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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number
WO 02/24958 A1

(51) International Patent Classification[7]: C12Q 1/68, C07H 21/04

(21) International Application Number: PCT/US01/29508

(22) International Filing Date:
19 September 2001 (19.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/233,520 19 September 2000 (19.09.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BIEGLECKI, Karyn, M.** [US/US]; 15 Waterhouse Lane, Chester, CT 06412 (US). **FINKEL, Kevin** [US/US]; 27 South Pond Circle, Cheshire, CT 06410 (US). **KAZEMI, Amir** [US/US]; 53 Southwind Drive, Wallingford, CT 06492 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US). **PARKS, Katie, E.** [US/US]; 139 Kingswood Drive, Naugatuck, CT 06770 (US). **SAUSKER, Elizabeth, Ann** [US/US]; 2122 Avalon Haven Drive, North Haven, CT 06473 (US).

(74) Agents: **FIELD, Gisela, M.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/24958 A1

(54) Title: HAPLOTYPES OF THE ADH7 GENE

(57) Abstract: Novel genetic variants of the Alcohol Dehydrogenase 7 (Class IV), Mu Or Sigma Polypeptide (ADH7) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the ADH7 gene. Compositions and methods for haplotyping and/or genotyping the ADH7 gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE ADH7 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/233,520 filed September 19, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in

clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* 4:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of cancers and Parkinson's disease (PD) is the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene or its encoded product. Human alcohol dehydrogenases consist of a family of five related classes of enzymes that function in the metabolism of alcohols and retinol. ADH7, a member of the class IV ADHs, is a retinol dehydrogenase involved in the synthesis of retinoic acid important for cellular differentiation.

ADH7 is the major type of class IV ADH expressed in the stomach (Zgombic-Knight *J Biol Chem* 1995 Mar 3;270(9):4305-11). Studies on the specific activity of recombinant ADH7 showed a high catalytic efficiency for the oxidation of all-trans retinol to all-trans-retinal, which is important for the synthesis of all-trans retinoic acid. Retinoic acid is an important hormone involved in cellular differentiation, which suggests that defects in ADH7 could play a role in some types of cancer (Satre et al., 1994; *J. Biol. Chem.* 269: 15606-15612). Also, ADH7 has a structure that allows the maximum binding of all-trans retinol to the zinc containing active site, thereby making ADH7 catalytically more efficient than other classes of ADHs (Kedishvili *J Biol Chem* 1995 Feb 24;270(8):3625-30).

The involvement of alcohol dehydrogenases in retinoid and dopamine metabolism has implicated these enzymes as potentially being involved in Parkinson's Disease (PD). This hypothesis is supported by an observed association between a cluster of alcohol dehdrogenase genes on chromosome 4 with an autosomal dominant form of PD. Buernevich et al. (*Mov Disord* 2000; 15(5):813-8) sequenced the ADH7 gene (formerly known as ADH class IV gene) in PD patients and discovered an association between a certain ADH7 polymorphism and PD, suggesting that genetic variations of ADH7 represent a risk factor for PD.

The alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide gene is located on chromosome 4q23-q24 and contains 9 exons that encode a 374 amino acid protein. A reference sequence for the ADH7 gene is shown in the contiguous lines of Figure 1 (Genaissance Reference No. 9403243; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_000673.2) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Because of the potential for variation in the ADH7 gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the ADH7 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of ADH7 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 15 novel polymorphic sites in the ADH7 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 3901 (PS1), 3996 (PS2), 4015 (PS3), 4019 (PS4), 4060 (PS5), 4148 (PS6), 9684 (PS7), 11281 (PS8), 11346 (PS9), 18624 (PS10), 18646 (PS11), 18667 (PS12), 18979 (PS13), 26036 (PS14) and 26195 (PS15). The polymorphisms at these sites are thymine or cytosine at PS1, thymine or cytosine at PS2, guanine or adenine at PS3, thymine or cytosine at PS4, thymine or cytosine at PS5, thymine or cytosine at PS6, cytosine or thymine at PS7, thymine or guanine at PS8, guanine or thymine at PS9, guanine or adenine at PS10, adenine or guanine at PS11, adenine or guanine at PS12, cytosine or thymine at PS13, adenine or thymine at PS14 and guanine or adenine at PS15. In addition, the

inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS15 in the ADH7 gene, which are shown below in Tables 5 and 4, respectively. Each of these ADH7 haplotypes constitutes a code that defines the variant nucleotides that exist in the human population at this set of polymorphic sites in the ADH7 gene. Thus each ADH7 haplotype also represents a naturally-occurring isoform (also referred to herein as an "isogene") of the ADH7 gene. The frequency of each haplotype and haplotype pair within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the ADH7 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15 in both copies of the ADH7 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel ADH7 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel ADH7 polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the ADH7 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the ADH7 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's ADH7 gene is defined by one of the ADH7 haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's ADH7 gene are defined by one of the ADH7 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. Establishing the ADH7 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with ADH7 activity, e.g., cancers and PD.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate ADH7 as a candidate target for treating a specific condition or disease predicted to be associated with ADH7 activity. Determining for a particular population the frequency of one or more of the individual ADH7 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue ADH7 as a target for treating the specific disease of interest. In particular, if

variable ADH7 activity is associated with the disease, then one or more ADH7 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed ADH7 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable ADH7 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any ADH7 haplotype or haplotype pair, apply the information derived from detecting ADH7 haplotypes in an individual to decide whether modulating ADH7 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting ADH7 to treat a specific condition or disease predicted to be associated with ADH7 activity. For example, detecting which of the ADH7 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the ADH7 isoforms present in the disease population, or for only the most frequent ADH7 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular ADH7 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

Haplotyping the ADH7 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with ADH7 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the ADH7 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute ADH7 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of an ADH7 haplotype or haplotype pair that is associated with response to the drug being studied in the trial, even if this association was previously unknown. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any ADH7 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and an ADH7 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the ADH7 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the ADH7 genotype or haplotype in a reference population. A higher frequency of the ADH7 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the ADH7 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the ADH7 haplotype is selected from the haplotypes shown in

Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for cancers and PD.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the ADH7 gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of cytosine at PS1, cytosine at PS2, adenine at PS3, cytosine at PS4, cytosine at PS5, cytosine at PS6, thymine at PS7, guanine at PS8, thymine at PS9, adenine at PS10, guanine at PS11, guanine at PS12, thymine at PS13, thymine at PS14 and adenine at PS15.

A particularly preferred polymorphic variant is an isogene of the ADH7 gene. An ADH7 isogene of the invention comprises thymine or cytosine at PS1, thymine or cytosine at PS2, guanine or adenine at PS3, thymine or cytosine at PS4, thymine or cytosine at PS5, thymine or cytosine at PS6, cytosine or thymine at PS7, thymine or guanine at PS8, guanine or thymine at PS9, guanine or adenine at PS10, adenine or guanine at PS11, adenine or guanine at PS12, cytosine or thymine at PS13, adenine or thymine at PS14 and guanine or adenine at PS15. The invention also provides a collection of ADH7 isogenes, referred to herein as an ADH7 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for an ADH7 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 355, adenine at a position corresponding to nucleotide 654, guanine at a position corresponding to nucleotide 676 and guanine at a position corresponding to nucleotide 697. A particularly preferred polymorphic cDNA variant comprises the coding sequence of an ADH7 isogene defined by haplotypes 1, 2, 4, 6, 9, 10, 14, 15 and 18.

Polynucleotides complementary to these ADH7 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the ADH7 gene will be useful in studying the expression and function of ADH7, and in expressing ADH7 protein for use in screening for candidate drugs to treat diseases related to ADH7 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic and cDNA variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express ADH7 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the ADH7 protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of cysteine at a position corresponding to amino acid position 119,

glutamic acid at a position corresponding to amino acid position 226 and valine at a position corresponding to amino acid position 233. A polymorphic variant of ADH7 is useful in studying the effect of the variation on the biological activity of ADH7 as well as on the binding affinity of candidate drugs targeting ADH7 for the treatment of cancers and PD.

The present invention also provides antibodies that recognize and bind to the above polymorphic ADH7 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one or more of the ADH7 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the ADH7 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against ADH7 protein, and for testing the efficacy of therapeutic agents and compounds for cancers and PD in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the ADH7 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes one or more of the following: the polymorphisms, the genotypes, the haplotypes, and the haplotype pairs identified for the ADH7 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing ADH7 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the ADH7 gene (Genaissance Reference No. 9403243; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:81 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS15, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:81 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the ADH7 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the

sequence.

Figure 3 illustrates a reference sequence for the ADH7 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

5

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the ADH7 gene. As described in more detail below, the inventors herein discovered 18 isogenes of the ADH7 gene by characterizing the ADH7 gene found in genomic DNAs isolated from an Index Repository that
10 contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified
15 ethnogeographic origin as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one
20 from each of North, Central and South America), one three-generation Caucasian family (from the

CEPH Utah cohort) and one two-generation African-American family.

The ADH7 isogenes present in the human reference population are defined by haplotypes for 15 polymorphic sites in the ADH7 gene, all of which are believed to be novel. The novel ADH7 polymorphic sites identified by the inventors are referred to as PS1-PS15 to designate the order in which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index Repository for PS1-PS15 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the ADH7 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the ADH7 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether ADH7 is a suitable target for drugs to treat cancers and PD, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms (e.g., alleles) of a gene found in a population. An isogene (or allele) contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, where physical features include polymorphic sites.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position on a chromosome or DNA molecule at which at least two alternative sequences are found in a population.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency

of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the ADH7 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel ADH7 polymorphisms, haplotypes and haplotype pairs identified herein.

The compositions comprise at least one oligonucleotide for detecting the variant nucleotide or nucleotide pair located at a novel ADH7 polymorphic site in one copy or two copies of the ADH7 gene. Such oligonucleotides are referred to herein as ADH7 haplotyping oligonucleotides or genotyping oligonucleotides, respectively, and collectively as ADH7 oligonucleotides. In one embodiment, an ADH7 haplotyping or genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that contains, or that is located close to, one of the novel polymorphic sites described herein.

As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives,

and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Haplotyping or genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of an ADH7 polynucleotide. Preferably, the target region is located in an ADH7 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with another region in the ADH7 polynucleotide or with a non-ADH7 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the ADH7 gene using the polymorphism information provided herein in conjunction with the known sequence information for the ADH7 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred haplotyping or genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one

allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention. ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent that the ASO contains either of the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting ADH7 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TTATAAGYTGGTCTG (SEQ ID NO:4) and its complement,
CTATCTAYGTGAAGG (SEQ ID NO:5) and its complement,
AGCTGCTRTTATATA (SEQ ID NO:6) and its complement,
GCTGTTAYATACAAC (SEQ ID NO:7) and its complement,
AAGTCTAYGTTTGCA (SEQ ID NO:8) and its complement,
ATCATGAYCTAATGA (SEQ ID NO:9) and its complement,
TACTTGCYTATTTCT (SEQ ID NO:10) and its complement,
TACATTTKTCCTATG (SEQ ID NO:11) and its complement,
TATTACTKGTCGTGG (SEQ ID NO:12) and its complement,
CATCTAGRATCATTG (SEQ ID NO:13) and its complement,
CCTCAACRAAGACAA (SEQ ID NO:14) and its complement,
GAAGGCCRTGGCTGT (SEQ ID NO:15) and its complement,
TTAAGAGYAGGTTCA (SEQ ID NO:16) and its complement,
CCAGAGGWTTTCCAA (SEQ ID NO:17) and its complement, and
AGTGGCARGAGGTCT (SEQ ID NO:18) and its complement.

A preferred ASO primer for detecting ADH7 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TTCTTTTTATAAGYT (SEQ ID NO:19); CCTTTGCAGACCARC (SEQ ID NO:20);
GCCAGACTATCTAYG (SEQ ID NO:21); CTTGTGCCTTCACRT (SEQ ID NO:22);
GGCACAAGCTGCTRT (SEQ ID NO:23); CTGTTGTATATAAYA (SEQ ID NO:24);
CAAGCTGCTGTTAYA (SEQ ID NO:25); CACTCTGTTGTATRT (SEQ ID NO:26);
CAGAAAAAGTCTAYG (SEQ ID NO:27); TATTTCTGCAAACRT (SEQ ID NO:28);
CTCCTCATCATGAYC (SEQ ID NO:29); CTCACATCATTAGRT (SEQ ID NO:30);
TTATAATACTTGCYT (SEQ ID NO:31); AACCTAAGAAATARG (SEQ ID NO:32);
TGTATTTACATTTKT (SEQ ID NO:33); AATTAGCATAGGAMA (SEQ ID NO:34);
AAACAGTATTACTKG (SEQ ID NO:35); AGTACTCCACGACMA (SEQ ID NO:36);
CTGGTGCATCTAGRA (SEQ ID NO:37); CAATCCCAATGATYC (SEQ ID NO:38);
GATTGACCTCAACRA (SEQ ID NO:39); TCAAATTTGTCTTYG (SEQ ID NO:40);
ATTTGAGAAGGCCRT (SEQ ID NO:41); GCACCTACAGCCAYG (SEQ ID NO:42);
AAACACTTAAGAGYA (SEQ ID NO:43); GTCCAGTGAACCTRC (SEQ ID NO:44);
TTGAGCCCAGAGGWT (SEQ ID NO:45); CTGTTATTGGAAAWC (SEQ ID NO:46);
ATCCAAAGTGGCARG (SEQ ID NO:47); and CAACACAGACCTCYT (SEQ ID NO:48).

Other oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting ADH7 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TTTTTATAAG (SEQ ID NO:49); TTGCAGACCA (SEQ ID NO:50);
AGACTATCTA (SEQ ID NO:51); GTGCCTTCAC (SEQ ID NO:52);
ACAAGCTGCT (SEQ ID NO:53); TTGTATATAA (SEQ ID NO:54);
GCTGCTGTTA (SEQ ID NO:55); TCTGTTGTAT (SEQ ID NO:56);
AAAAAGTCTA (SEQ ID NO:57); TTCTGCAAAC (SEQ ID NO:58);
CTCATCATGA (SEQ ID NO:59); ACATCATTAG (SEQ ID NO:60);
TAATACTTGC (SEQ ID NO:61); CTAAGAAATA (SEQ ID NO:62);
ATTTACATTT (SEQ ID NO:63); TAGCATAGGA (SEQ ID NO:64);
CAGTATTACT (SEQ ID NO:65); ACTCCACGAC (SEQ ID NO:66);
GTGCATCTAG (SEQ ID NO:67); TCCCAATGAT (SEQ ID NO:68);
TGACCTCAAC (SEQ ID NO:69); AATTTGTCTT (SEQ ID NO:70);
TGAGAAGGCC (SEQ ID NO:71); CCTACAGCCA (SEQ ID NO:72);
CACTTAAGAG (SEQ ID NO:73); CAGTGAACCT (SEQ ID NO:74);
AGCCCAGAGG (SEQ ID NO:75); TTATTGGAAA (SEQ ID NO:76);
CAAAGTGGCA (SEQ ID NO:77); and CACAGACCTC (SEQ ID NO:78).

In some embodiments, a composition contains two or more differently labeled ADH7 oligonucleotides for simultaneously probing the identity of nucleotides or nucleotide pairs at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

ADH7 oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized ADH7 oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two ADH7 oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the ADH7 gene in an individual. As used herein, the terms "ADH7 genotype" and "ADH7 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the ADH7 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of a genotyping method of the invention involves isolating from the individual a nucleic acid sample comprising the two copies of the ADH7 gene, mRNA transcripts thereof or cDNA copies thereof, or a fragment of any of the foregoing, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15 in the two copies to assign an ADH7 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene, mRNA or cDNA (or fragment of such ADH7 molecules) in an individual may be the same allele or may be different alleles. In another embodiment, a genotyping method of the invention comprises determining the identity of the nucleotide pair at each of PS1-PS15.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the ADH7 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions if not present in the mRNA or cDNA. If an ADH7 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of a haplotyping method of the invention comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the ADH7 gene, mRNA or cDNA, or a fragment of such ADH7 molecules, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15 in that copy to assign an ADH7 haplotype to the individual.

The nucleic acid used in the above haplotyping methods of the invention may be isolated using any method capable of separating the two copies of the ADH7 gene or fragment such as one of the methods described above for preparing ADH7 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will typically only provide haplotype information on one of the two ADH7 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional ADH7 clones will usually need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the ADH7 gene in an individual. In some cases, however, once the haplotype for one ADH7 allele is directly determined, the haplotype for the other allele may be inferred if the individual has a known genotype for the polymorphic sites of interest or if the haplotype frequency or haplotype pair frequency for the individual's population group is known. In a particularly preferred embodiment, the nucleotide at each of PS1-PS15 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the ADH7 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's ADH7 gene, the phased sequence of nucleotides present at each of PS1-PS15. This identifying step does not necessarily require that each of PS1-PS15 be directly examined. Typically only a subset of PS1-PS15 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* **97**:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* **22**:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* **4**:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In another embodiment of a haplotyping method of the invention, an ADH7 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15 in each copy of the ADH7 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS15 in each copy of the ADH7 gene.

When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the ADH7 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988). Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* **89**:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of

the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the ADH7 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, mRNA, cDNA or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by haplotyping or genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Detection of the allele(s) present at a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's ADH7 haplotype pair is predicted from its ADH7 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying an ADH7 genotype for the individual at two or more ADH7 polymorphic sites described herein, accessing data containing ADH7 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the genotype data. In one embodiment, the reference haplotype pairs include the ADH7 haplotype pairs shown in Table 4. The ADH7 haplotype pair can be assigned by comparing the individual's genotype with the genotypes corresponding to the haplotype pairs known to exist in the general population or in a specific population group, and determining which haplotype pair is consistent with the genotype of the individual. In some embodiments, comparison of the genotype of the individual to the haplotype pairs identified in a reference population and determination of which haplotype pair is consistent with the genotype of the individual may be performed by visual inspection (for example, by consulting Table 4). When the genotype of the individual is consistent with more than one haplotype pair, haplotype pair frequency data (such as that presented in Table 7) may be used to determine which of these haplotype pairs is most likely to be present in the individual. This determination may also be performed in some embodiments by visual inspection upon consulting Table 7. If a particular ADH7 haplotype pair consistent with the genotype of the individual is more frequent in the reference population than others consistent with the genotype, then that haplotype pair with the highest frequency is the most likely to be present in the individual. In other embodiments, the comparison may be made by a computer-implemented algorithm with the genotype of the individual and the reference haplotype data stored in computer-readable formats. For example, as described in PCT/US01/12831, filed April 18, 2001, one computer-implemented algorithm to perform this comparison entails enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing ADH7 haplotype pairs frequency data determined in a reference population to determine a probability that the individual has a possible haplotype pair, and analyzing the determined probabilities to assign a haplotype pair to the individual.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from

Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting an ADH7 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22; copending PCT/US01/12831 filed April 18, 2001) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible

haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of an ADH7 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel ADH7 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be e.g., a reference population, a family population, a same gender population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for ADH7 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and an ADH7 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. In one embodiment, the method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one or more of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by a predictive genotype to haplotype approach as described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular ADH7 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that ADH7 genotype, haplotype or haplotype pair. Preferably, the ADH7 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting ADH7 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified

diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and/or adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and an ADH7 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the ADH7 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and ADH7 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their ADH7 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression

calculation of which polymorphic sites in the ADH7 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in WO 01/01218, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between ADH7 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra,* Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in WO 01/01218.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the ADH7 gene. As described in WO 01/01218, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of ADH7 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the ADH7 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the ADH7 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying ADH7 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the ADH7 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant ADH7 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and

PS15. Similarly, the nucleotide sequence of a variant fragment of the ADH7 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the ADH7 gene, which is defined by haplotype 16, (or other reported ADH7 sequences) or to portions of the reference sequence (or other reported ADH7 sequences), except for the haplotyping and genotyping oligonucleotides described above.

The location of a polymorphism in a variant ADH7 gene or fragment is preferably identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of cytosine at PS1, cytosine at PS2, adenine at PS3, cytosine at PS4, cytosine at PS5, cytosine at PS6, thymine at PS7, guanine at PS8, thymine at PS9, adenine at PS10, guanine at PS11, guanine at PS12, thymine at PS13, thymine at PS14 and adenine at PS15. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the ADH7 gene which is defined by any one of haplotypes 1- 15 and 17 - 18 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the ADH7 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant or fragment thereof, that is claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art. Any particular ADH7 variant or fragment thereof may also be prepared using synthetic or semi-synthetic methods known in the art.

ADH7 isogenes, or fragments thereof, may be isolated using any method that allows separation of the two "copies" of the ADH7 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides ADH7 genome anthologies, which are collections of at least two ADH7 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same gender population. An ADH7 genome anthology may comprise individual ADH7 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the ADH7 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of such isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in

buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred ADH7 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below. An ADH7 genome anthology is useful in providing control nucleic acids for kits of the invention.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded ADH7 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant ADH7 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the ADH7 gene will produce ADH7 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of an ADH7 cDNA comprising a nucleotide sequence which is a polymorphic variant of the ADH7 reference coding sequence shown in Figure 2. Thus, the invention also provides ADH7 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig.

2), or its corresponding RNA sequence, for those regions of SEQ ID NO:2 that correspond to the examined portions of the ADH7 gene (as described in the Examples below), except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 355, adenine at a position corresponding to nucleotide 654, guanine at a position corresponding to nucleotide 676 and guanine at a position corresponding to nucleotide 697. A particularly preferred polymorphic cDNA variant comprises the coding sequence of an ADH7 isogene defined by any one of haplotypes 1, 2, 4, 6, 9, 10, 14, 15 and 18. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain one or more of the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized ADH7 mRNAs, cDNAs or fragments thereof. Polynucleotides comprising a variant ADH7 RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of an ADH7 gene, mRNA or cDNA fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the ADH7 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the ADH7 gene or cDNA may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the ADH7 genomic, mRNA and cDNA variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment of the invention may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular ADH7 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the ADH7 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular ADH7 isogene. Expression of an ADH7 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA or antisense RNA for the isogene or fragment thereof. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions

−10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of ADH7 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of ADH7 mRNA transcribed from a particular isogene.

The untranslated mRNA, antisense RNA or antisense oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, such molecules may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of (a) the reference ADH7 amino acid sequence shown in Figure 3 or (b) a fragment of this reference sequence. The location of a variant amino acid in an ADH7 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). An ADH7 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 for those regions of SEQ ID NO:3 that are encoded by examined portions of the ADH7 gene (as described in the Examples below), except for having one or more variant amino acids selected from the group consisting of cysteine at a position corresponding to amino acid position 119, glutamic acid at a position corresponding to amino acid position 226 and valine at a position corresponding to amino acid position 233. Thus, an ADH7 fragment of the invention, also referred to herein as an ADH7 peptide variant, is any fragment of an ADH7 protein variant that contains one or more of the amino acid variations shown in Table 2. The invention specifically excludes amino acid sequences identical to those previously identified for ADH7, including SEQ ID NO:3, and previously described fragments thereof. ADH7 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, an ADH7 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes, 1, 2, 4, 6, 9, 10, 14, 15 and 18, shown in Table 5.

Table 2. Novel Polymorphic Variants of ADH7

Polymorphic Variant Number	Amino Acid Position and Identities		
	119	226	233
1	G	K	V
2	G	E	M
3	G	E	V
4	C	K	M
5	C	K	V
6	C	E	M
7	C	E	V

An ADH7 peptide variant of the invention is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such ADH7 peptide variants may be useful as antigens to generate antibodies specific for one of the above ADH7 isoforms. In addition, the ADH7 peptide variants may be useful in drug screening assays.

An ADH7 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing an appropriate variant ADH7 genomic or cDNA sequence described above. Alternatively, the ADH7 protein variant may be isolated from a biological sample of an individual having an ADH7 isogene which encodes the variant protein. Where the sample contains two different ADH7 isoforms (i.e., the individual has different ADH7 isogenes), a particular ADH7 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular ADH7 isoform but does not bind to the other ADH7 isoform.

The expressed or isolated ADH7 protein or peptide may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the ADH7 protein or peptide as discussed further below. ADH7 variant proteins and peptides can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant ADH7 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric ADH7 protein. The non-ADH7 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the ADH7 and non-ADH7 portions so that the ADH7 protein may be cleaved and purified away from the non-ADH7 portion.

An additional embodiment of the invention relates to using a novel ADH7 protein isoform, or a fragment thereof, in any of a variety of drug screening assays. Such screening assays may be

performed to identify agents that bind specifically to all known ADH7 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The ADH7 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to an ADH7 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the ADH7 protein(s) of interest and then washed. Bound ADH7 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel ADH7 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the ADH7 protein.

In yet another embodiment, when a particular ADH7 haplotype or group of ADH7 haplotypes encodes an ADH7 protein variant with an amino acid sequence distinct from that of ADH7 protein isoforms encoded by other ADH7 haplotypes, then detection of that particular ADH7 haplotype or group of ADH7 haplotypes may be accomplished by detecting expression of the encoded ADH7 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel ADH7 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The ADH7 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the ADH7 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the ADH7 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the ADH7 protein variant from solution as well as react with ADH7 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect ADH7 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel ADH7 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the ADH7 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of ADH7 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the ADH7 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into ADH7 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired ADH7 isogene may be introduced

into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the ADH7 isogene is introduced into a cell in such a way that it recombines with the endogenous ADH7 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired ADH7 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the ADH7 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the ADH7 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant ADH7 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the ADH7 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human ADH7 isogene and producing the encoded human ADH7 protein can be used as biological models for studying diseases related to abnormal ADH7 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel ADH7 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel ADH7 isogenes; an antisense oligonucleotide directed against one of the novel ADH7 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel ADH7 isogene described herein. Preferably, the

composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel ADH7 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the ADH7 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The ADH7 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the

specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the ADH7 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the ADH7 gene were amplified using 'tailed' PCR primers, each of which includes a universal sequence forming a noncomplementary 'tail' attached to the 5′ end of each unique sequence in the PCR primer pairs. The universal 'tail' sequence for the forward PCR primers comprises the sequence 5′-TGTAAAACGACGGCCAGT-3′ (SEQ ID NO:79) and the universal 'tail' sequence for the reverse PCR primers comprises the sequence 5′-AGGAAACAGCTATGACCAT-3′ (SEQ ID NO:80). The nucleotide positions of the first and last nucleotide of the forward and reverse primers for each region amplified are presented below and correspond to positions in SEQ ID NO:1 (Figure 1).

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	3090-3113	complement of 3636-3613	547 nt
Fragment 2	3396-3418	complement of 3941-3920	546 nt
Fragment 3	3703-3726	complement of 4266-4244	564 nt
Fragment 4	9497-9516	complement of 9967-9944	471 nt
Fragment 5	11156-11180	complement of 11699-11680	544 nt
Fragment 6	18361-18383	complement of 18730-18710	370 nt
Fragment 7	18601-18624	complement of 19030-19005	430 nt
Fragment 8	19995-20016	complement of 20520-20497	526 nt
Fragment 9	23492-23514	complement of 24150-24129	659 nt
Fragment 10	25897-25919	complement of 26300-26280	404 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:
97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec. } 10 cycles
72°C - 45 sec.

97°C - 15 sec.
64°C - 45 sec. } 35 cycles
72°C - 45 sec.

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the appropriate universal 'tail' sequence as a primer. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the ADH7 reference genomic sequence (SEQ ID NO:1) are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the ADH7 Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	12777638	3901	T	C		
PS2	12163660	3996	T	C		
PS3	12163751	4015	G	A		
PS4	12163842	4019	T	C		
PS5	12163935	4060	T	C		
PS6	12164026	4148	T	C		
PS7	12164303	9684	C	T		
PS8	12509776	11281	T	G		
PS9	12510058	11346	G	T	355	G119C
PS10	11897699	18624	G	A	654	R218R
PS11	11897796	18646	A	G	676	K226E
PS12	11897893	18667	A	G	697	M233V
PS13	11901180	18979	C	T		
PS14	11898360	26036	A	T		
PS15	11898645	26195	G	A		

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

EXAMPLE 2

This example illustrates analysis of the ADH7 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in unrelated members of the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4 (Part1). Genotypes and Haplotype Pairs Observed for ADH7 Gene

Genotype Number	Polymorphic Sites PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	HAP	Pair
1	T	T	G	T	T	C	C	T	G	G	11	11
2	T	T	G	T	T	T	C	T	G	G	16	16
3	T	T	G	C	T	C	C	T	G	G	5	5
4	T/C	T	G	T	T	C	T/C	T/G	G	G/A	15	2
5	T	T	G	T	T	T/C	C/T	T	G	G	16	15
6	T	T	G	T	T	C	C/T	T	G	G	12	15
7	T	T	G	T	T	C	C/T	T	G	G	11	14
8	T	T	G	T	T	C	C	T	G	G/A	12	9
9	T	T	G	T	T	T/C	C	T	G	G/A	16	9
10	T	T	G	T/C	T	C/T	C	T	G	G	11	7
11	T/C	T	G	T	T	C	C	T/G	G	G/A	11	1
12	T	T	G	T	T	C	C	T	G	G	11	13
13	T	T	G	T	T/C	T	C	T	G	G	16	8
14	T	T	G	T/C	T	T/C	C	T	G	G	16	5
15	T	T	G	T/C	T	C/T	C	T	G	G	12	7
16	T/C	T	G	T	T	C	C	T/G	G	A	9	1
17	T/C	T	G	C/T	T	C	C	T/G	G	G/A	5	1
18	T	T	G	T/C	T	C	C	T	G	G	12	5
19	T/C	T	G	T	T	T/C	C	T/G	G	G/A	16	1
20	T	T	G	T	T	C/T	C	T	G	G	12	17
21	T	T	G	T/C	T	C	C	T	G/T	G/A	11	6
22	T	T	G	T	T	C/T	C	T	G	G	11	18
23	T	T	G	T	T/C	C/T	C	T	G	G	11	8
24	T	T	G/A	T	T	C	C	T	G	G/A	12	4
25	T	T	G	T	T	C	T/C	T	G	G/A	15	10
26	T	T	G	C/T	T	C	C	T	G	G/A	5	9
27	T	T	G	T	T	C	C	T	G	G/A	11	9
28	T	T	G	T/C	T	C	C	T	G	G	11	5
29	T	T	G	T	T	C	C	T	G	G	11	12
30	T	T	G	T	T	T/C	C	T	G	G	16	12
31	T	T/C	G	T	T	C/T	C	T	G	G	11	3
32	T	T	G	T	T	C/T	C	T	G	G	11	16

Table 4 (Part2). Genotypes and Haplotype Pairs Observed for ADH7 Gene

Genotype Number	Polymorphic Sites PS11	PS12	PS13	PS14	PS15	HAP	Pair
1	A	A	C	A	G	11	11
2	A	A	C	A	G	16	16
3	A	A	C	A	G	5	5
4	G	A	C	A	G	15	2
5	A/G	A	C	A	G	16	15
6	A/G	A	C	T/A	G	12	15
7	A/G	A	C	A	G/A	11	14
8	A	A	C	T/A	G	12	9
9	A	A	C	A	G	16	9
10	A	A	C	A/T	G	11	7
11	A	A	C	A	G	11	1
12	A	A	C/T	A	G	11	13
13	A	A	C	A	G	16	8
14	A	A	C	A	G	16	5
15	A	A	C	T	G	12	7
16	A	A	C	A	G	9	1
17	A	A	C	A	G	5	1
18	A	A	C	T/A	G	12	5
19	A	A	C	A	G	16	1
20	A	A	C	T	G	12	17
21	A	A	C	A/T	G	11	6
22	A	A/G	C	A	G	11	18
23	A	A	C	A	G	11	8
24	A	A	C	T	G	12	4
25	G	A	C	A	G	15	10
26	A	A	C	A	G	5	9
27	A	A	C	A	G	11	9
28	A	A	C	A	G	11	5
29	A	A	C	A/T	G	11	12
30	A	A	C	A/T	G	16	12
31	A	A	C	A/T	G	11	3
32	A	A	C	A	G	11	16

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* **7**, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in PCT/US01/12831, filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In the present analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 18 human ADH7 haplotypes shown in Table 5 below.

An ADH7 isogene defined by a full-haplotype shown in Table 5 below comprises the regions

of the SEQ ID NOS indicated in Table 5, with their corresponding set of polymorphic locations and identities, which are also set forth in Table 5.

Table 5. Haplotypes of the ADH7 Gene

Haplotype Number[a]										PS	PS	SEQ ID	Region
1	2	3	4	5	6	7	8	9	10	No.[b]	Pos.[c]	No.[d]	Examined[e]
C	C	T	T	T	T	T	T	T	T	1	3901	1/81	3090-4266
T	T	C	T	T	T	T	T	T	T	2	3996	1/81	3090-4266
G	G	G	A	G	G	G	G	G	G	3	4015	1/81	3090-4266
T	T	T	T	C	C	C	T	T	T	4	4019	1/81	3090-4266
T	T	T	T	T	T	T	C	T	T	5	4060	1/81	3090-4266
C	C	T	C	C	C	T	T	C	C	6	4148	1/81	3090-4266
C	C	C	C	C	C	C	C	C	C	7	9684	1/81	9497-9967
G	G	T	T	T	T	T	T	T	T	8	11281	1/81	11156-11699
G	G	G	G	G	T	G	G	G	G	9	11346	1/81	11156-11699
A	A	G	A	G	A	G	G	A	A	10	18624	1/81	18361-19030
A	G	A	A	A	A	A	A	A	G	11	18646	1/81	18361-19030
A	A	A	A	A	A	A	A	A	A	12	18667	1/81	18361-19030
C	C	C	C	C	C	C	C	C	C	13	18979	1/81	18361-19030
													19995-20520
													23492-24150
A	A	T	T	A	T	T	A	A	A	14	26036	1/81	25897-26300
G	G	G	G	G	G	G	G	G	G	15	26195	1/81	25897-26300

Haplotype Number[a]								PS	PS	SEQ ID	Region
11	12	13	14	15	16	17	18	No.[b]	Pos.[c]	No.[d]	Examined[e]
T	T	T	T	T	T	T	T	1	3901	1/81	3090-4266
T	T	T	T	T	T	T	T	2	3996	1/81	3090-4266
G	G	G	G	G	G	G	G	3	4015	1/81	3090-4266
T	T	T	T	T	T	T	T	4	4019	1/81	3090-4266
T	T	T	T	T	T	T	T	5	4060	1/81	3090-4266
C	C	C	C	C	T	T	T	6	4148	1/81	3090-4266
C	C	C	T	T	C	C	C	7	9684	1/81	9497-9967
T	T	T	T	T	T	T	T	8	11281	1/81	11156-11699
G	G	G	G	G	G	G	G	9	11346	1/81	11156-11699
G	G	G	G	G	G	G	G	10	18624	1/81	18361-19030
A	A	A	G	G	A	A	A	11	18646	1/81	18361-19030
A	A	A	A	A	A	A	G	12	18667	1/81	18361-19030
C	C	T	C	C	C	C	C	13	18979	1/81	18361-19030
											19995-20520
											23492-24150
A	T	A	A	A	A	T	A	14	26036	1/81	25897-26300
G	G	G	A	G	G	G	G	15	26195	1/81	25897-26300

[a]Alleles for ADH7 haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS within the indicated SEQ ID NO, with the 1st position number referring to the first SEQ ID NO and the 2nd position number referring to the 2nd SEQ ID NO;
[d]1st SEQ ID NO refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol; 2nd SEQ ID NO is a modified version of the 1st SEQ ID NO that comprises the context sequence of each polymorphic site, PS1-PS15, to facilitate electronic searching of the haplotypes;
[e]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:81 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS15 in a uniform format to facilitate electronic searching of the ADH7 haplotypes. For each polymorphic site, SEQ ID NO:81 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 6 below shows the percent of chromosomes characterized by a given ADH7 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given ADH7 haplotype pair is shown in Table 7. In Tables 6 and 7, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 6 and 7 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 6. Frequency of Observed ADH7 Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	12858962	4.27	9.52	0.0	0.0	8.33	0.0
2	12858969	0.61	0.0	0.0	0.0	2.78	0.0
3	12858973	0.61	0.0	0.0	2.5	0.0	0.0
4	12858974	0.61	0.0	2.5	0.0	0.0	0.0
5	12858960	8.54	19.05	0.0	12.5	2.78	0.0
6	12858967	0.61	0.0	2.5	0.0	0.0	0.0
7	12858963	3.05	2.38	0.0	0.0	11.11	0.0
8	12858966	1.22	0.0	2.5	2.5	0.0	0.0
9	12858961	6.71	4.76	10.0	10.0	2.78	0.0
10	12858970	0.61	0.0	2.5	0.0	0.0	0.0
11	12858957	35.37	26.19	40.0	37.5	38.89	33.33
12	12858959	9.15	9.52	15.0	5.0	8.33	0.0
13	12858971	0.61	0.0	2.5	0.0	0.0	0.0
14	12858968	0.61	0.0	2.5	0.0	0.0	0.0
15	12858964	2.44	0.0	5.0	0.0	2.78	16.67
16	12858958	23.17	26.19	10.0	30.0	22.22	50.0
17	12858965	1.22	2.38	2.5	0.0	0.0	0.0
18	12858972	0.61	0.0	2.5	0.0	0.0	0.0

Table 7. Frequency of Observed ADH7 Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
11	11	12.2	4.76	10.0	20.0	16.67	0.0
16	16	7.32	14.29	0.0	15.0	0.0	0.0
5	5	2.44	4.76	0.0	5.0	0.0	0.0
15	2	1.22	0.0	0.0	0.0	5.56	0.0
16	15	1.22	0.0	0.0	0.0	0.0	33.33
12	15	1.22	0.0	5.0	0.0	0.0	0.0
11	14	1.22	0.0	5.0	0.0	0.0	0.0
12	9	2.44	0.0	10.0	0.0	0.0	0.0
16	9	1.22	4.76	0.0	0.0	0.0	0.0
11	7	4.88	4.76	0.0	0.0	16.67	0.0
11	1	1.22	4.76	0.0	0.0	0.0	0.0
11	13	1.22	0.0	5.0	0.0	0.0	0.0
16	8	1.22	0.0	0.0	5.0	0.0	0.0
16	5	1.22	0.0	0.0	0.0	5.56	0.0
12	7	1.22	0.0	0.0	0.0	5.56	0.0
9	1	1.22	0.0	0.0	0.0	5.56	0.0
5	1	2.44	9.52	0.0	0.0	0.0	0.0
12	5	1.22	4.76	0.0	0.0	0.0	0.0
16	1	3.66	4.76	0.0	0.0	11.11	0.0
12	17	2.44	4.76	5.0	0.0	0.0	0.0
11	6	1.22	0.0	5.0	0.0	0.0	0.0
11	18	1.22	0.0	5.0	0.0	0.0	0.0
11	8	1.22	0.0	5.0	0.0	0.0	0.0
12	4	1.22	0.0	5.0	0.0	0.0	0.0
15	10	1.22	0.0	5.0	0.0	0.0	0.0
5	9	3.66	4.76	0.0	10.0	0.0	0.0
11	9	4.88	0.0	10.0	10.0	0.0	0.0
11	5	3.66	9.52	0.0	5.0	0.0	0.0
11	12	4.88	9.52	5.0	0.0	5.56	0.0
16	12	3.66	0.0	0.0	10.0	5.56	0.0
11	3	1.22	0.0	0.0	5.0	0.0	0.0
11	16	19.51	14.29	20.0	15.0	22.22	66.67

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the

relative frequencies determined therein for the haplotypes and haplotype pairs of the ADH7 gene are likely to be similar to the relative frequencies of these ADH7 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene of an individual, which comprises determining which of the ADH7 haplotypes shown in the table immediately below defines one copy of the individual's ADH7 gene, wherein each of the ADH7 haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Haplotype Number[a]										PS	PS
1	2	3	4	5	6	7	8	9	10	No.[b]	Position[c]
C	C	T	T	T	T	T	T	T	T	1	3901
T	T	C	T	T	T	T	T	T	T	2	3996
G	G	G	A	G	G	G	G	G	G	3	4015
T	T	T	T	C	C	C	T	T	T	4	4019
T	T	T	T	T	T	T	C	T	T	5	4060
C	C	T	C	C	C	T	T	C	C	6	4148
C	C	C	C	C	C	C	C	C	C	7	9684
G	G	T	T	T	T	T	T	T	T	8	11281
G	G	G	G	G	T	G	G	G	G	9	11346
A	A	G	A	G	A	G	G	A	A	10	18624
A	G	A	A	A	A	A	A	A	G	11	18646
A	A	A	A	A	A	A	A	A	A	12	18667
C	C	C	C	C	C	C	C	C	C	13	18979
A	A	T	T	A	T	T	A	A	A	14	26036
G	G	G	G	G	G	G	G	G	G	15	26195

Haplotype Number[a]								PS	PS
11	12	13	14	15	16	17	18	No.[b]	Position[c]
T	T	T	T	T	T	T	T	1	3901
T	T	T	T	T	T	T	T	2	3996
G	G	G	G	G	G	G	G	3	4015
T	T	T	T	T	T	T	T	4	4019
T	T	T	T	T	T	T	T	5	4060
C	C	C	C	C	T	T	T	6	4148
C	C	C	T	T	C	C	C	7	9684
T	T	T	T	T	T	T	T	8	11281
G	G	G	G	G	G	G	G	9	11346
G	G	G	G	G	G	G	G	10	18624
A	A	A	G	G	A	A	A	11	18646
A	A	A	A	A	A	A	G	12	18667
C	C	T	C	C	C	C	C	13	18979
A	T	A	A	A	A	T	A	14	26036
G	G	G	A	G	G	G	G	15	26195

[a]Alleles for haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS within SEQ ID NO:1.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS15 on at least one copy of the individual's ADH7 gene.

3. A method for haplotyping the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide

(ADH7) gene of an individual, which comprises determining which of the ADH7 haplotype pairs shown in the table immediately below defines both copies of the individual's ADH7 gene, wherein each of the ADH7 haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions and identities are set forth in the table immediately below:

Haplotype Pair[a]								PS No.[b]	PS Position[c]
11/11	16/16	5/5	15/2	16/15	12/15	11/14	12/9		
T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	T/T	C/C	C/C	T/C	C/C	C/C	C/C	6	4148
C/C	C/C	C/C	T/C	C/T	C/T	C/T	C/C	7	9684
T/T	T/T	T/T	T/G	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	G/G	A/G	A/G	A/G	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/A	T/A	A/A	T/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
16/9	11/7	11/1	11/13	16/8	16/5	12/7	9/1		
T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/C	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/C	T/T	T/T	T/T	T/C	T/C	T/T	4	4019
T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	5	4060
T/C	C/T	C/C	C/C	T/T	T/C	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/G	T/T	T/T	T/T	T/T	T/G	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/G	G/G	G/G	A/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	13	18979
A/A	A/T	A/A	A/A	A/A	A/A	T/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
5/1	12/5	16/1	12/17	11/6	11/18	11/8	12/4		
T/C	T/T	T/C	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	3	4015
C/T	T/C	T/T	T/T	T/C	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	5	4060
C/C	C/C	T/C	C/T	C/C	C/T	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/G	T/T	T/G	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/A	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	T/A	A/A	T/T	A/T	A/A	A/A	T/T	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
15/10	5/9	11/9	11/5	11/12	16/12	11/3	11/16		
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	C/T	T/T	T/C	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	C/C	C/C	C/C	C/C	T/C	C/T	C/T	6	4148
T/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/A	G/A	G/G	G/G	G/G	G/G	G/G	10	18624
G/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/T	A/T	A/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS15 on both copies of the individual's ADH7 gene.

5. A method for genotyping the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene of an individual, comprising determining for the two copies of the ADH7 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15, wherein the one or more PS have the position and alternative alleles shown in SEQ ID NO:1.

6. The method of claim 5, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the ADH7 gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for genotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the ADH7 gene present in the individual the identity of the nucleotide pair at each of PS1-PS15.

8. A method for haplotyping the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene of an individual which comprises determining, for one copy of the ADH7 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

9. The method of claim 8, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the ADH7 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid sample a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for haplotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

10. A method for predicting a haplotype pair for the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene of an individual comprising:

(a) identifying an ADH7 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting

of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1;

(b) comparing the genotype to the haplotype pair data set forth in the table immediately below; and

(c) determining which haplotype pair is consistent with the genotype of the individual and with the haplotype pair data

Haplotype Pair[a]								PS No.[b]	PS Position[c]
11/11	16/16	5/5	15/2	16/15	12/15	11/14	12/9		
T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	T/T	C/C	C/C	T/C	C/C	C/C	C/C	6	4148
C/C	C/C	C/C	T/C	C/T	C/T	C/T	C/C	7	9684
T/T	T/T	T/T	T/G	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	G/G	A/G	A/G	A/G	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/A	T/A	A/A	T/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
16/9	11/7	11/1	11/13	16/8	16/5	12/7	9/1		
T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/C	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/C	T/T	T/T	T/T	T/C	T/C	T/T	4	4019
T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	5	4060
T/C	C/T	C/C	C/C	T/T	T/C	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/G	T/T	T/T	T/T	T/T	T/G	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/G	G/G	G/G	A/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	13	18979
A/A	A/T	A/A	A/A	A/A	A/A	T/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
5/1	12/5	16/1	12/17	11/6	11/18	11/8	12/4		
T/C	T/T	T/C	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	3	4015
C/T	T/C	T/T	T/T	T/C	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	5	4060
C/C	C/C	T/C	C/T	C/C	C/T	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/G	T/T	T/G	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/A	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	T/A	A/A	T/T	A/T	A/A	A/A	T/T	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
15/10	5/9	11/9	11/5	11/12	16/12	11/3	11/16		
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	C/T	T/T	T/C	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	C/C	C/C	C/C	C/C	T/C	C/T	C/T	6	4148
T/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/A	G/A	G/G	G/G	G/G	G/G	G/G	10	18624
G/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/T	A/T	A/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS15, which have the position and alternative alleles shown in SEQ ID NO:1.

12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-18 shown in the table presented immediately below, wherein each of the haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Haplotype Number[a]										PS	PS
1	2	3	4	5	6	7	8	9	10	No.[b]	Position[c]
C	C	T	T	T	T	T	T	T	T	1	3901
T	T	C	T	T	T	T	T	T	T	2	3996
G	G	G	A	G	G	G	G	G	G	3	4015
T	T	T	T	C	C	C	T	T	T	4	4019
T	T	T	T	T	T	T	C	T	T	5	4060
C	C	T	C	C	C	T	T	C	C	6	4148
C	C	C	C	C	C	C	C	C	C	7	9684
G	G	T	T	T	T	T	T	T	T	8	11281
G	G	G	G	G	T	G	G	G	G	9	11346
A	A	G	A	G	A	G	G	A	A	10	18624
A	G	A	A	A	A	A	A	A	G	11	18646
A	A	A	A	A	A	A	A	A	A	12	18667
C	C	C	C	C	C	C	C	C	C	13	18979
A	A	T	T	A	T	T	A	A	A	14	26036
G	G	G	G	G	G	G	G	G	G	15	26195

Haplotype Number[a]								PS	PS
11	12	13	14	15	16	17	18	No.[b]	Position[c]
T	T	T	T	T	T	T	T	1	3901
T	T	T	T	T	T	T	T	2	3996
G	G	G	G	G	G	G	G	3	4015
T	T	T	T	T	T	T	T	4	4019
T	T	T	T	T	T	T	T	5	4060
C	C	C	C	C	T	T	T	6	4148
C	C	C	T	T	C	C	C	7	9684
T	T	T	T	T	T	T	T	8	11281
G	G	G	G	G	G	G	G	9	11346
G	G	G	G	G	G	G	G	10	18624
A	A	A	G	G	A	A	A	11	18646
A	A	A	A	A	A	A	G	12	18667
C	C	T	C	C	C	C	C	13	18979
A	T	A	A	A	A	T	A	14	26036
G	G	G	A	G	G	G	G	15	26195

[a]Alleles for haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the ADH7 haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions and identities are set forth in the table immediately below:

Haplotype Pair[a]								PS	PS
11/11	16/16	5/5	15/2	16/15	12/15	11/14	12/9	No.[b]	Position[c]
T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	T/T	C/C	C/C	T/C	C/C	C/C	C/C	6	4148
C/C	C/C	C/C	T/C	C/T	C/T	C/T	C/C	7	9684
T/T	T/T	T/T	T/G	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	G/G	A/G	A/G	A/G	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/A	T/A	A/A	T/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	15	26195

Haplotype Pair[a]								PS	PS
16/9	11/7	11/1	11/13	16/8	16/5	12/7	9/1	No.[b]	Position[c]
T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/C	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/C	T/T	T/T	T/T	T/C	T/C	T/T	4	4019
T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	5	4060
T/C	C/T	C/C	C/C	T/T	T/C	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/G	T/T	T/T	T/T	T/T	T/G	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/G	G/G	G/G	A/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	13	18979
A/A	A/T	A/A	A/A	A/A	A/A	T/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS	PS
5/1	12/5	16/1	12/17	11/6	11/18	11/8	12/4	No.[b]	Position[c]
T/C	T/T	T/C	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	3	4015
C/T	T/C	T/T	T/T	T/C	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	5	4060
C/C	C/C	T/C	C/T	C/C	C/T	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/G	T/T	T/G	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/A	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	T/A	A/A	T/T	A/T	A/A	A/A	T/T	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
15/10	5/9	11/9	11/5	11/12	16/12	11/3	11/16		
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	C/T	T/T	T/C	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	C/C	C/C	C/C	C/C	T/C	C/T	C/T	6	4148
T/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/A	G/A	G/G	G/G	G/G	G/G	G/G	10	18624
G/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/T	A/T	A/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

13. The method of claim 12, wherein the trait is a clinical response to a drug targeting ADH7.

14. An isolated oligonucleotide designed for detecting a polymorphism in the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

15. The isolated oligonucleotide of claim 14, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the ADH7 gene at a region containing the polymorphic site.

16. The allele-specific oligonucleotide of claim 15, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-18, the complements of SEQ ID NOS:4-18, and SEQ ID NOS:19-48.

17. The isolated oligonucleotide of claim 14, which is a primer-extension oligonucleotide.

18. The primer-extension oligonucleotide of claim 17,which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:49-78.

19. A kit for haplotyping or genotyping the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene of an individual, which comprises a set of oligonucleotides designed to haplotype or genotype each of polymorphic sites (PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14 and PS15, wherein the selected PS have the position

and alternative alleles shown in SEQ ID NO:1.

20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which comprises a alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) isogene, wherein the ADH7 isogene is selected from the group consisting of isogenes 1- 15 and 17 - 18 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1- 15 and 17 - 18 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below; and

Isogene Number[a]										PS	PS	Region
1	2	3	4	5	6	7	8	9	10	No.[b]	Pos.[c]	Examined[d]
C	C	T	T	T	T	T	T	T	T	1	3901	3090-4266
T	T	C	T	T	T	T	T	T	T	2	3996	3090-4266
G	G	G	A	G	G	G	G	G	G	3	4015	3090-4266
T	T	T	T	C	C	C	T	T	T	4	4019	3090-4266
T	T	T	T	T	T	T	C	T	T	5	4060	3090-4266
C	C	T	C	C	C	T	T	C	C	6	4148	3090-4266
C	C	C	C	C	C	C	C	C	C	7	9684	9497-9967
G	G	T	T	T	T	T	T	T	T	8	11281	11156-11699
G	G	G	G	G	T	G	G	G	G	9	11346	11156-11699
A	A	G	A	G	A	G	G	A	A	10	18624	18361-19030
A	G	A	A	A	A	A	A	A	G	11	18646	18361-19030
A	A	A	A	A	A	A	A	A	A	12	18667	18361-19030
C	C	C	C	C	C	C	C	C	C	13	18979	18361-19030 19995-20520 23492-24150
A	A	T	T	A	T	T	A	A	A	14	26036	25897-26300
G	G	G	G	G	G	G	G	G	G	15	26195	25897-26300

Isogene Number[a]							PS	PS	Region
11	12	13	14	15	17	18	No.[b]	Pos.[c]	Examined[d]
T	T	T	T	T	T	T	1	3901	3090-4266
T	T	T	T	T	T	T	2	3996	3090-4266
G	G	G	G	G	G	G	3	4015	3090-4266
T	T	T	T	T	T	T	4	4019	3090-4266
T	T	T	T	T	T	T	5	4060	3090-4266
C	C	C	C	C	T	T	6	4148	3090-4266
C	C	C	T	T	C	C	7	9684	9497-9967
T	T	T	T	T	T	T	8	11281	11156-11699
G	G	G	G	G	G	G	9	11346	11156-11699
G	G	G	G	G	G	G	10	18624	18361-19030
A	A	A	G	G	A	A	11	18646	18361-19030
A	A	A	A	A	A	G	12	18667	18361-19030
C	C	T	C	C	C	C	13	18979	18361-19030
									19995-20520
									23492-24150
A	T	A	A	A	T	A	14	26036	25897-26300
G	G	G	A	G	G	G	15	26195	25897-26300

[a]Alleles for isogenes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[e]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

21. The isolated polynucleotide of claim 20, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

22. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 20, wherein the organism expresses an ADH7 protein that is encoded by the first nucleotide sequence.

23. The recombinant nonhuman organism of claim 22, which is a transgenic animal.

24. An isolated fragment of a alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) isogene, wherein the fragment comprises at least 10 nucleotides in one of the regions of SEQ ID NO:1 shown in the table immediately below and wherein the fragment comprises one or more polymorphisms selected from the group consisting of cytosine at PS1, cytosine at PS2, adenine at PS3, cytosine at PS4, cytosine at PS5, cytosine at PS6, thymine at PS7, guanine at PS8, thymine at PS9, adenine at PS10, guanine at PS11, guanine at PS12, thymine at PS13, thymine at PS14 and adenine at PS15, wherein the selected polymorphism has the position set forth in the table immediately below:

Isogene Number[a]										PS	PS	Region
1	2	3	4	5	6	7	8	9	10	No.[b]	Pos.[c]	Examined[d]
C	C	T	T	T	T	T	T	T	T	1	3901	3090-4266
T	T	C	T	T	T	T	T	T	T	2	3996	3090-4266
G	G	G	A	G	G	G	G	G	G	3	4015	3090-4266
T	T	T	T	C	C	C	T	T	T	4	4019	3090-4266
T	T	T	T	T	T	T	C	T	T	5	4060	3090-4266
C	C	T	C	C	C	T	T	C	C	6	4148	3090-4266
C	C	C	C	C	C	C	C	C	C	7	9684	9497-9967
G	G	T	T	T	T	T	T	T	T	8	11281	11156-11699
G	G	G	G	G	T	G	G	G	G	9	11346	11156-11699
A	A	G	A	G	A	G	G	A	A	10	18624	18361-19030
A	G	A	A	A	A	A	A	A	G	11	18646	18361-19030
A	A	A	A	A	A	A	A	A	A	12	18667	18361-19030
C	C	C	C	C	C	C	C	C	C	13	18979	18361-19030
												19995-20520
												23492-24150
A	A	T	T	A	T	T	A	A	A	14	26036	25897-26300
G	G	G	G	G	G	G	G	G	G	15	26195	25897-26300

Isogene Number[a]							PS	PS	Region
11	12	13	14	15	17	18	No.[b]	Pos.[c]	Examined[d]
T	T	T	T	T	T	T	1	3901	3090-4266
T	T	T	T	T	T	T	2	3996	3090-4266
G	G	G	G	G	G	G	3	4015	3090-4266
T	T	T	T	T	T	T	4	4019	3090-4266
T	T	T	T	T	T	T	5	4060	3090-4266
C	C	C	C	C	T	T	6	4148	3090-4266
C	C	C	T	T	C	C	7	9684	9497-9967
T	T	T	T	T	T	T	8	11281	11156-11699
G	G	G	G	G	G	G	9	11346	11156-11699
G	G	G	G	G	G	G	10	18624	18361-19030
A	A	A	G	G	A	A	11	18646	18361-19030
A	A	A	A	A	A	G	12	18667	18361-19030
C	C	T	C	C	C	C	13	18979	18361-19030
									19995-20520
									23492-24150
A	T	A	A	A	T	A	14	26036	25897-26300
G	G	G	A	G	G	G	15	26195	25897-26300

[a]Alleles for isogenes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[e]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

25. An isolated polynucleotide comprising an ADH7 coding sequence, wherein the coding sequence is selected from the group consisting of 1, 2, 4, 6, 9, 10, 14, 15 and 18 shown in the table immediately below, and wherein each of the coding sequences comprises the regions of SEQ ID NO:2 that are defined by exons 1,2 and 6-9, except at each of the polymorphic sites which have the positions in SEQ ID NO:2 and polymorphisms set forth in the table immediately below:

Coding Sequence Haplotype Number									PS	PS
1c	2c	4c	6c	9c	10c	14c	15c	18c	No.[b]	Position[c]
G	G	G	T	G	G	G	G	G	9	355
A	A	A	A	A	A	G	G	G	10	654
A	G	A	A	A	G	G	G	A	11	676
A	A	A	A	A	A	A	A	G	12	697

[a]Alleles for the isogene coding sequence are presented 5′ to 3′ in each column; the numerical portion of the isogene coding sequence number represents the number of the parent full ADH7 isogene;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:2.

26. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 25, wherein the organism expresses a alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) protein that is encoded by the polymorphic variant sequence.
27. The recombinant nonhuman organism of claim 26, which is a transgenic animal.
28. An isolated fragment of an ADH7 cDNA, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 355, adenine at a position corresponding to nucleotide 654, guanine at a position corresponding to nucleotide 676 and guanine at a position corresponding to nucleotide 697 in SEQ ID NO:2.
29. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) protein, wherein the reference sequence comprises SEQ ID NO:3 for the regions encoded by exons 1,2 and 6-9, except the polymorphic variant comprises one or more variant amino acids selected from the group consisting of cysteine at a position corresponding to amino acid position 119, glutamic acid at a position corresponding to amino acid position 226 and valine at a position corresponding to amino acid position 233.
30. An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 29.
31. A method for screening for drugs targeting the isolated polypeptide of claim 29 which comprises contacting the ADH7 polymorphic variant with a candidate agent and assaying for binding activity.
32. An isolated fragment of the ADH7 polypeptide, wherein the fragment comprises one or more variant amino acids selected from the group consisting of cysteine at a position corresponding to amino acid position 119, glutamic acid at a position corresponding to amino acid position 226 and valine at a position corresponding to amino acid position 233 in SEQ ID NO:3.
33. A computer system for storing and analyzing polymorphism data for the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide gene, comprising:
    (a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises any one or more of the haplotypes set forth in the table immediately below:

Haplotype Number[a]										PS	PS
1	2	3	4	5	6	7	8	9	10	No.[b]	Position[c]
C	C	T	T	T	T	T	T	T	T	1	3901
T	T	C	T	T	T	T	T	T	T	2	3996
G	G	G	A	G	G	G	G	G	G	3	4015
T	T	T	T	C	C	C	T	T	T	4	4019
T	T	T	T	T	T	T	C	T	T	5	4060
C	C	T	C	C	C	T	T	C	C	6	4148
C	C	C	C	C	C	C	C	C	C	7	9684
G	G	T	T	T	T	T	T	T	T	8	11281
G	G	G	G	G	T	G	G	G	G	9	11346
A	A	G	A	G	A	G	G	A	A	10	18624
A	G	A	A	A	A	A	A	A	G	11	18646
A	A	A	A	A	A	A	A	A	A	12	18667
C	C	C	C	C	C	C	C	C	C	13	18979
A	A	T	T	A	T	T	A	A	A	14	26036
G	G	G	G	G	G	G	G	G	G	15	26195

Haplotype Number[a]								PS	PS
11	12	13	14	15	16	17	18	No.[b]	Position[c]
T	T	T	T	T	T	T	T	1	3901
T	T	T	T	T	T	T	T	2	3996
G	G	G	G	G	G	G	G	3	4015
T	T	T	T	T	T	T	T	4	4019
T	T	T	T	T	T	T	T	5	4060
C	C	C	C	C	T	T	T	6	4148
C	C	C	T	T	C	C	C	7	9684
T	T	T	T	T	T	T	T	8	11281
G	G	G	G	G	G	G	G	9	11346
G	G	G	G	G	G	G	G	10	18624
A	A	A	G	G	A	A	A	11	18646
A	A	A	A	A	A	A	G	12	18667
C	C	T	C	C	C	C	C	13	18979
A	T	A	A	A	A	T	A	14	26036
G	G	G	A	G	G	G	G	15	26195

[a]Alleles for haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

the haplotype pairs set forth in the table immediately below:

Haplotype Pair[a]								PS No.[b]	PS Position[c]
11/11	16/16	5/5	15/2	16/15	12/15	11/14	12/9		
T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/T	C/C	T/T	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	T/T	C/C	C/C	T/C	C/C	C/C	C/C	6	4148
C/C	C/C	C/C	T/C	C/T	C/T	C/T	C/C	7	9684
T/T	T/T	T/T	T/G	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	G/G	A/G	A/G	A/G	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/A	T/A	A/A	T/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
16/9	11/7	11/1	11/13	16/8	16/5	12/7	9/1		
T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/C	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	T/C	T/T	T/T	T/T	T/C	T/C	T/T	4	4019
T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	5	4060
T/C	C/T	C/C	C/C	T/T	T/C	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/G	T/T	T/T	T/T	T/T	T/G	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/G	G/G	G/G	A/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	13	18979
A/A	A/T	A/A	A/A	A/A	A/A	T/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
5/1	12/5	16/1	12/17	11/6	11/18	11/8	12/4		
T/C	T/T	T/C	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	3	4015
C/T	T/C	T/T	T/T	T/C	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	5	4060
C/C	C/C	T/C	C/T	C/C	C/T	C/T	C/C	6	4148
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/G	T/T	T/G	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	9	11346
G/A	G/G	G/A	G/G	G/A	G/G	G/G	G/A	10	18624
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	T/A	A/A	T/T	A/T	A/A	A/A	T/T	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

Haplotype Pair[a]								PS No.[b]	PS Position[c]
15/10	5/9	11/9	11/5	11/12	16/12	11/3	11/16		
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	1	3901
T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	2	3996
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	4015
T/T	C/T	T/T	T/C	T/T	T/T	T/T	T/T	4	4019
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	5	4060
C/C	C/C	C/C	C/C	C/C	T/C	C/T	C/T	6	4148
T/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	7	9684
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	8	11281
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	11346
G/A	G/A	G/A	G/G	G/G	G/G	G/G	G/G	10	18624
G/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A	11	18646
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	12	18667
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	18979
A/A	A/A	A/A	A/A	A/T	A/T	A/T	A/A	14	26036
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	26195

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS= polymorphic site;
[c]Position of PS in SEQ ID NO:1;

and the frequency data in Tables 6 and 7.

34. A genome anthology for the alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide (ADH7) gene which comprises two or more ADH7 isogenes selected from the group consisting of isogenes 1-18 shown in the table immediately below, and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-18 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Isogene Number[a]								PS	PS	Region	
1	2	3	4	5	6	7	8	9	10	No.[b]	Pos.[c]	Examined[d]
C	C	T	T	T	T	T	T	T	T	1	3901	3090-4266
T	T	C	T	T	T	T	T	T	T	2	3996	3090-4266
G	G	G	A	G	G	G	G	G	G	3	4015	3090-4266
T	T	T	T	C	C	C	T	T	T	4	4019	3090-4266
T	T	T	T	T	T	T	C	T	T	5	4060	3090-4266
C	C	T	C	C	C	T	T	C	C	6	4148	3090-4266
C	C	C	C	C	C	C	C	C	C	7	9684	9497-9967
G	G	T	T	T	T	T	T	T	T	8	11281	11156-11699
G	G	G	G	G	T	G	G	G	G	9	11346	11156-11699
A	A	G	A	G	A	G	G	A	A	10	18624	18361-19030
A	G	A	A	A	A	A	A	A	G	11	18646	18361-19030
A	A	A	A	A	A	A	A	A	A	12	18667	18361-19030
C	C	C	C	C	C	C	C	C	C	13	18979	18361-19030
												19995-20520
												23492-24150
A	A	T	T	A	T	T	A	A	A	14	26036	25897-26300
G	G	G	G	G	G	G	G	G	G	15	26195	25897-26300

Isogene Number[a]								PS	PS	Region
11	12	13	14	15	16	17	18	No.[b]	Pos.[c]	Examined[d]
T	T	T	T	T	T	T	T	1	3901	3090-4266
T	T	T	T	T	T	T	T	2	3996	3090-4266
G	G	G	G	G	G	G	G	3	4015	3090-4266
T	T	T	T	T	T	T	T	4	4019	3090-4266
T	T	T	T	T	T	T	T	5	4060	3090-4266
C	C	C	C	C	T	T	T	6	4148	3090-4266
C	C	C	T	T	C	C	C	7	9684	9497-9967
T	T	T	T	T	T	T	T	8	11281	11156-11699
G	G	G	G	G	G	G	G	9	11346	11156-11699
G	G	G	G	G	G	G	G	10	18624	18361-19030
A	A	A	G	G	A	A	A	11	18646	18361-19030
A	A	A	A	A	A	A	G	12	18667	18361-19030
C	C	T	C	C	C	C	C	13	18979	18361-19030
										19995-20520
										23492-24150
A	T	A	A	A	A	T	A	14	26036	25897-26300
G	G	G	A	G	G	G	G	15	26195	25897-26300

[a]Alleles for isogenes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[d]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

POLYMORPHISMS IN THE ADH7 GENE

```
AAACTTCAAG AGTTGAGGTG CGGTGGGGAG TCACTGAGGT TTCTTCCAGG
CTTGCCCCTG ACCACACTGC TGCAGCGTAA GGTGAGGTAC TAAGTACGTG      100
CAAAGCTCTG CAATGAAGAA TTGGTCCAAG GCCACTTCCG CTGCTACCAT
ACCATTCCAG GAGCCTGCTC TATATTTTAG TGACATGAGT CCTGAGGGGA      200
CCTTCGGAAA CAGATGCAAA GAGAGAAACT TTAACTTGAC TTAATGAGAT
ACTGAACACC TCCCAGAGCC ATGTCTGTCT TCTTCGCAAG GTTGTGCCTC      300
CCATCTTCCA GTTTTTAGGC ACGCTCCCTT TCTGATTGCT GTGTTTTGAT
GCCGGTCCTT CCCTGCCGGA AACAGCCTTT AGGCTCAGTC TCTGTAAAGA      400
ACTGCTAGGC CTTAGTGCCT TGGGGTCATT TTCAACAAAT CCTTAGTTGC
ATACATCACA AAACTGATTC TCCTGTGCTT TTTTTTTTT TTTTTGAGAT      500
AATGAGTAGG AAAAAGAAAA AAACTTGGGT TTATTTTTTT CACAATCCTT
GACCTCCTTG AAATGAAATT CCCATCAGCC AGCACACCGA TTTCATGAGC      600
TAAGTGAAAT GGAGAGTTTA TGGGAAGCAG ATGTGAAGCC CTCATCTTGT
GAAAAATCTG CTTTGGATTC TTTTCTTGAG TAGGGTGTGC ATGCAGAAGC      700
AAATTTTCTG AAACTACTAG GCACAATCTC CCATGGCAGT TCTGAAAGCT
GGCACAGTGT GCCTGAAAAA GGATTAGCAA AAGCGTGACA AATCTGGGCA      800
GAGCTGCAGA GGCCCGATCT CTGCCCTACT CGTGGCACAG CCTCATTCCC
AGTTGGCTTG GCAGAGCTGA TGGTATGAAA GGATGAGTTT CAGCTTGCAT      900
CTAGCCATAG GATTTACCAC CTGAGTTTGC AGTGTTGGAG AAAGAGAATT
TTAGAGACAA TGGCAAGAAT GAAACCCCAC CCTCAAGGTT TGTAGTTTAC     1000
TAAAGTGGAG TGTAAATGAG CAGCATTCAC CCTAATACAG GGTGTCTAGT
TATATCCCAC TCAGTAAGTA CAATTGCATA ACTATTTATG CTATTAAGAA     1100
TGATTTAGCA ATAAACAGAT CCTTTCTTCC CTTCTTAACC TTAGGGAAAT
ACTTATTTAA TTTTCAGTTA CTGAGTGTTT GACAGTTTAA TAAGCACTTC     1200
CGCACTTGTG CTTTACAGCA AGTCCTGAAG TGGGTATTAT TTTCTCCCTT
TGTGAAGGTT GGAGCATTTA GTTTTCTCGC CATAAGTGCC TACCAGTCCC     1300
TATGGAGCAG ATACCTAGCT CTTTTGACTG TAGTGACTCT TTCACTTCTC
TGCTGTCATT ATGAGACTGG AGTGAGAACA CTTCTGGACA GGGTCCATAA     1400
CAGAAAGCTG TTACCGAAGG GCGAAGGAGC AAGGCTGTAG GAGAGTAGTG
CTGATAACAT GACCCTGTAA GCCTCCCACA CTTACAGACT CCCACACTCA     1500
CAGGTGCTGC AGAGTATGGA CAGGGTAACC GACTAGTCTC CATTTGCCTG
GGACTCTCCT GGTTTGGGAA CTGAAAATCT TGCATCCTGG GAACCCCCTC     1600
AATTCTGTGC ACACCAATGG CTGGTCAACA CCTGTGTGGA TCATATTTTG
CCTCTTAGGT TTCATCTAAA TTTCTGGTAG CACTTCCAGG TTCAAAAGAC     1700
TTAGTCCTCA AACTCTTAGG AAATTCTGGC GGCTTCCAGG ACCTCACAGT
CGATTGCTGC AAGACAGCTT TACTGAGTCA GGACTCCGGG ACATTTGGTC     1800
TCATGTGGTT CCATATTTAA ATATTTGAGC AAACAATCCA TAATAATAAT
AATAATCCAA GCAGCCAACT GGAGTCAAGT TAATAATAAT GACATCTGCT     1900
ACTGATTATG AATCAGGCAG TATACTTCAG TGCTTTCTGA CCCTCACTAC
CAGTCCTCAG GACACCCATG CAATGTGGTT ATCATTAGTC TTATTTTATG     2000
AATATGGAAG CTGAAGTCAT ACATATTCAA GCAAGGCTCC AGGCCCTGAA
CCAAGTGACT CTAATTCCAA AGCCCATATT CCTGTAGTAC ACTTTTCAAA     2100
GTTCCTCTAA GTTCCCCTGT ATCTTGACCT GAGGAAATTA CAGTTGCATG
GGTGCTTTGA GGTAATCATT AAAGGCTTTT GTAAACTTTG GGAGTAGGAC     2200
CAGGCTCACA GTAACTAGGA GATTGAGTAC CCTGAGGGTT CTGGCCCCAG
AATTTCATAG AAATATGGTG TTTAGAAGAA AGAAGTCATC TGTTAAATTC     2300
TCTAAACCTA AAATTAAGAC TTAGCTGCTA TATATCCAAA CTGACCCATT
TTTTTTATTA CCTTCCAAGA TTCAGCTTTT GGCACTTTAT CATCCAATCA     2400
AGTTTCATGG AGACTGGATT AGGACCATAA GACTTATTCA GATATTAAGC
CTCACATGAT ACATTATTTT TTGTAAAAAA TTTTTTAGAT TAATATGATA     2500
CAAAGTTTTG CCAAGCCCTT AAGAGGGAGT TTATCAATTC AGCGCTTTTG
```

FIGURE 1A

```
CATTTATGCT GTCACCACCC ATGTTTGAAC TCATTTTCTT TTGTTCATAA      2600
GTAAACAAGT TTTTGTTTGA AAGTCGAGAC TAAGAATACC TGTCCTTGTC
CACACCTGTT TAACTTTGTC TTATAGTGCC AGCAAACATA CTGGAGTTCT      2700
TCAGCACTAC TGCTAAAGCA ATGGTGCTTA CTCACAAGCC AAATCAACTG
GCAATGGTTT TTCCTGGTGG ATTTGATTTG GTTAATGAGT TGTAGTAACA      2800
AATTCTAATT ACAAATTCTG AATTTGACAT AATATATGTT AAATTGGATG
TACATTTACT ATGTTATATT GAAAGAGAGA TGTGTGGAAA AAATGTATAC      2900
TTATTTTTGA AAAGAAACTT TTGTTTGTAT CTGAAAAATA AACTCTTAGC
AAATTTCATC TCTGGGTTGT TAAGAGTCCA GGTTCTGGAC TTCATAGCTT      3000
GAACAGCTTT GTCTGAAGAC TTAGAAAACC TGCTTCACTG GATCACTTAA
CAAAAGTAAA ATATCCTTTG TGTAAAGTTT TCTTGGAGGT GATGGTGTAC      3100
CAGTATGCTG ACCATCTGGT GTACCAGTAT GCTGACCATC TACTGTCAAA
AATAACATCT ATACTGGACA TTCTCCTATC AAATAATAGA TCCAGAAGCT      3200
TCTGGAGCCA TTGCAACTTA TATTAAGAAA TAATGGTCTG TAATGGTTAA
ACAGCTACTC AAGAAGACCT AGGTACTTGC CCAAGTAGAT GAAAACCCTC      3300
TGGCATTCGA TAGCCAGTAT GCAGGCTGGA TTTCAATTTC TCTTGAGCCC
CTGCACTGTG TTACATTGGT GCCGTAAACA AACTGTACAA CCATATGAGG      3400
CAACTCTGTG GTTAATGGGA TTTGGAGTCA GATGATAATT TCTCTATGTG
CAGACACAGA AAGTTTTACT TAACTTTCTA CACCTAAAAT ACTTAGAAAC      3500
CATAGTCATC CATATTGATA TTTTATATCA TCTCAACTTC TCTTGCCCAA
ACTTGACCTA CTGTTGCAAT TATTTTACAT TTCCTTGGCT CTGTTTTCAT      3600
TTATATTTAA TTCCAGAAAC CACATCAAGT CTTTGCAGAA TGAAGTAGAG
CATTAAGAAG TAGAGATGTA CACACGCATC TCTAAAATCA GCCATGCCTA      3700
GGCAAAGCAG CTTGCACTTA AACACCCAAT ACATTTTTCA TGATTGTGTT
GAAGTGAAGT AACCTAACCC GTTTTTATAT CCTTCAAAAT AAGGTGGATA      3800
GGAATGCTTT CAGCCCTTTT CAATAGCTTT GATTATCTTG TTTTTGTTAG
ATCCCTCCTC TTGGTTTGAT CATAGTAGTT ACTGTATTTC TTTTTATAAG      3900
TTGGTCTGCA AAGGGTAGGG CTTGCAGACC ATTGCAAAGT TGTGACGGCT
C
GTGAGTCATA TTGCTGAAGG TGGAACTCTG AAGCCAGACT ATCTATGTGA      4000
                                                    C
AGGCACAAGC TGCTGTTATA TACAACAGAG TGAACTGAGC ATCAGTCAGA
               A   C
AAAAGTCTAT GTTTGCAGAA ATACAGATCC AAGACAAAGA CAGGATGGGC      4100
      C
     [exon 1: 4095..
ACTGCTGGAA AAGTAAGTGG AACATTTCTG TCCCCTCCTC ATCATGATCT
                                                      C
                ..4112]
AATGATGTGA GGCTGATACT TAGAACTTTG AATCCATTAA AGTAATTAAA      4200
CACTTGGAGA TATTCCTTGA GGAATGAAAT GCTTGGTGAG CAGGCATACA
GTGAGGGAAA CACTGGATAT GGTGTTTCAG AGAATGTCAG TGGAAGCAGG      4300
GGTAAATTGG AAATAGTATA TCCAATGTCA GCAAGACATT GGTGGAAAAT
AGAGAAATTG ATTTATATTG GCATTATTCA TAATTTGATT TCCGGATAGC      4400
TGGGCAATTG CTCCAGGAAG CCAAATGTTT CCACAAGGAC ACTTCAATCG
TCTCAGTAGC ATGTGCTGCA CTCGCTGCAG TAGTTCAATG GGAAACCAGG      4500
GTCTCATGCC CAGATGTTGG AAGAAACGTT CCTGCCATTA AGAATGGAAA
ATGCTAATTC TGGGATGCAA AATGGCCAAG AAAGTGTCAT GAATTTTAGG      4600
TTTCTTACAA ATGTAATTTT TATCAAAATA TATATTTAAA AATTCTTTAA
AATCAAGACT ATATTATTGG TAAAGTATGT TTATCGTTCT AAAAACAACA      4700
GAGTTAGCAG AGTTGGTATT AACTTTAGAA ACAGCTTTGC TTTGTTTTTT
TTTATTTGAA AATAGGTCAG TATATTTCCT TACTTTAAAT GGGTACATAA      4800
TTGCTTAAAG GACTAAATTT TTTTGTTTGT TTTGTTTTAA CTTGAACAAT
```

FIGURE 1B

```
TTTGCATTTA GCTTGCTTAA GGTTTTGGAA AAATAAAGAT GACATGAGAA      4900
GTCTGGGTAA TGTGACAAGT CAGGAGACTC TATGATGTTA GAATTTGAAA
TAAGTTTTAC TAGAATTGGA TAGAACTGGA ATAAAACCAT CCTAATTTGG      5000
GTCAGAAATC CAGCAAAACT AGATTACTTG GGAAGTTAAT TTTGCACTAC
AGAAACTAGC AACTCTACCC ATGGATATTA ATGAATCTTA GAACTGAAAG      5100
GAATCTTAAA AATTTGCAAC TTTTACCACC CAACAGATGC TTCAGGTCTG
GTTACACTAT TTCTAGGAAG TCGATATTCA TCCACTGTAA GAACATTTCT      5200
GATGATGGAG AAAGTAGCCT CAAGGCAGCA CTTTCCAATT AATAATACAG
TATATAATAT ATAAAGAGTA TAAAAGTTAT ATGATTTTGA TATATGGTCA      5300
TGCAATAGTG TGGTTTGCCT AAGGATTATA TATTTCCCAA CTTTCTGTAG
TGTTCCAGTG TAGCTCTTGA AAGGAGTTTC CCTTTACATT GGGATTGATT      5400
TTTTCTCTTT ATAATAAATG TTTTCTTTTT CATTAAAAAA GTAAAAGATA
ACTGTATTAG TTTCCTAGGG CTTCCATAAC AAAATATCAC AGACTGGGTG      5500
GCTTAAACAG CAGAAATTTG TTTTCTTGTA CATCTGAAGG CTGCAATTCT
GAGATTAAGG TGTCATCAGG GTTGGTTTCT TCTGAGGCCT GTGTCCTTGG      5600
CTTGTGGATG ACCCTCTTCT CGCTGTCTTC ACATGATCTT CCCTTTTCCT
AATCTCCCCT TCTAATAAAG ACACCAGCCA TATTGGATCA GGTTCCACCC      5700
TGATTACCAC ATTTTAAATT AATTACTTCT TGAAAGACCC CATCTCCAAA
TACAGTCACC TTCTGAGGTA CGGGGGGTTA GGACTTCAAC ATACGAATTT      5800
GGAGGGAGAC ACAATTCATC CCATACCAAT GATTATTCTA GAAAATTTGG
AAGGTCAGAA AATATGAAGG AAAAAAATCA ATCATAATAT CACATCCCCA      5900
AACTACCATT TTTGTGTGTG TGTTTCCTTC TATTTTTCCA TCCATAGTTT
ATGTAGCTGT ACTCATACAA TATACAAATA TTTTATCCTA TTTTTCCATC      6000
TTAAACATTT TTCAATTTAT AACAATTTAT AAACATTTTT AAATGTCTAA
TTTCTGCATT TAGCACAATT TATTTCATCA TTCCTCTATT ATTGTTTATG      6100
TATTTTCCTT TGTAAATAGT GCTGTGATGA CATCTGCTAT ATAGAGAGGT
TTTTCTAAAA TTAATATTTC TTCGAGGTAG CCTTTAAAAA TGGAATTACA      6200
TGGGTAAACA TGCCACAACA TCCTTAATTA TCTTGATGTA CCATGCAAAT
TACTTTCTAA AGGGTTGACA TAGGTAAGAT GGCCCATTTT ACCATTTATC      6300
TGACAATACT GTGTATTCTC ATTAAGTTCT CATTAAGTAA AAAGTCATTT
TCAGAGCATG GTGGCTTGTG CCTGTAGTCC CAGCTAACTG GGGAGGCTGA      6400
GATGGGAAGA CCACTTGAGC CCAGGAACTA GAGTACAATC TGGGCAACAG
AGTGTGATAT TGTCTTAAAA AAATTTTTTT TTTTAATTCA TTACAGACTT      6500
GATAGGCACA GAATTGTATC TCATTTTTAT TTTTTTAAAA GGCGTTTTTA
TGAGTGAGAA CAAACATGTC ACAATCTATT CATTAATCTG TTGCAGTTTT      6600
TTCCATAGGT CCTTTGCCTG CTATGTATCT CCCAGGATGT AAGTGTGGTA
TGTGGAAGTG GGAAGGAGTG TTAGGAGTGA AGTTGTTAAC ACTTCCTCTG      6700
TAAATTTGAG TAAGCATTGT ATTTATCAAA TGTGGTAACT TTTATCTATT
TAATATTTTC CCTCAGTACT ATTTGATGTC AATGTGGAAA ATTTACTATT      6800
TCTCAACACA CAAAATTTTA ATTCTTAGAA GCTGTTTCCT TTATGACTTT
TTTTAACTTT TATTTTAGGT TCAGGGGTAC ATGTGCAGGT TTGCTACATA      6900
GGCAAATTGT GTATCATGGG GTTTGGTGTA CAAATTATTT TGTCACCCAG
ATAATAAGCA TGGTACCTAA TTGGTGGTTT TTTCATCCTT ACCCTCCCTC      7000
CACCTTCCAC CCTCAAATGG GCCCTGGTGT CCATTGTTCT TTTCTTTGTG
TCCATAGGTA CTCAATATTT AGCTTCCACT TATAAGTGAG AACATGTAGT      7100
ATTTAGTTTT CTGTTTCTGC ATTCGTTTGC TCAGGATAAT GGCCTCCAGC
TCCTTCCATG TCCCTGCAAA GAATATGATC TTATTTTTTT ATGGCTGCAT      7200
AGTATTCTGT GGCAAATATG TACCACATTT TCTTTATCTA GTCTAGCATT
AATGGGCGCT TAGGTTGATT CCATGTCTTT GCTATTGTGA ATAGTGCTTC      7300
CGTGAACATA CGCATGCATA TGTCTTTATG GTAGGATGAT TTATATTTCT
TTGGGTATAT ACCCAATAAT GGGATTGCTG GGTCAAATGT TAATCCTGTT      7400
TTATGTTCTT TGAGAAATGG CCAGACTGCT TTCCACAGTG GCTGAACTAA
TTAACCTTCC CACCATCAGT GCATAAGTAT TCCCTTTTCG CTCCAACCTC      7500
```

FIGURE 1C

```
ACCAGTTTGT TTTTTGACTT TTTAGCAATA GTCATTCTGA CTGATATGAG
ATAGTACCTC ATTGTGGTTT TGATTTGCAT TTCTCTAATG ATTAGTGATG      7600
TTGAGTATTT TTTCATATGC TTGTTGGCCA TAAGTATGTT TTGTTTTGAA
AAATGTCTGT TCATATCCTT TGCCCAGTTT TTTATGGCGT TGTTTGTGTT      7700
TTGCTTATTA TTTTGTTTAA GTTCCTTATA GATGCCGGAT ATTAGACCTT
TGTTAGATGC ATAGTTTGCA AATATTTTCT CCACTTCTAT AGGTTGTCTG      7800
TGTACTCTGT TGATAGTTTC TCTTGCTGTG CAGAAGCCCT TTAGTTAAAT
TAGGCCCCTT TTGTCAATTT TTGTTTTTGT TGCCATTGCT TTTGGCATCT      7900
TTGTCATTAA GTCTTTACCA GGTCCTATGT TCAGAATGGT ATTCCCTAAG
TTATCTTCCA GGGTTTTTAG AGTTTTGGGT TTTAGATTTA AGTCACATCT      8000
TAAGTTAACT TTTATGTATG ATATAAGGAA GGGGTCTAGC TTCAATCTTC
TGCATATGGC TAGCCAGTTA CCTCAGCACC ATTTATTAAG TAGAGCATTC      8100
TTTCCCCATG GCTTGTCCTT TGTTATTTTT TCTATCACTC TTAAGCATAA
AAGTTACTCC ATGCTCAGTA TGATCAATAT TTACATCTTC ACTGATTGAT      8200
AAAATAGTTT GACTTTAAAA ATAAATCTTG GGCCAGGCAC AGTGGCTCAC
ACCTGTAATC CCAGCTCTCT GGGAGGCCGA GGAGGGCAGA TAACAAGGTC      8300
AGGAGATCGA GACTATGCTG GCCAACATGG TGAAACCCTG TCTCTACTAA
AATACAAAAA CATATTAGCT GGGCGTGGTG GCACACACCT GTAGTCTCAG      8400
CTACTCGGGT GGCTGAGGCA GGGGACTCAC TGAACCTGGG AGGCAGAGGT
TGCAGTGAGC CCAGATTGCG CCACTACACT CCAGCCTAGC GACAGAGCAA      8500
GATTGCGTCT AAAAAAAAAT AATAAAAAAA TAAAAATCTT AACAATGAAT
CCAGCTTGAA TTTGTTTGGG AATGTGGTGT GAGATTTCTA ACGTCAACAT      8600
TGGGCCATAT TACACTGTAA CCCATAACAT TTTCATTAGT AAGGCAGAGG
GTAAATATTA AATTGAGCTA CTGATGGCTT TCAGCCGCCC ATCTGTTACC      8700
AGTGTATCAT TTTTCACATT TAACCTTTTT ACAAAAGAGC CACTTAAGAA
TCCATAAGAG CTCTCTAGGA CTGTTGTGTT TGTTGACAAG TTGAAGGTAA      8800
CTTATTTTGC ATCTGAACAA ATTAATTAAA ATATTTTATA GGAGTCCTAG
ACAGATTGGG AGGTAGGAAG GTGGAAACAA TTTCTCTATG TGTGTCTGTC      8900
TCTGTCTCTC TCTCTCTGTC ATAATTCAAT AATCTTAGCA AAACCATTTA
ATTGAATAAC TACAGCAAAA CCAGTTGAAC AACTGGATGC CAGTCACTGG      9000
TATCTCCACA CCTAGTGTCC ACTACTATAA GATGACAAGC TTTAACCTGC
CTATCAAGCC CTACCAGGTG CAAGAACTGG GCTAGGTGCA AATGGGATGC      9100
AGAGATGCAG ACACAAATTC AATCATCCAG GAACTCACAG TCATAATCTG
TATGCTTCAT GATGTTCATG GTTCTCTTCA TCCTTACCAT GCTAATAGTC      9200
TAATTTCTTC ACCATTGCTA TGTGATTATA TGCAGCTATT AAATACTGGT
TTAAATTCTT ACAGACAAGT CAGCAGCATT TTCTGTTTTA TAAATTTGTC      9300
TCTTGTGAAT AAGCGATACA CAAAGGACAA TTTTGTTAAT CTATTAGATT
TGAAATTGTA ACAAAATAAC CTATTTTTGG GTGGGGGTGT TATGTGAACA      9400
GCAAACAAAA ATCAATAAGT TGGGAAAAGT ACCTCAGGTC TAGTCCAGCT
CTGCATGTCA TGTGCTAGGT GACCATGGAT CCATCATTTA GTCACTCTGG      9500
GCTGCATTTC TCATGCCTTA AGCAAAGGCA CACACAAGAT GATTTATAAG
GCTTATAATA CTAGTGTTTG ATGAAAATCA TTCTGGAAAA CTTAAAATAT      9600
TATATATGCT CTTCTGTTCA TATTTATACT AGTCAGGGAG TCTGTAAGTT
AAAGGCATAA TTCAGACCCA TTATAATACT TGCCTATTTC TTAGGTTATT      9700
                                        T
     [exon 2: 9695..
AAATGCAAAG CAGCTGTGCT TTGGGAGCAG AAGCAACCCT TCTCCATTGA
GGAAATAGAA GTTGCCCCAC CAAAGACTAA AGAAGTTCGC ATTAAGGTAA      9800
               ..9796]
GCGTGAGCCC TAGAGAACTT AAGCCAAAAG CGTTATCAAA CTTATATTGA
ATATAGCACT GTTGTGAGGC TGGATTCTTA AATAAGGAGA TGCTTCCCAT      9900
GTTTGGAGTA TTAATTCCTT TACTGGCTCT TATATATTGT TAAGGATCTG
GTTGTACTCG ATGTGAAATT AATGAGTTCA TTATCCTTTT AAAAAACAAC     10000
```

FIGURE 1D

```
AGTGCTAACA TGCTTTATTA TACCTTTGAT TTGTTGAATT TTTATTCATA
TAGTGTCTAA TTTAAAATGG TTAATAATCA TAGCTAAAGA CTTCTTTTAA      10100
AATCTTCTGG GAGTCAGATT GTTTTGAGGT TACTTTTTGC AAAGAATGAG
GAACTCACTT GTGCCTGGAG AAATGACAAC TTCGTGGAAC GTTTTCCTCC      10200
CTTTCTTTGG ACAGAGCTTA ACCTTGGCTT TGGACACCAG TTGTTGAGTA
GGCATTGCTA CCAATCTTTC TTTGACCCTC TTACTGTTAC CTGAGCACAA      10300
CAGTAGCAGG AAGATTTTGC TCTTTGCTTA CCTTCTTGCA TGCCATTTCC
TCTGTTCTTT GTTAAGAGAG CCCAATAAGG GCCTAGCTAT AAAAATCTAC      10400
TTTGGGGCCA TCACATAGTA ACATCTGTAT TCTGGTTCCC TCACAAGAGC
ACCAATTTAC TACAGACAGG CACCAAGATT CAACCTAGGG TATAAGGGAG      10500
CAATATTATA AGTTCCGTCA GAGAGGGATC TTGTCTGATG GTTGATTTTT
GGCTCAGAAT AGTCTAAGAG TAAATGTTGT TATATTAGGC AAAGATATTG      10600
ATTAAACAAT TATAATTTCT TCAAGAAGTT AAATATTCTC CCAACAGTGA
AATGATCAGT TTGTTGATTG GTGCAATGTT TTTGTCTTTG AACACAGATT      10700
     [exon 3: 10698..
TTGGCCACAG GAATCTGTCG CACAGATGAC CATGTGATAA AAGGAACAAT
GGTGTCCAAG TTTCCAGTGA TTGTGGGACA TGAGGCAACT GGGATTGTAG      10800
AGAGCATTGG AGAAGGAGTG ACTACAGTGA AACCAGGTAT ATGCAGGTGT
                ..10836]
CAAACCACAA GTTTGAAATA ATTAGGCTTT GATTAGCCTA TCAAAGGAAA      10900
TAGCACACAC TAGGAATTAT TAGAGGGATC CTTAACACAA TTCAGGGAAT
TAAAAGGATA CAAGAATGGA TTTTACTGAG TAATGTTTAA TTTGGCAAAA      11000
CCTCAACCTT TAGAAGGCAA ACCTTACGGT GTTTATAAAC CTTAGAATAT
ATTTTTAAAA GTTTTACCTA TAGTATGGGC TCAATTCACA TTTGTTAATT      11100
TCATATTTTA ACATTAATGA ACAGCATCTT ATATCATGAT TTTTTTCCTG
TAGGTGACAA AGTCATCCCT CTCTTTCTGC CACAATGTAG AGAATGCAAT      11200
     [exon 4: 11154..
GCTTGTCGCA ACCCAGATGG CAACCTTTGC ATTAGGAGCG AGTAGGTTTC
                ..11241]
AGTCATTTTT ACTTTAATGT ATTTACATTT TTCCTATGCT AATTTTTGAA      11300
                                  G
TTGAATTAAT TAATACGTGT ATTTGATGTA TCAAACAGTA TTACTGGTCG
                                            T
     [exon 5: 11339..
TGGAGTACTG GCTGATGGCA CCACCAGATT TACATGCAAG GGCAAACCAG      11400
TACACCACTT CATGAACACC AGTACATTTA CCGAGTACAC AGTGGTGGAT
GAATCTTCTG TTGCTAAGAT TGATGATGCA GCTCCTCCTG AGAAAGTCTG      11500
TTTAATTGGC TGTGGGTTTT CCACTGGATA TGGCGCTGCT GTTAAAACTG
GCAAGGTAAG AAACAGGGTA GGCTAGTTTA TTTGTTCTTC TTTCCTTTTC      11600
                ..11555]
TTAAATGATT TGCTTGGAAT CTTGAAGAGT TTTGGAAAAA AAAAAAACAT
TTAAAAGAAT GTATTGCATA TTCAGGCTGC CAAGCAACAC AATGCCAAAT      11700
CAACTAAAAG AAAAATAAAA AATTCAGATT AGAAATTTTA TTTCTTCTTT
TTACACTTGA GAATATTTAT TGACTATATT TGCAAACTGC ATCATTTCTC      11800
CATCTTTAAA ATATACCCCT TCCCCATATG TTAAGTTTGT TAAGATAGGA
CAATCTCAGT ATTTTAGTTA CAGCAGAGGC TTTTTGTCAG AGCTTTTACT      11900
CCTAGTTTCT TGTCAGCTAA GTAAGATTTA TTGTATAACT ACAAAATACA
GTAGTAAGTA CATTACCAAC TTCTCAGAAA AAAGTACAAT TTTGTCAGAT      12000
CACCTGTTAT AGATAATATA AAGAGCAATA AAATCATTAA TCATTTTGAT
TACCACCTAT AATGCGTATC AGAAATAAAA TACAGTAGTG TTTGTTGCCT      12100
ATTATCAGTA ATGCTGTTAC TGTTCACCAA AACTGATTTC TCAGTATACA
TTTTTCCATC AACTGTGATG TTAGATAGGA TCCCTGGAAT TTTAAGTTAA      12200
TCACATGAAA TTTCACAAGT AAATTGGAGT TCAAGGAATT AAGAATCTTT
```

FIGURE 1E

GACCTCTTAT TATTTTGAGA TACGTCCCAT CAATACCTAG TTTATTGAGA 12300
GTTTTTAGCA TGAAGCGTTG TTGAATTTTG TCAAAGGCCT TTTCTGCATC
TATTGAGATA ATCATGTGGT TTTTGTCTTT GGTTCTGTTT ATATGCAGAT 12400
TACATTTATT GATTTGCATA TATTGAACCA GCCTTGCATC GCAGGTATGA
AAGCCCATTT GATCATGGTG GATAAGCTTT TTGATGTGCT GTTGGATTCG 12500
GTTTGCCAGT ATTTTATTGA GGATTTTTGC ATCAATGTTC ATCAAGGATA
TTGGTCTAAA ATTCTCTTTT TTGGTTGTGT CTCTACCCAG CTTTGATATC 12600
AGGATGATGC TGGCCTCATA AAATGAGTTA GGGAGGATTC CCTCTTTTTC
TATTGATTGG AATAGTTTCA GAAGGAATGG TACCAGTTCC TCCTTGTACC 12700
TCTTGTAGAA TTCGGCTGTG AATCCATCTG GCCCCGGACT CTTTTTGGTT
GGTAAGCTAT TGATTATTGC CACAATTTCA GATCCTGTTA TTGGTCTATT 12800
CAGAGATTCA ACTTCTTCCT GGTTTAGTCT TGGGAGGGTG TATGTGTCAA
GGAATTTATC CATTTCTTCT AGATTTTCTA GTTTATTTGC GTAGAGGTGT 12900
TTGTAGTATT CTCTGATGGT AGTTTGTATT TCTGTGGGAT CGGTGGTGAT
ATCCCCTTTA TCATTTTTTA TTGCATCTAT TTGATTCTCC TCTCTTTTTT 13000
TCTTTATTAG TCTTGCTAGC GGTCTATCAA TTTTGTTGAT CCTTTCAAAA
CACCAGCTCC ATGATTCATT AATTTTTTGA AGGGTTTTTT GTGTCTCTAT 13100
TTCCTTCAGT TCTGCTGTGA TTTTAGTTAT TTCTTGCCTT CTGCTAGCTT
TTGAATGTGT TTGCTCTTGT TTTTCTAGCT CTTTTAATTG TGATGTTAGG 13200
GTGTCAATTT TAGATCTTTC CTGCTTTCTC TTGTGGGCAT TTAGTGCTGT
AAATTTCTCT CTACACACTG CTTTGAATGT GTCCCAGAGA TTCTCGAATG 13300
CTGTGTATTT GTTTTCATTG GTTTCAAAGA ACATCTTTAT TTCTGCCTTC
ATTTCGTTAT GTACCCAGTA GTCATTCAGG AGCAGGTTGT TCAGTTTCCA 13400
TGTAGTTGAG CGGTTTTGAG TGAGTTTCTT AATCCTGAGT TCTAGTTTGA
TTGCACTGTG GTCTGAGAGA CAGTTTGTTA TAATTTCTGT TCTTTTACAT 13500
TTGCTGAGGA GAGCTTTACT TCCAACTATG TGGTCAATTT TGGAATAGGT
GTGGTGTGGT GCTGAAAAAA TGTATATTCT GTTGATTTGG GCTGGAGAGT 13600
TCTGTAGATG TCTATTAGGT CCACTTGGTG TGGAGCTGAG TTCAATTCCT
GGATATCCTT GTTAACTTTC TGTCTCGTTG ATCTGTCTAA TGTTGACAGT 13700
GGGGTGTTAA AGTCTCCCAT TATTATTGTG TGGGAGTCTA AGTCTCTTTG
TAGGTCACTC AGGGCTTGCT TTATGAATCT GGGTGCTCCT GTATTGGGTG 13800
CATATATATT TAGGTTAGTT AGCTCTTCTT GTTGAATTGA TCCCTTTACC
ACTATGTAAT GGCCTTCTTT GTCTCTTTTG ATCTTTGTTG GTTTAAAGTC 13900
TGTTTTATCA GAGACTAGGA TTGCAACCCC TGCCTTTTTT TGTTTTCCAT
TTGCTTGGTA GATCTTCCTC CATCCTTTTA TTTTGAGCCT ATGTGTGTCT 14000
CTGCACGTGA GATGGGTCTC CTGAATACAG CACACTGATG GGTCTCGACT
GTTTATCGAA TTTACCAGTC TGTGTCTTTT AATTGGAGCA TTTAGTCCAT 14100
TTACATTTAA AGTTAATATT GTTATGTGTG AATTTGATCC TGTCATGATG
ATGTTAGCTG GTTATTTTGC TCATTAGTTG ATGCAGTTTC TTCCTAGTCT 14200
CGATGGTCTT TACATTTTGG CATGATTTTG CAGCAGCTGG TACCGGTTGT
TCCTTTCTAT GTTTAGTTCT TCCTTCAGGA ACTCTTTTAG GGCAGGCCTG 14300
GTGCTATCTA TGACAAATCC ACAGCCAATG TCATACTGAA TGGGCAAAAA
CTGGAAGCAT TCCCTTTGAA AAGTGGCACA AGACAGGGAT ACCCTCTCTC 14400
ACCACTCCTA TTCAACATAG TGTTGGAAGT TCTGGCCAGG GCAATTAGGC
AGGAGAAGGA AATAAAGGGT ATTCAATTAG GAAAAGAGGA AGTCAAATTG 14500
TCCCTGTTTG CAGATGACAT GATTGTATAT CTAGAAAACC CCATTGTCTC
AGCCCAAAAT CTCCTTAAGC TGATAAGCAA CTTCAGCAAA GTCTCAGGAC 14600
ACAAAATCAA TGTACAAAAA TCACAAGCAT TCTTACACAC CAATAACAAA
CAGAGAGCCA AATCATGAGT GAACTCCCAT TCACAATTGC TTCAAAGAGA 14700
ATGAAACACC TAGGAATCCA ACTTACAAGG GATGTGAAGG ACCTCTTCAA
GGAGAACTAC AAACCACTGC TCAAGGAAAT AAAAGAGGAT ACAAAGAAAT 14800
GGAAGAACAT TCCATGCTCA TGGGTAGGAA GAATCAATAT CATGAAAATG
GCCATACTGC CCAAGGTAAT TTATAGATTC AATGCCATCC CCATCAAGCT 14900

FIGURE 1F

ACCAATGACT TTCTTCACAG AATTGGAAAA AACTACTTTC AAGTTCATAT
GGAACCAAAA AAGAGCATGC ATCGCCAAGT CAATCCTAAG CCAAAAGAAC 15000
AAAGCTGGAG GCATCACGCT ATGTGACTTC AAACTATACT ACAAGGCTAC
AGTAACAAAA ACAGCATGGT ACTGGTACCA AAACAGAGAT GTAGATCAAT 15100
GGAACAGATC AGAGCCCTCA GAAATAACGC CACATATCTA CAACTATCTG
ATCTTTGACA AACCTGAGAA AAACAAGAAA TGGGGAAGGG ATTCCCTATT 15200
TAATAAATGG TGCTGGGAAA ACTGGCTAGC CATATGTAGA AAGCTGAAAC
TGGATCCCTT CCCTACACCT TATACAAAAA TTAATTCAAG ATGGATTAAA 15300
GACTTAAACA TTAGACCTAA AACCATAAAA ACCCTAGAAG AAAACCTAGG
CATTACCATT CAGGACATAG GCATGGGCAA GGACTTCATG TCTAAAACAC 15400
CAAAAGCAAT GGCAACAAAA GCCAAAATTG ACAAATGGGA TCTAATTAAA
CTAAAGAGCT TCTGCACAGC AAAAGAAACT ACCATCAGAG TGAACAGACA 15500
ACCTATAAAA TGGGGGAAAA TTTTTGCAAC CTACTCATCT GACAAAGGGC
TAATATCCAG AATCTACAAT GAACTCAAAC AAATTTACAA GAAAAAAACA 15600
AACAACCCCA TCAAAAAGTG GACAAAGGAC AGGAACAGAC ACTTCTGAAA
AGAAGACATT TATGCAGCCA AAAAACATGT GGAAAAATGC TCACCATCAC 15700
TGGCCATCAG AGAAATGCAA ATCAAAACCA CAATGAGATA CCATCTCACA
CCAGTTAGAA TGGCAATCAT TAAAAAGTCA GGAAACAACA AGTGCTGGAG 15800
AGGATGTGGA GAAATAGGAA CACTTTTACA CTGTTGGTGG GACTGTAAAC
TAGTTCAACC ATTGTGGAAG TCAGTGTGGC GATTCCTCAG GGATCTAGAG 15900
CTAGAAATAC CATTTGACCC AGCCATCCCA TTACTGGGTA TATACCCAAA
GGACTATAAA TCATGCTGCT ATAAAGACAC ATGCACACGT ATGTTTATTG 16000
CGGCACTATT CACAATAGCA AAGACTTGGA ACCAATCCAA ATGTCCAACA
ATGATAGACT GGATTAAGAA AATGTGGCAC ATATACACCA TGGAATACTA 16100
TGCAGCCATA AAAAATGATG AGTTCATGTC CTTTGTAGGG ACATGGATGA
AGCTGGAAAC CATCATTCTC AGCAAACTAT CACAAGGACA AAAAAACAAA 16200
CACCGCATGT TCTCACTCAT AGATAGGAAT TGAACAATGA GAACACATGG
ACACAGGAAG GGGAACATCA CACTCTGGAG ACTGTTGTGG GGTGGGGGGA 16300
GGGGGGAGGG ATAGCATTAG GAGATATACC TAATGCTAAA TGACGAGTTA
ATGGGTGCAG CACACCAGCA TGGCACATGT ATACATATGT AACTAACCTG 16400
CACATTGTGC ACATGTACCC TAAAACTTAA AGTATAATAA TAATAATAAA
TAAATAAATA AAATAGATCT ATTTTATTAG AAAAAAAAAT AAAATAGGTC 16500
TATTTTATTA GAAAAAAAAA AGAATCTTTG ACCTAATACA TTATTAAAAG
AAAAACCTTT ATCAAGAAAT GCAATTGAAA ATGAAGTTTC TAGATTTAAT 16600
TTTTGTTTCT TCCACATCTC CTTCTGATTC AAATTTCTAA AATTTATATT
TGTATCTAAA TTTATTTACA AATGACTCAG AGTTATGATT TTTTATTTCT 16700
TCTATTAAAA AATGGGATAT ATGTGCAGAG CGTGCAGGTT TGTTACACAG
GTATACGTGC GCCATGGTGG TTTGCTGCAT CTGTTGACTC GTCCTCCAGG 16800
TTCCCTCCCC TCACCCCCCA CCTCCTAACA GGCCCTGATG TGTGTTGTTC
CCCTCTCTGT GTCCATGTGT TCTCATTGTT CAACTCCCAC TTATGAGTGA 16900
GAACATGTGG TGTTTGGTTT TCTGTTCTTG TGTTAGTTTG CTGAGAATGA
TGGCTTCCAG CTTCTTCCAC GTCCCTGCAA AGGACATGAT CTCATTTCTT 17000
TTTATGGCTG CATAATATTC CATGGTGTAT ATGTACTACA TTTTCTTTAT
CCAGTATATC ATTGATGGCC ATTTGGGTTA GTTCCAAGTC TTTGCTATTG 17100
TAAATAGTGC TGCAATAAAC ATACGTGTGC ATGTGTCTTT GCAGTGGAAT
GATTTATATT CCTTTGGGTG TATACCCAGT AATGGGATGG CTGGGTCAAA 17200
TGATATTTCT GGTTCTAGAT CCTTGAGGAA TTGCCATACT GTCTTCCACA
ATGGTTGAAC TAATTTATAT TCTCACCAAC AGTGTAAAAA GCGTTCTTAT 17300
TTCTCCATAG CATCACGAGC ATCGATTGTT TCCTGACTTT TTAATAATCG
CCATTCTGAC TGGCATGAGA TGGTATCACA CGGTGGTTTT GATTTGCATT 17400
TCTCTAATGA TCAGTGATGT TGAGCTTTTT TCATATGTTT ATTGGCCACA
TAAATGTCTT CTTTTGAGAA GTGTCTGTTC ATATCCTGTG ACCACTTTTG 17500
GATGGGGTTC TTCATTTTTT TTCTTGTATA TTTGTTTAAG TTCCTTGTAA

FIGURE 1G

```
ATTCTGCATA TTAGACCTTT GTCAGATGAG TACATCACAA AAATCTCCTC      17600
CCATTCTGTA GGTTGCCTGT TCACTCTGAC GATAGTTTCT TTTGCTGTGC
AGAAGCTCTT TAGTTTAATT AGATCCCATT TGTCAAGTTT GGCTTTTGTT      17700
GCAATTGCTT TTGGTGTTTT TGTCATGAAG TCTTTGCCTA TGTCTAAGTT
CTGAATGGTA TTGCCTAGGT TTTCTTCTAG GGTTTTTATG GTTTGGGGTT      17800
TTACATTTAA GTGTTTAATC CATCTTCAGT TAATTTTTGT GTAAGATGTA
AGGAAGGGGT CCAGTCTGAG TTTTCTGCAT ATGGCTAGCC AGTTTTCCTA      17900
GCATCATTTA TTAAATAGGG AATCCTTTCC CCATTGTTTG TTTTTGTCAG
ATTTGTTGAA GAGCAAATGA TTGTAGATGT GTGGTGTTAT TTCTGAGGCC      18000
TCTGTTCTGT TCCATTGGTC TATATATCTG TTTTGGTACC AGTACCATGC
TGTTTTTGTT ACTGTAGCCT TGTAGTATAG TTTGAAGGCT GGTAGCATGA      18100
TGCCTCCAGC TTTGTTTTTT TGGCTTAGGA TTGTCTTGGC TATATGGGCT
CTTTTTTGGT TCTAGATGAA ATTTAAAGTA GTTTTTTTCT AATTCTGTGT      18200
AGAAAGTAAA TGGTAGCTTG ATAGGGATAA CATTGAATCT ATAAATTACT
TTGGGCAGTA TGGCCATTTT CACGATATTG ATTCTTCCTA TGCATGAGCA      18300
TGGAATTTTT CCATTTGTTT GTGTCCTCTC TTATTTCCTT GAGCAGTGGT
TTGTAATTCT CCTTGAAGAG GTCCTTCACA TCCCTTGTAA GTTGTATTCC      18400
TAGATATTTT ATTCTCTTTG TAGCAATTGT GAATGGGAGT TCACTCATGA
TTTGTGTTTT TAACTGGAGG CCCTTTTCAG GTTTCACTTT TTGACCCTAA      18500
CACCTAACAT GTTCAAGAAC ATTCCTCTCC ACAGGTCAAA CCTGGTTCCA
     [exon 6: 18535..
CTTGCGTCGT CTTTGGCCTG GGAGGAGTTG GCCTGTCAGT CATCATGGGC      18600
TGTAAGTCAG CTGGTGCATC TAGGATCATT GGGATTGACC TCAACAAAGA
                     A                       G
CAAATTTGAG AAGGCCATGG CTGTAGGTGC CACTGAGTGT ATCAGTCCCA      18700
               G
AGGACTCTAC CAAACCCATC AGTGAGGTGC TGTCAGAAAT GACAGGCAAC
AACGTGGGAT ACACCTTTGA AGTTATTGGG CATCTTGAAA CCATGGTAAG      18800
              ..18795]
ACCCAAAATT TGAGAAGCCA CAAAATACCC AAGTTATTAG ATATGCACAA
GGTAAATTTA ATTGTCTATA ACTTTATTAA TAGTCAATAA TGGAAATATC      18900
TCTGCGATAG TCAGTTTTAC TCAGGCAAAA ACTGATAATC ACATAGCAAT
TTTGCCTAAA AAGTCAAACA CTTAAGAGCA GGTTCACTGG ACCAAAGTTA      19000
                        T
AATACCATAT CATACCACTC TCAGCTGACT AGATGGCAAG GTAGTCAACC
CTTCTAGATA ATTGTTTTAA GTAGGATCTT CCTTGTATAT GTAGATTTCT      19100
TGGCGAAAGC AATTCTGTCT CTTGAATGTC AGCAAAGTAA GCATCAGAAG
GTGACTGAGG AGAGGTGAAA AGGGATTATT CCAAATGTTG TGGTCTATTC      19200
CCCCAAATCT GTGATCCATC GACCTTCTCA CTCTTTACAT CTGAGCTTTG
TTTGCCCTCA TCTGAATTCC CTTAGAATAT GAACCTTCAA TTTTCTAACC      19300
ATGCCACCTG TATGGCAGCC CGCAGCACCA GACCTTCAGT TAGACCTGTA
AAACAGATCT GAGATTCCTG GAACTAGAAA GTTCCAACCT GGCATAGCGT      19400
AAAGAGACTT GGAAAAATGG AATAAAGCCA CTTTTGCTCA AGATTGAGTG
GCAGATTAAC TCACCTTAAT GCTTTTGGTA ATTTAAGAAT AAATAGTCCT      19500
TAATTTATCA TTGCTGTCAT ATTTCTTTCC TAGTGAATCA CTTAGATTTA
TTTCATATCA ACAAATAGCA TTCTAGGTCA TAAGTTTTTG AAGCCATATG      19600
TGATACTGGA TATATCATTA GAGCCTACTG ATTTCCAGTT GCTACTTGAT
GAAGGCTGGT CTCTAAAAAC AATGCCATCT GTTGCTCCAA TCTAGCAAGT      19700
ATTAAATTAT TGCTTAATAA CTGGATGGGT TAATGACCAG AAAATGTGAG
GTTAGTCTCA GAAATGCTCC CAAGAAAAGG GACTCTTTTG TCCCACATTT      19800
TAGTATTTTA TTTAATTTTA AGACTAAACT TAAAAAAAAA ACTACGCTAA
ATGATTTGGT ATACTTAATT CAATTTAGTT TGGTTCAAAT GGCTTTCTAA      19900
GGCTTATCTG AAGTGGTGAT ATTGGTATGA AAATATCACA AGAATTTCCC
```

FIGURE 1H

```
TTGTATTTTA GTCCATGGTA CACTAAAAAT GCCACAAACA CTGACCACAT      20000
TGATTGCTGT TTCTGCCTTA AAGGGCAAAT GTGTATAGAA ATTTAATTAA
GGACTAAAGT TGCACTGAAA TTTTTTTAAA GCAAGTAAAA AACAAGGTGG      20100
CTTAATTTAA TTTCATGACT TTGTCATAGT ACATAGGTCA TTTCAGAGCT
TATGTCACAC TGTAAGAGAC ACAGGAAGCA AACGATCTCC TCCGTTTTAA      20200
ACTCAGATTG ATGCCCTGGC ATCCTGCCAC ATGAACTATG GGACCAGCGT
     [exon 7: 20207..
GGTTGTAGGA GTTCCTCCAT CAGCCAAGAT GCTCACCTAT GACCCGATGT      20300
TGCTCTTCAC TGGACGCACA TGGAAGGGAT GTGTCTTTGG AGGTCAGGAA
              ..20342]
AGCAAAGCCT CTGGATGGGG AGTGATGGCT TTCACTCTGG TGCTTGGCAA      20400
GTGGGAGAAG CCTGTTTCCT CAGGCCTTTC TTCCAAGAAT GAGTATGAAG
TGATCTAAAA TGGAACTGTG ATCATAATAA CACAAAAGAT AATTCTCCAC      20500
TCCTCACCTT GATGTAGGAC TAATATTGGT GGACGCATTT TTAATATTAA
CATCTGCAAC TCAAGTTCCC CAAAGGCCTG AATGAGAGAC ACCAGTAGTG      20600
TTTCTATAGG GATGAATTAT GTAACAGGAA GTTCCGATTT AATCCATTAT
TATTAATATA CACAAGAATG TGTATTAAAG TGAATTTATT CAATTCATAA      20700
TAGCACCAAC AAAAATCCTG GACCCAAAGC CAAAACCAAC AACAACAACA
AAAGGGATTG TCACATTTTA TGGGAACAGT TCAGATTCAC ACTGGAAGGA      20800
CTTAGATTAT CACTCAACAA TTAATATTGT TTTAGGAGAG ATGGAGCCAG
GGCCACTCTG CTCAAGGTAG GTGACACAGT GAGTCCTTCT TTTTCCCTCC      20900
TTCACAAAGT CTCGTTTTCT AGTGCCTTCA TACAGAGTTG TACTGGGTGC
TTTGAAGGTA CTCCCAATAT TAGAGTCTGC TTTATACAAT ATAAGAGAAT      21000
AGTAACTATT CAAGTACTTG TTAAGAGTAT AAAAAGGAGA AAATGATTGT
TCTTCACTTA GAGAAGTACA TAATGTAGCG AAGGATACAA AAACAAATAT      21100
AAAAGAAAGA ATATAAAAAT AAGTGCAGCT ATAAATATTA TAAAACAAAT
TATAAGAGGA GTTTTTTAGC ATTTATTCTA TGTTAAAAGT TTTACATTAA      21200
TGTGTTTGAA GTTTATTTTA TTTCTCAGTA ACTCTCTGAG GAAGCTACTT
AATGAATAAT AAAACTGAGA CTTGGAACTA TTAAGTTTTT TGCCCCTATA      21300
GCTCTCAAGT TGAGAACTGA GATTTAAAGC TGGTCCATGG GAGAACAAAG
TTTTTCCTCT TTCTGCAACA TTACTTGGGG CATAAAAAAG CAAGAGGCTA      21400
GTTGGTATGG ATGAGGTTTT ATAGAGAAGG CTTTTTAGGG AAAATGAAAC
TTGAAGAACA AAACAATAAG AGATATAAGA AGTTGAACCT CCTTGCTTTG      21500
ACTATTTGTG GGAAGTGGGA AGATTATAGA TTATTCAGGA ATATTAGAGG
TAAAATTCAG TTATTCAGCA AATATTTGTT GAGTGTATAC TGAGTACAGG      21600
TACCATTGTG TACATAGTGT ACCATTCCAG GAGCTAGGAA CGGAGAAGAG
ACAAAACATA CCAAGTCCCT GATCTCAGGA AGTGTGAGAA TCCTCTCACT      21700
TTGGTATAGA GAGACTGACA ATAAATAAAT GAATGCAGAA TATACCATAA
GGTTACAAGT GCAAAGGATA ATAATAAAGG ATGGGTAACG GGCAGTGCCA      21800
GAGGACTGGG GTGGAGGTGG AGGTGAAAGC TATTTTATAA AGTTCAAGGA
AGGCCTCCTT GATAAGGTAA CATTGGAGCA AAAACATAAA TAAGATATTC      21900
AGAAATCTTG GGAAAGAATA TTCCAGGAAG GAGACGTAGT AAATACAACA
GGGTGTTGGC AAGAACCATA AGAACGGGGA AAAGGTCAGT ATGGCTACAA      22000
CAGAGTGAGG ACAGAAGAAC AGTCACAGCT GGGGTCAGAG GACCTTAAAT
TATATGGGGC CTTTGGGCCA TTGCAAGGTT TTAGGAGTTT ATTCTGAGAG      22100
AGAAATAAAA CCACAGAGGC TTCTGAGGGA GGTGTCACCT GATCTGACTT
CCATTTTAAC ACTCTCCTTC TGGCTGCTCT ATGGGGAATA CTCTAAATGG      22200
GGCAAAGGGC ACAAGCAGGG AGACCATGGA AGAAGTCATT ATAATGCTGG
AAGGAAAGGA TGCTGGCTTG GAGTGATGGG GCTGGTGAGA AGTAGTCAGC      22300
TCTTGGATAC ATTTTGAAGA TAGAGTCAAC AAAATTTGCC GAAATAGTTA
TTGTATGGCA AAAAGAGTAG CATCCAAGAT TTTTGGGCCT AAGCAACAGG      22400
AATCATGTCA CCATCATCAT TTATTGAGAA GGAAGACTAT AGAAGGAATA
GATTGGGGGC TAGGGTGGAA GTGGTAGAAA CCAGGAGTGG TTTGGGGCAG      22500
```

FIGURE 1I

```
GTTACACTTG TGCAGGTACC TGTGAGATAC CTGAATAAAG ATGTCTATAA
AATGTAGGCA ATTGGCTATG AGTCTGTAGT ACAGGTAGAG GTTTAGGATG      22600
GGTATAGAAA CCTAAGTATT TTAAGAGGAG TATATCGAAA GCCACAAGAC
CAGAGGAGGT CACCCTAGAA GTAAATGTAG ATAAAGAACA GAAATGGACC      22700
AAGGACCAAG CCCTAGGAAT CGTTGGAGTT AGAGGTCTGG GAGATGTAGA
GGACCCAGCA AAGATCTGTG GAAGAAGTAA CTAGTGAGCA AGGAAAGTGA      22800
CAATGATGAT GACGATGTAG TATCTAGAAG CTAAGGGAAC AGAGTGTTTC
AGAAAAGGGG AGCCACCAAC AGTATCAGTT ATACCTAACA GTTCATATAA      22900
GAGGAGGACT GAGAAATGAC CCCTGTCTGG GTCAAAATGG AGCTCAAAGA
TGACCTTGAT GATATCCATT ACAGGACAGT GCTGGGGGCA AGAGCGTAAT      23000
CAGAGTGGGC TCAGAGAGGG TGAGAGGAAA GGAACAGGAT CCAGAAACAG
TGAACTGAAG ACAATTCTTT CTGTAAAGTA AAGAAGTTTC TCTGCAATAG      23100
AAGCAAAGAT ATGGGGTGAC TGGAAGGAGA TGTGGGTATC AGAAGAGTTT
TTTTAACCTT TGTTTCTTTT TAATTTAATT TAATTTTTAT TTTACTTTAA      23200
GTTCTGGGAT ACATGTGCAG AACGTGCAGG TTCGTTACGT TGGAATACAA
GTGCCATGGT GGTTTGCTGC ACCCATCAAC CCATCATCTA GGTTTTAAGC      23300
CCCGCATGCA TTAGGTATTT GTCTTAATGC TCTCCCTCCC CTTGCCTGCC
ACCCACCGAC AGGCCCTGGT ATGTGTCATT CCCCTCCCTG TGTCCATGTG      23400
TTCTCATTGT TCAACTCCCA CTTACGAGTG AGAACGTGCG GTGTTTGGTT
TTTTGTTCCT GTGTTAGTTT GCTGAGAATG ATGGTTTCCA GCTTCATCCA      23500
TGTCCCTGCA AAGGACATGA ACTCATTCTT TTTTATGGCT GCATAACTTT
TGTTTCTTAT GCTGGGACAA ATACAGCTTA TTTGTACACT GCTGGGAATG      23600
ATCTAACACA GGGAAAGTCA TAGTGCAGGA AAGAGAAGGA ATAAATATAA
CATAATAAAA GATAAGGATT ATTTAGTAAT GTCTAAAGAG AAAATGTGTG      23700
CTTATTTGCA GGTTTGAAAA GCAGAGATGA TGTCCCAAAA CTAGTGACTG
     [exon 8: 23712..
AGTTCCTGGC AAAGAAATTT GACCTGGACC AGTTGATAAC TCATGTTTTA      23800
CCATTTAAAA AAATCAGTGA AGGATTTGAG CTGCTCAATT CAGGACAAAG
                ..23850]
GTAACTGTTT CTTATGATGA TCTTGTCTCT TCTCCACAGG CTACTTTTAA      23900
AATTTAGTAC TTATAAATGT GCCTGTTTTC TTCATATTTT ATTTCTAATT
ATTTAATCTA GCCAGATTGA TACAGCTAAT TACTTCTCTT TACCCACTGC      24000
ATTAGTGTAT TTTCATACTA CTATAAAAAA AAAAAACTGC CCGAGACTGG
GTCATTTATA AAGCAAACAG ATTCAATTGA CTCACAGTTC AGCATGGCTG      24100
GGGGCTTCAG GAAACTTATA ATCATGGTGG AAGGCAAAGA GGAAGCAAGG
CACCTTCTTC ACAGGCTGCA GGAAGGAGAA GTGTCGAGCA AAGCGGGAAG      24200
AGCCCCTTAC AAAACCATCA GATCTTGTGA AAACTCACTC ACTATCATGA
GAACAGCATG AGGGAAACTG CCCCATGATT CAATTACCTC CACCTGGTCT      24300
CTCCCTTGAC ATGTAGGGAT TATAGGGATT ATAATTCAAG ATGAGATTTG
GGTGGGGACA CAAAGTCTAA CCATATCACC AACTTAGTAT GCCATTGTAC      24400
TATATTAATC AATTTTTTAA AATCTTTTGG ATTTTGTTTA CCTAGTTGAT
TTCATTGTGG CTATTTTTAT TTCATCATAA AGCTGCTATT TATTTCAAGT      24500
AGGCCACAAA ATTTCCTTAT TTTACAGTTT TCAAAGTGCT TTCTCAAATG
TGCATTATTC AGATCCCTGT AAGCCAGGTA TTATTTTTAC CATTTTTAGA      24600
TGAAGACCAA GGTTGTGAAA GCAAATAAAT AACTGGTCAA ACCTCCTGGG
GTAGTGCCAG GATATAAACT AATTGCAAAA CTTGTGCTCT TTTGACCATA      24700
TAACAAACTT CTCTGACAGA GCTTGGCAAG CTTTGGGTAA ACATGGGTGT
GATGATTACT ACTCTTGTTG CTTCTGAGGG TTGGGCAAGG ATAGAGGTCA      24800
AGTATGAATT CATGCAAATT TTATCTACAA CATCTACCTC ATGGAGGTCT
CAAACAGCAA TCAAAAAGCC TGTCTACAAG ATGAAATGAC TTTCACTCAT      24900
TGGGCCTATA AATAATTATT GAGCTTCCAC TATGTGCCAA GGACGTGGAT
TTCATTAGTA AAAGGAACTG TATATTGAGA TCCAAATTTT TACTAATTTT      25000
TTTTTTTACC TGAGAGATTT TTCAATGTTG TGTCTTGTCT CCACATTTGG
```

FIGURE 1J

```
TCTAGCCTAC AGGATCATCA TATTATGAAC ATACATTTTT TTAAGCAAGT      25100
AAACAATGGG ATTTTGTAAG AAAAACTGTC AGAACCACTC AGTACTACTT
TACTCAGATA AGCAAAACTA ACAGCTTAGT ATGGCCTGCT TGGAAAATAA      25200
TTACTATTTG TTTTCCAAAG AAACATAAAT TCTGCTTTCA GGAATTTTGC
TTCAAATATC ACCTTAAAGC AGGGACATCC AAACTAACAT TTGTAAACCT      25300
AAACTCTTCA ATTTCCACCT ACTTCCTGCA CAGCTCTATT TGAATGTCCC
ACTTGGAAAT ATTTTTTCGC CAGTAATTTT TATCTCAGTC ATTATATTAT      25400
GCGTTCCTCT TTTATAGATG AGAAGCCATA GCATCATGTC TGACTTCTTT
TTCACTCTTG CCTACCACAT ACAAACCAAT TCTAAGCCTT GATGATGCAA      25500
CCCTTGATTT CTCCTTCAGA ATCCCTCATC TCATTGCCCA CGCTCATTGC
TTTAATTCAG TCTAATCATT TCTGACATGG ACATGGTAGT GAGATAATAG      25600
GTTTTCCACC TTCTGTATTA AAATACATTC CAGTTAGTCT TTCACTAAGC
TGCCAAGCTC AAGTTTTGCC AAAATGCCAA TGTCTTAGCA TGGTTTTGAA      25700
AGCCTTTATT ATTGAACTAA AACCTATTTT GTTTCAACAG TTAAGAATTT
GGCATTCATT CTATCACCCT CTGCAGCAGG CTTTTGGGAT AAAGTGAGGG      25800
AAGTAGGTCT CATCTACTTG CAGGATCTCT TGGGCACCCA GAAATCTAAA
CATATTGGAT TACCTTTCCT GTTCACATGT TCCTAGGCTT GTTCATGCTT      25900
TCTCCTCAAC CTGGAAACCA CTTCCCTCTT GTTGGCATAC TGAAATTATA
CTGATGCTTT CAGAATTAGC TCAGATACCC TTATCAGAGT TGCGTAACAA      26000
AGAAAACTAA TTGTTAAGGA CTTTGAGCCC AGAGGATTTC CAATAACAGA
                                  T
ATTTACTTGA AGTTTCTATG ACAGATAGAA AGGACAATTG TTTAAACTAT      26100
CCTTTCTTGA AAGATATGAA AACAAGTCAT TAAAAACTCT CATTTTACAT
TTCAGCATTC GAACGGTCCT GACGTTTTGA GATCCAAAGT GGCAGGAGGT      26200
                                                  A
     [exon 9: 26156..26180]
CTGTGTTGTC ATGGTGAACT GGAGTTTCTC TTGTGAGAGT TCCCTCATCT
GAAATCATGT ATCTGTCTCA CAAATACAAG CATAAGTAGA AGATTTGTTG      26300
AAGACATAGA ACCCTTATAA AGAATTATTA ACCTTTATAA ACATTTAAAG
TCTTGTGAGC ACCTGGGAAT TAGTATAATA ACAATGTTAA TATTTTTGAT      26400
TTACATTTTG TAAGGCTATA ATTGTATCTT TTAAGAAAAC ATACACTTGG
ATTTCTATGT TGAAATGGAG ATTTTTAAGA GTTTTAACCA GCTGCTGCAG      26500
ATATATAACT CAAAACAGAT ATAGCGTATA AAGATATAGT AAATGCATCT
CCCAGAGTAA TATTCACTTA ACACATTGAA ACTATTATTT TTTAGATTTG      26600
AATATAAATG TATTTTTTAA ACACTTGTTA TGAGTTAACT TGGATTACAT
TTTGAAATCA GTTCATTCCA TGATGCATAT TACTGGATTA GATTAAGAAA      26700
GACAGAAAAG ATTAAGGGAC GGGCACATTT TTCAACGATT AAGAATCATC
ATTACATAAC TTGGTGAAAC TGAAAAAGTA TATCATATGG GTACACAAGG      26800
CTATTTGCCA GCATATATTA ATATTTTAGA AAATATTCCT TTTGTAATAC
TGAATATAAA CATAGAGCTA GAGTCATATT ATCATACTTA TCATAATGTT      26900
CAATTTGATA CAGTAGAATT GCAAGTCCCT AAGTCCCTAT TCACTGTGCT
TAGTAGTGAC TCCATTTAAT AAAAAGTGTT TTTAGTTTTT AACAACTAAA      27000
CCGATGTATC TATATATATC TATAACATGT TAAAAATTCT TAAGAAAATT
AAAAATTATA TAAAATGTTT TGTGGTGATA TATTTTTAAA CAGGCATATG      27100
TTGTTATTAA GCTTTTATTT GTTATAAACA AAGTGTTGAA GCGCCTGAAA
ACATAAAAAT ACTTTATAAG ACATATAAGA AGTTGGGTGG TCATAGCAGA      27200
AATTGAGGCT TCCAATGTTA GGGCATAATA TAGATTGGAC ACTTCTACCA
TGGTAAATTG CAAAATGTCC AGGTATTTAT GAACCACTTT TACATGATGA      27300
ATATTTCCCA GTGAAAAGTG GCATGAACTC GAGGAGTAAA ATTTCCTTTT
TCTTTTACCA GTTTGTCAAT CTAGGCTTAA TCTAGGTTTA AAAAAAATAG      27400
CTGAAAAGTG TAAATAATTA TTTTGAAGTA TCTTCATCAT CAGCTGAGTT
TAGCTATGGC AATTCTCAAA ACACTATGAA TATTTACAGC TTATAAGAAT      27500
TATATTTTTA TGTATACATA TATATATCTA TATTTATGTC AATAAATAGA
```

FIGURE 1K

```
TACATAGATA GATGATAGAT AGATAGATAG ATAGATAGAT AGATAGATGA      27600
TAGATATAGA TAAGTATAGA AAATCTCTAT CCAAGAATCA TTTGGAGACA
TTTCCTCTGT CCATTAGCAA TATCTGTAGC TAGGATCTCA GAATCCATTA      27700
AGGAAGGAAA ACTGCCCCCC AAAAAAGCAT TTTTAACAAG AAAAATATGT
AGAATGGATT AATTCCTTTA TGTCTCCTGC CTCAAATTTT GAAAAACATG      27800
TAAATAAGGA TACTAAAAGT AATATGCAAA TATAGGTACA TCAAAAGAAA
TAAATCCAAG AAGGAATAAA CACATTTATA TTCTCAGATC ACAGCTTTTT      27900
AAATGAAATC CACTTTATGC TTAGATAAAA TTATATTTTG ACAAATATGC
TGCATATGTG ACAATTTTAT TTATTTAATA TTTATTATGA TACTTTGTGG      28000
AATCATTCCT CATGTCTGAG CTTGTCATGG CCAAATAATG AACAGAAAGA
CCTTCTACTA TCTTTCTCAG GGTAACCTGC ACATGAATAG CATTTTGAGT      28100
CATCAGAAAA CACTGGTTCT AAGAAGAGAT CAGGAAATCT GATTATTTTT
GCATATTTCT TATTTCTGGT CAATGAAACC ACAAAGCATT GTCAAACTGA      28200
GTTCTTGAAT ATAAATCAAA ATATAAGTAA GCAATAAATA GCTAGAGGAA
TGCTAAAAGA AATATCCAGC TTCTTCTTTA AAGAATAGTC TAGACTTCTA      28300
ATCATTTAAC TGATATGTGG AGAACATTCA TTGGGCAATT TTTTTTTTAA
TTTCAACCAG TATTAAGCAC TGAGCTAGGT GTTTGGTATA CAGTAATGAA      28400
TAAAACAGAG TAATGAATGT CCTTTGGAGC ATATGGTCTA GTAGTGATGG
CAGCTCCTCT CTGCTAGGAA CTGAACACTC TTCGGGACAC CCTGGCTGCA      28500
GAAAGGAGTG ACCCACTGCA GGTCTCCTCT GAGCTGTTCT ATCACTCAAT
AAAGCTCTTC TTCACCTTGC TCACTCTTCA GTTGTCTGCA TACCTCATTC      28600
TTCCAGTTCG CAGGACAAGA ACTTGGGGCC CACCAAATAG CCGGGCTAAA
AGAGCTGTAA CACCAACAGG GCTGAAACAC GCCCCTTGCT CACCATGTTG      28700
TGGGTGACAA GGAGAGAAAA GAGATGGAGA GAAGAGCTGC GACCCTTTGG
GGAGCCCAGA CCTAGGAGCT CCCTAAGCCA GGCTGTAACA CCCTCTTTGG      28800
AACTCTATGG TTCCTGATGT CTCCAAGCTT CCAGGCACCA CTGCATTCTT
GGTGTCAGCT GTGGAAGCTG CTTATGGTAT GCCTGGTCCA GCCACAGACT      28900
TGCAGGAAGC GGGTGCTCAT GCCAGCACCT GGAGCTGCCC ACCCTGCTGC
AGCCAGCATG CCTGGCTGTG TGCAGTGGCC AGACCCCATA CTTCCTCACA      29000
CACCCCTCAT CACTCCATGA CGGGCTCACC CTTGGCAGGT GTGGGACCCA
GACCAGTAGC ATGAGCTGAG CATAGCCTGC CAGGCCGAGT GGGCCGAATG      29100
AGCCCAGCAG GCCAGAGCAA AACATGAGCA AAGGCGCCAC TGGCCACAGA
GGTTTCTGGC TGGCAAAGTG ACACCCCAAG GATCCTGTAA CAGTAGGAGA      29200
TTTAATAAGT AAATAAACAT ATAAGTATAT TACTTCAATT TTTGATAAAT
GATAAATAAG GCAATGATCA AGGTGCTGTG TTAGAGAGAA AAAAGAGGGG      29300
AAACAGGAAG GTATTTCCAG CTTTCCATAA GGTACTTGAA GGCCTCTCTG
AGGAGGTAAA TTTAAGCTTA CATGGAAGGA TAAGCTAGAG CCAGTTCTGT      29400
GAAGAGAAGA GAAAATTATC TCAGGCAAAG GAAAAACATG GCCAAAGGCC
ATTAATTTGG GAAAAGCATA CCCATTTATT GTTGGGAACT ATAAAAGGGC      29500
CAGTGTAGCT GGTATTTGCT GCAAAACATT AATGATATTG TTTAGGTTCA
CGAAGCTGAA AAAGATGCTA AAACCTTGTG GACATTTTTA TACATTTACC      29600
TCAATAATCA GCTTCCTTAT GTATATTTCC CAAAGATAAT GTGATTTGAT
TAGTGCCAAA ATTGATATGA AATTATTTTA ATTTAGAGGA AGAAAAGGAC      29700
TATATAGAAA TAGCATTCCC AAGAGTACCC TCTTGCCCAC GACTTCAACT
CCAGTCTCTG CAAAGATTTC TTTCACCAGC AATTTCTTTT TTGTTTCTCA      29800
TTGCTATTTA CATCTAAAAT TATTTATATT TTTATGTAAT AAGTAACTAA
CCAAGGCCCC ATAAACTAGT TTATAGTATT CATCTTCTTC ATGCATAGGT      29900
ATTTGTCCAT TTTTTTATTC CTTACTCCCT CACAGAGTTA GATTCCCTTC
ATTAAGGAAA ATCTCGATTA CCTAGGCCAC CAGCTTTAGG TTCCAGTAGT      30000
TCTTTCCACC ATGTATCTGA CTTTGTCTGT TACCCTATAG GGTACAGAAA
AGTGACGGAT GCCTTCACTG AAAGTCCATT CCGTTTTACA TGTAAAGTGT      30100
CAGAAGATAA GAATTATGTT CTAAGTTCTT TCTCAAAACA GTACCTAGTA
TTAATAGAAC GTTTTGTACC CAATGGACA                             30179
```

FIGURE 1L

POLYMORPHISMS IN THE CODING SEQUENCE OF ADH7

```
ATGGGCACTG CTGGAAAAGT TATTAAATGC AAAGCAGCTG TGCTTTGGGA
GCAGAAGCAA CCCTTCTCCA TTGAGGAAAT AGAAGTTGCC CCACCAAAGA      100
CTAAAGAAGT TCGCATTAAG ATTTTGGCCA CAGGAATCTG TCGCACAGAT
GACCATGTGA TAAAAGGAAC AATGGTGTCC AAGTTTCCAG TGATTGTGGG      200
ACATGAGGCA ACTGGGATTG TAGAGAGCAT TGGAGAAGGA GTGACTACAG
TGAAACCAGG TGACAAAGTC ATCCCTCTCT TTCTGCCACA ATGTAGAGAA      300
TGCAATGCTT GTCGCAACCC AGATGGCAAC CTTTGCATTA GGAGCGATAT
TACTGGTCGT GGAGTACTGG CTGATGGCAC CACCAGATTT ACATGCAAGG      400
    T
GCAAACCAGT ACACCACTTC ATGAACACCA GTACATTTAC CGAGTACACA
GTGGTGGATG AATCTTCTGT TGCTAAGATT GATGATGCAG CTCCTCCTGA      500
GAAAGTCTGT TTAATTGGCT GTGGGTTTTC CACTGGATAT GGCGCTGCTG
TTAAAACTGG CAAGGTCAAA CCTGGTTCCA CTTGCGTCGT CTTTGGCCTG      600
GGAGGAGTTG GCCTGTCAGT CATCATGGGC TGTAAGTCAG CTGGTGCATC
TAGGATCATT GGGATTGACC TCAACAAAGA CAAATTTGAG AAGGCCATGG      700
  A                   G                   G
CTGTAGGTGC CACTGAGTGT ATCAGTCCCA AGGACTCTAC CAAACCCATC
AGTGAGGTGC TGTCAGAAAT GACAGGCAAC AACGTGGGAT ACACCTTTGA      800
AGTTATTGGG CATCTTGAAA CCATGATTGA TGCCCTGGCA TCCTGCCACA
TGAACTATGG GACCAGCGTG GTTGTAGGAG TTCCTCCATC AGCCAAGATG      900
CTCACCTATG ACCCGATGTT GCTCTTCACT GGACGCACAT GGAAGGGATG
TGTCTTTGGA GGTTTGAAAA GCAGAGATGA TGTCCCAAAA CTAGTGACTG     1000
AGTTCCTGGC AAAGAAATTT GACCTGGACC AGTTGATAAC TCATGTTTTA
CCATTTAAAA AAATCAGTGA AGGATTTGAG CTGCTCAATT CAGGACAAAG     1100
CATTCGAACG GTCCTGACGT TTTGA                                1125
```

FIGURE 2

ISOFORMS OF THE ADH7 PROTEIN

```
MGTAGKVIKC KAAVLWEQKQ PFSIEEIEVA PPKTKEVRIK ILATGICRTD
DHVIKGTMVS KFPVIVGHEA TGIVESIGEG VTTVKPGDKV IPLFLPQCRE      100
CNACRNPDGN LCIRSDITGR GVLADGTTRF TCKGKPVHHF MNTSTFTEYT
                  C
VVDESSVAKI DDAAPPEKVC LIGCGFSTGY GAAVKTGKVK PGSTCVVFGL      200
GGVGLSVIMG CKSAGASRII GIDLNKDKFE KAMAVGATEC ISPKDSTKPI
                        E       V
SEVLSEMTGN NVGYTFEVIG HLETMIDALA SCHMNYGTSV VVGVPPSAKM      300
LTYDPMLLFT GRTWKGCVFG GLKSRDDVPK LVTEFLAKKF DLDQLITHVL
PFKKISEGFE LLNSGQSIRT VLTF                                  374
```

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals
Bieglecki, Karyn M.
Finkel, Kevin
Kazemi, Amir
Koshy, Beena
Parks, Katie E.
Sausker, Elizabeth Ann

<120> Haplotypes of the ADH7 Gene

<130> MWH-1032PCT ADH7

<140> TBA
<141> 2001-09-19

<150> 60/233,520
<151> 2000-09-19

<160> 81

<170> PatentIn Ver. 2.1

<210> 1
<211> 30179
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (3901)
<223> PS1: polymorphic base T or C

<220>
<221> allele
<222> (3996)
<223> PS2: polymorphic base T or C

<220>
<221> allele
<222> (4015)
<223> PS3: polymorphic base G or A

<220>
<221> allele
<222> (4019)
<223> PS4: polymorphic base T or C

<220>
<221> allele
<222> (4060)
<223> PS5: polymorphic base T or C

<220>
<221> allele
<222> (4148)
<223> PS6: polymorphic base T or C

<220>
<221> allele
<222> (9684)
<223> PS7: polymorphic base C or T

<220>
<221> allele
<222> (11281)
<223> PS8: polymorphic base T or G

<220>
<221> allele
<222> (11346)
<223> PS9: polymorphic base G or T

<220>
<221> allele
<222> (18624)
<223> PS10: polymorphic base G or A

<220>
<221> allele
<222> (18646)
<223> PS11: polymorphic base A or G

<220>
<221> allele
<222> (18667)
<223> PS12: polymorphic base A or G

<220>
<221> allele
<222> (18979)
<223> PS13: polymorphic base C or T

<220>
<221> allele

<222> (26036)
<223> PS14: polymorphic base A or T

<220>
<221> allele
<222> (26195)
<223> PS15: polymorphic base G or A

<400> 1
aaacttcaag agttgaggtg cggtggggag tcactgaggt ttcttccagg cttgcccctg 60
accacactgc tgcagcgtaa ggtgaggtac taagtacgtg caaagctctg caatgaagaa 120
ttggtccaag gccacttccg ctgctaccat accattccag gagcctgctc tatattttag 180
tgacatgagt cctgagggga ccttcggaaa cagatgcaaa gagagaaact ttaacttgac 240
ttaatgagat actgaacacc tcccagagcc atgtctgtct tcttcgcaag gttgtgcctc 300
ccatcttcca gtttttaggc acgctccctt tctgattgct gtgttttgat gccggtcctt 360
ccctgccgga aacagccttt aggctcagtc tctgtaaaga actgctaggc cttagtgcct 420
tggggtcatt ttcaacaaat ccttagttgc atacatcaca aaactgattc tcctgtgctt 480
ttttttttt ttttgagat aatgagtagg aaaaagaaaa aaacttgggt ttatttttt 540
cacaatcctt gacctccttg aaatgaaatt cccatcagcc agcacaccga tttcatgagc 600
taagtgaaat ggagagttta tgggaagcag atgtgaagcc ctcatcttgt gaaaaatctg 660
ctttggattc ttttcttgag tagggtgtgc atgcagaagc aaatttttctg aaactactag 720
gcacaatctc ccatggcagt tctgaaagct ggcacagtgt gcctgaaaaa ggattagcaa 780
aagcgtgaca aatctgggca gagctgcaga ggcccgatct ctgccctact cgtggcacag 840
cctcattccc agttggcttg gcagagctga tggtatgaaa ggatgagttt cagcttgcat 900
ctagccatag gatttaccac ctgagtttgc agtgttggag aaagagaatt ttagagacaa 960
tggcaagaat gaaaccccac cctcaaggtt tgtagtttac taaagtggag tgtaaatgag 1020
cagcattcac cctaatacag ggtgtctagt tatatcccac tcagtaagta caattgcata 1080
actatttatg ctattaagaa tgatttagca ataaacagat cctttcttcc cttcttaacc 1140
ttagggaaat acttatttaa ttttcagtta ctgagtgttt gacagtttaa taagcacttc 1200
cgcacttgtg ctttacagca agtcctgaag tgggtattat tttctccctt tgtgaaggtt 1260
ggagcattta gttttctcgc cataagtgcc taccagtccc tatggagcag atacctagct 1320
cttttgactg tagtgactct ttcacttctc tgctgtcatt atgagactgg agtgagaaca 1380
cttctggaca gggtccataa cagaaagctg ttaccgaagg gcgaaggagc aaggctgtag 1440
gagagtagtg ctgataacat gaccctgtaa gcctcccaca cttacagact cccacactca 1500
caggtgctgc agagtatgga cagggtaacc gactagtctc catttgcctg ggactctcct 1560
ggtttgggaa ctgaaaatct tgcatcctgg gaacccccctc aattctgtgc acaccaatgg 1620
ctggtcaaca cctgtgtgga tcatattttg cctcttaggt ttcatctaaa tttctggtag 1680
cacttccagg ttcaaaagac ttagtcctca aactcttagg aaattctggc ggcttccagg 1740
acctcacagt cgattgctgc aagacagctt tactgagtca ggactccggg acatttggtc 1800
tcatgtggtt ccatatttaa atatttgagc aaacaatcca taataataat aataatccaa 1860
gcagccaact ggagtcaagt taataataat gacatctgct actgattatg aatcaggcag 1920
tatacttcag tgctttctga ccctcactac cagtcctcag gacacccatg caatgtggtt 1980
atcattagtc ttattttatg aatatggaag ctgaagtcat acatattcaa gcaaggctcc 2040
aggccctgaa ccaagtgact ctaattccaa agcccatatt cctgtagtac acttttcaaa 2100
gttcctctaa gttcccctgt atcttgacct gaggaaatta cagttgcatg ggtgctttga 2160
ggtaatcatt aaaggctttt gtaaactttg ggagtaggac caggctcaca gtaactagga 2220
gattgagtac cctgagggtt ctggccccag aatttcatag aaatatggtg tttagaagaa 2280
agaagtcatc tgttaaattc tctaaaccta aaattaagac ttagctgcta tatatccaaa 2340

ctgacccatt ttttttatta ccttccaaga ttcagctttt ggcactttat catccaatca 2400
agtttcatgg agactggatt aggaccataa gacttattca gatattaagc ctcacatgat 2460
acattatttt ttgtaaaaaa ttttttagat taatatgata caaagttttg ccaagccctt 2520
aagagggagt ttatcaattc agcgcttttg catttatgct gtcaccaccc atgtttgaac 2580
tcattttctt ttgttcataa gtaaacaagt ttttgtttga aagtcgagac taagaatacc 2640
tgtccttgtc cacacctgtt taactttgtc ttatagtgcc agcaaacata ctggagttct 2700
tcagcactac tgctaaagca atggtgctta ctcacaagcc aaatcaactg gcaatggttt 2760
ttcctggtgg atttgatttg gttaatgagt tgtagtaaca aattctaatt acaaattctg 2820
aatttgacat aatatatgtt aaattggatg tacatttact atgttatatt gaaagagaga 2880
tgtgtggaaa aaatgtatac ttatttttga aaagaaactt ttgtttgtat ctgaaaaata 2940
aactcttagc aaatttcatc tctgggttgt taagagtcca ggttctggac ttcatagctt 3000
gaacagcttt gtctgaagac ttagaaaacc tgcttcactg gatcacttaa caaaagtaaa 3060
atatcctttg tgtaaagttt tcttggaggt gatggtgtac cagtatgctg accatctggt 3120
gtaccagtat gctgaccatc tactgtcaaa aataacatct atactggaca ttctcctatc 3180
aaataataga tccagaagct tctggagcca ttgcaactta tattaagaaa taatggtctg 3240
taatggttaa acagctactc aagaagacct aggtacttgc ccaagtagat gaaaaccctc 3300
tggcattcga tagccagtat gcaggctgga tttcaatttc tcttgagccc ctgcactgtg 3360
ttacattggt gccgtaaaca aactgtacaa ccatatgagg caactctgtg gttaatggga 3420
tttggagtca gatgataatt tctctatgtg cagacacaga aagttttact taactttcta 3480
cacctaaaat acttagaaac catagtcatc catattgata ttttatatca tctcaacttc 3540
tcttgcccaa acttgaccta ctgttgcaat tattttacat ttccttggct ctgttttcat 3600
ttatatttaa ttccagaaac cacatcaagt ctttgcagaa tgaagtagag cattaagaag 3660
tagagatgta cacacgcatc tctaaaatca gccatgccta ggcaaagcag cttgcactta 3720
aacacccaat acatttttca tgattgtgtt gaagtgaagt aacctaaccc gtttttatat 3780
ccttcaaaat aaggtggata ggaatgcttt cagccctttt caatagcttt gattatcttg 3840
tttttgttag atccctcctc ttggtttgat catagtagtt actgtatttc tttttataag 3900
ytggtctgca aagggtaggg cttgcagacc attgcaaagt tgtgacggct gtgagtcata 3960
ttgctgaagg tggaactctg aagccagact atctaygtga aggcacaagc tgctrttaya 4020
tacaacagag tgaactgagc atcagtcaga aaaagtctay gtttgcagaa atacagatcc 4080
aagacaaaga caggatgggc actgctggaa aagtaagtgg aacatttctg tcccctcctc 4140
atcatgayct aatgatgtga ggctgatact tagaactttg aatccattaa agtaattaaa 4200
cacttggaga tattccttga ggaatgaaat gcttggtgag caggcataca gtgagggaaa 4260
cactggatat ggtgtttcag agaatgtcag tggaagcagg ggtaaattgg aaatagtata 4320
tccaatgtca gcaagacatt ggtggaaaat agagaaattg atttatattg gcattattca 4380
taatttgatt tccggatagc tgggcaattg ctccaggaag ccaaatgttt ccacaaggac 4440
acttcaatcg tctcagtagc atgtgctgca ctcgctgcag tagttcaatg ggaaaccagg 4500
gtctcatgcc cagatgttgg aagaaacgtt cctgccatta agaatggaaa atgctaattc 4560
tgggatgcaa aatggccaag aaagtgtcat gaattttagg tttcttacaa atgtaatttt 4620
tatcaaaata tatatttaaa aattctttaa aatcaagact atattattgg taaagtatgt 4680
ttatcgttct aaaaacaaca gagttagcag agttggtatt aactttagaa acagctttgc 4740
tttgtttttt tttatttgaa aataggtcag tatatttcct tactttaaat gggtacataa 4800
ttgcttaaag gactaaattt ttttgtttgt tttgttttaa cttgaacaat tttgcattta 4860
gcttgcttaa ggttttggaa aaataaagat gacatgagaa gtctgggtaa tgtgacaagt 4920
caggagactc tatgatgtta gaatttgaaa taagttttac tagaattgga tagaactgga 4980
ataaaaccat cctaatttgg gtcagaaatc cagcaaaact agattacttg ggaagttaat 5040
tttgcactac agaaactagc aactctaccc atggatatta atgaatctta gaactgaaag 5100
gaatcttaaa aatttgcaac ttttaccacc caacagatgc ttcaggtctg gttacactat 5160
ttctaggaag tcgatattca tccactgtaa gaacatttct gatgatggag aaagtagcct 5220

caaggcagca ctttccaatt aataatacag tatataatat ataaagagta taaaagttat 5280
atgattttga tatatggtca tgcaatagtg tggtttgcct aaggattata tatttcccaa 5340
ctttctgtag tgttccagtg tagctcttga aaggagtttc cctttacatt gggattgatt 5400
ttttctcttt ataataaatg ttttcttttt cattaaaaaa gtaaaagata actgtattag 5460
tttcctaggg cttccataac aaaatatcac agactgggtg gcttaaacag cagaaatttg 5520
ttttcttgta catctgaagg ctgcaattct gagattaagg tgtcatcagg gttggtttct 5580
tctgaggcct gtgtccttgg cttgtggatg accctcttct cgctgtcttc acatgatctt 5640
ccctttttcct aatctccccct tctaataaag acaccagcca tattggatca ggttccaccc 5700
tgattaccac attttaaatt aattacttct tgaaagaccc catctccaaa tacagtcacc 5760
ttctgaggta cgggggggtta ggacttcaac atacgaattt ggagggagac acaattcatc 5820
ccataccaat gattattcta gaaaatttgg aaggtcagaa aatatgaagg aaaaaaatca 5880
atcataatat cacatcccca aactaccatt tttgtgtgtg tgtttccttc tatttttcca 5940
tccatagttt atgtagctgt actcatacaa tatacaaata ttttatccta tttttccatc 6000
ttaaacattt ttcaatttat aacaatttat aaacattttt aaatgtctaa tttctgcatt 6060
tagcacaatt tatttcatca ttcctctatt attgtttatg tattttcctt tgtaaatagt 6120
gctgtgatga catctgctat atagagaggt ttttctaaaa ttaatatttc ttcgaggtag 6180
cctttaaaaa tggaattaca tgggtaaaca tgccacaaca tccttaatta tcttgatgta 6240
ccatgcaaat tactttctaa agggttgaca taggtaagat ggcccatttt accatttatc 6300
tgacaatact gtgtattctc attaagttct cattaagtaa aaagtcattt tcagagcatg 6360
gtggcttgtg cctgtagtcc cagctaactg gggaggctga gatgggaaga ccacttgagc 6420
ccaggaacta gagtacaatc tgggcaacag agtgtgatat tgtcttaaaa aaatttttt 6480
ttttaattca ttacagactt gataggcaca gaattgtatc tcatttttat ttttttaaaa 6540
ggcgttttta tgagtgagaa caaacatgtc acaatctatt cattaatctg ttgcagtttt 6600
ttccataggt cctttgcctg ctatgtatct cccaggatgt aagtgtggta tgtggaagtg 6660
ggaaggagtg ttaggagtga agttgttaac acttcctctg taaatttgag taagcattgt 6720
atttatcaaa tgtggtaact tttatctatt taatattttc cctcagtact atttgatgtc 6780
aatgtggaaa atttactatt tctcaacaca caaaatttta attcttagaa gctgtttcct 6840
ttatgacttt ttttaacttt tattttaggt tcaggggtac atgtgcaggt ttgctacata 6900
ggcaaattgt gtatcatggg gtttggtgta caaattattt tgtcacccag ataataagca 6960
tggtacctaa ttggtggttt tttcatcctt accctccctc caccttccac cctcaaatgg 7020
gccctggtgt ccattgttct tttctttgtg tccataggta ctcaatattt agcttccact 7080
tataagtgag aacatgtagt atttagtttt ctgtttctgc attcgtttgc tcaggataat 7140
ggcctccagc tccttccatg tccctgcaaa gaatatgatc ttatttttt atggctgcat 7200
agtattctgt ggcaaatatg taccacattt tctttatcta gtctagcatt aatgggcgct 7260
taggttgatt ccatgtcttt gctattgtga atagtgcttc cgtgaacata cgcatgcata 7320
tgtctttatg gtaggatgat ttatatttct ttgggtatat acccaataat gggattgctg 7380
ggtcaaatgt taatcctgtt ttatgttctt tgagaaatgg ccagactgct ttccacagtg 7440
gctgaactaa ttaaccttcc caccatcagt gcataagtat tccctttcg ctccaacctc 7500
accagtttgt tttttgactt tttagcaata gtcattctga ctgatatgag atagtacctc 7560
attgtggttt tgatttgcat ttctctaatg attagtgatg ttgagtattt tttcatatgc 7620
ttgttggcca taagtatgtt ttgttttgaa aaatgtctgt tcatatcctt tgcccagttt 7680
tttatggcgt tgtttgtgtt ttgcttatta ttttgtttaa gttccttata gatgccggat 7740
attagacctt tgttagatgc atagtttgca aatattttct ccacttctat aggttgtctg 7800
tgtactctgt tgatagtttc tcttgctgtg cagaagccct ttagttaaat taggcccctt 7860
ttgtcaattt ttgtttttgt tgccattgct tttggcatct ttgtcattaa gtctttacca 7920
ggtcctatgt tcagaatggt attccctaag ttatcttcca gggtttttag agttttgggt 7980
tttagattta agtcacatct taagttaact tttatgtatg atataaggaa ggggtctagc 8040
ttcaatcttc tgcatatggc tagccagtta cctcagcacc atttattaag tagagcattc 8100

tttccccatg gcttgtcctt tgttattttt tctatcactc ttaagcataa aagttactcc 8160
atgctcagta tgatcaatat ttacatcttc actgattgat aaaatagttt gactttaaaa 8220
ataaatcttg ggccaggcac agtggctcac acctgtaatc ccagctctct gggaggccga 8280
ggagggcaga taacaaggtc aggagatcga gactatgctg gccaacatgg tgaaaccctg 8340
tctctactaa aatacaaaaa catattagct gggcgtggtg gcacacacct gtagtctcag 8400
ctactcgggt ggctgaggca ggggactcac tgaacctggg aggcagaggt tgcagtgagc 8460
ccagattgcg ccactacact ccagcctagc gacagagcaa gattgcgtct aaaaaaaaat 8520
aataaaaaaa taaaaatctt aacaatgaat ccagcttgaa tttgtttggg aatgtggtgt 8580
gagatttcta acgtcaacat tgggccatat tacactgtaa cccataacat tttcattagt 8640
aaggcagagg gtaaatatta aattgagcta ctgatggctt tcagccgccc atctgttacc 8700
agtgtatcat ttttcacatt taacctttt acaaaagagc cacttaagaa tccataagag 8760
ctctctagga ctgttgtgtt tgttgacaag ttgaaggtaa cttattttgc atctgaacaa 8820
attaattaaa atattttata ggagtcctag acagattggg aggtaggaag gtggaaacaa 8880
tttctctatg tgtgtctgtc tctgtctctc tctctctgtc ataattcaat aatcttagca 8940
aaaccattta attgaataac tacagcaaaa ccagttgaac aactggatgc cagtcactgg 9000
tatctccaca cctagtgtcc actactataa gatgacaagc tttaacctgc ctatcaagcc 9060
ctaccaggtg caagaactgg gctaggtgca aatgggatgc agagatgcag acacaaattc 9120
aatcatccag gaactcacag tcataatctg tatgcttcat gatgttcatg gttctcttca 9180
tccttaccat gctaatagtc taatttcttc accattgcta tgtgattata tgcagctatt 9240
aaatactggt ttaaattctt acagacaagt cagcagcatt ttctgtttta taaatttgtc 9300
tcttgtgaat aagcgataca caaaggacaa ttttgttaat ctattagatt tgaaattgta 9360
acaaaataac ctatttttgg gtgggggtgt tatgtgaaca gcaaacaaaa atcaataagt 9420
tgggaaaagt acctcaggtc tagtccagct ctgcatgtca tgtgctaggt gaccatggat 9480
ccatcattta gtcactctgg gctgcatttc tcatgcctta agcaaaggca cacacaagat 9540
gatttataag gcttataata ctagtgtttg atgaaaatca ttctggaaaa cttaaaatat 9600
tatatatgct cttctgttca tatttatact agtcagggag tctgtaagtt aaaggcataa 9660
ttcagaccca ttataatact tgcytatttc ttaggttatt aaatgcaaag cagctgtgct 9720
ttgggagcag aagcaaccct tctccattga ggaaatagaa gttgccccac caaagactaa 9780
agaagttcgc attaaggtaa gcgtgagccc tagagaactt aagccaaaag cgttatcaaa 9840
cttatattga atatagcact gttgtgaggc tggattctta aataaggaga tgcttcccat 9900
gtttggagta ttaattcctt tactggctct tatatattgt taaggatctg gttgtactcg 9960
atgtgaaatt aatgagttca ttatcctttt aaaaaacaac agtgctaaca tgctttatta 10020
taccttttgat ttgttgaatt tttattcata tagtgtctaa tttaaaatgg ttaataatca 10080
tagctaaaga cttcttttaa aatcttctgg gagtcagatt gttttgaggt tacttttttgc 10140
aaagaatgag gaactcactt gtgcctggag aaatgacaac ttcgtggaac gttttcctcc 10200
ctttctttgg acagagctta accttggctt tggacaccag ttgttgagta ggcattgcta 10260
ccaatctttc tttgaccctc ttactgttac ctgagcacaa cagtagcagg aagattttgc 10320
tctttgctta ccttcttgca tgccatttcc tctgttcttt gttaagagag cccaataagg 10380
gcctagctat aaaaatctac tttggggcca tcacatagta acatctgtat tctggttccc 10440
tcacaagagc accaatttac tacagacagg caccaagatt caacctaggg tataagggag 10500
caatattata agttccgtca gagagggatc ttgtctgatg gttgattttt ggctcagaat 10560
agtctaagag taaatgttgt tatattaggc aaagatattg attaaacaat tataatttct 10620
tcaagaagtt aaatattctc ccaacagtga aatgatcagt ttgttgattg gtgcaatgtt 10680
tttgtctttg aacacagatt ttggccacag gaatctgtcg cacagatgac catgtgataa 10740
aaggaacaat ggtgtccaag tttccagtga ttgtgggaca tgaggcaact gggattgtag 10800
agagcattgg agaaggagtg actacagtga aaccaggtat atgcaggtgt caaaccacaa 10860
gtttgaaata attaggcttt gattagccta tcaaaggaaa tagcacacac taggaattat 10920
tagagggatc cttaacacaa ttcagggaat taaaaggata caagaatgga ttttactgag 10980

taatgtttaa tttggcaaaa cctcaacctt tagaaggcaa accttacggt gtttataaac 11040
cttagaatat atttttaaaa gttttaccta tagtatgggc tcaattcaca tttgttaatt 11100
tcatatttta acattaatga acagcatctt atatcatgat ttttttcctg taggtgacaa 11160
agtcatccct ctctttctgc cacaatgtag agaatgcaat gcttgtcgca acccagatgg 11220
caacctttgc attaggagcg agtaggtttc agtcattttt actttaatgt atttacattt 11280
ktcctatgct aatttttgaa ttgaattaat taatacgtgt atttgatgta tcaaacagta 11340
ttactkgtcg tggagtactg gctgatggca ccaccagatt tacatgcaag ggcaaaccag 11400
tacaccactt catgaacacc agtacattta ccgagtacac agtggtggat gaatcttctg 11460
ttgctaagat tgatgatgca gctcctcctg agaaagtctg tttaattggc tgtgggtttt 11520
ccactggata tggcgctgct gttaaaactg gcaaggtaag aaacagggta ggctagttta 11580
tttgttcttc tttccttttc ttaaatgatt tgcttggaat cttgaagagt tttggaaaaa 11640
aaaaaaacat ttaaaagaat gtattgcata ttcaggctgc caagcaacac aatgccaaat 11700
caactaaaag aaaaataaaa aattcagatt agaaatttta tttcttcttt ttacacttga 11760
gaatatttat tgactatatt tgcaaactgc atcatttctc catctttaaa atataccccct 11820
tccccatatg ttaagtttgt taagatagga caatctcagt attttagtta cagcagaggc 11880
tttttgtcag agcttttact cctagtttct tgtcagctaa gtaagattta ttgtataact 11940
acaaaatata gtagtaagta cattaccaac ttctcagaaa aaagtacaat tttgtcagat 12000
cacctgttat agataatata aagagcaata aaatcattaa tcattttgat taccacctat 12060
aatgcgtatc agaaataaaa tacagtagtg tttgttgcct attatcagta atgctgttac 12120
tgttcaccaa aactgatttc tcagtataca tttttccatc aactgtgatg ttagatagga 12180
tccctggaat tttaagttaa tcacatgaaa tttcacaagt aaattggagt tcaaggaatt 12240
aagaatcttt gacctcttat tattttgaga tacgtcccat caatacctag tttattgaga 12300
gtttttagca tgaagcgttg ttgaattttg tcaaaggcct tttctgcatc tattgagata 12360
atcatgtggt ttttgtcttt ggttctgttt atatgcagat tacatttatt gatttgcata 12420
tattgaacca gccttgcatc gcaggtatga aagcccattt gatcatggtg gataagcttt 12480
ttgatgtgct gttggattcg gtttgccagt attttattga ggatttttgc atcaatgttc 12540
atcaaggata ttggtctaaa attctctttt ttggttgtgt ctctacccag ctttgatatc 12600
aggatgatgc tggcctcata aaatgagtta gggaggattc cctctttttc tattgattgg 12660
aatagtttca gaaggaatgg taccagttcc tccttgtacc tcttgtagaa ttcggctgtg 12720
aatccatctg gccccggact ctttttggtt ggtaagctat tgattattgc cacaatttca 12780
gatcctgtta ttggtctatt cagagattca acttcttcct ggtttagtct tgggagggtg 12840
tatgtgtcaa ggaatttatc catttcttct agattttcta gtttatttgc gtagaggtgt 12900
ttgtagtatt ctctgatggt agtttgtatt tctgtgggat cggtggtgat atccccttta 12960
tcatttttta ttgcatctat ttgattctcc tctcttttct tctttattag tcttgctagc 13020
ggtctatcaa ttttgttgat cctttcaaaa caccagctcc atgattcatt aatttttga 13080
agggtttttt gtgtctctat ttccttcagt tctgctgtga ttttagttat ttcttgcctt 13140
ctgctagctt ttgaatgtgt ttgctcttgt ttttctagct cttttaattg tgatgttagg 13200
gtgtcaattt tagatctttc ctgctttctc ttgtgggcat ttagtgctgt aaatttctct 13260
ctacacactg ctttgaatgt gtcccagaga ttctcgaatg ctgtgtattt gttttcattg 13320
gtttcaaaga acatctttat ttctgccttc atttcgttat gtacccagta gtcattcagg 13380
agcaggttgt tcagtttcca tgtagttgag cggttttgag tgagtttctt aatcctgagt 13440
tctagtttga ttgcactgtg gtctgagaga cagtttgtta taatttctgt tcttttacat 13500
ttgctgagga gagctttact tccaactatg tggtcaattt tggaataggt gtggtgtggt 13560
gctgaaaaaa tgtatattct gttgatttgg gctggagagt tctgtagatg tctattaggt 13620
ccacttggtg tggagctgag ttcaattcct ggatatcctt gttaactttc tgtctcgttg 13680
atctgtctaa tgttgacagt ggggtgttaa agtctcccat tattattgtg tgggagtcta 13740
agtctctttg taggtcactc agggcttgct ttatgaatct gggtgctcct gtattgggtg 13800
catatatatt taggttagtt agctcttctt gttgaattga tccctttacc actatgtaat 13860

ggccttcttt gtctcttttg atctttgttg gtttaaagtc tgttttatca gagactagga 13920
ttgcaacccc tgccttttt tgttttccat ttgcttggta gatcttcctc catcctttta 13980
ttttgagcct atgtgtgtct ctgcacgtga gatgggtctc ctgaatacag cacactgatg 14040
ggtctcgact gtttatcgaa tttaccagtc tgtgtctttt aattggagca tttagtccat 14100
ttacatttaa agttaatatt gttatgtgtg aatttgatcc tgtcatgatg atgttagctg 14160
gttattttgc tcattagttg atgcagtttc ttcctagtct cgatggtctt tacattttgg 14220
catgattttg cagcagctgg taccggttgt tcctttctat gtttagttct tccttcagga 14280
actcttttag ggcaggcctg gtgctatcta tgacaaatcc acagccaatg tcatactgaa 14340
tgggcaaaaa ctggaagcat tccctttgaa aagtggcaca agacagggat accctctctc 14400
accactccta ttcaacatag tgttggaagt tctggccagg gcaattaggc aggagaagga 14460
aataaagggt attcaattag gaaaagagga agtcaaattg tccctgtttg cagatgacat 14520
gattgtatat ctagaaaacc ccattgtctc agcccaaaat ctccttaagc tgataagcaa 14580
cttcagcaaa gtctcaggac acaaaatcaa tgtacaaaaa tcacaagcat tcttacacac 14640
caataacaaa cagagagcca aatcatgagt gaactcccat tcacaattgc ttcaaagaga 14700
atgaaacacc taggaatcca acttacaagg gatgtgaagg acctcttcaa ggagaactac 14760
aaaccactgc tcaaggaaat aaaagaggat acaaagaaat ggaagaacat tccatgctca 14820
tgggtaggaa gaatcaatat catgaaaatg gccatactgc ccaaggtaat ttatagattc 14880
aatgccatcc ccatcaagct accaatgact ttcttcacag aattggaaaa aactactttc 14940
aagttcatat ggaaccaaaa aagagcatgc atcgccaagt caatcctaag ccaaaagaac 15000
aaagctggag gcatcacgct atgtgacttc aaactatact acaaggctac agtaacaaaa 15060
acagcatggt actggtacca aaacagagat gtagatcaat ggaacagatc agagccctca 15120
gaaataacgc cacatatcta caactatctg atctttgaca aacctgagaa aaacaagaaa 15180
tggggaaggg attccctatt taataaatgg tgctgggaaa actggctagc catatgtaga 15240
aagctgaaac tggatccctt ccctacacct tatacaaaaa ttaattcaag atggattaaa 15300
gacttaaaca ttagacctaa aaccataaaa accctagaag aaaacctagg cattaccatt 15360
caggacatag gcatgggcaa ggacttcatg tctaaaacac caaaagcaat ggcaacaaaa 15420
gccaaaattg acaaatggga tctaattaaa ctaaagagct tctgcacagc aaaagaaact 15480
accatcagag tgaacagaca acctataaaa tgggggaaaa tttttgcaac ctactcatct 15540
gacaaagggc taatatccag aatctacaat gaactcaaac aaatttacaa gaaaaaaaca 15600
aacaacccca tcaaaaagtg gacaaaggac aggaacagac acttctgaaa agaagacatt 15660
tatgcagcca aaaaacatgt ggaaaaatgc tcaccatcac tggccatcag agaaatgcaa 15720
atcaaaacca caatgagata ccatctcaca ccagttagaa tggcaatcat taaaaagtca 15780
ggaaacaaca agtgctggag aggatgtgga gaaataggaa cacttttaca ctgttggtgg 15840
gactgtaaac tagttcaacc attgtggaag tcagtgtggc gattcctcag ggatctagag 15900
ctagaaatac catttgaccc agccatccca ttactgggta tatacccaaa ggactataaa 15960
tcatgctgct ataaagacac atgcacacgt atgtttattg cggcactatt cacaatagca 16020
aagacttgga accaatccaa atgtccaaca atgatagact ggattaagaa aatgtggcac 16080
atatacacca tggaatacta tgcagccata aaaaatgatg agttcatgtc ctttgtaggg 16140
acatggatga agctggaaac catcattctc agcaaactat cacaaggaca aaaaaacaaa 16200
caccgcatgt tctcactcat agataggaat tgaacaatga gaacacatgg acacaggaag 16260
gggaacatca cactctggag actgttgtgg ggtgggggga ggggggaggg atagcattag 16320
gagatatacc taatgctaaa tgacgagtta atgggtgcag cacaccagca tggcacatgt 16380
atacatatgt aactaacctg cacattgtgc acatgtaccc taaaacttaa agtataataa 16440
taataataaa taaataaata aaatagatct attttattag aaaaaaaaat aaaataggtc 16500
tattttatta gaaaaaaaaa agaatctttg acctaataca ttattaaaag aaaaaccttt 16560
atcaagaaat gcaattgaaa atgaagtttc tagatttaat ttttgtttct tccacatctc 16620
cttctgattc aaatttctaa aatttatatt tgtatctaaa tttatttaca aatgactcag 16680
agttatgatt ttttatttct tctattaaaa aatgggatat atgtgcagag cgtgcaggtt 16740

tgttacacag gtatacgtgc gccatggtgg tttgctgcat ctgttgactc gtcctccagg 16800
ttccctcccc tcacccccca cctcctaaca ggccctgatg tgtgttgttc ccctctctgt 16860
gtccatgtgt tctcattgtt caactcccac ttatgagtga gaacatgtgg tgtttggttt 16920
tctgttcttg tgttagtttg ctgagaatga tggcttccag cttcttccac gtccctgcaa 16980
aggacatgat ctcatttctt tttatggctg cataatattc catggtgtat atgtactaca 17040
ttttctttat ccagtatatc attgatggcc atttgggtta gttccaagtc tttgctattg 17100
taaatagtgc tgcaataaac atacgtgtgc atgtgtcttt gcagtggaat gatttatatt 17160
cctttgggtg tatacccagt aatgggatgg ctgggtcaaa tgatatttct ggttctagat 17220
ccttgaggaa ttgccatact gtcttccaca atggttgaac taatttatat tctcaccaac 17280
agtgtaaaaa gcgttcttat ttctccatag catcacgagc atcgattgtt tcctgacttt 17340
ttaataatcg ccattctgac tggcatgaga tggtatcaca cggtggtttt gatttgcatt 17400
tctctaatga tcagtgatgt tgagcttttt tcatatgttt attggccaca taaatgtctt 17460
cttttgagaa gtgtctgttc atatcctgtg accacttttg gatggggttc ttcatttttt 17520
ttcttgtata tttgtttaag ttccttgtaa attctgcata ttagaccttt gtcagatgag 17580
tacatcacaa aaatctcctc ccattctgta ggttgcctgt tcactctgac gatagtttct 17640
tttgctgtgc agaagctctt tagtttaatt agatcccatt tgtcaagttt ggcttttgtt 17700
gcaattgctt ttggtgtttt tgtcatgaag tctttgccta tgtctaagtt ctgaatggta 17760
ttgcctaggt tttcttctag ggtttttatg gtttggggtt ttacatttaa gtgtttaatc 17820
catcttcagt taatttttgt gtaagatgta aggaaggggt ccagtctgag ttttctgcat 17880
atggctagcc agttttccta gcatcattta ttaaataggg aatcctttcc ccattgtttg 17940
tttttgtcag atttgttgaa gagcaaatga ttgtagatgt gtggtgttat ttctgaggcc 18000
tctgttctgt tccattggtc tatatatctg ttttggtacc agtaccatgc tgtttttgtt 18060
actgtagcct tgtagtatag tttgaaggct ggtagcatga tgcctccagc tttgtttttt 18120
tggcttagga ttgtcttggc tatatgggct ctttttggt tctagatgaa atttaaagta 18180
gtttttttct aattctgtgt agaaagtaaa tggtagcttg atagggataa cattgaatct 18240
ataaattact ttgggcagta tggccatttt cacgatattg attcttccta tgcatgagca 18300
tggaattttt ccatttgttt gtgtcctctc ttatttcctt gagcagtggt ttgtaattct 18360
ccttgaagag gtccttcaca tcccttgtaa gttgtattcc tagatatttt attctctttg 18420
tagcaattgt gaatgggagt tcactcatga tttgtgtttt taactggagg ccctttcag 18480
gtttcacttt ttgaccctaa cacctaacat gttcaagaac attcctctcc acaggtcaaa 18540
cctggttcca cttgcgtcgt ctttggcctg ggaggagttg gcctgtcagt catcatgggc 18600
tgtaagtcag ctggtgcatc tagratcatt gggattgacc tcaacraaga caaatttgag 18660
aaggccrtgg ctgtaggtgc cactgagtgt atcagtccca aggactctac caaacccatc 18720
agtgaggtgc tgtcagaaat gacaggcaac aacgtgggat acacctttga agttattggg 18780
catcttgaaa ccatggtaag acccaaaatt tgagaagcca caaaataccc aagttattag 18840
atatgcacaa ggtaaattta attgtctata actttattaa tagtcaataa tggaaatatc 18900
tctgcgatag tcagttttac tcaggcaaaa actgataatc acatagcaat tttgcctaaa 18960
aagtcaaaca cttaagagya ggttcactgg accaaagtta aataccatat cataccactc 19020
tcagctgact agatggcaag gtagtcaacc cttctagata attgttttaa gtaggatctt 19080
ccttgtatat gtagatttct tggcgaaagc aattctgtct cttgaatgtc agcaaagtaa 19140
gcatcagaag gtgactgagg agaggtgaaa agggattatt ccaaatgttg tggtctattc 19200
ccccaaatct gtgatccatc gaccttctca ctctttacat ctgagctttg tttgccctca 19260
tctgaattcc cttagaatat gaaccttcaa ttttctaacc atgccacctg tatggcagcc 19320
cgcagcacca gaccttcagt tagacctgta aaacagatct gagattcctg gaactagaaa 19380
gttccaacct ggcatagcgt aaagagactt ggaaaaatgg aataaagcca cttttgctca 19440
agattgagtg gcagattaac tcaccttaat gcttttggta atttaagaat aaatagtcct 19500
taatttatca ttgctgtcat atttctttcc tagtgaatca cttagattta tttcatatca 19560
acaaatagca ttctaggtca taagtttttg aagccatatg tgatactgga tatatcatta 19620

gagcctactg atttccagtt gctacttgat gaaggctggt ctctaaaaac aatgccatct 19680
gttgctccaa tctagcaagt attaaattat tgcttaataa ctggatgggt taatgaccag 19740
aaaatgtgag gttagtctca gaaatgctcc caagaaaagg gactcttttg tcccacattt 19800
tagtatttta tttaatttta agactaaact taaaaaaaaa actacgctaa atgatttggt 19860
atacttaatt caatttagtt tggttcaaat ggctttctaa ggcttatctg aagtggtgat 19920
attggtatga aaatatcaca agaatttccc ttgtatttta gtccatggta cactaaaaat 19980
gccacaaaca ctgaccacat tgattgctgt ttctgcctta aagggcaaat gtgtatagaa 20040
atttaattaa ggactaaagt tgcactgaaa ttttttaaaa gcaagtaaaa aacaaggtgg 20100
cttaatttaa tttcatgact ttgtcatagt acataggtca tttcagagct tatgtcacac 20160
tgtaagagac acaggaagca aacgatctcc tccgttttaa actcagattg atgccctggc 20220
atcctgccac atgaactatg ggaccagcgt ggttgtagga gttcctccat cagccaagat 20280
gctcacctat gacccgatgt tgctcttcac tggacgcaca tggaagggat gtgtctttgg 20340
aggtcaggaa agcaaagcct ctggatgggg agtgatggct ttcactctgg tgcttggcaa 20400
gtgggagaag cctgtttcct caggcctttc ttccaagaat gagtatgaag tgatctaaaa 20460
tggaactgtg atcataataa cacaaaagat aattctccac tcctcacctt gatgtaggac 20520
taatattggt ggacgcattt ttaatattaa catctgcaac tcaagttccc caaaggcctg 20580
aatgagagac accagtagtg tttctatagg gatgaattat gtaacaggaa gttccgattt 20640
aatccattat tattaatata cacaagaatg tgtattaaag tgaatttatt caattcataa 20700
tagcaccaac aaaaatcctg gacccaaagc caaaaccaac aacaacaaca aaagggattg 20760
tcacatttta tgggaacagt tcagattcac actggaagga cttagattat cactcaacaa 20820
ttaatattgt tttaggagag atggagccag ggccactctg ctcaaggtag gtgacacagt 20880
gagtccttct ttttccctcc ttcacaaagt ctcgttttct agtgccttca tacagagttg 20940
tactgggtgc tttgaaggta ctcccaatat tagagtctgc tttatacaat ataagagaat 21000
agtaactatt caagtacttg ttaagagtat aaaaaggaga aaatgattgt tcttcactta 21060
gagaagtaca taatgtagcg aaggatacaa aaacaaatat aaaagaaaga atataaaaat 21120
aagtgcagct ataaatatta taaaacaaat tataagagga gttttttagc atttattcta 21180
tgttaaaagt tttacattaa tgtgtttgaa gtttatttta tttctcagta actctctgag 21240
gaagctactt aatgaataat aaaactgaga cttggaacta ttaagttttt tgcccctata 21300
gctctcaagt tgagaactga gatttaaagc tggtccatgg gagaacaaag tttttcctct 21360
ttctgcaaca ttacttgggg cataaaaaag caagaggcta gttggtatgg atgaggtttt 21420
atagagaagg ctttttaggg aaaatgaaac ttgaagaaca aaacaataag agatataaga 21480
agttgaacct ccttgctttg actatttgtg ggaagtggga agattataga ttattcagga 21540
atattagagg taaaattcag ttattcagca aatatttgtt gagtgtatac tgagtacagg 21600
taccattgtg tacatagtgt accattccag gagctaggaa cggagaagag acaaaacata 21660
ccaagtccct gatctcagga agtgtgagaa tcctctcact ttggtataga gagactgaca 21720
ataaataaat gaatgcagaa tataccataa ggttacaagt gcaaaggata ataataaagg 21780
atgggtaacg ggcagtgcca gaggactggg gtggaggtgg aggtgaaagc tattttataa 21840
agttcaagga aggcctcctt gataaggtaa cattggagca aaaacataaa taagatattc 21900
agaaatcttg ggaaagaata ttccaggaag gagacgtagt aaatacaaca gggtgttggc 21960
aagaaccata agaacgggga aaaggtcagt atggctacaa cagagtgagg acagaagaac 22020
agtcacagct ggggtcagag gaccttaaat tatatggggc ctttgggcca ttgcaaggtt 22080
ttaggagttt attctgagag agaaataaaa ccacagaggc ttctgaggga ggtgtcacct 22140
gatctgactt ccattttaac actctccttc tggctgctct atggggaata ctctaaatgg 22200
ggcaaagggc acaagcaggg agaccatgga agaagtcatt ataatgctgg aaggaaagga 22260
tgctggcttg gagtgatggg gctggtgaga agtagtcagc tcttggatac attttgaaga 22320
tagagtcaac aaaatttgcc gaaatagtta ttgtatggca aaaagagtag catccaagat 22380
ttttgggcct aagcaacagg aatcatgtca ccatcatcat ttattgagaa ggaagactat 22440
agaaggaata gattgggggc tagggtggaa gtggtagaaa ccaggagtgg tttggggcag 22500

gttacacttg tgcaggtacc tgtgagatac ctgaataaag atgtctataa aatgtaggca 22560
attggctatg agtctgtagt acaggtagag gtttaggatg ggtatagaaa cctaagtatt 22620
ttaagaggag tatatcgaaa gccacaagac cagaggaggt caccctagaa gtaaatgtag 22680
ataaagaaca gaaatggacc aaggaccaag ccctaggaat cgttggagtt agaggtctgg 22740
gagatgtaga ggacccagca aagatctgtg gaagaagtaa ctagtgagca aggaaagtga 22800
caatgatgat gacgatgtag tatctagaag ctaagggaac agagtgtttc agaaaagggg 22860
agccaccaac agtatcagtt atacctaaca gttcatataa gaggaggact gagaaatgac 22920
ccctgtctgg gtcaaaatgg agctcaaaga tgaccttgat gatatccatt acaggacagt 22980
gctgggggca agagcgtaat cagagtgggc tcagagaggg tgagaggaaa ggaacaggat 23040
ccagaaacag tgaactgaag acaattcttt ctgtaaagta aagaagtttc tctgcaatag 23100
aagcaaagat atggggtgac tggaaggaga tgtgggtatc agaagagttt ttttaacctt 23160
tgtttctttt taatttaatt taatttttat tttactttaa gttctgggat acatgtgcag 23220
aacgtgcagg ttcgttacgt tggaatacaa gtgccatggt ggtttgctgc acccatcaac 23280
ccatcatcta ggtttttaagc cccgcatgca ttaggtattt gtcttaatgc tctccctccc 23340
cttgcctgcc acccaccgac aggccctggt atgtgtcatt cccctccctg tgtccatgtg 23400
ttctcattgt tcaactccca cttacgagtg agaacgtgcg gtgtttggtt ttttgttcct 23460
gtgttagttt gctgagaatg atggtttcca gcttcatcca tgtccctgca aaggacatga 23520
actcattctt ttttatggct gcataacttt tgtttcttat gctgggacaa atacagctta 23580
tttgtacact gctgggaatg atctaacaca gggaaagtca tagtgcagga aagagaagga 23640
ataaatataa cataataaaa gataaggatt atttagtaat gtctaaagag aaaatgtgtg 23700
cttatttgca ggtttgaaaa gcagagatga tgtcccaaaa ctagtgactg agttcctggc 23760
aaagaaattt gacctggacc agttgataac tcatgtttta ccatttaaaa aaatcagtga 23820
aggatttgag ctgctcaatt caggacaaag gtaactgttt cttatgatga tcttgtctct 23880
tctccacagg ctacttttaa aatttagtac ttataaatgt gcctgttttc ttcatatttt 23940
atttctaatt atttaatcta gccagattga tacagctaat tacttctctt tacccactgc 24000
attagtgtat tttcatacta ctataaaaaa aaaaaactgc ccgagactgg gtcatttata 24060
aagcaaacag attcaattga ctcacagttc agcatggctg ggggcttcag gaaacttata 24120
atcatggtgg aaggcaaaga ggaagcaagg caccttcttc acaggctgca ggaaggagaa 24180
gtgtcgagca aagcgggaag agccccttac aaaaccatca gatcttgtga aaactcactc 24240
actatcatga gaacagcatg agggaaactg ccccatgatt caattacctc cacctggtct 24300
ctcccttgac atgtagggat tatagggatt ataattcaag atgagatttg ggtggggaca 24360
caaagtctaa ccatatcacc aacttagtat gccattgtac tatattaatc aatttttaa 24420
aatcttttgg attttgttta cctagttgat ttcattgtgg ctatttttat ttcatcataa 24480
agctgctatt tatttcaagt aggccacaaa atttccttat tttacagttt tcaaagtgct 24540
ttctcaaatg tgcattattc agatccctgt aagccaggta ttatttttac catttttaga 24600
tgaagaccaa ggttgtgaaa gcaaataaat aactggtcaa acctcctggg gtagtgccag 24660
gatataaact aattgcaaaa cttgtgctct tttgaccata taacaaactt ctctgacaga 24720
gcttggcaag ctttgggtaa acatgggtgt gatgattact actcttgttg cttctgaggg 24780
ttgggcaagg atagaggtca agtatgaatt catgcaaatt ttatctacaa catctacctc 24840
atggaggtct caaacagcaa tcaaaaagcc tgtctacaag atgaaatgac tttcactcat 24900
tgggcctata aataattatt gagcttccac tatgtgccaa ggacgtggat ttcattagta 24960
aaaggaactg tatattgaga tccaaatttt tactaatttt ttttttacc tgagagattt 25020
ttcaatgttg tgtcttgtct ccacatttgg tctagcctac aggatcatca tattatgaac 25080
atacattttt ttaagcaagt aaacaatggg attttgtaag aaaaactgtc agaaccactc 25140
agtactactt tactcagata agcaaaacta acagcttagt atggcctgct tggaaaataa 25200
ttactatttg ttttccaaag aaacataaat tctgctttca ggaattttgc ttcaaatatc 25260
accttaaagc agggacatcc aaactaacat ttgtaaacct aaactcttca atttccacct 25320
acttcctgca cagctctatt tgaatgtccc acttggaaat atttttcgc cagtaatttt 25380

tatctcagtc attatattat gcgttcctct tttatagatg agaagccata gcatcatgtc 25440
tgacttcttt ttcactcttg cctaccacat acaaaccaat tctaagcctt gatgatgcaa 25500
cccttgattt ctccttcaga atccctcatc tcattgccca cgctcattgc tttaattcag 25560
tctaatcatt tctgacatgg acatggtagt gagataatag gttttccacc ttctgtatta 25620
aaatacattc cagttagtct ttcactaagc tgccaagctc aagttttgcc aaaatgccaa 25680
tgtcttagca tggttttgaa agcctttatt attgaactaa aacctatttt gtttcaacag 25740
ttaagaattt ggcattcatt ctatcaccct ctgcagcagg cttttgggat aaagtgaggg 25800
aagtaggtct catctacttg caggatctct tgggcaccca gaaatctaaa catattggat 25860
taccttttcct gttcacatgt tcctaggctt gttcatgctt tctcctcaac ctggaaacca 25920
cttccctctt gttggcatac tgaaattata ctgatgcttt cagaattagc tcagataccc 25980
ttatcagagt tgcgtaacaa agaaaactaa ttgttaagga ctttgagccc agaggwtttc 26040
caataacaga atttacttga agtttctatg acagatagaa aggacaattg tttaaactat 26100
cctttcttga aagatatgaa aacaagtcat taaaaactct cattttacat ttcagcattc 26160
gaacggtcct gacgttttga gatccaaagt ggcargaggt ctgtgttgtc atggtgaact 26220
ggagtttctc ttgtgagagt tccctcatct gaaatcatgt atctgtctca caaatacaag 26280
cataagtaga agatttgttg aagacataga acccttataa agaattatta acctttataa 26340
acatttaaag tcttgtgagc acctgggaat tagtataata acaatgttaa tatttttgat 26400
ttacattttg taaggctata attgtatctt ttaagaaaac atacacttgg atttctatgt 26460
tgaaatggag atttttaaga gttttaacca gctgctgcag atatataact caaaacagat 26520
atagcgtata aagatatagt aaatgcatct cccagagtaa tattcactta acacattgaa 26580
actattattt tttagatttg aatataaatg tatttttaa acacttgtta tgagttaact 26640
tggattacat tttgaaatca gttcattcca tgatgcatat tactggatta gattaagaaa 26700
gacagaaaag attaagggac gggcacattt ttcaacgatt aagaatcatc attacataac 26760
ttggtgaaac tgaaaaagta tatcatatgg gtacacaagg ctatttgcca gcatatatta 26820
atattttaga aaatattcct tttgtaatac tgaatataaa catagagcta gagtcatatt 26880
atcatactta tcataatgtt caatttgata cagtagaatt gcaagtccct aagtccctat 26940
tcactgtgct tagtagtgac tccatttaat aaaaagtgtt tttagttttt aacaactaaa 27000
ccgatgtatc tatatatatc tataacatgt taaaaattct taagaaaatt aaaaattata 27060
taaaatgttt tgtggtgata tatttttaaa caggcatatg ttgttattaa gcttttattt 27120
gttataaaca aagtgttgaa gcgcctgaaa acataaaaat actttataag acatataaga 27180
agttgggtgg tcatagcaga aattgaggct tccaatgtta gggcataata tagattggac 27240
acttctacca tggtaaattg caaaatgtcc aggtatttat gaaccacttt tacatgatga 27300
atatttccca gtgaaaagtg gcatgaactc gaggagtaaa atttcctttt tcttttacca 27360
gtttgtcaat ctaggcttaa tctaggttta aaaaaaatag ctgaaaagtg taaataatta 27420
ttttgaagta tcttcatcat cagctgagtt tagctatggc aattctcaaa acactatgaa 27480
tatttacagc ttataagaat tatattttta tgtatacata tatatatcta tatttatgtc 27540
aataaataga tacatagata gatgatagat agatagatag atagatagat agatagatga 27600
tagatataga taagtataga aaatctctat ccaagaatca tttggagaca tttcctctgt 27660
ccattagcaa tatctgtagc taggatctca gaatccatta aggaaggaaa actgcccccc 27720
aaaaaagcat ttttaacaag aaaaatatgt agaatggatt aattccttta tgtctcctgc 27780
ctcaaatttt gaaaaacatg taaataagga tactaaaagt aatatgcaaa tataggtaca 27840
tcaaaagaaa taaatccaag aaggaataaa cacatttata ttctcagatc acagcttttt 27900
aaatgaaatc cactttatgc ttagataaaa ttatattttg acaaatatgc tgcatatgtg 27960
acaattttat ttatttaata tttattatga tactttgtgg aatcattcct catgtctgag 28020
cttgtcatgg ccaaataatg aacagaaaga ccttctacta tctttctcag ggtaacctgc 28080
acatgaatag cattttgagt catcagaaaa cactggttct aagaagagat caggaaatct 28140
gattattttt gcatatttct tatttctggt caatgaaacc acaaagcatt gtcaaactga 28200
gttcttgaat ataaatcaaa atataagtaa gcaataaata gctagaggaa tgctaaaaga 28260

aatatccagc ttcttcttta aagaatagtc tagacttcta atcatttaac tgatatgtgg 28320
agaacattca ttgggcaatt ttttttttaa tttcaaccag tattaagcac tgagctaggt 28380
gtttggtata cagtaatgaa taaaacagag taatgaatgt cctttggagc atatggtcta 28440
gtagtgatgg cagctcctct ctgctaggaa ctgaacactc ttcgggacac cctggctgca 28500
gaaaggagtg acccactgca ggtctcctct gagctgttct atcactcaat aaagctcttc 28560
ttcaccttgc tcactcttca gttgtctgca tacctcattc ttccagttcg caggacaaga 28620
acttggggcc caccaaatag ccgggctaaa agagctgtaa caccaacagg gctgaaacac 28680
gccccttgct caccatgttg tgggtgacaa ggagagaaaa gagatggaga gaagagctgc 28740
gacccctttgg ggagcccaga cctaggagct ccctaagcca ggctgtaaca ccctctttgg 28800
aactctatgg ttcctgatgt ctccaagctt ccaggcacca ctgcattctt ggtgtcagct 28860
gtggaagctg cttatggtat gcctggtcca gccacagact tgcaggaagc gggtgctcat 28920
gccagcacct ggagctgccc accctgctgc agccagcatg cctggctgtg tgcagtggcc 28980
agaccccata cttcctcaca cacccctcat cactccatga cgggctcacc cttggcaggt 29040
gtgggaccca gaccagtagc atgagctgag catagcctgc caggccgagt gggccgaatg 29100
agcccagcag gccagagcaa aacatgagca aaggcgccac tggccacaga ggtttctggc 29160
tggcaaagtg acaccccaag gatcctgtaa cagtaggaga tttaataagt aaataaacat 29220
ataagtatat tacttcaatt tttgataaat gataaataag gcaatgatca aggtgctgtg 29280
ttagagagaa aaaagagggg aaacaggaag gtatttccag ctttccataa ggtacttgaa 29340
ggcctctctg aggaggtaaa tttaagctta catggaagga taagctagag ccagttctgt 29400
gaagagaaga gaaaattatc tcaggcaaag gaaaaacatg gccaaaggcc attaatttgg 29460
gaaaagcata cccatttatt gttgggaact ataaaagggc cagtgtagct ggtatttgct 29520
gcaaaacatt aatgatattg tttaggttca cgaagctgaa aaagatgcta aaaccttgtg 29580
gacattttta tacatttacc tcaataatca gcttccttat gtatatttcc caaagataat 29640
gtgatttgat tagtgccaaa attgatatga aattatttta atttagagga agaaaaggac 29700
tatatagaaa tagcattccc aagagtaccc tcttgcccac gacttcaact ccagtctctg 29760
caaagatttc tttcaccagc aatttctttt ttgtttctca ttgctattta catctaaaat 29820
tatttatatt tttatgtaat aagtaactaa ccaaggcccc ataaactagt ttatagtatt 29880
catcttcttc atgcataggt atttgtccat ttttttattc cttactccct cacagagtta 29940
gattcccttc attaaggaaa atctcgatta cctaggccac cagctttagg ttccagtagt 30000
tctttccacc atgtatctga ctttgtctgt taccctatag ggtacagaaa agtgacggat 30060
gccttcactg aaagtccatt ccgttttaca tgtaaagtgt cagaagataa gaattatgtt 30120
ctaagttctt tctcaaaaca gtacctagta ttaatagaac gttttgtacc caatggaca 30179

<210> 2
<211> 1125
<212> DNA
<213> Homo sapien

<400> 2
atgggcactg ctggaaaagt tattaaatgc aaagcagctg tgctttggga gcagaagcaa 60
cccttctcca ttgaggaaat agaagttgcc ccaccaaaga ctaaagaagt tcgcattaag 120
attttggcca caggaatctg tcgcacagat gaccatgtga taaaaggaac aatggtgtcc 180
aagtttccag tgattgtggg acatgaggca actgggattg tagagagcat tggagaagga 240
gtgactacag tgaaaccagg tgacaaagtc atccctctct ttctgccaca atgtagagaa 300
tgcaatgctt gtcgcaaccc agatggcaac ctttgcatta ggagcgatat tactggtcgt 360
ggagtactgg ctgatggcac caccagattt acatgcaagg gcaaaccagt acaccacttc 420
atgaacacca gtacatttac cgagtacaca gtggtggatg aatcttctgt tgctaagatt 480

gatgatgcag ctcctcctga gaaagtctgt ttaattggct gtgggttttc cactggatat 540
ggcgctgctg ttaaaactgg caaggtcaaa cctggttcca cttgcgtcgt ctttggcctg 600
ggaggagttg gcctgtcagt catcatgggc tgtaagtcag ctggtgcatc taggatcatt 660
gggattgacc tcaacaaaga caaatttgag aaggccatgg ctgtaggtgc cactgagtgt 720
atcagtccca aggactctac caaacccatc agtgaggtgc tgtcagaaat gacaggcaac 780
aacgtgggat acacctttga agttattggg catcttgaaa ccatgattga tgccctggca 840
tcctgccaca tgaactatgg gaccagcgtg gttgtaggag ttcctccatc agccaagatg 900
ctcacctatg acccgatgtt gctcttcact ggacgcacat ggaagggatg tgtctttgga 960
ggtttgaaaa gcagagatga tgtcccaaaa ctagtgactg agttcctggc aaagaaattt 1020
gacctggacc agttgataac tcatgtttta ccatttaaaa aaatcagtga aggatttgag 1080
ctgctcaatt caggacaaag cattcgaacg gtcctgacgt tttga 1125

<210> 3
<211> 374
<212> PRT
<213> Homo sapien

<400> 3

Met	Gly	Thr	Ala	Gly	Lys	Val	Ile	Lys	Cys	Lys	Ala	Ala	Val	Leu	Trp
1				5					10					15	
Glu	Gln	Lys	Gln	Pro	Phe	Ser	Ile	Glu	Glu	Ile	Glu	Val	Ala	Pro	Pro
			20					25					30		
Lys	Thr	Lys	Glu	Val	Arg	Ile	Lys	Ile	Leu	Ala	Thr	Gly	Ile	Cys	Arg
		35					40					45			
Thr	Asp	Asp	His	Val	Ile	Lys	Gly	Thr	Met	Val	Ser	Lys	Phe	Pro	Val
	50					55					60				
Ile	Val	Gly	His	Glu	Ala	Thr	Gly	Ile	Val	Glu	Ser	Ile	Gly	Glu	Gly
65					70					75					80
Val	Thr	Thr	Val	Lys	Pro	Gly	Asp	Lys	Val	Ile	Pro	Leu	Phe	Leu	Pro
				85					90					95	
Gln	Cys	Arg	Glu	Cys	Asn	Ala	Cys	Arg	Asn	Pro	Asp	Gly	Asn	Leu	Cys
			100					105					110		
Ile	Arg	Ser	Asp	Ile	Thr	Gly	Arg	Gly	Val	Leu	Ala	Asp	Gly	Thr	Thr
		115					120					125			
Arg	Phe	Thr	Cys	Lys	Gly	Lys	Pro	Val	His	His	Phe	Met	Asn	Thr	Ser
	130					135					140				
Thr	Phe	Thr	Glu	Tyr	Thr	Val	Val	Asp	Glu	Ser	Ser	Val	Ala	Lys	Ile
145					150					155					160

```
Asp Asp Ala Ala Pro Pro Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ala Ala Val Lys Thr Gly Lys Val Lys Pro Gly
            180                 185                 190

Ser Thr Cys Val Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ser Ala Gly Ala Ser Arg Ile Ile Gly Ile Asp Leu
    210                 215                 220

Asn Lys Asp Lys Phe Glu Lys Ala Met Ala Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Ile Ser Pro Lys Asp Ser Thr Lys Pro Ile Ser Glu Val Leu Ser Glu
                245                 250                 255

Met Thr Gly Asn Asn Val Gly Tyr Thr Phe Glu Val Ile Gly His Leu
            260                 265                 270

Glu Thr Met Ile Asp Ala Leu Ala Ser Cys His Met Asn Tyr Gly Thr
        275                 280                 285

Ser Val Val Val Gly Val Pro Pro Ser Ala Lys Met Leu Thr Tyr Asp
    290                 295                 300

Pro Met Leu Leu Phe Thr Gly Arg Thr Trp Lys Gly Cys Val Phe Gly
305                 310                 315                 320

Gly Leu Lys Ser Arg Asp Asp Val Pro Lys Leu Val Thr Glu Phe Leu
                325                 330                 335

Ala Lys Lys Phe Asp Leu Asp Gln Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Lys Lys Ile Ser Glu Gly Phe Glu Leu Leu Asn Ser Gly Gln Ser Ile
        355                 360                 365

Arg Thr Val Leu Thr Phe
    370
```

<210> 4
<211> 15
<212> DNA
<213> Homo sapien

<400> 4
ttataagytg gtctg 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapien

<400> 5
ctatctaygt gaagg 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapien

<400> 6
agctgctrtt atata 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapien

<400> 7
gctgttayat acaac 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapien

<400> 8
aagtctaygt ttgca 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapien

<400> 9
atcatgayct aatga 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapien

<400> 10
tacttgcyta tttct 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapien

<400> 11
tacatttktc ctatg 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapien

<400> 12
tattactkgt cgtgg 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapien

<400> 13
catctagrat cattg 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapien

<400> 14
cctcaacraa gacaa 15

<210> 15

<211> 15
<212> DNA
<213> Homo sapien

<400> 15
gaaggccrtg gctgt 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapien

<400> 16
ttaagagyag gttca 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapien

<400> 17
ccagaggwtt tccaa 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapien

<400> 18
agtggcarga ggtct 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapien

<400> 19
ttctttttat aagyt 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapien

<400> 20
cctttgcaga ccarc 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapien

<400> 21
gccagactat ctayg 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapien

<400> 22
cttgtgcctt cacrt 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapien

<400> 23
ggcacaagct gctrt 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapien

<400> 24
ctgttgtata taaya 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapien

<400> 25
caagctgctg ttaya 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapien

<400> 26
cactctgttg tatrt 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapien

<400> 27
cagaaaaagt ctayg 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapien

<400> 28
tatttctgca aacrt 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapien

<400> 29
ctcctcatca tgayc 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapien

<400> 30
ctcacatcat tagrt 15

<210> 31

<211> 15
<212> DNA
<213> Homo sapien

<400> 31
ttataatact tgcyt 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapien

<400> 32
aacctaagaa atarg 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapien

<400> 33
tgtatttaca tttkt 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapien

<400> 34
aattagcata ggama 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapien

<400> 35
aaacagtatt actkg 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapien

<400> 36
agtactccac gacma 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapien

<400> 37
ctggtgcatc tagra 15

<210> 38
<211> 15
<212> DNA
<213> Homo sapien

<400> 38
caatcccaat gatyc 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapien

<400> 39
gattgacctc aacra 15

<210> 40
<211> 15
<212> DNA
<213> Homo sapien

<400> 40
tcaaatttgt cttyg 15

<210> 41
<211> 15
<212> DNA
<213> Homo sapien

<400> 41
atttgagaag gccrt 15

<210> 42
<211> 15
<212> DNA
<213> Homo sapien

<400> 42
gcacctacag ccayg 15

<210> 43
<211> 15
<212> DNA
<213> Homo sapien

<400> 43
aaacacttaa gagya 15

<210> 44
<211> 15
<212> DNA
<213> Homo sapien

<400> 44
gtccagtgaa cctrc 15

<210> 45
<211> 15
<212> DNA
<213> Homo sapien

<400> 45
ttgagcccag aggwt 15

<210> 46
<211> 15
<212> DNA
<213> Homo sapien

<400> 46
ctgttattgg aaawc 15

<210> 47

<211> 15
<212> DNA
<213> Homo sapien

<400> 47
atccaaagtg gcarg 15

<210> 48
<211> 15
<212> DNA
<213> Homo sapien

<400> 48
caacacagac ctcyt 15

<210> 49
<211> 10
<212> DNA
<213> Homo sapien

<400> 49
tttttataag 10

<210> 50
<211> 10
<212> DNA
<213> Homo sapien

<400> 50
ttgcagacca 10

<210> 51
<211> 10
<212> DNA
<213> Homo sapien

<400> 51
agactatcta 10

<210> 52
<211> 10
<212> DNA
<213> Homo sapien

<400> 52
gtgccttcac 10

<210> 53
<211> 10
<212> DNA
<213> Homo sapien

<400> 53
acaagctgct 10

<210> 54
<211> 10
<212> DNA
<213> Homo sapien

<400> 54
ttgtatataa 10

<210> 55
<211> 10
<212> DNA
<213> Homo sapien

<400> 55
gctgctgtta 10

<210> 56
<211> 10
<212> DNA
<213> Homo sapien

<400> 56
tctgttgtat 10

<210> 57
<211> 10
<212> DNA
<213> Homo sapien

<400> 57
aaaaagtcta 10

<210> 58
<211> 10
<212> DNA
<213> Homo sapien

<400> 58
ttctgcaaac 10

<210> 59
<211> 10
<212> DNA
<213> Homo sapien

<400> 59
ctcatcatga 10

<210> 60
<211> 10
<212> DNA
<213> Homo sapien

<400> 60
acatcattag 10

<210> 61
<211> 10
<212> DNA
<213> Homo sapien

<400> 61
taatacttgc 10

<210> 62
<211> 10
<212> DNA
<213> Homo sapien

<400> 62
ctaagaaata 10

<210> 63

<211> 10
<212> DNA
<213> Homo sapien

<400> 63
atttacattt 10

<210> 64
<211> 10
<212> DNA
<213> Homo sapien

<400> 64
tagcatagga 10

<210> 65
<211> 10
<212> DNA
<213> Homo sapien

<400> 65
cagtattact 10

<210> 66
<211> 10
<212> DNA
<213> Homo sapien

<400> 66
actccacgac 10

<210> 67
<211> 10
<212> DNA
<213> Homo sapien

<400> 67
gtgcatctag 10

<210> 68
<211> 10
<212> DNA
<213> Homo sapien

<400> 68
tcccaatgat 10

<210> 69
<211> 10
<212> DNA
<213> Homo sapien

<400> 69
tgacctcaac 10

<210> 70
<211> 10
<212> DNA
<213> Homo sapien

<400> 70
aatttgtctt 10

<210> 71
<211> 10
<212> DNA
<213> Homo sapien

<400> 71
tgagaaggcc 10

<210> 72
<211> 10
<212> DNA
<213> Homo sapien

<400> 72
cctacagcca 10

<210> 73
<211> 10
<212> DNA
<213> Homo sapien

<400> 73
cacttaagag 10

<210> 74
<211> 10
<212> DNA
<213> Homo sapien

<400> 74
cagtgaacct 10

<210> 75
<211> 10
<212> DNA
<213> Homo sapien

<400> 75
agcccagagg 10

<210> 76
<211> 10
<212> DNA
<213> Homo sapien

<400> 76
ttattggaaa 10

<210> 77
<211> 10
<212> DNA
<213> Homo sapien

<400> 77
caaagtggca 10

<210> 78
<211> 10
<212> DNA
<213> Homo sapien

<400> 78
cacagacctc 10

<210> 79

<211> 18
<212> DNA
<213> Homo sapien

<400> 79
tgtaaaacga cggccagt 18

<210> 80
<211> 19
<212> DNA
<213> Homo sapien

<400> 80
aggaaacagc tatgaccat 19

<210> 81
<211> 1800
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (30)
<223> PS1: polymorphic base T or C

<220>
<221> misc_feature
<222> (61)..(120)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (150)
<223> PS2: polymophic base T or C

<220>
<221> misc_feature
<222> (181)..(240)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (270)
<223> PS3: polymorphic base G or A

<220>

<221> misc_feature
<222> (301)..(360)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (390)
<223> PS4: polymorphic base T or C

<220>
<221> misc_feature
<222> (421)..(480)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (510)
<223> PS5: polymorphic base T or C

<220>
<221> misc_feature
<222> (541)..(600)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (630)
<223> PS6: polymorphic base T or C

<220>
<221> misc_feature
<222> (661)..(720)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (750)
<223> PS7: polymorphic base C or T

<220>
<221> misc_feature
<222> (781)..(840)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (870)
<223> PS8: polymorphic base T or G

<220>
<221> misc_feature
<222> (901)..(960)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (990)
<223> PS9: polymorphic base G or T

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1110)
<223> PS10: polymorphic base G or A

<220>
<221> misc_feature
<222> (1141)..(1200)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1230)
<223> PS11: polymorphic base A or G

<220>
<221> misc_feature
<222> (1261)..(1320)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1350)
<223> PS12: polymorphic base A or G

<220>
<221> misc_feature
<222> (1381)..(1440)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele

<222> (1470)
<223> PS13: polymorphic base C or T

<220>
<221> misc_feature
<222> (1501)..(1560)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1590)
<223> PS14: polymorphic base A or T

<220>
<221> misc_feature
<222> (1621)..(1680)
<223> Ns represent sequence between polymorphic sites

<220>
<221> allele
<222> (1710)
<223> PS15: polymorphic base G or A

<220>
<221> misc_feature
<222> (1741)..(1800)
<223> Ns represent sequence 3' to PS15

<400> 81
atagtagtta ctgtatttct ttttataagy tggtctgcaa agggtagggc ttgcagacca 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
aaggtggaac tctgaagcca gactatctay gtgaaggcac aagctgctgt tatatacaac 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
agactatcta tgtgaaggca caagctgctr ttatatacaa cagagtgaac tgagcatcag 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
tatctatgtg aaggcacaag ctgctgttay atacaacaga gtgaactgag catcagtcag 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
tgaactgagc atcagtcaga aaaagtctay gtttgcagaa atacagatcc aagacaaaga 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
ggaacatttc tgtcccctcc tcatcatgay ctaatgatgt gaggctgata cttagaactt 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
gcataattca gacccattat aatacttgcy tatttcttag gttattaaat gcaaagcagc 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
gtcattttta ctttaatgta tttacatttk tcctatgcta atttttgaat tgaattaatt 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
gtgtatttga tgtatcaaac agtattactk gtcgtggagt actggctgat ggcaccacca 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
atgggctgta agtcagctgg tgcatctagr atcattggga ttgacctcaa caaagacaaa 1140

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
catctaggat cattgggatt gacctcaacr aagacaaatt tgagaaggcc atggctgtag 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
acctcaacaa agacaaattt gagaaggccr tggctgtagg tgccactgag tgtatcagtc 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
ttttgcctaa aaagtcaaac acttaagagy aggttcactg gaccaaagtt aaataccata 1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
ctaattgtta aggactttga gcccagaggw tttccaataa cagaatttac ttgaagtttc 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680
gtcctgacgt tttgagatcc aaagtggcar gaggtctgtg ttgtcatggt gaactggagt 1740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1800

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29508

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/04

US CL : 435/6, 91.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00/00621 A1 (OLSON et al.) 6 Janurary 2000 (06.01.2000), pages 23-24 and table 4.	1-2
X	ZGOMBIC-KNIGHT et al. Genomic Structure and Expression of the ADH7 Gene Encoding Human Class IV Alcohol Dehydrogenase, the Form Most Efficient for Retinol Metabolism in Vitro. The Journal of Biological Chemistry. March 3, 1995, Vol. 270, No. 9, pages 4305-4311, especially page 4306.	1-2
X	NCBI Entrez database nucleotide sequences, National Center for Biotechnology Information, National Library of Medicine, NIH (Bethesda, MD, USA) Accession number AP002027, 'The Complete Genomic DNA Sequence of the Human ADH Gene Complex 1', Tsai, S. F., 19 July 2000.	1-2
X	NCBI Entrez database nucleotide sequences, National Center for Biotechnology Information, National Library of Medicine, NIH (Bethesda, MD, USA) Accession number AC027065, 'Homo sapiens, clone RP11-801P9', Birren et at., 26 May 2000.	1-2

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

24 January 2002 (24.01.2002)

Date of mailing of the international search report

15 FEB 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Felicia D. Roberts for
Juliet C. Einsmann

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/29508

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1-2, haplotypes 1 and 11

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29508

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Groups 1-18, claim(s) 1-2, in part, drawn to methods for haplotyping the alcohol dehydrogenase 7 gene (ADH7) comprising determining which of the 18 haplotypes listed in the table in claim 1 defines one copy of the individual's ADH7 gene. It is noted that Groups 1-18 correspond to the 18 haplotypes listed in the claim. Therefore, the first mentioned invention is the methods of claims 1-2 to the extent that they apply to haplotype 1. **Group 1, the first mentioned invention, is the invention which will be searched in accordance with PCT Article 17(3)(a).** Additional groups may be elected. For example, if Group 2 is elected, and the proper fees are paid, then claims 1-2 will be examined to the extent that they apply to methods of haplotyping comprising a step of determining whether the individual has the second haplotype recited in claim 1. Upon election of an invention to be searched in addition to group 1, please identify the number of the haplotype to be searched in the method claims.

Groups 19-50, claim(s), 3-4, in part, drawn to methods of haplotyping the ADH7 comprising determining which of the 32 haplotype pairs listed in the table in claim 3 defines both copies of an individual's ADH7 gene. Groups 19-50 correspond the 32 haplotype pairs listed in the table in claim 3. If group 19 is elected, then claims 3-4 will be examined to the extent that they apply to methods of haplotyping comprising a step of determining whether the individual has the first listed haplotype pair. Upon election of one or more inventions in this group, please specify the haplotype pairs requested for search.

Groups 51-65, claim(s) 5-7, in part, drawn to a method for genotyping the ADH7 gene which comprises identifying the nucleotide present for one or more of fifteen polymorphic sites. It is noted that Groups 51-65 correspond to polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, and PS15, respectively. For example, if Group 51 is elected, claims 5-7 will be examined to the extent that they apply are limited to method of genotyping comprising a step of identifying the nucleotide pair at PS1.

Groups 66-170, claim(s) 8-9, in part, drawn to a method for haplotyping the ADH7 gene by identifying a ADH7 genotype for the individual at two or more polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, and PS15. It is noted that the claims encompass methods requiring identification of 105 possible combinations of two of the recited polymorphic sites, and that Groups 66-170 each correspond to one of these possible pairs, in the order recited in the claim. For example, if Group 66 is selected, then claims 8-9 will be examined to the extent that they apply to a combination of PS1 and PS2. If Group 170 is selected, then claims 8-9 will be examined to the extent that it applies to a combination of PS14 and PS15. If applicants elect any of these groups, please specify the two sites to be examined in the method for predicting a haplotype pair.

Groups 171-275, claim(s) 10-11, in part, drawn to a method for predicting a haplotype pair for the ADH7 gene by identifying a ADH7 genotype for the individual at two or more polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, and PS15. It is noted that the claims encompass methods requiring identification of 105 possible combinations of two of the recited polymorphic sites, and that Groups 171-275 each correspond to one of these possible pairs, in the order recited in the claim. For example, if Group 171 is selected, then claims 10-11 will be examined to the extent that they apply to a combination of PS1 and PS2. If Group 275 is selected, then claims 10-11 will be examined to the extent that it applies to a combination of PS14 and PS15. If applicants elect any of these groups, please specify the two sites to be examined in the method for predicting a haplotype pair.

Groups 276-325 claim(s) 12-13, in part, drawn to a method for identifying an association between a trait and a haplotype between one of the 18 haplotypes and 32 haplotype pairs of ADH7 gene. Groups 276-325 each correspond to one of the 50 particular combinations of the polymorphic sites, haplotypes, and the haplotype pairs encompassed by the claims (i.e., the 18 different haplotypes, as well as the 32 different haplotype pairs). For example if Group 276 is selected, the claims will be examined to the extent that they apply to the first haplotype.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29508

Groups 326-340, claim(s) 14-18, in part, drawn to a composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the ADH7 gene. Claims 14-18 encompass polynucleotides for the identification of nucleotides present at fifteen different polymorphic sites. Claims 16 and 18 identify specific oligonucleotides by sequence ID numbers. If any one of groups 326-340 is elected, applicant is requested to identify the relevant PS site as well as the relevant sequence ID numbers.

Group 341, claim 19, drawn to a kit comprising a set of oligonucleotides designed to genotype each of the polymorphic sites.

Groups 342-373, claims 20-21 and 24, in part, drawn to a polynucleotide which is a polymorphic variant of a reference sequence for ADH7 gene or a fragment thereof. Claims 20-21 and 24 recite 17 different isogenes and 15 fragments comprising polymorphisms.

Group 374-390, claim(s) 22-23, in part, drawn to a recombinant nonhuman organism comprising a polynucleotide which is a polymorphic variant of a reference sequence for ADH7 gene or a fragment thereof (as recited in claim 20).

Groups 391-408, claims 25 and 28, in part, drawn to isolated polynucleotides comprising ADH7 coding sequences or a fragment thereof. Claims 25 and 28 recite 9 different coding sequences and 9 fragments comprising polymorphisms.

Groups 409-417, claim(s) 26-27, in part, drawn to a recombinant nonhuman organism comprising a polynucleotide which is a polymorphic variant of a reference sequence for ADH7 gene or a fragment thereof (as recited in claim 25).

Groups 418-420, claim(s) 29 and 32, drawn to a polypeptide comprising an amino acid sequence which is a polymorphic variants of a reference sequence for the ADH7 protein or a fragment thereof. Claim 29 recites a total of 3 distinct variant polypeptides, and in the case of claim 32, fragments thereof.

Group 421-423, claim(s) 32, drawn to an antibody which binds to the polypeptide of Claim 29.

Group 424-426, claim(s) 33, drawn to a method for screening for drugs targeting the ADH7 polypeptides listed in claim 29.

Group 427-476, claim(s) 33, drawn to a computer system comprising polymorphism data wherein the data comprises any one or more of the 18 haplotypes and 32 haplotype pairs listed in the claim. It is noted that groups 427-476 correspond to the 50 haplotypes and haplotype pairs listed in the claim. If any of these groups is elected, applicant is requested to identify the particular haplotypes or haplotype pairs desired for search on the computer system.

Groups 477-629, claim(s) 34, in part, drawn to genome anthologies comprising any two or more ADH7 isogene having any one of the haplotypes set forth in the table. It is noted that Groups 477-629 correspond to anthologies comprising any two of the isogenes 1-18 listed in the table in claim 34. For example, Group 477 is drawn to an anthology comprising isogenes having haplotypes 1 and 2.

The inventions listed in the instant application lack unity for a number of reasons.

The haplotyping methods of claims 1-2 comprises haplotyping the ADH7 gene to determine the presence of haplotype 16. Haplotype 16 is the reference or "wild type." Haplotype 11 represents an oligonucleotide which is "variant" only at PS6. Methods for "haplotyping" encompass methods in which the gene in question is sequenced, since during such a sequencing the nucleotide present at each position in the gene is determined. Thus, the claimed methods read on sequencing the ADH7 gene. This method does not provide a special technical feature over the art, since the genes having haplotypes 11 and 16 were known in the art (See GenBank Accession AC027065 and GenBank Accession AP002027), as were methods for sequencing the gene, as is admitted by the specification. Thus, there is no special technical feature linking the recited groups, as would be necessary to fulfill the requirement for unity of invention.

Furthermore, Olson *et al.* (WO 00/00621) teach ADH7 polynucleotides, specifically disclosing polymorphisms in the ADH7 gene (see Fig. 4). Polymorphisms M1, M2, and M7 taught by Olson *et al.* are identical to PS1, PS4, and PS10, respectively, in the instant application. Olson *et al.* further teach methods for the identifictaion of these polymorphisms, and oligonucleotides useful in such methods (see, for example, pages 12-14 and the sequence listing of Olson *et al.*). Olson *et al.* specifically mention minisequencing as a methodology useful for identifying the alleles present in a sample. Thus, the teachings of Olson *et al.* clearly anticipate at least claims 5-9, 14, 15, 17, 24 and 28. Thus, there is a lack of unity since the broadest claims do not provide a special technical feature over the prior art. PCT Rule 13.2 states "The expression "special technical features" shall mean those technical features that define a contribution which each of the claimed inventions, considered as a whole, makes ***over the prior art.*** (emphasis added)" Since at least claims 5-9, 14, 15, 17, 24 and 28 are anticipated by Olson *et al.*, the claims are not joined by a special technical feature.

The first claimed inventions, claims 1-2 (groups 1-19) lack unity because they represent methods which have different results depending on the nucleic acid present in the sample. There is no special technical feature which joins all of the different methods for identification. That is, depending on the nucleic acid present in the sample, the special technical feature of each part of the invention would be the haplotype listed in the table. For each of these groups, the distinct haplotype is considered to be the special technical feature of the group.

Each polymorphic site and each molecule containing said polymorphic site is structurally and functionally distinct from and has a different special technical feature than each other polymorphic site and molecules containing said site. The chemical structure of each polymorphism and of each molecule containing the same differ from each other. For example, a polynucleotide comprising PS1 is chemically, structurally, and functionally different from a molecule comprising PS4. As the products and methods encompassed by the claims do not share a special technical feature, the distinct products and methods may not properly be presented in the alternative. Accordingly, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed by the claims, and the claims will be examined only as they read upon the invention of the elected group. For the same reasons, the remainder of the claims have been separated in a number of groups corresponding to the number of different inventions encompassed thereby.

The haplotypes and genotypes encompassed by the instantly recited method claims are also distinct from each other and from the single polymorphisms recited in e.g., claims 5-7. For example, a molecule of haplotype 1, comprising a particular combination of polymorphisms, differs chemically, structurally, and functionally from a molecule of haplotype 2 and from a molecule comprising a single polymorphism (e.g., PS1). The special technical feature of each haplotype or genotype is the combination of polymorphisms contained therein, which feature is lacking from and not shared with each other haplotype or genotype or with, e.g., a molecule comprising any single polymorphism set forth in the claims. Similarly, with respect to the pairs of polymorphisms of Claim 8, each combination of polymorphisms differs from each other combination and from each of the other combinations discussed above (i.e., haplotypes, genotypes, and single polymorphic sites). The methods utilizing and investigating these nucleic acids are distinct for the same reasons that the nucleic acids themselves are distinct. Thus, the claims have been separated into a number of groups corresponding to the number of different inventions encompassed thereby, and the claims will be examined only as they read upon the invention of the elected group.

Further, the different methods have different objectives and require different process steps. The haplotyping methods require steps of identifying haplotypes and haplotype pairs to achieve the objectives of haplotyping. The methods of genotyping require steps of identifying a single nucleotide on one gene copy to achieve the objective of genotyping. The methods of predicting a haplotype pair require steps of identifying two polymorphisms in a gene to achieve the objective of "predicting a haplotype pair". The methods of identifying an association requires steps of comparing frequencies of haplotypes in a population to achieve the objective of "identifying an association between a trait" and a haplotype. The methods of assaying for binding activity require steps of assaying for binding activity for candidate agents. In addition to differences in objectives, effects, and method steps, it is again noted that the claims of the present Groups are not directed to the detection or identification of molecules having the same or common special technical feature, for the reasons discussed above.

The groups comprising polynucleotides, kits, recombinant organisms, polypeptides, antibodies, computer systems and genome anthologies are additionally drawn to multiple, distinct products lacking the same or corresponding special technical features. The nucleic acids are composed of nucleotides and function in, e.g., methods of nucleic acid

hybridization or amplification. These groups are directed to different combinations of nucleic acids which are different from one another and may be employed in different methods. The recombinant organisms are complex organisms that are employed in, e.g. animal research methods. Such organisms cannot be employed as, e.g., probes or primers and they differ in both structure and function from the nucleic acids. The polypeptides differ in both structure and function from either the nucleic acids or the transgenic organisms. The polypeptides are composed of amino acids linked by peptide bonds and arranged in a complex combination of alpha helices, beta pleated sheets, hydrophobic and hydrophilic domains. The polypeptides also differ in function, e.g., fusion proteins with an enzymatic functions. The antibodies are composed of amino acids linked by peptide bonds, antibodies are glycosylated and their tertiary structure is unique, where four subunits (2 light chains and 2 heavy chains) associated via disulfide bonds into a Y-shaped symmetric dimer. The antibodies function in immunoassays. Further the computer systems are composed of, e.g., a CPU, a display device, an input device, etc., as recited in Claim 31, and function in, e.g., methods of electronic sequence comparison. Accordingly, the products differ structurally and functionally from one another. As products of different sets of Groups differ from each other in structure, function, and effect, they do not belong to a recognized class of chemical compound, or have both a "common property or activity" and a common structure as would be required to show that the inventions are "of a similar nature".

Continuation of* B. *FIELDS SEARCHED Item 3:

EMBL, GENBANK, US PATENTS, GENSEQ, EST sequence databases for portions of SEQ ID 1 containing polymorphic sites

BIOSIS, MEDLINE, CAPLUS word search for ADH7 and synonyms combined with polymorphism or allele or variant or mutation