(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 May 2009 (14.05.2009)

PCT

(10) International Publication Number
WO 2009/061799 A2

(51) International Patent Classification:
*C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/US2008/082455

(22) International Filing Date:
5 November 2008 (05.11.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/002,068 5 November 2007 (05.11.2007) US

(71) Applicant *(for all designated States except US)*: **CELERA CORPORATION** [US/US]; c/o Victor K. Lee, 1401 Harbor Bay Parkway, Alameda, California 94502 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LI, Yonghong** [US/US]; c/o Celera Corporation, 1401 Harbor Bay Parkway, Alameda, California 95402 (US). **HUANG, Hongjin** [CA/US]; c/o Celera Corporation, 1401 Harbor Bay Parkway, Alameda, California 94502 (US).

(74) Agent: **LEE, Victor K.**; c/o Celera Corporation, 1401 Harbor Bay Parkway, Alameda, California 94502 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

WO 2009/061799 A2

(54) **Title:** GENETIC POLYMORPHISMS ASSOCIATED WITH LIVER FIBROSIS, METHODS OF DETECTION AND USES THEREOF

(57) **Abstract:** The present invention is based on the discovery of genetic polymorphisms that are associated with liver fibrosis and related pathologies. In particular, the present invention relates to nucleic acid molecules containing the polymorphisms, including groups of nucleic acid molecules that may be used as a signature marker set, variant proteins encoded by such nucleic acid molecules, reagents for detecting the polymorphic nucleic acid molecules and proteins, and methods of using the nucleic acid and proteins as well as methods of using reagents for their detection.

# GENETIC POLYMORPHISMS ASSOCIATED WITH LIVER FIBROSIS, METHODS OF DETECTION AND USES THEREOF

## FIELD OF THE INVENTION

The present invention is in the field of fibrosis diagnosis and therapy and in particular liver fibrosis diagnosis and therapy, and more particularly, liver fibrosis associated with hepatitis C virus (HCV) infection. More specifically, the present invention relates to specific single nucleotide polymorphisms (SNPs) in the human genome, or combinations of such SNPs, and their association with liver fibrosis and related pathologies. Based on differences in allele frequencies in the patient population with advanced or bridging fibrosis/cirrhosis relative to individuals with no or minimal fibrosis, the naturally-occurring SNPs disclosed herein can be used as targets for the design of diagnostic reagents and the development of therapeutic agents, as well as for disease association and linkage analysis. In particular, the SNPs of the present invention are useful for identifying an individual who is at an increased or decreased risk of developing liver fibrosis and for early detection of the disease, for providing clinically important information for the prevention and/or treatment of liver fibrosis, and for screening and selecting therapeutic agents. The SNPs disclosed herein are also useful for human identification applications. Methods, assays, kits, and reagents for detecting the presence of these polymorphisms and their encoded products are provided.

## BACKGROUND OF THE INVENTION

Fibrosis is a quantitative and qualitative change in the extracellular matrix that surrounds cells as a response to tissue injury. The trauma that generates fibrosis is varied and includes radiological trauma (*i.e.,* x-ray, gamma ray, etc.), chemical trauma (*ie.,* radicals, ethanol, phenols, etc.) viral infection and physical trauma. Fibrosis encompasses pathological conditions in a variety of tissues such as pulmonary fibrosis, retroperitoneal fibrosis, epidural fibrosis, congenital fibrosis, focal fibrosis, muscle fibrosis, massive fibrosis, radiation fibrosis (*e.g.* radiation induced lung fibrosis), liver fibrosis and cardiac fibrosis.

Liver Fibrosis in HCV-Infected Subjects

HCV affects about 4 million people in the United States and more than 170 million people worldwide. Approximately 85% of the infected individuals develop chronic hepatitis, and up to 20% progress to bridging fibrosis/cirrhosis, which is end-stage severe liver fibrosis and is generally irreversible (Lauer et al. 2001, *N Eng J Med* 345: 41-52). HCV infection is the major

cause of cirrhosis and hepatocellular carcinoma (HCC), and accounts for one third of liver transplantations. The interval between infection and the development of cirrhosis may exceed 30 years but varies widely among individuals. Based on fibrosis progression rate, chronic HCV patients can be roughly divided into three groups (Poynard *et al* 1997, *Lancet* 349: 825-832): rapid, median, and slow fibrosers.

Previous studies have indicated that host factors may play a role in the progression of fibrosis, and these include age at infection, duration of infection, alcohol consumption, and gender. However, these host factors account for only 17%-29% of the variability in fibrosis progression (Poynard *et al.*, 1997, *Lancet* 349: 825-832; Wright *et al Gut.* 2003, 52(4):574-9). Viral load or viral genotype has not shown significant correlation with fibrosis progression (Poynard *et al.*, 1997, *Lancet* 349: 825-832). Thus, other factors, such as host genetic factors, are likely to play an important role in determining the rate of fibrosis progression.

Recent studies suggest that some genetic polymorphisms influence the progression of fibrosis in patients with HCV infection (Powell *et al. Hepatology* 31(4): 828-33, 2000), autoimmune chronic cholestasis (Tanaka *et al. J. Infec. Dis.* 187:1822-5, 2003), alcohol induced liver diseases (Yamauchi *et al., J. Hepatology* 23(5):519-23, 1995), and nonalcoholic fatty liver diseases (Bernard *et al. Diabetologia* 2000, 43(8):995-9). However, none of these genetic polymorphisms have been integrated into clinical practice for various reasons (Bataller *et al Hepatology*. 2003, 37(3):493-503). For example, limitations in study design, such as small study populations, lack of replication sample sets, and lack of proper control groups have contributed to contradictory results; an example being the conflicting results reported on the role of mutations in the hemochromatosis gene (HFE) on fibrosis progression in HCV-infected patients (Smith *et al.*, *Hepatology*. 1998, 27(6):1695-9; Thorburn *et al., Gut.* 2002, 50(2):248-52).

Currently, there is no diagnostic test that can identify patients who are predisposed to developing liver damage from chronic HCV infection, despite the large variability in fibrosis progression rate among HCV patients. Furthermore, diagnosis of fibrosis stage (early, middle or late) and monitoring of fibrosis progression is currently accomplished by liver biopsy, which is invasive, painful, and costly, and generally must be performed multiple times to assess fibrosis status. The discovery of genetic markers which are useful in identifying HCV-infected individuals who are at increased risk for advancing from early stage fibrosis to cirrhosis and/or HCC may lead to, for example, better therapeutic strategies, economic models, and health care policy decisions.

SNPs

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution, generating variant forms of progenitor genetic sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). A variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. Additionally, the effects of a variant form may be both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. In many cases, both progenitor and variant forms survive and co-exist in a species population. The coexistence of multiple forms of a genetic sequence gives rise to genetic polymorphisms, including SNPs.

Approximately 90% of all polymorphisms in the human genome are SNPs. SNPs are single base positions in DNA at which different alleles, or alternative nucleotides, exist in a population. The SNP position (interchangeably referred to herein as SNP, SNP site, SNP locus, SNP marker, or marker) is usually preceded by and followed by highly conserved sequences of the allele (*e.g.*, sequences that vary in less than 1/100 or 1/1000 members of the populations). An individual may be homozygous or heterozygous for an allele at each SNP position. A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP is an amino acid coding sequence.

A SNP may arise from a substitution of one nucleotide for another at the polymorphic site. Substitutions can be transitions or transversions. A transition is the replacement of one purine nucleotide by another purine nucleotide, or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine, or vice versa. A SNP may also be a single base insertion or deletion variant referred to as an “indel” (Weber *et al.*, “Human diallelic insertion/deletion polymorphisms”, *Am J Hum Genet* 2002 Oct; 71(4):854-62).

A synonymous codon change, or silent mutation/SNP (terms such as “SNP”, ”polymorphism”, “mutation”, ”mutant”, “variation”, and ”variant” are used herein interchangeably), is one that does not result in a change of amino acid due to the degeneracy of the genetic code. A substitution that changes a codon coding for one amino acid to a codon coding for a different amino acid (*i.e.*, a non-synonymous codon change) is referred to as a missense mutation. A nonsense mutation results in a type of non-synonymous codon change in which a stop codon is formed, thereby leading to premature termination of a polypeptide chain and a truncated protein. A read-through mutation is another type of non-synonymous codon

change that causes the destruction of a stop codon, thereby resulting in an extended polypeptide product. While SNPs can be bi-, tri-, or tetra- allelic, the vast majority of the SNPs are bi-allelic, and are thus often referred to as "bi-allelic markers", or "di-allelic markers".

As used herein, references to SNPs and SNP genotypes include individual SNPs and/or haplotypes, which are groups of SNPs that are generally inherited together. Haplotypes can have stronger correlations with diseases or other phenotypic effects compared with individual SNPs, and therefore may provide increased diagnostic accuracy in some cases (Stephens *et al. Science* 293, 489-493, 20 July 2001).

Causative SNPs are those SNPs that produce alterations in gene expression or in the expression, structure, and/or function of a gene product, and therefore are most predictive of a possible clinical phenotype. One such class includes SNPs falling within regions of genes encoding a polypeptide product, *i.e.* cSNPs. These SNPs may result in an alteration of the amino acid sequence of the polypeptide product (*i.e.*, non-synonymous codon changes) and give rise to the expression of a defective or other variant protein. Furthermore, in the case of nonsense mutations, a SNP may lead to premature termination of a polypeptide product. Such variant products can result in a pathological condition, *e.g.*, genetic disease. Examples of genes in which a SNP within a coding sequence causes a genetic disease include sickle cell anemia and cystic fibrosis.

Causative SNPs do not necessarily have to occur in coding regions; causative SNPs can occur in, for example, any genetic region that can ultimately affect the expression, structure, and/or activity of the protein encoded by a nucleic acid. Such genetic regions include, for example, those involved in transcription, such as SNPs in transcription factor binding domains, SNPs in promoter regions, in areas involved in transcript processing, such as SNPs at intron-exon boundaries that may cause defective splicing, or SNPs in mRNA processing signal sequences such as polyadenylation signal regions. Some SNPs that are not causative SNPs nevertheless are in close association with, and therefore segregate with, a disease-causing sequence In this situation, the presence of a SNP correlates with the presence of, or predisposition to, or an increased risk in developing the disease. These SNPs, although not causative, are nonetheless also useful for diagnostics, disease predisposition screening, and other uses.

An association study of a SNP and a specific disorder involves determining the presence or frequency of the SNP allele in biological samples from individuals with the disorder of interest, such as liver fibrosis and related pathologies and comparing the information to that of controls (*i.e.*, individuals who do not have the disorder; controls may be also referred to as "healthy" or "normal" individuals) who are preferably of similar age and race. The appropriate

selection of patients and controls is important to the success of SNP association studies. Therefore, a pool of individuals with well-characterized phenotypes is extremely desirable.

A SNP may be screened in diseased tissue samples or any biological sample obtained from a diseased individual, and compared to control samples, and selected for its increased (or decreased) occurrence in a specific pathological condition, such as pathologies related to liver fibrosis, increased or decreased risk of developing bridging fibrosis/cirrhosis, and progression of liver fibrosis. Once a statistically significant association is established between one or more SNP(s) and a pathological condition (or other phenotype) of interest, then the region around the SNP can optionally be thoroughly screened to identify the causative genetic locus/sequence(s) (*e.g.*, causative SNP/mutation, gene, regulatory region, etc.) that influences the pathological condition or phenotype. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies).

Clinical trials have shown that patient response to treatment with pharmaceuticals is often heterogeneous. There is a continuing need to improve pharmaceutical agent design and therapy. In that regard, SNPs can be used to identify patients most suited to therapy with particular pharmaceutical agents (this is often termed "pharmacogenomics"). Similarly, SNPs can be used to exclude patients from certain treatment due to the patient's increased likelihood of developing toxic side effects or their likelihood of not responding to the treatment. Pharmacogenomics can also be used in pharmaceutical research to assist the drug development and selection process. (Linder *et al.* (1997), *Clinical Chemistry*, 43, 254; Marshall (1997), *Nature Biotechnology*, 15, 1249; International Patent Application WO 97/40462, Spectra Biomedical; and Schafer *et al.* (1998), *Nature Biotechnology*, 16: 3).

## SUMMARY OF THE INVENTION

The present invention relates to the identification of novel SNPs, unique combinations of such SNPs, and haplotypes of SNPs that are associated with liver fibrosis and in particular the increased or decreased risk of developing bridging fibrosis/cirrhosis, and the rate of progression of liver fibrosis. The polymorphisms disclosed herein are directly useful as targets for the design of diagnostic reagents and the development of therapeutic agents for use in the diagnosis and treatment of liver fibrosis and related pathologies.

Based on the identification of SNPs associated with liver fibrosis, the present invention also provides methods of detecting these variants as well as the design and preparation of detection reagents needed to accomplish this task. The invention specifically provides, for example, novel SNPs in genetic sequences involved in liver fibrosis and related pathologies,

isolated nucleic acid molecules (including, for example, DNA and RNA molecules) containing these SNPs, variant proteins encoded by nucleic acid molecules containing such SNPs, antibodies to the encoded variant proteins, computer-based and data storage systems containing the novel SNP information, methods of detecting these SNPs in a test sample, methods of identifying individuals who have an altered (*i.e.*, increased or decreased) risk of developing liver fibrosis based on the presence or absence of one or more particular nucleotides (alleles) at one or more SNP sites disclosed herein or the detection of one or more encoded variant products (*e.g.*, variant mRNA transcripts or variant proteins), methods of identifying individuals who are more or less likely to respond to a treatment (or more or less likely to experience undesirable side effects from a treatment, etc.), methods of screening for compounds useful in the treatment of a disorder associated with a variant gene/protein, compounds identified by these methods, methods of treating disorders mediated by a variant gene/protein, methods of using the novel SNPs of the present invention for human identification, etc.

In Tables 1-2, the present invention provides gene information, transcript sequences (SEQ ID NOS:1-16), encoded amino acid sequences (SEQ ID NOS:17-32), genomic sequence (SEQ ID NO: 65-90), transcript-based context sequences (SEQ ID NOS:33-64) and genomic-based context sequences (SEQ ID NOS:91-358) that contain the SNPs of the present invention, and extensive SNP information that includes observed alleles, allele frequencies, populations/ethnic groups in which alleles have been observed, information about the type of SNP and corresponding functional effect, and, for cSNPs, information about the encoded polypeptide product. The transcript sequences (SEQ ID NOS:1-16), amino acid sequences (SEQ ID NOS:17-32), genomic sequence (SEQ ID NO:65-90), transcript-based SNP context sequences (SEQ ID NOS: 33-64), and genomic-based SNP context sequences (SEQ ID NOS:91-358) are also provided in the Sequence Listing. The Sequence Listing also provides the primer sequences (SEQ ID NOS:359-757).

In a specific embodiment of the present invention, SNPs that occur naturally in the human genome are provided as isolated nucleic acid molecules. These SNPs are associated with liver fibrosis and related pathologies. In particular the SNPs are associated with either an increased or decreased risk of developing bridging fibrosis/cirrhosis and affect the rate of progression of liver fibrosis. As such, they can have a variety of uses in the diagnosis and/or treatment of liver fibrosis and related pathologies. One aspect of the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence in which at least one nucleotide is a SNP disclosed in Tables 1-2. In an alternative embodiment, a nucleic acid of the invention is an amplified polynucleotide, which is produced by amplification of a SNP-containing nucleic acid

template. In another embodiment, the invention provides for a variant protein that is encoded by a nucleic acid molecule containing a SNP disclosed herein.

In certain exemplary embodiments, the invention provides polymorphisms set forth in at least one of Tables 7-11 and 13-20, and methods of using these polymorphisms, particularly for determining an individual's risk for developing liver fibrosis or risk for progressing rapidly from minimal fibrosis to bridging fibrosis/cirrhosis (particularly for an HCV-infected individual), or for other uses related to liver fibrosis/cirrhosis.

In certain exemplary embodiments, the invention provides haplotypes set forth in at least one of Tables 8-10, 16, and 18, and methods of using these polymorphisms, particularly for determining an individual's risk for developing liver fibrosis or risk for progressing rapidly from minimal fibrosis to bridging fibrosis/cirrhosis (particularly for an HCV-infected individual), or for other uses related to liver fibrosis/cirrhosis.

In yet another embodiment of the invention, a reagent for detecting a SNP in the context of its naturally-occurring flanking nucleotide sequences (which can be, *e.g.*, either DNA or mRNA) is provided. In particular, such a reagent may be in the form of, for example, a hybridization probe or an amplification primer that is useful in the specific detection of a SNP of interest. In an alternative embodiment, a protein detection reagent is used to detect a variant protein that is encoded by a nucleic acid molecule containing a SNP disclosed herein. A preferred embodiment of a protein detection reagent is an antibody or an antigen-reactive antibody fragment.

Various embodiments of the invention also provide kits comprising SNP detection reagents, and methods for detecting the SNPs disclosed herein by employing detection reagents. In a specific embodiment, the present invention provides for a method of identifying a human having an alterered (increased or decreased) risk of developing liver fibrosis by detecting the presence or absence of one or more SNP alleles disclosed herein in said human's nucleic acids wherein the presence of the SNP is indicative of an altered risk for developing liver fibrosis in said human. In another embodiment, a method for diagnosis of liver fibrosis and related pathologies by detecting the presence or absence of one or more SNP alleles disclosed herein is provided.

The nucleic acid molecules of the invention can be inserted in an expression vector, such as to produce a variant protein in a host cell. Thus, the present invention also provides for a vector comprising a SNP-containing nucleic acid molecule, genetically-engineered host cells containing the vector, and methods for expressing a recombinant variant protein using such host cells. In another specific embodiment, the host cells, SNP-containing nucleic acid molecules,

and/or variant proteins can be used as targets in a method for screening and identifying therapeutic agents or pharmaceutical compounds useful in the treatment of liver fibrosis and related pathologies.

An aspect of this invention is a method for treating liver fibrosis in a human subject wherein said human subject harbors a SNP, gene, transcript, and/or encoded protein identified in Tables 1-2, which method comprises administering to said human subject a therapeutically or prophylactically effective amount of one or more agents counteracting the effects of the disease, such as by inhibiting (or stimulating) the activity of the gene, transcript, and/or encoded protein identified in Tables 1-2.

Another aspect of this invention is a method for identifying an agent useful in therapeutically or prophylactically treating liver fibrosis and related pathologies in a human subject wherein said human subject harbors a SNP, gene, transcript, and/or encoded protein identified in Tables 1-2, which method comprises contacting the gene, transcript, or encoded protein with a candidate agent under conditions suitable to allow formation of a binding complex between the gene, transcript, or encoded protein and the candidate agent and detecting the formation of the binding complex, wherein the presence of the complex identifies said agent.

Another aspect of this invention is a method for treating liver fibrosis and related pathologies in a human subject, which method comprises:

(i) determining that said human subject harbors a SNP, gene, transcript, and/or encoded protein identified in Tables 1-2, and

(ii) administering to said subject a therapeutically or prophylactically effective amount of one or more agents counteracting the effects of the disease.

Yet another aspect of this invention is a method for evaluating the suitability of a patient for HCV treatment comprising determining the genotype of said patient with respect to a particular set of SNP markers, said SNP markers comprising a plurality of individual SNPs ranging from two to seven SNPs in Table 1 or Table 2, and calculating a score using an appropriate algorithm based on the genotype of said patient, the resulting score being indicative of the suitability of said patient undergoing HCV treatment.

Another aspect of the invention is a method of treating an HCV patient comprising administering an appropriate drug such as interferon, or interferon and ribavirin, in a therapeutically effective amount to said HCV patient whose genotype has been shown to contain a SNP (or a plurality of SNPs) disclosed herein, such in any of Tables 1-20 and/or the Examples sections below.

Many other uses and advantages of the present invention will be apparent to those skilled in the art upon review of the detailed description of the preferred embodiments herein. Solely for clarity of discussion, the invention is described in the sections below by way of non-limiting examples.

**DESCRIPTION OF THE SEQUENCE LISTING**

File SEQLIST_CD000020PCT.txt provides the Sequence Listing in text (ASCII) format. The Sequence Listing provides the transcript sequences (SEQ ID NOS:1-16) and protein sequences (SEQ ID NOS:17-32) as shown in Table 1, and genomic sequence (SEQ ID NO:65-90) as shown in Table 2, for each liver fibrosis-associated gene that contains one or more SNPs of the present invention. Also provided in the Sequence Listing are context sequences flanking each SNP, including both transcript-based context sequences as shown in Table 1 (SEQ ID NOS:33-64) and genomic-based context sequences as shown in Table 2 (SEQ ID NOS:91-358). The context sequences generally provide 100bp upstream (5') and 100bp downstream (3') of each SNP, with the SNP in the middle of the context sequence, for a total of 200bp of context sequence surrounding each SNP.

File SEQLIST_CD000020PCT.txt is 2,468 KB in size, and was created on October 22, 2008.

The Sequence Listing is hereby incorporated by reference pursuant to 37 CFR 1.77(b)(4).

**DESCRIPTION OF TABLE 1 AND TABLE 2**

Table 1 and Table 2 disclose the SNP and associated gene/transcript/protein information of the present invention. For each gene, Table 1 provides a header containing gene, transcript and protein information, followed by a transcript and protein sequence identifier (SEQ ID NO), and then SNP information regarding each SNP found in that gene/transcript including the transcript context sequence. For each gene in Table 2, a header is provided that contains gene and genomic information, followed by a genomic sequence identifier (SEQ ID NO) and then SNP information regarding each SNP found in that gene, including the genomic context sequence.

Note that SNP markers may be included in both Table 1 and Table 2; Table 1 presents the SNPs relative to their transcript sequences and encoded protein sequences, whereas Table 2 presents the SNPs relative to their genomic sequences. In some instances Table 2 may also include, after the last gene sequence, genomic sequences of one or more intergenic regions, as well as SNP context sequences and other SNP information for any SNPs that lie within these

intergenic regions. Additionally, in either Table 1 or 2 a "Related Interrogated SNP" may be listed following a SNP which is determined to be in LD with that interrogated SNP according to the given Power value. SNPs can be readily cross-referenced between all Tables based on their Celera hCV (or, in some instances, hDV) identification numbers and/or public rs identification numbers, and to the Sequence Listing based on their corresponding SEQ ID NOs.

The gene/transcript/protein information includes:

- a gene number (1 through n, where n = the total number of genes in the Table),

- a gene symbol, along with an Entrez gene identification number (Entrez Gene database, National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health)

- a gene name,

- an accession number for the transcript (e.g., RefSeq NM number, or a Celera hCT identification number if no RefSeq NM number is available) (Table 1 only),

- an accession number for the protein (e.g., RefSeq NP number, or a Celera hCP identification number if no RefSeq NP number is available) (Table 1 only),

- the chromosome number of the chromosome on which the gene is located,

- an OMIM ("Online Mendelian Inheritance in Man" database, Johns Hopkins University/NCBI) public reference number for the gene, and OMIM information such as alternative gene/protein name(s) and/or symbol(s) in the OMIM entry.

Note that, due to the presence of alternative splice forms, multiple transcript/protein entries may be provided for a single gene entry in Table 1; i.e., for a single Gene Number, multiple entries may be provided in series that differ in their transcript/protein information and sequences.

Following the gene/transcript/protein information is a transcript context sequence (Table 1), or a genomic context sequence (Table 2), for each SNP within that gene.

After the last gene sequence, Table 2 may include additional genomic sequences of intergenic regions (in such instances, these sequences are identified as "Intergenic region:" followed by a numerical identification number), as well as SNP context sequences and other SNP information for any SNPs that lie within each intergenic region (such SNPs are identified as "INTERGENIC" for SNP type).

Note that the transcript, protein, and transcript-based SNP context sequences are all provided in the Sequence Listing. The transcript-based SNP context sequences are provided in both Table 1 and also in the Sequence Listing. The genomic and genomic-based SNP context sequences are provided in the Sequence Listing. The genomic-based SNP context sequences are

provided in both Table 2 and in the Sequence Listing. SEQ ID NOs are indicated in Table 1 for the transcript-based context sequences (SEQ ID NOS:33-64); SEQ ID NOs are indicated in Table 2 for the genomic-based context sequences (SEQ ID NOS:91-358).

The SNP information includes:

- Context sequence (taken from the transcript sequence in Table 1, the genomic sequence in Table 2) with the SNP represented by its IUB code, including 100bp upstream (5') of the SNP position plus 100bp downstream (3') of the SNP position (the transcript-based SNP context sequences in Table 1 are provided in the Sequence Listing as SEQ ID NOS:33-64; the genomic-based SNP context sequences in Table 2 are provided in the Sequence Listing as SEQ ID NOS:91-358).

- Celera hCV internal identification number for the SNP (in some instances, an "hDV" number is given instead of an "hCV" number).

- The corresponding public identification number for the SNP, the rs number.

- "SNP Chromosome Position" indicates the nucleotide position of the SNP along the entire sequence of the chromosome as provided in NCBI Genome Build 36.

- SNP position (nucleotide position of the SNP within the given transcript sequence (Table 1) or within the given genomic sequence (Table 2)).

- "Related Interrogated SNP" is the interrogated SNP with which the listed SNP is in LD at the given value of Power.

- SNP source (may include any combination of one or more of the following five codes, depending on which internal sequencing projects and/or public databases the SNP has been observed in: "Applera" = SNP observed during the re-sequencing of genes and regulatory regions of 39 individuals, "Celera" = SNP observed during shotgun sequencing and assembly of the Celera human genome sequence, "Celera Diagnostics" = SNP observed during re-sequencing of nucleic acid samples from individuals who have a disease, "dbSNP" = SNP observed in the dbSNP public database, "HGBASE" = SNP observed in the HGBASE public database, "HGMD" = SNP observed in the Human Gene Mutation Database (HGMD) public database, "HapMap" = SNP observed in the International HapMap Project public database, "CSNP" = SNP observed in an internal Applied Biosystems (Foster City, CA) database of coding SNPS (cSNPs).

Note that multiple "Applera" source entries for a single SNP indicate that the same SNP was covered by multiple overlapping amplification products and the re-sequencing results (e.g., observed allele counts) from each of these amplification products is being provided.

Certain SNPs from Tables 1 and 2 are SNPs for which the SNP source falls into one of the following three categories: 1) SNPs for which the SNP source is only "Applera" and none

other, 2) SNPs for which the SNP source is only "Celera Diagnostics" and none other, and 3) SNPs for which the SNP source is both "Applera" and "Celera Diagnostics" but none other. These SNPs have not been observed in any of the public databases (dbSNP, HGBASE, and HGMD), and were also not observed during shotgun sequencing and assembly of the Celera human genome sequence (i.e., "Celera" SNP source). These SNPs include: hCV25597248 (transcript-based context sequence SEQ ID NO:39 in Table 1, genomic-based context sequence SEQ ID NO:113 in Table 2) and hCV25635059 (transcript-based context sequences SEQ ID NOS:42 and 45 in Table 1; genomic-based context sequences SEQ ID NO:123 and 335 in Table 2).

- Population/allele/allele count information in the format of [population1(first_allele,count|second_allele,count)population2(first_allele,count|second_allele,count) total (first_allele,total count|second_allele,total count)]. The information in this field includes populations/ethnic groups in which particular SNP alleles have been observed ("cau" = Caucasian, "his" = Hispanic, "chn" = Chinese, and "afr" = African-American, "jpn" = Japanese, "ind" = Indian, "mex" = Mexican, "ain" = "American Indian, "cra" = Celera donor, "no_pop" = no population information available), identified SNP alleles, and observed allele counts (within each population group and total allele counts), where available ["-" in the allele field represents a deletion allele of an insertion/deletion ("indel") polymorphism (in which case the corresponding insertion allele, which may be comprised of one or more nucleotides, is indicated in the allele field on the opposite side of the "|"); "-"in the count field indicates that allele count information is not available]. For certain SNPs from the public dbSNP database, population/ethnic information is indicated as follows (this population information is publicly available in dbSNP): "HISP1" = human individual DNA (anonymized samples) from 23 individuals of self-described HISPANIC heritage; "PAC1" = human individual DNA (anonymized samples) from 24 individuals of self-described PACIFIC RIM heritage; "CAUC1" = human individual DNA (anonymized samples) from 31 individuals of self-described CAUCASIAN heritage; "AFR1" = human individual DNA (anonymized samples) from 24 individuals of self-described AFRICAN/AFRICAN AMERICAN heritage; "P1" = human individual DNA (anonymized samples) from 102 individuals of self-described heritage; "PA130299515"; "SC_12_A" = SANGER 12 DNAs of Asian origin from Corielle cell repositories, 6 of which are male and 6 female; "SC_12_C" = SANGER 12 DNAs of Caucasian origin from Corielle cell repositories from the CEPH/UTAH library, six male and six female; "SC_12_AA" = SANGER 12 DNAs of African-American origin from Corielle cell repositories 6 of which are male and 6 female; "SC_95_C" = SANGER 95 DNAs of Caucasian origin from Corielle cell repositories from the

CEPH/UTAH library; and "SC_12_CA" = Caucasians - 12 DNAs from Corielle cell repositories that are from the CEPH/UTAH library, six male and six female.

Note that for SNPs of "Applera" SNP source, genes/regulatory regions of 39 individuals (20 Caucasians and 19 African Americans) were re-sequenced and, since each SNP position is represented by two chromosomes in each individual (with the exception of SNPs on X and Y chromosomes in males, for which each SNP position is represented by a single chromosome), up to 78 chromosomes were genotyped for each SNP position. Thus, the sum of the African-American ("afr") allele counts is up to 38, the sum of the Caucasian allele counts ("cau") is up to 40, and the total sum of all allele counts is up to 78.

Note that semicolons separate population/allele/count information corresponding to each indicated SNP source; i.e., if four SNP sources are indicated, such as "Celera," "dbSNP," "HGBASE," and "HGMD," then population/allele/count information is provided in four groups which are separated by semicolons and listed in the same order as the listing of SNP sources, with each population/allele/count information group corresponding to the respective SNP source based on order; thus, in this example, the first population/allele/count information group would correspond to the first listed SNP source (Celera) and the third population/allele/count information group separated by semicolons would correspond to the third listed SNP source (HGBASE); if population/allele/count information is not available for any particular SNP source, then a pair of semicolons is still inserted as a place-holder in order to maintain correspondence between the list of SNP sources and the corresponding listing of population/allele/count information.

- SNP type (e.g., location within gene/transcript and/or predicted functional effect) ["MIS-SENSE MUTATION" = SNP causes a change in the encoded amino acid (i.e., a non-synonymous coding SNP); "SILENT MUTATION" = SNP does not cause a change in the encoded amino acid (i.e., a synonymous coding SNP); "STOP CODON MUTATION" = SNP is located in a stop codon; "NONSENSE MUTATION" = SNP creates or destroys a stop codon; "UTR 5" = SNP is located in a 5' UTR of a transcript; "UTR 3" = SNP is located in a 3' UTR of a transcript; "PUTATIVE UTR 5" = SNP is located in a putative 5' UTR; "PUTATIVE UTR 3" = SNP is located in a putative 3' UTR; "DONOR SPLICE SITE" = SNP is located in a donor splice site (5' intron boundary); "ACCEPTOR SPLICE SITE" = SNP is located in an acceptor splice site (3' intron boundary); "CODING REGION" = SNP is located in a protein-coding region of the transcript; "EXON" = SNP is located in an exon; "INTRON" = SNP is located in an intron; "hmCS" = SNP is located in a human-mouse conserved segment; "TFBS" = SNP is

located in a transcription factor binding site; "UNKNOWN" = SNP type is not defined; "INTERGENIC" = SNP is intergenic, i.e., outside of any gene boundary].

- Protein coding information (Table 1 only), where relevant, in the format of [protein SEQ ID NO, amino acid position, (amino acid-1, codon1) (amino acid-2, codon2)]. The information in this field includes SEQ ID NO of the encoded protein sequence, position of the amino acid residue within the protein identified by the SEQ ID NO that is encoded by the codon containing the SNP, amino acids (represented by one-letter amino acid codes) that are encoded by the alternative SNP alleles (in the case of stop codons, "X" is used for the one-letter amino acid code), and alternative codons containing the alternative SNP nucleotides which encode the amino acid residues (thus, for example, for missense mutation-type SNPs, at least two different amino acids and at least two different codons are generally indicated; for silent mutation-type SNPs, one amino acid and at least two different codons are generally indicated, etc.). In instances where the SNP is located outside of a protein-coding region (e.g., in a UTR region), "None" is indicated following the protein SEQ ID NO.

Note that SNPs can be cross-referenced between all tables herein based on the hCV/hDV and/or rs identification number of each SNP. However, eight of the SNPs may possess two different hCV/hDV identification numbers, as follows:

hDV71101101is equivalent to hCV12023155;

hDV71115804 is equivalent to hCV1113705;

hDV71153302 is equivalent to hCV8847946;

hDV71161271 is equivalent to hCV11245312;

hDV71170845is equivalent to hCV12023147;

hDV71206854is equivalent to hCV27253292;

hDV71210383 is equivalent to hCV27495012; and

hDV71564063 is equivalent to hCV31784034

File CD000020PCT_Table1.txt provides Table 1 of the present application. File CD000020PCT_Table1.txt is 30 KB in size, and was created on October 21, 2008.

File CD000020PCT_Table2.txt provides Table 2 of the present application. File CD000020PCT_Table2.txt is 202 KB in size, and was created on October 21, 2008.

**DESCRIPTION OF TABLE 3**

Table 3 provides sequences (SEQ ID NOS:359-757) of primers that may be used to assay the SNPs disclosed herein by allele-specific PCR or other methods, such as for uses related to liver fibrosis.

Table 3 provides the following:

- the column labeled "Marker" provides an identification number (e.g., a public "rs" number or internal "hCV" number) for each SNP site.

- the column labeled "Alleles" designates the two alternative alleles (i.e., nucleotides) at the SNP site. These alleles are targeted by the allele-specific primers (the allele-specific primers are shown as Primer 1 and Primer 2). Note that alleles may be presented in Table 3 based on a different orientation (i.e., the reverse complement) relative to how the same alleles are presented in Tables 1-2.

- the column labeled "Primer 1 (Allele-Specific Primer)" provides an allele-specific primer that is specific for an allele designated in the "Alleles" column.

- the column labeled "Primer 2 (Allele-Specific Primer)" provides an allele-specific primer that is specific for the other allele designated in the "Alleles" column.

- the column labeled "Common Primer" provides a common primer that is used in conjunction with each of the allele-specific primers (i.e., Primer 1 and Primer 2) and which hybridizes at a site away from the SNP position.

All primer sequences are given in the 5' to 3' direction.

Each of the nucleotides designated in the "Alleles" column matches or is the reverse complement of (depending on the orientation of the primer relative to the designated allele) the 3' nucleotide of the allele-specific primer (i.e., either Primer 1 or Primer 2) that is specific for that allele. For instance, Primer 1 for hCV11722237 can be used to detect the reverse complement of the C allele listed in the "Allele" column in Table 3, whereas Primer 1 for hCV29292005 can be used to detect the G allele listed in the "Allele" column in Table 3.

**DESCRIPTION OF TABLE 4**

Table 4 provides a list of linkage disequilibrium (LD) SNPs that are related to and derived from certain interrogated SNPs. The interrogated SNPs, which are shown in column 1 (which indicates the hCV identification numbers of each interrogated SNP) and column 2 (which indicates the public rs identification numbers of each interrogated SNP) of Table 4, are statistically significantly associated with liver fibrosis, as described and shown herein, particularly in Tables 5-20 and in the Examples section below. The LD SNPs are provided as an example of SNPs which can also serve as markers for disease association based on their being in LD with an interrogated SNP. The criteria and process of selecting such LD SNPs, including the calculation of the $r^2$ value and the $r^2$ threshold value, are described in Example Four, below.

In Table 4, the column labeled "Interrogated SNP" presents each marker as identified by its unique hCV identification number. The column labeled "Interrogated rs" presents the publicly known rs identification number for the corresponding hCV number. The column labeled "LD SNP" presents the hCV numbers of the LD SNPs that are derived from their corresponding interrogated SNPs. The column labeled "LD SNP rs" presents the publicly known rs identification number for the corresponding hCV number. The column labeled "Power" presents the level of power where the $r^2$ threshold is set. For example, when power is set at .51, the threshold $r^2$ value calculated therefrom is the minimum $r^2$ that an LD SNP must have in reference to an interrogated SNP, in order for the LD SNP to be classified as a marker capable of being associated with a disease phenotype at greater than 51% probability. The column labeled "Threshold $r^2$" presents the minimum value of $r^2$ that an LD SNP must meet in reference to an interrogated SNP in order to qualify as an LD SNP. The column labeled "$r^2$" presents the actual $r^2$ value of the LD SNP in reference to the interrogated SNP to which it is related.

**BRIEF DESCRIPTION OF TABLES 5-20**

Tables 5-20 provide the results of statistical analyses for SNPs disclosed in Tables 1 and 2 (SNPs can be cross-referenced between all the tables herein based on their hCV and/or rs identification numbers). The results shown in Tables 5-20 provide support for the association of these SNPs with liver fibrosis/cirrhosis (cirrhosis is severe fibrosis, which may also be referred to as bridging fibrosis).

Table 5 provides a summary of clinical data for subjects in the analysis described in Example One below.

Table 6 provides inclusion/exclusion criteria for enrollment of subjects in the analysis described in Example One below.

Table 7 provides SNPs in the *TLR4* region significantly associated with cirrhosis risk (see Example One below).

Tables 8-10 provide haplotypes in the *TLR4* region significantly associated with cirrhosis risk (see Example One below).

Table 11 provides data for causal SNP analysis (see Example One below).

Table 12 provides clinical characteristics of subjects in the analysis described in Example Two below.

Table 13 provides significant SNPs in the *TLR4* region (P<=0.01) (see Example Two below).

Table 14 provides fine mapping coverage for CRS7 (see Example Three below).

Table 15 provides markers significantly associated with liver fibrosis risk that are in high linkage disequilibrium with the *TLR4* SNP rs4986791 or independently significant (see Example Three below).

Table 16 provides haplotypes in the *TLR4* region associated with liver fibrosis risk (see Example Three below).

Table 17 provides SNPs in the *STXBP5L* locus that are independently associated with liver fibrosis risk (see Example Three below).

Table 18 provides haplotype in the *STXBP5L* region associated with liver fibrosis risk (see Example Three below).

Table 19 provides markers significantly associated with liver fibrosis risk (see Example Three below).

Table 20 provides further information regarding the SNPs disclosed herein, as well as additional SNPs associated with liver fibrosis risk (see Example Three below).

Throughout Tables 5-20, "OR" refers to the odds ratio (and "95% CI" refers to the 95% confidence interval for the odds ratio). Odds ratios (OR) that are greater than one indicate that a given allele (or combination of alleles such as a haplotype or diplotype) is a risk allele (which may also be referred to as a susceptibility allele), whereas odds ratios that are less than one indicate that a given allele is a non-risk allele (which may also be referred to as a protective allele). For a given risk allele, the other alternative allele at the SNP position (which can be derived from the information provided in Tables 1-2, for example) may be considered a non-risk allele. For a given non-risk allele, the other alternative allele at the SNP position may be considered a risk allele.

Thus, with respect to disease risk (e.g., liver fibrosis), if the odds ratio for a particular allele at a SNP position is greater than one, this indicates that an individual with this particular allele has a higher risk for the disease than an individual who has the other allele at the SNP position. In contrast, if the odds ratio for a particular allele is less than one, this indicates that an individual with this particular allele has a reduced risk for the disease compared with an individual who has the other allele at the SNP position.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides SNPs associated with liver fibrosis and related pathologies, nucleic acid molecules containing SNPs, methods and reagents for the detection of the SNPs disclosed herein, uses of these SNPs for the development of detection reagents, and

assays or kits that utilize such reagents. The liver fibrosis-associated SNPs disclosed herein are useful for diagnosing, screening for, and evaluating predisposition to liver fibrosis, including an increased or decreased risk of developing bridging fibrosis/cirrhosis, the rate of progression of fibrosis, and related pathologies in humans. Furthermore, such SNPs and their encoded products are useful targets for the development of therapeutic agents.

A large number of SNPs have been identified from re-sequencing DNA from 39 individuals, and they are indicated as "Applera" SNP source in Tables 1-2. Their allele frequencies observed in each of the Caucasian and African-American ethnic groups are provided. Additional SNPs included herein were previously identified during shotgun sequencing and assembly of the human genome, and they are indicated as "Celera" SNP source in Tables 1-2. Furthermore, the information provided in Table 1-2, particularly the allele frequency information obtained from 39 individuals and the identification of the precise position of each SNP within each gene/transcript, allows haplotypes (*i.e.*, groups of SNPs that are co-inherited) to be readily inferred. The present invention encompasses SNP haplotypes, as well as individual SNPs.

Thus, the present invention provides individual SNPs associated with liver fibrosis, as well as combinations of SNPs and haplotypes in genetic regions associated with liver fibrosis, polymorphic/variant transcript sequences (SEQ ID NOS:1-16) and genomic sequence (SEQ ID NO:65-90) containing SNPs, encoded amino acid sequences (SEQ ID NOS: 17-32), and both transcript-based SNP context sequences (SEQ ID NOS: 33-64) and genomic-based SNP context sequences (SEQ ID NOS:91-358) (transcript sequences, protein sequences, and transcript-based SNP context sequences are provided in Table 1 and the Sequence Listing; genomic sequences and genomic-based SNP context sequences are provided in Table 2 and the Sequence Listing), methods of detecting these polymorphisms in a test sample, methods of determining the risk of an individual of having or developing liver fibrosis, methods of screening for compounds useful for treating disorders associated with a variant gene/protein such as liver fibrosis, compounds identified by these screening methods, methods of using the disclosed SNPs to select a treatment strategy, methods of treating a disorder associated with a variant gene/protein (*i.e.*, therapeutic methods), and methods of using the SNPs of the present invention for human identification.

The present invention provides novel SNPs associated with liver fibrosis and related pathologies, as well as SNPs that were previously known in the art, but were not previously known to be associated with liver fibrosis. Accordingly, the present invention provides novel compositions and methods based on the novel SNPs disclosed herein, and also provides novel methods of using the known, but previously unassociated, SNPs in methods relating to liver fibrosis (*e.g.*, for diagnosing liver fibrosis, etc.). In Tables 1-2, known SNPs are identified based

on the public database in which they have been observed, which is indicated as one or more of the following SNP types: "dbSNP" = SNP observed in dbSNP, "HGBASE" = SNP observed in HGBASE, and "HGMD" = SNP observed in the Human Gene Mutation Database (HGMD). Novel SNPs for which the SNP source is only "Applera" and none other, *i.e.*, those that have not been observed in any public databases and which were also not observed during shotgun sequencing and assembly of the Celera human genome sequence (*i.e.*, "Celera" SNP source), include hCV25597248 (transcript-based context sequence SEQ ID NO:39 in Table 1, genomic-based context sequence SEQ ID NO:113 in Table 2) and hCV25635059 (transcript-based context sequences SEQ ID NOS:42 and 45 in Table 1; genomic-based context sequences SEQ ID NO:123 and 335 in Table 2).

Particular SNP alleles of the present invention can be associated with either an increased risk of having or developing liver fibrosis and related pathologies, or a decreased risk of having or developing liver fibrosis. SNP alleles that are associated with a decreased risk of having or developing liver fibrosis may be referred to as "protective" alleles, and SNP alleles that are associated with an increased risk of having or developing liver fibrosis may be referred to as "susceptibility" alleles, "risk" alleles, or "risk factors". Thus, whereas certain SNPs (or their encoded products) can be assayed to determine whether an individual possesses a SNP allele that is indicative of an increased risk of having or developing liver fibrosis (*i.e.*, a susceptibility allele), other SNPs (or their encoded products) can be assayed to determine whether an individual possesses a SNP allele that is indicative of a decreased risk of having or developing liver fibrosis (*i.e.*, a protective allele). Similarly, particular SNP alleles of the present invention can be associated with either an increased or decreased likelihood of responding to a particular treatment or therapeutic compound, or an increased or decreased likelihood of experiencing toxic effects from a particular treatment or therapeutic compound. The term "altered" may be used herein to encompass either of these two possibilities (*e.g.*, an increased or a decreased risk/likelihood).

Those skilled in the art will readily recognize that nucleic acid molecules may be double-stranded molecules and that reference to a particular site on one strand refers, as well, to the corresponding site on a complementary strand. In defining a SNP position, SNP allele, or nucleotide sequence, reference to an adenine, a thymine (uridine), a cytosine, or a guanine at a particular site on one strand of a nucleic acid molecule also defines the thymine (uridine), adenine, guanine, or cytosine (respectively) at the corresponding site on a complementary strand of the nucleic acid molecule. Thus, reference may be made to either strand in order to refer to a particular SNP position, SNP allele, or nucleotide sequence. Probes and primers, may be designed to hybridize to either strand and SNP genotyping methods disclosed herein may

generally target either strand. Throughout the specification, in identifying a SNP position, reference is generally made to the protein-encoding strand, only for the purpose of convenience.

References to variant peptides, polypeptides, or proteins of the present invention include peptides, polypeptides, proteins, or fragments thereof, that contain at least one amino acid residue that differs from the corresponding amino acid sequence of the art-known peptide/polypeptide/protein (the art-known protein may be interchangeably referred to as the "wild-type", "reference", or "normal" protein). Such variant peptides/polypeptides/proteins can result from a codon change caused by a nonsynonymous nucleotide substitution at a protein-coding SNP position (*i.e.*, a missense mutation) disclosed by the present invention. Variant peptides/polypeptides/proteins of the present invention can also result from a nonsense mutation, *i.e.,* a SNP that creates a premature stop codon, a SNP that generates a read-through mutation by abolishing a stop codon, or due to any SNP disclosed by the present invention that otherwise alters the structure, function/activity, or expression of a protein, such as a SNP in a regulatory region (*e.g.* a promoter or enhancer) or a SNP that leads to alternative or defective splicing, such as a SNP in an intron or a SNP at an exon/intron boundary. As used herein, the terms "polypeptide", "peptide", and "protein" are used interchangeably.

As used herein, an "allele" may refer to a nucleotide at a SNP position (wherein at least two alternative nucleotides are present in the population at the SNP position, in accordance with the inherent definition of a SNP) or may refer to an amino acid residue that is encoded by the codon which contains the SNP position (where the alternative nucleotides that are present in the population at the SNP position form alternative codons that encode different amino acid residues). An "allele" may also be referred to herein as a "variant". Also, an amino acid residue that is encoded by a codon containing a particular SNP may simply be referred to as being encoded by the SNP.

A phrase such as "as reprented by", "as shown by", "as symbolized by", or "as designated by" may be used herein to refer to a SNP within a sequence (*e.g.*, a polynucleotide context sequence surrounding a SNP), such as in the context of "a polymorphism as represented by position 101 of SEQ ID NO:X or its complement". Typically, the sequence surrounding a SNP may be recited when referring to a SNP, however the sequence is not intended as a structural limitation beyond the specific SNP position itself. Rather, the sequence is recited merely as a way of referring to the SNP (in this example, "SEQ ID NO:X or its complement" is recited in order to refer to the SNP located at position 101 of SEQ ID NO:X, but SEQ ID NO:X or its complement is not intended as a structural limitation beyond the specific SNP position itself). A SNP is a variation at a single nucleotide position and therefore it is customary to refer to context

sequence (*e.g.*, SEQ ID NO:X in this example) surrounding a particular SNP position in order to uniquely identify and refer to the SNP. Alternatively, a SNP can be referred to by a unique identification number such as a public "rs" identification number or an internal "hCV" identification number, such as provided herein for each SNP (*e.g.*, in Tables 1-2).

With respect to an individual's risk for a disease (*e.g.*, based on the presence or absence of one or more SNPs disclosed herein in the individual's nucleic acid), terms such as "assigning" or "designating" may be used herein to characterize the individual's risk for the disease.

As used herein, the term "benefit" (with respect to a preventive or therapeutic drug treatment) is defined as achieving a reduced risk for a disease that the drug is intended to treat or prevent (*e.g.*, liver fibrosis) by administrating the drug treatment, compared with the risk for the disease in the absence of receiving the drug treatment (or receiving a placebo in lieu of the drug treatment) for the same genotype. The term "benefit" may be used herein interchangeably with terms such as "respond positively" or "positively respond".

As used herein, the terms "drug" and "therapeutic agent" are used interchangeably, and may include, but are not limited to, small molecule compounds, biologics (*e.g.*, antibodies, proteins, protein fragments, fusion proteins, glycoproteins, etc.), nucleic acid agents (*e.g.*, antisense, RNAi/siRNA, and microRNA molecules, etc.), vaccines, etc., which may be used for therapeutic and/or preventive treatment of a disease (*e.g.*, liver fibrosis).

The various methods described herein, such as correlating the presence or absence of a polymorphism with an altered (*e.g.*, increased or decreased) risk (or no altered risk) for liver fibrosis (and/or correlating the presence or absence of a polymorphism with the predicted response of an individual to a drug such as a statin), can be carried out by automated methods such as by using a computer (or other apparatus/devices such as biomedical devices, laboratory instrumentation, or other apparatus/devices having a computer processor) programmed to carry out any of the methods described herein. For example, computer software (which may be interchangeably referred to herein as a computer program) can perform the step of correlating the presence or absence of a polymorphism in an individual with an altered (*e.g.*, increased or decreased) risk (or no altered risk) for liver fibrosis for the individual. Computer software can also perform the step of correlating the presence or absence of a polymorphism in an individual with the predicted response of the invididual to a drug such as a statin.

*Reports, Transmission of Reports, Programmed Computers, and Business Methods*

The results of a test (*e.g.*, an individual's risk for liver fibrosis, or an individual's predicted drug responsiveness, based on assaying one or more SNPs disclosed herein, and/or an

individual's allele(s)/genotype at one or more SNPs disclosed herein, etc.), and/or any other information pertaining to a test, may be referred to herein as a "report". A tangible report can optionally be generated as part of a testing process (which may be interchangeably referred to herein as "reporting", or as "providing" a report, "producing" a report, or "generating" a report).

Examples of tangible reports may include, but are not limited to, reports in paper (such as computer-generated printouts of test results) or equivalent formats and reports stored on computer readable medium (such as a CD, USB flash drive or other removable storage device, computer hard drive, or computer network server, etc.). Reports, particularly those stored on computer readable medium, can be part of a database, which may optionally be accessible via the internet (such as a database of patient records or genetic information stored on a computer network server, which may be a "secure database" that has security features that limit access to the report, such as to allow only the patient and the patient's medical practioners to view the report while preventing other unauthorized individuals from viewing the report, for example). In addition to, or as an alternative to, generating a tangible report, reports can also be displayed on a computer screen (or the display of another electronic device or instrument).

A report can include, for example, an individual's risk for liver fibrosis, or may just include the allele(s)/genotype that an individual carries at one or more SNPs disclosed herein, which may optionally be linked to information regarding the significance of having the allele(s)/genotype at the SNP (for example, a report on computer readable medium such as a network server may include hyperlink(s) to one or more journal publications or websites that describe the medical/biological implications, such as increased or decreased disease risk, for individuals having a certain allele/genotype at the SNP). Thus, for example, the report can include disease risk or other medical/biological significance (e.g., drug responsiveness, etc.) as well as optionally also including the allele/genotype information, or the report may just include allele/genotype information without including disease risk or other medical/biological significance (such that an individual viewing the report can use the allele/genotype information to determine the associated disease risk or other medical/biological significance from a source outside of the report itself, such as from a medical practioner, publication, website, etc., which may optionally be linked to the report such as by a hyperlink).

A report can further be "transmitted" or "communicated" (these terms may be used herein interchangeably), such as to the individual who was tested, a medical practitioner (*e.g.*, a doctor, nurse, clinical laboratory practitioner, genetic counselor, etc.), a healthcare organization, a clinical laboratory, and/or any other party or requester intended to view or possess the report. The act of "transmitting" or "communicating" a report can be by any means known in the art, based

on the format of the report. Furthermore, "transmitting" or "communicating" a report can include delivering a report ("pushing") and/or retrieving ("pulling") a report. For example, reports can be transmitted/communicated by various means, including being physically transferred between parties (such as for reports in paper format) such as by being physically delivered from one party to another, or by being transmitted electronically or in signal form (*e.g.*, via e-mail or over the internet, by facsimile, and/or by any wired or wireless communication methods known in the art) such as by being retrieved from a database stored on a computer network server, etc.

In certain exemplary embodiments, the invention provides computers (or other apparatus/devices such as biomedical devices or laboratory instrumentation) programmed to carry out the methods described herein. For example, in certain embodiments, the invention provides a computer programmed to receive (*i.e.*, as input) the identity (*e.g.*, the allele(s) or genotype at a SNP) of one or more SNPs disclosed herein and provide (*i.e.*, as output) the disease risk (*e.g.*, an individual's risk for liver fibrosis) or other result (*e.g.*, disease diagnosis or prognosis, drug responsiveness, *etc.*) based on the identity of the SNP(s). Such output (*e.g.*, communication of disease risk, disease diagnosis or prognosis, drug responsiveness, *etc.*) may be, for example, in the form of a report on computer readable medium, printed in paper form, and/or displayed on a computer screen or other display.

In various exemplary embodiments, the invention further provides methods of doing business (with respect to methods of doing business, the terms "individual" and "customer" are used herein interchangeably). For example, exemplary methods of doing business can comprise assaying one or more SNPs disclosed herein and providing a report that includes, for example, a customer's risk for liver fibrosis (based on which allele(s)/genotype is present at the assayed SNP(s)) and/or that includes the allele(s)/genotype at the assayed SNP(s) which may optionally be linked to information (*e.g.*, journal publications, websites, *etc.*) pertaining to disease risk or other biological/medical significance such as by means of a hyperlink (the report may be provided, for example, on a computer network server or other computer readable medium that is internet-accessible, and the report may be included in a secure database that allows the customer to access their report while preventing other unauthorized individuals from viewing the report), and optionally transmitting the report. Customers (or another party who is associated with the customer, such as the customer's doctor, for example) can request/order (*e.g.*, purchase) the test online via the internet (or by phone, mail order, at an outlet/store, *etc.*), for example, and a kit can be sent/delivered (or otherwise provided) to the customer (or another party on behalf of the customer, such as the customer's doctor, for example) for collection of a biological sample from the customer (*e.g.*, a buccal swab for collecting buccal cells), and the customer (or a party who

collects the customer's biological sample) can submit their biological samples for assaying (*e.g.*, to a laboratory or party associated with the laboratory such as a party that accepts the customer samples on behalf of the laboratory, a party for whom the laboratory is under the control of (*e.g.*, the laboratory carries out the assays by request of the party or under a contract with the party, for example), and/or a party that receives at least a portion of the customer's payment for the test). The report (*e.g.*, results of the assay including, for example, the customer's disease risk and/or allele(s)/genotype at the assayed SNP(s)) may be provided to the customer by, for example, the laboratory that assays the SNP(s) or a party associated with the laboratory (*e.g.*, a party that receives at least a portion of the customer's payment for the assay, or a party that requests the laboratory to carry out the assays or that contracts with the laboratory for the assays to be carried out) or a doctor or other medical practitioner who is associated with (*e.g.*, employed by or having a consulting or contracting arrangement with) the laboratory or with a party associated with the laboratory, or the report may be provided to a third party (*e.g.*, a doctor, genetic counselor, hospital, *etc.*) which optionally provides the report to the customer. In further embodiments, the customer may be a doctor or other medical practitioner, or a hospital, laboratory, medical insurance organization, or other medical organization that requests/orders (*e.g.*, purchases) tests for the purposes of having other individuals (*e.g.*, their patients or customers) assayed for one or more SNPs disclosed herein and optionally obtaining a report of the assay results.

In certain exemplary methods of doing business, kits for collecting a biological sample from a customer (*e.g.*, a buccal swab for collecting buccal cells) are provided (*e.g.*, for sale), such as at an outlet (*e.g.*, a drug store, pharmacy, general merchandise store, or any other desirable outlet), online via the internet, by mail order, *etc.*, whereby customers can obtain (*e.g.*, purchase) the kits, collect their own biological samples, and submit (*e.g.*, send/deliver via mail) their samples to a laboratory which assays the samples for one or more SNPs disclosed herein (such as to determine the customer's risk for liver fibrosis) and optionally provides a report to the customer (of the customer's disease risk based on their SNP genotype(s), for example) or provides the results of the assay to another party (*e.g.*, a doctor, genetic counselor, hospital, *etc.*) which optionally provides a report to the customer (of the customer's disease risk based on their SNP genotype(s), for example).

## ISOLATED NUCLEIC ACID MOLECULES AND SNP DETECTION REAGENTS & KITS

Tables 1 and 2 provide a variety of information about each SNP of the present invention that is associated with liver fibrosis, including the transcript sequences (SEQ ID NOS:1-16),

genomic sequence (SEQ ID NO:65-90), and protein sequences (SEQ ID NOS:17-32) of the encoded gene products (with the SNPs indicated by IUB codes in the nucleic acid sequences). In addition, Tables 1 and 2 include SNP context sequences, which generally include 100 nucleotide upstream (5') plus 100 nucleotides downstream (3') of each SNP position (SEQ ID NOS:33-64 correspond to transcript-based SNP context sequences disclosed in Table 1, and SEQ ID NOS:91-358 correspond to genomic-based context sequences disclosed in Table 2), the alternative nucleotides (alleles) at each SNP position, and additional information about the variant where relevant, such as SNP type (coding, missense, splice site, UTR, etc.), human populations in which the SNP was observed, observed allele frequencies, information about the encoded protein, etc.

**Isolated Nucleic Acid Molecules**

In exemplary embodiments, the present invention provides isolated nucleic acid molecules that contain one or more SNPs disclosed herein, such as in any of Tables 1-20 and/or in the Examples sections below. Isolated nucleic acid molecules containing one or more SNPs disclosed herein, such as in any one of Tables 1-20 and/or in the Examples sections below, may be interchangeably referred to throughout the present text as "SNP-containing nucleic acid molecules". Isolated nucleic acid molecules may optionally encode a full-length variant protein or fragment thereof. The isolated nucleic acid molecules of the present invention also include probes and primers (which are described in greater detail below in the section entitled "SNP Detection Reagents"), which may be used for assaying the disclosed SNPs, and isolated full-length genes, transcripts, cDNA molecules, and fragments thereof, which may be used for such purposes as expressing an encoded protein.

As used herein, an "isolated nucleic acid molecule" generally is one that contains a SNP of the present invention or one that hybridizes to such molecule such as a nucleic acid with a complementary sequence, and is separated from most other nucleic acids present in the natural source of the nucleic acid molecule. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule containing a SNP of the present invention, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered "isolated". Nucleic acid molecules present in non-human transgenic animals, which do not naturally occur in the animal, are also considered "isolated". For example, recombinant DNA molecules contained in a vector are considered "isolated". Further examples of "isolated" DNA molecules include recombinant DNA molecules

maintained in heterologous host cells, and purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of the isolated SNP-containing DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Generally, an isolated SNP-containing nucleic acid molecule comprises one or more SNP positions disclosed by the present invention with flanking nucleotide sequences on either side of the SNP positions. A flanking sequence can include nucleotide residues that are naturally associated with the SNP site and/or heterologous nucleotide sequences. Preferably the flanking sequence is up to about 500, 300, 100, 60, 50, 30, 25, 20, 15, 10, 8, or 4 nucleotides (or any other length in-between) on either side of a SNP position, or as long as the full-length gene or entire protein-coding sequence (or any portion thereof such as an exon), especially if the SNP-containing nucleic acid molecule is to be used to produce a protein or protein fragment.

For full-length genes and entire protein-coding sequences, a SNP flanking sequence can be, for example, up to about 5KB, 4KB, 3KB, 2KB, 1KB on either side of the SNP. Furthermore, in such instances, the isolated nucleic acid molecule comprises exonic sequences (including protein-coding and/or non-coding exonic sequences), but may also include intronic sequences. Thus, any protein coding sequence may be either contiguous or separated by introns. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences and is of appropriate length such that it can be subjected to the specific manipulations or uses described herein such as recombinant protein expression, preparation of probes and primers for assaying the SNP position, and other uses specific to the SNP-containing nucleic acid sequences.

An isolated SNP-containing nucleic acid molecule can comprise, for example, a full-length gene or transcript, such as a gene isolated from genomic DNA (*e.g.*, by cloning or PCR amplification), a cDNA molecule, or an mRNA transcript molecule. Polymorphic transcript sequences are provided in Table 1 and in the Sequence Listing (SEQ ID NOS:1-16), and polymorphic genomic sequence is provided in Table 2 and in the Sequence Listing (SEQ ID NO:65-90). Furthermore, fragments of such full-length genes and transcripts that contain one or more SNPs disclosed herein are also encompassed by the present invention, and such fragments may be used, for example, to express any part of a protein, such as a particular functional domain or an antigenic epitope.

Thus, the present invention also encompasses fragments of the nucleic acid sequences provided in Tables 1-2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-16, genomic sequence is provided in Table 2 as SEQ ID NO:65-90, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NOS:33-64, and genomic-based SNP context sequences are

provided in Table 2 as SEQ ID NOS:91-358) and their complements. A fragment typically comprises a contiguous nucleotide sequence at least about 8 or more nucleotides, more preferably at least about 12 or more nucleotides, and even more preferably at least about 16 or more nucleotides. Further, a fragment could comprise at least about 18, 20, 22, 25, 30, 40, 50, 60, 80, 100, 150, 200, 250 or 500 (or any other number in-between) nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope-bearing regions of a variant peptide or regions of a variant peptide that differ from the normal/wild-type protein, or can be useful as a polynucleotide probe or primer. Such fragments can be isolated using the nucleotide sequences provided in Table 1 and/or Table 2 for the synthesis of a polynucleotide probe. A labeled probe can then be used, for example, to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in amplification reactions, such as for purposes of assaying one or more SNPs sites or for cloning specific regions of a gene.

An isolated nucleic acid molecule of the present invention further encompasses a SNP-containing polynucleotide that is the product of any one of a variety of nucleic acid amplification methods, which are used to increase the copy numbers of a polynucleotide of interest in a nucleic acid sample. Such amplification methods are well known in the art, and they include but are not limited to, polymerase chain reaction (PCR) (U.S. Patent Nos. 4,683,195; and 4,683,202; *PCR Technology: Principles and Applications for DNA Amplification*, ed. H.A. Erlich, Freeman Press, NY, NY, 1992), ligase chain reaction (LCR) (Wu and Wallace, *Genomics* 4:560, 1989; Landegren *et al.*, *Science* 241:1077, 1988), strand displacement amplification (SDA) (U.S. Patent Nos. 5,270,184; and 5,422,252), transcription-mediated amplification (TMA) (U.S. Patent No. 5,399,491), linked linear amplification (LLA) (U.S. Patent No. 6,027,923), and the like, and isothermal amplification methods such as nucleic acid sequence based amplification (NASBA), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Natl. Acad. Sci. USA* 87: 1874, 1990). Based on such methodologies, a person skilled in the art can readily design primers in any suitable regions 5' and 3' to a SNP disclosed herein. Such primers may be used to amplify DNA of any length so long that it contains the SNP of interest in its sequence.

As used herein, an "amplified polynucleotide" of the invention is a SNP-containing nucleic acid molecule whose amount has been increased at least two fold by any nucleic acid amplification method performed *in vitro* as compared to its starting amount in a test sample. In other preferred embodiments, an amplified polynucleotide is the result of at least ten fold, fifty fold, one hundred fold, one thousand fold, or even ten thousand fold increase as compared to its starting amount in a test sample. In a typical PCR amplification, a polynucleotide of interest is

often amplified at least fifty thousand fold in amount over the unamplified genomic DNA, but the precise amount of amplification needed for an assay depends on the sensitivity of the subsequent detection method used.

Generally, an amplified polynucleotide is at least about 16 nucleotides in length. More typically, an amplified polynucleotide is at least about 20 nucleotides in length. In a preferred embodiment of the invention, an amplified polynucleotide is at least about 30 nucleotides in length. In a more preferred embodiment of the invention, an amplified polynucleotide is at least about 32, 40, 45, 50, or 60 nucleotides in length. In yet another preferred embodiment of the invention, an amplified polynucleotide is at least about 100, 200, 300, 400, or 500 nucleotides in length. While the total length of an amplified polynucleotide of the invention can be as long as an exon, an intron or the entire gene where the SNP of interest resides, an amplified product is typically up to about 1,000 nucleotides in length (although certain amplification methods may generate amplified products greater than 1000 nucleotides in length). More preferably, an amplified polynucleotide is not greater than about 600-700 nucleotides in length. It is understood that irrespective of the length of an amplified polynucleotide, a SNP of interest may be located anywhere along its sequence.

In a specific embodiment of the invention, the amplified product is at least about 201 nucleotides in length, comprises one of the transcript-based context sequences or the genomic-based context sequences shown in Tables 1-2. Such a product may have additional sequences on its 5' end or 3' end or both. In another embodiment, the amplified product is about 101 nucleotides in length, and it contains a SNP disclosed herein. Preferably, the SNP is located at the middle of the amplified product (*e.g.*, at position 101 in an amplified product that is 201 nucleotides in length, or at position 51 in an amplified product that is 101 nucleotides in length), or within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, or 20 nucleotides from the middle of the amplified product (however, as indicated above, the SNP of interest may be located anywhere along the length of the amplified product).

The present invention provides isolated nucleic acid molecules that comprise, consist of, or consist essentially of one or more polynucleotide sequences that contain one or more SNPs disclosed herein, complements thereof, and SNP-containing fragments thereof.

Accordingly, the present invention provides nucleic acid molecules that consist of any of the nucleotide sequences shown in Table 1 and/or Table 2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-16, genomic sequence is provided in Table 2 as SEQ ID NO:65-90, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NOS:33-64, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NOS:91-358), or any nucleic acid

molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:17-32). A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of any of the nucleotide sequences shown in Table 1 and/or Table 2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-16, genomic sequence is provided in Table 2 as SEQ ID NO:65-90, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NOS:33-64, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NOS:91-358), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:17-32). A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleotide residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise any of the nucleotide sequences shown in Table 1 and/or Table 2 or a SNP-containing fragment thereof (transcript sequences are provided in Table 1 as SEQ ID NOS:1-16, genomic sequence is provided in Table 2 as SEQ ID NO:65-90, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NOS:33-64, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NOS:91-358), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:17-32). A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleotide residues, such as residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have one to a few additional nucleotides or can comprise many more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made and isolated is provided below, and such techniques are well known to those of ordinary skill in the art (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY).

The isolated nucleic acid molecules can encode mature proteins plus additional amino or carboxyl-terminal amino acids or both, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life, or facilitate manipulation of a protein for assay or production. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

Thus, the isolated nucleic acid molecules include, but are not limited to, nucleic acid molecules having a sequence encoding a peptide alone, a sequence encoding a mature peptide and additional coding sequences such as a leader or secretory sequence (*e.g.*, a pre-pro or pro-protein sequence), a sequence encoding a mature peptide with or without additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but untranslated sequences that play a role in, for example, transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding, and/or stability of mRNA. In addition, the nucleic acid molecules may be fused to heterologous marker sequences encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA, which may be obtained, for example, by molecular cloning or produced by chemical synthetic techniques or by a combination thereof (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY). Furthermore, isolated nucleic acid molecules, particularly SNP detection reagents such as probes and primers, can also be partially or completely in the form of one or more types of nucleic acid analogs, such as peptide nucleic acid (PNA) (U.S. Patent Nos. 5,539,082; 5,527,675; 5,623,049; 5,714,331). The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the complementary non-coding strand (anti-sense strand). DNA, RNA, or PNA segments can be assembled, for example, from fragments of the human genome (in the case of DNA or RNA) or single nucleotides, short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic nucleic acid molecule. Nucleic acid molecules can be readily synthesized using the sequences provided herein as a reference; oligonucleotide and PNA oligomer synthesis techniques are well known in the art (see, *e.g.*, Corey, "Peptide nucleic acids: expanding the scope of nucleic acid recognition", *Trends Biotechnol.* 1997 Jun;15(6):224-9, and Hyrup et al., "Peptide nucleic acids (PNA): synthesis, properties and potential applications", *Bioorg Med Chem.* 1996 Jan;4(1):5-23). Furthermore, large-scale automated oligonucleotide/PNA synthesis (including synthesis on an array or bead surface or other solid support) can readily be accomplished using commercially available nucleic acid synthesizers, such as the Applied Biosystems (Foster City, CA) 3900 High-Throughput DNA Synthesizer or Expedite 8909 Nucleic Acid Synthesis System, and the sequence information provided herein.

The present invention encompasses nucleic acid analogs that contain modified, synthetic, or non-naturally occurring nucleotides or structural elements or other alternative/modified nucleic acid chemistries known in the art. Such nucleic acid analogs are useful, for example, as

detection reagents (*e.g.*, primers/probes) for detecting one or more SNPs identified in Table 1 and/or Table 2. Furthermore, kits/systems (such as beads, arrays, etc.) that include these analogs are also encompassed by the present invention. For example, PNA oligomers that are based on the polymorphic sequences of the present invention are specifically contemplated. PNA oligomers are analogs of DNA in which the phosphate backbone is replaced with a peptide-like backbone (Lagriffoul *et al.*, Bioorganic & Medicinal Chemistry Letters, 4: 1081-1082 (1994), Petersen *et al.*, Bioorganic & Medicinal Chemistry Letters, 6: 793-796 (1996), Kumar *et al.*, Organic Letters 3(9): 1269-1272 (2001), WO96/04000). PNA hybridizes to complementary RNA or DNA with higher affinity and specificity than conventional oligonucleotides and oligonucleotide analogs. The properties of PNA enable novel molecular biology and biochemistry applications unachievable with traditional oligonucleotides and peptides.

Additional examples of nucleic acid modifications that improve the binding properties and/or stability of a nucleic acid include the use of base analogs such as inosine, intercalators (U.S. Patent No. 4,835,263) and the minor groove binders (U.S. Patent No. 5,801,115). Thus, references herein to nucleic acid molecules, SNP-containing nucleic acid molecules, SNP detection reagents (*e.g.*, probes and primers), oligonucleotides/polynucleotides include PNA oligomers and other nucleic acid analogs. Other examples of nucleic acid analogs and alternative/modified nucleic acid chemistries known in the art are described in *Current Protocols in Nucleic Acid Chemistry*, John Wiley & Sons, N.Y. (2002).

The present invention further provides nucleic acid molecules that encode fragments of the variant polypeptides disclosed herein as well as nucleic acid molecules that encode obvious variants of such variant polypeptides. Such nucleic acid molecules may be naturally occurring, such as paralogs (different locus) and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, the variants can contain nucleotide substitutions, deletions, inversions and insertions (in addition to the SNPs disclosed in Tables 1-2). Variation can occur in either or both the coding and non-coding regions. The variations can produce conservative and/or non-conservative amino acid substitutions.

Further variants of the nucleic acid molecules disclosed in Tables 1-2, such as naturally occurring allelic variants (as well as orthologs and paralogs) and synthetic variants produced by mutagenesis techniques, can be identified and/or produced using methods well known in the art. Such further variants can comprise a nucleotide sequence that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with a nucleic

acid sequence disclosed in Table 1 and/or Table 2 (or a fragment thereof) and that includes a novel SNP allele disclosed in Table 1 and/or Table 2. Further, variants can comprise a nucleotide sequence that encodes a polypeptide that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with a polypeptide sequence disclosed in Table 1 (or a fragment thereof) and that includes a novel SNP allele disclosed in Table 1 and/or Table 2. Thus, an aspect of the present invention that is specifically contemplated are isolated nucleic acid molecules that have a certain degree of sequence variation compared with the sequences shown in Tables 1-2, but that contain a novel SNP allele disclosed herein. In other words, as long as an isolated nucleic acid molecule contains a novel SNP allele disclosed herein, other portions of the nucleic acid molecule that flank the novel SNP allele can vary to some degree from the specific transcript, genomic, and context sequences shown in Tables 1-2, and can encode a polypeptide that varies to some degree from the specific polypeptide sequences shown in Table 1.

To determine the percent identity of two amino acid sequences or two nucleotide sequences of two molecules that share sequence homology, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a

preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch algorithm (*J. Mol. Biol.* (48):444-453 (1970)) which has been incorporated into the GAP program in the GCG software package, using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res. 12(1)*:387 (1984)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4.

The nucleotide and amino acid sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. In addition to BLAST, examples of other search and sequence comparison programs used in the art include, but are not limited to, FASTA (Pearson, *Methods Mol. Biol.* 25, 365-389 (1994)) and KERR (Dufresne *et al.*, *Nat Biotechnol* 2002 Dec;20(12):1269-71). For further information regarding bioinformatics techniques, see *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., N.Y.

The present invention further provides non-coding fragments of the nucleic acid molecules disclosed in Table 1 and/or Table 2. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, intronic sequences, 5' untranslated regions (UTRs), 3' untranslated regions, gene modulating sequences and gene termination

sequences. Such fragments are useful, for example, in controlling heterologous gene expression and in developing screens to identify gene-modulating agents.

**SNP Detection Reagents**

In a specific aspect of the present invention, the SNPs disclosed in Table 1 and/or Table 2, and their associated transcript sequences (provided in Table 1 as SEQ ID NOS:1-16), genomic sequence (provided in Table 2 as SEQ ID NO:65-90), and context sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:33-64; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:91-358), can be used for the design of SNP detection reagents. As used herein, a "SNP detection reagent" is a reagent that specifically detects a specific target SNP position disclosed herein, and that is preferably specific for a particular nucleotide (allele) of the target SNP position (*i.e.*, the detection reagent preferably can differentiate between different alternative nucleotides at a target SNP position, thereby allowing the identity of the nucleotide present at the target SNP position to be determined). Typically, such detection reagent hybridizes to a target SNP-containing nucleic acid molecule by complementary base-pairing in a sequence specific manner, and discriminates the target variant sequence from other nucleic acid sequences such as an art-known form in a test sample. An example of a detection reagent is a probe that hybridizes to a target nucleic acid containing one or more of the SNPs provided in Table 1 and/or Table 2. In a preferred embodiment, such a probe can differentiate between nucleic acids having a particular nucleotide (allele) at a target SNP position from other nucleic acids that have a different nucleotide at the same target SNP position. In addition, a detection reagent may hybridize to a specific region 5' and/or 3' to a SNP position, particularly a region corresponding to the context sequences provided in Table 1 and/or Table 2 (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:33-64; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:91-358). Another example of a detection reagent is a primer which acts as an initiation point of nucleotide extension along a complementary strand of a target polynucleotide. The SNP sequence information provided herein is also useful for designing primers, *e.g.* allele-specific primers, to amplify (*e.g.*, using PCR) any SNP of the present invention.

In one preferred embodiment of the invention, a SNP detection reagent is an isolated or synthetic DNA or RNA polynucleotide probe or primer or PNA oligomer, or a combination of DNA, RNA and/or PNA, that hybridizes to a segment of a target nucleic acid molecule containing a SNP identified in Table 1 and/or Table 2. A detection reagent in the form of a polynucleotide may optionally contain modified base analogs, intercalators or minor groove binders. Multiple detection reagents such as probes may be, for example, affixed to a solid

support (*e.g.*, arrays or beads) or supplied in solution (*e.g.*, probe/primer sets for enzymatic reactions such as PCR, RT-PCR, TaqMan assays, or primer-extension reactions) to form a SNP detection kit.

A probe or primer typically is a substantially purified oligonucleotide or PNA oligomer. Such oligonucleotide typically comprises a region of complementary nucleotide sequence that hybridizes under stringent conditions to at least about 8, 10, 12, 16, 18, 20, 22, 25, 30, 40, 50, 55, 60, 65, 70, 80, 90, 100, 120 (or any other number in-between) or more consecutive nucleotides in a target nucleic acid molecule. Depending on the particular assay, the consecutive nucleotides can either include the target SNP position, or be a specific region in close enough proximity 5' and/or 3' to the SNP position to carry out the desired assay.

Other preferred primer and probe sequences can readily be determined using the transcript sequences (SEQ ID NOS:1-16), genomic sequence (SEQ ID NO:65-90), and SNP context sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:33-64; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:91-358) disclosed in the Sequence Listing and in Tables 1-2. It will be apparent to one of skill in the art that such primers and probes are directly useful as reagents for genotyping the SNPs of the present invention, and can be incorporated into any kit/system format.

In order to produce a probe or primer specific for a target SNP-containing sequence, the gene/transcript and/or context sequence surrounding the SNP of interest is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene/SNP context sequence, have a GC content within a range suitable for hybridization, lack predicted secondary structure that may interfere with hybridization, and/or possess other desired characteristics or that lack other undesired characteristics.

A primer or probe of the present invention is typically at least about 8 nucleotides in length. In one embodiment of the invention, a primer or a probe is at least about 10 nucleotides in length. In a preferred embodiment, a primer or a probe is at least about 12 nucleotides in length. In a more preferred embodiment, a primer or probe is at least about 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length. While the maximal length of a probe can be as long as the target sequence to be detected, depending on the type of assay in which it is employed, it is typically less than about 50, 60, 65, or 70 nucleotides in length. In the case of a primer, it is typically less than about 30 nucleotides in length. In a specific preferred embodiment of the invention, a primer or a probe is within the length of about 18 and about 28 nucleotides. However, in other embodiments, such as nucleic acid arrays and other embodiments in which

probes are affixed to a substrate, the probes can be longer, such as on the order of 30-70, 75, 80, 90, 100, or more nucleotides in length (see the section below entitled "SNP Detection Kits and Systems").

For analyzing SNPs, it may be appropriate to use oligonucleotides specific for alternative SNP alleles. Such oligonucleotides which detect single nucleotide variations in target sequences may be referred to by such terms as "allele-specific oligonucleotides", "allele-specific probes", or "allele-specific primers". The design and use of allele-specific probes for analyzing polymorphisms is described in, *e.g.*, *Mutation Detection A Practical Approach,* ed. Cotton *et al.* Oxford University Press, 1998; *Saiki et al.*, *Nature* 324, 163-166 (1986); Dattagupta, EP235,726; and Saiki, WO 89/11548.

While the design of each allele-specific primer or probe depends on variables such as the precise composition of the nucleotide sequences flanking a SNP position in a target nucleic acid molecule, and the length of the primer or probe, another factor in the use of primers and probes is the stringency of the condition under which the hybridization between the probe or primer and the target sequence is performed. Higher stringency conditions utilize buffers with lower ionic strength and/or a higher reaction temperature, and tend to require a more perfect match between probe/primer and a target sequence in order to form a stable duplex. If the stringency is too high, however, hybridization may not occur at all. In contrast, lower stringency conditions utilize buffers with higher ionic strength and/or a lower reaction temperature, and permit the formation of stable duplexes with more mismatched bases between a probe/primer and a target sequence. By way of example and not limitation, exemplary conditions for high stringency hybridization conditions using an allele-specific probe are as follows: Prehybridization with a solution containing 5X standard saline phosphate EDTA (SSPE), 0.5% $NaDodSO_4$ (SDS) at 55°C, and incubating probe with target nucleic acid molecules in the same solution at the same temperature, followed by washing with a solution containing 2X SSPE, and 0.1%SDS at 55°C or room temperature.

Moderate stringency hybridization conditions may be used for allele-specific primer extension reactions with a solution containing, *e.g.*, about 50mM KCl at about 46°C. Alternatively, the reaction may be carried out at an elevated temperature such as 60°C. In another embodiment, a moderately stringent hybridization condition suitable for oligonucleotide ligation assay (OLA) reactions wherein two probes are ligated if they are completely complementary to the target sequence may utilize a solution of about 100mM KCl at a temperature of 46°C.

In a hybridization-based assay, allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms (*e.g.*, alternative SNP alleles/nucleotides) in the respective DNA segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant detectable difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles or significantly more strongly to one allele. While a probe may be designed to hybridize to a target sequence that contains a SNP site such that the SNP site aligns anywhere along the sequence of the probe, the probe is preferably designed to hybridize to a segment of the target sequence such that the SNP site aligns with a central position of the probe (*e.g.*, a position within the probe that is at least three nucleotides from either end of the probe). This design of probe generally achieves good discrimination in hybridization between different allelic forms.

In another embodiment, a probe or primer may be designed to hybridize to a segment of target DNA such that the SNP aligns with either the 5' most end or the 3' most end of the probe or primer. In a specific preferred embodiment which is particularly suitable for use in a oligonucleotide ligation assay (U.S. Patent No. 4,988,617), the 3'most nucleotide of the probe aligns with the SNP position in the target sequence.

Oligonucleotide probes and primers may be prepared by methods well known in the art. Chemical synthetic methods include, but are limited to, the phosphotriester method described by Narang *et al.*, 1979, Methods in Enzymology 68:90; the phosphodiester method described by Brown *et al.,* 1979, Methods in Enzymology 68:109, the diethylphosphoamidate method described by Beaucage *et al.*, 1981, *Tetrahedron Letters* 22:1859; and the solid support method described in U.S. Patent No. 4,458,066.

Allele-specific probes are often used in pairs (or, less commonly, in sets of 3 or 4, such as if a SNP position is known to have 3 or 4 alleles, respectively, or to assay both strands of a nucleic acid molecule for a target SNP allele), and such pairs may be identical except for a one nucleotide mismatch that represents the allelic variants at the SNP position. Commonly, one member of a pair perfectly matches a reference form of a target sequence that has a more common SNP allele (*i.e.*, the allele that is more frequent in the target population) and the other member of the pair perfectly matches a form of the target sequence that has a less common SNP allele (*i.e.*, the allele that is rarer in the target population). In the case of an array, multiple pairs of probes can be immobilized on the same support for simultaneous analysis of multiple different polymorphisms.

In one type of PCR-based assay, an allele-specific primer hybridizes to a region on a target nucleic acid molecule that overlaps a SNP position and only primes amplification of an allelic form to which the primer exhibits perfect complementarity (Gibbs, 1989, *Nucleic Acid Res.* 17 2427-2448). Typically, the primer's 3'-most nucleotide is aligned with and complementary to the SNP position of the target nucleic acid molecule. This primer is used in conjunction with a second primer that hybridizes at a distal site. Amplification proceeds from the two primers, producing a detectable product that indicates which allelic form is present in the test sample. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification or substantially reduces amplification efficiency, so that either no detectable product is formed or it is formed in lower amounts or at a slower pace. The method generally works most effectively when the mismatch is at the 3'-most position of the oligonucleotide (*i.e.*, the 3'-most position of the oligonucleotide aligns with the target SNP position) because this position is most destabilizing to elongation from the primer (see, *e.g.*, WO 93/22456). This PCR-based assay can be utilized as part of the TaqMan assay, described below.

In a specific embodiment of the invention, a primer of the invention contains a sequence substantially complementary to a segment of a target SNP-containing nucleic acid molecule except that the primer has a mismatched nucleotide in one of the three nucleotide positions at the 3'-most end of the primer, such that the mismatched nucleotide does not base pair with a particular allele at the SNP site. In a preferred embodiment, the mismatched nucleotide in the primer is the second from the last nucleotide at the 3'-most position of the primer. In a more preferred embodiment, the mismatched nucleotide in the primer is the last nucleotide at the 3'-most position of the primer.

In another embodiment of the invention, a SNP detection reagent of the invention is labeled with a fluorogenic reporter dye that emits a detectable signal. While the preferred reporter dye is a fluorescent dye, any reporter dye that can be attached to a detection reagent such as an oligonucleotide probe or primer is suitable for use in the invention. Such dyes include, but are not limited to, Acridine, AMCA, BODIPY, Cascade Blue, Cy2, Cy3, Cy5, Cy7, Dabcyl, Edans, Eosin, Erythrosin, Fluorescein, 6-Fam, Tet, Joe, Hex, Oregon Green, Rhodamine, Rhodol Green, Tamra, Rox, and Texas Red.

In yet another embodiment of the invention, the detection reagent may be further labeled with a quencher dye such as Tamra, especially when the reagent is used as a self-quenching probe such as a TaqMan (U.S. Patent Nos. 5,210,015 and 5,538,848) or Molecular Beacon probe (U.S. Patent Nos. 5,118,801 and 5,312,728), or other stemless or linear beacon probe (Livak *et al.*, 1995,

PCR Method Appl. 4:357-362; Tyagi *et al.*, 1996, Nature Biotechnology 14: 303-308; Nazarenko *et al.*, 1997, Nucl. Acids Res. 25:2516-2521; U.S. Patent Nos. 5,866,336 and 6,117,635).

The detection reagents of the invention may also contain other labels, including but not limited to, biotin for streptavidin binding, hapten for antibody binding, and oligonucleotide for binding to another complementary oligonucleotide such as pairs of zipcodes.

The present invention also contemplates reagents that do not contain (or that are complementary to) a SNP nucleotide identified herein but that are used to assay one or more SNPs disclosed herein. For example, primers that flank, but do not hybridize directly to a target SNP position provided herein are useful in primer extension reactions in which the primers hybridize to a region adjacent to the target SNP position (*i.e.*, within one or more nucleotides from the target SNP site). During the primer extension reaction, a primer is typically not able to extend past a target SNP site if a particular nucleotide (allele) is present at that target SNP site, and the primer extension product can be detected in order to determine which SNP allele is present at the target SNP site. For example, particular ddNTPs are typically used in the primer extension reaction to terminate primer extension once a ddNTP is incorporated into the extension product (a primer extension product which includes a ddNTP at the 3'-most end of the primer extension product, and in which the ddNTP is a nucleotide of a SNP disclosed herein, is a composition that is specifically contemplated by the present invention). Thus, reagents that bind to a nucleic acid molecule in a region adjacent to a SNP site and that are used for assaying the SNP site, even though the bound sequences do not necessarily include the SNP site itself, are also contemplated by the present invention.

**SNP Detection Kits and Systems**

A person skilled in the art will recognize that, based on the SNP and associated sequence information disclosed herein, detection reagents can be developed and used to assay any SNP of the present invention individually or in combination, and such detection reagents can be readily incorporated into one of the established kit or system formats which are well known in the art. The terms "kits" and "systems", as used herein in the context of SNP detection reagents, are intended to refer to such things as combinations of multiple SNP detection reagents, or one or more SNP detection reagents in combination with one or more other types of elements or components (*e.g.*, other types of biochemical reagents, containers, packages such as packaging intended for commercial sale, substrates to which SNP detection reagents are attached, electronic hardware components, etc.). Accordingly, the present invention further provides SNP detection kits and systems, including but not limited to, packaged probe and primer sets (*e.g.*, TaqMan

probe/primer sets), arrays/microarrays of nucleic acid molecules, and beads that contain one or more probes, primers, or other detection reagents for detecting one or more SNPs of the present invention. The kits/systems can optionally include various electronic hardware components; for example, arrays ("DNA chips") and microfluidic systems ("lab-on-a-chip" systems) provided by various manufacturers typically comprise hardware components. Other kits/systems (*e.g.*, probe/primer sets) may not include electronic hardware components, but may be comprised of, for example, one or more SNP detection reagents (along with, optionally, other biochemical reagents) packaged in one or more containers.

In some embodiments, a SNP detection kit typically contains one or more detection reagents and other components (*e.g.*, a buffer, enzymes such as DNA polymerases or ligases, chain extension nucleotides such as deoxynucleotide triphosphates, and in the case of Sanger-type DNA sequencing reactions, chain terminating nucleotides, positive control sequences, negative control sequences, and the like) necessary to carry out an assay or reaction, such as amplification and/or detection of a SNP-containing nucleic acid molecule. A kit may further contain means for determining the amount of a target nucleic acid, and means for comparing the amount with a standard, and can comprise instructions for using the kit to detect the SNP-containing nucleic acid molecule of interest. In one embodiment of the present invention, kits are provided which contain the necessary reagents to carry out one or more assays to detect one or more SNPs disclosed herein. In a preferred embodiment of the present invention, SNP detection kits/systems are in the form of nucleic acid arrays, or compartmentalized kits, including microfluidic/lab-on-a-chip systems.

SNP detection kits/systems may contain, for example, one or more probes, or pairs of probes, that hybridize to a nucleic acid molecule at or near each target SNP position. Multiple pairs of allele-specific probes may be included in the kit/system to simultaneously assay large numbers of SNPs, at least one of which is a SNP of the present invention. In some kits/systems, the allele-specific probes are immobilized to a substrate such as an array or bead. For example, the same substrate can comprise allele-specific probes for detecting at least 1; 10; 100; 1000; 10,000; 100,000 (or any other number in-between) or substantially all of the SNPs shown in Table 1 and/or Table 2.

The terms "arrays", "microarrays", and "DNA chips" are used herein interchangeably to refer to an array of distinct polynucleotides affixed to a substrate, such as glass, plastic, paper, nylon or other type of membrane, filter, chip, or any other suitable solid support. The polynucleotides can be synthesized directly on the substrate, or synthesized separate from the substrate and then affixed to the substrate. In one embodiment, the microarray is prepared and

used according to the methods described in U.S. Patent No. 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; *Nat. Biotech.* 14: 1675-1680) and Schena, M. *et al.* (1996; *Proc. Natl. Acad. Sci.* 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, U.S. Patent No. 5,807,522.

Nucleic acid arrays are reviewed in the following references: Zammatteo *et al.*, "New chips for molecular biology and diagnostics", *Biotechnol Annu Rev.* 2002;8:85-101; Sosnowski *et al.*, "Active microelectronic array system for DNA hybridization, genotyping and pharmacogenomic applications", *Psychiatr Genet.* 2002 Dec;12(4):181-92; Heller, "DNA microarray technology: devices, systems, and applications", *Annu Rev Biomed Eng.* 2002;4:129-53. Epub 2002 Mar 22; Kolchinsky *et al.*, "Analysis of SNPs and other genomic variations using gel-based chips", *Hum Mutat.* 2002 Apr;19(4):343-60; and McGall *et al.*, "High-density genechip oligonucleotide probe arrays", *Adv Biochem Eng Biotechnol.* 2002;77:21-42.

Any number of probes, such as allele-specific probes, may be implemented in an array, and each probe or pair of probes can hybridize to a different SNP position. In the case of polynucleotide probes, they can be synthesized at designated areas (or synthesized separately and then affixed to designated areas) on a substrate using a light-directed chemical process. Each DNA chip can contain, for example, thousands to millions of individual synthetic polynucleotide probes arranged in a grid-like pattern and miniaturized (*e.g.,* to the size of a dime). Preferably, probes are attached to a solid support in an ordered, addressable array.

A microarray can be composed of a large number of unique, single-stranded polynucleotides, usually either synthetic antisense polynucleotides or fragments of cDNAs, fixed to a solid support. Typical polynucleotides are preferably about 6-60 nucleotides in length, more preferably about 15-30 nucleotides in length, and most preferably about 18-25 nucleotides in length. For certain types of microarrays or other detection kits/systems, it may be preferable to use oligonucleotides that are only about 7-20 nucleotides in length. In other types of arrays, such as arrays used in conjunction with chemiluminescent detection technology, preferred probe lengths can be, for example, about 15-80 nucleotides in length, preferably about 50-70 nucleotides in length, more preferably about 55-65 nucleotides in length, and most preferably about 60 nucleotides in length. The microarray or detection kit can contain polynucleotides that cover the known 5' or 3' sequence of a gene/transcript or target SNP site, sequential polynucleotides that cover the full-length sequence of a gene/transcript; or unique polynucleotides selected from particular areas along the length of a target gene/transcript sequence, particularly areas corresponding to one or more SNPs disclosed in Table 1 and/or

Table 2. Polynucleotides used in the microarray or detection kit can be specific to a SNP or SNPs of interest (*e.g.*, specific to a particular SNP allele at a target SNP site, or specific to particular SNP alleles at multiple different SNP sites), or specific to a polymorphic gene/transcript or genes/transcripts of interest.

Hybridization assays based on polynucleotide arrays rely on the differences in hybridization stability of the probes to perfectly matched and mismatched target sequence variants. For SNP genotyping, it is generally preferable that stringency conditions used in hybridization assays are high enough such that nucleic acid molecules that differ from one another at as little as a single SNP position can be differentiated (*e.g.*, typical SNP hybridization assays are designed so that hybridization will occur only if one particular nucleotide is present at a SNP position, but will not occur if an alternative nucleotide is present at that SNP position). Such high stringency conditions may be preferable when using, for example, nucleic acid arrays of allele-specific probes for SNP detection. Such high stringency conditions are described in the preceding section, and are well known to those skilled in the art and can be found in, for example, *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

In other embodiments, the arrays are used in conjunction with chemiluminescent detection technology. The following patents and patent applications, which are all hereby incorporated by reference, provide additional information pertaining to chemiluminescent detection: U.S. patent applications 10/620332 and 10/620333 describe chemiluminescent approaches for microarray detection; U.S. Patent Nos. 6124478, 6107024, 5994073, 5981768, 5871938, 5843681, 5800999, and 5773628 describe methods and compositions of dioxetane for performing chemiluminescent detection; and U.S. published application US2002/0110828 discloses methods and compositions for microarray controls.

In one embodiment of the invention, a nucleic acid array can comprise an array of probes of about 15-25 nucleotides in length. In further embodiments, a nucleic acid array can comprise any number of probes, in which at least one probe is capable of detecting one or more SNPs disclosed in Table 1 and/or Table 2, and/or at least one probe comprises a fragment of one of the sequences selected from the group consisting of those disclosed in Table 1, Table 2, the Sequence Listing, and sequences complementary thereto, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 12, 15, 16, 18, 20, more preferably 22, 25, 30, 40, 47, 50, 55, 60, 65, 70, 80, 90, 100, or more consecutive nucleotides (or any other number in-between) and containing (or being complementary to) a novel SNP allele disclosed in Table 1 and/or Table 2. In some embodiments, the nucleotide complementary to the SNP site is within 5, 4, 3, 2, or 1 nucleotide from the center of the probe, more preferably at the center of said probe.

A polynucleotide probe can be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more polynucleotides, or any other number which lends itself to the efficient use of commercially available instrumentation.

Using such arrays or other kits/systems, the present invention provides methods of identifying the SNPs disclosed herein in a test sample. Such methods typically involve incubating a test sample of nucleic acids with an array comprising one or more probes corresponding to at least one SNP position of the present invention, and assaying for binding of a nucleic acid from the test sample with one or more of the probes. Conditions for incubating a SNP detection reagent (or a kit/system that employs one or more such SNP detection reagents) with a test sample vary. Incubation conditions depend on such factors as the format employed in the assay, the detection methods employed, and the type and nature of the detection reagents used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification and array assay formats can readily be adapted to detect the SNPs disclosed herein.

A SNP detection kit/system of the present invention may include components that are used to prepare nucleic acids from a test sample for the subsequent amplification and/or detection of a SNP-containing nucleic acid molecule. Such sample preparation components can be used to produce nucleic acid extracts (including DNA and/or RNA), proteins or membrane extracts from any bodily fluids (such as blood, serum, plasma, urine, saliva, phlegm, gastric juices, semen, tears, sweat, etc.), skin, hair, cells (especially nucleated cells), biopsies, buccal swabs or tissue specimens. The test samples used in the above-described methods will vary based on such factors as the assay format, nature of the detection method, and the specific tissues, cells or extracts used as the test sample to be assayed. Methods of preparing nucleic acids, proteins, and cell extracts are well known in the art and can be readily adapted to obtain a sample that is compatible with the system utilized. Automated sample preparation systems for extracting nucleic acids from a test sample are commercially available, and examples are Qiagen's BioRobot 9600, Applied Biosystems' PRISM™ 6700 sample preparation system, and Roche Molecular Systems' COBAS AmpliPrep System.

Another form of kit contemplated by the present invention is a compartmentalized kit. A compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include, for example, small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allow one to efficiently transfer reagents from one compartment to another compartment such that the test samples and reagents are not cross-contaminated, or from one container to another vessel not included in the kit, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another or to another vessel. Such containers may include, for example, one or more containers which will accept the test sample, one or more containers which contain at least one probe or other SNP detection reagent for detecting one or more SNPs of the present invention, one or more containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and one or more containers which contain the reagents used to reveal the presence of the bound probe or other SNP detection reagents. The kit can optionally further comprise compartments and/or reagents for, for example, nucleic acid amplification or other enzymatic reactions such as primer extension reactions, hybridization, ligation, electrophoresis (preferably capillary electrophoresis), mass spectrometry, and/or laser-induced fluorescent detection. The kit may also include instructions for using the kit. Exemplary compartmentalized kits include microfluidic devices known in the art (see, *e.g.*, Weigl *et al.*, "Lab-on-a-chip for drug development", *Adv Drug Deliv Rev.* 2003 Feb 24;55(3):349-77). In such microfluidic devices, the containers may be referred to as, for example, microfluidic "compartments", "chambers", or "channels".

Microfluidic devices, which may also be referred to as "lab-on-a-chip" systems, biomedical micro-electro-mechanical systems (bioMEMs), or multicomponent integrated systems, are exemplary kits/systems of the present invention for analyzing SNPs. Such systems miniaturize and compartmentalize processes such as probe/target hybridization, nucleic acid amplification, and capillary electrophoresis reactions in a single functional device. Such microfluidic devices typically utilize detection reagents in at least one aspect of the system, and such detection reagents may be used to detect one or more SNPs of the present invention. One example of a microfluidic system is disclosed in U.S. Patent No. 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips. Exemplary microfluidic systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples may be controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage can be used as a means to control the liquid flow at intersections between the micro-machined channels and

to change the liquid flow rate for pumping across different sections of the microchip. See, for example, U.S. Patent Nos. 6,153,073, Dubrow *et al.*, and 6,156,181, Parce *et al.*

For genotyping SNPs, an exemplary microfluidic system may integrate, for example, nucleic acid amplification, primer extension, capillary electrophoresis, and a detection method such as laser induced fluorescence detection. In a first step of an exemplary process for using such an exemplary system, nucleic acid samples are amplified, preferably by PCR. Then, the amplification products are subjected to automated primer extension reactions using ddNTPs (specific fluorescence for each ddNTP) and the appropriate oligonucleotide primers to carry out primer extension reactions which hybridize just upstream of the targeted SNP. Once the extension at the 3' end is completed, the primers are separated from the unincorporated fluorescent ddNTPs by capillary electrophoresis. The separation medium used in capillary electrophoresis can be, for example, polyacrylamide, polyethyleneglycol or dextran. The incorporated ddNTPs in the single nucleotide primer extension products are identified by laser-induced fluorescence detection. Such an exemplary microchip can be used to process, for example, at least 96 to 384 samples, or more, in parallel.

## USES OF NUCLEIC ACID MOLECULES

The nucleic acid molecules of the present invention have a variety of uses, especially in the diagnosis and treatment of liver fibrosis and related pathologies. For example, the nucleic acid molecules are useful as hybridization probes, such as for genotyping SNPs in messenger RNA, transcript, cDNA, genomic DNA, amplified DNA or other nucleic acid molecules, and for isolating full-length cDNA and genomic clones encoding the variant peptides disclosed in Table 1 as well as their orthologs.

A probe can hybridize to any nucleotide sequence along the entire length of a nucleic acid molecule provided in Table 1 and/or Table 2. Preferably, a probe of the present invention hybridizes to a region of a target sequence that encompasses a SNP position indicated in Table 1 and/or Table 2. More preferably, a probe hybridizes to a SNP-containing target sequence in a sequence-specific manner such that it distinguishes the target sequence from other nucleotide sequences which vary from the target sequence only by which nucleotide is present at the SNP site. Such a probe is particularly useful for detecting the presence of a SNP-containing nucleic acid in a test sample, or for determining which nucleotide (allele) is present at a particular SNP site (*i.e.*, genotyping the SNP site).

A nucleic acid hybridization probe may be used for determining the presence, level, form, and/or distribution of nucleic acid expression. The nucleic acid whose level is determined can be

DNA or RNA. Accordingly, probes specific for the SNPs described herein can be used to assess the presence, expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in gene expression relative to normal levels. *In vitro* techniques for detection of mRNA include, for example, Northern blot hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern blot hybridizations and *in situ* hybridizations (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY).

Probes can be used as part of a diagnostic test kit for identifying cells or tissues in which a variant protein is expressed, such as by measuring the level of a variant protein-encoding nucleic acid (*e.g.*, mRNA) in a sample of cells from a subject or determining if a polynucleotide contains a SNP of interest.

Thus, the nucleic acid molecules of the invention can be used as hybridization probes to detect the SNPs disclosed herein, thereby determining whether an individual with the polymorphisms is at risk for liver fibrosis and related pathologies or has developed early stage liver fibrosis. Detection of a SNP associated with a disease phenotype provides a diagnostic tool for an active disease and/or genetic predisposition to the disease.

Furthermore, the nucleic acid molecules of the invention are therefore useful for detecting a gene (gene information is disclosed in Table 2, for example) which contains a SNP disclosed herein and/or products of such genes, such as expressed mRNA transcript molecules (transcript information is disclosed in Table 1, for example), and are thus useful for detecting gene expression. The nucleic acid molecules can optionally be implemented in, for example, an array or kit format for use in detecting gene expression.

The nucleic acid molecules of the invention are also useful as primers to amplify any given region of a nucleic acid molecule, particularly a region containing a SNP identified in Table 1 and/or Table 2.

The nucleic acid molecules of the invention are also useful for constructing recombinant vectors (described in greater detail below). Such vectors include expression vectors that express a portion of, or all of, any of the variant peptide sequences provided in Table 1. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced SNPs.

The nucleic acid molecules of the invention are also useful for expressing antigenic portions of the variant proteins, particularly antigenic portions that contain a variant amino acid sequence (*e.g.*, an amino acid substitution) caused by a SNP disclosed in Table 1 and/or Table 2.

The nucleic acid molecules of the invention are also useful for constructing vectors containing a gene regulatory region of the nucleic acid molecules of the present invention.

The nucleic acid molecules of the invention are also useful for designing ribozymes corresponding to all, or a part, of an mRNA molecule expressed from a SNP-containing nucleic acid molecule described herein.

The nucleic acid molecules of the invention are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and variant peptides.

The nucleic acid molecules of the invention are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and variant peptides. The production of recombinant cells and transgenic animals having nucleic acid molecules which contain the SNPs disclosed in Table 1 and/or Table 2 allow, for example, effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules of the invention are also useful in assays for drug screening to identify compounds that, for example, modulate nucleic acid expression.

The nucleic acid molecules of the invention are also useful in gene therapy in patients whose cells have aberrant gene expression. Thus, recombinant cells, which include a patient's cells that have been engineered *ex vivo* and returned to the patient, can be introduced into an individual where the recombinant cells produce the desired protein to treat the individual.

**SNP Genotyping Methods**

The process of determining which specific nucleotide (*i.e.*, allele) is present at each of one or more SNP positions, such as a SNP position in a nucleic acid molecule disclosed in Table 1 and/or Table 2, is referred to as SNP genotyping. The present invention provides methods of SNP genotyping, such as for use in screening for liver fibrosis or related pathologies, or determining predisposition thereto, or determining responsiveness to a form of treatment, or in genome mapping or SNP association analysis, etc.

Nucleic acid samples can be genotyped to determine which allele(s) is/are present at any given genetic region (*e.g.*, SNP position) of interest by methods well known in the art. The neighboring sequence can be used to design SNP detection reagents such as oligonucleotide probes, which may optionally be implemented in a kit format. Exemplary SNP genotyping methods are described in Chen *et al.*, "Single nucleotide polymorphism genotyping: biochemistry,

protocol, cost and throughput", *Pharmacogenomics J.* 2003;3(2):77-96; Kwok *et al.*, "Detection of single nucleotide polymorphisms", *Curr Issues Mol Biol.* 2003 Apr;5(2):43-60; Shi, "Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes", *Am J Pharmacogenomics.* 2002;2(3):197-205; and Kwok, "Methods for genotyping single nucleotide polymorphisms", *Annu Rev Genomics Hum Genet* 2001;2:235-58. Exemplary techniques for high-throughput SNP genotyping are described in Marnellos, "High-throughput SNP analysis for genetic association studies", *Curr Opin Drug Discov Devel.* 2003 May;6(3):317-21. Common SNP genotyping methods include, but are not limited to, TaqMan assays, molecular beacon assays, nucleic acid arrays, allele-specific primer extension, allele-specific PCR, arrayed primer extension, homogeneous primer extension assays, primer extension with detection by mass spectrometry, pyrosequencing, multiplex primer extension sorted on genetic arrays, ligation with rolling circle amplification, homogeneous ligation, OLA (U.S. Patent No. 4,988,167), multiplex ligation reaction sorted on genetic arrays, restriction-fragment length polymorphism, single base extension-tag assays, and the Invader assay. Such methods may be used in combination with detection mechanisms such as, for example, luminescence or chemiluminescence detection, fluorescence detection, time-resolved fluorescence detection, fluorescence resonance energy transfer, fluorescence polarization, mass spectrometry, and electrical detection.

Various methods for detecting polymorphisms include, but are not limited to, methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985); Cotton *et al.*, *PNAS 85*:4397 (1988); and Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), comparison of the electrophoretic mobility of variant and wild type nucleic acid molecules (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79 (1992)), and assaying the movement of polymorphic or wild-type fragments in polyacrylamide gels containing a gradient of denaturant using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.*, *Nature 313*:495 (1985)). Sequence variations at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or chemical cleavage methods.

In a preferred embodiment, SNP genotyping is performed using the TaqMan assay, which is also known as the 5' nuclease assay (U.S. Patent Nos. 5,210,015 and 5,538,848). The TaqMan assay detects the accumulation of a specific amplified product during PCR. The TaqMan assay utilizes an oligonucleotide probe labeled with a fluorescent reporter dye and a quencher dye. The reporter dye is excited by irradiation at an appropriate wavelength, it transfers energy to the quencher dye in the same probe via a process called fluorescence resonance energy transfer (FRET). When attached to the probe, the excited reporter dye does not emit a signal. The

proximity of the quencher dye to the reporter dye in the intact probe maintains a reduced fluorescence for the reporter. The reporter dye and quencher dye may be at the 5' most and the 3' most ends, respectively, or vice versa. Alternatively, the reporter dye may be at the 5' or 3' most end while the quencher dye is attached to an internal nucleotide, or vice versa. In yet another embodiment, both the reporter and the quencher may be attached to internal nucleotides at a distance from each other such that fluorescence of the reporter is reduced.

During PCR, the 5' nuclease activity of DNA polymerase cleaves the probe, thereby separating the reporter dye and the quencher dye and resulting in increased fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of the reporter dye. The DNA polymerase cleaves the probe between the reporter dye and the quencher dye only if the probe hybridizes to the target SNP-containing template which is amplified during PCR, and the probe is designed to hybridize to the target SNP site only if a particular SNP allele is present.

Preferred TaqMan primer and probe sequences can readily be determined using the SNP and associated nucleic acid sequence information provided herein. A number of computer programs, such as Primer Express (Applied Biosystems, Foster City, CA), can be used to rapidly obtain optimal primer/probe sets. It will be apparent to one of skill in the art that such primers and probes for detecting the SNPs of the present invention are useful in diagnostic assays for liver fibrosis and related pathologies, and can be readily incorporated into a kit format. The present invention also includes modifications of the Taqman assay well known in the art such as the use of Molecular Beacon probes (U.S. Patent Nos. 5,118,801 and 5,312,728) and other variant formats (U.S. Patent Nos. 5,866,336 and 6,117,635).

Another preferred method for genotyping the SNPs of the present invention is the use of two oligonucleotide probes in an OLA (see, *e.g.*, U.S. Patent No. 4,988,617). In this method, one probe hybridizes to a segment of a target nucleic acid with its 3' most end aligned with the SNP site. A second probe hybridizes to an adjacent segment of the target nucleic acid molecule directly 3' to the first probe. The two juxtaposed probes hybridize to the target nucleic acid molecule, and are ligated in the presence of a linking agent such as a ligase if there is perfect complementarity between the 3' most nucleotide of the first probe with the SNP site. If there is a mismatch, ligation would not occur. After the reaction, the ligated probes are separated from the target nucleic acid molecule, and detected as indicators of the presence of a SNP.

The following patents, patent applications, and published international patent applications, which are all hereby incorporated by reference, provide additional information pertaining to techniques for carrying out various types of OLA: U.S. Patent Nos. 6027889,

6268148, 5494810, 5830711, and 6054564 describe OLA strategies for performing SNP detection; WO 97/31256 and WO 00/56927 describe OLA strategies for performing SNP detection using universal arrays, wherein a zipcode sequence can be introduced into one of the hybridization probes, and the resulting product, or amplified product, hybridized to a universal zip code array; U.S. application US01/17329 (and 09/584,905) describes OLA (or LDR) followed by PCR, wherein zipcodes are incorporated into OLA probes, and amplified PCR products are determined by electrophoretic or universal zipcode array readout; U.S. applications 60/427818, 60/445636, and 60/445494 describe SNPlex methods and software for multiplexed SNP detection using OLA followed by PCR, wherein zipcodes are incorporated into OLA probes, and amplified PCR products are hybridized with a zipchute reagent, and the identity of the SNP determined from electrophoretic readout of the zipchute. In some embodiments, OLA is carried out prior to PCR (or another method of nucleic acid amplification). In other embodiments, PCR (or another method of nucleic acid amplification) is carried out prior to OLA.

Another method for SNP genotyping is based on mass spectrometry. Mass spectrometry takes advantage of the unique mass of each of the four nucleotides of DNA. SNPs can be unambiguously genotyped by mass spectrometry by measuring the differences in the mass of nucleic acids having alternative SNP alleles. MALDI-TOF (Matrix Assisted Laser Desorption Ionization – Time of Flight) mass spectrometry technology is preferred for extremely precise determinations of molecular mass, such as SNPs. Numerous approaches to SNP analysis have been developed based on mass spectrometry. Preferred mass spectrometry-based methods of SNP genotyping include primer extension assays, which can also be utilized in combination with other approaches, such as traditional gel-based formats and microarrays.

Typically, the primer extension assay involves designing and annealing a primer to a template PCR amplicon upstream (5') from a target SNP position. A mix of dideoxynucleotide triphosphates (ddNTPs) and/or deoxynucleotide triphosphates (dNTPs) are added to a reaction mixture containing template (*e.g.*, a SNP-containing nucleic acid molecule which has typically been amplified, such as by PCR), primer, and DNA polymerase. Extension of the primer terminates at the first position in the template where a nucleotide complementary to one of the ddNTPs in the mix occurs. The primer can be either immediately adjacent (*i.e.*, the nucleotide at the 3' end of the primer hybridizes to the nucleotide next to the target SNP site) or two or more nucleotides removed from the SNP position. If the primer is several nucleotides removed from the target SNP position, the only limitation is that the template sequence between the 3' end of the primer and the SNP position cannot contain a nucleotide of the same type as the one to be detected, or this will cause premature termination of the extension primer. Alternatively, if all

four ddNTPs alone, with no dNTPs, are added to the reaction mixture, the primer will always be extended by only one nucleotide, corresponding to the target SNP position. In this instance, primers are designed to bind one nucleotide upstream from the SNP position (*i.e.*, the nucleotide at the 3' end of the primer hybridizes to the nucleotide that is immediately adjacent to the target SNP site on the 5' side of the target SNP site). Extension by only one nucleotide is preferable, as it minimizes the overall mass of the extended primer, thereby increasing the resolution of mass differences between alternative SNP nucleotides. Furthermore, mass-tagged ddNTPs can be employed in the primer extension reactions in place of unmodified ddNTPs. This increases the mass difference between primers extended with these ddNTPs, thereby providing increased sensitivity and accuracy, and is particularly useful for typing heterozygous base positions. Mass-tagging also alleviates the need for intensive sample-preparation procedures and decreases the necessary resolving power of the mass spectrometer.

The extended primers can then be purified and analyzed by MALDI-TOF mass spectrometry to determine the identity of the nucleotide present at the target SNP position. In one method of analysis, the products from the primer extension reaction are combined with light absorbing crystals that form a matrix. The matrix is then hit with an energy source such as a laser to ionize and desorb the nucleic acid molecules into the gas-phase. The ionized molecules are then ejected into a flight tube and accelerated down the tube towards a detector. The time between the ionization event, such as a laser pulse, and collision of the molecule with the detector is the time of flight of that molecule. The time of flight is precisely correlated with the mass-to-charge ratio (m/z) of the ionized molecule. Ions with smaller m/z travel down the tube faster than ions with larger m/z and therefore the lighter ions reach the detector before the heavier ions. The time-of-flight is then converted into a corresponding, and highly precise, m/z. In this manner, SNPs can be identified based on the slight differences in mass, and the corresponding time of flight differences, inherent in nucleic acid molecules having different nucleotides at a single base position. For further information regarding the use of primer extension assays in conjunction with MALDI-TOF mass spectrometry for SNP genotyping, see, *e.g.*, Wise *et al.*, "A standard protocol for single nucleotide primer extension in the human genome using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", *Rapid Commun Mass Spectrom.* 2003;17(11):1195-202.

The following references provide further information describing mass spectrometry-based methods for SNP genotyping: Bocker, "SNP and mutation discovery using base-specific cleavage and MALDI-TOF mass spectrometry", *Bioinformatics.* 2003 Jul;19 Suppl 1:I44-I53; Storm *et al.*, "MALDI-TOF mass spectrometry-based SNP genotyping", *Methods Mol Biol.* 2003;212:241-62;

Jurinke *et al.*, "The use of Mass ARRAY technology for high throughput genotyping", *Adv Biochem Eng Biotechnol.* 2002;77:57-74; and Jurinke *et al.*, "Automated genotyping using the DNA MassArray technology", *Methods Mol Biol.* 2002;187:179-92.

SNPs can also be scored by direct DNA sequencing. A variety of automated sequencing procedures can be utilized ((1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)). The nucleic acid sequences of the present invention enable one of ordinary skill in the art to readily design sequencing primers for such automated sequencing procedures. Commercial instrumentation, such as the Applied Biosystems 377, 3100, 3700, 3730, and 3730xl DNA Analyzers (Foster City, CA), is commonly used in the art for automated sequencing.

Other methods that can be used to genotype the SNPs of the present invention include single-strand conformational polymorphism (SSCP), and denaturing gradient gel electrophoresis (DGGE) (Myers *et al.*, *Nature* 313:495 (1985)). SSCP identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Nat. Acad.* Single-stranded PCR products can be generated by heating or otherwise denaturing double stranded PCR products. Single-stranded nucleic acids may refold or form secondary structures that are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products are related to base-sequence differences at SNP positions. DGGE differentiates SNP alleles based on the different sequence-dependent stabilities and melting properties inherent in polymorphic DNA and the corresponding differences in electrophoretic migration patterns in a denaturing gradient gel (Erlich, ed., PCR Technology, *Principles and Applications for DNA Amplification*, W.H. Freeman and Co, New York, 1992, Chapter 7).

Sequence-specific ribozymes (U.S.Patent No. 5,498,531) can also be used to score SNPs based on the development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature. If the SNP affects a restriction enzyme cleavage site, the SNP can be identified by alterations in restriction enzyme digestion patterns, and the corresponding changes in nucleic acid fragment lengths determined by gel electrophoresis

SNP genotyping can include the steps of, for example, collecting a biological sample from a human subject (*e.g.*, sample of tissues, cells, fluids, secretions, etc.), isolating nucleic acids (*e.g.*, genomic DNA, mRNA or both) from the cells of the sample, contacting the nucleic acids with one or more primers which specifically hybridize to a region of the isolated nucleic

acid containing a target SNP under conditions such that hybridization and amplification of the target nucleic acid region occurs, and determining the nucleotide present at the SNP position of interest, or, in some assays, detecting the presence or absence of an amplification product (assays can be designed so that hybridization and/or amplification will only occur if a particular SNP allele is present or absent). In some assays, the size of the amplification product is detected and compared to the length of a control sample; for example, deletions and insertions can be detected by a change in size of the amplified product compared to a normal genotype.

SNP genotyping is useful for numerous practical applications, as described below. Examples of such applications include, but are not limited to, SNP-disease association analysis, disease predisposition screening, disease diagnosis, disease prognosis, disease progression monitoring, determining therapeutic strategies based on an individual's genotype ("pharmacogenomics"), developing therapeutic agents based on SNP genotypes associated with a disease or likelihood of responding to a drug, stratifying a patient population for clinical trial for a treatment regimen, predicting the likelihood that an individual will experience toxic side effects from a therapeutic agent, and human identification applications such as forensics.

**Analysis of Genetic Association Between SNPs and Phenotypic Traits**

SNP genotyping for disease diagnosis, disease predisposition screening, disease prognosis, determining drug responsiveness (pharmacogenomics), drug toxicity screening, and other uses described herein, typically relies on initially establishing a genetic association between one or more specific SNPs and the particular phenotypic traits of interest.

Different study designs may be used for genetic association studies (*Modern Epidemiology*, Lippincott Williams & Wilkins (1998), 609-622). Observational studies are most frequently carried out in which the response of the patients is not interfered with. The first type of observational study identifies a sample of persons in whom the suspected cause of the disease is present and another sample of persons in whom the suspected cause is absent, and then the frequency of development of disease in the two samples is compared. These sampled populations are called cohorts, and the study is a prospective study. The other type of observational study is case-control or a retrospective study. In typical case-control studies, samples are collected from individuals with the phenotype of interest (cases) such as certain manifestations of a disease, and from individuals without the phenotype (controls) in a population (target population) that conclusions are to be drawn from. Then the possible causes of the disease are investigated retrospectively. As the time and costs of collecting samples in case-control studies are considerably less than those for prospective studies, case-control studies are

the more commonly used study design in genetic association studies, at least during the exploration and discovery stage.

In both types of observational studies, there may be potential confounding factors that should be taken into consideration. Confounding factors are those that are associated with both the real cause(s) of the disease and the disease itself, and they include demographic information such as age, gender, ethnicity as well as environmental factors. When confounding factors are not matched in cases and controls in a study, and are not controlled properly, spurious association results can arise. If potential confounding factors are identified, they should be controlled for by analysis methods explained below.

In a genetic association study, the cause of interest to be tested is a certain allele or a SNP or a combination of alleles or a haplotype from several SNPs. Thus, tissue specimens (*e.g.*, whole blood) from the sampled individuals may be collected and genomic DNA genotyped for the SNP(s) of interest. In addition to the phenotypic trait of interest, other information such as demographic (*e.g.*, age, gender, ethnicity, etc.), clinical, and environmental information that may influence the outcome of the trait can be collected to further characterize and define the sample set. In many cases, these factors are known to be associated with diseases and/or SNP allele frequencies. There are likely gene-environment and/or gene-gene interactions as well. Analysis methods to address gene-environment and gene-gene interactions (for example, the effects of the presence of both susceptibility alleles at two different genes can be greater than the effects of the individual alleles at two genes combined) are discussed below.

After all the relevant phenotypic and genotypic information has been obtained, statistical analyses are carried out to determine if there is any significant correlation between the presence of an allele or a genotype with the phenotypic characteristics of an individual. Preferably, data inspection and cleaning are first performed before carrying out statistical tests for genetic association. Epidemiological and clinical data of the samples can be summarized by descriptive statistics with tables and graphs. Data validation is preferably performed to check for data completion, inconsistent entries, and outliers. Chi-squared tests and t-tests (Wilcoxon rank-sum tests if distributions are not normal) may then be used to check for significant differences between cases and controls for discrete and continuous variables, respectively. To ensure genotyping quality, Hardy-Weinberg disequilibrium tests can be performed on cases and controls separately. Significant deviation from Hardy-Weinberg equilibrium (HWE) in both cases and controls for individual markers can be indicative of genotyping errors. If HWE is violated in a majority of markers, it is indicative of population substructure that should be further investigated.

Moreover, Hardy-Weinberg disequilibrium in cases only can indicate genetic association of the markers with the disease (*Genetic Data Analysis*, Weir B., Sinauer (1990)).

To test whether an allele of a single SNP is associated with the case or control status of a phenotypic trait, one skilled in the art can compare allele frequencies in cases and controls. Standard chi-squared tests and Fisher exact tests can be carried out on a 2x2 table (2 SNP alleles x 2 outcomes in the categorical trait of interest). To test whether genotypes of a SNP are associated, chi-squared tests can be carried out on a 3x2 table (3 genotypes x 2 outcomes). Score tests are also carried out for genotypic association to contrast the three genotypic frequencies (major homozygotes, heterozygotes and minor homozygotes) in cases and controls, and to look for trends using 3 different modes of inheritance, namely dominant (with contrast coefficients 2, –1, –1), additive (with contrast coefficients 1, 0, –1) and recessive (with contrast coefficients 1, 1, –2). Odds ratios for minor versus major alleles, and odds ratios for heterozygote and homozygote variants versus the wild type genotypes are calculated with the desired confidence limits, usually 95%.

In order to control for confounders and to test for interaction and effect modifiers, stratified analyses may be performed using stratified factors that are likely to be confounding, including demographic information such as age, ethnicity, and gender, or an interacting element or effect modifier, such as a known major gene (*e.g.*, APOE for Alzheimer's disease or HLA genes for autoimmune diseases), or environmental factors such as smoking in lung cancer. Stratified association tests may be carried out using Cochran-Mantel-Haenszel tests that take into account the ordinal nature of genotypes with 0, 1, and 2 variant alleles. Exact tests by StatXact may also be performed when computationally possible. Another way to adjust for confounding effects and test for interactions is to perform stepwise multiple logistic regression analysis using statistical packages such as SAS or R. Logistic regression is a model-building technique in which the best fitting and most parsimonious model is built to describe the relation between the dichotomous outcome (for instance, getting a certain disease or not) and a set of independent variables (for instance, genotypes of different associated genes, and the associated demographic and environmental factors). The most common model is one in which the logit transformation of the odds ratios is expressed as a linear combination of the variables (main effects) and their cross-product terms (interactions) (Applied Logistic Regression, Hosmer and Lemeshow, Wiley (2000)). To test whether a certain variable or interaction is significantly associated with the outcome, coefficients in the model are first estimated and then tested for statistical significance of their departure from zero.

In addition to performing association tests one marker at a time, haplotype association analysis may also be performed to study a number of markers that are closely linked together. Haplotype association tests can have better power than genotypic or allelic association tests when the tested markers are not the disease-causing mutations themselves but are in linkage disequilibrium with such mutations. The test will even be more powerful if the disease is indeed caused by a combination of alleles on a haplotype (*e.g.*, APOE is a haplotype formed by 2 SNPs that are very close to each other). In order to perform haplotype association effectively, marker-marker linkage disequilibrium measures, both D' and $R^2$, are typically calculated for the markers within a gene to elucidate the haplotype structure. Recent studies (Daly *et al, Nature Genetics*, 29, 232-235, 2001) in linkage disequilibrium indicate that SNPs within a gene are organized in block pattern, and a high degree of linkage disequilibrium exists within blocks and very little linkage disequilibrium exists between blocks. Haplotype association with the disease status can be performed using such blocks once they have been elucidated.

Haplotype association tests can be carried out in a similar fashion as the allelic and genotypic association tests. Each haplotype in a gene is analogous to an allele in a multi-allelic marker. One skilled in the art can either compare the haplotype frequencies in cases and controls or test genetic association with different pairs of haplotypes. It has been proposed (Schaid *et al*, *Am. J. Hum. Genet.*, 70, 425-434, 2002) that score tests can be done on haplotypes using the program "haplo.score". In that method, haplotypes are first inferred by EM algorithm and score tests are carried out with a generalized linear model (GLM) framework that allows the adjustment of other factors.

An important decision in the performance of genetic association tests is the determination of the significance level at which significant association can be declared when the p-value of the tests reaches that level. In an exploratory analysis where positive hits will be followed up in subsequent confirmatory testing, an unadjusted p-value <0.2 (a significance level on the lenient side), for example, may be used for generating hypotheses for significant association of a SNP with certain phenotypic characteristics of a disease. It is preferred that a p-value < 0.05 (a significance level traditionally used in the art) is achieved in order for a SNP to be considered to have an association with a disease. It is more preferred that a p-value <0.01 (a significance level on the stringent side) is achieved for an association to be declared. When hits are followed up in confirmatory analyses in more samples of the same source or in different samples from different sources, adjustment for multiple testing will be performed as to avoid excess number of hits while maintaining the experiment-wise error rates at 0.05. While there are different methods to adjust for multiple testing to control for different kinds of error rates, a commonly used but rather

conservative method is Bonferroni correction to control the experiment-wise or family-wise error rate (*Multiple comparisons and multiple tests*, Westfall *et al*, SAS Institute (1999)). Permutation tests to control for the false discovery rates, FDR, can be more powerful (Benjamini and Hochberg, Journal of the Royal Statistical Society, Series B 57, 1289-1300, 1995, *Resampling-based Multiple Testing*, Westfall and Young, Wiley (1993)). Such methods to control for multiplicity would be preferred when the tests are dependent and controlling for false discovery rates is sufficient as opposed to controlling for the experiment-wise error rates.

In replication studies using samples from different populations after statistically significant markers have been identified in the exploratory stage, meta-analyses can then be performed by combining evidence of different studies (*Modern Epidemiology*, Lippincott Williams & Wilkins, 1998, 643-673). If available, association results known in the art for the same SNPs can be included in the meta-analyses.

Since both genotyping and disease status classification can involve errors, sensitivity analyses may be performed to see how odds ratios and p-values would change upon various estimates on genotyping and disease classification error rates.

It has been well known that subpopulation-based sampling bias between cases and controls can lead to spurious results in case-control association studies (Ewens and Spielman, *Am. J. Hum. Genet.* 62, 450-458, 1995) when prevalence of the disease is associated with different subpopulation groups. Such bias can also lead to a loss of statistical power in genetic association studies. To detect population stratification, Pritchard and Rosenberg (Pritchard *et al. Am. J. Hum. Gen.* 1999, 65:220-228) suggested typing markers that are unlinked to the disease and using results of association tests on those markers to determine whether there is any population stratification. When stratification is detected, the genomic control (GC) method as proposed by Devlin and Roeder (Devlin *et al. Biometrics* 1999, 55:997-1004) can be used to adjust for the inflation of test statistics due to population stratification. GC method is robust to changes in population structure levels as well as being applicable to DNA pooling designs (Devlin *et al. Genet. Epidem.* 20001, 21:273-284).

While Pritchard's method recommended using 15-20 unlinked microsatellite markers, it suggested using more than 30 biallelic markers to get enough power to detect population stratification. For the GC method, it has been shown (Bacanu *et al. Am. J. Hum. Genet.* 2000, 66:1933-1944) that about 60-70 biallelic markers are sufficient to estimate the inflation factor for the test statistics due to population stratification. Hence, 70 intergenic SNPs can be chosen in unlinked regions as indicated in a genome scan (Kehoe *et al. Hum. Mol. Genet.* 1999, 8:237-245).

Once individual risk factors, genetic or non-genetic, have been found for the predisposition to disease, the next step is to set up a classification/prediction scheme to predict the category (for instance, disease or no-disease) that an individual will be in depending on his genotypes of associated SNPs and other non-genetic risk factors. Logistic regression for discrete trait and linear regression for continuous trait are standard techniques for such tasks (*Applied Regression Analysis*, Draper and Smith, Wiley (1998)). Moreover, other techniques can also be used for setting up classification. Such techniques include, but are not limited to, MART, CART, neural network, and discriminant analyses that are suitable for use in comparing the performance of different methods (*The Elements of Statistical Learning*, Hastie, Tibshirani & Friedman, Springer (2002)).

**Disease Diagnosis and Predisposition Screening**

Information on association/correlation between genotypes and disease-related phenotypes can be exploited in several ways. For example, in the case of a highly statistically significant association between one or more SNPs with predisposition to a disease for which treatment is available, detection of such a genotype pattern in an individual may justify immediate administration of treatment, or at least the institution of regular monitoring of the individual. Detection of the susceptibility alleles associated with serious disease in a couple contemplating having children may also be valuable to the couple in their reproductive decisions. In the case of a weaker but still statistically significant association between a SNP and a human disease, immediate therapeutic intervention or monitoring may not be justified after detecting the susceptibility allele or SNP. Nevertheless, the subject can be motivated to begin simple life-style changes (*e.g.*, diet, exercise) that can be accomplished at little or no cost to the individual but would confer potential benefits in reducing the risk of developing conditions for which that individual may have an increased risk by virtue of having the susceptibility allele(s).

The SNPs of the invention may contribute to liver fibrosis and related pathologies in an individual in different ways. Some polymorphisms occur within a protein coding sequence and contribute to disease phenotype by affecting protein structure. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on, for example, replication, transcription, and/or translation. A single SNP may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by multiple SNPs in different genes.

As used herein, the terms "diagnose", "diagnosis", and "diagnostics" include, but are not limited to any of the following: detection of liver fibrosis that an individual may presently have, predisposition/susceptibility screening (*i.e.*, determining the increased risk of an individual in

developing liver fibrosis in the future, or determining whether an individual has a decreased risk of developing liver fibrosis in the future, determining the rate of progression of fibrosis to bridging fibrosis/cirrhosis), determining a particular type or subclass of liver fibrosis in an individual known to have liver fibrosis, confirming or reinforcing a previously made diagnosis of liver fibrosis, pharmacogenomic evaluation of an individual to determine which therapeutic strategy that individual is most likely to positively respond to or to predict whether a patient is likely to respond to a particular treatment, predicting whether a patient is likely to experience toxic effects from a particular treatment or therapeutic compound, and evaluating the future prognosis of an individual having liver fibrosis. Such diagnostic uses are based on the SNPs individually or in a unique combination or SNP haplotypes of the present invention.

Haplotypes are particularly useful in that, for example, fewer SNPs can be genotyped to determine if a particular genomic region harbors a locus that influences a particular phenotype, such as in linkage disequilibrium-based SNP association analysis.

Linkage disequilibrium (LD) refers to the co-inheritance of alleles (*e.g.*, alternative nucleotides) at two or more different SNP sites at frequencies greater than would be expected from the separate frequencies of occurrence of each allele in a given population. The expected frequency of co-occurrence of two alleles that are inherited independently is the frequency of the first allele multiplied by the frequency of the second allele. Alleles that co-occur at expected frequencies are said to be in "linkage equilibrium". In contrast, LD refers to any non-random genetic association between allele(s) at two or more different SNP sites, which is generally due to the physical proximity of the two loci along a chromosome. LD can occur when two or more SNPs sites are in close physical proximity to each other on a given chromosome and therefore alleles at these SNP sites will tend to remain unseparated for multiple generations with the consequence that a particular nucleotide (allele) at one SNP site will show a non-random association with a particular nucleotide (allele) at a different SNP site located nearby. Hence, genotyping one of the SNP sites will give almost the same information as genotyping the other SNP site that is in LD.

Various degrees of LD can be encountered between two or more SNPs with the result being that some SNPs are more closely associated (*i.e.*, in stronger LD) than others. Furthermore, the physical distance over which LD extends along a chromosome differs between different regions of the genome, and therefore the degree of physical separation between two or more SNP sites necessary for LD to occur can differ between different regions of the genome.

For diagnostic purposes and similar uses, if a particular SNP site is found to be useful for diagnosing liver fibrosis and related pathologies (*e.g.*, has a significant statistical association with

the condition and/or is recognized as a causative polymorphism for the condition), then the skilled artisan would recognize that other SNP sites which are in LD with this SNP site would also be useful for diagnosing the condition. Thus, polymorphisms (*e.g.*, SNPs and/or haplotypes) that are not the actual disease-causing (causative) polymorphisms, but are in LD with such causative polymorphisms, are also useful. In such instances, the genotype of the polymorphism(s) that is/are in LD with the causative polymorphism is predictive of the genotype of the causative polymorphism and, consequently, predictive of the phenotype (*e.g.*, liver fibrosis) that is influenced by the causative SNP(s). Therefore, polymorphic markers that are in LD with causative polymorphisms are useful as diagnostic markers, and are particularly useful when the actual causative polymorphism(s) is/are unknown.

Examples of polymorphisms that can be in LD with one or more causative polymorphisms (and/or in LD with one or more polymorphisms that have a significant statistical association with a condition) and therefore useful for diagnosing the same condition that the causative/associated SNP(s) is used to diagnose, include, for example, other SNPs in the same gene, protein-coding, or mRNA transcript-coding region as the causative/associated SNP, other SNPs in the same exon or same intron as the causative/associated SNP, other SNPs in the same haplotype block as the causative/associated SNP, other SNPs in the same intergenic region as the causative/associated SNP, SNPs that are outside but near a gene (*e.g.*, within 6kb on either side, 5' or 3', of a gene boundary) that harbors a causative/associated SNP, etc. Such useful LD SNPs can be selected from among the SNPs disclosed in Table 4, for example.

Linkage disequilibrium in the human genome is reviewed in the following references: Wall *et al.et al.*, "Haplotype blocks and linkage disequilibrium in the human genome,", *Nat Rev Genet.* 2003 Aug;4(8):587-97 (Aug. 2003); Garner *et al.et al.*, "On selecting markers for association studies: patterns of linkage disequilibrium between two and three diallelic loci,", *Genet Epidemiol.* 2003 Jan;24(1):57-67 (Jan. 2003); Ardlie *et al.et al.*, "Patterns of linkage disequilibrium in the human genome,", *Nat Rev Genet.* 2002 Apr;3(4):299-309 (Apr. 2002); (erratum in *Nat Rev Genet* 2002 Jul;3(7):566 (Jul. 2002); and Remm *et al.et al.*, "High-density genotyping and linkage disequilibrium in the human genome using chromosome 22 as a model,"; *Curr Opin Chem Biol.* 2002 Feb;6(1):24-30 (Feb. 2002); J.B.S. Haldane, " JBS (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors," . *J Genet* 8:299-309 (1919); G. Mendel, , *G. (1866) Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereines in Brünn [(Proceedings of the Natural History Society of Brünn)]* (1866); Lewin B (1990) *Genes IV,* B. Lewin, ed., . Oxford University Press, N.Y.ew York, USA (1990); D.L. Hartl DL and A.G. Clark AG (1989) *Principles of Population*

*Genetics 2nd ed.*, . Sinauer Associates, Inc., Ma Sunderland, Mass., USA (1989); J.H. Gillespie JH (2004) *Population Genetics: A Concise Guide.2nd ed.*,. Johns Hopkins University Press. (2004) USA; R.C. Lewontin, "RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models,". *Genetics* 49:49-67 (1964); P.G. Hoel, PG (1954) *Introduction to Mathematical Statistics 2nd ed.*, John Wiley & Sons, Inc., N.Y. New York, USA (1954); R.R. Hudson, RR "(2001) Two-locus sampling distributions and their application,". *Genetics* 159:1805-1817 (2001); A.P. Dempster AP, N.M. Laird, D.B. NM, Rubin, "DB (1977) Maximum likelihood from incomplete data via the EM algorithm,". *J R Stat Soc* 39:1-38 (1977); L. Excoffier L, M. Slatkin, M "(1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population,". *Mol Biol Evol* 12(5):921-927 (1995); D.A. Tregouet DA, S. Escolano S, L. Tiret L, A. Mallet A, J.L. Golmard, JL "(2004) A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm," . *Ann Hum Genet* 68(Pt 2):165-177 (2004); A.D. Long AD and C.H. Langley CH, " (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits," . *Genome Research* 9:720-731 (1999); A. Agresti, A (1990) *Categorical Data Analysis*,. John Wiley & Sons, Inc., N.Y. New York, USA (1990); K. Lange, K *(1997) Mathematical and Statistical Methods for Genetic Analysis*, . Springer-Verlag New York, Inc., N.Y. New York, USA (1997); The International HapMap Consortium, "(2003) The International HapMap Project," . *Nature* 426:789-796 (2003); The International HapMap Consortium, " (2005) A haplotype map of the human genome,". *Nature* 437:1299-1320 (2005); G.A. Thorisson GA, A.V. Smith AV, L. Krishnan L, L.D. Stein LD (2005), "The International HapMap Project Web Site,". *Genome Research* 15:1591-493 (2005); G. McVean, C.C.A. G, Spencer CCA, R. Chaix R (2005), "Perspectives on human genetic variation from the HapMap project,". *PLoS Genetics* 1(4):413-418 (2005); J.N. Hirschhorn JN, M.J. Daly, MJ "(2005) Genome-wide association studies for common diseases and complex traits,". *Nat Genet* 6:95-108 (2005); S.J. Schrodi, " SJ (2005) A probabilistic approach to large-scale association scans: a semi-Bayesian method to detect disease-predisposing alleles,". *SAGMB* 4(1):31 (2005); W.Y.S. Wang WYS, B.J. Barratt BJ, D.G. Clayton DG, J.A. Todd, "JA (2005) Genome-wide association studies: theoretical and practical concerns,". *Nat Rev Genet* 6:109-118 (2005); J.K. . Pritchard JK, M. Przeworski, " M (2001) Linkage disequilibrium in humans: models and data,". *Am J Hum Genet* 69:1-14 (2001).

As discussed above, one aspect of the present invention is the discovery that SNPs which are in certain LD distance with the interrogated SNP can also be used as valid markers for identifying an increased or decreased risks of having or developing CHD. As used herein, the term “interrogated SNP” refers to SNPs that have been found to be associated with an increased

or decreased risk of disease using genotyping results and analysis, or other appropriate experimental method as exemplified in the working examples described in this application. As used herein, the term "LD SNP" refers to a SNP that has been characterized as a SNP associating with an increased or decreased risk of diseases due to their being in LD with the "interrogated SNP" under the methods of calculation described in the application. Below, applicants describe the methods of calculation with which one of ordinary skilled in the art may determine if a particular SNP is in LD with an interrogated SNP. The parameter $r^2$ is commonly used in the genetics art to characterize the extent of linkage disequilibrium between markers (Hudson, 2001). As used herein, the term "in LD with" refers to a particular SNP that is measured at above the threshold of a parameter such as $r^2$ with an interrogated SNP.

It is now common place to directly observe genetic variants in a sample of chromosomes obtained from a population. Suppose one has genotype data at two genetic markers located on the same chromosome, for the markers $A$ and $B$. Further suppose that two alleles segregate at each of these two markers such that alleles $A_1$ and $A_2$ can be found at marker $A$ and alleles $B_1$ and $B_2$ at marker $B$. Also assume that these two markers are on a human autosome. If one is to examine a specific individual and find that they are heterozygous at both markers, such that their two-marker genotype is $A_1A_2B_1B_2$, then there are two possible configurations: the individual in question could have the alleles $A_1B_1$ on one chromosome and $A_2B_2$ on the remaining chromosome; alternatively, the individual could have alleles $A_1B_2$ on one chromosome and $A_2B_1$ on the other. The arrangement of alleles on a chromosome is called a haplotype. In this illustration, the individual could have haplotypes $A_1B_1/A_2B_2$ or $A_1B_2/A_2B_1$ (see Hartl and Clark (1989) for a more complete description). The concept of linkage equilibrium relates the frequency of haplotypes to the allele frequencies.

Assume that a sample of individuals is selected from a larger population. Considering the two markers described above, each having two alleles, there are four possible haplotypes: $A_1B_1$, $A_1B_2$, $A_2B_1$ and $A_2B_2$. Denote the frequencies of these four haplotypes with the following notation.

$$P_{11} = freq(A_1B_1) \quad (1)$$

$$P_{12} = freq(A_1B_2) \quad (2)$$

$$P_{21} = freq(A_2B_1) \quad (3)$$

$$P_{22} = freq(A_2B_2) \quad (4)$$

The allele frequencies at the two markers are then the sum of different haplotype frequencies, it is straightforward to write down a similar set of equations relating single-marker allele frequencies to two-marker haplotype frequencies:

$$p_1 = freq(A_1) = P_{11} + P_{12} \quad (5)$$

$$p_2 = freq(A_2) = P_{21} + P_{22} \quad (6)$$

$$q_1 = freq(B_1) = P_{11} + P_{21} \quad (7)$$

$$q_2 = freq(B_2) = P_{12} + P_{22} \quad (8)$$

Note that the four haplotype frequencies and the allele frequencies at each marker must sum to a frequency of 1.

$$P_{11} + P_{12} + P_{21} + P_{22} = 1 \quad (9)$$

$$p_1 + p_2 = 1 \quad (10)$$

$$q_1 + q_2 = 1 \quad (11)$$

If there is no correlation between the alleles at the two markers, one would expect that the frequency of the haplotypes would be approximately the product of the composite alleles. Therefore,

$$P_{11} \approx p_1 q_1 \quad (12)$$

$$P_{12} \approx p_1 q_2 \quad (13)$$

$$P_{21} \approx p_2 q_1 \quad (14)$$

$$P_{22} \approx p_2 q_2 \quad (15)$$

These approximating equations (12)-(15) represent the concept of linkage equilibrium where there is independent assortment between the two markers – the alleles at the two markers occur together at random. These are represented as approximations because linkage equilibrium and linkage disequilibrium are concepts typically thought of as properties of a sample of chromosomes; and as such they are susceptible to stochastic fluctuations due to the sampling process. Empirically, many pairs of genetic markers will be in linkage equilibrium, but certainly not all pairs.

Having established the concept of linkage equilibrium above, applicants can now describe the concept of linkage disequilibrium (LD), which is the deviation from linkage equilibrium. Since the frequency of the $A_1B_1$ haplotype is approximately the product of the allele frequencies for $A_1$ and $B_1$ under the assumption of linkage equilibrium as stated mathematically in (12), a simple measure for the amount of departure from linkage equilibrium is the difference in these two quantities, $D$,

$$D = P_{11} - p_1 q_1 \tag{16}$$

$D = 0$ indicates perfect linkage equilibrium. Substantial departures from $D = 0$ indicates LD in the sample of chromosomes examined. Many properties of $D$ are discussed in Lewontin (1964) including the maximum and minimum values that $D$ can take. Mathematically, using basic algebra, it can be shown that $D$ can also be written solely in terms of haplotypes:

$$D = P_{11}P_{22} - P_{12}P_{21} \tag{17}$$

If one transforms $D$ by squaring it and subsequently dividing by the product of the allele frequencies of $A_1$, $A_2$, $B_1$ and $B_2$, the resulting quantity, called $r^2$, is equivalent to the square of the Pearson's correlation coefficient commonly used in statistics (*e.g.* Hoel, 1954).

$$r^2 = \frac{D^2}{p_1 p_2 q_1 q_2} \tag{18}$$

As with $D$, values of $r^2$ close to 0 indicate linkage equilibrium between the two markers examined in the sample set. As values of $r^2$ increase, the two markers are said to be in linkage disequilibrium. The range of values that $r^2$ can take are from 0 to 1. $r^2 = 1$ when there is a perfect correlation between the alleles at the two markers.

In addition, the quantities discussed above are sample-specific. And as such, it is necessary to formulate notation specific to the samples studied. In the approach discussed here, three types of samples are of primary interest: (i) a sample of chromosomes from individuals affected by a disease-related phenotype (cases), (ii) a sample of chromosomes obtained from individuals not affected by the disease-related phenotype (controls), and (iii) a standard sample set used for the construction of haplotypes and calculation pairwise linkage disequilibrium. For

the allele frequencies used in the development of the method described below, an additional subscript will be added to denote either the case or control sample sets.

$$p_{1,cs} = freq(A_1 \text{ in cases}) \quad (19)$$

$$p_{2,cs} = freq(A_2 \text{ in cases}) \quad (20)$$

$$q_{1,cs} = freq(B_1 \text{ in cases}) \quad (21)$$

$$q_{2,cs} = freq(B_2 \text{ in cases}) \quad (22)$$

Similarly,

$$p_{1,ct} = freq(A_1 \text{ in controls}) \quad (23)$$

$$p_{2,ct} = freq(A_2 \text{ in controls}) \quad (24)$$

$$q_{1,ct} = freq(B_1 \text{ in controls}) \quad (25)$$

$$q_{2,ct} = freq(B_2 \text{ in controls}) \quad (26)$$

As a well-accepted sample set is necessary for robust linkage disequilibrium calculations, data obtained from the International HapMap project (The International HapMap Consortium 2003, 2005; Thorisson et al, 2005; McVean et al, 2005) can be used for the calculation of pairwise $r^2$ values. Indeed, the samples genotyped for the International HapMap Project were selected to be representative examples from various human sub-populations with sufficient numbers of chromosomes examined to draw meaningful and robust conclusions from the patterns of genetic variation observed. The International HapMap project website (*hapmap.org*) contains a description of the project, methods utilized and samples examined. It is useful to examine empirical data to get a sense of the patterns present in such data.

Haplotype frequencies were explicit arguments in equation (18) above. However, knowing the 2-marker haplotype frequencies requires that phase to be determined for doubly heterozygous samples. When phase is unknown in the data examined, various algorithms can be used to infer phase from the genotype data. This issue was discussed earlier where the doubly heterozygous individual with a 2-SNP genotype of $A_1A_2B_1B_2$ could have one of two different sets of chromosomes: $A_1B_1 / A_2B_2$ or $A_1B_2 / A_2B_1$. One such algorithm to estimate haplotype frequencies is the expectation-maximization (EM) algorithm first formalized by Dempster et al*et al.* (1977). This algorithm is often used in genetics to infer haplotype frequencies from genotype data (e.g.*e.g.* Excoffier and Slatkin (, 1995); Tregouet et al*et al.*, (2004)). It should be noted that

for the two-SNP case explored here, EM algorithms have very little error provided that the allele frequencies and sample sizes are not too small. The impact on $r^2$ values is typically negligible.

As correlated genetic markers share information, interrogation of SNP markers in LD with a disease-associated SNP marker can also have sufficient power to detect disease association (Long and Langley (, 1999)). The relationship between the power to directly find disease-associated alleles and the power to indirectly detect disease-association was investigated by Pritchard and Przeworski (2001). In a straight-forward derivation, it can be shown that the power to detect disease association indirectly at a marker locus in linkage disequilibrium with a disease-association locus is approximately the same as the power to detect disease-association directly at the disease- association locus if the sample size is increased by a factor of $\frac{1}{r^2}$ (the reciprocal of equation 18) at the marker in comparison with the disease- association locus.

Therefore, if one calculated the power to detect disease-association indirectly with an experiment having $N$ samples, then equivalent power to directly detect disease-association (at the actual disease-susceptibility locus) would necessitate an experiment using approximately $r^2N$ samples. This elementary relationship between power, sample size and linkage disequilibrium can be used to derive an $r^2$ threshold value useful in determining whether or not genotyping markers in linkage disequilibrium with a SNP marker directly associated with disease status has enough power to indirectly detect disease-association.

To commence a derivation of the power to detect disease-associated markers through an indirect process, define the effective chromosomal sample size as

$$n = \frac{4N_{cs}N_{ct}}{N_{cs} + N_{ct}}; \qquad (27)$$

where $N_{cs}$ and $N_{ct}$ are the numbers of diploid cases and controls, respectively. This is necessary to handle situations where the numbers of cases and controls are not equivalent. For equal case and control sample sizes, $N_{cs} = N_{ct} = N$, the value of the effective number of chromosomes is simply $n = 2N$ – as expected. Let power be calculated for a significance level $\alpha$ (such that traditional P-values below $\alpha$ will be deemed statistically significant). Define the standard Gaussian distribution function as $\Phi(\bullet)$. Mathematically,

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{\theta^2}{2}}d\theta \tag{28}$$

Alternatively, the following error function notation (Erf) may also be used,

$$\Phi(x)=\frac{1}{2}\left[1+Erf\left(\frac{x}{\sqrt{2}}\right)\right] \tag{29}$$

For example, $\Phi(1.644854)=0.95$. The value of $r^2$ may be derived to yield a pre-specified minimum amount of power to detect disease association though indirect interrogation. Noting that the LD SNP marker could be the one that is carrying the disease- association allele, therefore that this approach constitutes a lower-bound model where all indirect power results are expected to be at least as large as those interrogated.

Denote by $\beta$ the error rate for not detecting truly disease-associated markers. Therefore, $1-\beta$ is the classical definition of statistical power. Substituting the Pritchard-Pzreworski result into the sample size, the power to detect disease association at a significance level of $\alpha$ is given by the approximation

$$1-\beta\cong\Phi\left[\frac{\left|q_{1,cs}-q_{1,ct}\right|}{\sqrt{\frac{q_{1,cs}(1-q_{1,cs})+q_{1,ct}(1-q_{1,ct})}{r^2n}}}-Z_{1-\alpha/2}\right]; \tag{30}$$

where $Z_u$ is the inverse of the standard normal cumulative distribution evaluated at $u$ ($u\in(0,1)$). $Z_u=\Phi^{-1}(u)$, where $\Phi(\Phi^{-1}(u))=\Phi^{-1}(\Phi(u))=u$. For example, setting $\alpha=0.05$, and therefore $1-\alpha/2=0.975$, $Z_{0.975}=1.95996$ is obtained. Next, setting power equal to a threshold of a minimum power of $T$,

$$T=\Phi\left[\frac{\left|q_{1,cs}-q_{1,ct}\right|}{\sqrt{\frac{q_{1,cs}(1-q_{1,cs})+q_{1,ct}(1-q_{1,ct})}{r^2n}}}-Z_{1-\alpha/2}\right] \tag{31}$$

and solving for $r^2$, the following threshold $r^2$ is obtained:

$$r_T^2 = \frac{\left[q_{1,cs}\left(1-q_{1,cs}\right)+q_{1,ct}\left(1-q_{1,ct}\right)\right]}{n\left(q_{1,cs}-q_{1,ct}\right)^2}\left[\Phi^{-1}(T)+Z_{1-\alpha/2}\right]^2 \quad (32)$$

Or,

$$r_T^2 = \frac{\left(Z_T+Z_{1-\alpha/2}\right)^2}{n}\left[\frac{q_{1,cs}-\left(q_{1,cs}\right)^2+q_{1,ct}-\left(q_{1,ct}\right)^2}{\left(q_{1,cs}-q_{1,ct}\right)^2}\right] \quad (33)$$

Suppose that $r^2$ is calculated between an interrogated SNP and a number of other SNPs with varying levels of LD with the interrogated SNP. The threshold value $r_T^2$ is the minimum value of linkage disequilibrium between the interrogated SNP and the potential LD SNPs such that the LD SNP still retains a power greater or equal to $T$ for detecting disease-association. For example, suppose that SNP rs200 is genotyped in a case-control disease-association study and it is found to be associated with a disease phenotype. Further suppose that the minor allele frequency in 1,000 case chromosomes was found to be 16% in contrast with a minor allele frequency of 10% in 1,000 control chromosomes. Given those measurements one could have predicted, prior to the experiment, that the power to detect disease association at a significance level of 0.05 was quite high – approximately 98% using a test of allelic association. Applying equation (32) one can calculate a minimum value of $r^2$ to indirectly assess disease association assuming that the minor allele at SNP rs200 is truly disease-predisposing for a threshold level of power. If one sets the threshold level of power to be 80%, then $r_T^2 = 0.489$ given the same significance level and chromosome numbers as above. Hence, any SNP with a pairwise $r^2$ value with rs200 greater than 0.489 is expected to have greater than 80% power to detect the disease association. Further, this is assuming the conservative model where the LD SNP is disease-associated only through linkage disequilibrium with the interrogated SNP rs200.

In general, the threshold $r_T^2$ value can be set such that one with ordinary skill in the art would consider that any two SNPs having an $r^2$ value greater than or equal to the threshold $r_T^2$ value would be in sufficient LD with each other such that either SNP is useful for the same utilities, such as determing an individual's risk for developing liver fibrosis, for example. For example, in various embodiments, the threshold $r_T^2$ value used to classify SNPs as being in

sufficient LD with an interrogated SNP (such that these LD SNPs can be used for the same utilities as the interrogated SNP, for example) can be set at, for example, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 1, *etc.* (or any other $r^2$ value in-between these values). Threshold $r_T^2$ values may be utilized with or without considering power or other calculations.

The contribution or association of particular SNPs and/or SNP haplotypes with disease phenotypes, such as liver fibrosis, enables the SNPs of the present invention to be used to develop superior diagnostic tests capable of identifying individuals who express a detectable trait, such as liver fibrosis, as the result of a specific genotype, or individuals whose genotype places them at an increased or decreased risk of developing a detectable trait at a subsequent time as compared to individuals who do not have that genotype. As described herein, diagnostics may be based on a single SNP or a group of SNPs. Combined detection of a plurality of SNPs (for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 24, 25, 30, 32, 48, 50, 64, 96, 100, or any other number in-between, or more, of the SNPs provided in Table 1 and/or Table 2) typically increases the probability of an accurate diagnosis. For example, the presence of a single SNP known to correlate with liver fibrosis might indicate a probability of 20% that an individual has or is at risk of developing liver fibrosis, whereas detection of five SNPs, each of which correlates with liver fibrosis, might indicate a probability of 80% that an individual has or is at risk of developing liver fibrosis. To further increase the accuracy of diagnosis or predisposition screening, analysis of the SNPs of the present invention can be combined with that of other polymorphisms or other risk factors of liver fibrosis, such as disease symptoms, pathological characteristics, family history, diet, environmental factors or lifestyle factors.

It will, of course, be understood by practitioners skilled in the treatment or diagnosis of liver fibrosis that the present invention generally does not intend to provide an absolute identification of individuals who are at risk (or less at risk) of developing liver fibrosis, and/or pathologies related to liver fibrosis, but rather to indicate a certain increased (or decreased) degree or likelihood of developing the disease based on statistically significant association results. However, this information is extremely valuable as it can be used to, for example, initiate preventive treatments or to allow an individual carrying one or more significant SNPs or SNP haplotypes to foresee warning signs such as minor clinical symptoms, or to have regularly scheduled physical exams to monitor for appearance of a condition in order to identify and begin treatment of the condition at an early stage. Particularly with diseases that are extremely debilitating or fatal if not treated on time, the knowledge of a potential predisposition, even if this predisposition is not absolute, would likely contribute in a very significant manner to treatment efficacy.

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a SNP or a SNP pattern associated with an increased or decreased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular polymorphism/mutation, including, for example, methods which enable the analysis of individual chromosomes for haplotyping, family studies, single sperm DNA analysis, or somatic hybrids. The trait analyzed using the diagnostics of the invention may be any detectable trait that is commonly observed in pathologies and disorders related to liver fibrosis.

Another aspect of the present invention relates to a method of determining whether an individual is at risk (or less at risk) of developing one or more traits or whether an individual expresses one or more traits as a consequence of possessing a particular trait-causing or trait-influencing allele. These methods generally involve obtaining a nucleic acid sample from an individual and assaying the nucleic acid sample to determine which nucleotide(s) is/are present at one or more SNP positions, wherein the assayed nucleotide(s) is/are indicative of an increased or decreased risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular trait-causing or trait-influencing allele.

In another embodiment, the SNP detection reagents of the present invention are used to determine whether an individual has one or more SNP allele(s) affecting the level (*e.g.*, the concentration of mRNA or protein in a sample, etc.) or pattern (*e.g.*, the kinetics of expression, rate of decomposition, stability profile, Km, Vmax, etc.) of gene expression (collectively, the "gene response" of a cell or bodily fluid). Such a determination can be accomplished by screening for mRNA or protein expression (*e.g.*, by using nucleic acid arrays, RT-PCR, TaqMan assays, or mass spectrometry), identifying genes having altered expression in an individual, genotyping SNPs disclosed in Table 1 and/or Table 2 that could affect the expression of the genes having altered expression (*e.g.*, SNPs that are in and/or around the gene(s) having altered expression, SNPs in regulatory/control regions, SNPs in and/or around other genes that are involved in pathways that could affect the expression of the gene(s) having altered expression, or all SNPs could be genotyped), and correlating SNP genotypes with altered gene expression. In this manner, specific SNP alleles at particular SNP sites can be identified that affect gene expression.

**Pharmacogenomics and Therapeutics/Drug Development**

The present invention provides methods for assessing the pharmacogenomics of a subject harboring particular SNP alleles or haplotypes to a particular therapeutic agent or pharmaceutical

compound, or to a class of such compounds. Pharmacogenomics deals with the roles which clinically significant hereditary variations (*e.g.*, SNPs) play in the response to drugs due to altered drug disposition and/or abnormal action in affected persons. See, *e.g.*, Roses, *Nature* 405, 857-865 (2000); Gould Rothberg, *Nature Biotechnology* 19, 209-211 (2001); Eichelbaum, *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996); and Linder, *Clin. Chem.* 43(2):254-266 (1997). The clinical outcomes of these variations can result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the SNP genotype of an individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. For example, SNPs in drug metabolizing enzymes can affect the activity of these enzymes, which in turn can affect both the intensity and duration of drug action, as well as drug metabolism and clearance.

The discovery of SNPs in drug metabolizing enzymes, drug transporters, proteins for pharmaceutical agents, and other drug targets has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. SNPs can be expressed in the phenotype of the extensive metabolizer and in the phenotype of the poor metabolizer. Accordingly, SNPs may lead to allelic variants of a protein in which one or more of the protein functions in one population are different from those in another population. SNPs and the encoded variant peptides thus provide targets to ascertain a genetic predisposition that can affect treatment modality. For example, in a ligand-based treatment, SNPs may give rise to amino terminal extracellular domains and/or other ligand-binding regions of a receptor that are more or less active in ligand binding, thereby affecting subsequent protein activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing particular SNP alleles or haplotypes.

As an alternative to genotyping, specific variant proteins containing variant amino acid sequences encoded by alternative SNP alleles could be identified. Thus, pharmacogenomic characterization of an individual permits the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic uses based on the individual's SNP genotype, thereby enhancing and optimizing the effectiveness of the therapy. Furthermore, the production of recombinant cells and transgenic animals containing particular SNPs/haplotypes allow effective clinical design and testing of treatment compounds and dosage regimens. For example, transgenic animals can be produced that differ only in specific SNP alleles in a gene that is orthologous to a human disease susceptibility gene.

Pharmacogenomic uses of the SNPs of the present invention provide several significant advantages for patient care, particularly in treating liver fibrosis. Pharmacogenomic characterization

of an individual, based on an individual's SNP genotype, can identify those individuals unlikely to respond to treatment with a particular medication and thereby allows physicians to avoid prescribing the ineffective medication to those individuals. On the other hand, SNP genotyping of an individual may enable physicians to select the appropriate medication and dosage regimen that will be most effective based on an individual's SNP genotype. This information increases a physician's confidence in prescribing medications and motivates patients to comply with their drug regimens. Furthermore, pharmacogenomics may identify patients predisposed to toxicity and adverse reactions to particular drugs or drug dosages. Adverse drug reactions lead to more than 100,000 avoidable deaths per year in the United States alone and therefore represent a significant cause of hospitalization and death, as well as a significant economic burden on the healthcare system (Pfost *et. al., Trends in Biotechnology*, Aug. 2000.). Thus, pharmacogenomics based on the SNPs disclosed herein has the potential to both save lives and reduce healthcare costs substantially. Current treatments of patients suffering liver fibrosis include interferon (including pegylated-interferon) treatment, and combination treatment with interferon and ribavirin. See Lauer et al for detailed description on HCV treatments. These treatments can have drastic side effects on patients receiving the treatment, and should be used in a targeted manner. In that regard, embodiments of the present invention can be very useful in assisting clinicians select patients who are more likely develop severe form of fibrosis, and thus warrant the application of HCV treatments on such patients. In the mean time, patients who are deemed to have low risk of developing severe form of fibrosis/cirrhosis, using SNP markers discovered herein, can be spared of the aggravation of the HCV treatment due to the reduced benefit of such treatment in view of its cost.

Pharmacogenomics in general is discussed further in Rose *et al.*, "Pharmacogenetic analysis of clinically relevant genetic polymorphisms", *Methods Mol Med.* 2003;85:225-37. Pharmacogenomics as it relates to Alzheimer's disease and other neurodegenerative disorders is discussed in Cacabelos, "Pharmacogenomics for the treatment of dementia", *Ann Med.* 2002;34(5):357-79, Maimone *et al.*, "Pharmacogenomics of neurodegenerative diseases", *Eur J Pharmacol.* 2001 Feb 9;413(1):11-29, and Poirier, "Apolipoprotein E: a pharmacogenetic target for the treatment of Alzheimer's disease", *Mol Diagn.* 1999 Dec;4(4):335-41. Pharmacogenomics as it relates to cardiovascular disorders is discussed in Siest *et al.*, "Pharmacogenomics of drugs affecting the cardiovascular system", *Clin Chem Lab Med.* 2003 Apr;41(4):590-9, Mukherjee *et al.*, "Pharmacogenomics in cardiovascular diseases", *Prog Cardiovasc Dis.* 2002 May-Jun;44(6):479-98, and Mooser *et al.*, "Cardiovascular pharmacogenetics in the SNP era", *J Thromb Haemost.* 2003 Jul;1(7):1398-402. Pharmacogenomics as it relates to cancer is discussed in McLeod *et al.*, "Cancer pharmacogenomics: SNPs, chips, and the individual patient", *Cancer*

*Invest.* 2003;21(4):630-40 and Watters *et al.*, "Cancer pharmacogenomics: current and future applications", *Biochim Biophys Acta.* 2003 Mar 17; 1603(2):99-111.

The SNPs of the present invention also can be used to identify novel therapeutic targets for liver fibrosis. For example, genes containing the disease-associated variants ("variant genes") or their products, as well as genes or their products that are directly or indirectly regulated by or interacting with these variant genes or their products, can be targeted for the development of therapeutics that, for example, treat the disease or prevent or delay disease onset. The therapeutics may be composed of, for example, small molecules, proteins, protein fragments or peptides, antibodies, nucleic acids, or their derivatives or mimetics which modulate the functions or levels of the target genes or gene products.

The SNP-containing nucleic acid molecules disclosed herein, and their complementary nucleic acid molecules, may be used as antisense constructs to control gene expression in cells, tissues, and organisms. Antisense technology is well established in the art and extensively reviewed in *Antisense Drug Technology: Principles, Strategies, and Applications*, Crooke (ed.), Marcel Dekker, Inc.: New York (2001). An antisense nucleic acid molecule is generally designed to be complementary to a region of mRNA expressed by a gene so that the antisense molecule hybridizes to the mRNA and thereby blocks translation of mRNA into protein. Various classes of antisense oligonucleotides are used in the art, two of which are cleavers and blockers. Cleavers, by binding to target RNAs, activate intracellular nucleases (*e.g.*, RNaseH or RNase L) that cleave the target RNA. Blockers, which also bind to target RNAs, inhibit protein translation through steric hindrance of ribosomes. Exemplary blockers include peptide nucleic acids, morpholinos, locked nucleic acids, and methylphosphonates (see, *e.g.*, Thompson, *Drug Discovery Today*, 7 (17): 912-917 (2002)). Antisense oligonucleotides are directly useful as therapeutic agents, and are also useful for determining and validating gene function (*e.g.*, in gene knock-out or knock-down experiments).

Antisense technology is further reviewed in: Lavery *et al.*, "Antisense and RNAi: powerful tools in drug target discovery and validation", *Curr Opin Drug Discov Devel.* 2003 Jul;6(4):561-9; Stephens *et al.*, "Antisense oligonucleotide therapy in cancer", *Curr Opin Mol Ther.* 2003 Apr;5(2):118-22; Kurreck, "Antisense technologies. Improvement through novel chemical modifications", *Eur J Biochem.* 2003 Apr;270(8):1628-44; Dias *et al.*, "Antisense oligonucleotides: basic concepts and mechanisms", *Mol Cancer Ther.* 2002 Mar;1(5):347-55; Chen, "Clinical development of antisense oligonucleotides as anti-cancer therapeutics", *Methods Mol Med.* 2003;75:621-36; Wang *et al.*, "Antisense anticancer oligonucleotide therapeutics",

*Curr Cancer Drug Targets*. 2001 Nov;1(3):177-96; and Bennett, "Efficiency of antisense oligonucleotide drug discovery", *Antisense Nucleic Acid Drug Dev*. 2002 Jun;12(3):215-24.

The SNPs of the present invention are particularly useful for designing antisense reagents that are specific for particular nucleic acid variants. Based on the SNP information disclosed herein, antisense oligonucleotides can be produced that specifically target mRNA molecules that contain one or more particular SNP nucleotides. In this manner, expression of mRNA molecules that contain one or more undesired polymorphisms (*e.g.*, SNP nucleotides that lead to a defective protein such as an amino acid substitution in a catalytic domain) can be inhibited or completely blocked. Thus, antisense oligonucleotides can be used to specifically bind a particular polymorphic form (*e.g.*, a SNP allele that encodes a defective protein), thereby inhibiting translation of this form, but which do not bind an alternative polymorphic form (*e.g.*, an alternative SNP nucleotide that encodes a protein having normal function).

Antisense molecules can be used to inactivate mRNA in order to inhibit gene expression and production of defective proteins. Accordingly, these molecules can be used to treat a disorder, such as liver fibrosis, characterized by abnormal or undesired gene expression or expression of certain defective proteins. This technique can involve cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible mRNA regions include, for example, protein-coding regions and particularly protein-coding regions corresponding to catalytic activities, substrate/ligand binding, or other functional activities of a protein.

The SNPs of the present invention are also useful for designing RNA interference reagents that specifically target nucleic acid molecules having particular SNP variants. RNA interference (RNAi), also referred to as gene silencing, is based on using double-stranded RNA (dsRNA) molecules to turn genes off. When introduced into a cell, dsRNAs are processed by the cell into short fragments (generally about 21, 22, or 23 nucleotides in length) known as small interfering RNAs (siRNAs) which the cell uses in a sequence-specific manner to recognize and destroy complementary RNAs (Thompson, *Drug Discovery Today*, 7 (17): 912-917 (2002)). Accordingly, an aspect of the present invention specifically contemplates isolated nucleic acid molecules that are about 18-26 nucleotides in length, preferably 19-25 nucleotides in length, and more preferably 20, 21, 22, or 23 nucleotides in length, and the use of these nucleic acid molecules for RNAi. Because RNAi molecules, including siRNAs, act in a sequence-specific manner, the SNPs of the present invention can be used to design RNAi reagents that recognize and destroy nucleic acid molecules having specific SNP alleles/nucleotides (such as deleterious alleles that lead to the production of defective proteins), while not affecting nucleic acid

molecules having alternative SNP alleles (such as alleles that encode proteins having normal function). As with antisense reagents, RNAi reagents may be directly useful as therapeutic agents (*e.g.*, for turning off defective, disease-causing genes), and are also useful for characterizing and validating gene function (*e.g.*, in gene knock-out or knock-down experiments).

The following references provide a further review of RNAi: Reynolds *et al.*, "Rational siRNA design for RNA interference", *Nat Biotechnol.* 2004 Mar;22(3):326-30. Epub 2004 Feb 01; Chi *et al.*, "Genomewide view of gene silencing by small interfering RNAs", *PNAS* 100(11):6343-6346, 2003; Vickers *et al.*, "Efficient Reduction of Target RNAs by Small Interfering RNA and RNase H-dependent Antisense Agents", *J. Biol. Chem.* 278: 7108-7118, 2003; Agami, "RNAi and related mechanisms and their potential use for therapy", *Curr Opin Chem Biol.* 2002 Dec;6(6):829-34; Lavery *et al.*, "Antisense and RNAi: powerful tools in drug target discovery and validation", *Curr Opin Drug Discov Devel.* 2003 Jul;6(4):561-9; Shi, "Mammalian RNAi for the masses", *Trends Genet* 2003 Jan;19(1):9-12), Shuey *et al.*, "RNAi: gene-silencing in therapeutic intervention", *Drug Discovery Today* 2002 Oct;7(20):1040-1046; McManus *et al.*, *Nat Rev Genet* 2002 Oct;3(10):737-47; Xia *et al.*, *Nat Biotechnol* 2002 Oct;20(10):1006-10; Plasterk *et al., Curr Opin Genet Dev* 2000 Oct;10(5):562-7; Bosher *et al., Nat Cell Biol* 2000 Feb;2(2):E31-6; and Hunter, *Curr Biol* 1999 Jun 17;9(12):R440-2).

A subject suffering from a pathological condition, such as liver fibrosis, ascribed to a SNP may be treated so as to correct the genetic defect (see Kren *et al., Proc. Natl. Acad. Sci. USA* 96:10349-10354 (1999)). Such a subject can be identified by any method that can detect the polymorphism in a biological sample drawn from the subject. Such a genetic defect may be permanently corrected by administering to such a subject a nucleic acid fragment incorporating a repair sequence that supplies the normal/wild-type nucleotide at the position of the SNP. This site-specific repair sequence can encompass an RNA/DNA oligonucleotide that operates to promote endogenous repair of a subject's genomic DNA. The site-specific repair sequence is administered in an appropriate vehicle, such as a complex with polyethylenimine, encapsulated in anionic liposomes, a viral vector such as an adenovirus, or other pharmaceutical composition that promotes intracellular uptake of the administered nucleic acid. A genetic defect leading to an inborn pathology may then be overcome, as the chimeric oligonucleotides induce incorporation of the normal sequence into the subject's genome. Upon incorporation, the normal gene product is expressed, and the replacement is propagated, thereby engendering a permanent repair and therapeutic enhancement of the clinical condition of the subject.

In cases in which a cSNP results in a variant protein that is ascribed to be the cause of, or a contributing factor to, a pathological condition, a method of treating such a condition can include administering to a subject experiencing the pathology the wild-type/normal cognate of the variant protein. Once administered in an effective dosing regimen, the wild-type cognate provides complementation or remediation of the pathological condition.

The invention further provides a method for identifying a compound or agent that can be used to treat liver fibrosis. The SNPs disclosed herein are useful as targets for the identification and/or development of therapeutic agents. A method for identifying a therapeutic agent or compound typically includes assaying the ability of the agent or compound to modulate the activity and/or expression of a SNP-containing nucleic acid or the encoded product and thus identifying an agent or a compound that can be used to treat a disorder characterized by undesired activity or expression of the SNP-containing nucleic acid or the encoded product. The assays can be performed in cell-based and cell-free systems. Cell-based assays can include cells naturally expressing the nucleic acid molecules of interest or recombinant cells genetically engineered to express certain nucleic acid molecules.

Variant gene expression in a liver fibrosis patient can include, for example, either expression of a SNP-containing nucleic acid sequence (for instance, a gene that contains a SNP can be transcribed into an mRNA transcript molecule containing the SNP, which can in turn be translated into a variant protein) or altered expression of a normal/wild-type nucleic acid sequence due to one or more SNPs (for instance, a regulatory/control region can contain a SNP that affects the level or pattern of expression of a normal transcript).

Assays for variant gene expression can involve direct assays of nucleic acid levels (*e.g.*, mRNA levels), expressed protein levels, or of collateral compounds involved in a signal pathway. Further, the expression of genes that are up- or down-regulated in response to the signal pathway can also be assayed. In this embodiment, the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Modulators of variant gene expression can be identified in a method wherein, for example, a cell is contacted with a candidate compound/agent and the expression of mRNA determined. The level of expression of mRNA in the presence of the candidate compound is compared to the level of expression of mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of variant gene expression based on this comparison and be used to treat a disorder such as liver fibrosis that is characterized by variant gene expression (*e.g.*, either expression of a SNP-containing nucleic acid or altered expression of a normal/wild-type nucleic acid molecule due to one or more SNPs that affect expression of the nucleic acid molecule) due to one or

more SNPs of the present invention. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the SNP or associated nucleic acid domain (*e.g.*, catalytic domain, ligand/substrate-binding domain, regulatory/control region, etc.) or gene, or the encoded mRNA transcript, as a target, using a compound identified through drug screening as a gene modulator to modulate variant nucleic acid expression. Modulation can include either up-regulation (*i.e.*, activation or agonization) or down-regulation (*i.e.*, suppression or antagonization) of nucleic acid expression.

Expression of mRNA transcripts and encoded proteins, either wild type or variant, may be altered in individuals with a particular SNP allele in a regulatory/control element, such as a promoter or transcription factor binding domain, that regulates expression. In this situation, methods of treatment and compounds can be identified, as discussed herein, that regulate or overcome the variant regulatory/control element, thereby generating normal, or healthy, expression levels of either the wild type or variant protein.

The SNP-containing nucleic acid molecules of the present invention are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of a variant gene, or encoded product, in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as an indicator for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance, as well as an indicator for toxicities. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

In another aspect of the present invention, there is provided a pharmaceutical pack comprising a therapeutic agent (*e.g.*, a small molecule drug, antibody, peptide, antisense or RNAi nucleic acid molecule, etc.) and a set of instructions for administration of the therapeutic agent to humans diagnostically tested for one or more SNPs or SNP haplotypes provided by the present invention.

The SNPs/haplotypes of the present invention are also useful for improving many different aspects of the drug development process. For instance, an aspect of the present

invention includes selecting individuals for clinical trials based on their SNP genotype. For example, individuals with SNP genotypes that indicate that they are likely to positively respond to a drug can be included in the trials, whereas those individuals whose SNP genotypes indicate that they are less likely to or would not respond to the drug, or who are at risk for suffering toxic effects or other adverse reactions, can be excluded from the clinical trials. This not only can improve the safety of clinical trials, but also can enhance the chances that the trial will demonstrate statistically significant efficacy. Furthermore, the SNPs of the present invention may explain why certain previously developed drugs performed poorly in clinical trials and may help identify a subset of the population that would benefit from a drug that had previously performed poorly in clinical trials, thereby "rescuing" previously developed drugs, and enabling the drug to be made available to a particular liver fibrosis patient population that can benefit from it.

SNPs have many important uses in drug discovery, screening, and development. A high probability exists that, for any gene/protein selected as a potential drug target, variants of that gene/protein will exist in a patient population. Thus, determining the impact of gene/protein variants on the selection and delivery of a therapeutic agent should be an integral aspect of the drug discovery and development process. (Jazwinska, *A Trends Guide to Genetic Variation and Genomic Medicine*, 2002 Mar; S30-S36).

Knowledge of variants (*e.g.*, SNPs and any corresponding amino acid polymorphisms) of a particular therapeutic target (*e.g.*, a gene, mRNA transcript, or protein) enables parallel screening of the variants in order to identify therapeutic candidates (*e.g.*, small molecule compounds, antibodies, antisense or RNAi nucleic acid compounds, etc.) that demonstrate efficacy across variants (Rothberg, *Nat Biotechnol* 2001 Mar;19(3):209-11). Such therapeutic candidates would be expected to show equal efficacy across a larger segment of the patient population, thereby leading to a larger potential market for the therapeutic candidate.

Furthermore, identifying variants of a potential therapeutic target enables the most common form of the target to be used for selection of therapeutic candidates, thereby helping to ensure that the experimental activity that is observed for the selected candidates reflects the real activity expected in the largest proportion of a patient population (Jazwinska, *A Trends Guide to Genetic Variation and Genomic Medicine*, 2002 Mar; S30-S36).

Additionally, screening therapeutic candidates against all known variants of a target can enable the early identification of potential toxicities and adverse reactions relating to particular variants. For example, variability in drug absorption, distribution, metabolism and excretion (ADME) caused by, for example, SNPs in therapeutic targets or drug metabolizing genes, can be

identified, and this information can be utilized during the drug development process to minimize variability in drug disposition and develop therapeutic agents that are safer across a wider range of a patient population. The SNPs of the present invention, including the variant proteins and encoding polymorphic nucleic acid molecules provided in Tables 1-2, are useful in conjunction with a variety of toxicology methods established in the art, such as those set forth in *Current Protocols in Toxicology*, John Wiley & Sons, Inc., N.Y.

Furthermore, therapeutic agents that target any art-known proteins (or nucleic acid molecules, either RNA or DNA) may cross-react with the variant proteins (or polymorphic nucleic acid molecules) disclosed in Table 1, thereby significantly affecting the pharmacokinetic properties of the drug. Consequently, the protein variants and the SNP-containing nucleic acid molecules disclosed in Tables 1-2 are useful in developing, screening, and evaluating therapeutic agents that target corresponding art-known protein forms (or nucleic acid molecules). Additionally, as discussed above, knowledge of all polymorphic forms of a particular drug target enables the design of therapeutic agents that are effective against most or all such polymorphic forms of the drug target.

**Pharmaceutical Compositions and Administration Thereof**

Any of the liver fibrosis-associated proteins, and encoding nucleic acid molecules, disclosed herein can be used as therapeutic targets (or directly used themselves as therapeutic compounds) for treating liver fibrosis and related pathologies, and the present disclosure enables therapeutic compounds (*e.g.*, small molecules, antibodies, therapeutic proteins, RNAi and antisense molecules, etc.) to be developed that target (or are comprised of) any of these therapeutic targets.

In general, a therapeutic compound will be administered in a therapeutically effective amount by any of the accepted modes of administration for agents that serve similar utilities. The actual amount of the therapeutic compound of this invention, *i.e.*, the active ingredient, will depend upon numerous factors such as the severity of the disease to be treated, the age and relative health of the subject, the potency of the compound used, the route and form of administration, and other factors.

Therapeutically effective amounts of therapeutic compounds may range from, for example, approximately 0.01-50 mg per kilogram body weight of the recipient per day; preferably about 0.1-20 mg/kg/day. Thus, as an example, for administration to a 70 kg person, the dosage range would most preferably be about 7 mg to 1.4 g per day.

In general, therapeutic compounds will be administered as pharmaceutical compositions by any one of the following routes: oral, systemic (*e.g.*, transdermal, intranasal, or by suppository), or parenteral (*e.g.*, intramuscular, intravenous, or subcutaneous) administration. The preferred manner of administration is oral or parenteral using a convenient daily dosage regimen, which can be adjusted according to the degree of affliction. Oral compositions can take the form of tablets, pills, capsules, semisolids, powders, sustained release formulations, solutions, suspensions, elixirs, aerosols, or any other appropriate compositions.

The choice of formulation depends on various factors such as the mode of drug administration (*e.g.,* for oral administration, formulations in the form of tablets, pills, or capsules are preferred) and the bioavailability of the drug substance. Recently, pharmaceutical formulations have been developed especially for drugs that show poor bioavailability based upon the principle that bioavailability can be increased by increasing the surface area, *i.e.*, decreasing particle size. For example, U.S. Patent No. 4,107,288 describes a pharmaceutical formulation having particles in the size range from 10 to 1,000 nm in which the active material is supported on a cross-linked matrix of macromolecules. U.S. Patent No. 5,145,684 describes the production of a pharmaceutical formulation in which the drug substance is pulverized to nanoparticles (average particle size of 400 nm) in the presence of a surface modifier and then dispersed in a liquid medium to give a pharmaceutical formulation that exhibits remarkably high bioavailability.

Pharmaceutical compositions are comprised of, in general, a therapeutic compound in combination with at least one pharmaceutically acceptable excipient. Acceptable excipients are non-toxic, aid administration, and do not adversely affect the therapeutic benefit of the therapeutic compound. Such excipients may be any solid, liquid, semi-solid or, in the case of an aerosol composition, gaseous excipient that is generally available to one skilled in the art.

Solid pharmaceutical excipients include starch, cellulose, talc, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, magnesium stearate, sodium stearate, glycerol monostearate, sodium chloride, dried skim milk and the like. Liquid and semisolid excipients may be selected from glycerol, propylene glycol, water, ethanol and various oils, including those of petroleum, animal, vegetable or synthetic origin, *e.g.*, peanut oil, soybean oil, mineral oil, sesame oil, etc. Preferred liquid carriers, particularly for injectable solutions, include water, saline, aqueous dextrose, and glycols.

Compressed gases may be used to disperse a compound of this invention in aerosol form. Inert gases suitable for this purpose are nitrogen, carbon dioxide, etc.

Other suitable pharmaceutical excipients and their formulations are described in Remington's Pharmaceutical Sciences, edited by E. W. Martin (Mack Publishing Company, 18th ed., 1990).

The amount of the therapeutic compound in a formulation can vary within the full range employed by those skilled in the art. Typically, the formulation will contain, on a weight percent (wt %) basis, from about 0.01-99.99 wt % of the therapeutic compound based on the total formulation, with the balance being one or more suitable pharmaceutical excipients. Preferably, the compound is present at a level of about 1-80 wt %.

Therapeutic compounds can be administered alone or in combination with other therapeutic compounds or in combination with one or more other active ingredient(s). For example, an inhibitor or stimulator of a liver fibrosis-associated protein can be administered in combination with another agent that inhibits or stimulates the activity of the same or a different liver fibrosis-associated protein to thereby counteract the affects of liver fibrosis.

For further information regarding pharmacology, see *Current Protocols in Pharmacology*, John Wiley & Sons, Inc., N.Y.

**Human Identification Applications**

In addition to their diagnostic and therapeutic uses in liver fibrosis and related pathologies, the SNPs provided by the present invention are also useful as human identification markers for such applications as forensics, paternity testing, and biometrics (see, *e.g.*, Gill, "An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes", *Int J Legal Med.* 2001;114(4-5):204-10). Genetic variations in the nucleic acid sequences between individuals can be used as genetic markers to identify individuals and to associate a biological sample with an individual. Determination of which nucleotides occupy a set of SNP positions in an individual identifies a set of SNP markers that distinguishes the individual. The more SNP positions that are analyzed, the lower the probability that the set of SNPs in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked (*i.e.*, inherited independently). Thus, preferred sets of SNPs can be selected from among the SNPs disclosed herein, which may include SNPs on different chromosomes, SNPs on different chromosome arms, and/or SNPs that are dispersed over substantial distances along the same chromosome arm.

Furthermore, among the SNPs disclosed herein, preferred SNPs for use in certain forensic/human identification applications include SNPs located at degenerate codon positions (*i.e.*, the third position in certain codons which can be one of two or more alternative nucleotides

and still encode the same amino acid), since these SNPs do not affect the encoded protein. SNPs that do not affect the encoded protein are expected to be under less selective pressure and are therefore expected to be more polymorphic in a population, which is typically an advantage for forensic/human identification applications. However, for certain forensics/human identification applications, such as predicting phenotypic characteristics (*e.g.*, inferring ancestry or inferring one or more physical characteristics of an individual) from a DNA sample, it may be desirable to utilize SNPs that affect the encoded protein.

For many of the SNPs disclosed in Tables 1-2 (which are identified as "Applera" SNP source), Tables 1-2 provide SNP allele frequencies obtained by re-sequencing the DNA of chromosomes from 39 individuals (Tables 1-2 also provide allele frequency information for "Celera" source SNPs and, where available, public SNPs from dbEST, HGBASE, and/or HGMD). The allele frequencies provided in Tables 1-2 enable these SNPs to be readily used for human identification applications. Although any SNP disclosed in Table 1 and/or Table 2 could be used for human identification, the closer that the frequency of the minor allele at a particular SNP site is to 50%, the greater the ability of that SNP to discriminate between different individuals in a population since it becomes increasingly likely that two randomly selected individuals would have different alleles at that SNP site. Using the SNP allele frequencies provided in Tables 1-2, one of ordinary skill in the art could readily select a subset of SNPs for which the frequency of the minor allele is, for example, at least 1%, 2%, 5%, 10%, 20%, 25%, 30%, 40%, 45%, or 50%, or any other frequency in-between. Thus, since Tables 1-2 provide allele frequencies based on the re-sequencing of the chromosomes from 39 individuals, a subset of SNPs could readily be selected for human identification in which the total allele count of the minor allele at a particular SNP site is, for example, at least 1, 2, 4, 8, 10, 16, 20, 24, 30, 32, 36, 38, 39, 40, or any other number in-between.

Furthermore, Tables 1-2 also provide population group (interchangeably referred to herein as ethnic or racial groups) information coupled with the extensive allele frequency information. For example, the group of 39 individuals whose DNA was re-sequenced was made-up of 20 Caucasians and 19 African-Americans. This population group information enables further refinement of SNP selection for human identification. For example, preferred SNPs for human identification can be selected from Tables 1-2 that have similar allele frequencies in both the Caucasian and African-American populations; thus, for example, SNPs can be selected that have equally high discriminatory power in both populations. Alternatively, SNPs can be selected for which there is a statistically significant difference in allele frequencies between the Caucasian and African-American populations (as an extreme example, a particular allele may be observed

only in either the Caucasian or the African-American population group but not observed in the other population group); such SNPs are useful, for example, for predicting the race/ethnicity of an unknown perpetrator from a biological sample such as a hair or blood stain recovered at a crime scene. For a discussion of using SNPs to predict ancestry from a DNA sample, including statistical methods, see Frudakis *et al.*, "A Classifier for the SNP-Based Inference of Ancestry", *Journal of Forensic Sciences* 2003; 48(4):771-782.

SNPs have numerous advantages over other types of polymorphic markers, such as short tandem repeats (STRs). For example, SNPs can be easily scored and are amenable to automation, making SNPs the markers of choice for large-scale forensic databases. SNPs are found in much greater abundance throughout the genome than repeat polymorphisms. Population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci. SNPs are mutationaly more stable than repeat polymorphisms. SNPs are not susceptible to artefacts such as stutter bands that can hinder analysis. Stutter bands are frequently encountered when analyzing repeat polymorphisms, and are particularly troublesome when analyzing samples such as crime scene samples that may contain mixtures of DNA from multiple sources. Another significant advantage of SNP markers over STR markers is the much shorter length of nucleic acid needed to score a SNP. For example, STR markers are generally several hundred base pairs in length. A SNP, on the other hand, comprises a single nucleotide, and generally a short conserved region on either side of the SNP position for primer and/or probe binding. This makes SNPs more amenable to typing in highly degraded or aged biological samples that are frequently encountered in forensic casework in which DNA may be fragmented into short pieces.

SNPs also are not subject to microvariant and "off-ladder" alleles frequently encountered when analyzing STR loci. Microvariants are deletions or insertions within a repeat unit that change the size of the amplified DNA product so that the amplified product does not migrate at the same rate as reference alleles with normal sized repeat units. When separated by size, such as by electrophoresis on a polyacrylamide gel, microvariants do not align with a reference allelic ladder of standard sized repeat units, but rather migrate between the reference alleles. The reference allelic ladder is used for precise sizing of alleles for allele classification; therefore alleles that do not align with the reference allelic ladder lead to substantial analysis problems. Furthermore, when analyzing multi-allelic repeat polymorphisms, occasionally an allele is found that consists of more or less repeat units than has been previously seen in the population, or more or less repeat alleles than are included in a reference allelic ladder. These alleles will migrate outside the size range of known alleles in a reference allelic ladder, and therefore are referred to

as "off-ladder" alleles. In extreme cases, the allele may contain so few or so many repeats that it migrates well out of the range of the reference allelic ladder. In this situation, the allele may not even be observed, or, with multiplex analysis, it may migrate within or close to the size range for another locus, further confounding analysis.

SNP analysis avoids the problems of microvariants and off-ladder alleles encountered in STR analysis. Importantly, microvariants and off-ladder alleles may provide significant problems, and may be completely missed, when using analysis methods such as oligonucleotide hybridization arrays, which utilize oligonucleotide probes specific for certain known alleles. Furthermore, off-ladder alleles and microvariants encountered with STR analysis, even when correctly typed, may lead to improper statistical analysis, since their frequencies in the population are generally unknown or poorly characterized, and therefore the statistical significance of a matching genotype may be questionable. All these advantages of SNP analysis are considerable in light of the consequences of most DNA identification cases, which may lead to life imprisonment for an individual, or re-association of remains to the family of a deceased individual.

DNA can be isolated from biological samples such as blood, bone, hair, saliva, or semen, and compared with the DNA from a reference source at particular SNP positions. Multiple SNP markers can be assayed simultaneously in order to increase the power of discrimination and the statistical significance of a matching genotype. For example, oligonucleotide arrays can be used to genotype a large number of SNPs simultaneously. The SNPs provided by the present invention can be assayed in combination with other polymorphic genetic markers, such as other SNPs known in the art or STRs, in order to identify an individual or to associate an individual with a particular biological sample.

Furthermore, the SNPs provided by the present invention can be genotyped for inclusion in a database of DNA genotypes, for example, a criminal DNA databank such as the FBI's Combined DNA Index System (CODIS) database. A genotype obtained from a biological sample of unknown source can then be queried against the database to find a matching genotype, with the SNPs of the present invention providing nucleotide positions at which to compare the known and unknown DNA sequences for identity. Accordingly, the present invention provides a database comprising novel SNPs or SNP alleles of the present invention (*e.g.*, the database can comprise information indicating which alleles are possessed by individual members of a population at one or more novel SNP sites of the present invention), such as for use in forensics, biometrics, or other human identification applications. Such a database typically comprises a computer-based system in which the SNPs or SNP alleles of the present invention are recorded

on a computer readable medium (see the section of the present specification entitled "Computer-Related Embodiments").

The SNPs of the present invention can also be assayed for use in paternity testing. The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child, with the SNPs of the present invention providing nucleotide positions at which to compare the putative father's and child's DNA sequences for identity. If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the putative father, it can be concluded, barring experimental error, that the putative father is not the father of the child. If the set of polymorphisms in the child attributable to the father match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match, and a conclusion drawn as to the likelihood that the putative father is the true biological father of the child.

In addition to paternity testing, SNPs are also useful for other types of kinship testing, such as for verifying familial relationships for immigration purposes, or for cases in which an individual alleges to be related to a deceased individual in order to claim an inheritance from the deceased individual, etc. For further information regarding the utility of SNPs for paternity testing and other types of kinship testing, including methods for statistical analysis, see Krawczak, "Informativity assessment for biallelic single nucleotide polymorphisms", *Electrophoresis* 1999 Jun;20(8):1676-81.

The use of the SNPs of the present invention for human identification further extends to various authentication systems, commonly referred to as biometric systems, which typically convert physical characteristics of humans (or other organisms) into digital data. Biometric systems include various technological devices that measure such unique anatomical or physiological characteristics as finger, thumb, or palm prints; hand geometry; vein patterning on the back of the hand; blood vessel patterning of the retina and color and texture of the iris; facial characteristics; voice patterns; signature and typing dynamics; and DNA. Such physiological measurements can be used to verify identity and, for example, restrict or allow access based on the identification. Examples of applications for biometrics include physical area security, computer and network security, aircraft passenger check-in and boarding, financial transactions, medical records access, government benefit distribution, voting, law enforcement, passports, visas and immigration, prisons, various military

applications, and for restricting access to expensive or dangerous items, such as automobiles or guns (see, for example, O'Connor, *Stanford Technology Law Review* and U.S. Patent No. 6,119,096).

Groups of SNPs, particularly the SNPs provided by the present invention, can be typed to uniquely identify an individual for biometric applications such as those described above. Such SNP typing can readily be accomplished using, for example, DNA chips/arrays. Preferably, a minimally invasive means for obtaining a DNA sample is utilized. For example, PCR amplification enables sufficient quantities of DNA for analysis to be obtained from buccal swabs or fingerprints, which contain DNA-containing skin cells and oils that are naturally transferred during contact. Further information regarding techniques for using SNPs in forensic/human identification applications can be found in, for example, *Current Protocols in Human Genetics*, John Wiley & Sons, N.Y. (2002), 14.1-14.7.

## VARIANT PROTEINS, ANTIBODIES, VECTORS & HOST CELLS, & USES THEREOF

### Variant Proteins Encoded by SNP-Containing Nucleic Acid Molecules

The present invention provides SNP-containing nucleic acid molecules, many of which encode proteins having variant amino acid sequences as compared to the art-known (*i.e.*, wild-type) proteins. Amino acid sequences encoded by the polymorphic nucleic acid molecules of the present invention are provided as SEQ ID NOS: 17-32 in Table 1 and the Sequence Listing. These variants will generally be referred to herein as variant proteins/peptides/polypeptides, or polymorphic proteins/peptides/polypeptides of the present invention. The terms "protein", "peptide", and "polypeptide" are used herein interchangeably.

A variant protein of the present invention may be encoded by, for example, a nonsynonymous nucleotide substitution at any one of the cSNP positions disclosed herein. In addition, variant proteins may also include proteins whose expression, structure, and/or function is altered by a SNP disclosed herein, such as a SNP that creates or destroys a stop codon, a SNP that affects splicing, and a SNP in control/regulatory elements, *e.g.* promoters, enhancers, or transcription factor binding domains.

As used herein, a protein or peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or chemical precursors or other chemicals. The variant proteins of the present invention can be purified to homogeneity or other lower degrees of purity. The level of purification will be based on the intended use. The key feature is that the preparation allows for the desired function of the variant protein, even if in the presence of considerable amounts of other components.

As used herein, "substantially free of cellular material" includes preparations of the variant protein having less than about 30% (by dry weight) other proteins (*i.e.*, contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the variant protein is recombinantly produced, it can also be substantially free of culture medium, *i.e.*, culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the variant protein in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the variant protein having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

An isolated variant protein may be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant host cells), or synthesized using known protein synthesis methods. For example, a nucleic acid molecule containing SNP(s) encoding the variant protein can be cloned into an expression vector, the expression vector introduced into a host cell, and the variant protein expressed in the host cell. The variant protein can then be isolated from the cells by any appropriate purification scheme using standard protein purification techniques. Examples of these techniques are described in detail below (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

The present invention provides isolated variant proteins that comprise, consist of or consist essentially of amino acid sequences that contain one or more variant amino acids encoded by one or more codons which contain a SNP of the present invention.

Accordingly, the present invention provides variant proteins that consist of amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein consists of an amino acid sequence when the amino acid sequence is the entire amino acid sequence of the protein.

The present invention further provides variant proteins that consist essentially of amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues in the final protein.

The present invention further provides variant proteins that comprise amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein may contain only the variant amino acid sequence or have additional amino acid residues, such as a contiguous encoded sequence that is naturally associated with it or heterologous amino acid residues. Such a protein can have a few additional amino acid residues or can comprise many more additional amino acids. A brief description of how various types of these proteins can be made and isolated is provided below.

The variant proteins of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a variant protein operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the variant protein. "Operatively linked" indicates that the coding sequences for the variant protein and the heterologous protein are ligated in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the variant protein. In another embodiment, the fusion protein is encoded by a fusion polynucleotide that is synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST protein). A variant protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the variant protein.

In many uses, the fusion protein does not affect the activity of the variant protein. The fusion protein can include, but is not limited to, enzymatic fusion proteins, for example, beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate their purification following recombinant expression. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Fusion proteins are further described in, for example, Terpe, "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems", *Appl Microbiol Biotechnol.* 2003 Jan;60(5):523-33. Epub 2002 Nov 07; Graddis *et al.*, "Designing proteins that work using recombinant technologies", *Curr Pharm Biotechnol.* 2002 Dec;3(4):285-97; and Nilsson *et al.*,

"Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins", *Protein Expr Purif.* 1997 Oct;11(1):1-16.

The present invention also relates to further obvious variants of the variant polypeptides of the present invention, such as naturally-occurring mature forms (*e.g.*, alleleic variants), non-naturally occurring recombinantly-derived variants, and orthologs and paralogs of such proteins that share sequence homology. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude those known in the prior art before the present invention.

Further variants of the variant polypeptides disclosed in Table 1 can comprise an amino acid sequence that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with an amino acid sequence disclosed in Table 1 (or a fragment thereof) and that includes a novel amino acid residue (allele) disclosed in Table 1 (which is encoded by a novel SNP allele). Thus, an aspect of the present invention that is specifically contemplated are polypeptides that have a certain degree of sequence variation compared with the polypeptide sequences shown in Table 1, but that contain a novel amino acid residue (allele) encoded by a novel SNP allele disclosed herein. In other words, as long as a polypeptide contains a novel amino acid residue disclosed herein, other portions of the polypeptide that flank the novel amino acid residue can vary to some degree from the polypeptide sequences shown in Table 1.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the amino acid sequences disclosed herein can readily be identified as having complete sequence identity to one of the variant proteins of the present invention as well as being encoded by the same genetic locus as the variant proteins provided herein.

Orthologs of a variant peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of a variant peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from non-human mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs can be encoded by a nucleic acid sequence that hybridizes to a variant peptide-encoding nucleic acid molecule under moderate to stringent conditions depending on the degree of relatedness of the two organisms yielding the homologous proteins.

Variant proteins include, but are not limited to, proteins containing deletions, additions and substitutions in the amino acid sequence caused by the SNPs of the present invention. One class of substitutions is conserved amino acid substitutions in which a given amino acid in a polypeptide is substituted for another amino acid of like characteristics. Typical conservative substitutions are

replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in, for example, Bowie *et al.*, *Science 247*:1306-1310 (1990).

Variant proteins can be fully functional or can lack function in one or more activities, *e.g.* ability to bind another molecule, ability to catalyze a substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variations or variations in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, truncations or extensions, or a substitution, insertion, inversion, or deletion of a critical residue or in a critical region.

Amino acids that are essential for function of a protein can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)), particularly using the amino acid sequence and polymorphism information provided in Table 1. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as enzyme activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

Polypeptides can contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Accordingly, the variant proteins of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (*e.g.*, polyethylene glycol), or in which additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known protein modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such protein modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993); Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.*, *Meth. Enzymol. 182*: 626-646 (1990); and Rattan *et al.*, *Ann. N.Y. Acad. Sci. 663*:48-62 (1992).

The present invention further provides fragments of the variant proteins in which the fragments contain one or more amino acid sequence variations (*e.g.*, substitutions, or truncations or extensions due to creation or destruction of a stop codon) encoded by one or more SNPs disclosed herein. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that have been disclosed in the prior art before the present invention.

As used herein, a fragment may comprise at least about 4, 8, 10, 12, 14, 16, 18, 20, 25, 30, 50, 100 (or any other number in-between) or more contiguous amino acid residues from a variant protein, wherein at least one amino acid residue is affected by a SNP of the present invention, *e.g.*, a variant amino acid residue encoded by a nonsynonymous nucleotide substitution at a cSNP position provided by the present invention. The variant amino acid encoded by a cSNP may occupy any residue position along the sequence of the fragment. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the variant protein or the ability to perform a function, *e.g.*, act as an immunogen. Particularly important fragments are biologically active fragments. Such fragments will typically comprise a domain or motif of a variant protein of the present invention, *e.g.*, active site, transmembrane domain, or ligand/substrate binding domain. Other fragments include, but are not limited to, domain or motif-containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and

functional sites are readily identifiable by computer programs well known to those of skill in the art (*e.g.*, PROSITE analysis) (*Current Protocols in Protein Science*, John Wiley & Sons, N.Y. (2002)).

**Uses of Variant Proteins**

The variant proteins of the present invention can be used in a variety of ways, including but not limited to, in assays to determine the biological activity of a variant protein, such as in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another type of immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the variant protein (or its binding partner) in biological fluids; as a marker for cells or tissues in which it is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); as a target for screening for a therapeutic agent; and as a direct therapeutic agent to be administered into a human subject. Any of the variant proteins disclosed herein may be developed into reagent grade or kit format for commercialization as research products. Methods for performing the uses listed above are well known to those skilled in the art (see, *e.g.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Sambrook and Russell, 2000, and Methods in Enzymology: Guide to Molecular Cloning Techniques, Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987).

In a specific embodiment of the invention, the methods of the present invention include detection of one or more variant proteins disclosed herein. Variant proteins are disclosed in Table 1 and in the Sequence Listing as SEQ ID NOS: 17-32. Detection of such proteins can be accomplished using, for example, antibodies, small molecule compounds, aptamers, ligands/substrates, other proteins or protein fragments, or other protein-binding agents. Preferably, protein detection agents are specific for a variant protein of the present invention and can therefore discriminate between a variant protein of the present invention and the wild-type protein or another variant form. This can generally be accomplished by, for example, selecting or designing detection agents that bind to the region of a protein that differs between the variant and wild-type protein, such as a region of a protein that contains one or more amino acid substitutions that is/are encoded by a non-synonymous cSNP of the present invention, or a region of a protein that follows a nonsense mutation-type SNP that creates a stop codon thereby leading to a shorter polypeptide, or a region of a protein that follows a read-through mutation-type SNP that destroys a stop codon thereby leading to a longer polypeptide in which a portion of the polypeptide is present in one version of the polypeptide but not the other.

In another specific aspect of the invention, the variant proteins of the present invention are used as targets for diagnosing liver fibrosis or for determining predisposition to liver fibrosis in a human. Accordingly, the invention provides methods for detecting the presence of, or levels of, one or more variant proteins of the present invention in a cell, tissue, or organism. Such methods typically involve contacting a test sample with an agent (*e.g.*, an antibody, small molecule compound, or peptide) capable of interacting with the variant protein such that specific binding of the agent to the variant protein can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an array, for example, an antibody or aptamer array (arrays for protein detection may also be referred to as "protein chips"). The variant protein of interest can be isolated from a test sample and assayed for the presence of a variant amino acid sequence encoded by one or more SNPs disclosed by the present invention. The SNPs may cause changes to the protein and the corresponding protein function/activity, such as through non-synonymous substitutions in protein coding regions that can lead to amino acid substitutions, deletions, insertions, and/or rearrangements; formation or destruction of stop codons; or alteration of control elements such as promoters. SNPs may also cause inappropriate post-translational modifications.

One preferred agent for detecting a variant protein in a sample is an antibody capable of selectively binding to a variant form of the protein (antibodies are described in greater detail in the next section). Such samples include, for example, tissues, cells, and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

*In vitro* methods for detection of the variant proteins associated with liver fibrosis that are disclosed herein and fragments thereof include, but are not limited to, enzyme linked immunosorbent assays (ELISAs), radioimmunoassays (RIA), Western blots, immunoprecipitations, immunofluorescence, and protein arrays/chips (*e.g.*, arrays of antibodies or aptamers). For further information regarding immunoassays and related protein detection methods, see *Current Protocols in Immunology*, John Wiley & Sons, N.Y., and Hage, "Immunoassays", *Anal Chem.* 1999 Jun 15;71(12):294R-304R.

Additional analytic methods of detecting amino acid variants include, but are not limited to, altered electrophoretic mobility, altered tryptic peptide digest, altered protein activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, and direct amino acid sequencing.

Alternatively, variant proteins can be detected *in vivo* in a subject by introducing into the subject a labeled antibody (or other type of detection reagent) specific for a variant protein. For

example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

Other uses of the variant peptides of the present invention are based on the class or action of the protein. For example, proteins isolated from humans and their mammalian orthologs serve as targets for identifying agents (*e.g.*, small molecule drugs or antibodies) for use in therapeutic applications, particularly for modulating a biological or pathological response in a cell or tissue that expresses the protein. Pharmaceutical agents can be developed that modulate protein activity.

As an alternative to modulating gene expression, therapeutic compounds can be developed that modulate protein function. For example, many SNPs disclosed herein affect the amino acid sequence of the encoded protein (*e.g.*, non-synonymous cSNPs and nonsense mutation-type SNPs). Such alterations in the encoded amino acid sequence may affect protein function, particularly if such amino acid sequence variations occur in functional protein domains, such as catalytic domains, ATP-binding domains, or ligand/substrate binding domains. It is well established in the art that variant proteins having amino acid sequence variations in functional domains can cause or influence pathological conditions. In such instances, compounds (*e.g.*, small molecule drugs or antibodies) can be developed that target the variant protein and modulate (*e.g.*, up- or down-regulate) protein function/activity.

The therapeutic methods of the present invention further include methods that target one or more variant proteins of the present invention. Variant proteins can be targeted using, for example, small molecule compounds, antibodies, aptamers, ligands/substrates, other proteins, or other protein-binding agents. Additionally, the skilled artisan will recognize that the novel protein variants (and polymorphic nucleic acid molecules) disclosed in Table 1 may themselves be directly used as therapeutic agents by acting as competitive inhibitors of corresponding art-known proteins (or nucleic acid molecules such as mRNA molecules).

The variant proteins of the present invention are particularly useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can utilize cells that naturally express the protein, a biopsy specimen, or cell cultures. In one embodiment, cell-based assays involve recombinant host cells expressing the variant protein. Cell-free assays can be used to detect the ability of a compound to directly bind to a variant protein or to the corresponding SNP-containing nucleic acid fragment that encodes the variant protein.

A variant protein of the present invention, as well as appropriate fragments thereof, can be used in high-throughput screening assays to test candidate compounds for the ability to bind and/or modulate the activity of the variant protein. These candidate compounds can be further screened

against a protein having normal function (*e.g.*, a wild-type/non-variant protein) to further determine the effect of the compound on the protein activity. Furthermore, these compounds can be tested in animal or invertebrate systems to determine *in vivo* activity/effectiveness. Compounds can be identified that activate (agonists) or inactivate (antagonists) the variant protein, and different compounds can be identified that cause various degrees of activation or inactivation of the variant protein.

Further, the variant proteins can be used to screen a compound for the ability to stimulate or inhibit interaction between the variant protein and a target molecule that normally interacts with the protein. The target can be a ligand, a substrate or a binding partner that the protein normally interacts with (for example, epinephrine or norepinephrine). Such assays typically include the steps of combining the variant protein with a candidate compound under conditions that allow the variant protein, or fragment thereof, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the variant protein and the target, such as any of the associated effects of signal transduction.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, *e.g.*, Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (*e.g.*, members of random and partially degenerate, directed phosphopeptide libraries, see, *e.g.*, Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (*e.g.*, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (*e.g.*, molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the variant protein that competes for ligand binding. Other candidate compounds include mutant proteins or appropriate fragments containing mutations that affect variant protein function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) variant protein activity. The assays typically involve an assay of events in the signal transduction pathway that indicate protein activity. Thus, the expression of genes that are up or down-regulated in response to the variant protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is

easily detectable, such as luciferase. Alternatively, phosphorylation of the variant protein, or a variant protein target, could also be measured. Any of the biological or biochemical functions mediated by the variant protein can be used as an endpoint assay. These include all of the biochemical or biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art.

Binding and/or activating compounds can also be screened by using chimeric variant proteins in which an amino terminal extracellular domain or parts thereof, an entire transmembrane domain or subregions, and/or the carboxyl terminal intracellular domain or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate than that which is normally recognized by a variant protein. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the variant protein is derived.

The variant proteins are also useful in competition binding assays in methods designed to discover compounds that interact with the variant protein. Thus, a compound can be exposed to a variant protein under conditions that allow the compound to bind or to otherwise interact with the variant protein. A binding partner, such as ligand, that normally interacts with the variant protein is also added to the mixture. If the test compound interacts with the variant protein or its binding partner, it decreases the amount of complex formed or activity from the variant protein. This type of assay is particularly useful in screening for compounds that interact with specific regions of the variant protein (Hodgson, *Bio/technology*, 1992, Sept 10(9), 973-80).

To perform cell-free drug screening assays, it is sometimes desirable to immobilize either the variant protein or a fragment thereof, or its target molecule, to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Any method for immobilizing proteins on matrices can be used in drug screening assays. In one embodiment, a fusion protein containing an added domain allows the protein to be bound to a matrix. For example, glutathione-S-transferase/$^{125}$I fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (*e.g.*, $^{35}$S-labeled) and a candidate compound, such as a drug candidate, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads can be washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated

from the matrix, separated by SDS-PAGE, and the level of bound material found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

Either the variant protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Alternatively, antibodies reactive with the variant protein but which do not interfere with binding of the variant protein to its target molecule can be derivatized to the wells of the plate, and the variant protein trapped in the wells by antibody conjugation. Preparations of the target molecule and a candidate compound are incubated in the variant protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the protein target molecule, or which are reactive with variant protein and compete with the target molecule, and enzyme-linked assays that rely on detecting an enzymatic activity associated with the target molecule.

Modulators of variant protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the protein pathway, such as liver fibrosis. These methods of treatment typically include the steps of administering the modulators of protein activity in a pharmaceutical composition to a subject in need of such treatment.

The variant proteins, or fragments thereof, disclosed herein can themselves be directly used to treat a disorder characterized by an absence of, inappropriate, or unwanted expression or activity of the variant protein. Accordingly, methods for treatment include the use of a variant protein disclosed herein or fragments thereof.

In yet another aspect of the invention, variant proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300) to identify other proteins that bind to or interact with the variant protein and are involved in variant protein activity. Such variant protein-binding proteins are also likely to be involved in the propagation of signals by the variant proteins or variant protein targets as, for example, elements of a protein-mediated signaling pathway. Alternatively, such variant protein-binding proteins are inhibitors of the variant protein.

The two-hybrid system is based on the modular nature of most transcription factors, which typically consist of separable DNA-binding and activation domains. Briefly, the assay typically utilizes two different DNA constructs. In one construct, the gene that codes for a variant protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences,

that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a variant protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected, and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein that interacts with the variant protein.

**Antibodies Directed to Variant Proteins**

The present invention also provides antibodies that selectively bind to the variant proteins disclosed herein and fragments thereof. Such antibodies may be used to quantitatively or qualitatively detect the variant proteins of the present invention. As used herein, an antibody selectively binds a target variant protein when it binds the variant protein and does not significantly bind to non-variant proteins, *i.e.*, the antibody does not significantly bind to normal, wild-type, or art-known proteins that do not contain a variant amino acid sequence due to one or more SNPs of the present invention (variant amino acid sequences may be due to, for example, nonsynonymous cSNPs, nonsense SNPs that create a stop codon, thereby causing a truncation of a polypeptide or SNPs that cause read-through mutations resulting in an extension of a polypeptide).

As used herein, an antibody is defined in terms consistent with that recognized in the art: they are multi-subunit proteins produced by an organism in response to an antigen challenge. The antibodies of the present invention include both monoclonal antibodies and polyclonal antibodies, as well as antigen-reactive proteolytic fragments of such antibodies, such as Fab, F(ab)'$_2$, and Fv fragments. In addition, an antibody of the present invention further includes any of a variety of engineered antigen-binding molecules such as a chimeric antibody (U.S. Patent Nos. 4,816,567 and 4,816,397; Morrison *et al., Proc. Natl. Acad. Sci. USA*, 81:6851, 1984; Neuberger *et al., Nature* 312:604, 1984), a humanized antibody (U.S. Patent Nos. 5,693,762; 5,585,089; and 5,565,332), a single-chain Fv (U.S. Patent No. 4,946,778; Ward *et al., Nature* 334:544, 1989), a bispecific antibody with two binding specificities (Segal *et al., J. Immunol. Methods* 248:1, 2001; Carter, *J. Immunol.* Methods 248:7, 2001), a diabody, a triabody, and a tetrabody (Todorovska *et al., J. Immunol. Methods*, 248:47, 2001), as well as a Fab conjugate (dimer or trimer), and a minibody.

Many methods are known in the art for generating and/or identifying antibodies to a given target antigen (Harlow, Antibodies, Cold Spring Harbor Press, (1989)). In general, an isolated peptide (*e.g.*, a variant protein of the present invention) is used as an immunogen and is

administered to a mammalian organism, such as a rat, rabbit, hamster or mouse. Either a full-length protein, an antigenic peptide fragment (*e.g.*, a peptide fragment containing a region that varies between a variant protein and a corresponding wild-type protein), or a fusion protein can be used. A protein used as an immunogen may be naturally-occurring, synthetic or recombinantly produced, and may be administered in combination with an adjuvant, including but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substance such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and the like.

Monoclonal antibodies can be produced by hybridoma technology (Kohler and Milstein, *Nature*, 256:495, 1975), which immortalizes cells secreting a specific monoclonal antibody. The immortalized cell lines can be created *in vitro* by fusing two different cell types, typically lymphocytes, and tumor cells. The hybridoma cells may be cultivated *in vitro* or *in vivo*. Additionally, fully human antibodies can be generated by transgenic animals (He *et al.*, *J. Immunol.*, 169:595, 2002). Fd phage and Fd phagemid technologies may be used to generate and select recombinant antibodies *in vitro* (Hoogenboom and Chames, *Immunol. Today* 21:371, 2000; Liu *et al.*, *J. Mol. Biol.* 315:1063, 2002). The complementarity-determining regions of an antibody can be identified, and synthetic peptides corresponding to such regions may be used to mediate antigen binding (U.S. Patent No. 5,637,677).

Antibodies are preferably prepared against regions or discrete fragments of a variant protein containing a variant amino acid sequence as compared to the corresponding wild-type protein (*e.g.*, a region of a variant protein that includes an amino acid encoded by a nonsynonymous cSNP, a region affected by truncation caused by a nonsense SNP that creates a stop codon, or a region resulting from the destruction of a stop codon due to read-through mutation caused by a SNP). Furthermore, preferred regions will include those involved in function/activity and/or protein/binding partner interaction. Such fragments can be selected on a physical property, such as fragments corresponding to regions that are located on the surface of the protein, *e.g.*, hydrophilic regions, or can be selected based on sequence uniqueness, or based on the position of the variant amino acid residue(s) encoded by the SNPs provided by the present invention. An antigenic fragment will typically comprise at least about 8-10 contiguous amino acid residues in which at least one of the amino acid residues is an amino acid affected by a SNP disclosed herein. The antigenic peptide can comprise, however, at least 12, 14, 16, 20, 25, 50, 100 (or any other number in-between) or more amino acid residues, provided that at least one amino acid is affected by a SNP disclosed herein.

Detection of an antibody of the present invention can be facilitated by coupling (*i.e.*, physically linking) the antibody or an antigen-reactive fragment thereof to a detectable substance. Detectable substances include, but are not limited to, various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include $^{125}I$, $^{131}I$, $^{35}S$ or $^{3}H$.

Antibodies, particularly the use of antibodies as therapeutic agents, are reviewed in: Morgan, "Antibody therapy for Alzheimer's disease", *Expert Rev Vaccines.* 2003 Feb;2(1):53-9; Ross *et al.*, "Anticancer antibodies", *Am J Clin Pathol.* 2003 Apr;119(4):472-85; Goldenberg, "Advancing role of radiolabeled antibodies in the therapy of cancer", *Cancer Immunol Immunother.* 2003 May;52(5):281-96. Epub 2003 Mar 11; Ross *et al.*, "Antibody-based therapeutics in oncology", *Expert Rev Anticancer Ther.* 2003 Feb;3(1):107-21; Cao *et al.*, "Bispecific antibody conjugates in therapeutics", *Adv Drug Deliv Rev.* 2003 Feb 10;55(2):171-97; von Mehren *et al.*, "Monoclonal antibody therapy for cancer", *Annu Rev Med.* 2003;54:343-69. Epub 2001 Dec 03; Hudson *et al.*, "Engineered antibodies", *Nat Med.* 2003 Jan;9(1):129-34; Brekke *et al.*, "Therapeutic antibodies for human diseases at the dawn of the twenty-first century", *Nat Rev Drug Discov.* 2003 Jan;2(1):52-62 (Erratum in: *Nat Rev Drug Discov.* 2003 Mar;2(3):240); Houdebine, "Antibody manufacture in transgenic animals and comparisons with other systems", *Curr Opin Biotechnol.* 2002 Dec;13(6):625-9; Andreakos *et al.*, "Monoclonal antibodies in immune and inflammatory diseases", *Curr Opin Biotechnol.* 2002 Dec;13(6):615-20; Kellermann *et al.*, "Antibody discovery: the use of transgenic mice to generate human monoclonal antibodies for therapeutics", *Curr Opin Biotechnol.* 2002 Dec;13(6):593-7; Pini *et al.*, "Phage display and colony filter screening for high-throughput selection of antibody libraries", *Comb Chem High Throughput Screen.* 2002 Nov;5(7):503-10; Batra *et al.*, "Pharmacokinetics and biodistribution of genetically engineered antibodies", *Curr Opin Biotechnol.* 2002 Dec;13(6):603-8; and Tangri *et al.*, "Rationally engineered proteins or antibodies with absent or reduced immunogenicity", *Curr Med Chem.* 2002 Dec;9(24):2191-9.

**Uses of Antibodies**

Antibodies can be used to isolate the variant proteins of the present invention from a natural cell source or from recombinant host cells by standard techniques, such as affinity chromatography or immunoprecipitation. In addition, antibodies are useful for detecting the presence of a variant protein of the present invention in cells or tissues to determine the pattern of expression of the variant protein among various tissues in an organism and over the course of normal development or disease progression. Further, antibodies can be used to detect variant protein *in situ*, *in vitro*, in a bodily fluid, or in a cell lysate or supernatant in order to evaluate the amount and pattern of expression. Also, antibodies can be used to assess abnormal tissue distribution, abnormal expression during development, or expression in an abnormal condition, such as liver fibrosis. Additionally, antibody detection of circulating fragments of the full-length variant protein can be used to identify turnover.

Antibodies to the variant proteins of the present invention are also useful in pharmacogenomic analysis. Thus, antibodies against variant proteins encoded by alternative SNP alleles can be used to identify individuals that require modified treatment modalities.

Further, antibodies can be used to assess expression of the variant protein in disease states such as in active stages of the disease or in an individual with a predisposition to a disease related to the protein's function, particularly liver fibrosis. Antibodies specific for a variant protein encoded by a SNP-containing nucleic acid molecule of the present invention can be used to assay for the presence of the variant protein, such as to screen for predisposition to liver fibrosis as indicated by the presence of the variant protein.

Antibodies are also useful as diagnostic tools for evaluating the variant proteins in conjunction with analysis by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays well known in the art.

Antibodies are also useful for tissue typing. Thus, where a specific variant protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

Antibodies can also be used to assess aberrant subcellular localization of a variant protein in cells in various tissues. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting the expression level or the presence of variant protein or aberrant tissue distribution or developmental expression of a variant protein, antibodies directed against the variant protein or relevant fragments can be used to monitor therapeutic efficacy.

The antibodies are also useful for inhibiting variant protein function, for example, by blocking the binding of a variant protein to a binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting a variant protein's function. An antibody can be used, for example, to block or competitively inhibit binding, thus modulating (agonizing or antagonizing) the activity of a variant protein. Antibodies can be prepared against specific variant protein fragments containing sites required for function or against an intact variant protein that is associated with a cell or cell membrane. For *in vivo* administration, an antibody may be linked with an additional therapeutic payload such as a radionuclide, an enzyme, an immunogenic epitope, or a cytotoxic agent. Suitable cytotoxic agents include, but are not limited to, bacterial toxin such as diphtheria, and plant toxin such as ricin. The *in vivo* half-life of an antibody or a fragment thereof may be lengthened by pegylation through conjugation to polyethylene glycol (Leong *et al., Cytokine* 16:106, 2001).

The invention also encompasses kits for using antibodies, such as kits for detecting the presence of a variant protein in a test sample. An exemplary kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting variant proteins in a biological sample; means for determining the amount, or presence/absence of variant protein in the sample; means for comparing the amount of variant protein in the sample with a standard; and instructions for use.

**Vectors and Host Cells**

The present invention also provides vectors containing the SNP-containing nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport a SNP-containing nucleic acid molecule. When the vector is a nucleic acid molecule, the SNP-containing nucleic acid molecule can be covalently linked to the vector nucleic acid. Such vectors include, but are not limited to, a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, or MAC.

A vector can be maintained in a host cell as an extrachromosomal element where it replicates and produces additional copies of the SNP-containing nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the SNP-containing nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the SNP-containing nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors typically contain cis-acting regulatory regions that are operably linked in the vector to the SNP-containing nucleic acid molecules such that transcription of the SNP-containing nucleic acid molecules is allowed in a host cell. The SNP-containing nucleic acid molecules can also be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the SNP-containing nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequences to which the SNP-containing nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region, a ribosome-binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. A person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors (see, *e.g.*, Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

A variety of expression vectors can be used to express a SNP-containing nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors can also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, *e.g.*, cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook and

Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The regulatory sequence in a vector may provide constitutive expression in one or more host cells (*e.g.*, tissue specific expression) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor, *e.g.*, a hormone or other ligand. A variety of vectors that provide constitutive or inducible expression of a nucleic acid sequence in prokaryotic and eukaryotic host cells are well known to those of ordinary skill in the art.

A SNP-containing nucleic acid molecule can be inserted into the vector by methodology well-known in the art. Generally, the SNP-containing nucleic acid molecule that will ultimately be expressed is joined to an expression vector by cleaving the SNP-containing nucleic acid molecule and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial host cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic host cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the variant peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the variant peptides. Fusion vectors can, for example, increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting, for example, as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired variant peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes suitable for such use include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology*: *Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in a bacterial host by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic

Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the SNP-containing nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example, *E. coli* (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The SNP-containing nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast (*e.g.*, *S. cerevisiae*) include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, Cell 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The SNP-containing nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the SNP-containing nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The invention also encompasses vectors in which the SNP-containing nucleic acid molecules described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to the SNP-containing nucleic acid sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include, for example, prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells can be prepared by introducing the vector constructs described herein into the cells by techniques readily available to persons of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those described in Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Host cells can contain more than one vector. Thus, different SNP-containing nucleotide sequences can be introduced in different vectors into the same cell. Similarly, the SNP-containing nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the SNP-containing nucleic acid molecules, such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced, or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication can occur in host cells that provide functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be inserted in the same vector that contains the SNP-containing nucleic acid molecules described herein or may be in a separate vector. Markers include, for example, tetracycline or ampicillin-resistance genes for prokaryotic host cells, and dihydrofolate reductase or neomycin resistance genes for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait can be effective.

While the mature variant proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these variant proteins using RNA derived from the DNA constructs described herein.

Where secretion of the variant protein is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as G-protein-coupled receptors (GPCRs), appropriate secretion signals can be incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the variant protein is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze/thaw, sonication, mechanical disruption, use of lysing agents, and the like. The variant protein can then be recovered and purified by well-known purification methods including, for example, ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that, depending upon the host cell in which recombinant production of the variant proteins described herein occurs, they can have various glycosylation patterns, or

may be non-glycosylated, as when produced in bacteria. In addition, the variant proteins may include an initial modified methionine in some cases as a result of a host-mediated process.

For further information regarding vectors and host cells, see *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y.

**Uses of Vectors and Host Cells, and Transgenic Animals**

Recombinant host cells that express the variant proteins described herein have a variety of uses. For example, the cells are useful for producing a variant protein that can be further purified into a preparation of desired amounts of the variant protein or fragments thereof. Thus, host cells containing expression vectors are useful for variant protein production.

Host cells are also useful for conducting cell-based assays involving the variant protein or variant protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a variant protein is useful for assaying compounds that stimulate or inhibit variant protein function. Such an ability of a compound to modulate variant protein function may not be apparent from assays of the compound on the native/wild-type protein, or from cell-free assays of the compound. Recombinant host cells are also useful for assaying functional alterations in the variant proteins as compared with a known function.

Genetically-engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a non-human mammal, for example, a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA containing a SNP of the present invention which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more of its cell types or tissues. Such animals are useful for studying the function of a variant protein *in vivo*, and identifying and evaluating modulators of variant protein activity. Other examples of transgenic animals include, but are not limited to, non-human primates, sheep, dogs, cows, goats, chickens, and amphibians. Transgenic non-human mammals such as cows and goats can be used to produce variant proteins which can be secreted in the animal's milk and then recovered.

A transgenic animal can be produced by introducing a SNP-containing nucleic acid molecule into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any nucleic acid molecules that contain one or more SNPs of the present invention can potentially be introduced as a transgene into the genome of a non-human animal.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the variant protein in particular cells or tissues.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described in, for example, U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.*, and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes a non-human animal in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1 (Lakso *et al. PNAS* 89:6232-6236 (1992)). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science* 251:1351-1355 (1991)). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are generally needed. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected variant protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in, for example, Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter $G_o$ phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then

transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell (*e.g.*, a somatic cell) is isolated.

Transgenic animals containing recombinant cells that express the variant proteins described herein are useful for conducting the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could influence ligand or substrate binding, variant protein activation, signal transduction, or other processes or interactions, may not be evident from *in vitro* cell-free or cell-based assays. Thus, non-human transgenic animals of the present invention may be used to assay *in vivo* variant protein function as well as the activities of a therapeutic agent or compound that modulates variant protein function/activity or expression. Such animals are also suitable for assessing the effects of null mutations (*i.e.*, mutations that substantially or completely eliminate one or more variant protein functions).

For further information regarding transgenic animals, see Houdebine, "Antibody manufacture in transgenic animals and comparisons with other systems", *Curr Opin Biotechnol.* 2002 Dec;13(6):625-9; Petters *et al.*, "Transgenic animals as models for human disease", *Transgenic Res.* 2000;9(4-5):347-51; discussion 345-6; Wolf *et al.*, "Use of transgenic animals in understanding molecular mechanisms of toxicity", *J Pharm Pharmacol.* 1998 Jun;50(6):567-74; Echelard, "Recombinant protein production in transgenic animals", *Curr Opin Biotechnol.* 1996 Oct;7(5):536-40; Houdebine, "Transgenic animal bioreactors", *Transgenic Res.* 2000;9(4-5):305-20; Pirity *et al.*, "Embryonic stem cells, creating transgenic animals", *Methods Cell Biol.* 1998;57:279-93; and Robl *et al.*, "Artificial chromosome vectors and expression of complex proteins in transgenic animals", *Theriogenology.* 2003 Jan 1; 59 (1):107-13.

## EXAMPLES

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

### Example One: Statistical Analysis of SNPs Associated with Liver Fibrosis

A case-control genetic study was performed to determine the association of SNPs in the human genome with liver fibrosis and in particular the increased or decreased risk of developing bridging fibrosis/cirrhosis, and the rate of progression of fibrosis in HCV infected patients. The stage of fibrosis in each individual was determined according to the system of Batts *et al., Am J. Surg. Pathol.* 19:1409-1417 (1995) as reviewed by Brunt *Hepatology* 31:241-246 (2000). In particular, this study relates to additional genetic markers and haplotypes associated with advanced fibrosis in HCV-infected patients in the TLR4 region.

The study involved genotyping SNPs in DNA samples obtained from >500 HCV-infected patients. The study populations came from two clinic sites, the University of California, San Francisco (UCSF) and Virginia Commonwealth University (VCU).

I. Sample description

The first sample set was collected from the clinic site at the University of California San Francisco (UCSF). This sample set consisted of 537 patients. The percentage for No fibrosis (Stage 0), Portal/Periportal fibrosis (Stage 1-2) and Bridging fibrosis/cirrhosis (Stage 3-4) are 25.9%, 49.0%, and 25.1% respectively.

The second sample set was collected from the clinic site at the Virginia Commonwealth University (VCU). This sample set consisted of 483 patients. No fibrosis (Stage 0), Portal/Periportal fibrosis (Stage 1-2) and Bridging fibrosis/cirrhosis (Stage 3-4) are 17.4%, 34.2%, and 48.4% respectively.

See Table 5 for detailed clinical data. All patients from the sample set met the inclusion/exclusion criteria in study protocol (see Table 6). All subjects had signed informed, written consent and the study protocol was IRB approved.

II. Selection of SNPs

A total of 157 unique SNPs around the original TLR4 SNP (hCV11722237 – T399I) in the multi-gene signature CRS7 as previously disclosed (see Huang et al, Gastroenterology 2006; 130: 1679-1687; and Huang et al. Hepatology 2007; 46: 297-306) were individually genotyped in both UCSF and VCU sample sets.

The boundary of the FDM region was determined by considering the linkage disequilibrium (LD) blocks as well as characterized genes located around the original CRS7 SNP. The FDM region covered the 98 kb upstream to 161 downstream from the original T399I SNP. The SNPs selected for genotyping included 1) all tagging SNPs, 2) all SNPs with Whole-Genome-Scan (WGS) quality, 3) additional SNPs to fill-in characterized genes for coverage of 1 SNP per every ~3kb, 4) additional SNPs to fill-in the region +/-50kb from the original SNP for coverage of 1 SNP per every 10kb, 5) additional SNPs to fill-in the region further than +/-50kb from the original SNP for coverage of 1 SNP per every 20kb.

Statistical analysis was done using only the Caucasian CHC patients from the 2 US centers. Cases were defined as subjects with Bridging fibrosis/Cirrhosis (F3/F4, N=263), while Controls were No-fibrosis (F0, N=157) on liver biopsy.

III. Univariate Analysis & Results

First, univariate allelic analysis was done on the individually typed markers to determine which ones were significantly associated in the combined dataset of UCSF+VCU. For the CAUCASIAN and OTHER strata, Fisher's Exact p-value was used.

Of the 157 SNPs in the TLR4 region, nine were significantly associated with cirrhosis risk with $p<0.005$. Their odds ratios (OR=1.68-2.96) were comparable to that of the original T399I SNP. The overall frequency of the risk alleles is 23.8-94.3%. Among the nine SNPs, one missense (D299G), three intronic and four intergenic SNPs were located in or near the TLR gene; one intronic SNP was in an uncharacterized gene (47kb from TLR4). The two missense TLR4 SNPs T399I and D299G are in nearly complete linkage disequilibrium ($R^2=0.934$).

See Table 7 for data on the 9 additional SNPs in the TLR4 region significantly associated with cirrhosis risk. See Table 21 below for cross-references for the SNPs.

Table 21

| Celera SNP ID | RS number | Known As | Chromosome | B36 Position | In HapMap? | Gene Symbol |
|---|---|---|---|---|---|---|
| hCV11722141 | rs1927911 | | 9 | 119509875 | Y | TLR4 |
| hCV11722237 | rs4986791 | T399I | 9 | 119515423 | Y | TLR4 |
| hCV11722238 | rs4986790 | D299G | 9 | 119515123 | Y | TLR4 |
| hCV29292005 | rs7864330 | | 9 | 119509371 | Y | TLR4 |
| hCV29816566 | rs10448253 | | 9 | 119469292 | Y | |
| hCV31783925 | rs12375686 | | 9 | 119456993 | Y | |
| hCV31783982 | rs10818069 | | 9 | 119493792 | Y | |
| hCV31783985 | rs10818070 | | 9 | 119496316 | Y | |
| hCV31784008 | rs10759933 | | 9 | 119510193 | Y | TLR4 |
| hCV8788444 | rs1252039 | | 9 | 119485355 | Y | |
| hDV71564063 | rs11536889 | | 9 | 119517952 | Y | TLR4 |

IV. Haplotype Analysis & Results

Next, haplotype analysis was done to determine significant haplotypes in this region, using the Haplo.Score program. Flanking regions with high linkage disequilibrium (LD) were identified from the HaploView's standard LD display of pairwise D' and LOD score measures, derived from HAPMAP data (Barrett et al 2005). A sliding window of various sizes (3, 5, or 7 adjacent SNPs) was applied and a haplotype score test was performed within each window (Schaid et al, 2002). The haplotype association global p-values in the sliding windows were used to narrow down the sub-region(s) most significantly associated with cirrhosis risk.

With a window size of 3, the region with 3 SNPs in the TLR4 gene that includes the original T399I SNP had the most significant global p-value (1.37E-05), which showed a 24-fold stronger association with cirrhosis risk than the original T399I SNP by itself (3.34E-04). Since

one of the SNPs in the window (hCV11722238) is in complete LD with the original T399I SNP ($R^2$ = 0.934), it is essentially a 2-SNP haplotype.

With a window size of 5, the region with 5 SNPs in the TLR4 gene that includes the original T399I SNP had the most significant global p-value (5.37E-06), which is a 62-fold stronger association than the original SNP by itself. See Tables 8-10 for haplotype data.

V. Causal SNP Analysis & Results

Finally, logistic regression was used to determine the relative importance of each SNP conditional on the effect of other SNPs for all possible pairs in the region.

Potential causal SNPs were defined as those which, after adjusting for all other SNPs in the region, maintained a p-value <0.01 across 85% of the SNPs in the region. These 9 additional SNPs in the TLR4 region that qualified under this criteria are the same as the 9 that are defined in Table 7 (Univariate Analysis).

Using values such as allele frequencies and $R^2$ with the original SNP to measure the linkage disequilibrium (LD) between SNPs, 4 different LD blocks were identified in the TLR4 region under which these 9 SNPs fall. See Table 11 for data. This may suggest that those associated with cirrhosis risk independent of T399I may be responsible for the cirrhosis risk in this region, in addition to the original SNP.

VI. Discussion

HCV infects about 4 million people in the United States and 170 million people worldwide. Almost 85 percent of the infection becomes chronic, and up to 20 percent will progress to cirrhosis, which is the end-stage fibrosis and generally irreversible (Lauer et al. 2001). HCV is the major cause of cirrhosis and hepatocellular carcinoma (HCC), and accounts for one third of the liver transplantation. The interval between infection and the development of cirrhosis exceeds 30 years but varies widely among individuals.

Currently, there is no diagnostic assay which can identify patients predisposing to develop liver damage from chronic HCV infection. Furthermore, the current diagnosis of fibrosis stage and monitoring fibrosis progression employs liver biopsy, which is invasive, painful and expensive. The major goal of this research is to define a panel of Single Nucleotide Polymorphisms (SNPs) able to distinguish HCV infected individuals at increased risk for advancing from early stage fibrosis to cirrhosis and hepatocellular carcinoma. Such diagnostic markers can lead to better therapeutic strategies, economic models and health care policy decisions.

Based on fibrosis progression rate, chronic HCV patients can be roughly divided into three groups (Poynard at al 1997): slow, moderate and rapid fibrosers. Some host factors, such as age at infection, estimated duration of infection, alcohol consumption and gender, have been found to be associated with the progression risk. However, these host factors only account for 17% - 29% of the variability in fibrosis progression, and conflicting results have been reported in literature (Poynard at al 1997; Wright et al 2003). It is clear that other unknown factors, such as host genetic factors, may play an important role in determining the rate of fibrosis.

Recent studies suggest some gene polymorphisms influence the progression of fibrosis in patients with HCV infection (Powell et al 2003), autoimmune chronic cholestasis (Tanaka et al 1999), alcohol induced liver diseases (Yamauchi 1995) and nonalcoholic fatty liver diseases (Bernard et al 2000). However, none of them have been integrated into clinic practice for a few reasons (Bataller et al 2003). First, the limitations in study design, such as small study population, lack of replication sample set, and lack of proper control group, contributed to the contradictory results. One example is the conflicting results reported on the role of mutations in hemochromatosis gene (HFE) on fibrosis progression in HCV patients (Smith et al 2000; Thorburn 2002). Second, the genetic studies restricted to selected known genes, no genome wide scan has been performed. Third, interactions between other host factors and genetic polymorphisms remain to be determined. The approach described herein, by carrying out discovery and replication study on well designed, multi-center large cohorts, employing both candidate gene and genome-wide scan should address all these challenges.

Understanding the pathogenesis of hepatic fibrosis is key to the genetic study. In liver, hepatic stellate cells play a central role in the fibrosis progression. Stellate cells comprise 15% of the total number of resident liver cells. Stellate cells constitute a heterogeneous group of cells that are functionally and anatomically similar but different in their expression of cytoskeletal filaments, their retinoid content, and their potential for ECM production (Knittel et al 2000). Under normal conditions, stellate cells are quiesecent and provide the principal storage site for retinoids (Wang 1999). Following liver injury of any etiology, hepatic stellate cells undergo a response known as activation, which is the transition of quiescent cells into proliferative, fibrogenic and contractile myofibroblasts. The early stage of activation, also termed as initiation, is associated with transcriptional events and induction of immediate early genes, among those transforming growth factor b (TGFb) and Kruppel-like factor 6 (KLF6) have been clearly identified. The second stage of activation, perpetuation involves key phenotypic response mediated by increased cytokine effects and remodeling of extracellular matrix (ECM). The major phenotypic changes after activation include proliferation, contractility, fibrogenesis, matrix

degradation, chemotaxis, retinoid loss, and white blood cell chemoattration. The key mediators for each process have been identified (Friedman 2000). All the above genes and other genes in related pathways are included in the present study.

Since the pathogenesis of all liver diseases can be explained by the 'two hit' model: the 'first hit' provides the necessary setting for the development of inflammation, injury and fibrosis, which can be caused by viral infection (HCV or HBV), dysregulation of lipid transport and lipid metabolism (NASH), drug usage (drug-induced liver diseases), alcohol (ALD), autoimmune disorders (AIH, PBC and PSC) and other unknown factors (cryptogenic cirrhosis). The damage from 'first hit' will trigger the 'second hit', which involves the initiation and perpetuation of stellate cells, which can be a common pathway for all liver fibrosis. Therefore the genes and markers identified herein as being associated with HCV fibrosis may also be associated with other liver diseases.

HCV infects about 4 million people in the United States and 170 million people worldwide. It is the major cause of cirrhosis and hepatocellular carcinoma (HCC), and accounts for one third of the liver transplantation. Despite the large variability in fibrosis progression rate among HCV patients, currently there are no diagnostic tests that can differentiate these patients. The fibrosis status can only be established through liver biopsy, which is invasive, risky and expensive and must be performed multiple times to assess fibrosis rate.

The genetic markers listed, alone, or in combination with other risk factors, such as age, gender and alcohol consumption, can provide a non-invasive test that enables physicians to assess the fibrosis risk in HCV patients. Such a test offers several advantages: 1) better treatment strategy: people with high risk will be given higher priority for treatment, while the treatment for those with low risk will be delayed, alleviating them from the side effects and high cost; 2) reducing the need for repeated liver biopsy for all patients.

The invention could be used in diagnostic kits to assess the fibrosis progression risk for all HCV patients. Depending on the genotypes of one or multiple markers listed, alone or in combination with other risk factors, physicians will be able to categorize the HCV patients into slow, median and rapid fibrosers.

The invention could be used in diagnostic kits to assess the fibrosis progression risk for patients with other liver diseases, such as hepatitis B, any co-infection with other virus (such as HIV), non-alcoholic fatty liver diseases (NAFLD), drug-induced liver diseases, alcoholic liver diseases (ALD), primary biliary cirrhosis (PBC), primary sclerosing cholangitis (PSC), autoimmune hepatitis (AIH) and cryptogenic cirrhosis. Depending on the genotypes of one or

multiple markers listed in Table 7, alone or in combination with other risk factors, physicians will be able to categorize these patients into slow, median and rapid fibrosers.

The invention could be used to identify other markers that are associated with fibrosis progression risk in HCV and other liver disease patients. The markers listed can be used to identify other mutations (preferably single nucleotide polymorphisms) with similar or better predictive value, in the identified genes, and/or all genes in the region ranging from 500 Kb upstream to 500 Kb downstream of the hit.

The invention could be used to select novel targets for the treatment of HCV and other liver diseases. Genes that contain associated markers or the products of these genes can be targeted for the development of novel medicines that treat HCV and other liver diseases. Such treatments may prevent or delay disease onset, or reverse or slow down the progression of the disease. The novel medicines may be composed of small molecules, proteins, protein fragments or peptides, antibodies, or nucleic acids that modulate the function of the target gene or its products.

Genes or their products that are directly or indirectly being regulated or interact with the genes relating to these markers can be targeted for the development of novel medicines that treat HCV and other liver diseases. Such treatments may prevent or delay disease onset, or reverse or slow down the progression of the disease. The novel medicines may be composed of small molecules, proteins, protein fragments or peptides, antibodies, or nucleic acids that modulate the function of the target gene or its products.

References (corresponding to Example One)

Bataller R, North KE, Brenner DA. Genetic polymorphisms and the progression of liver fibrosis: a critical appraisal. Hepatology. 2003, 37(3):493-503.

Bernard S, Touzet S, Personne I, Lapras V, Bondon PJ, Berthezene F, Moulin P. Association between microsomal triglyceride transfer protein gene polymorphism and the biological features of liver steatosis in patients with type II diabetes. Diabetologia 2000, 43(8):995-9.

Friedman SL. Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. J Biol Chem. 2000, 275(4):2247-50.

Knittel T, Kobold D, Saile B, Grundmann A, Neubauer K, Piscaglia F, Ramadori G. Rat liver myofibroblasts and hepatic stellate cells: different cell populations of the fibroblast lineage with fibrogenic potential. Gastroenterology. 1999, 117(5):1205-21.

Lauer GM, Walker BD. Hepatitis C virus infection. N Engl J Med. 2001,345(1):41-52.

Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet. 1997, 349(9055):825-32.

Smith BC, Gorve J, Guzail MA, Day CP, Daly AK, Burt AD, Bassendine MF. Heterozygosity for hereditary hemochromatosis is associated with more fibrosis in chronic hepatitis C. Hepatology. 1998, 27(6):1695-9.

Thorburn D, Curry G, Spooner R, Spence E, Oien K, Halls D, Fox R, McCruden EA, MacSween RN, Mills PR. The role of iron and haemochromatosis gene mutations in the progression of liver disease in chronic hepatitis C. Gut. 2002, 50(2):248-52.

Wang XD. Chronic alcohol intake interferes with retinoid metabolism and signaling. Nutr Rev. 1999, 57(2):51-9.

Wright M, Goldin R, Fabre A, Lloyd J, Thomas H, Trepo C, Pradat P, Thursz M; HENCORE collaboration. Measurement and determinants of the natural history of liver fibrosis in hepatitis C virus infection: a cross sectional and longitudinal study. Gut. 2003, 52(4):574-9.

Poynard et al. Lancet 2001; 358: 958-965.

Huang et al. Gastroenterology 2006; 130: 1679-1687.

Huang et al. Hepatology 2007; 46: 297-306.

Schwabe et al. Gastroenterology 2006; 130: 1886-1990.

Machida et al. Journal of Virology 2006; 80: 866-874.

Sinwell JS and Schaid DJ. haplo.stats R package version 1.2.0

**Example Two: Multiple SNPs in the TLR4 Region are Associated with Cirrhosis Risk in Chronic Hepatitis C (CHC) Patients**

Overview

To identify genetic markers in or near the TLR4 locus that are associated with cirrhosis risk, 157 SNPs in the TLR4 region around the T399I SNP (hCV11722237), covering the region from 98kb upstream to 161kb downstream from the T399I SNP, were genotyped in Caucasians.

Clinical Characteristics

Clinical characteristics of the subjects in this analysis are shown in Table 12.

Data Analysis

Clinical endpoints were as follows: controls were subjects with no-fibrosis (F0), and cases were subjects with bridging fibrosis/cirrhosis (F3 and F4)

For statistical analysis, significant individual markers were identified by univariate analysis, causal SNPs were identified by logistic regression, and significant haplotypes were identified using haplo.stats (Sinwell JS and Schaid DJ. haplo.stats R package version 1.2.0).

Results: SNPs in the TLR4 Region Significantly Associated with Cirrhosis Risk

SNP in the TLR4 identified as being significantly associated (P<=0.01) with cirrhosis risk are shown in Table 13 (in Table 13, SNPs are numbered numerically by their positions from 5' to 3' end, "Dis" is distance (kb) between each SNP and the original T399I, $R^2$ measures the linkage disequilibrium between each SNP and the T399I SNP (hCV11722237), and causal SNPs are identified with a circle in the last column).

Sliding Window Analysis with Three-SNP Haplotypes in the TLR4 Region

Sliding window analysis was performed to evaluate haplotypes consisting of three consecutive SNPs across the entire TLR4 region. Global p value was used to measure the association of each haplotype window with cirrhosis risk. Allelic p value was used to measure the association of each individual SNP with cirrhosis risk and served as a comparison.

This analysis identified a haplotype of the following three SNPs as significantly associated with cirrhosis risk (p value 1.37E-05): D299G (hCV11722238), T399I (hCV11722237), and hDV71564063 (D299G and T399I are in complete LD).

Conclusions

Of 157 SNPs in the TLR4 region, 10 SNPs were identified as being significantly associated with cirrhosis risk with p<=0.01.

Causal analysis in the entire region identified one intronic SNP in TLR4 and one intronic SNP in JORLAW that were associated with cirrhosis risk independent of T399I, suggesting these three SNPs may be responsible for the cirrhosis risk in this region.

A three-SNP haplotype (D299G, T399I, and hDV71564063) containing the risk allele of the T399I SNP showed a 24-fold stronger association with cirrhosis risk than T399I alone (p value 1.37E-05 for the three-SNP haplotype, compared with 3.34E-04 for T399I alone).

**Example Three: Genetic Variants Associated with Risk of Developing Liver Fibrosis**

Overview

Seven genomic loci (referred to herein as "CRS7" or "CRS"), implicated by single nucleotide polymorphisms (SNPs), have recently been associated with fibrosis risk in patients

with chronic hepatitis C (CHC) (Huang et al., "A 7 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis C", *Hepatology*. 2007 Aug;46(2):297-306, which is incorporated herein by reference in its entirety). This Example relates to the analysis of other SNPs in the genomic regions of these 7 genes for association with fibrosis risk.

Dense genotyping and association testing of additional SNPs in each of the 7 regions was carried out. This led to the identification of several SNPs in the toll-like receptor 4 (*TLR4*) and syntaxin binding protein 5-like (*STXBP5L*) loci that were associated with fibrosis risk independently of the original significant SNPs. Haplotypes consisting of these independent SNPs in *TLR4* or *STXBP5L* are highly significantly associated with fibrosis risk (global $P=3.04\text{x}10^{-5}$ and $4.49\text{x}10^{-6}$, respectively). Furthermore, in addition to their association with fibrosis risk, the SNPs in the *TLR4* locus identified herein may also be associated with infectious and inflammatory diseases.

Results

The individual CRS7 predictor SNPs reside in 7 distinct chromosomal regions (Table 14), where linkage disequilibrium (LD) extends from ~20 to ~662 kbp according to the HapMap CEPH dataset. To thoroughly examine whether other SNPs in these regions associate with cirrhosis risk more strongly than and/or independently from the original markers, dense SNP genotyping was carried out in the Caucasian samples used to build the CRS signature. Common HapMap SNPs (of ≥5% allele frequency) in these regions can be efficiently tested with a minimal of 14 to 34 SNPs that are capable of tagging other untested markers at $r^2 \geq 0.8$; for the CRS7 predictor 6 region on chromosome 3, marker-marker LD is extensive (~662 kbp), but only a 163 kbp region that contains *STXBP5* and *POLQ* genes (the only two within this entire LD block) was targeted to determine which of these genes was more likely to be causal. A total of 23 to 71 SNPs was genotyped for each region; these include tagging, putative functional and other SNPs such as those in high LD with the original marker (Table 14). Coverage of the tagging SNPs by the HapMap markers that were tested ranged from 64 to 92% but was likely to be higher since additional non-HapMap markers were genotyped as well.

In one of the 7 regions, implicated by the original CRS7 predictor rs4986791 in *TLR4*, located at ~120Mbp on chromosome 9, extensive marker-marker LD was discernable across a region of ~76 kbp encompassing rs4986791. No other genes are located in this block. For fine mapping, an additional 61 SNPs were tested, and 15 of these SNPs were identified that were significantly associated with cirrhosis risk at allelic $P<0.05$; the original marker had the strongest effect although two other markers, both in high LD with rs4986791, were more significantly

associated with fibrosis risk (Table 15). Pair-wise SNP regression analysis revealed that significance of some markers could be adjusted away by other significant markers, suggesting that all were not independently associated with disease risk, as expected from marker-marker LD. Attempting to derive a most parsimonious set of independently significant markers, three groups of SNPs were identified in the *TLR4* region that were associated with fibrosis risk (Table 15). Group 1 contains 9 SNPs including the original *TLR4* marker rs4986791 and 8 other significant fine-mapping SNPs that are in relatively high LD with rs4986791; none of the 8 markers remained significant after adjustment for rs4986791, nor did rs4986791 after adjustment for any of the 8 markers, thus a causal relationship cannot be assigned by this genetic analysis. Of the other significant markers in the *TLR4* region, 5 survived adjustment for rs4986791 (regression $P<0.05$), four of which are in relatively moderate to high LD (Group 2). An intergenic SNP, rs960312, had the strongest effect, although causal relationship between this and the other 3 SNPs could not be determined as their significance could be adjusted away by each other. The third group contains only one marker, rs11536889, that shared little LD with Group 1 or 2 markers. This marker trended to significance ($P=0.086$) after adjustment for Group 2 marker rs960312 ($P=0.086$), while rs960312 remained significant after adjustment for the Group 3 marker. Because there was almost no LD between markers in these two groups and they are present in distinct haplotypes, Groups 2 and 3 markers were considered to be independently associated with disease risk. A haplotype analysis using markers with the highest effect size in each group resulted in the identification of three significant and common haplotypes, each distinguishable by one of the three independent markers (Table 16). At the global level, the haplotypes were highly significantly associated with fibrosis risk (global $P= 3.04 \times 10^{-5}$).

Other regions were similarly analyzed as above. A number of the tested markers were significant at $P<0.05$ (Table 19). Among all 58 SNPs tested in the *STXBP5L* region on chromosome 3, two markers, rs17740066 and rs2169302, remained significant after adjustment for any of the other markers (Table 17), indicating that no other marker could account for association of these two markers. When other markers were adjusted for rs17740066, only 3 markers still remained significant, one of which was rs2169302. The other two were rs13086038 and rs35827958; these two SNPs were nearly perfectly concordant ($r^2=0.97$), and neither remained significant when adjusted for the other (regression $P=0.98$ for rs13086038 after adjustment for rs35827958 and rs35827958 after adjustment for rs13086038). Similarly, certain other markers remained significant when adjusted for rs2169302 but they could be accounted for by rs17740066. Thus the most parsimonious set of independently significant markers among the markers tested would include rs17740066, rs2169302 and rs13086038/rs35827958. LD between

these independent markers was low (Table 17; $r^2$<0.02 between any pairs). In addition, haplotype analysis identified three common haplotypes, each distinguishable by one of the three independent markers (Table 18); haplotype-disease risk association was further strengthened at the global level (global P= $4.49 \times 10^{-6}$) compared with disease association at the single marker level.

With respect to other CRS7 regions, in the case of SNP predictor 5, association of the original marker rs4290029 could not be accounted for by any other fine-mapping marker and all fine-mapping markers could be accounted for by rs4290029, suggesting that rs4290029 is likely to be the sole "causal" marker; rs4290029 is located in the intergenic region between *DEGS1* encoding lipid desaturase and *NVL1* encoding nuclear VCP-like protein. For other chromosomal regions, causal relationship between the initial CRS7 markers and other similarly significant and high LD markers that were tested could not be teased apart by regression analysis. However, for the SNP predictors 3 and 7 regions, the significant and independent SNPs are located in *TRMP5* and *AQP2* genes, encoding a transient receptor potential cation channel and aquaporin 2, respectively, although both regions are gene-rich. Additionally, markers in the regions adjacent to the main LD block that contain the individual CRS7 markers were also tested.

Table 20 provides further information regarding the SNPs disclosed herein, as well as additional SNPs associated with liver fibrosis risk.

Discussion

The analysis described in the Example indicates that multiple SNPs in each of the 7 chromosomal loci that were investigated are significantly associated with fibrosis risk and that risk allele profiles are locus/gene-specific. For the *TLR4* and *STXBP5L* loci, each has three independently significant sets of SNPs, which together give rise to highly significant haplotypes modulating risk of liver fibrosis. For the other 5 loci, each appears to have one set of independent markers, one of which, rs4290029, may be deemed "causal" (defined so if its significance cannot be explained by other variants) while causality of other original and fine-mapping markers cannot be distinguished as their significance can be accounted for by each other.

The two missense variants, T339I (rs4986791) and D299G (rs4986790) in Group 1 (Table 15), attenuate receptor signaling, NFκB activation and pro-inflammatory cytokine production and impact cell growth and survival (1, 2). These variants appear to be associated with susceptibility to infectious and inflammatory diseases and other conditions, such as endotoxin hyporesponsiveness (1), Gram-negative infections and septic shock (3, 4), malaria (5),

inflammatory bowel disease (6), atherogenesis (7), gastric cancer (8) and others (9). Both of the highly correlated missense SNPs are however very rare or absent in Asians (10, 11); thus, genetic variation in *TLR4* may play a role in disease susceptibility in this population. Since allelic heterogeneity contributes to disease risk and because TLR4 is involved in pathogen recognition and activation of innate immunity, variants other than T339I or D299G may modulate risk of aforementioned and other conditions as well, which may be particularly pronounced in the Asian population (both Group 2 SNP rs960312 and Group 3 SNP rs11536889 are common in Asians, with an allelic frequency of ~25% in the HapMap). The variants described herein, particularly the variants in Groups 2 and 3, may affect disease risk by perturbing gene expression.

Materials and Methods

*Study design*

The individual SNPs comprised of the CRS7 signature were initially identified from a gene-centric, genome-wide associations study of ~25,000 SNPs (14). Additional SNPs were tested in this follow-up study to provide better coverage of each region implicated by the signature SNPs so that other potentially causal or independently significant markers could be identified. The extent of fine-mapping regions was determined by examining the block structure of LD in the HapMap CEPH dataset; markers that are present in the same LD block ("main block") as the individual CRS7 markers were primarily targeted, although some markers in the adjacent regions were also tested. Markers tested included tagging SNPs, putative functional SNPs, and others such as those in high LD with the individual CRS7 markers. Markers capable of tagging SNP diversity in the main block were selected with the tagger program under the following criteria: minor allele frequency $\geq 0.05$ and $r^2 > 0.8$; the sample set had 80% power to detect a variant of 0.05 frequency that has an effect size of 2.2 at the allelic level. The putative functional markers, such as non-synonymous SNPs and those in a putative transcription factor binding site, were selected based on both public and Celera annotation. Additional information for the selected SNPs can be found in Table 14.

*Study samples*

The 420 Caucasian samples used in this study were collected from the University of California at San Francisco (N=187) and the Virginia Commonwealth University (VCU) (N=233). They consisted of 263 cases and 157 controls where patients with fibrosis stages 3 or 4 were defined as cases and those with fibrosis stage 0 were used as controls; samples with fibrosis stage 1 or 2 were excluded from the study. Fibrosis stages were determined by biopsies read by

liver pathologists; the Batts-Ludwig scoring system was utilized in UCSF and the Knodell system in VCU (15). Participants ranged from 23-83 years of age (Mean= 53, SD ±8) with males representing 75% of cases and 61% of controls. Age at infection was estimated at 24.9±9.0 years, and duration of infection at 24.1±7.9 years. All patients provided written informed consents, and the study was approved by institutional review boards of UCSF and VCU. Additional information can be found in previous publications (14).

*Genotyping*

Cases and controls were individually genotyped in duplicate by allele-specific kinetic PCR (16). For each allele-specific PCR reaction, 0.3 ng of DNA was amplified. Genotypes were automatically called by an in-house software program followed by manual curation without any knowledge of case/control status. Genotyping accuracy is approximately 99% as described in a previous publication (17).

*Statistical Analysis*

Allelic association of the SNPs with fibrosis risk was determined by the $\chi^2$ test. Logistic regression models for each possible pair of SNPs assumed an additive effect of each additional risk allele on the log odds of fibrosis risk. Linkage disequilibrium ($r^2$) were calculated from the unphased genotype data using LDMax in the GOLD package (18).

<u>References (corresponding to Example Three)</u>

1. Arbour, N.C., Lorenz, E., Schutte, B.C., Zabner, J., Kline, J.N., Jones, M., Frees, K., Watt, J.L. and Schwartz, D.A. (2000) TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet,* 25, 187-191.
2. Guo, J., Loke, J., Zheng, F., Yea, S., Fugita, M., Tarocchi, M., Abar, O.T., Huang, H., Sninsky, J.J. and Friedman, S.L. (2009) Functional linkage of cirrhosis-predictive single nucleotide polymorphisms of Toll-like receptor 4 to hepatic stellate cell response. *Hepatology,* in press.
3. Agnese, D.M., Calvano, J.E., Hahm, S.J., Coyle, S.M., Corbett, S.A., Calvano, S.E. and Lowry, S.F. (2002) Human toll-like receptor 4 mutations but not CD14 polymorphisms are associated with an increased risk of gram-negative infections. *J Infect Dis,* 186, 1522-1525.

4. Lorenz, E., Mira, J.P., Frees, K.L. and Schwartz, D.A. (2002) Relevance of mutations in the TLR4 receptor in patients with gram-negative septic shock. *Arch Intern Med,* 162, 1028-1032.

5. Mockenhaupt, F.P., Cramer, J.P., Hamann, L., Stegemann, M.S., Eckert, J., Oh, N.R., Otchwemah, R.N., Dietz, E., Ehrhardt, S., Schroder, N.W. *et al.* (2006) Toll-like receptor (TLR) polymorphisms in African children: Common TLR-4 variants predispose to severe malaria. *Proc Natl Acad Sci U S A,* 103, 177-182.

6. Browning, B.L., Huebner, C., Petermann, I., Gearry, R.B., Barclay, M.L., Shelling, A.N. and Ferguson, L.R. (2007) Has toll-like receptor 4 been prematurely dismissed as an inflammatory bowel disease gene? Association study combined with meta-analysis shows strong evidence for association. *Am J Gastroenterol,* 102, 2504-2512.

7. Kiechl, S., Lorenz, E., Reindl, M., Wiedermann, C.J., Oberhollenzer, F., Bonora, E., Willeit, J. and Schwartz, D.A. (2002) Toll-like receptor 4 polymorphisms and atherogenesis. *N Engl J Med,* 347, 185-192.

8. Hold, G.L., Rabkin, C.S., Chow, W.H., Smith, M.G., Gammon, M.D., Risch, H.A., Vaughan, T.L., McColl, K.E., Lissowska, J., Zatonski, W. *et al.* (2007) A functional polymorphism of toll-like receptor 4 gene increases risk of gastric carcinoma and its precursors. *Gastroenterology,* 132, 905-912.

9. Ferwerda, B., McCall, M.B., Verheijen, K., Kullberg, B.J., van der Ven, A.J., Van der Meer, J.W. and Netea, M.G. (2008) Functional consequences of toll-like receptor 4 polymorphisms. *Mol Med,* 14, 346-352.

10. Hang, J., Zhou, W., Zhang, H., Sun, B., Dai, H., Su, L. and Christiani, D.C. (2004) TLR4 Asp299Gly and Thr399Ile polymorphisms are very rare in the Chinese population. *J Endotoxin Res,* 10, 238-240.

11. Yoon, H.J., Choi, J.Y., Kim, C.O., Park, Y.S., Kim, M.S., Kim, Y.K., Shin, S.Y., Kim, J.M. and Song, Y.G. (2006) Lack of Toll-like receptor 4 and 2 polymorphisms in Korean patients with bacteremia. *J Korean Med Sci,* 21, 979-982.

12. Zheng, S.L., Augustsson-Balter, K., Chang, B., Hedelin, M., Li, L., Adami, H.O., Bensen, J., Li, G., Johnasson, J.E., Turner, A.R. *et al.* (2004) Sequence variants of toll-like receptor 4 are associated with prostate cancer risk: results from the CAncer Prostate in Sweden Study. *Cancer Res,* 64, 2918-2922.

13. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S. *et al.* (2008) Genetics of gene expression and its effect on disease. *Nature,* 452, 423-428.

14. Huang, H., Shiffman, M.L., Friedman, S., Venkatesh, R., Bzowej, N., Abar, O.T., Rowland, C.M., Catanese, J.J., Leong, D.U., Sninsky, J.J. *et al.* (2007) A 7 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis C. *Hepatology,* 46, 297-306.

15. Brunt, E.M. (2000) Grading and staging the histopathological lesions of chronic hepatitis: the Knodell histology activity index and beyond. *Hepatology,* 31, 241-246.

16. Germer, S., Holland, M.J. and Higuchi, R. (2000) High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res,* 10, 258-266.

17. Li, Y., Rowland, C., Catanese, J., Morris, J., Lovestone, S., O'Donovan, M.C., Goate, A., Owen, M., Williams, J. and Grupe, A. (2008) SORL1 variants and risk of late-onset Alzheimer's disease. *Neurobiol Dis,* 29, 293-296.

18. Abecasis, G.R. and Cookson, W.O. (2000) GOLD--graphical overview of linkage disequilibrium. *Bioinformatics,* 16, 182-183.

**Example Four: Calculated Linkage Disequilibrium (LD) SNPs associated with Liver Fibrosis**

Another investigation was conducted to identify additional SNPs that are calculated to be in linkage disequilibrium (LD) with certain "interrogated SNPs" that have been found to be associated with liver fibrosis, as described herein (particularly in Examples One, Two, and Three above) and shown in the tables. The interrogated SNPs are shown in column 1 (which indicates the hCV identification numbers of each interrogated SNP) and column 2 (which indicates the public rs identification numbers of each interrogated SNP) of Table 4. The methodology is described earlier in the instant application. To summarize briefly, the power threshold ($T$) was set at an appropriate level, such as 51%, for detecting disease association using LD markers. This power threshold is based on equation (31) above, which incorporates allele frequency data from previous disease association studies, the predicted error rate for not detecting truly disease-associated markers, and a significance level of 0.05. Using this power calculation and the sample size, a threshold level of LD, or $r^2$ value, was derived for each interrogated SNP ( $r_T^2$ , equations (32) and (33) above). The threshold value $r_T^2$ is the minimum value of linkage disequilibrium between the interrogated SNP and its LD SNPs possible such that the non-interrogated SNP still retains a power greater or equal to $T$ for detecting disease association.

Based on the above methodology, LD SNPs were found for the interrogated SNPs. Several exemplary LD SNPs for the interrogated SNPs are listed in Table 4; each LD SNP is associated with its respective interrogated SNP. Also shown are the public SNP IDs (rs numbers)

for the interrogated and LD SNPs, when available, and the threshold $r^2$ value and the power used to determine this, and the $r^2$ value of linkage disequilibrium between the interrogated SNP and its corresponding LD SNP. As an example in Table 4, the interrogated SNP rs2570950 (hCV1113678) was calculated to be in LD with rs2679741 (hCV1113685) at an $r^2$ value of 0.9634, based on a 51% power calculation, thus establishing the latter SNP as a marker associated with liver fibrosis as well.

**TABLE 3**

| Marker | Alleles | Primer 1 (Allele-specific Primer) | Primer 2 (Allele-specific Primer) | Common Primer |
|---|---|---|---|---|
| hCV1002613 | C/T | ACAGAGAGGGTAAGTAACTTGTC (SEQ ID NO:359) | ACAGAGAGGGTAAGTAACTTGTT (SEQ ID NO:360) | CTTGTCAGTGCGGGGTTAAG (SEQ ID NO:361) |
| hCV1002616 | C/T | TGCCTTGGGGGTTCC (SEQ ID NO:362) | CTGCCTTGGGGGTTCT (SEQ ID NO:363) | AGGGGCATGCACAACTCT (SEQ ID NO:364) |
| hCV1113678 | A/C | GAATGGTGTGCATGACAAAA (SEQ ID NO:365) | GAATGGTGTGCATGACAAAC (SEQ ID NO:366) | ATGCTGACCCCCATCTACT (SEQ ID NO:367) |
| hCV1113693 | C/T | GAAATGCATCGACAAATAAGGC (SEQ ID NO:368) | TGAAATGCATCGACAAATAAGGT (SEQ ID NO:369) | ATTATCGGTGGTGATAAGTGAGG T (SEQ ID NO:370) |
| hCV1113699 | A/G | CTGTTTTCAAGTAAAATAATCTGAC AGTA (SEQ ID NO:371) | GTTTTCAAGTAAAATAATCTGACAG TG (SEQ ID NO:372) | ATCCTGGTGAGATACACTTTAGG TA (SEQ ID NO:373) |
| hCV1113700 | A/G | ATATTTGAGGAGTAAAAGAGTGTT CT (SEQ ID NO:374) | TTTGAGGAGTAAAAGAGTGTTCC (SEQ ID NO:375) | CCCTCACCAACCTAGTGTATATA CAAGA (SEQ ID NO:376) |
| hCV1113702 | A/G | GAGTTTGCTCTATTGGACACT (SEQ ID NO:377) | GAGTTTGCTCTATTGGACACC (SEQ ID NO:378) | GCTACATGCTTTGCAATGCTATG G (SEQ ID NO:379) |
| hCV1113704 | A/G | CAGAGGATCCTTGCAGTAAAA (SEQ ID NO:380) | CAGAGGATCCTTGCAGTAAAG (SEQ ID NO:381) | CCACCATGCTTGGCAGTATG (SEQ ID NO:382) |
| hCV1113711 | G/T | AGAAAGGTTTATGTTGTGACATTG (SEQ ID NO:383) | AAGAAAGGTTTATGTTGTGACATTT (SEQ ID NO:384) | TTCATGCTTGCTCCCTAGAAC (SEQ ID NO:385) |
| hCV1113790 | A/G | ACTTAGTTCCGACCTTTAGTTCT (SEQ ID NO:386) | CTTAGTTCCGACCTTTAGTTCC (SEQ ID NO:387) | AAGGCCTAGGAACAGAGAGATT CAAAG (SEQ ID NO:388) |
| hCV1113793 | A/G | ACCCTTTGCAGTTGATTTGTT (SEQ ID NO:389) | ACCCTTTGCAGTTGATTTGTC (SEQ ID NO:390) | ATCTACACAACCCCTTTGCTCTA CT (SEQ ID NO:391) |
| hCV1113798 | A/G | CACTACAAAACATCAGAGAGCA (SEQ ID NO:392) | ACTACAAAACATCAGAGAGCG (SEQ ID NO:393) | ACAGTGAGGCCCTGTCTC (SEQ ID NO:394) |
| hCV1113799 | C/G | CTTACCTCACATGGGTCAG (SEQ ID NO:395) | CTTACCTCACATGGGTCAC (SEQ ID NO:396) | GCCAACAAGATGGTACTGTCTAG TG (SEQ ID NO:397) |
| hCV1113800 | C/T | AGAAAGTTGGATTTGGAAGTCTTA C (SEQ ID NO:398) | AGAAAGTTGGATTTGGAAGTCTTA T (SEQ ID NO:399) | GCCAAGATTGTGCCACTGGA (SEQ ID NO:400) |
| hCV1113802 | A/G | ACTCTCTTCTAGACTTATGCAAGTA A (SEQ ID NO:401) | CTCTCTTCTAGACTTATGCAAGTA G (SEQ ID NO:402) | GCCTGTGTGAATTACTCTGACTA CA (SEQ ID NO:403) |
| hCV1113803 | C/T | AGCCCATTGCTACTTTTACTG (SEQ ID NO:404) | CAGCCCATTGCTACTTTTACTA (SEQ ID NO:405) | CAGCATGGGAGACAGAATGAGA TAC (SEQ ID NO:406) |

| | | | | |
|---|---|---|---|---|
| hCV11238745 | A/T | AACATGTTATTGGAAGGGCATAT (SEQ ID NO:407) | ACATGTTATTGGAAGGGCATAA (SEQ ID NO:408) | GTGATAATCTACAGGTAGGTCAG GAGTTAC (SEQ ID NO:409) |
| hCV11238766 | C/T | GCACAGTGGAATGAAGTTAGAAG (SEQ ID NO:410) | GCACAGTGGAATGAAGTTAGAAA (SEQ ID NO:411) | CCTTGATCCATTGGTTAAGAGAG TGTT (SEQ ID NO:412) |
| hCV11240023 | C/T | GACATGCAAAATGAGAAGATTACA C (SEQ ID NO:413) | GACATGCAAAATGAGAAGATTACA T (SEQ ID NO:414) | GCAGTCAGAATAACTGGCTTTTA ACT (SEQ ID NO:415) |
| hCV11240063 | C/G | TCAACAGCTAAAACTCTGTATCAG (SEQ ID NO:416) | TCAACAGCTAAAACTCTGTATCAC (SEQ ID NO:417) | TCCTGGACTCAAGTGATCCACAT AC (SEQ ID NO:418) |
| hCV11245274 | C/T | AGGCTTTCCTCCCCAG (SEQ ID NO:419) | AAGGCTTTCCTCCCCAA (SEQ ID NO:420) | TGATCGTGAAAGAGTGTGGATG AGA (SEQ ID NO:421) |
| hCV11245299 | G/T | ACAATGACGGTATTATCTCCATTG (SEQ ID NO:422) | ACAATGACGGTATTATCTCCATTT (SEQ ID NO:423) | CTGGGGAAGCAAGGTTTTCATAT (SEQ ID NO:424) |
| hCV11245301 | C/T | TCAATAACCAAAGGTTTTCAACTAA G (SEQ ID NO:425) | TCAATAACCAAAGGTTTTCAACTAA A (SEQ ID NO:426) | TCGCAGTTTCCAAGAACCTACTG A (SEQ ID NO:427) |
| hCV11367836 | A/T | ATGTCGTCGGGCTAATCT (SEQ ID NO:428) | TGTCGTCGGGCTAATCA (SEQ ID NO:429) | GAGGTGGGGCTGGTTCACTA (SEQ ID NO:430) |
| hCV11367838 | C/T | CCCTACCTCCAAGGTGC (SEQ ID NO:431) | CCCTACCTCCAAGGTGT (SEQ ID NO:432) | AATGTGGTGGCCTGAGAG (SEQ ID NO:433) |
| hCV11524424 | A/C | ACCTGGCTAGAAACCCTT (SEQ ID NO:434) | ACCTGGCTAGAAACCCTG (SEQ ID NO:435) | GTTTCTTTGTGTAGCCCGTGTTC A (SEQ ID NO:436) |
| hCV11722141 | A/G | TTTGACAACTGCATTCTTTTT (SEQ ID NO:437) | TTTGACAACTGCATTCTTTTC (SEQ ID NO:438) | GCATAAAGCACTTGACCTTATTT A (SEQ ID NO:439) |
| hCV11722160 | C/T | GCATACACTTTTGGAATAAATGAG AG (SEQ ID NO:440) | GCATACACTTTTGGAATAAATGAG AA (SEQ ID NO:441) | GCGGTGTAATATACACTAATAGC TTCATGT (SEQ ID NO:442) |
| hCV11722237 | C/T | CAAAGTGATTTTGGGACAAC (SEQ ID NO:443) | CAAAGTGATTTTGGGACAAT (SEQ ID NO:444) | GAATACTGAAAACTCACTCATTT GT (SEQ ID NO:445) |
| hCV11722238 | A/G | CATACTTAGACTACTACCTCGATG A (SEQ ID NO:446) | ACTTAGACTACTACCTCGATGG (SEQ ID NO:447) | ACCCTTTCAATAGTCACACTCAC CAG (SEQ ID NO:448) |
| hCV12021443 | A/G | CAATAGAAAAGAGCTCTTAGCTTTT ATA (SEQ ID NO:449) | CAATAGAAAAGAGCTCTTAGCTTTT ATG (SEQ ID NO:450) | GTCTCAAACTCCTGTTCAAGTGA T (SEQ ID NO:451) |
| hCV12023148 | T/C | TGAATCCTGTACAGCCTTACTT (SEQ ID NO:452) | TGAATCCTGTACAGCCTTACTC (SEQ ID NO:453) | CAGTTGAGCAGTTTCTAATGTGA (SEQ ID NO:454) |
| hCV130085 | A/G | TGAAGAGGTTAGTGTATTAGTGTG T (SEQ ID NO:455) | GAAGAGGTTAGTGTATTAGTGTGC (SEQ ID NO:456) | GGCACAACCTTTACAATACACTC AGAT (SEQ ID NO:457) |
| hCV1381359 | A/G | CCTAGGAAAGGGCTTCAGAA (SEQ ID NO:458) | CCTAGGAAAGGGCTTCAGAG (SEQ ID NO:459) | GTCGTTACCCTTTGGCTCATTAA TAG (SEQ ID NO:460) |

| | | | | |
|---|---|---|---|---|
| hCV1381363 | A/G | AAACAACAGAAGTTGGCTGATA (SEQ ID NO:461) | ACAACAGAAGTTGGCTGATG (SEQ ID NO:462) | CCTGTGTTGGTTTGTGACTTTATCT (SEQ ID NO:463) |
| hCV1381377 | A/G | CAGCGTGGCTGTCTGTT (SEQ ID NO:464) | CAGCGTGGCTGTCTGTC (SEQ ID NO:465) | TCCCTCTGGGACTCTCATACT (SEQ ID NO:466) |
| hCV1381379 | A/G | CACTGGTCCCCAGATTCA (SEQ ID NO:467) | ACTGGTCCCCAGATTCG (SEQ ID NO:468) | CGTCTGAAGACATAGAAGAGCAACTAG (SEQ ID NO:469) |
| hCV1420652 | A/G | CAACGCCTACATTACAGTGAA (SEQ ID NO:470) | CAACGCCTACATTACAGTGAG (SEQ ID NO:471) | CTCTCTTGAGAACCCTGTGATGAT (SEQ ID NO:472) |
| hCV1420653 | C/G | AACACTCACTCTGTACCCC (SEQ ID NO:473) | ACACTCACTCTGTACCCG (SEQ ID NO:474) | TGCGGGACTCATCAAAGACAT (SEQ ID NO:475) |
| hCV1420655 | C/T | CTCTCTCCATAGGCTTCTCC (SEQ ID NO:476) | GCTCTCTCCATAGGCTTCTCT (SEQ ID NO:477) | GGGTGTCTGTTCCTCTTGAG (SEQ ID NO:478) |
| hCV1420722 | A/G | TGGGGTAGGGACTCTGA (SEQ ID NO:479) | TGGGGTAGGGACTCTGG (SEQ ID NO:480) | GTGTGCTCAGTAAACATGTGTTGT (SEQ ID NO:481) |
| hCV1451716 | A/G | CCAAAAGTAGAGGTAACAATGAAAAT (SEQ ID NO:482) | CAAAAGTAGAGGTAACAATGAAAAC (SEQ ID NO:483) | CCACAATGCCCTCTCACATTC (SEQ ID NO:484) |
| hCV15819007 | A/G | GCTGTTGCCCAACTGT (SEQ ID NO:485) | GCTGTTGCCCAACTGC (SEQ ID NO:486) | GCAGTCCAAACCCCTGGAG (SEQ ID NO:487) |
| hCV15826462 | G/T | CTACGTTTATGTAGTGAAAGAAAGAAG (SEQ ID NO:488) | CTACGTTTATGTAGTGAAAGAAAGAAT (SEQ ID NO:489) | GCTATATTTTACTACCTCTTGACAACCTCT (SEQ ID NO:490) |
| hCV15850218 | A/C | GTTACTCAGCTTGATTAATGTTCATTTT (SEQ ID NO:491) | ACTCAGCTTGATTAATGTTCATTTG (SEQ ID NO:492) | GCACTACCAAATCGTATCACATTCAGT (SEQ ID NO:493) |
| hCV15974376 | C/T | AGATAGCATATGAACCAAGTACTTC (SEQ ID NO:494) | CAGATAGCATATGAACCAAGTACTTT (SEQ ID NO:495) | GGTAAGTGCCATTTCTCAAATACAGT (SEQ ID NO:496) |
| hCV15974378 | C/T | GTACCTCCTTCCCCACAG (SEQ ID NO:497) | GTACCTCCTTCCCCACAA (SEQ ID NO:498) | GCAAATCCCCTCCCATCCAATTTA (SEQ ID NO:499) |
| hCV1716155 | A/G | GATAACATGAGGTAATGTGACTTCTA (SEQ ID NO:500) | AACATGAGGTAATGTGACTTCTG (SEQ ID NO:501) | CCTGGCTCCTCTGAGAGAAC (SEQ ID NO:502) |
| hCV1721645 | C/G | CAGCTGGTCCAGACTG (SEQ ID NO:503) | CAGCTGGTCCAGACTC (SEQ ID NO:504) | GGCAGGGCCTTTCATCTTCT (SEQ ID NO:505) |
| hCV228233 | C/T | GCCAAAATTGAAAGTATCATTTGTG (SEQ ID NO:506) | TGCCAAAATTGAAAGTATCATTTGTA (SEQ ID NO:507) | TCTCTGAGTTCCTGATTCAAGATCTCTGTA (SEQ ID NO:508) |
| hCV231892 | G/T | TGGGGCTGCCACATTAG (SEQ ID NO:509) | TGGGGCTGCCACATTAT (SEQ ID NO:510) | TCCTTAGGGAAGATCCGTACAGT (SEQ ID NO:511) |
| hCV2430514 | G/T | TGGTTTTGCATAACAGGTATTCC (SEQ ID NO:512) | TGGTTTTGCATAACAGGTATTCA (SEQ ID NO:513) | CATCCTGCTTCAGCTGTTACCTAATTA (SEQ ID NO:514) |

| | | | | |
|---|---|---|---|---|
| hCV2430569 | G/T | AACGTTAGAATTTCCACACATTG (SEQ ID NO:515) | AAACGTTAGAATTTCCACACATTT (SEQ ID NO:516) | TCTGATTTTCCACTCAGCTTCTTA T (SEQ ID NO:517) |
| hCV25597248 | A/G | CCAGGAACTCTGCGAAT (SEQ ID NO:518) | CCAGGAACTCTGCGAAC (SEQ ID NO:519) | CCCAGAGGCCTTGAGAAAGAG (SEQ ID NO:520) |
| hCV25635059 | A/G | AACAGAGGCTCATCTTCCTTA (SEQ ID NO:521) | ACAGAGGCTCATCTTCCTTG (SEQ ID NO:522) | TGAGCACATTGGATCTGAGACA G (SEQ ID NO:523) |
| hCV25647150 | C/T | GCTGTGACTGCTTTGAGAAAC (SEQ ID NO:524) | GCTGTGACTGCTTTGAGAAAT (SEQ ID NO:525) | GGTTGCACAGTGTCTCTGATACA T (SEQ ID NO:526) |
| hCV25647151 | C/T | CAAATTTATTCCCAGCATCACTAC (SEQ ID NO:527) | CCAAATTTATTCCCAGCATCACTAT (SEQ ID NO:528) | TTGCACAGTGTCTCTGATACATC T (SEQ ID NO:529) |
| hCV25647179 | A/G | CCACTTTCCATGTTGGTAAAT (SEQ ID NO:530) | CCACTTTCCATGTTGGTAAAC (SEQ ID NO:531) | ACGTTTCTGTGTATGTGTATTTG TGTGTG (SEQ ID NO:532) |
| hCV25647188 | A/G | CATCTCTCCTTGTCTGTTCA (SEQ ID NO:533) | ATCTCTCCTTGTCTGTTCG (SEQ ID NO:534) | CAGTGAAGGAATGAATGAATACA A (SEQ ID NO:535) |
| hCV25765485 | C/G | TCTTCCCTTTCCCTTTTG (SEQ ID NO:536) | CTCTTCCCTTTCCCTTTTC (SEQ ID NO:537) | AAGAGCCCATAAATGTTGTTAAC (SEQ ID NO:538) |
| hCV26919823 | C/T | GGCAGATATTTTATAGAATGTCACT CAG (SEQ ID NO:539) | GGCAGATATTTTATAGAATGTCACT CAA (SEQ ID NO:540) | CCCAGTATCACCTCTGCAATATT CC (SEQ ID NO:541) |
| hCV26919853 | C/T | CAGGGACAATTTGACTCTTCC (SEQ ID NO:542) | ACAGGGACAATTTGACTCTTCT (SEQ ID NO:543) | ACCCTCTCTCACCACTCCTATTT A (SEQ ID NO:544) |
| hCV26954819 | A/G | GCCTAGGTAAATGTAATGTGTTCA (SEQ ID NO:545) | GCCTAGGTAAATGTAATGTGTTCG (SEQ ID NO:546) | TTCATACCAGGCAACGTAGTGT (SEQ ID NO:547) |
| hCV26954831 | A/G | GTCACACAAAATTGTGATTATACAT T (SEQ ID NO:548) | GTCACACAAAATTGTGATTATACAT C (SEQ ID NO:549) | CCAGTCATTTATCTTGGCTACCA ACTAAC (SEQ ID NO:550) |
| hCV2703984 | A/C | CCACAAATAGATATTTGCATTTCTG A (SEQ ID NO:551) | CCACAAATAGATATTTGCATTTCTG C (SEQ ID NO:552) | TGAAACAGCAGACAAGGTACTCA (SEQ ID NO:553) |
| hCV27253261 | C/T | GTATATCGAGAGGGCTTAAGAATA TG (SEQ ID NO:554) | GGTATATCGAGAGGGCTTAAGAAT ATA (SEQ ID NO:555) | CAGATAGAATTCATTTCCGGAAA CACACAT (SEQ ID NO:556) |
| hCV27481000 | A/G | CTCATTTTAATCCAGTCTCCTCTAT (SEQ ID NO:557) | CTCATTTTAATCCAGTCTCCTCTAC (SEQ ID NO:558) | TCACTGCTTGTTTTGGGCTGAAT (SEQ ID NO:559) |
| hCV27502059 | A/T | CTGTATCAATCACGCTGGAA (SEQ ID NO:560) | CTGTATCAATCACGCTGGAT (SEQ ID NO:561) | GGCTCTGGCCAGTTATACC (SEQ ID NO:562) |
| hCV27915384 | A/G | TGAGAAAGAGAAGAGTGCCAT (SEQ ID NO:563) | TGAGAAAGAGAAGAGTGCCAC (SEQ ID NO:564) | CTTCCACCTTCCAGCTTACT (SEQ ID NO:565) |
| hCV27940202 | C/T | CTGTGTCCTATTCTAAGACTTGG (SEQ ID NO:566) | ACTGTGTCCTATTCTAAGACTTGA (SEQ ID NO:567) | GCACTGCTAACAGCTCTCTTGA (SEQ ID NO:568) |

| | | | | |
|---|---|---|---|---|
| hCV27983683 | C/T | GGGAGCAATTGTCAGCTTC (SEQ ID NO:569) | GGGAGCAATTGTCAGCTTT (SEQ ID NO:570) | TGAAGTTCATACCCCACCTCATTTAT (SEQ ID NO:571) |
| hCV27998434 | A/G | CATAGAGTATAAGAGCCAAAGAGTT (SEQ ID NO:572) | CATAGAGTATAAGAGCCAAAGAGTC (SEQ ID NO:573) | CTGGATTTGTGACCCTCAACATAACATTTA (SEQ ID NO:574) |
| hCV27999672 | C/G | TGGCTAGGAGTGTACTTTACTG (SEQ ID NO:575) | TGGCTAGGAGTGTACTTTACTC (SEQ ID NO:576) | GAGCCTAGAGCACATAGAATGTACTTGA (SEQ ID NO:577) |
| hCV29230371 | A/G | GCAGGACTCAGCTCCTTT (SEQ ID NO:578) | GCAGGACTCAGCTCCTTC (SEQ ID NO:579) | CTGGGCTGCTTGACAGTTTCTAG (SEQ ID NO:580) |
| hCV29292005 | G/T | GCGCAAAGTCAGAGTTCC (SEQ ID NO:581) | GCGCAAAGTCAGAGTTCA (SEQ ID NO:582) | GTGCCTGGCACTGTTTCTATTAGAC (SEQ ID NO:583) |
| hCV29292008 | C/G | AGTGCATAATACAGTATTGTTCATTATAG (SEQ ID NO:584) | AGTGCATAATACAGTATTGTTCATTATAC (SEQ ID NO:585) | GGGAGAAATGAGGAAATAGGGAGTTGT (SEQ ID NO:586) |
| hCV2945565 | G/T | CCTCAGGCCCCAATCTAAG (SEQ ID NO:587) | CCTCAGGCCCCAATCTAAT (SEQ ID NO:588) | TGCAACGACTGATCCCTTCTT (SEQ ID NO:589) |
| hCV29690012 | A/G | GGATGTTATGCTTGCTCTAAGAT (SEQ ID NO:590) | GGATGTTATGCTTGCTCTAAGAC (SEQ ID NO:591) | CCTTCAGGGCAGAAGACACTACT (SEQ ID NO:592) |
| hCV29780197 | C/T | AGATGTGATGATAGCTGTTGAAG (SEQ ID NO:593) | GTAGATGTGATGATAGCTGTTGAAA (SEQ ID NO:594) | CATCATGCCTGGCTAACAATTACGTA (SEQ ID NO:595) |
| hCV29816566 | A/G | TGCTGAATGAATACATTGAGTCA (SEQ ID NO:596) | TGCTGAATGAATACATTGAGTCG (SEQ ID NO:597) | GCTCTTTGGAGAAGGGATCTTTAGT (SEQ ID NO:598) |
| hCV2990649 | C/T | AACAGAAGCTGGAAATGAGAC (SEQ ID NO:599) | AAACAGAAGCTGGAAATGAGAT (SEQ ID NO:600) | AGTTCCCTGTGGGTGATTCTA (SEQ ID NO:601) |
| hCV2990660 | G/T | AGGAGGCACGAGACTG (SEQ ID NO:602) | CAGGAGGCACGAGACTT (SEQ ID NO:603) | CTCGAGGCCCAGGTGTTA (SEQ ID NO:604) |
| hCV3016893 | A/G | CCCCCGGACAACACA (SEQ ID NO:605) | CCCCCGGACAACACG (SEQ ID NO:606) | TTTCCATTGACTCGTTTCCTCTTGA (SEQ ID NO:607) |
| hCV30194915 | C/T | AGATGTGAAGTAGCCAATCAC (SEQ ID NO:608) | CAGATGTGAAGTAGCCAATCAT (SEQ ID NO:609) | GGAATTGGTAGCAGTAACAGGATATT (SEQ ID NO:610) |
| hCV30338809 | C/T | GGGATCACTTGATCAAAAGATAGTTC (SEQ ID NO:611) | GGGATCACTTGATCAAAAGATAGTTT (SEQ ID NO:612) | CGCTTATACAGTATTGGTGGGAATTAGT (SEQ ID NO:613) |
| hCV30555357 | C/T | TGTTATTTGTTTCTTTTCTCTTGCC (SEQ ID NO:614) | TGTTATTTGTTTCTTTTCTCTTGCT (SEQ ID NO:615) | TTTAACTCCAAGAATACTGCCTCAAG (SEQ ID NO:616) |
| hCV3100643 | A/C | GCTAGGGGAGTAGAGGTT (SEQ ID NO:617) | GCTAGGGGAGTAGAGGTG (SEQ ID NO:618) | GTGGTTTTACTGAACAGCAGGATCAA (SEQ ID NO:619) |
| hCV3100650 | A/C | GAGTTGTATTTTGGCTCCTCA (SEQ ID NO:620) | GAGTTGTATTTTGGCTCCTCC (SEQ ID NO:621) | CTGACAACAATGCTCCAAAGTAGA (SEQ ID NO:622) |



WO 2009/061799 PCT/US2008/082455

| | | | | |
|---|---|---|---|---|
| hCV3100651 | A/G | ACTCCCAATTCTGCACTTTT (SEQ ID NO:623) | ACTCCCAATTCTGCACTTTC (SEQ ID NO:624) | GCACAACTGACTTCCAAATGGTA AAC (SEQ ID NO:625) |
| hCV31367919 | C/G | CAGCCCATAACCTGTTCTTG (SEQ ID NO:626) | CAGCCCATAACCTGTTCTTC (SEQ ID NO:627) | CTAATCAGAAGTCAGTGGGAGC TCTAC (SEQ ID NO:628) |
| hCV31456696 | C/T | CCCATGTGAGGCAACTTC (SEQ ID NO:629) | CCCATGTGAGGCAACTTT (SEQ ID NO:630) | TGCTCCCCCGACAACATC (SEQ ID NO:631) |
| hCV31456704 | A/G | AGAGTGCTCAGTTGACCA (SEQ ID NO:632) | GAGTGCTCAGTTGACCG (SEQ ID NO:633) | AACTCCTGATCTCAGGTGATCT (SEQ ID NO:634) |
| hCV31590419 | C/T | TGCGTGATCTCAGGGC (SEQ ID NO:635) | GTGCGTGATCTCAGGGT (SEQ ID NO:636) | TGTGGGGAGAGGGAGGTTTAG (SEQ ID NO:637) |
| hCV31590424 | A/G | TGCCTTGTAGGTTACTGTGT (SEQ ID NO:638) | GCCTTGTAGGTTACTGTGC (SEQ ID NO:639) | CCTCTGCAAGTCTCAGTGATAAG AG (SEQ ID NO:640) |
| hCV31711270 | A/T | GTATGTATGTAGTAATAGGAACAT GTGT (SEQ ID NO:641) | GTATGTATGTAGTAATAGGAACAT GTGA (SEQ ID NO:642) | CTGAACCTGAAATGCAGACTCTG T (SEQ ID NO:643) |
| hCV31711313 | A/G | GCAAATGGAAGGGAAGATCAT (SEQ ID NO:644) | GCAAATGGAAGGGAAGATCAC (SEQ ID NO:645) | TCCCCTTGAACAGGCTCTGAA (SEQ ID NO:646) |
| hCV31746803 | C/G | GGGGCAACCATTGCTG (SEQ ID NO:647) | GGGGCAACCATTGCTC (SEQ ID NO:648) | TGCAACCTTCAATGGTACCTCTT CT (SEQ ID NO:649) |
| hCV31746822 | C/T | CCTATCTCCTTCCCTACCC (SEQ ID NO:650) | TCCTATCTCCTTCCCTACCT (SEQ ID NO:651) | AGAATGAAGCATGCTACAATATC TGG (SEQ ID NO:652) |
| hCV31746823 | C/T | ACCTCTACATCTCACCATATACG (SEQ ID NO:653) | GACCTCTACATCTCACCATATACA (SEQ ID NO:654) | GCTTGTGGGGTGTTGCATAAGA (SEQ ID NO:655) |
| hCV31783925 | G/T | AGGACCTCTCTAATAAGCTGTAC (SEQ ID NO:656) | ATAGGACCTCTCTAATAAGCTGTA A (SEQ ID NO:657) | CGTTCATCCCTTGAGTCATCACA (SEQ ID NO:658) |
| hCV31783950 | C/T | ACCATCCTCACGTCTTCC (SEQ ID NO:659) | CACCATCCTCACGTCTTCT (SEQ ID NO:660) | TGCTGTTTTGCACCTTTGCTTAT A (SEQ ID NO:661) |
| hCV31783982 | A/C | CTACTAGCAGCAATGTATCAAGT (SEQ ID NO:662) | ACTAGCAGCAATGTATCAAGG (SEQ ID NO:663) | TTGCAAAGATTCCTGGCATCATT AGT (SEQ ID NO:664) |
| hCV31783985 | A/G | TCCTGGGTCTTATGCATACTT (SEQ ID NO:665) | TCCTGGGTCTTATGCATACTC (SEQ ID NO:666) | GGCAAAGTTCATATCCTGAGGG AATTC (SEQ ID NO:667) |
| hCV31784008 | A/C | AATATCCAATATCGTGCTTGCT (SEQ ID NO:668) | TCCAATATCGTGCTTGCG (SEQ ID NO:669) | CGCATCATGGATTTGTGTGTCAT C (SEQ ID NO:670) |
| hCV3187664 | C/G | GCATGGAGACAGGTAGACAC (SEQ ID NO:671) | GCATGGAGACAGGTAGACAG (SEQ ID NO:672) | CAACATCACAGCTTAAAGATACA A (SEQ ID NO:673) |
| hCV32148849 | A/C | CCTTTATGTCAGTATAGTATGTGCT T (SEQ ID NO:674) | CCTTTATGTCAGTATAGTATGTGCT G (SEQ ID NO:675) | TCTGTACATCCTGAGCCAGTTAC TAAGTA (SEQ ID NO:676) |

| | | | | |
|---|---|---|---|---|
| hCV481173 | C/T | CAGGCGATGACTTAGTTTGTC (SEQ ID NO:677) | CAGGCGATGACTTAGTTTGTT (SEQ ID NO:678) | ACCCCGACCACCTCTATGTATA (SEQ ID NO:679) |
| hCV509813 | A/G | CCGTGCCATGCAAGA (SEQ ID NO:680) | CCGTGCCATGCAAGG (SEQ ID NO:681) | TCTCAGCCCACTAAGAACTAACA T (SEQ ID NO:682) |
| hCV670794 | A/G | CACAGACCCTCCTCCAA (SEQ ID NO:683) | CACAGACCCTCCTCCAG (SEQ ID NO:684) | GCTAATGAGGCCCCAGAGACTA (SEQ ID NO:685) |
| hCV670833 | A/G | GCGGTGCAGGGTTAT (SEQ ID NO:686) | GCGGTGCAGGGTTAC (SEQ ID NO:687) | CTTCCTTCAGTGCTGGGAGAGA A (SEQ ID NO:688) |
| hCV8247698 | A/T | AAAGCATGCTAGGTACTTTGT (SEQ ID NO:689) | AAAGCATGCTAGGTACTTTGA (SEQ ID NO:690) | CCACTAGCAACTTGGTGAGAAAT TTGT (SEQ ID NO:691) |
| hCV8788422 | C/T | TCCTGATCTATAAATGCTTCATCAG (SEQ ID NO:692) | TCCTGATCTATAAATGCTTCATCAA (SEQ ID NO:693) | GCAGTGTGAGAACAGATGAATA CATACAG (SEQ ID NO:694) |
| hCV8788434 | A/G | GTTTCTGAAGTCTATGGTTTATTTT AGTA (SEQ ID NO:695) | GTTTCTGAAGTCTATGGTTTATTTT AGTG (SEQ ID NO:696) | ACTGTGTGCATGCATTAACAGT (SEQ ID NO:697) |
| hCV8788444 | A/T | GAAATTCCCATTACTTCCCTTAAGT (SEQ ID NO:698) | GAAATTCCCATTACTTCCCTTAAGA (SEQ ID NO:699) | GGCAGGGCTTATTATGAGTCAAT GA (SEQ ID NO:700) |
| hCV8846755 | A/G | CTTTCTCCCCCAAGGTTTTATT (SEQ ID NO:701) | CTTTCTCCCCCAAGGTTTTATC (SEQ ID NO:702) | GGAGGGGTCTCCAACTTTACTTT G (SEQ ID NO:703) |
| hCV8847939 | A/G | GCATTTCTGGACCACAACAA (SEQ ID NO:704) | GCATTTCTGGACCACAACAG (SEQ ID NO:705) | CCATCAAATTGCCCTTGGACTTT TC (SEQ ID NO:706) |
| hCV8847948 | C/T | TGTGCCTGGCCTTACC (SEQ ID NO:707) | CTGTGCCTGGCCTTACT (SEQ ID NO:708) | CCTGCCTCTGCCATTTACTGT (SEQ ID NO:709) |
| hCV919213 | C/T | TCCTAGAGCCCACAATCTC (SEQ ID NO:710) | ATCCTAGAGCCCACAATCTT (SEQ ID NO:711) | TGGCTGGGAGGAATCTCA (SEQ ID NO:712) |
| hCV9290199 | A/G | ACTTACAGAAGGTACGAAATAAAC A (SEQ ID NO:713) | ACTTACAGAAGGTACGAAATAAAC G (SEQ ID NO:714) | TTCTGGCAGACATTTCAGTACAT TC (SEQ ID NO:715) |
| hDV70753259 | C/T | CTTTCCTCAGTTCTGCCAG (SEQ ID NO:716) | TCTTTCCTCAGTTCTGCCAA (SEQ ID NO:717) | GAGAGCTCAAAGCAGAAATACAA GAACTT (SEQ ID NO:718) |
| hDV70753261 | C/G | TCTTGACTATTTTCAGGGGTTG (SEQ ID NO:719) | CTCTTGACTATTTTCAGGGGTTC (SEQ ID NO:720) | CAACATTTTAGTGAGTGAAAGTC GGTGTTT (SEQ ID NO:721) |
| hDV70753270 | A/C | CAGGTATGTTAACTCGCACA (SEQ ID NO:722) | CAGGTATGTTAACTCGCACC (SEQ ID NO:723) | AGTGAGGGACAGACAGTTTGT (SEQ ID NO:724) |
| hDV70919856 | A/G | GCACTTGGATCCAACATCCTA (SEQ ID NO:725) | CACTTGGATCCAACATCCTG (SEQ ID NO:726) | CCATCTGGAAGAACACCGATTTT G (SEQ ID NO:727) |
| hDV70919911 | A/G | TCCAGTCCCCAACCCA (SEQ ID NO:728) | CCAGTCCCCAACCCG (SEQ ID NO:729) | ACCATCTTGTGACTAGCCATTTC A (SEQ ID NO:730) |

| | | | | |
|---|---|---|---|---|
| hDV71005086 | A/C | CTTAATCCAGCCAGTTTCAGA (SEQ ID NO:731) | AATCCAGCCAGTTTCAGC (SEQ ID NO:732) | TCCTGCCTGGACTATTTCAGTA (SEQ ID NO:733) |
| hDV71005095 | C/G | AGCCTGAGTTTATTGTTAACGG (SEQ ID NO:734) | GCCTGAGTTTATTGTTAACGC (SEQ ID NO:735) | GGAAGACACCTTCTGCAATATCTG (SEQ ID NO:736) |
| hDV71101101 | C/T | CTCCAAAACTTTAGATGATAAAATGGAG (SEQ ID NO:737) | CTCCAAAACTTTAGATGATAAAATGGAA (SEQ ID NO:738) | CTTGGTATCACCGAGCTCTTACTCT (SEQ ID NO:739) |
| hDV71115804 | A/C | GCTGAACCTACATGAGCTAAT (SEQ ID NO:740) | GCTGAACCTACATGAGCTAAG (SEQ ID NO:741) | CCATGCTTGGCAGTATGAGCTTTAT (SEQ ID NO:742) |
| hDV71153302 | A/T | TGCAGGAGAGCTGTTCTATAT (SEQ ID NO:743) | TGCAGGAGAGCTGTTCTATAA (SEQ ID NO:744) | GGCAAACAGTCTTCAAATACAATACAGACA (SEQ ID NO:745) |
| hDV71161271 | A/T | GCTGATACCAGAGTACCTTTAAAATA (SEQ ID NO:746) | GCTGATACCAGAGTACCTTTAAAATT (SEQ ID NO:747) | AGTGATTATGTGCCCCTCAATATGT (SEQ ID NO:748) |
| hDV71170845 | C/T | CCAGCTACTTGAGACCAAC (SEQ ID NO:749) | ACCAGCTACTTGAGACCAAT (SEQ ID NO:750) | CTCACATTAGAAACTGCTCAACTGA (SEQ ID NO:751) |
| hDV71206854 | C/T | CTGGATTACAGTCATGAGCTAC (SEQ ID NO:752) | CTGGATTACAGTCATGAGCTAT (SEQ ID NO:753) | AGACTTTGAAGTCATAAGAGGTAACAG (SEQ ID NO:754) |
| hDV71210383 | A/G | GCTCCTGTCCAAAACCAA (SEQ ID NO:755) | GCTCCTGTCCAAAACCAG (SEQ ID NO:756) | AGGGAATGTTGAGCTAGAGATCTAT (SEQ ID NO:757) |

**TABLE 4**

| Interrogated SNP | Interrogated rs | LD SNP | LD SNP rs | Power | Threshold $r^2$ | $r^2$ |
|---|---|---|---|---|---|---|
| hCV1113678 | rs2570950 | hCV1113685 | rs2679741 | 0.51 | 0.65378598 | 0.9634 |
| hCV1113678 | rs2570950 | hCV11240063 | rs2679742 | 0.51 | 0.65378598 | 0.9634 |
| hCV1113678 | rs2570950 | hCV11250855 | rs2679759 | 0.51 | 0.65378598 | 0.9263 |
| hCV1113678 | rs2570950 | hCV19792 | rs601588 | 0.51 | 0.65378598 | 0.9634 |
| hCV1113678 | rs2570950 | hCV20071 | rs665298 | 0.51 | 0.65378598 | 0.7907 |
| hCV1113678 | rs2570950 | hCV27253292 | rs1806299 | 0.51 | 0.65378598 | 1 |
| hCV1113678 | rs2570950 | hCV297822 | rs548488 | 0.51 | 0.65378598 | 0.7907 |
| hCV1113678 | rs2570950 | hCV502300 | rs678839 | 0.51 | 0.65378598 | 0.7907 |
| hCV1113678 | rs2570950 | hCV502301 | rs667927 | 0.51 | 0.65378598 | 1 |
| hCV1113678 | rs2570950 | hCV506571 | rs7833202 | 0.51 | 0.65378598 | 0.9634 |
| hCV1113678 | rs2570950 | hCV8846764 | rs1055376 | 0.51 | 0.65378598 | 0.963 |
| hCV1113693 | rs892484 | hCV1113699 | rs2679758 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV1113700 | rs2679757 | 0.51 | 0.69785058 | 0.9662 |
| hCV1113693 | rs892484 | hCV1113702 | rs2679754 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113704 | rs2164061 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113707 | rs2679752 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113711 | rs2570942 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113717 | rs2436845 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV1113772 | rs2513910 | 0.51 | 0.69785058 | 0.8688 |
| hCV1113693 | rs892484 | hCV1113786 | rs2513922 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113787 | rs2513921 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113793 | rs972142 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113797 | rs2304346 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV1113798 | rs1434235 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113799 | rs34655485 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV1113800 | rs2513935 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV11240023 | rs1138 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV11245299 | rs1991927 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV11245300 | rs2679749 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV11245301 | rs2679748 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV11245312 | rs10808382 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV15819007 | rs2117313 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV15918184 | rs2679746 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV15918185 | rs2679756 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV15974378 | rs2256440 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV16020129 | rs2436857 | 0.51 | 0.69785058 | 0.966 |
| hCV1113693 | rs892484 | hCV16025141 | rs2513919 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV27253261 | rs2436844 | 0.51 | 0.69785058 | 1 |
| hCV1113693 | rs892484 | hCV8846754 | rs892485 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV8846755 | rs892486 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV8847946 | rs6921 | 0.51 | 0.69785058 | 0.9664 |
| hCV1113693 | rs892484 | hCV8847948 | rs974759 | 0.51 | 0.69785058 | 1 |
| hCV1113699 | rs2679758 | hCV1113693 | rs892484 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113700 | rs2679757 | 0.51 | 0.653551679 | 0.9337 |
| hCV1113699 | rs2679758 | hCV1113702 | rs2679754 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113704 | rs2164061 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113707 | rs2679752 | 0.51 | 0.653551679 | 0.9664 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV1113699 | rs2679758 | hCV1113711 | rs2570942 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113717 | rs2436845 | 0.51 | 0.653551679 | 0.9328 |
| hCV1113699 | rs2679758 | hCV1113772 | rs2513910 | 0.51 | 0.653551679 | 0.9005 |
| hCV1113699 | rs2679758 | hCV1113786 | rs2513922 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113787 | rs2513921 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113793 | rs972142 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113797 | rs2304346 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV1113798 | rs1434235 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113799 | rs34655485 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV1113800 | rs2513935 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV11240023 | rs1138 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV11245299 | rs1991927 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV11245300 | rs2679749 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV11245301 | rs2679748 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV11245312 | rs10808382 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV15819007 | rs2117313 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV15918184 | rs2679746 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV15918185 | rs2679756 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV15974378 | rs2256440 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV16020129 | rs2436857 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV16025141 | rs2513919 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV27253261 | rs2436844 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113699 | rs2679758 | hCV8846754 | rs892485 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV8846755 | rs892486 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV8847946 | rs6921 | 0.51 | 0.653551679 | 1 |
| hCV1113699 | rs2679758 | hCV8847948 | rs974759 | 0.51 | 0.653551679 | 0.9664 |
| hCV1113700 | rs2679757 | hCV1113693 | rs892484 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113699 | rs2679758 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV1113702 | rs2679754 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113704 | rs2164061 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113707 | rs2679752 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113711 | rs2570942 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113717 | rs2436845 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV1113772 | rs2513910 | 0.51 | 0.673736854 | 0.838 |
| hCV1113700 | rs2679757 | hCV1113786 | rs2513922 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113787 | rs2513921 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113793 | rs972142 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113797 | rs2304346 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV1113798 | rs1434235 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113799 | rs34655485 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV1113800 | rs2513935 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV11240023 | rs1138 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV11245299 | rs1991927 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV11245300 | rs2679749 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV11245301 | rs2679748 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV11245312 | rs10808382 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV15819007 | rs2117313 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV15918184 | rs2679746 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV15918185 | rs2679756 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV15974378 | rs2256440 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV16020129 | rs2436857 | 0.51 | 0.673736854 | 0.9329 |
| hCV1113700 | rs2679757 | hCV16025141 | rs2513919 | 0.51 | 0.673736854 | 0.9662 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV1113700 | rs2679757 | hCV27253261 | rs2436844 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113700 | rs2679757 | hCV8846754 | rs892485 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV8846755 | rs892486 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV8847938 | rs1062315 | 0.51 | 0.673736854 | 0.6977 |
| hCV1113700 | rs2679757 | hCV8847946 | rs6921 | 0.51 | 0.673736854 | 0.9337 |
| hCV1113700 | rs2679757 | hCV8847948 | rs974759 | 0.51 | 0.673736854 | 0.9662 |
| hCV1113702 | rs2679754 | hCV1113693 | rs892484 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113699 | rs2679758 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV1113700 | rs2679757 | 0.51 | 0.664941205 | 0.9662 |
| hCV1113702 | rs2679754 | hCV1113704 | rs2164061 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113707 | rs2679752 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113711 | rs2570942 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113717 | rs2436845 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV1113772 | rs2513910 | 0.51 | 0.664941205 | 0.8688 |
| hCV1113702 | rs2679754 | hCV1113786 | rs2513922 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113787 | rs2513921 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113793 | rs972142 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113797 | rs2304346 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV1113798 | rs1434235 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113799 | rs34655485 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV1113800 | rs2513935 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV11240023 | rs1138 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV11245299 | rs1991927 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV11245300 | rs2679749 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV11245301 | rs2679748 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV11245312 | rs10808382 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV15819007 | rs2117313 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV15918184 | rs2679746 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV15918185 | rs2679756 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV15974378 | rs2256440 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV16020129 | rs2436857 | 0.51 | 0.664941205 | 0.966 |
| hCV1113702 | rs2679754 | hCV16025141 | rs2513919 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV27253261 | rs2436844 | 0.51 | 0.664941205 | 1 |
| hCV1113702 | rs2679754 | hCV8846754 | rs892485 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV8846755 | rs892486 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV8846764 | rs1055376 | 0.51 | 0.664941205 | 0.6741 |
| hCV1113702 | rs2679754 | hCV8847938 | rs1062315 | 0.51 | 0.664941205 | 0.6741 |
| hCV1113702 | rs2679754 | hCV8847946 | rs6921 | 0.51 | 0.664941205 | 0.9664 |
| hCV1113702 | rs2679754 | hCV8847948 | rs974759 | 0.51 | 0.664941205 | 1 |
| hCV1113704 | rs2164061 | hCV1113693 | rs892484 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113699 | rs2679758 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV1113700 | rs2679757 | 0.51 | 0.668971451 | 0.9662 |
| hCV1113704 | rs2164061 | hCV1113702 | rs2679754 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113707 | rs2679752 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113711 | rs2570942 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113717 | rs2436845 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV1113772 | rs2513910 | 0.51 | 0.668971451 | 0.8688 |
| hCV1113704 | rs2164061 | hCV1113786 | rs2513922 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113787 | rs2513921 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113793 | rs972142 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113797 | rs2304346 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV1113798 | rs1434235 | 0.51 | 0.668971451 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV1113704 | rs2164061 | hCV1113799 | rs34655485 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV1113800 | rs2513935 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV11240023 | rs1138 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV11245299 | rs1991927 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV11245300 | rs2679749 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV11245301 | rs2679748 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV11245312 | rs10808382 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV15819007 | rs2117313 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV15918184 | rs2679746 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV15918185 | rs2679756 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV15974378 | rs2256440 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV16020129 | rs2436857 | 0.51 | 0.668971451 | 0.966 |
| hCV1113704 | rs2164061 | hCV16025141 | rs2513919 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV27253261 | rs2436844 | 0.51 | 0.668971451 | 1 |
| hCV1113704 | rs2164061 | hCV8846754 | rs892485 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV8846755 | rs892486 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV8846764 | rs1055376 | 0.51 | 0.668971451 | 0.6741 |
| hCV1113704 | rs2164061 | hCV8847938 | rs1062315 | 0.51 | 0.668971451 | 0.6741 |
| hCV1113704 | rs2164061 | hCV8847946 | rs6921 | 0.51 | 0.668971451 | 0.9664 |
| hCV1113704 | rs2164061 | hCV8847948 | rs974759 | 0.51 | 0.668971451 | 1 |
| hCV1113711 | rs2570942 | hCV1113693 | rs892484 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113699 | rs2679758 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV1113700 | rs2679757 | 0.51 | 0.669665918 | 0.9662 |
| hCV1113711 | rs2570942 | hCV1113702 | rs2679754 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113704 | rs2164061 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113707 | rs2679752 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113717 | rs2436845 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV1113772 | rs2513910 | 0.51 | 0.669665918 | 0.8688 |
| hCV1113711 | rs2570942 | hCV1113786 | rs2513922 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113787 | rs2513921 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113793 | rs972142 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113797 | rs2304346 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV1113798 | rs1434235 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113799 | rs34655485 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV1113800 | rs2513935 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV11240023 | rs1138 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV11245299 | rs1991927 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV11245300 | rs2679749 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV11245301 | rs2679748 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV11245312 | rs10808382 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV15819007 | rs2117313 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV15918184 | rs2679746 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV15918185 | rs2679756 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV15974378 | rs2256440 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV16020129 | rs2436857 | 0.51 | 0.669665918 | 0.966 |
| hCV1113711 | rs2570942 | hCV16025141 | rs2513919 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV27253261 | rs2436844 | 0.51 | 0.669665918 | 1 |
| hCV1113711 | rs2570942 | hCV8846754 | rs892485 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV8846755 | rs892486 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113711 | rs2570942 | hCV8846764 | rs1055376 | 0.51 | 0.669665918 | 0.6741 |
| hCV1113711 | rs2570942 | hCV8847938 | rs1062315 | 0.51 | 0.669665918 | 0.6741 |
| hCV1113711 | rs2570942 | hCV8847946 | rs6921 | 0.51 | 0.669665918 | 0.9664 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV1113711 | rs2570942 | hCV8847948 | rs974759 | 0.51 | 0.669665918 | 1 |
| hCV1113790 | rs12546520 | hCV11245274 | rs12545210 | 0.51 | 0.919151226 | 0.9381 |
| hCV1113790 | rs12546520 | hCV15974376 | rs2304344 | 0.51 | 0.919151226 | 1 |
| hCV1113790 | rs12546520 | hCV506570 | rs6985891 | 0.51 | 0.919151226 | 0.9353 |
| hCV1113793 | rs972142 | hCV1113693 | rs892484 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113699 | rs2679758 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV1113700 | rs2679757 | 0.51 | 0.692900819 | 0.9662 |
| hCV1113793 | rs972142 | hCV1113702 | rs2679754 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113704 | rs2164061 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113707 | rs2679752 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113711 | rs2570942 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113717 | rs2436845 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV1113772 | rs2513910 | 0.51 | 0.692900819 | 0.8688 |
| hCV1113793 | rs972142 | hCV1113786 | rs2513922 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113787 | rs2513921 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113797 | rs2304346 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV1113798 | rs1434235 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113799 | rs34655485 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV1113800 | rs2513935 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV11240023 | rs1138 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV11245299 | rs1991927 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV11245300 | rs2679749 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV11245301 | rs2679748 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV11245312 | rs10808382 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV15819007 | rs2117313 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV15918184 | rs2679746 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV15918185 | rs2679756 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV15974378 | rs2256440 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV16020129 | rs2436857 | 0.51 | 0.692900819 | 0.966 |
| hCV1113793 | rs972142 | hCV16025141 | rs2513919 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV27253261 | rs2436844 | 0.51 | 0.692900819 | 1 |
| hCV1113793 | rs972142 | hCV8846754 | rs892485 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV8846755 | rs892486 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV8847946 | rs6921 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113793 | rs972142 | hCV8847948 | rs974759 | 0.51 | 0.692900819 | 1 |
| hCV1113798 | rs1434235 | hCV1113693 | rs892484 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113699 | rs2679758 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV1113700 | rs2679757 | 0.51 | 0.669665918 | 0.9662 |
| hCV1113798 | rs1434235 | hCV1113702 | rs2679754 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113704 | rs2164061 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113707 | rs2679752 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113711 | rs2570942 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113717 | rs2436845 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV1113772 | rs2513910 | 0.51 | 0.669665918 | 0.8688 |
| hCV1113798 | rs1434235 | hCV1113786 | rs2513922 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113787 | rs2513921 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113793 | rs972142 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113797 | rs2304346 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV1113799 | rs34655485 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV1113800 | rs2513935 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV11240023 | rs1138 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV11245299 | rs1991927 | 0.51 | 0.669665918 | 0.9664 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV1113798 | rs1434235 | hCV11245300 | rs2679749 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV11245301 | rs2679748 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV11245312 | rs10808382 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV15819007 | rs2117313 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV15918184 | rs2679746 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV15918185 | rs2679756 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV15974378 | rs2256440 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV16020129 | rs2436857 | 0.51 | 0.669665918 | 0.966 |
| hCV1113798 | rs1434235 | hCV16025141 | rs2513919 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV27253261 | rs2436844 | 0.51 | 0.669665918 | 1 |
| hCV1113798 | rs1434235 | hCV8846754 | rs892485 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV8846755 | rs892486 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV8846764 | rs1055376 | 0.51 | 0.669665918 | 0.6741 |
| hCV1113798 | rs1434235 | hCV8847938 | rs1062315 | 0.51 | 0.669665918 | 0.6741 |
| hCV1113798 | rs1434235 | hCV8847946 | rs6921 | 0.51 | 0.669665918 | 0.9664 |
| hCV1113798 | rs1434235 | hCV8847948 | rs974759 | 0.51 | 0.669665918 | 1 |
| hCV1113799 | rs34655485 | hCV1113693 | rs892484 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113699 | rs2679758 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV1113700 | rs2679757 | 0.51 | 0.692900819 | 0.9662 |
| hCV1113799 | rs34655485 | hCV1113702 | rs2679754 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113704 | rs2164061 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113707 | rs2679752 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113711 | rs2570942 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113717 | rs2436845 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV1113772 | rs2513910 | 0.51 | 0.692900819 | 0.8688 |
| hCV1113799 | rs34655485 | hCV1113786 | rs2513922 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113787 | rs2513921 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113793 | rs972142 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113797 | rs2304346 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV1113798 | rs1434235 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV1113800 | rs2513935 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV11240023 | rs1138 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV11245299 | rs1991927 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV11245300 | rs2679749 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV11245301 | rs2679748 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV11245312 | rs10808382 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV15819007 | rs2117313 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV15918184 | rs2679746 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV15918185 | rs2679756 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV15974378 | rs2256440 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV16020129 | rs2436857 | 0.51 | 0.692900819 | 0.966 |
| hCV1113799 | rs34655485 | hCV16025141 | rs2513919 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV27253261 | rs2436844 | 0.51 | 0.692900819 | 1 |
| hCV1113799 | rs34655485 | hCV8846754 | rs892485 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV8846755 | rs892486 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV8847946 | rs6921 | 0.51 | 0.692900819 | 0.9664 |
| hCV1113799 | rs34655485 | hCV8847948 | rs974759 | 0.51 | 0.692900819 | 1 |
| hCV1113800 | rs2513935 | hCV1113693 | rs892484 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113699 | rs2679758 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV1113700 | rs2679757 | 0.51 | 0.670215224 | 0.9662 |
| hCV1113800 | rs2513935 | hCV1113702 | rs2679754 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113704 | rs2164061 | 0.51 | 0.670215224 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV1113800 | rs2513935 | hCV1113707 | rs2679752 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113711 | rs2570942 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113717 | rs2436845 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV1113772 | rs2513910 | 0.51 | 0.670215224 | 0.8688 |
| hCV1113800 | rs2513935 | hCV1113786 | rs2513922 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113787 | rs2513921 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113793 | rs972142 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113797 | rs2304346 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV1113798 | rs1434235 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV1113799 | rs34655485 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV11240023 | rs1138 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV11245299 | rs1991927 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV11245300 | rs2679749 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV11245301 | rs2679748 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV11245312 | rs10808382 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV15819007 | rs2117313 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV15918184 | rs2679746 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV15918185 | rs2679756 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV15974378 | rs2256440 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV16020129 | rs2436857 | 0.51 | 0.670215224 | 0.966 |
| hCV1113800 | rs2513935 | hCV16025141 | rs2513919 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV27253261 | rs2436844 | 0.51 | 0.670215224 | 1 |
| hCV1113800 | rs2513935 | hCV8846754 | rs892485 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV8846755 | rs892486 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV8846764 | rs1055376 | 0.51 | 0.670215224 | 0.6741 |
| hCV1113800 | rs2513935 | hCV8847938 | rs1062315 | 0.51 | 0.670215224 | 0.6741 |
| hCV1113800 | rs2513935 | hCV8847946 | rs6921 | 0.51 | 0.670215224 | 0.9664 |
| hCV1113800 | rs2513935 | hCV8847948 | rs974759 | 0.51 | 0.670215224 | 1 |
| hCV11240023 | rs1138 | hCV1113678 | rs2570950 | 0.51 | 0.592214674 | 0.6171 |
| hCV11240023 | rs1138 | hCV1113685 | rs2679741 | 0.51 | 0.592214674 | 0.6422 |
| hCV11240023 | rs1138 | hCV1113693 | rs892484 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113699 | rs2679758 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV1113700 | rs2679757 | 0.51 | 0.592214674 | 0.9337 |
| hCV11240023 | rs1138 | hCV1113702 | rs2679754 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113704 | rs2164061 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113707 | rs2679752 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113711 | rs2570942 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113717 | rs2436845 | 0.51 | 0.592214674 | 0.9328 |
| hCV11240023 | rs1138 | hCV1113768 | rs2436867 | 0.51 | 0.592214674 | 0.63 |
| hCV11240023 | rs1138 | hCV1113772 | rs2513910 | 0.51 | 0.592214674 | 0.9005 |
| hCV11240023 | rs1138 | hCV1113786 | rs2513922 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113787 | rs2513921 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113793 | rs972142 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113797 | rs2304346 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV1113798 | rs1434235 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113799 | rs34655485 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV1113800 | rs2513935 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV11240063 | rs2679742 | 0.51 | 0.592214674 | 0.6422 |
| hCV11240023 | rs1138 | hCV11245299 | rs1991927 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV11245300 | rs2679749 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV11245301 | rs2679748 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV11245312 | rs10808382 | 0.51 | 0.592214674 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV11240023 | rs1138 | hCV11250855 | rs2679759 | 0.51 | 0.592214674 | 0.6171 |
| hCV11240023 | rs1138 | hCV15819007 | rs2117313 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV15918184 | rs2679746 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV15918185 | rs2679756 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV15974378 | rs2256440 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV16020129 | rs2436857 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV16025141 | rs2513919 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV19792 | rs601588 | 0.51 | 0.592214674 | 0.6422 |
| hCV11240023 | rs1138 | hCV27253261 | rs2436844 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240023 | rs1138 | hCV27253292 | rs1806299 | 0.51 | 0.592214674 | 0.6171 |
| hCV11240023 | rs1138 | hCV502301 | rs667927 | 0.51 | 0.592214674 | 0.6171 |
| hCV11240023 | rs1138 | hCV506571 | rs7833202 | 0.51 | 0.592214674 | 0.6422 |
| hCV11240023 | rs1138 | hCV8846754 | rs892485 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV8846755 | rs892486 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV8846764 | rs1055376 | 0.51 | 0.592214674 | 0.6514 |
| hCV11240023 | rs1138 | hCV8847938 | rs1062315 | 0.51 | 0.592214674 | 0.6514 |
| hCV11240023 | rs1138 | hCV8847946 | rs6921 | 0.51 | 0.592214674 | 1 |
| hCV11240023 | rs1138 | hCV8847948 | rs974759 | 0.51 | 0.592214674 | 0.9664 |
| hCV11240063 | rs2679742 | hCV1113678 | rs2570950 | 0.51 | 0.735977565 | 0.9634 |
| hCV11240063 | rs2679742 | hCV1113685 | rs2679741 | 0.51 | 0.735977565 | 1 |
| hCV11240063 | rs2679742 | hCV11250855 | rs2679759 | 0.51 | 0.735977565 | 0.9634 |
| hCV11240063 | rs2679742 | hCV19792 | rs601588 | 0.51 | 0.735977565 | 1 |
| hCV11240063 | rs2679742 | hCV20071 | rs665298 | 0.51 | 0.735977565 | 0.7618 |
| hCV11240063 | rs2679742 | hCV27253292 | rs1806299 | 0.51 | 0.735977565 | 0.9634 |
| hCV11240063 | rs2679742 | hCV297822 | rs548488 | 0.51 | 0.735977565 | 0.7618 |
| hCV11240063 | rs2679742 | hCV502300 | rs678839 | 0.51 | 0.735977565 | 0.7618 |
| hCV11240063 | rs2679742 | hCV502301 | rs667927 | 0.51 | 0.735977565 | 0.9634 |
| hCV11240063 | rs2679742 | hCV506571 | rs7833202 | 0.51 | 0.735977565 | 1 |
| hCV11240063 | rs2679742 | hCV8846764 | rs1055376 | 0.51 | 0.735977565 | 0.9277 |
| hCV11245274 | rs12545210 | hCV1113790 | rs12546520 | 0.51 | 0.838739339 | 0.9381 |
| hCV11245274 | rs12545210 | hCV15974376 | rs2304344 | 0.51 | 0.838739339 | 0.9319 |
| hCV11245274 | rs12545210 | hCV506570 | rs6985891 | 0.51 | 0.838739339 | 0.8774 |
| hCV11245299 | rs1991927 | hCV1113693 | rs892484 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113699 | rs2679758 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV1113700 | rs2679757 | 0.51 | 0.685330357 | 0.9337 |
| hCV11245299 | rs1991927 | hCV1113702 | rs2679754 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113704 | rs2164061 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113707 | rs2679752 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113711 | rs2570942 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113717 | rs2436845 | 0.51 | 0.685330357 | 0.9328 |
| hCV11245299 | rs1991927 | hCV1113772 | rs2513910 | 0.51 | 0.685330357 | 0.9005 |
| hCV11245299 | rs1991927 | hCV1113786 | rs2513922 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113787 | rs2513921 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113793 | rs972142 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113797 | rs2304346 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV1113798 | rs1434235 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113799 | rs34655485 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV1113800 | rs2513935 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV11240023 | rs1138 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV11245300 | rs2679749 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV11245301 | rs2679748 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV11245312 | rs10808382 | 0.51 | 0.685330357 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV11245299 | rs1991927 | hCV15819007 | rs2117313 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV15918184 | rs2679746 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV15918185 | rs2679756 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV15974378 | rs2256440 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV16020129 | rs2436857 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV16025141 | rs2513919 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV27253261 | rs2436844 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245299 | rs1991927 | hCV8846754 | rs892485 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV8846755 | rs892486 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV8847946 | rs6921 | 0.51 | 0.685330357 | 1 |
| hCV11245299 | rs1991927 | hCV8847948 | rs974759 | 0.51 | 0.685330357 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113685 | rs2679741 | 0.51 | 0.631705966 | 0.6422 |
| hCV11245301 | rs2679748 | hCV1113693 | rs892484 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113699 | rs2679758 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV1113700 | rs2679757 | 0.51 | 0.631705966 | 0.9337 |
| hCV11245301 | rs2679748 | hCV1113702 | rs2679754 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113704 | rs2164061 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113707 | rs2679752 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113711 | rs2570942 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113717 | rs2436845 | 0.51 | 0.631705966 | 0.9328 |
| hCV11245301 | rs2679748 | hCV1113772 | rs2513910 | 0.51 | 0.631705966 | 0.9005 |
| hCV11245301 | rs2679748 | hCV1113786 | rs2513922 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113787 | rs2513921 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113793 | rs972142 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113797 | rs2304346 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV1113798 | rs1434235 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113799 | rs34655485 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV1113800 | rs2513935 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV11240023 | rs1138 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV11240063 | rs2679742 | 0.51 | 0.631705966 | 0.6422 |
| hCV11245301 | rs2679748 | hCV11245299 | rs1991927 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV11245300 | rs2679749 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV11245312 | rs10808382 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV15819007 | rs2117313 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV15918184 | rs2679746 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV15918185 | rs2679756 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV15974378 | rs2256440 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV16020129 | rs2436857 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV16025141 | rs2513919 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV19792 | rs601588 | 0.51 | 0.631705966 | 0.6422 |
| hCV11245301 | rs2679748 | hCV27253261 | rs2436844 | 0.51 | 0.631705966 | 0.9664 |
| hCV11245301 | rs2679748 | hCV506571 | rs7833202 | 0.51 | 0.631705966 | 0.6422 |
| hCV11245301 | rs2679748 | hCV8846754 | rs892485 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV8846755 | rs892486 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV8846764 | rs1055376 | 0.51 | 0.631705966 | 0.6514 |
| hCV11245301 | rs2679748 | hCV8847938 | rs1062315 | 0.51 | 0.631705966 | 0.6514 |
| hCV11245301 | rs2679748 | hCV8847946 | rs6921 | 0.51 | 0.631705966 | 1 |
| hCV11245301 | rs2679748 | hCV8847948 | rs974759 | 0.51 | 0.631705966 | 0.9664 |
| hCV11367836 | rs2074236 | hCV11367838 | rs886277 | 0.51 | 0.439209961 | 1 |
| hCV11367836 | rs2074236 | hCV2990649 | rs753138 | 0.51 | 0.439209961 | 1 |
| hCV11367836 | rs2074236 | hCV2990660 | rs2301698 | 0.51 | 0.439209961 | 0.606 |
| hCV11367836 | rs2074236 | hCV600141 | rs757091 | 0.51 | 0.439209961 | 0.962 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV11367838 | rs886277 | hCV11367836 | rs2074236 | 0.51 | 0.410641749 | 1 |
| hCV11367838 | rs886277 | hCV2990649 | rs753138 | 0.51 | 0.410641749 | 1 |
| hCV11367838 | rs886277 | hCV2990660 | rs2301698 | 0.51 | 0.410641749 | 0.606 |
| hCV11367838 | rs886277 | hCV600141 | rs757091 | 0.51 | 0.410641749 | 0.962 |
| hCV11524424 | rs11638418 | hCV1381363 | rs12591948 | 0.51 | 0.534944271 | 1 |
| hCV11524424 | rs11638418 | hCV1381379 | rs34549499 | 0.51 | 0.534944271 | 0.9669 |
| hCV11524424 | rs11638418 | hCV25769381 | rs17240471 | 0.51 | 0.534944271 | 1 |
| hCV11524424 | rs11638418 | hCV3016886 | rs11635652 | 0.51 | 0.534944271 | 1 |
| hCV11524424 | rs11638418 | hCV3016887 | rs11639336 | 0.51 | 0.534944271 | 0.9641 |
| hCV11524424 | rs11638418 | hCV31590424 | rs11639214 | 0.51 | 0.534944271 | 0.9665 |
| hCV11524424 | rs11638418 | hCV9290199 | rs12197 | 0.51 | 0.534944271 | 1 |
| hCV11524424 | rs11638418 | hDV71583036 | rs12914541 | 0.51 | 0.534944271 | 1 |
| hCV1381363 | rs12591948 | hCV11524424 | rs11638418 | 0.51 | 0.53824862 | 1 |
| hCV1381363 | rs12591948 | hCV1381379 | rs34549499 | 0.51 | 0.53824862 | 0.9669 |
| hCV1381363 | rs12591948 | hCV25769381 | rs17240471 | 0.51 | 0.53824862 | 0.9338 |
| hCV1381363 | rs12591948 | hCV3016886 | rs11635652 | 0.51 | 0.53824862 | 1 |
| hCV1381363 | rs12591948 | hCV3016887 | rs11639336 | 0.51 | 0.53824862 | 0.8932 |
| hCV1381363 | rs12591948 | hCV31590424 | rs11639214 | 0.51 | 0.53824862 | 0.9665 |
| hCV1381363 | rs12591948 | hCV9290199 | rs12197 | 0.51 | 0.53824862 | 1 |
| hCV1381363 | rs12591948 | hDV71583036 | rs12914541 | 0.51 | 0.53824862 | 1 |
| hCV1381379 | rs34549499 | hCV11524424 | rs11638418 | 0.51 | 0.481184653 | 0.9669 |
| hCV1381379 | rs34549499 | hCV1381363 | rs12591948 | 0.51 | 0.481184653 | 0.9669 |
| hCV1381379 | rs34549499 | hCV2121926 | rs16943673 | 0.51 | 0.481184653 | 0.5143 |
| hCV1381379 | rs34549499 | hCV2121930 | rs8030258 | 0.51 | 0.481184653 | 0.4902 |
| hCV1381379 | rs34549499 | hCV25769381 | rs17240471 | 0.51 | 0.481184653 | 0.9017 |
| hCV1381379 | rs34549499 | hCV26774463 | rs4932255 | 0.51 | 0.481184653 | 0.4829 |
| hCV1381379 | rs34549499 | hCV3016884 | rs4932252 | 0.51 | 0.481184653 | 0.5137 |
| hCV1381379 | rs34549499 | hCV3016886 | rs11635652 | 0.51 | 0.481184653 | 0.9669 |
| hCV1381379 | rs34549499 | hCV3016887 | rs11639336 | 0.51 | 0.481184653 | 0.8594 |
| hCV1381379 | rs34549499 | hCV31590424 | rs11639214 | 0.51 | 0.481184653 | 0.934 |
| hCV1381379 | rs34549499 | hCV3237758 | rs10520684 | 0.51 | 0.481184653 | 0.5033 |
| hCV1381379 | rs34549499 | hCV9290199 | rs12197 | 0.51 | 0.481184653 | 0.9669 |
| hCV1381379 | rs34549499 | hDV71583036 | rs12914541 | 0.51 | 0.481184653 | 0.9669 |
| hCV15819007 | rs2117313 | hCV1113693 | rs892484 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113699 | rs2679758 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV1113700 | rs2679757 | 0.51 | 0.704392273 | 0.9337 |
| hCV15819007 | rs2117313 | hCV1113702 | rs2679754 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113704 | rs2164061 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113707 | rs2679752 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113711 | rs2570942 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113717 | rs2436845 | 0.51 | 0.704392273 | 0.9328 |
| hCV15819007 | rs2117313 | hCV1113772 | rs2513910 | 0.51 | 0.704392273 | 0.9005 |
| hCV15819007 | rs2117313 | hCV1113786 | rs2513922 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113787 | rs2513921 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113793 | rs972142 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113797 | rs2304346 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV1113798 | rs1434235 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113799 | rs34655485 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV1113800 | rs2513935 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV11240023 | rs1138 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV11245299 | rs1991927 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV11245300 | rs2679749 | 0.51 | 0.704392273 | 0.9664 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV15819007 | rs2117313 | hCV11245301 | rs2679748 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV11245312 | rs10808382 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV15918184 | rs2679746 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV15918185 | rs2679756 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV15974378 | rs2256440 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV16020129 | rs2436857 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV16025141 | rs2513919 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV27253261 | rs2436844 | 0.51 | 0.704392273 | 0.9664 |
| hCV15819007 | rs2117313 | hCV8846754 | rs892485 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV8846755 | rs892486 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV8847946 | rs6921 | 0.51 | 0.704392273 | 1 |
| hCV15819007 | rs2117313 | hCV8847948 | rs974759 | 0.51 | 0.704392273 | 0.9664 |
| hCV15974376 | rs2304344 | hCV1113790 | rs12546520 | 0.51 | 0.941697483 | 1 |
| hCV15974376 | rs2304344 | hCV506570 | rs6985891 | 0.51 | 0.941697483 | 1 |
| hCV15974378 | rs2256440 | hCV1113693 | rs892484 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113699 | rs2679758 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV1113700 | rs2679757 | 0.51 | 0.64794844 | 0.9662 |
| hCV15974378 | rs2256440 | hCV1113702 | rs2679754 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113704 | rs2164061 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113707 | rs2679752 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113711 | rs2570942 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113717 | rs2436845 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV1113772 | rs2513910 | 0.51 | 0.64794844 | 0.8688 |
| hCV15974378 | rs2256440 | hCV1113786 | rs2513922 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113787 | rs2513921 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113793 | rs972142 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113797 | rs2304346 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV1113798 | rs1434235 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113799 | rs34655485 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV1113800 | rs2513935 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV11240023 | rs1138 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV11245299 | rs1991927 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV11245300 | rs2679749 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV11245301 | rs2679748 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV11245312 | rs10808382 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV15819007 | rs2117313 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV15918184 | rs2679746 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV15918185 | rs2679756 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV16020129 | rs2436857 | 0.51 | 0.64794844 | 0.966 |
| hCV15974378 | rs2256440 | hCV16025141 | rs2513919 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV27253261 | rs2436844 | 0.51 | 0.64794844 | 1 |
| hCV15974378 | rs2256440 | hCV8846754 | rs892485 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV8846755 | rs892486 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV8846764 | rs1055376 | 0.51 | 0.64794844 | 0.6741 |
| hCV15974378 | rs2256440 | hCV8847938 | rs1062315 | 0.51 | 0.64794844 | 0.6741 |
| hCV15974378 | rs2256440 | hCV8847946 | rs6921 | 0.51 | 0.64794844 | 0.9664 |
| hCV15974378 | rs2256440 | hCV8847948 | rs974759 | 0.51 | 0.64794844 | 1 |
| hCV1721645 | rs4290029 | hCV31711270 | rs6426060 | 0.51 | 0.792667081 | 0.903 |
| hCV1721645 | rs4290029 | hCV31711301 | rs10916489 | 0.51 | 0.792667081 | 0.951 |
| hCV1721645 | rs4290029 | hCV31711306 | rs10916512 | 0.51 | 0.792667081 | 0.951 |
| hCV1721645 | rs4290029 | hCV31711311 | rs10916515 | 0.51 | 0.792667081 | 0.9476 |
| hCV2703984 | rs10513311 | hCV2704001 | rs17334309 | 0.51 | 0.920150753 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV2703984 | rs10513311 | hCV29169927 | rs7856175 | 0.51 | 0.920150753 | 1 |
| hCV2703984 | rs10513311 | hDV70726273 | rs16906276 | 0.51 | 0.920150753 | 1 |
| hCV27253261 | rs2436844 | hCV1113693 | rs892484 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113699 | rs2679758 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV1113700 | rs2679757 | 0.51 | 0.668971451 | 0.9662 |
| hCV27253261 | rs2436844 | hCV1113702 | rs2679754 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113704 | rs2164061 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113707 | rs2679752 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113711 | rs2570942 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113717 | rs2436845 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV1113772 | rs2513910 | 0.51 | 0.668971451 | 0.8688 |
| hCV27253261 | rs2436844 | hCV1113786 | rs2513922 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113787 | rs2513921 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113793 | rs972142 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113797 | rs2304346 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV1113798 | rs1434235 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113799 | rs34655485 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV1113800 | rs2513935 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV11240023 | rs1138 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV11245299 | rs1991927 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV11245300 | rs2679749 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV11245301 | rs2679748 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV11245312 | rs10808382 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV15819007 | rs2117313 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV15918184 | rs2679746 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV15918185 | rs2679756 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV15974378 | rs2256440 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV16020129 | rs2436857 | 0.51 | 0.668971451 | 0.966 |
| hCV27253261 | rs2436844 | hCV16025141 | rs2513919 | 0.51 | 0.668971451 | 1 |
| hCV27253261 | rs2436844 | hCV8846754 | rs892485 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV8846755 | rs892486 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV8846764 | rs1055376 | 0.51 | 0.668971451 | 0.6741 |
| hCV27253261 | rs2436844 | hCV8847938 | rs1062315 | 0.51 | 0.668971451 | 0.6741 |
| hCV27253261 | rs2436844 | hCV8847946 | rs6921 | 0.51 | 0.668971451 | 0.9664 |
| hCV27253261 | rs2436844 | hCV8847948 | rs974759 | 0.51 | 0.668971451 | 1 |
| hCV27502059 | rs3845856 | hCV31746823 | rs11716257 | 0.51 | 0.924147366 | 1 |
| hCV2990649 | rs753138 | hCV11367836 | rs2074236 | 0.51 | 0.410641749 | 1 |
| hCV2990649 | rs753138 | hCV11367838 | rs886277 | 0.51 | 0.410641749 | 1 |
| hCV2990649 | rs753138 | hCV2990660 | rs2301698 | 0.51 | 0.410641749 | 0.606 |
| hCV2990649 | rs753138 | hCV600141 | rs757091 | 0.51 | 0.410641749 | 0.962 |
| hCV31590424 | rs11639214 | hCV11524424 | rs11638418 | 0.51 | 0.520239484 | 0.9665 |
| hCV31590424 | rs11639214 | hCV1381363 | rs12591948 | 0.51 | 0.520239484 | 0.9665 |
| hCV31590424 | rs11639214 | hCV1381379 | rs34549499 | 0.51 | 0.520239484 | 0.934 |
| hCV31590424 | rs11639214 | hCV25769381 | rs17240471 | 0.51 | 0.520239484 | 0.9665 |
| hCV31590424 | rs11639214 | hCV3016886 | rs11635652 | 0.51 | 0.520239484 | 0.9665 |
| hCV31590424 | rs11639214 | hCV3016887 | rs11639336 | 0.51 | 0.520239484 | 1 |
| hCV31590424 | rs11639214 | hCV9290199 | rs12197 | 0.51 | 0.520239484 | 0.9665 |
| hCV31590424 | rs11639214 | hDV71583036 | rs12914541 | 0.51 | 0.520239484 | 0.9665 |
| hCV8846755 | rs892486 | hCV1113693 | rs892484 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113699 | rs2679758 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV1113700 | rs2679757 | 0.51 | 0.743899438 | 0.9337 |
| hCV8846755 | rs892486 | hCV1113702 | rs2679754 | 0.51 | 0.743899438 | 0.9664 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV8846755 | rs892486 | hCV1113704 | rs2164061 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113707 | rs2679752 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113711 | rs2570942 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113717 | rs2436845 | 0.51 | 0.743899438 | 0.9328 |
| hCV8846755 | rs892486 | hCV1113772 | rs2513910 | 0.51 | 0.743899438 | 0.9005 |
| hCV8846755 | rs892486 | hCV1113786 | rs2513922 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113787 | rs2513921 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113793 | rs972142 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113797 | rs2304346 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV1113798 | rs1434235 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113799 | rs34655485 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV1113800 | rs2513935 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV11240023 | rs1138 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV11245299 | rs1991927 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV11245300 | rs2679749 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV11245301 | rs2679748 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV11245312 | rs10808382 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV15819007 | rs2117313 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV15918184 | rs2679746 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV15918185 | rs2679756 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV15974378 | rs2256440 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV16020129 | rs2436857 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV16025141 | rs2513919 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV27253261 | rs2436844 | 0.51 | 0.743899438 | 0.9664 |
| hCV8846755 | rs892486 | hCV8846754 | rs892485 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV8847946 | rs6921 | 0.51 | 0.743899438 | 1 |
| hCV8846755 | rs892486 | hCV8847948 | rs974759 | 0.51 | 0.743899438 | 0.9664 |
| hCV8847948 | rs974759 | hCV1113678 | rs2570950 | 0.51 | 0.62716649 | 0.6396 |
| hCV8847948 | rs974759 | hCV1113693 | rs892484 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113699 | rs2679758 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV1113700 | rs2679757 | 0.51 | 0.62716649 | 0.9662 |
| hCV8847948 | rs974759 | hCV1113702 | rs2679754 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113704 | rs2164061 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113707 | rs2679752 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113711 | rs2570942 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113717 | rs2436845 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV1113772 | rs2513910 | 0.51 | 0.62716649 | 0.8688 |
| hCV8847948 | rs974759 | hCV1113786 | rs2513922 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113787 | rs2513921 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113793 | rs972142 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113797 | rs2304346 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV1113798 | rs1434235 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113799 | rs34655485 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV1113800 | rs2513935 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV11240023 | rs1138 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV11245299 | rs1991927 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV11245300 | rs2679749 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV11245301 | rs2679748 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV11245312 | rs10808382 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV15819007 | rs2117313 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV15918184 | rs2679746 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV15918185 | rs2679756 | 0.51 | 0.62716649 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hCV8847948 | rs974759 | hCV15974378 | rs2256440 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV16020129 | rs2436857 | 0.51 | 0.62716649 | 0.966 |
| hCV8847948 | rs974759 | hCV16025141 | rs2513919 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV27253261 | rs2436844 | 0.51 | 0.62716649 | 1 |
| hCV8847948 | rs974759 | hCV27253292 | rs1806299 | 0.51 | 0.62716649 | 0.6396 |
| hCV8847948 | rs974759 | hCV502301 | rs667927 | 0.51 | 0.62716649 | 0.6396 |
| hCV8847948 | rs974759 | hCV8846754 | rs892485 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV8846755 | rs892486 | 0.51 | 0.62716649 | 0.9664 |
| hCV8847948 | rs974759 | hCV8846764 | rs1055376 | 0.51 | 0.62716649 | 0.6741 |
| hCV8847948 | rs974759 | hCV8847938 | rs1062315 | 0.51 | 0.62716649 | 0.6741 |
| hCV8847948 | rs974759 | hCV8847946 | rs6921 | 0.51 | 0.62716649 | 0.9664 |
| hCV9290199 | rs12197 | hCV11524424 | rs11638418 | 0.51 | 0.55300105 | 1 |
| hCV9290199 | rs12197 | hCV1381363 | rs12591948 | 0.51 | 0.55300105 | 1 |
| hCV9290199 | rs12197 | hCV1381379 | rs34549499 | 0.51 | 0.55300105 | 0.9669 |
| hCV9290199 | rs12197 | hCV25769381 | rs17240471 | 0.51 | 0.55300105 | 0.9338 |
| hCV9290199 | rs12197 | hCV3016886 | rs11635652 | 0.51 | 0.55300105 | 1 |
| hCV9290199 | rs12197 | hCV3016887 | rs11639336 | 0.51 | 0.55300105 | 0.8932 |
| hCV9290199 | rs12197 | hCV31590424 | rs11639214 | 0.51 | 0.55300105 | 0.9665 |
| hCV9290199 | rs12197 | hDV71583036 | rs12914541 | 0.51 | 0.55300105 | 1 |
| hDV71153302 | rs6921 | hCV1113693 | rs892484 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113699 | rs2679758 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV1113700 | rs2679757 | 0.51 | 0.685330357 | 0.9337 |
| hDV71153302 | rs6921 | hCV1113702 | rs2679754 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113704 | rs2164061 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113707 | rs2679752 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113711 | rs2570942 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113717 | rs2436845 | 0.51 | 0.685330357 | 0.9328 |
| hDV71153302 | rs6921 | hCV1113772 | rs2513910 | 0.51 | 0.685330357 | 0.9005 |
| hDV71153302 | rs6921 | hCV1113786 | rs2513922 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113787 | rs2513921 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113793 | rs972142 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113797 | rs2304346 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV1113798 | rs1434235 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113799 | rs34655485 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV1113800 | rs2513935 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV11240023 | rs1138 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV11245299 | rs1991927 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV11245300 | rs2679749 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV11245301 | rs2679748 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV11245312 | rs10808382 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV15819007 | rs2117313 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV15918184 | rs2679746 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV15918185 | rs2679756 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV15974378 | rs2256440 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV16020129 | rs2436857 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV16025141 | rs2513919 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV27253261 | rs2436844 | 0.51 | 0.685330357 | 0.9664 |
| hDV71153302 | rs6921 | hCV8846754 | rs892485 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV8846755 | rs892486 | 0.51 | 0.685330357 | 1 |
| hDV71153302 | rs6921 | hCV8847948 | rs974759 | 0.51 | 0.685330357 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113693 | rs892484 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113699 | rs2679758 | 0.51 | 0.710277929 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hDV71161271 | rs10808382 | hCV1113700 | rs2679757 | 0.51 | 0.710277929 | 0.9337 |
| hDV71161271 | rs10808382 | hCV1113702 | rs2679754 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113704 | rs2164061 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113707 | rs2679752 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113711 | rs2570942 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113717 | rs2436845 | 0.51 | 0.710277929 | 0.9328 |
| hDV71161271 | rs10808382 | hCV1113772 | rs2513910 | 0.51 | 0.710277929 | 0.9005 |
| hDV71161271 | rs10808382 | hCV1113786 | rs2513922 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113787 | rs2513921 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113793 | rs972142 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113797 | rs2304346 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV1113798 | rs1434235 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113799 | rs34655485 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV1113800 | rs2513935 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV11240023 | rs1138 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV11245299 | rs1991927 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV11245300 | rs2679749 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV11245301 | rs2679748 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV15819007 | rs2117313 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV15918184 | rs2679746 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV15918185 | rs2679756 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV15974378 | rs2256440 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV16020129 | rs2436857 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV16025141 | rs2513919 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV27253261 | rs2436844 | 0.51 | 0.710277929 | 0.9664 |
| hDV71161271 | rs10808382 | hCV8846754 | rs892485 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV8846755 | rs892486 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV8847946 | rs6921 | 0.51 | 0.710277929 | 1 |
| hDV71161271 | rs10808382 | hCV8847948 | rs974759 | 0.51 | 0.710277929 | 0.9664 |
| hDV71206854 | rs1806299 | hCV1113678 | rs2570950 | 0.51 | 0.615922951 | 1 |
| hDV71206854 | rs1806299 | hCV1113685 | rs2679741 | 0.51 | 0.615922951 | 0.9634 |
| hDV71206854 | rs1806299 | hCV1113693 | rs892484 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113699 | rs2679758 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV1113702 | rs2679754 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113704 | rs2164061 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113707 | rs2679752 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113711 | rs2570942 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113717 | rs2436845 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV1113786 | rs2513922 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113787 | rs2513921 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113793 | rs972142 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113797 | rs2304346 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV1113798 | rs1434235 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113799 | rs34655485 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV1113800 | rs2513935 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV11240023 | rs1138 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV11240063 | rs2679742 | 0.51 | 0.615922951 | 0.9634 |
| hDV71206854 | rs1806299 | hCV11245299 | rs1991927 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV11245300 | rs2679749 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV11245301 | rs2679748 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV11245312 | rs10808382 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV11250855 | rs2679759 | 0.51 | 0.615922951 | 0.9263 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hDV71206854 | rs1806299 | hCV11250859 | rs10955294 | 0.51 | 0.615922951 | 0.6374 |
| hDV71206854 | rs1806299 | hCV15819007 | rs2117313 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV15918184 | rs2679746 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV15918185 | rs2679756 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV15974378 | rs2256440 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV16025141 | rs2513919 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV19792 | rs601588 | 0.51 | 0.615922951 | 0.9634 |
| hDV71206854 | rs1806299 | hCV20071 | rs665298 | 0.51 | 0.615922951 | 0.7907 |
| hDV71206854 | rs1806299 | hCV27253261 | rs2436844 | 0.51 | 0.615922951 | 0.6396 |
| hDV71206854 | rs1806299 | hCV297822 | rs548488 | 0.51 | 0.615922951 | 0.7907 |
| hDV71206854 | rs1806299 | hCV502300 | rs678839 | 0.51 | 0.615922951 | 0.7907 |
| hDV71206854 | rs1806299 | hCV502301 | rs667927 | 0.51 | 0.615922951 | 1 |
| hDV71206854 | rs1806299 | hCV506571 | rs7833202 | 0.51 | 0.615922951 | 0.9634 |
| hDV71206854 | rs1806299 | hCV8846754 | rs892485 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV8846755 | rs892486 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV8846764 | rs1055376 | 0.51 | 0.615922951 | 0.963 |
| hDV71206854 | rs1806299 | hCV8847946 | rs6921 | 0.51 | 0.615922951 | 0.6171 |
| hDV71206854 | rs1806299 | hCV8847948 | rs974759 | 0.51 | 0.615922951 | 0.6396 |

## TABLE 5: Summary of clinical data (corresponding to Example One)

| | UCSF_Wright | | VCU_Shiffman | |
|---|---|---|---|---|
| | Number | (SD)/Range/% | Number | (SD)/Range/% |
| **# of patients** | 537 | | 483 | |
| **Gender (M/F)** | | | | |
| Male | 453 | 84.4% | 279 | 57.8% |
| Female | 84 | 15.6% | 204 | 42.2% |
| **Age** | | | | |
| Mean | 52.6 | (7.3) | 50.8 | (8.3) |
| Median | 53 | 22 -- 81 | 50 | 19 -- 74 |
| **Age at Bx** | | | | |
| Mean | 50.35 | (7.51) | 47.7 | (8.3) |
| **Fibrosis score*** | | | | |
| 0 | 139 | 25.9% | 84 | 17.4% |
| 1 | 141 | 26.3% | 165 | 34.2% |
| 2 | 122 | 22.7% | | |
| 3 | 79 | 14.7% | 114 | 23.6% |
| 4 | 56 | 10.4% | 120 | 24.8% |
| Missing | 0 | 0.0% | 0 | 0.0% |
| Mean | 1.6 | (1.3) | 2.04 | (1.5) |
| Median | 1 | 0 -- 4 | 1 | 0 -- 4 |
| **Inflammation score** | | | | |
| Mean | 1.77 | (0.74) | 6.30 | (2.52) |
| Median | 2 | 0 -- 4 | 7 | 1--12 |
| **Ethnicity** | | | | |
| Caucasian | 367 | 68.3% | 341 | 70.6% |
| African-American | 78 | 14.5% | 127 | 26.3% |
| Others | 92 | 17.1% | 15 | 3.1% |
| Missing | 0 | 0.0% | 0 | 0.0% |
| **Age at infection** | | | | |
| Mean | 22.9 | (7.6) | 26.0 | (9.3) |
| Median | 21 | 0 -- 61 | 24.5 | 0 -- 58.5 |
| **Duration of infection** | | | | |
| Patients with duration | 395 | 73.6% | 347 | 71.8% |
| Mean | 27.6 | (7.4) | 21.9 | (8.5) |
| Median | 29 | 1 -- 39 | 21 | 0 -- 58.5 |
| **Fibrosis rate** | | | | |
| Mean | 0.068 | (0.158) | 0.087 | (0.446) |
| Median | 0.06 | 0 -- 3 | 0.085 | 0 -- 1.11 |
| **Risk factors** | | | | |
| Unknown | 138 | 25.7% | 149 | 30.8% |
| IVDU | 277 | 51.6% | 179 | 37.1% |
| BT | 63 | 11.7% | 122 | 25.3% |
| IVDU/BT | 59 | 11.0% | 33 | 6.8% |
| Other | 0 | 0.0% | 0 | 0.0% |
| **Alcohol (daily)** | | | | |
| Unknown | 0 | 0.0% | 0 | 0.0% |
| =0 | 161 | 30.0% | 74 | 15.3% |
| <20 | 56 | 10.4% | 234 | 48.4% |
| =20 to 50 | 73 | 13.6% | 97 | 20.1% |
| =50 to 80 | 52 | 9.7% | 47 | 9.7% |
| =>80 | 195 | 36.3% | 31 | 6.4% |
| Men | | | | |
| Mean | 89.9 | (110.50) | 29.5 | (33.0) |
| Median | 45 | 0 -- 600 | 18.4 | 0 -- 236.71 |
| Women | | | | |
| Mean | 58.1 | (101.94) | 16.3 | (28.3) |
| Median | 15 | 0 -- 513 | 4.9 | 0 -- 151.74 |
| CAGE >2 (3 or 4) | 220 | 41.0% | N/A | N/A |
| **Viral Genotype** | | | | |
| Unknown | 52 | 9.7% | 49 | 10.1% |
| Type 1 | 338 | 62.9% | 347 | 71.8% |
| 2 & 3 | 141 | 26.3% | 80 | 16.6% |
| Mixed and others | 6 | 1.1% | 7 | 1.4% |

* UCSF uses Batts-Ludwig scoring system, VCU uses Knodell scoring systems

**TABLE 6: Sample enrollment - inclusion/exclusion criteria (corresponding to Example One)**

Inclusion criteria:

- Adults (Age 18 ~ 75).
- HCV positive patients who have undergone a full course (at least 24 weeks) of Interferon treatment (any formulation +/- ribavirin) and for whom six month follow-up viral load data is available/potentially available.

Exclusion criteria:

- Discontinuation of IFN Rx secondary to poor tolerance of side effects
- Evidence of other chronic active viral hepatitis including positive HbsAg,
- Evidence of co-infection with HIV, e.g. Positive anti-HIV antibody.
- Evidence of other serious liver disease: e.g. Wilson's, Hemochromatosis, etc
- Other serious medical conditions: Rheumatic/renal/lung diseases, CVD, cancer

Additional information required for data analysis:

- Age
- Race
- Gender
- HCV genotype
- Viral load
- EtOH use
- IVDU
- Other medications
- Exact treatment regimen
- ALT levels
- Response to IFN treatment
- Other medical history, including serious medical illness such as kidney disease, cardiovascular disease, autoimmune disease, cancer

**TABLE 7: Nine SNPs in the TLR4 region significantly associated with cirrhosis risk (corresponding to Example One)**

| Marker | Chrom | Position | SNP Type | Public Gene | Allele | CT AF | CS AF | pExact/MHp | OR/MHOR | OR > 1 | Overall Risk Allele Freq |
|---|---|---|---|---|---|---|---|---|---|---|---|
| hCV31783925 | 9 | 119456993 | Intron | JORLAW | G | 12.1% | 5.1% | 0.000437 | 0.39 | 2.54 | 92.3% |
| hCV29816566 | 9 | 119469292 | Intergenic/Unknown | N/A | A | 12.2% | 5.2% | 0.000432 | 0.39 | 2.55 | 92.2% |
| hCV8788444 | 9 | 119485355 | Intergenic/Unknown | N/A | T | 11.5% | 4.8% | 0.000494 | 0.39 | 2.60 | 92.7% |
| hCV31783982 | 9 | 119493792 | Intergenic/Unknown | N/A | A | 11.5% | 4.4% | 0.000143 | 0.35 | 2.83 | 93.0% |
| hCV31783985 | 9 | 119496316 | Intergenic/Unknown | N/A | A | 11.5% | 4.6% | 0.000266 | 0.37 | 2.71 | 92.9% |
| hCV29292005 | 9 | 119509371 | Intron | TLR4 | G | 9.9% | 3.8% | 0.000535 | 0.36 | 2.76 | 93.9% |
| hCV11722141 | 9 | 119509875 | Intron | TLR4 | A | 18.2% | 27.2% | 0.00332 | 1.68 | 1.68 | 23.8% |
| hCV31784008 | 9 | 119510193 | Intron | TLR4 | C | 9.9% | 3.8% | 0.000521 | 0.36 | 2.77 | 93.9% |
| hCV11722238 | 9 | 119515123 | Missense | TLR4 | G | 9.6% | 3.5% | 0.000349 | 0.34 | 2.96 | 94.3% |

**TABLE 8: Haplotype Windows in the TLR4 region significantly associated with cirrhosis risk (corresponding to Example One)**

| Marker | Public Gene | Chrom | Position | Relative Position | SNP Type | Hap Window (n=3) | | Global Pvalue |
|---|---|---|---|---|---|---|---|---|
| hCV26954831 | TLR4 | 9 | 119514103 | -1.3 | Intron | | | |
| hCV11722236 | TLR4 | 9 | 119515123 | -0.3 | Missense | | 1 | 9.20E-05 |
| hCV11722237 | TLR4 | 9 | 119515423 | 0 | Missense | | 2 | 1.37E-05 |
| hDV71564063 | TLR4 | 9 | 119517952 | 2.5 | UTR3 | | 3 | 1.17E-04 |
| hCV29292008 | TLR4 | 9 | 119518757 | 3.3 | UTR3 | | | |
| | | | | | | | | |
| Marker | Public Gene | Chrom | Position | Relative Position | SNP Type | Hap Window (n=5) | | Global Pvalue |
| hCV26954831 | TLR4 | 9 | 119514103 | -1.3 | Intron | | | |
| hCV11722236 | TLR4 | 9 | 119515123 | -0.3 | Missense | | | |
| hCV11722237 | TLR4 | 9 | 119515423 | 0 | Missense | | | 5.37E-06 |
| hDV71564063 | TLR4 | 9 | 119517952 | 2.5 | UTR3 | | | |
| hCV29292008 | TLR4 | 9 | 119518757 | 3.3 | UTR3 | | | |

153

**TABLE 9: 3-SNP Haplotypes in the TLR4 region (corresponding to Example One)**

| hCV11722238 | hCV11722237 | hDV71564063 | Stratum | Haplotype | ALL_freq | control_freq | case_freq | Hap.Score | Hap P-value | Global P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| A | C | G | CAUC | ACG | 78.0% | 82.5% | 70.4% | -4.234226 | 2.29E-05 | 1.37E-05 |
| A | C | C | CAUC | ACC | 16.1% | 14.1% | 19.4% | 2.122911 | 0.0338 | 1.37E-05 |
| G | T | G | CAUC | GTG | 5.5% | 3.4% | 9.2% | 3.54956 | 0.0004 | 1.37E-05 |
| | | | | | | | | | | |
| r^2 with original SNP | | | | | | | | | | |
| hCV11722238 | 0.934 | | | | | | | | | |
| hDV71564063 | 0.011 | | | | | | | | | |

10

**TABLE 10: 5-SNP Haplotypes in TLR4 region (corresponding to Example One)**

| hCV26954831 | hCV11722238 | hCV11722237 | hDV71564063 | hCV29292008 | Stratum | Haplotype | ALL_freq | control_freq | case_freq | Hap.Score | Hap P-value | Global P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G | A | C | G | G | CAUC | GACGG | 36.4% | 40.4% | 29.5% | -3.2406138 | 0.0012 | 5.37E-06 |
| G | A | C | C | G | CAUC | GACCG | 15.9% | 14.1% | 18.9% | 2.0140898 | 0.0440 | 5.37E-06 |
| A | A | C | G | G | CAUC | AACGG | 29.1% | 28.1% | 30.8% | 0.8651657 | 0.3869 | 5.37E-06 |
| G | G | T | G | G | CAUC | GGTGG | 5.6% | 3.4% | 9.2% | 3.5724179 | 0.0004 | 5.37E-06 |
| G | A | C | G | C | CAUC | GACGC | 12.5% | 13.9% | 10.0% | -1.6275129 | 0.1036 | 5.37E-06 |
| | | | | | | | | | | | | |
| r^2 with original SNP | | | | | | | | | | | | |
| hCV26954831 | 0.026 | | | | | | | | | | | |
| hCV11722238 | 0.934 | | | | | | | | | | | |
| hDV71564063 | 0.011 | | | | | | | | | | | |
| hCV29292008 | 0.007 | | | | | | | | | | | |

**TABLE 11: Causal SNP Analysis data (corresponding to Example One)**

| SNP | Gene Name | Position | SNP Type | LD Block | Distance from Original SNP | CAUC freq | R^2 with Original SNP |
|---|---|---|---|---|---|---|---|
| hCV31783925 | JORLAW | 119456993 | Intron | 1 | -58.4 | 7.7% | 0.507 |
| hCV29816566 | N/A | 119469292 | Intergenic/Unknown | 1 | -46.1 | 7.8% | 0.506 |
| hCV8788444 | N/A | 119485355 | Intergenic/Unknown | 2 | -30.1 | 7.3% | 0.718 |
| hCV31783982 | N/A | 119493792 | Intergenic/Unknown | 2 | -21.6 | 7.0% | 0.745 |
| hCV31783985 | N/A | 119496316 | Intergenic/Unknown | 2 | -19.1 | 7.1% | 0.731 |
| hCV29292005 | TLR4 | 119509371 | Intron | 3 | -6.1 | 6.1% | 0.874 |
| hCV11722141 | TLR4 | 119509875 | Intron | 4 | -5.5 | 23.8% | 0.019 |
| hCV31784008 | TLR4 | 119510193 | Intron | 3 | -5.2 | 6.1% | 0.874 |
| hCV11722238 | TLR4 | 119515123 | Missense | 3 | -0.3 | 5.7% | 0.934 |
| hCV11722237 | TLR4 | 119515423 | Missense | 3 | 0.0 | 5.7% | |

**TABLE 12: Clinical Characteristics of Subjects (corresponding to Example Two)**

| | **All subjects** (Caucasians only) (N = 420) | **Controls** No Fibrosis (N =157) | **Cases** Bridging Fib./Cirrhosis (N = 263) | ***P*** **Value** |
|---|---|---|---|---|
| **Age_Bx** | | | | |
| mean +/- SD | 48.6 +/- 8.2 | 47.25 +/- 9.1 | 49.4 +/- 7.4 | 0.011 |
| **Sex** | | | | |
| % Male | 70.0% | 61.2% | 75.3% | 0.002 |
| **Daily alcohol** | | | | |
| % > 50 g/day | 30.2% | 28.7% | 31.2% | NS[b] |
| mean +/- SD | 47.8 +/- 70.6 | 50.5 +/- 76.2 | 46.2 +/- 67.2 | NS[b] |
| **Duration of infection** | | | | |
| mean +/- SD | 24.1 +/- 7.9 | 23.5 +/- 8.0 | 24.4 +/- 7.9 | NS[b] |

5 "NS" = non significant

**TABLE 13: Significant SNPs in TLR4 Region (P<=0.01) (corresponding to Example Two)**

| SNP# [b] | SNP ID | Gene | SNP Type | Dis [c] | P value | Freq | R^2 [d] | Causal SNPs [e] |
|---|---|---|---|---|---|---|---|---|
| 19 | rs12375686 | JORLAW | Intron | -58.4 | 0.0006 | 92.3% | 0.507 | ● |
| 29 | rs10448253 | N/A | Intergenic/ | -46.1 | 0.0006 | 92.2% | 0.506 | |
| 40 | rs1252039 | N/A | Intergenic/ | -30.1 | 0.0007 | 92.7% | 0.718 | |
| 46 | rs10818069 | N/A | Intergenic/ | -21.6 | 0.0002 | 93.0% | 0.745 | |
| 47 | rs10818070 | N/A | Intergenic/ | -19.1 | 0.0005 | 92.9% | 0.731 | |
| 52 | rs7864330 | TLR4 | Intron | -6.1 | 0.0010 | 93.9% | 0.874 | |
| 53 | rs1927911 | TLR4 | Intron | -5.5 | 0.0123 | 23.8% | 0.019 | ● |
| 54 | rs10759933 | TLR4 | Intron | -5.2 | 0.0010 | 93.9% | 0.874 | |
| 59 | D299G | TLR4 | Missense | -0.3 | 0.0006 | 94.3% | 0.934 | |
| * 60 | T399I | TLR4 | Missense | 0.0 | 0.0004 | 94.3% | | ● |

**TABLE 14: Fine mapping coverage for CRS7 (corresponding to Example Three)**

| SNP Predictor* | RS Number | Gene | Chr | Position (bp) | Targeted Region (kbp)** | # Tags Required*** | # HapMap SNPs Typed**** | % Coverage of Tags | # Non-HapMap SNPs Typed | # Total SNPs Typed |
|---|---|---|---|---|---|---|---|---|---|---|
| SNP1 | hCV25635059 | AZIN1 | 8 | 103,910,885 | 159 | 26 | 55 | 92 | 16 | 71 |
| SNP2 | rs4986791 | TLR4 | 9 | 119,515,423 | 76 | 34 | 52 | 88 | 10 | 62 |
| SNP3 | rs886277 | TRPM5 | 11 | 2,396,343 | 97 | 22 | 22 | 64 | 28 | 50 |
| SNP4 | rs2290351 | C15orf38 | 15 | 88,175,785 | 101 | 21 | 32 | 86 | 14 | 46 |
| SNP5 | rs4290029 | | 1 | 222,467,263 | 181 | 29 | 41 | 69 | 8 | 49 |
| SNP6 | rs17740066 | STXBP5L | 3 | 122,582,973 | 163 | 27 | 46 | 74 | 12 | 58 |
| SNP7 | rs2878771 | AQP2 | 12 | 48,638,660 | 20 | 14 | 9 | 79 | 14 | 23 |

*based on Huang et al. Hepatology 46, 297, 2007

**region identified based on the linkage disequilibrium structure in the HapMap dataset. For the SNP6 region, marker-marker LD extends ~660kbp covering STXBP5L and POLQ genes; we only tested markers in part of the large block that includes POLQ gene for the purpose of determining whether STXBP5L or POLQ is more likely to be the causal gene.

***determined by the HapMap tagger with $r^2$>0.8 and minor allele frequency≥0.05

****including tagging and putative functional SNPs and other SNPs such as those in high linkage disequilibrium with the original marker.

**TABLE 15: Markers significantly associated with liver fibrosis risk in high linkage disequilibrium with the *TLR4* SNP rs4986791 or independently significant (corresponding to Example Three)**

| | | | | | | | Regression P-value of | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | RS Number | SNP Type* | Case Allele Frequency | Control Allele Frequency | OR (95% CI)** | Allelic P-Value | marker adjusted for rs4986791 | rs4986791 adjusted for marker | marker adjusted for rs960312 | rs960312 adjusted for marker | r2 with rs4986791 | r2 with rs960312 |
| 1 | rs4986791 | missense (T399I) | 0.034 | 0.096 | 0.33 (0.18 - 0.61) | 0.00019 | N/A | N/A | 0.00096 | 0.027 | N/A | 0.01 |
| | rs4986790 | missense (D299G) | 0.034 | 0.096 | 0.34 (0.19 - 0.62) | 0.00024 | 0.89 | 0.30 | 0.0014 | 0.025 | 0.93 | 0.01 |
| | rs10818069 | intergenic | 0.044 | 0.115 | 0.35 (0.21 - 0.61) | 0.00010 | 0.30 | 0.33 | 0.00056 | 0.025 | 0.74 | 0.01 |
| | rs10759933 | intron | 0.038 | 0.099 | 0.36 (0.2 - 0.64) | 0.00037 | 0.73 | 0.12 | 0.0019 | 0.021 | 0.87 | 0.01 |
| | rs7864330 | intron | 0.038 | 0.099 | 0.36 (0.2 - 0.65) | 0.00039 | 0.73 | 0.12 | 0.0019 | 0.027 | 0.87 | 0.01 |
| | rs10818070 | intergenic | 0.046 | 0.115 | 0.37 (0.22 - 0.63) | 0.00017 | 0.41 | 0.23 | 0.00077 | 0.022 | 0.73 | 0.01 |
| | rs10448253 | intergenic | 0.052 | 0.122 | 0.39 (0.23 - 0.66) | 0.00024 | 0.17 | 0.11 | 0.0015 | 0.018 | 0.51 | 0.01 |
| | rs12375686 | intergenic | 0.051 | 0.121 | 0.39 (0.23 - 0.66) | 0.00026 | 0.17 | 0.11 | 0.0015 | 0.026 | 0.51 | 0.01 |
| | rs1252039 | intergenic | 0.048 | 0.115 | 0.39 (0.23 - 0.66) | 0.00029 | 0.63 | 0.13 | 0.0015 | 0.021 | 0.72 | 0.01 |
| 2 | rs960312 | intergenic | 0.148 | 0.086 | 1.85 (1.17 - 2.94) | 0.0083 | 0.027 | 0.00096 | N/A | N/A | 0.01 | N/A |
| | rs1329060 | intergenic | 0.153 | 0.093 | 1.76 (1.12 - 2.76) | 0.013 | 0.036 | 0.0013 | 0.89 | 0.59 | 0.01 | 0.94 |
| | rs4837494 | intergenic | 0.160 | 0.099 | 1.74 (1.12 - 2.7) | 0.012 | 0.042 | 0.0011 | 0.82 | 0.47 | 0.01 | 0.89 |
| | rs1927911 | intron | 0.272 | 0.182 | 1.68 (1.19 - 2.38) | 0.0029 | 0.016 | 0.0012 | 0.13 | 0.32 | 0.02 | 0.46 |
| 3 | rs11536889 | 3'UTR | 0.141 | 0.194 | 0.68 (0.47 - 0.99) | 0.043 | 0.011 | 0.00015 | 0.086 | 0.017 | 0.01 | 0.03 |

* missense, intron and 3'UTR SNPs are all in TLR4, and intergenic SNPs are in 5' upstream TLR4

** sorted by effect size within each group

**TABLE 16: Haplotypes association in the *TLR4* region with fibrosis risk (corresponding to Example Three)**

| Haplotype # | Haplotype | rs960312 | rs4986791 | rs11536889 | Case Freqency | Control Freqency | *P*-Haplotype | *P*-Global | OR |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ACG | A | C | G | 0.676 | 0.623 | 0.11 | 3.04E-05 | 1.26 |
| 2 | ACC | A | C | C | 0.141 | 0.194 | 3.15E-02 | 3.04E-05 | 0.68 |
| 3 | ATG | A | T | G | 0.034 | 0.096 | 1.63E-04 | 3.04E-05 | 0.33 |
| 4 | GCG | G | C | G | 0.148 | 0.086 | 7.20E-03 | 3.04E-05 | 1.85 |

**TABLE 17: SNPs in the *STXBP5L* locus that are independently associated with fibrosis risk (corresponding to Example Three)**

| RS number | SNP type* | Case Allele Frequency | Control Allele Frequency | OR (95% CI) | P-value | Regression P-value of | | | | r2 with rs17740066 | r2 with rs13086038 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | marker adjusted for rs17740066 | rs17740066 adjusted for marker | marker adjusted for rs13086038 | rs13086038 adjusted for marker | | |
| rs17740066 | Missense (V855I) | 0.120 | 0.045 | 2.92 (1.61 - 5.30) | 0.00026 | N/A | N/A | 0.0010 | 0.017 | N/A | 0.004 |
| rs13086038 | intron | 0.029 | 0.067 | 0.41 (0.21 - 0.81) | 0.0085 | 0.017 | 0.0010 | N/A | N/A | 0.004 | N/A |
| rs2169302 | Intergenic | 0.107 | 0.186 | 0.52 (0.35 - 0.78) | 0.0013 | 0.011 | 0.0031 | 0.00029 | 0.0018 | 0.016 | 0.007 |

*Both the intron and missense SNPs are in STXBP5L, and the intergenic SNP is in 3' downstream of STXBP5L

** in nearly perfect linkage disequilibrium with rs35827958 ($r^2$=0.97)

**TABLE 18: Haplotype association in the *STXBP5L* region with fibrosis risk (corresponding to Example Three)**

| Haplotype # | Haplotype | rs13086038 | rs17740066 | rs2169302 | Case freqency | Control freqency | *P*-Haplotype | *P*-Global | OR |
|---|---|---|---|---|---|---|---|---|---|
| 1 | TGT | T | G | T | 0.746 | 0.702 | 0.18 | 4.49E-06 | 1.24 |
| 2 | TGC | T | G | C | 0.107 | 0.186 | 1.62E-03 | 4.49E-06 | 0.52 |
| 3 | CGT | C | G | T | 0.027 | 0.067 | 8.93E-03 | 4.49E-06 | 0.39 |
| 4 | TAT | T | A | T | 0.118 | 0.045 | 4.93E-04 | 4.49E-06 | 2.88 |

**TABLE 19: Markers that were significantly associated with liver fibrosis risk (corresponding to Example Three)**

| SNP Predictor | RS Number | Gene | Chr | Location | Case Allele Frequency | Control Allele Frequency | Allelic P-Value | OR (95% CI) |
|---|---|---|---|---|---|---|---|---|
| SNP1 | rs17775450 | | 8 | 103,897,346 | 0.086 | 0.173 | 1.61E-04 | 0.45 (0.29 - 0.69) |
| SNP1 | rs17775510 | | 8 | 103,897,938 | 0.086 | 0.172 | 1.71E-04 | 0.45 (0.3 - 0.69) |
| SNP1 | rs12546520 | | 8 | 103,901,432 | 0.128 | 0.223 | 3.46E-04 | 0.51 (0.35 - 0.74) |
| SNP1 | rs10808382 | | 8 | 103,903,768 | 0.406 | 0.494 | 1.39E-02 | 0.7 (0.53 - 0.93) |
| SNP1 | rs2117313 | | 8 | 103,904,230 | 0.403 | 0.490 | 1.31E-02 | 0.7 (0.53 - 0.93) |
| SNP1 | rs972142 | | 8 | 103,905,429 | 0.395 | 0.484 | 1.17E-02 | 0.7 (0.52 - 0.92) |
| SNP1 | rs974759 | | 8 | 103,907,088 | 0.397 | 0.490 | 8.40E-03 | 0.69 (0.52 - 0.91) |
| SNP1 | rs6921 | AZIN1 | 8 | 103,908,052 | 0.405 | 0.494 | 1.22E-02 | 0.7 (0.53 - 0.93) |
| SNP1 | rs13123 | AZIN1 | 8 | 103,908,179 | 0.087 | 0.170 | 3.52E-04 | 0.47 (0.31 - 0.71) |
| SNP1 | hCV25635059 | AZIN1 | 8 | 103,910,885 | 0.036 | 0.099 | 2.07E-04 | 0.34 (0.19 - 0.62) |
| SNP1 | rs1434235 | AZIN1 | 8 | 103,912,208 | 0.394 | 0.484 | 1.03E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | rs34655485 | AZIN1 | 8 | 103,913,878 | 0.395 | 0.484 | 1.17E-02 | 0.7 (0.52 - 0.92) |
| SNP1 | rs2513935 | AZIN1 | 8 | 103,914,024 | 0.397 | 0.487 | 1.06E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | rs2304349 | AZIN1 | 8 | 103,915,877 | 0.087 | 0.172 | 2.52E-04 | 0.46 (0.3 - 0.7) |
| SNP1 | rs2916558 | AZIN1 | 8 | 103,916,575 | 0.397 | 0.328 | 4.43E-02 | 1.35 (1.01 - 1.81) |
| SNP1 | rs12546636 | AZIN1 | 8 | 103,917,129 | 0.051 | 0.124 | 1.46E-04 | 0.38 (0.23 - 0.64) |
| SNP1 | rs2256440 | AZIN1 | 8 | 103,917,439 | 0.395 | 0.487 | 9.30E-03 | 0.69 (0.52 - 0.91) |
| SNP1 | rs2436844 | AZIN1 | 8 | 103,919,268 | 0.393 | 0.484 | 1.02E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | rs2304344 | AZIN1 | 8 | 103,924,867 | 0.130 | 0.223 | 4.32E-04 | 0.52 (0.36 - 0.75) |
| SNP1 | rs12546634 | AZIN1 | 8 | 103,926,301 | 0.087 | 0.175 | 1.56E-04 | 0.45 (0.3 - 0.69) |
| SNP1 | rs2679748 | AZIN1 | 8 | 103,927,484 | 0.401 | 0.494 | 8.91E-03 | 0.69 (0.52 - 0.91) |
| SNP1 | rs1991927 | AZIN1 | 8 | 103,927,924 | 0.405 | 0.494 | 1.22E-02 | 0.7 (0.53 - 0.93) |
| SNP1 | rs2570942 | AZIN1 | 8 | 103,930,216 | 0.394 | 0.484 | 1.03E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | rs1019975 | AZIN1 | 8 | 103,933,586 | 0.087 | 0.172 | 2.52E-04 | 0.46 (0.3 - 0.7) |
| SNP1 | rs2164061 | AZIN1 | 8 | 103,933,649 | 0.393 | 0.484 | 1.02E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | rs2679754 | AZIN1 | 8 | 103,937,176 | 0.393 | 0.484 | 9.99E-03 | 0.69 (0.52 - 0.92) |
| SNP1 | rs1138 | AZIN1 | 8 | 103,937,662 | 0.405 | 0.500 | 7.25E-03 | 0.68 (0.51 - 0.9) |
| SNP1 | rs2679757 | AZIN1 | 8 | 103,939,994 | 0.394 | 0.484 | 1.05E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | rs2679758 | AZIN1 | 8 | 103,940,382 | 0.403 | 0.494 | 1.02E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | rs892484 | AZIN1 | 8 | 103,943,737 | 0.395 | 0.484 | 1.20E-02 | 0.7 (0.53 - 0.92) |
| SNP1 | rs892486 | AZIN1 | 8 | 103,943,805 | 0.405 | 0.490 | 1.59E-02 | 0.71 (0.53 - 0.94) |
| SNP1 | rs2679742 | | 8 | 103,948,973 | 0.333 | 0.424 | 8.19E-03 | 0.68 (0.51 - 0.91) |
| SNP1 | rs2570950 | | 8 | 103,953,041 | 0.327 | 0.424 | 4.86E-03 | 0.66 (0.5 - 0.88) |
| SNP1 | rs17197341 | | 8 | 103,958,888 | 0.087 | 0.175 | 1.56E-04 | 0.45 (0.3 - 0.69) |
| SNP1 | rs12546479 | | 8 | 103,966,297 | 0.052 | 0.127 | 8.88E-05 | 0.37 (0.22 - 0.62) |
| SNP1 | rs1806299 | | 8 | 103,967,707 | 0.334 | 0.433 | 4.12E-03 | 0.66 (0.49 - 0.88) |
| SNP1 | rs12550185 | | 8 | 103,973,095 | 0.051 | 0.127 | 8.27E-05 | 0.37 (0.22 - 0.62) |
| SNP1 | rs17197736 | | 8 | 103,979,898 | 0.051 | 0.127 | 8.27E-05 | 0.37 (0.22 - 0.62) |
| SNP1 | rs12545210 | | 8 | 103,994,953 | 0.127 | 0.226 | 1.87E-04 | 0.5 (0.35 - 0.72) |
| SNP2 | rs12375686 | | 9 | 119,456,993 | 0.051 | 0.121 | 2.55E-04 | 0.39 (0.23 - 0.66) |
| SNP2 | rs1410851 | | 9 | 119,464,417 | 0.183 | 0.124 | 2.60E-02 | 1.57 (1.05 - 2.35) |

| SNP Predictor | RS Number | Gene | Chr | Location | Case Allele Frequency | Control Allele Frequency | Allelic P-Value | OR (95% CI) |
|---|---|---|---|---|---|---|---|---|
| SNP2 | rs7037542 | | 9 | 119,466,039 | 0.188 | 0.128 | 2.40E-02 | 1.58 (1.06 - 2.35) |
| SNP2 | rs10448253 | | 9 | 119,469,292 | 0.052 | 0.122 | 2.43E-04 | 0.39 (0.23 - 0.66) |
| SNP2 | rs2039124 | | 9 | 119,473,338 | 0.174 | 0.122 | 4.23E-02 | 1.52 (1.01 - 2.29) |
| SNP2 | rs4837494 | | 9 | 119,475,962 | 0.160 | 0.099 | 1.22E-02 | 1.74 (1.12 - 2.7) |
| SNP2 | rs1329060 | | 9 | 119,478,298 | 0.153 | 0.093 | 1.31E-02 | 1.76 (1.12 - 2.76) |
| SNP2 | rs960312 | | 9 | 119,483,600 | 0.148 | 0.086 | 8.25E-03 | 1.85 (1.17 - 2.94) |
| SNP2 | rs1252039 | | 9 | 119,485,355 | 0.048 | 0.115 | 2.87E-04 | 0.39 (0.23 - 0.66) |
| SNP2 | rs10818069 | | 9 | 119,493,792 | 0.044 | 0.115 | 9.96E-05 | 0.35 (0.21 - 0.61) |
| SNP2 | rs10818070 | | 9 | 119,496,316 | 0.046 | 0.115 | 1.71E-04 | 0.37 (0.22 - 0.63) |
| SNP2 | rs7864330 | TLR4 | 9 | 119,509,371 | 0.038 | 0.099 | 3.86E-04 | 0.36 (0.2 - 0.65) |
| SNP2 | rs1927911 | TLR4 | 9 | 119,509,875 | 0.272 | 0.182 | 2.94E-03 | 1.68 (1.19 - 2.38) |
| SNP2 | rs10759933 | TLR4 | 9 | 119,510,193 | 0.038 | 0.099 | 3.65E-04 | 0.36 (0.2 - 0.64) |
| SNP2 | rs4986790 | TLR4 | 9 | 119,515,123 | 0.034 | 0.096 | 2.38E-04 | 0.34 (0.19 - 0.62) |
| SNP2 | rs4986791 | TLR4 | 9 | 119,515,423 | 0.034 | 0.096 | 1.92E-04 | 0.33 (0.18 - 0.61) |
| SNP2 | rs11536889 | TLR4 | 9 | 119,517,952 | 0.141 | 0.194 | 4.32E-02 | 0.68 (0.47 - 0.99) |
| SNP3 | rs2301698 | TRPM5 | 11 | 2,394,001 | 0.517 | 0.424 | 8.66E-03 | 1.46 (1.1 - 1.93) |
| SNP3 | rs11601537 | TRPM5 | 11 | 2,394,078 | 0.196 | 0.258 | 3.51E-02 | 0.7 (0.5 - 0.98) |
| SNP3 | rs2074236 | TRPM5 | 11 | 2,396,197 | 0.412 | 0.287 | 2.57E-04 | 1.75 (1.29 - 2.36) |
| SNP3 | rs886277 | TRPM5 | 11 | 2,396,343 | 0.416 | 0.287 | 1.62E-04 | 1.78 (1.32 - 2.4) |
| SNP3 | rs753138 | TRPM5 | 11 | 2,396,852 | 0.416 | 0.287 | 1.62E-04 | 1.78 (1.32 - 2.4) |
| SNP3 | rs11607369 | TRPM5 | 11 | 2,397,136 | 0.243 | 0.318 | 1.78E-02 | 0.69 (0.5 - 0.94) |
| SNP3 | rs4910756 | HBE1 | 11 | 5,329,432 | 0.179 | 0.240 | 3.15E-02 | 0.69 (0.49 - 0.97) |
| SNP3 | rs7929412 | HBE1 | 11 | 5,381,006 | 0.160 | 0.217 | 3.83E-02 | 0.69 (0.48 - 0.98) |
| SNP3 | rs11037445 | HBE1 | 11 | 5,418,831 | 0.184 | 0.248 | 2.71E-02 | 0.68 (0.49 - 0.96) |
| SNP4 | rs12148357 | | 15 | 88,166,426 | 0.487 | 0.385 | 4.08E-03 | 1.52 (1.14 - 2.02) |
| SNP4 | rs11638418 | | 15 | 88,170,179 | 0.449 | 0.545 | 7.12E-03 | 0.68 (0.51 - 0.9) |
| SNP4 | rs7173483 | | 15 | 88,172,552 | 0.249 | 0.169 | 6.58E-03 | 1.63 (1.14 - 2.33) |
| SNP4 | rs34549499 | | 15 | 88,173,627 | 0.447 | 0.548 | 4.60E-03 | 0.67 (0.5 - 0.88) |
| SNP4 | rs11639214 | | 15 | 88,174,111 | 0.454 | 0.551 | 6.69E-03 | 0.68 (0.51 - 0.9) |
| SNP4 | rs2290351 | AP3S2 | 15 | 88,175,785 | 0.253 | 0.166 | 3.16E-03 | 1.71 (1.19 - 2.44) |
| SNP4 | rs4932145 | AP3S2 | 15 | 88,177,463 | 0.257 | 0.166 | 2.15E-03 | 1.74 (1.22 - 2.48) |
| SNP4 | rs12197 | AP3S2 | 15 | 88,177,804 | 0.451 | 0.545 | 8.21E-03 | 0.68 (0.52 - 0.91) |
| SNP4 | rs16943647 | AP3S2 | 15 | 88,186,315 | 0.249 | 0.169 | 6.51E-03 | 1.63 (1.14 - 2.33) |
| SNP4 | rs12591948 | AP3S2 | 15 | 88,189,847 | 0.452 | 0.548 | 7.51E-03 | 0.68 (0.52 - 0.9) |
| SNP4 | rs16943651 | AP3S2 | 15 | 88,190,467 | 0.272 | 0.185 | 4.18E-03 | 1.65 (1.17 - 2.32) |
| SNP4 | rs10852122 | AP3S2 | 15 | 88,192,925 | 0.254 | 0.172 | 5.88E-03 | 1.64 (1.15 - 2.33) |
| SNP4 | rs16943661 | AP3S2 | 15 | 88,205,127 | 0.100 | 0.041 | 2.33E-03 | 2.56 (1.37 - 4.78) |
| SNP5 | rs908804 | DEGS1 | 1 | 222,446,857 | 0.069 | 0.125 | 5.87E-03 | 0.52 (0.32 - 0.83) |
| SNP5 | rs908803 | DEGS1 | 1 | 222,447,730 | 0.071 | 0.128 | 5.35E-03 | 0.52 (0.32 - 0.83) |
| SNP5 | rs908802 | | 1 | 222,447,873 | 0.068 | 0.124 | 6.11E-03 | 0.52 (0.32 - 0.83) |
| SNP5 | rs6426060 | | 1 | 222,451,334 | 0.095 | 0.162 | 3.68E-03 | 0.54 (0.36 - 0.82) |
| SNP5 | rs4290029 | | 1 | 222,467,263 | 0.110 | 0.213 | 4.86E-05 | 0.46 (0.31 - 0.67) |

| SNP Predictor | RS Number | Gene | Chr | Location | Case Allele Frequency | Control Allele Frequency | Allelic P-Value | OR (95% CI) |
|---|---|---|---|---|---|---|---|---|
| SNP5 | rs10495222 | | 1 | 222,470,392 | 0.071 | 0.146 | 3.71E-04 | 0.44 (0.28 - 0.7) |
| SNP5 | rs12134258 | | 1 | 222,478,780 | 0.084 | 0.166 | 3.04E-04 | 0.46 (0.3 - 0.71) |
| SNP5 | rs3767732 | NVL | 1 | 222,482,132 | 0.046 | 0.106 | 7.95E-04 | 0.4 (0.23 - 0.7) |
| SNP5 | rs1401318 | NVL | 1 | 222,487,789 | 0.084 | 0.166 | 3.04E-04 | 0.46 (0.3 - 0.71) |
| SNP5 | rs10799551 | NVL | 1 | 222,493,794 | 0.099 | 0.162 | 6.58E-03 | 0.57 (0.37 - 0.86) |
| SNP5 | rs4654009 | NVL | 1 | 222,494,884 | 0.099 | 0.166 | 4.49E-03 | 0.55 (0.37 - 0.84) |
| SNP5 | rs4654010 | NVL | 1 | 222,499,843 | 0.099 | 0.166 | 4.49E-03 | 0.55 (0.37 - 0.84) |
| SNP6 | rs13086038 | STXBP5L | 3 | 122,554,186 | 0.029 | 0.067 | 8.51E-03 | 0.41 (0.21 - 0.81) |
| SNP6 | rs3845856 | STXBP5L | 3 | 122,558,322 | 0.260 | 0.162 | 1.08E-03 | 1.81 (1.26 - 2.59) |
| SNP6 | rs11716257 | STXBP5L | 3 | 122,561,157 | 0.246 | 0.162 | 4.26E-03 | 1.68 (1.18 - 2.41) |
| SNP6 | rs7617178 | STXBP5L | 3 | 122,562,741 | 0.244 | 0.159 | 3.58E-03 | 1.71 (1.19 - 2.45) |
| SNP6 | rs7649908 | STXBP5L | 3 | 122,562,939 | 0.242 | 0.160 | 4.94E-03 | 1.68 (1.17 - 2.41) |
| SNP6 | rs35827958 | STXBP5L | 3 | 122,567,628 | 0.031 | 0.067 | 1.32E-02 | 0.44 (0.23 - 0.86) |
| SNP6 | rs17740066 | STXBP5L | 3 | 122,582,973 | 0.120 | 0.045 | 2.59E-04 | 2.92 (1.6 - 5.3) |
| SNP6 | rs3898024 | STXBP5L | 3 | 122,586,525 | 0.057 | 0.096 | 3.60E-02 | 0.57 (0.34 - 0.97) |
| SNP6 | rs4676719 | STXBP5L | 3 | 122,594,516 | 0.236 | 0.156 | 5.72E-03 | 1.67 (1.16 - 2.4) |
| SNP6 | rs6766105 | STXBP5L | 3 | 122,601,791 | 0.285 | 0.354 | 3.84E-02 | 0.73 (0.54 - 0.98) |
| SNP6 | rs12495471 | STXBP5L | 3 | 122,602,635 | 0.236 | 0.156 | 5.72E-03 | 1.67 (1.16 - 2.4) |
| SNP6 | rs9878946 | STXBP5L | 3 | 122,604,648 | 0.057 | 0.096 | 3.60E-02 | 0.57 (0.34 - 0.97) |
| SNP6 | rs11914436 | STXBP5L | 3 | 122,605,913 | 0.059 | 0.096 | 4.79E-02 | 0.59 (0.35 - 1) |
| SNP6 | rs4440066 | STXBP5L | 3 | 122,607,087 | 0.057 | 0.096 | 3.74E-02 | 0.57 (0.34 - 0.97) |
| SNP6 | rs9828910 | STXBP5L | 3 | 122,607,697 | 0.057 | 0.096 | 3.60E-02 | 0.57 (0.34 - 0.97) |
| SNP6 | rs2127024 | STXBP5L | 3 | 122,610,787 | 0.239 | 0.156 | 4.37E-03 | 1.69 (1.18 - 2.44) |
| SNP6 | rs7650658 | STXBP5L | 3 | 122,614,066 | 0.236 | 0.154 | 4.55E-03 | 1.7 (1.17 - 2.45) |
| SNP6 | rs9849118 | STXBP5L | 3 | 122,614,958 | 0.236 | 0.159 | 8.12E-03 | 1.63 (1.13 - 2.34) |
| SNP6 | rs7609781 | STXBP5L | 3 | 122,615,507 | 0.236 | 0.159 | 8.12E-03 | 1.63 (1.13 - 2.34) |
| SNP6 | rs7634611 | STXBP5L | 3 | 122,616,215 | 0.236 | 0.159 | 8.12E-03 | 1.63 (1.13 - 2.34) |
| SNP6 | rs3732404 | STXBP5L | 3 | 122,617,681 | 0.236 | 0.157 | 6.65E-03 | 1.65 (1.15 - 2.39) |
| SNP6 | rs6782025 | STXBP5L | 3 | 122,620,666 | 0.240 | 0.159 | 5.66E-03 | 1.66 (1.16 - 2.39) |
| SNP6 | rs6782033 | STXBP5L | 3 | 122,620,705 | 0.240 | 0.159 | 5.66E-03 | 1.66 (1.16 - 2.39) |
| SNP6 | rs2169302 | | 3 | 122,628,751 | 0.107 | 0.186 | 1.28E-03 | 0.52 (0.35 - 0.78) |
| SNP6 | rs9862879 | | 3 | 122,631,699 | 0.111 | 0.170 | 1.47E-02 | 0.61 (0.41 - 0.91) |
| SNP6 | rs3821367 | POLQ | 3 | 122,685,766 | 0.279 | 0.350 | 3.11E-02 | 0.72 (0.53 - 0.97) |
| SNP6 | rs693403 | POLQ | 3 | 122,710,082 | 0.235 | 0.158 | 8.18E-03 | 1.63 (1.13 - 2.36) |
| SNP7 | hCV25597248 | AQP2 | 12 | 48,630,919 | 0.013 | 0.000 | 4.01E-02 | |
| SNP7 | rs467199 | AQP2 | 12 | 48,631,532 | 0.160 | 0.239 | 4.60E-03 | 0.61 (0.43 - 0.86) |
| SNP7 | rs34119994 | AQP2 | 12 | 48,631,978 | 0.158 | 0.239 | 3.63E-03 | 0.6 (0.42 - 0.85) |
| SNP7 | hCV2945565 | AQP2 | 12 | 48,633,955 | 0.158 | 0.242 | 2.57E-03 | 0.59 (0.41 - 0.83) |
| SNP7 | rs426496 | AQP2 | 12 | 48,634,345 | 0.158 | 0.240 | 3.41E-03 | 0.59 (0.42 - 0.84) |
| SNP7 | rs439779 | AQP2 | 12 | 48,635,196 | 0.158 | 0.239 | 3.63E-03 | 0.6 (0.42 - 0.85) |
| SNP7 | rs467323 | AQP2 | 12 | 48,636,032 | 0.254 | 0.324 | 2.95E-02 | 0.71 (0.52 - 0.97) |
| SNP7 | rs2878771 | AQP2 | 12 | 48,638,660 | 0.143 | 0.232 | 9.37E-04 | 0.55 (0.38 - 0.79) |

**TABLE 20: SNPs associated with liver fibrosis risk (corresponding to Example Three).**

| SNP Predictor | Marker | RS Number | Gene | Chr | Location | Minor Allele Count CASE | Minor Allele Count CONTROL | Major Allele Count CASE | Major Allele Count CONTROL | Allelic P-Value | OR (95% CI) | Independent markers |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP1 | hDV71005086 | rs17775450 | | 8 | 103897346 | 45 | 54 | 479 | 258 | 1.61E-04 | 0.45 (0.29 - 0.69) | |
| SNP1 | hDV71005095 | rs17775510 | | 8 | 103897938 | 45 | 54 | 481 | 260 | 1.71E-04 | 0.45 (0.3 - 0.69) | |
| SNP1 | hCV1113790 | rs12546520 | | 8 | 103901432 | 67 | 69 | 457 | 241 | 3.46E-04 | 0.51 (0.35 - 0.74) | |
| SNP1 | hDV71161271 | rs10808382 | | 8 | 103903768 | 213 | 155 | 311 | 159 | 1.39E-02 | 0.7 (0.53 - 0.93) | |
| SNP1 | hCV15819007 | rs2117313 | | 8 | 103904230 | 211 | 154 | 313 | 160 | 1.31E-02 | 0.7 (0.53 - 0.93) | |
| SNP1 | hCV1113793 | rs972142 | | 8 | 103905429 | 207 | 152 | 317 | 162 | 1.17E-02 | 0.7 (0.52 - 0.92) | |
| SNP1 | hCV8847948 | rs974759 | | 8 | 103907088 | 209 | 154 | 317 | 160 | 8.40E-03 | 0.69 (0.52 - 0.91) | |
| SNP1 | hDV71153302 | rs6921 | | 8 | 103908052 | 213 | 155 | 313 | 159 | 1.22E-02 | 0.7 (0.53 - 0.93) | |
| SNP1 | hCV8847939 | rs13123 | AZIN1 | 8 | 103908179 | 46 | 53 | 480 | 259 | 3.52E-04 | 0.47 (0.31 - 0.71) | |
| SNP1 | hCV25635059 | | | 8 | 103910885 | 19 | 31 | 507 | 283 | 2.07E-04 | 0.34 (0.19 - 0.62) | Yes |
| SNP1 | hCV1113798 | rs1434235 | AZIN1 | 8 | 103912208 | 207 | 152 | 319 | 162 | 1.03E-02 | 0.69 (0.52 - 0.92) | |
| SNP1 | hCV1113799 | rs34655485 | AZIN1 | 8 | 103913878 | 207 | 152 | 317 | 162 | 1.17E-02 | 0.7 (0.52 - 0.92) | |
| SNP1 | hCV1113800 | rs2513935 | AZIN1 | 8 | 103914024 | 208 | 153 | 316 | 161 | 1.06E-02 | 0.69 (0.52 - 0.92) | |
| SNP1 | hCV1113802 | rs2304349 | AZIN1 | 8 | 103915877 | 46 | 54 | 480 | 260 | 2.52E-04 | 0.46 (0.3 - 0.7) | |
| SNP1 | hCV15850218 | rs2916558 | AZIN1 | 8 | 103916575 | 209 | 103 | 317 | 211 | 4.43E-02 | 1.35 (1.01 - 1.81) | |
| SNP1 | hCV1113803 | rs12546636 | AZIN1 | 8 | 103917129 | 27 | 39 | 499 | 275 | 1.46E-04 | 0.38 (0.23 - 0.64) | |
| SNP1 | hCV15974378 | rs2256440 | AZIN1 | 8 | 103917439 | 208 | 153 | 318 | 161 | 9.30E-03 | 0.69 (0.52 - 0.91) | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP1 | hCV27253261 | rs2436844 | AZIN1 | 8 | 103919268 | 206 | 151 | 318 | 161 | 1.02E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | hCV15974376 | rs2304344 | AZIN1 | 8 | 103924867 | 68 | 70 | 456 | 244 | 4.32E-04 | 0.52 (0.36 - 0.75) |
| SNP1 | hCV32148849 | rs12546634 | | 8 | 103926301 | 46 | 55 | 480 | 259 | 1.56E-04 | 0.45 (0.3 - 0.69) |
| SNP1 | hCV11245301 | rs2679748 | AZIN1 | 8 | 103927484 | 211 | 155 | 315 | 159 | 8.91E-03 | 0.69 (0.52 - 0.91) |
| SNP1 | hCV11245299 | rs1991927 | AZIN1 | 8 | 103927924 | 213 | 155 | 313 | 159 | 1.22E-02 | 0.7 (0.53 - 0.93) |
| SNP1 | hCV1113711 | rs2570942 | AZIN1 | 8 | 103930216 | 207 | 152 | 319 | 162 | 1.03E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | hDV71115804 | rs1019975 | | 8 | 103933586 | 46 | 54 | 480 | 260 | 2.52E-04 | 0.46 (0.3 - 0.7) |
| SNP1 | hCV1113704 | rs2164061 | AZIN1 | 8 | 103933649 | 206 | 151 | 318 | 161 | 1.02E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | hCV1113702 | rs2679754 | AZIN1 | 8 | 103937176 | 206 | 152 | 318 | 162 | 9.99E-03 | 0.69 (0.52 - 0.92) |
| SNP1 | hCV11240023 | rs1138 | AZIN1 | 8 | 103937662 | 213 | 157 | 313 | 157 | 7.25E-03 | 0.68 (0.51 - 0.9) |
| SNP1 | hCV1113700 | rs2679757 | AZIN1 | 8 | 103939994 | 207 | 151 | 319 | 161 | 1.05E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | hCV1113699 | rs2679758 | AZIN1 | 8 | 103940382 | 211 | 155 | 313 | 159 | 1.02E-02 | 0.69 (0.52 - 0.92) |
| SNP1 | hCV1113693 | rs892484 | AZIN1 | 8 | 103943737 | 208 | 152 | 318 | 162 | 1.20E-02 | 0.7 (0.53 - 0.92) |
| SNP1 | hCV8846755 | rs892486 | AZIN1 | 8 | 103943805 | 213 | 153 | 313 | 159 | 1.59E-02 | 0.71 (0.53 - 0.94) |
| SNP1 | hCV11240063 | rs2679742 | | 8 | 103948973 | 175 | 133 | 351 | 181 | 8.19E-03 | 0.68 (0.51 - 0.91) |
| SNP1 | hCV1113678 | rs2570950 | | 8 | 103953041 | 172 | 133 | 354 | 181 | 4.86E-03 | 0.66 (0.5 - 0.88) |
| SNP1 | hDV70919856 | rs17197341 | | 8 | 103958888 | 46 | 55 | 480 | 259 | 1.56E-04 | 0.45 (0.3 - 0.69) |
| SNP1 | hCV1716155 | rs12546479 | | 8 | 103966297 | 27 | 40 | 497 | 274 | 8.88E-05 | 0.37 (0.22 - 0.62) |
| SNP1 | hDV71206854 | rs1806299 | | 8 | 103967707 | 173 | 136 | 345 | 178 | 4.12E-03 | 0.66 (0.49 - 0.88) |
| SNP1 | hCV228233 | rs12550185 | | 8 | 103973095 | 27 | 40 | 499 | 274 | 8.27E-05 | 0.37 (0.22 - 0.62) |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP1 | hDV70919911 | rs17197736 | | 8 | 103979898 | 27 | 40 | 499 | 274 | 8.27E-05 | 0.37 (0.22 - 0.62) | |
| SNP1 | hCV11245274 | rs12545210 | | 8 | 103994953 | 67 | 71 | 459 | 243 | 1.87E-04 | 0.5 (0.35 - 0.72) | |
| SNP2 | hCV2430514 | rs10983717 | | 9 | 119417293 | 37 | 36 | 489 | 276 | 2.54E-02 | 0.58 (0.36 - 0.94) | |
| SNP2 | hCV31783925 | rs12375686 | | 9 | 119456993 | 27 | 38 | 499 | 276 | 2.55E-04 | 0.39 (0.23 - 0.66) | |
| SNP2 | hCV2430569 | rs1410851 | | 9 | 119464417 | 96 | 39 | 430 | 275 | 2.60E-02 | 1.57 (1.05 - 2.35) | |
| SNP2 | hCV26954819 | rs7037542 | | 9 | 119466039 | 99 | 40 | 427 | 272 | 2.40E-02 | 1.58 (1.06 - 2.35) | |
| SNP2 | hCV29816566 | rs10448253 | | 9 | 119469292 | 27 | 38 | 497 | 274 | 2.43E-04 | 0.39 (0.23 - 0.66) | |
| SNP2 | hCV11722160 | rs2039124 | | 9 | 119473338 | 91 | 38 | 431 | 274 | 4.23E-02 | 1.52 (1.01 - 2.29) | |
| SNP2 | hCV31783950 | rs4837494 | | 9 | 119475962 | 84 | 31 | 440 | 283 | 1.22E-02 | 1.74 (1.12 - 2.7) | |
| SNP2 | hCV8788422 | rs1329060 | | 9 | 119478298 | 80 | 29 | 444 | 283 | 1.31E-02 | 1.76 (1.12 - 2.76) | |
| SNP2 | hCV8788434 | rs960312 | | 9 | 119483600 | 78 | 27 | 448 | 287 | 8.25E-03 | 1.85 (1.17 - 2.94) | Yes |
| SNP2 | hCV8788444 | rs1252039 | | 9 | 119485355 | 25 | 36 | 501 | 278 | 2.87E-04 | 0.39 (0.23 - 0.66) | |
| SNP2 | hCV31783982 | rs10818069 | | 9 | 119493792 | 23 | 36 | 503 | 278 | 9.96E-05 | 0.35 (0.21 - 0.61) | |
| SNP2 | hCV31783985 | rs10818070 | | 9 | 119496316 | 24 | 36 | 502 | 278 | 1.71E-04 | 0.37 (0.22 - 0.63) | |
| SNP2 | hCV29292005 | rs7864330 | TLR4 | 9 | 119509371 | 20 | 31 | 504 | 283 | 3.86E-04 | 0.36 (0.2 - 0.65) | |
| SNP2 | hCV11722141 | rs1927911 | TLR4 | 9 | 119509875 | 143 | 57 | 383 | 257 | 2.94E-03 | 1.68 (1.19 - 2.38) | |
| SNP2 | hCV31784008 | rs10759933 | | 9 | 119510193 | 20 | 31 | 506 | 283 | 3.65E-04 | 0.36 (0.2 - 0.64) | |
| SNP2 | hCV11722238 | rs4986790 | TLR4 | 9 | 119515123 | 18 | 30 | 504 | 284 | 2.38E-04 | 0.34 (0.19 - 0.62) | |
| SNP2 | hCV11722237 | rs4986791 | TLR4 | 9 | 119515423 | 18 | 30 | 508 | 282 | 1.92E-04 | 0.33 (0.18 - 0.61) | Yes |
| SNP2 | hDV71564063 | rs11536889 | | 9 | 119517952 | 74 | 61 | 450 | 253 | 4.32E-02 | 0.68 (0.47 - 0.99) | Yes |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP2 | hCV31367919 | rs12335791 | | 9 | 119560233 | 80 | 66 | 444 | 248 | 3.36E-02 | 0.68 (0.47 - 0.97) | |
| SNP2 | hCV2703984 | rs10513311 | TLR4 | 9 | 119648541 | 233 | 163 | 291 | 151 | 3.67E-02 | 0.74 (0.56 - 0.98) | |
| SNP3 | hCV2990660 | rs2301698 | TRPM5 | 11 | 2394001 | 271 | 133 | 253 | 181 | 8.66E-03 | 1.46 (1.1 - 1.93) | |
| SNP3 | hCV31456696 | rs11601537 | | 11 | 2394078 | 103 | 81 | 423 | 233 | 3.51E-02 | 0.7 (0.5 - 0.98) | |
| SNP3 | hCV11367836 | rs2074236 | TRPM5 | 11 | 2396197 | 216 | 90 | 308 | 224 | 2.57E-04 | 1.75 (1.29 - 2.36) | |
| SNP3 | hCV11367838 | rs886277 | TRPM5 | 11 | 2396343 | 219 | 90 | 307 | 224 | 1.62E-04 | 1.78 (1.32 - 2.4) | Yes |
| SNP3 | hCV2990649 | rs753138 | TRPM5 | 11 | 2396852 | 219 | 90 | 307 | 224 | 1.62E-04 | 1.78 (1.32 - 2.4) | |
| SNP3 | hCV31456704 | rs11607369 | | 11 | 2397136 | 128 | 100 | 398 | 214 | 1.78E-02 | 0.69 (0.5 - 0.94) | |
| SNP3 | hCV27915384 | rs4910756 | HBE1 | 11 | 5329432 | 94 | 75 | 432 | 237 | 3.15E-02 | 0.69 (0.49 - 0.97) | |
| SNP3 | hCV1451716 | rs7929412 | HBE1 | 11 | 5381006 | 84 | 68 | 442 | 246 | 3.83E-02 | 0.69 (0.48 - 0.98) | |
| SNP3 | hCV25765485 | rs11037445 | HBE1 | 11 | 5418831 | 97 | 78 | 429 | 236 | 2.71E-02 | 0.68 (0.49 - 0.96) | |
| SNP4 | hCV3016893 | rs1439120 | ANPEP | 15 | 88139197 | 179 | 84 | 347 | 228 | 3.21E-02 | 1.4 (1.03 - 1.91) | |
| SNP4 | hCV31590419 | rs12148357 | | 15 | 88166426 | 256 | 120 | 270 | 192 | 4.08E-03 | 1.52 (1.14 - 2.02) | |
| SNP4 | hCV11524424 | rs11638418 | | 15 | 88170179 | 236 | 171 | 290 | 143 | 7.12E-03 | 0.68 (0.51 - 0.9) | |
| SNP4 | hCV29230371 | rs7173483 | | 15 | 88172552 | 130 | 53 | 392 | 261 | 6.58E-03 | 1.63 (1.14 - 2.33) | |
| SNP4 | hCV1381379 | rs34549499 | | 15 | 88173627 | 235 | 172 | 291 | 142 | 4.60E-03 | 0.67 (0.5 - 0.88) | |
| SNP4 | hCV31590424 | rs11639214 | | 15 | 88174111 | 238 | 173 | 286 | 141 | 6.69E-03 | 0.68 (0.51 - 0.9) | |
| SNP4 | hCV1381377 | rs2290351 | C15orf38 | 15 | 88175785 | 133 | 52 | 393 | 262 | 3.16E-03 | 1.71 (1.19 - 2.44) | Yes |
| SNP4 | hCV27998434 | rs4932145 | AP3S2 | 15 | 88177463 | 135 | 52 | 391 | 262 | 2.15E-03 | 1.74 (1.22 - 2.48) | |
| SNP4 | hCV9290199 | rs12197 | AP3S2 | 15 | 88177804 | 237 | 169 | 289 | 141 | 8.21E-03 | 0.68 (0.52 - 0.91) | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP4 | hDV70753259 | rs16943647 | AP3S2 | 15 | 88186315 | 131 | 53 | 395 | 261 | 6.51E-03 | 1.63 (1.14 - 2.33) | |
| SNP4 | hCV1381363 | rs12591948 | AP3S2 | 15 | 88189847 | 238 | 172 | 288 | 142 | 7.51E-03 | 0.68 (0.52 - 0.9) | |
| SNP4 | hDV70753261 | rs16943651 | AP3S2 | 15 | 88190467 | 143 | 58 | 383 | 256 | 4.18E-03 | 1.65 (1.17 - 2.32) | |
| SNP4 | hCV1381359 | rs10852122 | AP3S2 | 15 | 88192925 | 133 | 54 | 391 | 260 | 5.88E-03 | 1.64 (1.15 - 2.33) | |
| SNP4 | hDV70753270 | rs16943661 | AP3S2 | 15 | 88205127 | 52 | 13 | 470 | 301 | 2.33E-03 | 2.56 (1.37 - 4.78) | |
| SNP5 | hDV71101101 | rs908804 | | 1 | 222446857 | 36 | 39 | 488 | 273 | 5.87E-03 | 0.52 (0.32 - 0.83) | |
| SNP5 | hCV12023148 | rs908803 | DEGS1 | 1 | 222447730 | 37 | 40 | 487 | 272 | 5.35E-03 | 0.52 (0.32 - 0.83) | |
| SNP5 | hDV71170845 | rs908802 | | 1 | 222447873 | 36 | 39 | 490 | 275 | 6.11E-03 | 0.52 (0.32 - 0.83) | |
| SNP5 | hCV31711270 | rs6426060 | | 1 | 222451334 | 50 | 51 | 476 | 263 | 3.68E-03 | 0.54 (0.36 - 0.82) | |
| SNP5 | hCV1721645 | rs4290029 | | 1 | 222467263 | 58 | 67 | 468 | 247 | 4.86E-05 | 0.46 (0.31 - 0.67) | Yes |
| SNP5 | hCV509813 | rs10495222 | | 1 | 222470392 | 37 | 46 | 487 | 268 | 3.71E-04 | 0.44 (0.28 - 0.7) | |
| SNP5 | hCV31711313 | rs12134258 | | 1 | 222478780 | 44 | 52 | 482 | 262 | 3.04E-04 | 0.46 (0.3 - 0.71) | |
| SNP5 | hCV27481000 | rs3767732 | NVL | 1 | 222482132 | 24 | 33 | 500 | 277 | 7.95E-04 | 0.4 (0.23 - 0.7) | |
| SNP5 | hCV12021443 | rs1401318 | NVL | 1 | 222487789 | 44 | 52 | 482 | 262 | 3.04E-04 | 0.46 (0.3 - 0.71) | |
| SNP5 | hCV231892 | rs10799551 | NVL | 1 | 222493794 | 52 | 51 | 474 | 263 | 6.58E-03 | 0.57 (0.37 - 0.86) | |
| SNP5 | hCV27983683 | rs4654009 | NVL | 1 | 222494884 | 52 | 52 | 474 | 262 | 4.49E-03 | 0.55 (0.37 - 0.84) | |
| SNP5 | hCV27940202 | rs4654010 | NVL | 1 | 222499843 | 52 | 52 | 474 | 262 | 4.49E-03 | 0.55 (0.37 - 0.84) | |
| SNP6 | hCV26919853 | rs13086038 | STXBP5L | 3 | 122554186 | 15 | 21 | 507 | 293 | 8.51E-03 | 0.41 (0.21 - 0.81) | Yes |
| SNP6 | hCV27502059 | rs3845856 | STXBP5L | 3 | 122558322 | 136 | 51 | 388 | 263 | 1.08E-03 | 1.81 (1.26 - 2.59) | |
| SNP6 | hCV31746823 | rs11716257 | | 3 | 122561157 | 129 | 51 | 395 | 263 | 4.26E-03 | 1.68 (1.18 - 2.41) | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP6 | hCV30555357 | rs7617178 | | 3 | 122562741 | 128 | 50 | 396 | 264 | 3.58E-03 | 1.71 (1.19 - 2.45) | |
| SNP6 | hCV31746822 | rs7649908 | | 3 | 122562939 | 127 | 50 | 397 | 262 | 4.94E-03 | 1.68 (1.17 - 2.41) | |
| SNP6 | hCV11238766 | rs35827958 | STXBP5L | 3 | 122567628 | 16 | 21 | 508 | 293 | 1.32E-02 | 0.44 (0.23 - 0.86) | |
| SNP6 | hCV25647188 | rs17740066 | STXBP5L | 3 | 122582973 | 63 | 14 | 463 | 300 | 2.59E-04 | 2.92 (1.6 - 5.3) | Yes |
| SNP6 | hCV8247698 | rs3898024 | STXBP5L | 3 | 122586525 | 30 | 30 | 496 | 284 | 3.60E-02 | 0.57 (0.34 - 0.97) | |
| SNP6 | hCV30338809 | rs4676719 | | 3 | 122594516 | 124 | 49 | 402 | 265 | 5.72E-03 | 1.67 (1.16 - 2.4) | |
| SNP6 | hCV481173 | rs6766105 | STXBP5L | 3 | 122601791 | 150 | 111 | 376 | 203 | 3.84E-02 | 0.73 (0.54 - 0.98) | |
| SNP6 | hCV31746803 | rs12495471 | STXBP5L | 3 | 122602635 | 124 | 49 | 402 | 265 | 5.72E-03 | 1.67 (1.16 - 2.4) | |
| SNP6 | hCV29780197 | rs9878946 | | 3 | 122604648 | 30 | 30 | 496 | 284 | 3.60E-02 | 0.57 (0.34 - 0.97) | |
| SNP6 | hCV130085 | rs11914436 | STXBP5L | 3 | 122605913 | 31 | 30 | 495 | 284 | 4.79E-02 | 0.59 (0.35 - 1) | |
| SNP6 | hCV27999672 | rs4440066 | STXBP5L | 3 | 122607087 | 30 | 30 | 494 | 284 | 3.74E-02 | 0.57 (0.34 - 0.97) | |
| SNP6 | hCV3100643 | rs9828910 | STXBP5L | 3 | 122607697 | 30 | 30 | 496 | 284 | 3.60E-02 | 0.57 (0.34 - 0.97) | |
| SNP6 | hCV15826462 | rs2127024 | STXBP5L | 3 | 122610787 | 125 | 49 | 399 | 265 | 4.37E-03 | 1.69 (1.18 - 2.44) | |
| SNP6 | hCV29690012 | rs7650658 | | 3 | 122614066 | 124 | 48 | 402 | 264 | 4.55E-03 | 1.7 (1.17 - 2.45) | |
| SNP6 | hCV11238745 | rs9849118 | STXBP5L | 3 | 122614958 | 124 | 50 | 402 | 264 | 8.12E-03 | 1.63 (1.13 - 2.34) | |
| SNP6 | hCV3100650 | rs7609781 | STXBP5L | 3 | 122615507 | 124 | 50 | 402 | 264 | 8.12E-03 | 1.63 (1.13 - 2.34) | |
| SNP6 | hCV3100651 | rs7634611 | STXBP5L | 3 | 122616215 | 124 | 50 | 402 | 264 | 8.12E-03 | 1.63 (1.13 - 2.34) | |
| SNP6 | hCV25647179 | rs3732404 | STXBP5L | 3 | 122617681 | 123 | 49 | 399 | 263 | 6.65E-03 | 1.65 (1.15 - 2.39) | |
| SNP6 | hCV25647150 | rs6782025 | STXBP5L | 3 | 122620666 | 126 | 50 | 400 | 264 | 5.66E-03 | 1.66 (1.16 - 2.39) | |
| SNP6 | hCV25647151 | rs6782033 | STXBP5L | 3 | 122620705 | 126 | 50 | 400 | 264 | 5.66E-03 | 1.66 (1.16 - 2.39) | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP6 | hCV26919823 | rs2169302 | | 3 | 122628751 | 56 | 58 | 468 | 254 | 1.28E-03 | 0.52 (0.35 - 0.78) | Yes |
| SNP6 | hCV30194915 | rs9862879 | | 3 | 122631699 | 58 | 53 | 466 | 259 | 1.47E-02 | 0.61 (0.41 - 0.91) | |
| SNP6 | hDV71210383 | rs3821367 | | 3 | 122685766 | 147 | 110 | 379 | 204 | 3.11E-02 | 0.72 (0.53 - 0.97) | |
| SNP6 | hCV919213 | rs693403 | POLQ | 3 | 122710082 | 123 | 49 | 401 | 261 | 8.18E-03 | 1.63 (1.13 - 2.36) | |
| SNP7 | hCV670794 | rs632151 | FAIM2 | 12 | 48575490 | 56 | 49 | 468 | 265 | 3.74E-02 | 0.65 (0.43 - 0.98) | |
| SNP7 | hCV1420722 | rs297907 | | 12 | 48603085 | 159 | 75 | 367 | 239 | 4.73E-02 | 1.38 (1 - 1.9) | |
| SNP7 | hCV670833 | rs385988 | | 12 | 48609549 | 7 | 0 | 519 | 310 | 4.14E-02 | | |
| SNP7 | hCV25597248 | | | 12 | 48630919 | 7 | 0 | 519 | 314 | 4.01E-02 | | |
| SNP7 | hCV1420652 | rs467199 | AQP2 | 12 | 48631532 | 84 | 75 | 442 | 239 | 4.60E-03 | 0.61 (0.43 - 0.86) | |
| SNP7 | hCV1420653 | rs34119994 | AQP2 | 12 | 48631978 | 83 | 75 | 443 | 239 | 3.63E-03 | 0.6 (0.42 - 0.85) | |
| SNP7 | hCV2945565 | | | 12 | 48633955 | 83 | 76 | 443 | 238 | 2.57E-03 | 0.59 (0.41 - 0.83) | |
| SNP7 | hCV1420655 | rs426496 | AQP2 | 12 | 48634345 | 83 | 75 | 441 | 237 | 3.41E-03 | 0.59 (0.42 - 0.84) | |
| SNP7 | hCV1002613 | rs439779 | AQP2 | 12 | 48635196 | 83 | 75 | 443 | 239 | 3.63E-03 | 0.6 (0.42 - 0.85) | |
| SNP7 | hCV1002616 | rs467323 | AQP2 | 12 | 48636032 | 133 | 101 | 391 | 211 | 2.95E-02 | 0.71 (0.52 - 0.97) | |
| SNP7 | hCV3187664 | rs2878771 | AQP2 | 12 | 48638660 | 75 | 73 | 451 | 241 | 9.37E-04 | 0.55 (0.38 - 0.79) | Yes |

All publications and patents cited in this specification are herein incorporated by reference in their entirety. Various modifications and variations of the described compositions, methods and systems 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 and certain working examples, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention that are obvious to those skilled in the field of molecular biology, genetics and related fields are intended to be within the scope of the following claims.

# CLAIMS

**What Is Claimed Is:**

1. A method of determining whether a human has an altered risk for developing liver fibrosis, comprising testing nucleic acid from said human for the presence or absence of a polymorphism selected from the group consisting of the polymorphisms represented by position 101 of any one of the nucleotide sequences of SEQ ID NOS:91-358 or its complement, wherein the polymorphism indicates an altered risk for developing liver fibrosis.

2. The method of claim 1, wherein said polymorphism is selected from the group consisting of the polymorphisms set forth in at least one of Tables 7-11 and 13-20.

3. The method of claim 2, wherein said polymorphism comprises a haplotype selected from the group consisting of the haplotypes set forth in at least one of Tables 8-10, 16, and 18.

4. The method of claim 1, wherein said polymorphism comprises a polymorphism selected from the group consisting of hCV11722141 (rs1927911), hCV11722237 (rs4986791/T399I), hCV11722238 (rs4986790/D299G), hCV29292005 (rs7864330), hCV29816566 (rs10448253), hCV31783925(rs12375686), hCV31783982 (rs10818069), hCV31783985(rs10818070), hCV31784008 (rs10759933), hCV8788444 (rs1252039), hDV71564063 (rs11536889), hCV26954831 (rs5030728), and hCV29292008 (rs7873784).

5. The method of claim 4, wherein said polymorphism comprises a haplotype selected from the group consisting of:

(a) haplotypes comprising polymorphisms hCV11722237 (rs4986791/T399I), hCV11722238 (rs4986790/D299G), and hDV71564063 (rs11536889); and

(b) the haplotypes of (a), further comprising polymorphisms hCV26954831 (rs5030728) and hCV29292008 (rs7873784).

6. The method of claim 1, wherein the altered risk is an increased risk.

7. The method of claim 1, wherein the altered risk is a decreased risk.

8. The method of claim 1, wherein said nucleic acid is a nucleic acid extract from a biological sample from said human.

9. The method of claim 8, wherein said biological sample is blood, saliva, or buccal cells.

10. The method of claim 8, further comprising preparing said nucleic acid extract from said biological sample prior to said testing step.

11. The method of claim 8, further comprising obtaining said biological sample from said human prior to said preparing step.

12. The method of claim 1, wherein said testing step comprises nucleic acid amplification.

13. The method of claim 12, wherein said nucleic acid amplification is carried out by polymerase chain reaction.

14. The method of claim 1, further comprising correlating the presence or absence of the polymorphism with an altered risk for developing liver fibrosis.

15. The method of claim 14, wherein said correlating step is performed by computer software.

16. The method of claim 1, wherein said testing is performed using sequencing, 5' nuclease digestion, molecular beacon assay, oligonucleotide ligation assay, size analysis, single-stranded conformation polymorphism analysis, or denaturing gradient gel electrophoresis (DGGE).

17. The method of any one of claim 1, wherein said testing is performed using an allele-specific method.

18. The method of claim 17, wherein said allele-specific method is allele-specific probe hybridization, allele-specific primer extension, or allele-specific amplification.

19. The method of claim 17, wherein the method is performed using an allele-specific primer set forth in Table 3.

20. The method of claim 1 which is an automated method.

21. A method of determining whether a human has an increased risk for progressing rapidly from minimal fibrosis to bridging fibrosis/cirrhosis, comprising testing nucleic acid from said human for the presence or absence of a polymorphism selected from the group consisting of the polymorphisms represented by position 101 of any one of the nucleotide sequences of SEQ

ID NOS:91-358 or its complement, wherein the polymorphism indicates an altered risk for progressing rapidly from minimal fibrosis to bridging fibrosis/cirrhosis.

22. The method of claim 21, further comprising correlating the presence or absence of the polymorphism with an increased risk for progressing rapidly from minimal fibrosis to bridging fibrosis/cirrhosis.

23. The method of claim 22, wherein said correlating step is performed by computer software.

24. A method of identifying a human having an increased risk for developing liver fibrosis, comprising testing a nucleic acid sample from said human for the presence or absence of a first polymorphism which is in linkage disequilibrium with a second polymorphism, wherein the second polymorphism is a polymorphism selected from the group consisting of the polymorphisms represented by position 101 of any one of the nucleotide sequences of SEQ ID NOS:91-358 or its complement, and wherein the first polymorphism identifies said human as having an increased risk for developing liver fibrosis.

25. The method of claim 24, wherein the linkage disequilibrium is $r^2 = 1$.

26. The method of claim 24, wherein the first polymorphism is selected from the group consisting of the polymorphisms set forth in Table 4.

27. The method of claim 24, further comprising correlating the presence or absence of said first polymorphism with an increased risk for developing liver fibrosis.

28. The method of claim 27, wherein said correlating step is performed by computer software.

29. A method for reducing risk of developing liver fibrosis in a human, comprising administering to said human an effective amount of a therapeutic agent, said human having been identified as having an increased risk for developing liver fibrosis due to the presence or absence of a polymorphisms selected from the group consisting of the polymorphisms represented by position 101 of any one of the nucleotide sequences of SEQ ID NOS:91-358.

30. The method of claim 29, wherein the method comprises testing nucleic acid from said human for the presence or absence of said polymorphism.

31. The method of claim 1, further comprising selecting said human for inclusion in a clinical trial of a therapeutic agent.

32. A kit for carrying out the method of claim 1, wherein the kit comprises an enzyme, a buffer, and at least one polynucleotide detection reagent, and wherein the polynucleotide detection reagent selectively hybridizes to said nucleic acid in the presence of said polymorphism and does not hybridize to said nucleic acid in the absence of said polymorphism.