiopurine class of drugs.

[0013] FIG. 1b shows a quantile-quantile plot (or a "Q-Q plot") demonstrating association statistic deviation from random effect for development of pancreatitis upon administration of the thiopurine class of drugs at each tested SNP.

### DETAILED DESCRIPTION

[0014] For the purposes of promoting an understanding of the principles of the invention, reference will now be made to specific embodiments and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the invention is thereby intended, and that such alterations and further modifications of the invention, and such further applications of the principles of the invention as illustrated herein as would normally occur to one skilled in the art to which the invention relates, are contemplated as within the scope of the invention.

[0015] All terms as used herein are defined according to the ordinary meanings they have acquired in the art. Such definitions can be found in any technical dictionary or reference known to the skilled artisan, such as the *McGraw-Hill Dictionary of Scientific and Technical Terms* (McGraw-Hill, Inc.), *Molecular Cloning: A Laboratory Manual* (Cold Springs Harbor, New York), Remington's *Pharmaceutical Sciences* (Mack Publishing, PA), and *Stedman's Medical Dictionary* (Williams and Wilkins, MD). These references, along with those references, patents, and patent applications cited herein are hereby incorporated by reference in their entirety.

[0016] The term "marker" as used herein refers to any morphological, biochemical, or nucleic acid-based phenotypic difference which reveals a DNA polymorphism. The presence of markers in a sample may be useful to determine the phenotypic status of a subject (e.g., whether an individual has or has not been afflicted with TIP), or may be predictive of a physiological outcome (e.g., whether an individual is likely to develop TIP). The markers may be differentially present in a biological sample or fluid, such as blood plasma or serum. The markers may be isolated by any method known in the art, including methods based on mass, binding characteristics, or other physicochemical characteristics. As used herein, the term "detecting" includes determining the presence, the absence, or a combination thereof, of one or more markers.

[0017] Non-limiting examples of nucleic acid-based, genetic markers include alleles, microsatellites, single nucleotide polymorphisms (SNPs), haplotypes, copy number variants (CNVs), insertions, and deletions.

[0018] The term "allele" as used herein refers to an observed class of DNA polymorphism at a genetic marker locus. Alleles may be classified based on different types of polymorphism, for example, DNA fragment size or DNA sequence. Individuals with the same observed fragment size or same sequence at a marker locus have the same genetic marker allele and thus are of the same allelic class.

**[0019]** The term “locus” as used herein refers to a genetically defined location for a collection of one or more DNA polymorphisms revealed by a morphological, biochemical or nucleic acid-bred analysis.

**[0020]** The term “genotype” as used herein refers to the allelic composition of an individual at genetic marker loci under study, and “genotyping” refers to the process of determining the genetic composition of individuals using genetic markers.

**[0021]** The term “single nucleotide polymorphism” (SNP) as used herein refers to a DNA sequence variation occurring when a single nucleotide in the genome or other shared sequence differs between members of a species or between paired chromosomes in an individual. The difference in the single nucleotide is referred to as an allele. A “haplotype” as used herein refers to a set of single SNPs on a single chromatid that are statistically associated.

**[0022]** The term “microsatellite” as used herein refers to polymorphic loci present in DNA that comprise repeating units of 1-6 base pairs in length.

**[0023]** An aspect of the invention provides a method for predicting the risk of a patient for developing adverse drug reactions, particularly TIP. As used herein, an “adverse drug reaction” is as an undesired and unintended effect of a drug. A “drug” as used herein is any compound or agent that is administered to a patient for prophylactic, diagnostic or therapeutic purposes.

**[0024]** TIP may be caused by the thiopurine class of drugs such as mercaptopurine and azathioprine.

**[0025]** Another aspect of the invention provides a method of identifying a subject afflicted with or at risk of developing TIP comprising (a) obtaining a nucleic acid-containing sample from the subject; and (b) analyzing the sample to detect the presence of at least one genetic marker, wherein the presence of the at least one genetic marker indicates that the subject is afflicted with or at risk of developing TIP. The method may further comprise treating the subject based on the results of step (b). The method may further comprise taking a clinical history from the subject. Genetic markers that are useful for the invention include, but are not limited to, alleles, microsatellites, SNPs, haplotypes, CNVs, insertions, and deletions.

**[0026]** In some embodiments of the invention, the genetic markers are one or more SNPs selected from those listed in Table 1.

**[0027]** Each person’s genetic material contains a unique SNP pattern that is made up of many different genetic variations. SNPs may serve as biological markers for pinpointing a disease on the human genome map, because they are usually located near a gene found to be associated with a certain disease. Occasionally, a SNP may actually cause a disease and, therefore, can be used to search for and isolate the disease-causing gene.

**[0028]** In accordance with the invention, at least one marker may be detected. It is to be understood, and is described herein, that one or more markers may be detected and subsequently analyzed, including several or all of the markers identified. Further, it is to be understood that the failure to detect one or more of the markers of the invention, or the detection thereof at levels or quantities that may correlate with TIP, may be useful as a means of selecting the individuals afflicted with or at risk for developing TIP, and that the same forms a contemplated aspect of the invention.

**[0029]** In addition to the SNPs listed in Table 1, genetic markers that are linked to each of the SNPs may be used to predict the corresponding TIP risk as well. The presence of equivalent genetic markers may be indicative of the presence of the allele or SNP of interest, which, in turn, is indicative of a risk for TIP. For example, equivalent markers may co-segregate or show linkage disequilibrium with the marker of interest. Equivalent markers may also be alleles or haplotypes based on combinations of SNPs.

**[0030]** The equivalent genetic marker may be any marker, including alleles, microsatellites, SNPs, and haplotypes. In some embodiments, the useful genetic markers are about 200 kb or less from the locus of interest. In other embodiments, the markers are about 100 kb, 80 kb, 60 kb, 40 kb, or 20 kb or less from the locus of interest.

**[0031]** To further increase the accuracy of risk prediction, the marker of interest and/or its equivalent marker may be determined along with the markers of accessory molecules and co-stimulatory molecules which are involved in the interaction between antigen-presenting cell and T-cell interaction. For example, the accessory and co-stimulatory molecules include cell surface molecules (e.g., CD80, CD86, CD28, CD4, CD8, T cell receptor (TCR), ICAM-1, CD11a, CD58, CD2, etc.), and inflammatory or pro-inflammatory cytokines, chemokines (e.g., TNF- $\alpha$ ), and mediators (e.g., complements, apoptosis proteins, enzymes, extracellular matrix components, etc.). Also of interest are genetic markers of drug metabolizing enzymes which are involved in the bioactivation and detoxification of drugs. Non-limiting examples of drug metabolizing enzymes include phase I enzymes (e.g., cytochrome P450 superfamily), and phase II enzymes (e.g., microsomal epoxide hydrolase, arylamine N-acetyltransferase, UDP-glucuronosyl-transferase, etc.).

**[0032]** Another aspect of the invention provides a method for pharmacogenomic profiling. Accordingly, a panel of genetic factors is determined for a given individual, and each genetic factor is associated with the predisposition for a disease or medical condition, including adverse drug reactions. In some embodiments, the panel of genetic factors may include at least one SNP selected from Table 1. The panel may include equivalent markers to the markers in Table 1. The genetic markers for accessory molecules, co-stimulatory molecules and/or drug metabolizing enzymes described above may also be included.

**[0033]** Yet another aspect of the invention provides a method of screening and/or identifying agents that can be used to treat TIP by using any of the genetic markers of the invention as a target in drug development. For example, cells expressing any of the SNPs or equivalents thereof may be contacted with putative drug agents, and the agents that bind to the SNP or equivalent are likely to inhibit the expression and/or function of the SNP. The efficacy of the candidate drug agent in treating TIP may then be further tested.

**[0034]** In some embodiments, it may be useful to amplify the target sequence before evaluating the genetic marker. Nucleic acids used as a template for amplification may be isolated from cells, tissues or other samples according to standard methodologies such as are described, for example, in Sambrook et al., 1989. In certain embodiments, analysis is performed on whole cell or tissue homogenates or biological fluid samples without substantial purification of the template nucleic acid. The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may

be desired to first convert the RNA to a complementary DNA. The DNA also may be from a cloned source or synthesized *in vitro*.

**[0035]** The term “primer,” refers to any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty or thirty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded or single-stranded form.

**[0036]** For amplification of SNPs, pairs of primers designed to selectively hybridize to nucleic acids flanking the polymorphic site may be contacted with the template nucleic acid under conditions that permit selective hybridization. Depending upon the desired application, high stringency hybridization conditions may be selected that will only allow hybridization to sequences that are completely complementary to the primers. In other embodiments, hybridization may occur under reduced stringency to allow for amplification of nucleic acids containing one or more mismatches with the primer sequences. Once hybridized, the template-primer complex may be contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as “cycles,” are conducted until a sufficient amount of amplification product is produced.

**[0037]** It is also possible that multiple target sequences will be amplified in a single reaction. Primers designed to expand specific sequences located in different regions of the target genome, thereby identifying different polymorphisms, would be mixed together in a single reaction mixture. The resulting amplification mixture would contain multiple amplified regions, and could be used as the source template for polymorphism detection using the methods described in this application.

**[0038]** Any known template dependent process may be advantageously employed to amplify the oligonucleotide sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (PCR), which is described in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis et al., 1988, each of which is incorporated herein by reference in their entirety.

**[0039]** A reverse transcriptase PCR amplification procedure may be performed when the source of nucleic acid is fractionated or whole cell RNA. Methods of reverse transcribing RNA into cDNA are well known and are described in, for example, Sambrook et al., 1989. Alternative exemplary methods for reverse polymerization utilize thermostable DNA polymerases. These methods are described, for example, in International Publication WO 90/07641. Polymerase chain reaction methodologies are well known in the art. Representative methods of RT-PCR are described, for example, in U.S. Pat. No. 5,882,864.

**[0040]** Another method for amplification is ligase chain reaction (LCR), disclosed, for example, in European Application No. 320 308, incorporated herein by reference in its entirety. U.S. Pat. No. 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence. A method based on PCR and oligonucleotide ligase assay (OLA), disclosed, for example, in U.S. Pat. No. 5,912,148, may also be used.

**[0041]** Another ligase-mediated reaction is disclosed by Guilfoyle et al. (1997). Genomic DNA is digested with a restriction enzyme and universal linkers are then ligated onto the restriction fragments. Primers to the universal linker

sequence are then used in PCR to amplify the restriction fragments. By varying the conditions of the PCR, one can specifically amplify fragments of a certain size (e.g., fewer than 1000 bases). A benefit to using this approach is that each individual region would not have to be amplified separately. There would be the potential to screen thousands of SNPs from the single PCR reaction.

**[0042]** Q-beta Replicase, described, for example, in International Application No. PCT/US87/00880, may also be used as an amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence, which may then be detected.

**[0043]** An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention (Walker et al., 1992). Strand Displacement Amplification (SDA), disclosed, for example, in U.S. Pat. No. 5,916,779, is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, e.g., nick translation.

**[0044]** Other nucleic acid amplification procedures include polymerization-based amplification systems (TAS), for example, nucleic acid sequence based amplification (NASBA) and 3SR (Kwoh et al., 1989; International Application WO 88/10315, incorporated herein by reference in their entirety). European Application No. 329 822 discloses a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA (ssRNA), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention.

**[0045]** International Application WO 89/06700 discloses a nucleic acid sequence amplification scheme based on the hybridization of a promoter region/primer sequence to a target single-stranded DNA (ssDNA) followed by polymerization of many RNA copies of the sequence. This scheme is not cyclic, i.e., new templates are not produced from the resultant RNA transcripts. Other amplification methods include “race” and “one-sided PCR” (Frohman, 1990; Ohara et al., 1989).

#### Methods of Detection

**[0046]** The genetic markers of the invention may be detected using any method known in the art. For example, genomic DNA may be hybridized to a probe that is specific for the allele of interest. The probe may be labeled for direct detection, or contacted by a second, detectable molecule that specifically binds to the probe. Alternatively, cDNA, RNA, or the protein product of the allele may be detected. For example, serotyping or microcytotoxicity methods may be used to determine the protein product of the allele. Similarly, equivalent genetic markers may be detected by any methods known in the art.

**[0047]** It is within the purview of one of skill in the art to design genetic tests to screen for TIP or a predisposition for TIP based on analysis of the genetic markers of the invention. For example, a genetic test may be based on the analysis of DNA for SNP patterns. Samples may be collected from a group of individuals affected by TIP due to drug treatment and the DNA analyzed for SNP patterns. Non-limiting examples of sample sources include blood, sputum, saliva, mucosal scraping or tissue biopsy samples. These SNP pat-

terns may then be compared to patterns obtained by analyzing the DNA from a group of individuals unaffected by TIP due to drug treatment. This type of comparison, called an “association study,” can detect differences between the SNP patterns of the two groups, thereby indicating which pattern is most likely associated with TIP. Eventually, SNP profiles that are characteristic of a variety of diseases will be established. These profiles can then be applied to the population at general, or those deemed to be at particular risk of developing TIP.

**[0048]** Various techniques may be used to assess genetic markers. Non-limiting examples of a few of these techniques are discussed here and also described in US Patent Publication 2007/026827, the disclosure of which is herein incorporated by reference in its entirety. In accordance with the invention, any of these methods may be used to design genetic tests for affliction with or predisposition to TIP. Additionally, these methods are continually being improved and new methods are being developed. It is contemplated that one of skill in the art will be able to use any improved or new methods, in addition to any existing method, for detecting and analyzing the genetic markers of the invention.

**[0049]** Restriction Fragment Length Polymorphism (RFLP) is a technique in which different DNA sequences may be differentiated by analysis of patterns derived from cleavage of that DNA. If two sequences differ in the distance between sites of cleavage of a particular restriction endonuclease, the length of the fragments produced will differ when the DNA is digested with a restriction enzyme. The similarity of the patterns generated can be used to differentiate species (and even individual species members) from one another.

**[0050]** Restriction endonucleases are the enzymes that cleave DNA molecules at specific nucleotide sequences depending on the particular enzyme used. Enzyme recognition sites are usually 4 to 6 base pairs in length. Generally, the shorter the recognition sequence, the greater the number of fragments generated. If molecules differ in nucleotide sequence, fragments of different sizes may be generated. The fragments can be separated by gel electrophoresis. Restriction enzymes are isolated from a wide variety of bacterial genera and are thought to be part of the cell’s defenses against invading bacterial viruses. Use of RFLP and restriction endonucleases in genetic marker analysis, such as SNP analysis, requires that the SNP affect cleavage of at least one restriction enzyme site.

**[0051]** Primer Extension is a technique in which the primer and no more than three NTPs may be combined with a polymerase and the target sequence, which serves as a template for amplification. By using fewer than all four NTPs, it is possible to omit one or more of the polymorphic nucleotides needed for incorporation at the polymorphic site. The amplification may be designed such that the omitted nucleotide(s) is(are) not required between the 3’ end of the primer and the target polymorphism. The primer is then extended by a nucleic acid polymerase, such as Taq polymerase. If the omitted NTP is required at the polymorphic site, the primer is extended up to the polymorphic site, at which point the polymerization ceases. However, if the omitted NTP is not required at the polymorphic site, the primer will be extended beyond the polymorphic site, creating a longer product. Detection of the extension products is based on, for example, separation by size/length which will thereby reveal which polymorphism is present.

**[0052]** Oligonucleotide Hybridization is a technique in which oligonucleotides may be designed to hybridize directly to a target site of interest. The hybridization can be performed on any useful format. For example, oligonucleotides may be arrayed on a chip or plate in a microarray. Microarrays comprise a plurality of oligos spatially distributed over, and stably associated with, the surface of a substantially planar substrate, e.g., a biochip. Microarrays of oligonucleotides have been developed and find use in a variety of applications, such as screening and DNA sequencing.

**[0053]** In gene analysis with microarrays, an array of “probe” oligonucleotides is contacted with a nucleic acid sample of interest, i.e., a target. Contact is carried out under hybridization conditions and unbound nucleic acid is then removed. The resultant pattern of hybridized nucleic acid provides information regarding the genetic profile of the sample tested. Methodologies of gene analysis on microarrays are capable of providing both qualitative and quantitative information.

**[0054]** A variety of different arrays which may be used is known in the art. The probe molecules of the arrays which are capable of sequence-specific hybridization with target nucleic acid may be polynucleotides or hybridizing analogues or mimetics thereof, including: nucleic acids in which the phosphodiester linkage has been replaced with a substitute linkage, such as phosphorothioate, methylimino, methylphosphonate, phosphoramidate, guanidine and the like; and nucleic acids in which the ribose subunit has been substituted, e.g., hexose phosphodiester, peptide nucleic acids, and the like. The length of the probes will generally range from 10 to 1000 nts, wherein in some embodiments the probes will be oligonucleotides and usually range from 15 to 150 nts and more usually from 15 to 100 nts in length, and in other embodiments the probes will be longer, usually ranging in length from 150 to 1000 nts, where the polynucleotide probes may be single- or double-stranded, usually single-stranded, and may be PCR fragments amplified from cDNA.

**[0055]** Probe molecules arrayed on the surface of a substrate may correspond to selected genes being analyzed and be positioned on the array at a known location so that positive hybridization events may be correlated to expression of a particular gene in the physiological source from which the target nucleic acid sample is derived. The substrate with which the probe molecules are stably associated may be fabricated from a variety of materials, including plastics, ceramics, metals, gels, membranes, glasses, and the like. The arrays may be produced according to any convenient methodology, such as preforming the probes and then stably associating them with the surface of the support or growing the probes directly on the support. Different array configurations and methods for their production and use are known to those of skill in the art and disclosed, for example, in U.S. Pat. Nos. 5,445,934, 5,532,128, 5,556,752, 5,242,974, 5,384,261, 5,405,783, 5,412,087, 5,424,186, 5,429,807, 5,436,327, 5,472,672, 5,527,681, 5,529,756, 5,545,531, 5,554,501, 5,561,071, 5,571,639, 5,593,839, 5,599,695, 5,624,711, 5,658,734, 5,700,637, and 6,004,755, the disclosures of which are herein incorporated by reference in their entireties.

**[0056]** Following hybridization, where non-hybridized labeled nucleic acid is capable of emitting a signal during the detection step, a washing step is employed in which unhybridized labeled nucleic acid is removed from the support surface, generating a pattern of hybridized nucleic acid on the



substrate surface. Various wash solutions and protocols for their use are known to those of skill in the art and may be used.

**[0057]** Where the label on the target nucleic acid is not directly detectable, the array comprising bound target may be contacted with the other member(s) of the signal producing system that is being employed. For example, where the target is biotinylated, the array may be contacted with streptavidin-fluorescer conjugate under conditions sufficient for binding between the specific binding member pairs to occur. Following contact, any unbound members of the signal producing system will then be removed, e.g., by washing. The specific wash conditions employed will depend on the specific nature of the signal producing system that is employed, as will be known to those of skill in the art familiar with the particular signal producing system employed.

**[0058]** The resultant hybridization pattern(s) of labeled nucleic acids may be visualized or detected in a variety of ways, with the particular manner of detection being chosen based on the particular label of the nucleic acid, where representative detection means include scintillation counting, autoradiography, fluorescence measurement, calorimetric measurement, light emission measurement and the like.

**[0059]** Prior to detection or visualization, the potential for a mismatch hybridization event that could potentially generate a false positive signal on the pattern may be reduced by treating the array of hybridized target/probe complexes with an endonuclease under conditions sufficient such that the endonuclease degrades single stranded, but not double stranded, DNA. Various different endonucleases are known and may be used, including but not limited to mung bean nuclease, S1 nuclease, and the like. Where such treatment is employed in an assay in which the target nucleic acids are not labeled with a directly detectable label, e.g., in an assay with biotinylated target nucleic acids, the endonuclease treatment will generally be performed prior to contact of the array with the other member(s) of the signal producing system, e.g., fluorescent-streptavidin conjugate. Endonuclease treatment, as described above, ensures that only end-labeled target/probe complexes having a substantially complete hybridization at the 3' end of the probe are detected in the hybridization pattern.

**[0060]** Following hybridization and any washing step(s) and/or subsequent treatments, as described herein, the resultant hybridization pattern may be detected. In detecting or visualizing the hybridization pattern, the intensity or signal value of the label may also be quantified, such that the signal from each spot of the hybridization will be measured and compared to a unit value corresponding the signal emitted by known number of labeled target nucleic acids to obtain a count or absolute value of the copy number of each end-labeled target that is hybridized to a particular spot on the array in the hybridization pattern.

**[0061]** It will be appreciated that any useful system for detecting nucleic acids may be used in accordance with the invention. For example, mass spectrometry, hybridization, sequencing, labeling, and separation analysis may be used individually or in combination, and may also be used in combination with other known methods of detecting nucleic acids.

**[0062]** Electrospray ionization (ESI) is a type of mass spectrometry that is used to produce gaseous ions from highly polar, mostly nonvolatile biomolecules, including lipids. The sample is typically injected as a liquid at low flow rates (1-10  $\mu\text{L}/\text{min}$ ) through a capillary tube to which a strong electric

field is applied. The field charges the liquid in the capillary and produces a fine spray of highly charged droplets that are electrostatically attracted to the mass spectrometer inlet. The evaporation of the solvent from the surface of a droplet as it travels through the desolvation chamber increases its charge density substantially. When this increase exceeds the Rayleigh stability limit, ions are ejected and ready for MS analysis.

**[0063]** A typical conventional ESI source consists of a metal capillary of typically 0.1-0.3 mm in diameter, with a tip held approximately 0.5 to 5 cm (but more usually 1 to 3 cm) away from an electrically grounded circular interface having at its center the sampling orifice. A potential difference of between 1 to 5 kV (but more typically 2 to 3 kV) is applied to the capillary by power supply to generate a high electrostatic field ( $10^6$  to  $10^7$  V/m) at the capillary tip. A sample liquid, carrying the analyte to be analyzed by the mass spectrometer, is delivered to the tip through an internal passage from a suitable source (such as from a chromatograph or directly from a sample solution via a liquid flow controller). By applying pressure to the sample in the capillary, the liquid leaves the capillary tip as small highly electrically charged droplets and further undergoes desolvation and breakdown to form single or multi-charged gas phase ions in the form of an ion beam. The ions are then collected by the grounded (or oppositely-charged) interface plate and led through an the orifice into an analyzer of the mass spectrometer. During this operation, the voltage applied to the capillary is held constant. Aspects of construction of ESI sources are described, for example, in U.S. Pat. Nos. 5,838,002; 5,788,166; 5,757,994; RE 35,413; and 5,986,258.

**[0064]** In ESI tandem mass spectroscopy (ESI/MS/MS), one is able to simultaneously analyze both precursor ions and product ions, thereby monitoring a single precursor product reaction and producing (through selective reaction monitoring (SRM)) a signal only when the desired precursor ion is present. When the internal standard is a stable isotope-labeled version of the analyte, this is known as quantification by the stable isotope dilution method. This approach has been used to accurately measure pharmaceuticals and bioactive peptides.

**[0065]** Secondary ion mass spectroscopy (SIMS) is an analytical method that uses ionized particles emitted from a surface for mass spectroscopy at a sensitivity of detection of a few parts per billion. The sample surface is bombarded by primary energetic particles, such as electrons, ions (e.g., O, Cs), neutrals or photons, forcing atomic and molecular particles to be ejected from the surface, a process called sputtering. Since some of these sputtered particles carry a charge, a mass spectrometer can be used to measure their mass and charge. Continued sputtering permits measuring of the exposed elements as material is removed. This in turn permits one to construct elemental depth profiles. Although the majority of secondary ionized particles are electrons, it is the secondary ions which are detected and analyzed by the mass spectrometer in this method.

**[0066]** Laser desorption mass spectroscopy (LD-MS) involves the use of a pulsed laser, which induces desorption of sample material from a sample site, and effectively, vaporizes sample off of the sample substrate. This method is usually used in conjunction with a mass spectrometer, and can be performed simultaneously with ionization by adjusting the laser radiation wavelength.

**[0067]** When coupled with Time-of-Flight (TOF) measurement, LD-MS is referred to as LDLPMS (Laser Desorption Laser Photoionization Mass Spectroscopy). The LDLPMS method of analysis gives instantaneous volatilization of the sample, and this form of sample fragmentation permits rapid analysis without any wet extraction chemistry. The LDLPMS instrumentation provides a profile of the species present while the retention time is low and the sample size is small. In LDLPMS, an impactor strip is loaded into a vacuum chamber. The pulsed laser is fired upon a certain spot of the sample site, and species present are desorbed and ionized by the laser radiation. This ionization also causes the molecules to break up into smaller fragment-ions. The positive or negative ions made are then accelerated into the flight tube, being detected at the end by a microchannel plate detector. Signal intensity, or peak height, is measured as a function of travel time. The applied voltage and charge of the particular ion determines the kinetic energy, and separation of fragments is due to their different sizes causing different velocities. Each ion mass will thus have a different flight-time to the detector.

**[0068]** Other advantages of the LDLPMS method include the possibility of constructing the system to give a quiet baseline of the spectra because one can prevent coevolved neutrals from entering the flight tube by operating the instrument in a linear mode. Also, in environmental analysis, the salts in the air and as deposits will not interfere with the laser desorption and ionization. This instrumentation also is very sensitive and robust, and has been shown to be capable of detecting trace levels in natural samples without any prior extraction preparations.

**[0069]** Matrix Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) is a type of mass spectrometry useful for analyzing molecules across an extensive mass range with high sensitivity, minimal sample preparation and rapid analysis times. MALDI-TOF also enables non-volatile and thermally labile molecules to be analyzed with relative ease. One important application of MALDI-TOF is in the area of quantification of peptides and proteins, such as in biological tissues and fluids.

**[0070]** Surface Enhanced Laser Desorption and Ionization (SELDI) is another type of desorption/ionization gas phase ion spectrometry in which an analyte is captured on the surface of a SELDI mass spectrometry probe. There are several known versions of SELDI.

**[0071]** One version of SELDI is affinity capture mass spectrometry, also called Surface-Enhanced Affinity Capture (SEAC). This version involves the use of probes that have a material on the probe surface that captures analytes through a non-covalent affinity interaction (adsorption) between the material and the analyte. The material is variously called an "adsorbent," a "capture reagent," an "affinity reagent" or a "binding moiety." The capture reagent may be any material capable of binding an analyte. The capture reagent may be attached directly to the substrate of the selective surface, or the substrate may have a reactive surface that carries a reactive moiety that is capable of binding the capture reagent, e.g., through a reaction forming a covalent or coordinate covalent bond. Epoxide and carbodiimidazole are useful reactive moieties to covalently bind polypeptide capture reagents such as antibodies or cellular receptors. Nitriloacetic acid and iminodiacetic acid are useful reactive moieties that function as chelating agents to bind metal ions that interact non-co-

valently with histidine containing peptides. Adsorbents are generally classified as chromatographic adsorbents and bio-specific adsorbents.

**[0072]** Another version of SELDI is Surface-Enhanced Neat Desorption (SEND), which involves the use of probes comprising energy absorbing molecules that are chemically bound to the probe surface. Energy absorbing molecules (EAM) refer to molecules that are capable of absorbing energy from a laser desorption/ionization source and, thereafter, of contributing to desorption and ionization of analyte molecules in contact therewith. The EAM category includes molecules used in MALDI, frequently referred to as "matrix," and is exemplified by cinnamic acid derivatives such as sinapinic acid (SPA), cyano-hydroxy-cinnamic acid (CHCA) and dihydroxybenzoic acid, ferulic acid, and hydroxyaceto-phenone derivatives. In certain versions, the energy absorbing molecule is incorporated into a linear or cross-linked polymer, e.g., a polymethacrylate. For example, the composition may be a co-polymer of  $\alpha$ -cyano-4-methacryloyloxy-cinnamic acid and acrylate. In another version, the composition may be a co-polymer of  $\alpha$ -cyano-4-methacryloyloxy-cinnamic acid, acrylate and 3-(tri-ethoxy)silyl propyl methacrylate. In another version, the composition may be a co-polymer of  $\alpha$ -cyano-4-methacryloyloxy-cinnamic acid and octadecyl-methacrylate ("C18 SEND").

**[0073]** SEAC/SEND is a version of SELDI in which both a capture reagent and an energy absorbing molecule are attached to the sample presenting surface. SEAC/SEND probes therefore allow the capture of analytes through affinity capture and ionization/desorption without the need to apply external matrix.

**[0074]** Another version of SELDI, called Surface-Enhanced Photolabile Attachment and Release (SEPAR), involves the use of probes having moieties attached to the surface that can covalently bind an analyte, and then release the analyte through breaking a photolabile bond in the moiety after exposure to light, e.g., to laser light. SEPAR and other forms of SELDI are readily adapted to detecting a marker or marker profile, in accordance with the present invention.

**[0075]** In accordance with the invention, nucleic acid hybridization is another useful method of analyzing genetic markers. Nucleic acid hybridization is generally understood as the ability of a nucleic acid to selectively form duplex molecules with complementary stretches of DNAs and/or RNAs. Depending on the application, varying conditions of hybridization may be used to achieve varying degrees of selectivity of the probe or primers for the target sequence.

**[0076]** Typically, a probe or primer of between 10 and 100 nucleotides, and up to 1-2 kilobases or more in length, will allow the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over contiguous stretches greater than 20 bases in length may be used to increase stability and selectivity of the hybrid molecules obtained. Nucleic acid molecules for hybridization may be readily prepared, for example, by directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

**[0077]** For applications requiring high selectivity, relatively high stringency conditions may be used to form the hybrids. For example, relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50° C. to about 70° C. Such high stringency conditions tolerate little, if any, mis-

match between the probe or primers and the template or target strand and would be particularly suitable for isolating specific genes or for detecting specific mRNA transcripts. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide.

**[0078]** For certain applications, lower stringency conditions may be used. Under these conditions, hybridization may occur even though the sequences of the hybridizing strands are not perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and/or decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37° C. to about 55° C., while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20° C. to about 55° C. Hybridization conditions can be readily manipulated by those of skill depending on the desired results.

**[0079]** It is within the purview of the skilled artisan to design and select the appropriate primers, probes, and enzymes for any of the methods of genetic marker analysis. For example, for detection of SNPs, the skilled artisan will generally use agents that are capable of detecting single nucleotide changes in DNA. These agents may hybridize to target sequences that contain the change. Or, these agents may hybridize to target sequences that are adjacent to (e.g., upstream or 5' to) the region of change.

**[0080]** In general, it is envisioned that the probes or primers described herein will be useful as reagents in solution hybridization for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The conditions selected will depend on the particular circumstances (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Optimization of hybridization conditions for the particular application of interest, as described herein, is well known to those of skill in the art. After washing of the hybridized molecules to remove non-specifically bound probe molecules, hybridization is detected, and/or quantified, by determining the amount of bound label. Representative solid phase hybridization methods are disclosed in U.S. Pat. Nos. 5,843,663, 5,900,481 and 5,919,626. Other methods of hybridization that may be used in the practice of the present invention are disclosed in U.S. Pat. Nos. 5,849,481, 5,849,486 and 5,851,772. The relevant portions of these and other references identified in this section are incorporated herein by reference.

**[0081]** The synthesis of oligonucleotides for use as primers and probes is well known to those of skill in the art. Chemical synthesis can be achieved, for example, by the diester method, the triester method, the polynucleotide phosphorylase method and by solid-phase chemistry. Various mechanisms of oligonucleotide synthesis have been disclosed, for example, in U.S. Pat. Nos. 4,659,774, 4,816,571, 5,141,813, 5,264,566, 4,959,463, 5,428,148, 5,554,744, 5,574,146, and 5,602,244, each of which is incorporated herein by reference in its entirety.

**[0082]** In certain embodiments, nucleic acid products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods such as those

described, for example, in Sambrook et al., 1989. Separated products may be cut out and eluted from the gel for further manipulation. Using low melting point agarose gels, the skilled artisan may remove the separated band by heating the gel, followed by extraction of the nucleic acid.

**[0083]** Separation of nucleic acids may also be effected by chromatographic techniques known in the art. There are many kinds of chromatography that may be used in the practice of the present invention, non-limiting examples of which include capillary adsorption, partition, ion-exchange, hydroxylapatite, molecular sieve, reverse-phase, column, paper, thin-layer, and gas chromatography, as well as HPLC.

**[0084]** A number of the above separation platforms may be coupled to achieve separations based on two different properties. For example, some of the primers may be coupled with a moiety that allows affinity capture, and some primers remain unmodified. Modifications may include a sugar (for binding to a lectin column), a hydrophobic group (for binding to a reverse-phase column), biotin (for binding to a streptavidin column), or an antigen (for binding to an antibody column). Samples may be run through an affinity chromatography column. The flow-through fraction is collected, and the bound fraction eluted (by chemical cleavage, salt elution, etc.). Each sample may then be further fractionated based on a property, such as mass, to identify individual components.

**[0085]** In certain aspects, it will be advantageous to employ nucleic acids of defined sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. Various appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In the case of enzyme tags, colorimetric indicator substrates are known that may be employed to provide a detection means that is visibly or spectrophotometrically detectable, to identify specific hybridization with complementary nucleic acid containing samples. In yet other embodiments, the primer has a mass label that can be used to detect the molecule amplified. Other embodiments also contemplate the use of Taqman™ and Molecular Beacon™ probes.

**[0086]** Radioactive isotopes useful for the invention include, but are not limited to, tritium, <sup>14</sup>C and <sup>32</sup>P. Among the fluorescent labels contemplated for use as conjugates include Alexa 350, Alexa 430, AMCA, BODIPY 630/650, BODIPY 650/665, BODIPY-FL, BODIPY-R6G, BODIPY-TMR, BODIPY-TRX, Cascade Blue, Cy3, Cy5,6-FAM, Fluorescein Isothiocyanate, HEX, 6-JOE, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, REG, Rhodamine Green, Rhodamine Red, Renographin, ROX, TAMRA, TET, Tetramethylrhodamine, and/or Texas Red.

**[0087]** The choice of label may vary, depending on the method used for analysis. When using capillary electrophoresis, microfluidic electrophoresis, HPLC, or LC separations, either incorporated or intercalated fluorescent dyes may be used to label and detect the amplification products. Samples are detected dynamically, in that fluorescence is quantitated as a labeled species moves past the detector. If an electrophoretic method, HPLC, or LC is used for separation, products can be detected by absorption of UV light. If polyacrylamide gel or slab gel electrophoresis is used, the primer for the extension reaction can be labeled with a fluorophore, a chromophore or a radioisotope, or by associated enzymatic reaction. Alternatively, if polyacrylamide gel or slab gel electrophoresis is used, one or more of the NTPs in the extension

reaction can be labeled with a fluorophore, a chromophore or a radioisotope, or by associated enzymatic reaction. Enzymatic detection involves binding an enzyme to a nucleic acid, e.g., via a biotin:avidin interaction, following separation of the amplification products on a gel, then detection by chemical reaction, such as chemiluminescence generated with luminol. A fluorescent signal may be monitored dynamically. Detection with a radioisotope or enzymatic reaction may require an initial separation by gel electrophoresis, followed by transfer of DNA molecules to a solid support (blot) prior to analysis. If blots are made, they can be analyzed more than once by probing, stripping the blot, and then reprobing. If the extension products are separated using a mass spectrometer, no label is required because nucleic acids are detected directly.

**[0088]** While whole genome association (WGA) studies allow examination of many common SNPs in different individuals to identify associations between SNPs and traits like major diseases, exome sequencing studies can increase efficiency by allowing selective sequencing of at least the coding regions (i.e., the exons that are translated into proteins) of the genome, in which most functional variation is thought to occur. Some benefits of exome sequencing can include the detection of traits without traditional genetic linkage, with fewer available case studies (e.g., rare Mendelian diseases), with causal variants in different genes (i.e., genetic heterogeneity), and with diverse clinical features (i.e., phenotypic heterogeneity). The exome constitutes only about 1% of the entire human genome, and a large number of rare mutations have weak or no effects in non-coding sequences.

**[0089]** Target-enrichment methods like direct genomic selection (DGS) allow selective capture of genomic regions of interest from a DNA sample prior to sequencing. Other target-enrichment methods can include, but are not limited to, at least one of polymerase chain reaction (PCR) to amplify target-specific DNA sequences; molecular inversion probes of single-stranded DNA oligonucleotides that undergo an enzymatic reaction with target-specific DNA sequences to form circular DNA fragments; hybrid capture microarrays that contain fixed, tiled single-stranded DNA oligonucleotides with target-specific DNA sequences to hybridize sheared double-stranded fragments of genomic DNA; in-solution capture with single-stranded DNA oligonucleotides with target-specific DNA sequences synthesized in solution to hybridize sheared double-stranded fragments of genomic DNA in the solution; and methods using sequencing platforms, such as Sanger sequencing, 454™ sequencing (available from Roche Diagnostics Corp. (Branford, Conn.)), the Genome Analyzer™ (available from Illumina, Inc. (San Diego, Calif.)), and SOLiD® and Ion Torrent™ technologies (available from Life Technologies Corp. (Carlsbad, Calif.)).

**[0090]** Other methods of nucleic acid detection that may be used in the practice of the instant invention are disclosed in U.S. Pat. Nos. 5,840,873, 5,843,640, 5,843,651, 5,846,708, 5,846,717, 5,846,726, 5,846,729, 5,849,487, 5,853,990, 5,853,992, 5,853,993, 5,856,092, 5,861,244, 5,863,732, 5,863,753, 5,866,331, 5,905,024, 5,910,407, 5,912,124, 5,912,145, 5,919,630, 5,925,517, 5,928,862, 5,928,869, 5,929,227, 5,932,413 and 5,935,791, each of which is incorporated herein by reference in its entirety.

**[0091]** While the foregoing specification teaches the principles of the invention, with examples provided for the purpose of illustration, it will be appreciated by one skilled in the art from reading this disclosure that various changes in form and detail can be made without departing from the true scope of the invention.

## Examples

### Whole-Genome Association Study

**[0092]** A whole-genome association (WGA) study was undertaken in which the case group comprised 172 cases. Patients were identified who developed a sudden onset of severe abdominal pain, a greater than or equal to two fold rise in amylase or lipase serum level and a physician assessment that determined the cause of the pancreatitis was thiopurine administration such that the agent was stopped. An expert adjudication panel assessed causality for all cases to exclude confounders and other possible causes of pancreatitis. Standardized phenotypic definitions for TIP are described in Heap, G A et al. *Nat Genet.* 2014 October; 46(10):1131-4. doi: 10.1038/ng.3093. Epub 2014 Sep. 14, the contents of which are incorporated by reference.

**[0093]** The control group comprised 2035 samples that match the cases for disease, sex, and race.

**[0094]** Case genotyping was performed using the Illumina HumaCoreExome, with additional SNP information derived from imputation of SNP genotypes using standard bioinformatics approaches as described in the Heap, G A et al. reference described above.

**[0095]** Principle component analysis (PCA) was done on all TIP cases and controls to detect population structure. Standard quality control procedures were applied to the case-control genotype data set (based on SNP call rates, Hardy-Weinberg Equilibrium, and minor allele frequency) to exclude from downstream analysis low quality SNPs that could generate potentially false positive associations. Genetically-matched controls were selected for each case group, resulting in 172 cases. The results were then replicated in a further cohort of 78 cases and 472 additional control samples matched for sex, disease status and drug exposure.

**[0096]** Associations were tested using Fisher's exact test under additive, dominant, and recessive models through PLINK. The cohorts analyzed against the 2035 controls in the WGA study were thiopurine induced pancreatitis only.

**[0097]** FIG. 1a is a Manhattan plot that depicts association p values at each tested SNP for development of pancreatitis upon administration of the thiopurine class of drugs. Line x is  $P=5 \times 10^{-8}$  and line y is  $P=1 \times 10^{-5}$ . FIG. 1b is a quantile-quantile plot (or a "Q-Q plot") that shows the association statistic deviation from random effect for development of pancreatitis upon administration of the thiopurine class of drugs at each tested SNP.

**[0098]** Table 1 shows the SNPs found to be the most strongly associated with TIP and have a p-value smaller than  $10^{-5}$  in the data set.



TABLE 1-continued

SNP Name	Allele 1	Allele 2	Frequency	Odds Ratio	Standard Error	p-value
AA_DRB1_13_32660109_YG	P	A	0.2289	2.1236	0.1166	1.04E-10
rs3104406	A	G	0.408	2.2141	0.1254	2.35E-10
rs2858331	C	T	0.4151	2.085	0.1162	2.60E-10
SNP_DRB1_32659976	G	A	0.3857	1.9785	0.1114	8.92E-10
rs2647015	C	A	0.1271	2.2608	0.1338	1.10E-09
rs2858308	A	C	0.1285	2.2522	0.1339	1.33E-09
rs2856705	A	G	0.1285	2.2522	0.1339	1.33E-09
AA_DRB1_57_32659977_DA	P	A	0.7525	0.502	0.1148	1.95E-09
SNP_DRB1_32659977_TG	P	A	0.7525	0.5021	0.1148	1.95E-09
SNP_DQA1_32718414	T	C	0.4122	1.9683	0.1132	2.20E-09
SNP_DRB1_32659926_CG	P	A	0.7423	0.5093	0.1159	5.77E-09
AA_DRB1_74_32659926_RA	P	A	0.7423	0.5093	0.1159	5.77E-09
AA_DRB1_9_32660121_W	P	A	0.4432	1.9651	0.116	5.80E-09
SNP_DRB1_32660121	C	T	0.4432	1.9651	0.116	5.80E-09
SNP_DRB1_32660122_A	P	A	0.4432	1.9651	0.116	5.80E-09
AA_DRB1_140_32657458_T	P	A	0.5448	0.5073	0.1166	5.93E-09
AA_DRB1_11_32660115_SVD	P	A	0.557	0.5094	0.116	6.15E-09
AA_DRB1_11_32660115_SV	P	A	0.5452	0.5092	0.1166	6.99E-09
AA_DRB1_9_32660121_E	P	A	0.5452	0.5092	0.1166	6.99E-09
SNP_DRB1_32660122_C	P	A	0.5452	0.5092	0.1166	6.99E-09
AA_DRB1_67_32659947_L	P	A	0.4116	0.475	0.1286	7.02E-09
SNP_DRB1_32659948_G	P	A	0.4116	0.475	0.1286	7.05E-09
AA_DRB1_140_32657458_A	P	A	0.4548	1.9623	0.1164	7.08E-09
SNP_DRB1_32657459	C	T	0.4548	1.9623	0.1164	7.08E-09
AA_DRB1_67_32659947_I	P	A	0.4772	1.9536	0.117	1.05E-08
SNP_DRB1_32659948_T	P	A	0.4772	1.9535	0.117	1.05E-08
AA_DRB1_13_32660109_SHG	P	A	0.5401	0.5124	0.1169	1.06E-08
rs3097648	T	A	0.1202	2.2889	0.1463	1.52E-08
rs6913309	A	T	0.3179	1.9166	0.1154	1.73E-08
AA_DRB1_57_32659977_D	P	A	0.7294	0.5241	0.1148	1.84E-08
SNP_DRB1_32659977_T	P	A	0.7294	0.5241	0.1148	1.85E-08
AA_DRB1_37_32660037_SF	P	A	0.4664	1.9273	0.1174	2.27E-08
SNP_DRB1_32660037_T	P	A	0.5167	0.5233	0.1178	3.90E-08
AA_DRB1_37_32660037_NY	P	A	0.5167	0.5234	0.1178	3.91E-08
rs401775	G	A	0.2114	1.9347	0.1218	6.05E-08
rs3104404	A	C	0.1848	2.0151	0.1297	6.65E-08
rs28732201	A	G	0.0571	2.6395	0.1801	7.10E-08
AA_DRB1_11_32660115_GL	P	A	0.305	1.8488	0.1147	8.45E-08
AA_DRB1_30_32660058_IC	P	A	0.3051	1.8458	0.1147	9.14E-08
rs396960	A	T	0.2628	1.8692	0.1179	1.13E-07
AA_DRB1_11_32660115_SPV	P	A	0.6833	0.545	0.1146	1.18E-07
AA_DRB1_30_32660058_YHR	P	A	0.6833	0.5451	0.1146	1.19E-07
SNP_DRB1_32660063_T	P	A	0.31	1.819	0.1144	1.68E-07
HLA_DQB1_0202	P	A	0.1167	2.0825	0.1402	1.68E-07
AA_DRB1_30_32660058_YHG	P	A	0.6902	0.55	0.1143	1.70E-07
AA_DRB1_13_32660109_SFG	P	A	0.5162	0.5475	0.1156	1.88E-07
SNP_DRB1_32657340	A	G	0.3216	1.8142	0.1143	1.88E-07
AA_DQB1_135_32737883_G	P	A	0.1167	2.0772	0.1403	1.90E-07
SNP_DQB1_32737883	C	T	0.1167	2.077	0.1403	1.91E-07
AA_DRB1_13_32660109_YF	P	A	0.322	1.813	0.1144	1.96E-07
SNP_DRB1_32660058_T	P	A	0.6783	0.5527	0.1143	2.13E-07
AA_DRB1_30_32660058_YH	P	A	0.6783	0.5527	0.1143	2.14E-07
AA_DRB1_30_32660058_YGR	P	A	0.6777	0.5551	0.1138	2.31E-07
rs12153855	C	T	0.1295	2.0612	0.1404	2.60E-07
AA_DRB1_11_32660115_SLD	P	A	0.5112	0.5517	0.1158	2.79E-07
AA_DRB1_71_32659935_KE	P	A	0.3899	0.5102	0.131	2.81E-07
SNP_DRB1_32659935_T	P	A	0.3899	0.5102	0.131	2.81E-07
AA_DRB1_13_32660109_SH	P	A	0.4855	0.544	0.1188	2.96E-07
AA_DRB1_30_32660058_YR	P	A	0.6658	0.5589	0.1136	3.01E-07
SNP_DRB1_32659937	G	C	0.4441	1.8004	0.1149	3.10E-07
SNP_DRB1_32659939	C	G	0.4441	1.8003	0.1149	3.11E-07
AA_DRB1_70_32659938_D	P	A	0.4441	1.8002	0.1149	3.12E-07
AA_DQB1_135_32737883_D	P	A	0.8821	0.4881	0.1403	3.19E-07
rs7775228	C	T	0.1572	1.9149	0.1271	3.22E-07
rs9391734	A	G	0.1303	2.0441	0.1403	3.46E-07
rs13211318	C	A	0.1304	2.043	0.1403	3.56E-07
rs2858332	C	A	0.4919	0.5442	0.1197	3.71E-07
AA_DRB1_11_32660115_SL	P	A	0.4994	0.5564	0.1154	3.74E-07
AA_DRB1_28_32660064_E	P	A	0.3276	1.7804	0.1137	3.92E-07
SNP_DRB1_32660063_G	P	A	0.6726	0.5619	0.1137	3.97E-07
AA_DRB1_30_32660058_YG	P	A	0.6727	0.5624	0.1136	4.02E-07
rs28366191	G	A	0.0795	2.2342	0.1591	4.38E-07
rs28366174	C	T	0.0862	2.1975	0.156	4.48E-07
rs13199524	T	C	0.1115	2.1314	0.1501	4.58E-07

TABLE 1-continued

SNP Name	Allele 1	Allele 2	Frequency	Odds Ratio	Standard Error	p-value
HLA_DQB1_0303	P	A	0.0671	2.4032	0.1745	5.05E-07
AA_DRB1_28_32660064_D	P	A	0.6607	0.5655	0.1135	5.06E-07
AA_DRB1_30_32660058_Y	P	A	0.6608	0.5659	0.1134	5.12E-07
rs9275141	T	G	0.4935	0.5583	0.1163	5.43E-07
rs4642516	G	T	0.4935	0.5583	0.1163	5.43E-07
rs6916062	T	C	0.0592	2.4452	0.1805	7.33E-07
AA_DRB1_71_32659935_K	P	A	0.2397	0.4312	0.1703	7.83E-07
rs3129867	G	C	0.3359	1.7289	0.111	8.11E-07
rs9469220	A	G	0.5358	1.7973	0.119	8.29E-07
rs12211410	T	C	0.1129	2.0837	0.1491	8.48E-07
rs12198173	A	G	0.1134	2.0786	0.1493	9.51E-07
rs9272105	G	A	0.4971	1.7558	0.1149	9.58E-07
rs8192591	A	G	0.0547	2.6104	0.1959	9.74E-07
rs2395166	C	T	0.4104	1.7263	0.1144	1.82E-06
AA_DRB1_70_32659938_Q	P	A	0.516	0.5761	0.1161	2.04E-06
AA_DRB1_13_32660109_HY	P	A	0.3459	1.7109	0.1143	2.63E-06
SNP_DRB1_32660109_T	P	A	0.3459	1.7109	0.1143	2.63E-06
AA_DRB1_13_32660109_SF	P	A	0.4616	0.5741	0.1187	2.93E-06
SNP_DRB1_32660109_GA	P	A	0.4616	0.5741	0.1187	2.93E-06
rs9268853	C	T	0.3697	1.6961	0.1131	3.00E-06
rs477515	T	C	0.3542	1.7018	0.114	3.10E-06
rs7356880	T	C	0.0662	2.2657	0.1761	3.42E-06
rs28732193	C	T	0.0622	2.3064	0.1801	3.48E-06
AA_DRB1_104_32657566_S	P	A	0.6419	0.5925	0.1131	3.69E-06
rs3129300	C	A	0.0343	2.8882	0.2292	3.69E-06
AA_DRB1_98_32657584_K	P	A	0.642	0.5929	0.1131	3.79E-06
rs9267845	T	A	0.3688	1.6874	0.1133	3.88E-06
rs1475961	C	T	0.3688	1.6874	0.1133	3.88E-06
rs3096691	A	G	0.3688	1.6872	0.1133	3.91E-06
SNP_DRB1_32657585	C	T	0.3579	1.6838	0.1131	4.06E-06
SNP_DRB1_32657567	C	A	0.3579	1.6837	0.1131	4.07E-06
AA_DRB1_104_32657566_A	P	A	0.3578	1.6838	0.1131	4.07E-06
AA_DRB1_98_32657584_E	P	A	0.3578	1.6838	0.1131	4.07E-06
rs3130609	T	C	0.0343	2.8767	0.2294	4.10E-06
rs9268923	T	C	0.3686	1.682	0.113	4.19E-06
rs2395185	T	G	0.3686	1.682	0.113	4.19E-06
rs9268969	T	C	0.3686	1.682	0.113	4.19E-06
rs9368726	C	T	0.3686	1.682	0.113	4.19E-06
rs9405108	T	C	0.3686	1.682	0.113	4.19E-06
rs3135338	G	A	0.3577	1.6564	0.1099	4.40E-06
rs3135335	G	C	0.3577	1.6564	0.1099	4.40E-06
rs984778	G	A	0.3577	1.6563	0.1099	4.40E-06
AA_DRB1_11_32660115_VG	P	A	0.3508	1.6868	0.1139	4.41E-06
rs2395173	A	G	0.3578	1.6534	0.1099	4.78E-06
rs2395178	G	C	0.3577	1.6531	0.1099	4.80E-06
rs3135395	A	C	0.3581	1.6502	0.1099	5.14E-06
AA_DQA1_47_32717191_RK	P	A	0.5583	1.731	0.1213	6.07E-06
AA_DQA1_52_32717206_R	P	A	0.4418	0.578	0.1213	6.19E-06
SNP_DRB1_32660116_C	P	A	0.3627	1.6641	0.1127	6.27E-06
SNP_DRB1_32660090	G	A	0.3628	1.6641	0.1128	6.29E-06
rs2856992	T	C	0.0319	2.9188	0.2374	6.43E-06
AA_DRB1_11_32660115_SPL	P	A	0.6375	0.6021	0.1127	6.72E-06
rs522308	A	G	0.3509	1.676	0.115	7.05E-06
rs8192583	T	C	0.0626	2.3057	0.1866	7.59E-06
rs2858312	G	C	0.3296	1.6882	0.1174	8.13E-06
rs241405	T	C	0.4962	1.674	0.1156	8.25E-06
rs241399	A	G	0.4965	1.6708	0.1155	8.77E-06
rs9272143	C	T	0.4706	0.5987	0.1154	8.85E-06
rs9271858	G	A	0.4706	0.5987	0.1154	8.85E-06
rs2187688	T	C	0.4966	1.6701	0.1155	8.90E-06

Table Nomenclature:

1. Classical SNP: rs[number]
2. Classical HLA alleles: HLA\_[GENE]\_[ALLELE]
3. HLA Amino Acids: AA\_[GENE]\_[AMINO ACID POSITION]\_[GENETIC POSITION]\_[ALLELE].
4. HLA intragenic SNPs: SNP\_[GENE]\_[POSITION]\_[ALLELE]
5. Insertions/deletions: [VARIANT]\_[GENE]\_[POSITION]\_[INSERTION/x = DELETION]

REFERENCES

[0099] Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989.

[0100] Innis et al., Proc. Natl. Acad. Sci. USA, 85(24): 9436-9449, 1988.

[0101] Guilfoyle et al., Nucleic Acids Research, 25: 1854-1858, 1997.

[0102] Walker et al., Proc. Natl. Acad. Sci. USA, 89: 392-396, 1992.

[0103] Kwoh et al., Proc. Natl. Acad. Sci. USA, 86: 1173, 1989.

**[0104]** Frohman, PCR Protocols: A Guide to Methods and Applications, Academic Press, N. Y., 1990.

**[0105]** Ohara et al., Proc. Natl. Acad. Sci. USA, 86: 5673-5677, 1989.

1. A method of identifying a subject afflicted with, or at risk of developing, thiopurine induced pancreatitis (TIP) comprising:

(a) obtaining a nucleic-acid containing sample from the subject; and

(b) analyzing the sample to detect the presence of at least one genetic marker, or an equivalent to at least one genetic marker, selected from those in Table 1,

wherein the presence of at least genetic marker, or an equivalent to at least one genetic marker, from Table 1 in the sample indicates that the subject is afflicted with, or at risk of, developing TIP.

2. The method of claim 1, wherein the genetic marker is any of alleles, microsatellites, SNPs, or haplotypes.

\* \* \* \* \*