

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 December 2001 (13.12.2001)

PCT

(10) International Publication Number
WO 01/94361 A2

(51) International Patent Classification⁷:

C07H

(74) Agents: **SHANER, Sandra, L.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(21) International Application Number:

PCT/US01/18321

(22) International Filing Date:

6 June 2001 (06.06.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/209,564 6 June 2000 (06.06.2000) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KLIEM, Stefanie, E.** [DE/US]; 1298 Hartford Turnpike, Apt. 4E, North Haven, CT 06473 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US). **TANGUAY, Debra, A.** [US/US]; 902 Aspen Glen Drive, Hamden, CT 06518 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, 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, 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, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

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 01/94361 A2

(54) Title: HAPLOTYPES OF THE PCDH2 GENE

(57) Abstract: Novel genetic variants of the Protocadherin 2 (Cadherin-Like 2) (PCDH2) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the PCDH2 gene. Compositions and methods for haplotyping and/or genotyping the PCDH2 gene in an individual are also disclosed. Polynucleotides defined by the sequence of the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE PCDH2 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/209,564 filed
5 June 6, 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 protocadherin 2 (cadherin-like 2) 10 (PCDH2) 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, 15 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. 20 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 25 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 30 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 35 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 cancer is the protocadherin 2 (cadherin-like 2) (PCDH2) gene or its encoded product. Protocadherins, such as PCDH2, are members of the cadherin superfamily. They contain six or more cadherin ectodomain repeats, and their cytoplasmic domains differ within the subgroup (Wu and Maniatis, Proc. Natl. Acad. Sci. U. S. A 2000; 97:3124-3129). Studies have suggested that protocadherins are involved in cell-cell interactions critical in the development of the central nervous system (Blanco et al., *Mamm. Genome* 2000; 11:906-914).

For example, expression of protocadherins is spatiotemporally regulated and they are localized at synapses in the CNS (Suzuki, Exp. Cell Res. 2000; 261:13-18). Thus protocadherins may play a central role in the CNS as related to synaptic function and may be involved in the development of neurological disorders.

5 The protocadherin 2 (cadherin-like 2) gene is located on chromosome 5q31 and contains 4 exons that encode a 934 amino acid protein. A reference sequence for the PCDH2 gene comprises nucleotides 1-7652, 7713-19184, and 19245-30244 of Figure 1 below (Genaissance Contig No. 3433164; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. AF152337.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

10 Because of the potential for variation in the PCDH2 gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the PCDH2 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 PCDH2 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

15

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 24 novel polymorphic sites in the PCDH2 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 2027 (PS1), 2137 (PS2), 2160 (PS3), 2393 (PS4), 2427 (PS5), 3096 (PS6), 3253 (PS7), 3258 (PS8), 3375 (PS9), 4629 (PS10), 5076 (PS11), 11834 (PS12), 22396 (PS13), 22670 (PS14), 28129 (PS15), 28318 (PS16), 28423 (PS17), 28479 (PS18), 28743 (PS19), 28980 (PS20), 29407 (PS21), 29500 (PS22), 29614 (PS23) and 29796 (PS24). The polymorphisms at these sites are cytosine or thymine at PS1, thymine or cytosine at PS2, guanine or adenine at PS3, cytosine or adenine at PS4, adenine or guanine at PS5, guanine or cytosine at PS6, thymine or cytosine at PS7, guanine or adenine at PS8, cytosine or thymine at PS9, thymine or cytosine at PS10, cytosine or thymine at PS11, guanine or adenine at PS12, guanine or cytosine at PS13, cytosine or thymine at PS14, adenine or guanine at PS15, guanine or adenine at PS16, guanine or adenine at PS17, thymine or cytosine at PS18, guanine or cytosine at PS19, guanine or adenine at PS20, cytosine or thymine at PS21, adenine or guanine at PS22, guanine or adenine at PS23 and cytosine or thymine at PS24. 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-PS24 in the PCDH2 gene, which are shown below in Tables 5 and 4, respectively. Each of these PCDH2 haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the PCDH2 gene that exists in the human population. The frequency with which each haplotype and haplotype pair occurs 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 PCDH2 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, PS15, PS16, PS17, PS18, PS19, PS20, 5 PS21, PS22, PS23 and PS24 in both copies of the PCDH2 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 PCDH2 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel PCDH2 polymorphic sites. The genotyping method, composition, and kit 10 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 PCDH2 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the PCDH2 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, 15 PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's PCDH2 gene is defined by one of the PCDH2 20 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 PCDH2 gene are defined by one of the PCDH2 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the PCDH2 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 PCDH2 activity, e.g., cancer.

For example, the haplotyping method can be used by the pharmaceutical research scientist to 25 validate PCDH2 as a candidate target for treating a specific condition or disease predicted to be associated with PCDH2 activity. Determining for a particular population the frequency of one or more of the individual PCDH2 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue PCDH2 as a target for treating the specific disease of interest. In particular, if variable PCDH2 activity is associated with the disease, then one or more PCDH2 haplotypes or 30 haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed PCDH2 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable PCDH2 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 PCDH2 haplotype or haplotype pair, apply the 35 information derived from detecting PCDH2 haplotypes in an individual to decide whether modulating PCDH2 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting PCDH2 to treat a

specific condition or disease predicted to be associated with PCDH2 activity. For example, detecting which of the PCDH2 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 most frequent PCDH2 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular PCDH2 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.

The method for haplotyping the PCDH2 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 PCDH2 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 PCDH2 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute PCDH2 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 a PCDH2 haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. 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 PCDH2 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a PCDH2 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the PCDH2 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the PCDH2 genotype or haplotype in a reference population. A higher frequency of the PCDH2 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the PCDH2 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 PCDH2 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 cancer.

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 PCDH2 gene or a fragment thereof. The reference sequence comprises nucleotides 1-7652, 7713-19184, and 19245-30244 shown in Figure 1 (SEQ ID NO:1) and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, cytosine at PS2, adenine at PS3, adenine at PS4, guanine at PS5, cytosine at PS6, cytosine at PS7, adenine at PS8, thymine at PS9, cytosine at PS10, thymine at PS11, adenine at PS12, cytosine at PS13, thymine at PS14, guanine at

PS15, adenine at PS16, adenine at PS17, cytosine at PS18, cytosine at PS19, adenine at PS20, thymine at PS21, guanine at PS22, adenine at PS23 and thymine at PS24.

A particularly preferred polymorphic variant is an isogene of the PCDH2 gene. A PCDH2 isogene of the invention comprises cytosine or thymine at PS1, thymine or cytosine at PS2, guanine or adenine at PS3, cytosine or adenine at PS4, adenine or guanine at PS5, guanine or cytosine at PS6, thymine or cytosine at PS7, guanine or adenine at PS8, cytosine or thymine at PS9, thymine or cytosine at PS10, cytosine or thymine at PS11, guanine or adenine at PS12, guanine or cytosine at PS13, cytosine or thymine at PS14, adenine or guanine at PS15, guanine or adenine at PS16, guanine or adenine at PS17, thymine or cytosine at PS18, guanine or cytosine at PS19, guanine or adenine at PS20, cytosine or thymine at PS21, adenine or guanine at PS22, guanine or adenine at PS23 and cytosine or thymine at PS24. The invention also provides a collection of PCDH2 isogenes, referred to herein as a PCDH2 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a PCDH2 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 cytosine at a position corresponding to nucleotide 582, cytosine at a position corresponding to nucleotide 739, adenine at a position corresponding to nucleotide 744, thymine at a position corresponding to nucleotide 861, cytosine at a position corresponding to nucleotide 2115, adenine at a position corresponding to nucleotide 2479 and guanine at a position corresponding to nucleotide 2646. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a PCDH2 isogene defined by haplotypes 1- 20 and 22 - 31.

Polynucleotides complementary to these PCDH2 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the PCDH2 gene will be useful in studying the expression and function of PCDH2, and in expressing PCDH2 protein for use in screening for candidate drugs to treat diseases related to PCDH2 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic 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 PCDH2 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 PCDH2 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 proline at a position corresponding to amino acid position 247 and serine at a position corresponding to amino acid position 827. A polymorphic variant of PCDH2 is useful in studying the effect of the variation on the biological activity of PCDH2 as well as on the binding affinity of candidate drugs targeting PCDH2 for the treatment of cancer.

The present invention also provides antibodies that recognize and bind to the above

polymorphic PCDH2 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 of the PCDH2 polymorphic genomic variants described herein and methods for producing such animals. The 5 transgenic animals are useful for studying expression of the PCDH2 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against PCDH2 protein, and for testing the efficacy of therapeutic agents and compounds for cancer in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the PCDH2 gene. The computer system comprises a computer 10 processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the PCDH2 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing PCDH2 haplotypes organized according to their evolutionary relationships.

15 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the PCDH2 gene (Genaissance Reference No. 3433164; nucleotides 1-7652, 7713-19184, and 19245-30244; SEQ ID NO:1), 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 20 reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:124 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:125 is a modified version 25 of SEQ ID NO:124 that shows the context sequence of each polymorphic site, PS1-24, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:125 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 PCDH2 coding sequence (contiguous lines; 30 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 PCDH2 protein (contiguous lines; SEQ ID 35 NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the PCDH2 gene. As described in more detail below, the inventors herein discovered 31 isogenes of the PCDH2 gene by characterizing the PCDH2 gene found in genomic DNAs isolated from an Index Repository that 5 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 10 reference population were organized into population subgroups by their self-identified 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 15 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 PCDH2 isogenes present in the human reference population are defined by haplotypes for 24 polymorphic sites in the PCDH2 gene, all of which are believed to be novel. The novel PCDH2 polymorphic sites identified by the inventors are referred to as PS1-PS24 to designate the order in 20 which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index

Repository for PS1-PS24 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the PCDH2 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the PCDH2 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed 5 herein are useful for validating whether PCDH2 is a suitable target for drugs to treat cancer, 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 10 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 15 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 20 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 25 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 30 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 35 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 of a gene found in a population. An isogene contains all of the 5 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 10 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.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-15 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 20 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 within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or 25 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: 30 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 35 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 5 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 10 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.

15 As discussed above, information on the identity of genotypes and haplotypes for the PCDH2 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel PCDH2 polymorphisms and haplotypes identified herein.

20 The compositions comprise at least one PCDH2 genotyping oligonucleotide. In one embodiment, a PCDH2 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, 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 25 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 30 backbone, which may be comprised of linkages such as carboxymethyl, acetamide, 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 35 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.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a PCDH2 polynucleotide, i.e., a PCDH2 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 a non-target region or a non-PCDH2 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 PCDH2 gene using the polymorphism information provided herein in conjunction with the known sequence information for the PCDH2 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 oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred 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 the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting PCDH2 gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

TTCGATCYCCTTCT	(SEQ ID NO:4)	and its complement,
CCTAAGGYAGGTTTC	(SEQ ID NO:5)	and its complement,
15 CGATACTRGCAAGGC	(SEQ ID NO:6)	and its complement,
AAATCAAMGGCATCC	(SEQ ID NO:7)	and its complement,
GGCGCCRGCGCCCA	(SEQ ID NO:8)	and its complement,
AGTACGCSGAGCTGG	(SEQ ID NO:9)	and its complement,
CAACCAGYCCTTGTA	(SEQ ID NO:10)	and its complement,
20 AGTCCTRTTACCGGG	(SEQ ID NO:11)	and its complement,
GCCACAAYCGCGCCG	(SEQ ID NO:12)	and its complement,
TGGGGTTYGTGGTCA	(SEQ ID NO:13)	and its complement,
AACTGGCYCTTCCTA	(SEQ ID NO:14)	and its complement,
GAGACCCRGCACCAG	(SEQ ID NO:15)	and its complement,
25 AGGTCTSGGCATGG	(SEQ ID NO:16)	and its complement,
CCATGCCYACGGACT	(SEQ ID NO:17)	and its complement,
GCTACGGRCCCCAGT	(SEQ ID NO:18)	and its complement,
ACAGGGCRGCCTCTC	(SEQ ID NO:19)	and its complement,
CCCCTTGRGAAACAG	(SEQ ID NO:20)	and its complement,
30 GGTTGAAYATGCAAA	(SEQ ID NO:21)	and its complement,
AAAGTTGSAAGGGCA	(SEQ ID NO:22)	and its complement,
GTGGACRTTTCCAC	(SEQ ID NO:23)	and its complement,
GGGTCAGYGGAGGCC	(SEQ ID NO:24)	and its complement,
AACAGAARGTCTCAG	(SEQ ID NO:25)	and its complement,
35 GTCCCCARTGCGCCC	(SEQ ID NO:26)	and its complement, and
AGAATAGYCAGTAGT	(SEQ ID NO:27)	and its complement.

A preferred ASO primer for detecting PCDH2 gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

40 GTTTGGTCGATCYC	(SEQ ID NO:28); ATCACAAAGAAAGGRG	(SEQ ID NO:29);
GCTTCCCTAACCGYA	(SEQ ID NO:30); GCATAGGAAACCTR	(SEQ ID NO:31);
ATGCACCGATACTRG	(SEQ ID NO:32); CAAAGCGCCTTGCYA	(SEQ ID NO:33);
GCGAGCAAATCAAMG	(SEQ ID NO:34); CTTTCTGGATGCCKT	(SEQ ID NO:35);
GGACTCGGCGCCRG	(SEQ ID NO:36); AGCGCTTGGCGCYG	(SEQ ID NO:37);
45 GCACCAAGTACGCSG	(SEQ ID NO:38); CCAACACCAGCTCSG	(SEQ ID NO:39);
TGTCTTCAACCAGYC	(SEQ ID NO:40); GCCCGGTACAAGGR	(SEQ ID NO:41);
TCAACCAGTCCTTRT	(SEQ ID NO:42); CGCGCGCCCGGTAYA	(SEQ ID NO:43);
TCGGCAGGCCACAAYC	(SEQ ID NO:44); GCACGCCGGCGCRT	(SEQ ID NO:45);
TTCTGTGGGGTTYG	(SEQ ID NO:46); ACACGTGACCACRA	(SEQ ID NO:47);

ACTTTCAACTGGCYC (SEQ ID NO:48); TTGATCTAGGAAGRG (SEQ ID NO:49);
 GGCCCAGAGACCCRG (SEQ ID NO:50); TACCCGCTGGTGCY (SEQ ID NO:51);
 AGAAACAGGTCTTSG (SEQ ID NO:52); CCTACCCCATGCCSA (SEQ ID NO:53);
 CAGCCACCATGCCYA (SEQ ID NO:54); ACATCCAGTCGTRG (SEQ ID NO:55);
 5 GCGCCCGCTACGGRC (SEQ ID NO:56); GGGTGAACGGGGYC (SEQ ID NO:57);
 AGAGCCACAGGGCRG (SEQ ID NO:58); GTTGGGAGAGGCY (SEQ ID NO:59);
 CATGCTCCCCTTGRG (SEQ ID NO:60); TTGTTCTGTTCYC (SEQ ID NO:61);
 CCAGGGGTTGAAYA (SEQ ID NO:62); ACTGCTTTGCATRT (SEQ ID NO:63);
 TTCTGAAAAGTTGSA (SEQ ID NO:64); TCATGATGCCCTTSC (SEQ ID NO:65);
 10 TTCATTGTTGACRT (SEQ ID NO:66); CATGCAGTGGAAAYG (SEQ ID NO:67);
 GTGGAAGGGTCAGYG (SEQ ID NO:68); CTGCTGGGCTCCRC (SEQ ID NO:69);
 CAAATGAACAGAARG (SEQ ID NO:70); CCTGGGCTGAGACYT (SEQ ID NO:71);
 GCCCAGGTCCCCART (SEQ ID NO:72); ACTAGGGGGCGCAYT (SEQ ID NO:73);
 TGGTGTAGAATAGYC (SEQ ID NO:74); and CACTACACTACTGRC (SEQ ID NO:75).
 15

Other genotyping 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 genotyping oligonucleotides are referred to 20 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 PCDH2 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5' to 3', selected from the group consisting 25 of:

TGGTTCGATC	(SEQ ID NO:76);	ACAAGAAAGG	(SEQ ID NO:77);
TTCCCTAACGG	(SEQ ID NO:78);	TAGGAAACCT	(SEQ ID NO:79);
CACCGATACT	(SEQ ID NO:80);	AGCGCCTTGC	(SEQ ID NO:81);
AGCAAATCAA	(SEQ ID NO:82);	TCTGGATGCC	(SEQ ID NO:83);
30 CTCGGCGCCC	(SEQ ID NO:84);	GCTTGGGCAG	(SEQ ID NO:85);
CCAAGTACGC	(SEQ ID NO:86);	ACACCAAGCTC	(SEQ ID NO:87);
CTTCAACCAG	(SEQ ID NO:88);	CGGTACAAGG	(SEQ ID NO:89);
ACCAGTCCTT	(SEQ ID NO:90);	GCGCCGGGTA	(SEQ ID NO:91);
GCAGCCACAA	(SEQ ID NO:92);	CGCCGGCGCG	(SEQ ID NO:93);
35 CTGTGGGTT	(SEQ ID NO:94);	CTGTGACCAC	(SEQ ID NO:95);
TTCAACTGGC	(SEQ ID NO:96);	ATCTAGGAAG	(SEQ ID NO:97);
CCAGAGACCC	(SEQ ID NO:98);	CCGCTGGTGC	(SEQ ID NO:99);
AACAGGTCTT	(SEQ ID NO:100);	ACCCCATGCC	(SEQ ID NO:101);
40 CCACCATGCC	(SEQ ID NO:102);	TCCAGTCCGT	(SEQ ID NO:103);
CCCGCTACGG	(SEQ ID NO:104);	TGAACGGGG	(SEQ ID NO:105);
GCCACAGGGC	(SEQ ID NO:106);	GGGGAGAGGC	(SEQ ID NO:107);
GCTCCCCTTG	(SEQ ID NO:108);	TTTCTGTTTC	(SEQ ID NO:109);
GGGGGTTGAA	(SEQ ID NO:110);	GCTTTTGCAT	(SEQ ID NO:111);
TGAAAAGTTG	(SEQ ID NO:112);	TGATGCCCTT	(SEQ ID NO:113);
45 ATTGTTGAC	(SEQ ID NO:114);	GCAGTGGAAA	(SEQ ID NO:115);
GAAGGGTCAG	(SEQ ID NO:116);	CTGGGGCTCC	(SEQ ID NO:117);
ATGAACAGAA	(SEQ ID NO:118);	GGGCTGAGAC	(SEQ ID NO:119);
CAGGTCCCCA	(SEQ ID NO:120);	AGGGGGCGCA	(SEQ ID NO:121);
TGTAGAATAG	(SEQ ID NO:122); and	TACACTACTG	(SEQ ID NO:123).

50

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides 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.

PCDH2 genotyping 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 genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized PCDH2 genotyping 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 genotyping 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 PCDH2 gene in an individual. As used herein, the terms "PCDH2 genotype" and "PCDH2 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 PCDH2 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the PCDH2 gene, or a fragment thereof, 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, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24 in the two copies to assign a PCDH2 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-PS24.

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 PCDH2 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 a PCDH2 gene fragment is isolated, it must 5 contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the PCDH2 gene, or a fragment thereof, 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, 10 PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24 in that copy to assign a PCDH2 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the PCDH2 gene or fragment such as one of the methods described above for preparing PCDH2 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 15 only provide haplotype information on one of the two PCDH2 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional PCDH2 clones will 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 PCDH2 gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-PS24 is identified.

20 In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the PCDH2 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's PCDH2 gene, the phased sequence of nucleotides present at each of PS1-PS24. The present invention also contemplates that typically only a subset of PS1-PS24 will need to be directly examined to assign to an individual one or more of the haplotypes shown in 25 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 30 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 a preferred embodiment, a PCDH2 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, 35 PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24 in each copy of the PCDH2 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-PS24 in each copy of the

PCDH2 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 5 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.

10 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 PCDH2 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 15 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 20 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 25 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).

30 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 35 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 5 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 10 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 PCDH2 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, 15 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 20 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) 25 (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 30 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.* 35 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 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 5 genomic regions not examined herein. Genotyping of 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 PCDH2 haplotype pair is predicted from its PCDH2 genotype using information on haplotype pairs known to exist in a reference population. In its 10 broadest embodiment, the haplotyping prediction method comprises identifying a PCDH2 genotype for the individual at two or more PCDH2 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing PCDH2 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the PCDH2 15 haplotype pairs shown in Table 4.

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 20 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 25 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 30 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 35 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,

- 5 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 a PCDH2 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

10 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

15 the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) 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

20 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*).

A preferred process for predicting PCDH2 haplotype pairs from PCDH2 genotypes is described in U.S. Provisional Application Serial No. 60/198,340 and the corresponding International Application filed April 18, 2001.

25 The invention also provides a method for determining the frequency of a PCDH2 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 PCDH2 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex

30 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 PCDH2 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 a PCDH2 genotype, haplotype, or haplotype pair. The trait may be any

35 detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment.

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 of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the 5 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 10 the populations are compared. If a particular PCDH2 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 PCDH2 genotype, haplotype or haplotype pair. Preferably, the PCDH2 genotype, haplotype, or haplotype pair being compared in the trait and reference 15 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 PCDH2 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 20 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 adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a PCDH2 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited 25 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 30 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 35 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 PCDH2 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 PCDH2 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their PCDH2 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 PCDH2 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between PCDH2 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 PCT Application Serial No. PCT/US00/17540.

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 PCDH2 gene. As described in PCT Application Serial No. PCT/US00/17540, 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 PCDH2 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

5 The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the PCDH2 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 10 of the polymorphic sites in the PCDH2 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 PCDH2 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.

15 In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the PCDH2 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 PCDH2 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, 20 PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24. Similarly, the nucleotide sequence of a variant fragment of the PCDH2 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 PCDH2 gene, which is defined by 25 haplotype 21, (or other reported PCDH2 sequences) or to portions of the reference sequence (or other reported PCDH2 sequences), except for genotyping oligonucleotides as described above.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS1, cytosine at PS2, adenine at PS3, adenine at PS4, guanine at PS5, cytosine at PS6, cytosine at PS7, adenine at PS8, thymine at PS9, cytosine at PS10, thymine at PS11, adenine at PS12, cytosine at PS13, thymine at PS14, guanine at PS15, adenine at PS16, adenine at PS17, cytosine at PS18, cytosine at PS19, adenine at PS20, thymine at PS21, guanine at PS22, adenine at PS23 and thymine at PS24. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the PCDH2 gene which is defined by any one of haplotypes 1- 20 and 22 - 31 shown in Table 5 below.

35 Polymorphic variants of the invention may be prepared by isolating a clone containing the PCDH2 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 claimed herein

could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

PCDH2 isogenes may be isolated using any method that allows separation of the two "copies" of the PCDH2 gene present in an individual, which, as readily understood by the skilled artisan, may be 5 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. 10 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 PCDH2 genome anthologies, which are collections of PCDH2 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 15 population, and a same sex population. A PCDH2 genome anthology may comprise individual PCDH2 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 PCDH2 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of 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, 20 freeze-dried preparations and the like. A preferred PCDH2 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

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 PCDH2 protein in a 25 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 30 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 35 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 PCDH2 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, 5 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 10 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 PCDH2 gene will produce PCDH2 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 a 15 PCDH2 cDNA comprising a nucleotide sequence which is a polymorphic variant of the PCDH2 reference coding sequence shown in Figure 2. Thus, the invention also provides PCDH2 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of cytosine at a position corresponding to nucleotide 582, cytosine at a position 20 corresponding to nucleotide 739, adenine at a position corresponding to nucleotide 744, thymine at a position corresponding to nucleotide 861, cytosine at a position corresponding to nucleotide 2115, adenine at a position corresponding to nucleotide 2479 and guanine at a position corresponding to nucleotide 2646. A particularly preferred polymorphic cDNA variant comprises the coding sequence of 25 a PCDH2 isogene defined by haplotypes 1- 20 and 22 - 31. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized PCDH2 cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

30 As used herein, a polymorphic variant of a PCDH2 gene 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.

35 In describing the PCDH2 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 PCDH2 gene 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 5 polynucleotides which are complementary to the sense strand of the PCDH2 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular PCDH2 protein isoform, an expression vector encoding the isoform may be 10 administered to the patient. The patient may be one who lacks the PCDH2 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 PCDH2 isogene. Expression of a PCDH2 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 for the 15 isogene. 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 20 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 PCDH2 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of PCDH2 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector 25 introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides 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 30 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 the reference PCDH2 amino acid sequence shown in Figure 3. The location of a variant amino acid in a PCDH2 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID 35 NO:3 (Fig. 3). A PCDH2 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of proline at a position corresponding to amino acid position 247 and serine at a position corresponding

to amino acid position 827. The invention specifically excludes amino acid sequences identical to those previously identified for PCDH2, including SEQ ID NO:3, and previously described fragments thereof. PCDH2 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 5 embodiments, a PCDH2 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

Table 2. Novel Polymorphic Variants of PCDH2

Polymorphic Variant	Amino Acid Position and Identities	
Number	247	827
1	S	S
2	P	G
3	P	S

15 The invention also includes PCDH2 peptide variants, which are any fragments of a PCDH2 protein variant that contain one or more of the amino acid variations shown in Table 2. A PCDH2 peptide variant 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 20 long. Such PCDH2 peptide variants may be useful as antigens to generate antibodies specific for one of the above PCDH2 isoforms. In addition, the PCDH2 peptide variants may be useful in drug screening assays.

A PCDH2 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant PCDH2 genomic and cDNA sequences as described above. 25 Alternatively, the PCDH2 protein variant may be isolated from a biological sample of an individual having a PCDH2 isogene which encodes the variant protein. Where the sample contains two different PCDH2 isoforms (i.e., the individual has different PCDH2 isogenes), a particular PCDH2 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular PCDH2 isoform but does not bind to the other PCDH2 isoform.

30 The expressed or isolated PCDH2 protein 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 PCDH2 protein as discussed further below. PCDH2 variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel 35 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.

40 A polymorphic variant PCDH2 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric PCDH2 protein. The non-PCDH2 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 PCDH2 and non-PCDH2 portions so that the PCDH2 protein may be cleaved and purified away from the non-PCDH2 portion.

An additional embodiment of the invention relates to using a novel PCDH2 protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents 5 that bind specifically to all known PCDH2 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 PCDH2 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 a PCDH2 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of 10 test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the PCDH2 protein(s) of interest and then washed. Bound PCDH2 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel PCDH2 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the PCDH2 protein.

15 In yet another embodiment, when a particular PCDH2 haplotype or group of PCDH2 haplotypes encodes a PCDH2 protein variant with an amino acid sequence distinct from that of PCDH2 protein isoforms encoded by other PCDH2 haplotypes, then detection of that particular PCDH2 haplotype or group of PCDH2 haplotypes may be accomplished by detecting expression of the encoded 20 PCDH2 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 PCDH2 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The PCDH2 protein or peptide variant used to generate the 25 antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the PCDH2 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 30 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 PCDH2 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

35 Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the PCDH2 protein variant from solution as well as react with PCDH2 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or

substrates. In another preferred embodiment, the antibodies will detect PCDH2 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.

- 5 In another embodiment, an antibody specifically immunoreactive with one of the novel PCDH2 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 PCDH2 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, 10 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. 15 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 Oellrich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological 20 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 25 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 30 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 PCDH2 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant 35 animals, containing a polymorphic variant of the PCDH2 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 PCDH2 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 PCDH2 isogene may be introduced 5 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 PCDH2 isogene is introduced into a cell in such a way that it recombines with the endogenous PCDH2 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired PCDH2 gene polymorphism. Vectors for the introduction of genes both 10 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 PCDH2 isogene may be introduced include, but 15 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 PCDH2 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 PCDH2 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant 20 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 transfected 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 25 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 PCDH2 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. 30 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 PCDH2 isogene and producing human PCDH2 protein can be used as biological models for studying diseases related to abnormal PCDH2 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 35 disorders affected by expression or function of a novel PCDH2 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide

comprising one of these novel PCDH2 isogenes; an antisense oligonucleotide directed against one of the novel PCDH2 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel PCDH2 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 PCDH2 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 PCDH2 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 PCDH2 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).

15

EXAMPLE 1

This example illustrates examination of various regions of the PCDH2 gene for polymorphic sites.

Amplification of Target Regions

20

The following target regions of the PCDH2 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1.

PCR Primer Pairs

	Fragment No.	Forward Primer	Reverse Primer	PCR Product
25	Fragment 1	1904-1925	complement of 2483-2462	580 nt
	Fragment 2	2179-2201	complement of 2786-2764	608 nt
	Fragment 3	2271-2293	complement of 2779-2757	509 nt
	Fragment 4	2505-2527	complement of 3160-3136	656 nt
	Fragment 5	2770-2791	complement of 3423-3401	654 nt
30	Fragment 6	3074-3094	complement of 3717-3694	644 nt
	Fragment 7	3388-3409	complement of 3968-3948	581 nt
	Fragment 8	3635-3656	complement of 4235-4215	601 nt
	Fragment 9	3929-3949	complement of 4520-4497	592 nt
	Fragment 10	4081-4103	complement of 4850-4830	770 nt
35	Fragment 11	4503-4526	complement of 5118-5098	616 nt
	Fragment 12	11517-11538	complement of 12060-12038	544 nt
	Fragment 13	22354-22378	complement of 22775-22753	422 nt
	Fragment 14	27912-27937	complement of 28529-28507	618 nt
	Fragment 15	28185-28204	complement of 28792-28770	608 nt
40	Fragment 16	28502-28522	complement of 29091-29070	590 nt
	Fragment 17	28775-28797	complement of 29379-29358	605 nt
	Fragment 18	29071-29092	complement of 29673-29651	603 nt
	Fragment 19	29307-29329	complement of 29995-29973	689 nt
	Fragment 20	29652-29674	complement of 30306-30325	595 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
5 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
10 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

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

20 97°C - 15 sec.
 70°C - 45 sec.
 72°C - 45 sec. } 10 cycles

25 97°C - 15 sec.
 64°C - 45 sec.
 72°C - 45 sec. } 35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltrronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled
30 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 primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA
35 Analyzer.

Sequencing Primer Pairs

	Fragment No.	Forward Primer	Reverse Primer
5	Fragment 1	1910-1931	complement of 2421-2403
	Fragment 2	2220-2238	complement of 2735-2716
	Fragment 3	2281-2300	complement of 2761-2741
	Fragment 4	2551-2570	complement of 3066-3047
10	Fragment 5	2854-2872	complement of 3339-3321
	Fragment 6	3147-3166	complement of 3631-3611
	Fragment 7	3417-3436	complement of 3933-3914
	Fragment 8	3690-3708	complement of 4206-4187
	Fragment 9	4003-4023	complement of 4494-4475
	Fragment 10	4270-4289	complement of 4784-4765
15	Fragment 11	4540-4559	complement of 5077-5056
	Fragment 12	11651-11671	complement of 11973-11953
	Fragment 13	22384-22403	complement of 22718-22698
	Fragment 14	27939-27958	complement of 28452-28433
	Fragment 15	28239-28257	complement of 28767-28748
20	Fragment 16	28528-28547	complement of 29050-29029
	Fragment 17	28813-28932	complement of 29323-29404
	Fragment 18	29102-29121	complement of 29650-29630
	Fragment 19	29383-29401	complement of 29920-29901
	Fragment 20	29676-29695	complement of 30208-30189

25

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 PCDH2 gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the PCDH2 Gene

	Polymorphic Site Number	PolyId ^a	Nucleotide Position	Reference Allele	Variant Allele	CDS Position	AA Variant
5	PS1	3433191	2027	C	T		
	PS2	3433195	2137	T	C		
	PS3	3433199	2160	G	A		
	PS4	3433204	2393	C	A		
10	PS5	3433206	2427	A	G		
	PS6	3433208	3096	G	C	582	A194A
	PS7	3433211	3253	T	C	739	S247P
	PS8	3433213	3258	G	A	744	L248L
	PS9	3433217	3375	C	T	861	N287N
15	PS10	3433328	4629	T	C	2115	F705F
	PS11	3433330	5076	C	T		
	PS12	3433334	11834	G	A	2479	G827S
	PS13	3433341	22396	G	C		
	PS14	3433347	22670	C	T		
20	PS15	3433351	28129	A	G	2646	G882G
	PS16	3433355	28318	G	A		
	PS17	3433361	28423	G	A		
	PS18	3433365	28479	T	C		
	PS19	3433375	28743	G	C		
25	PS20	3433377	28980	G	A		
	PS21	3433386	29407	C	T		
	PS22	3433388	29500	A	G		
	PS23	3433390	29614	G	A		
	PS24	3433392	29796	C	T		

30 ^aPolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

EXAMPLE 2

This example illustrates analysis of the PCDH2 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

35 The different genotypes containing these polymorphisms that were observed in 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
40 based on linkage disequilibrium and/or Mendelian inheritance.

Table 4 (Part1). Genotypes and Haplotype Pairs Observed for PCDH2 Gene
 Genotype Polymorphic Sites

	Number	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	HAP	Pair
5	1	C	T	G	C	A	G	T	G	C	T	21	21
	2	C	T	G	C	A	G	T	G	C	C	16	16
	3	C/T	T	G	C	A	C/G	T	G	C	C	8	28
	4	C	T	G	C	A	C/G	T	G	C	C	8	17
	5	C	T	G	C	A	G/C	T	G	C	T/C	21	10
10	6	C	T	G/A	C	A	G	T	G/A	C	T/C	21	2
	7	T/C	T	G	C	A	G/C	T	G	C	C	28	9
	8	C	T	G/A	C	A	G	T	G	C	C	19	5
	9	C	T	G	C/A	A	G	T	G	C	T/C	21	6
	10	C	T	G	C	A	G	T	G	C	T/C	21	12
15	11	C	T/C	G	C	A	G	T	G	C	T	21	1
	12	C	T	G	C	A	G/C	T	G	C	T/C	21	9
	13	C	T	G	C	A	C	T	G	C	C	8	9
	14	C	T	G/A	C	A	C/G	T	G	C	C	8	3
	15	C	T	G	C	A	G	T	G	C	C	16	19
20	16	C/T	T	G	C	A	C/G	T	G	C	C	8	31
	17	C	T	G	C	A/G	G/C	T	G	C	T/C	21	24
	18	C	T	G	C	A	G	T	G	C	C	16	12
	19	C	T	G	C	A	C/G	T	G	C	C	8	14
	20	C	T	G	C	A	C/G	T	G	C	C	8	19
25	21	C/T	T	G	C	A	G	T	G	C	T/C	21	29
	22	C/T	T	G	C	A	G	T	G	C	T/C	21	30
	23	C/T	T	G/A	C	A	G	T	G	C	C	13	25
	24	C	T	G	C/A	A	G	T	G	C/T	T/C	21	7
	25	C	T	G	C	A	C/G	T	G	C	C	8	18
30	26	C/T	T	G	C	A	G	T	G	C	T/C	21	28
	27	C/T	T	G	C	A	G	T/C	G	C	C	13	27
	28	T/C	T	G	C	A	G	T	G	C	C	30	11
	29	C	T	G	C	A	G	T	G	C	C	16	13
	30	C	T	G	C	A	G	T	G	C/T	C	19	23
35	31	C	T	G	C	A	C/G	T	G	C	C	8	13
	32	C/T	T	G/A	C	A	G	T	G	C	C	13	26
	33	C	T	G	C	A	G	T	G	C	T	21	20
	34	C	T	G	C	A	G/C	T	G	C	T/C	21	8
	35	C/T	T	G	C	A	C/G	T	G	C	C	8	30
40	36	C	T	G	C	A	G	T	G	C	T/C	21	16
	37	C	T	G	C	A	C/G	T	G	C	C	8	16
	38	C	T	G	C	A	G	T	G	C	T/C	21	13
	39	C	T	A/G	C	A	G	T	G	C	C	4	15
	40	C	T	A/G	C	A	G	T	G	C/T	C	4	22
45	41	C	T/C	G	C	A	G/C	T	G	C	T/C	8	1
	42	C	T	G	C	A	G/C	T	G	C	C	16	9
	43	C	T/C	G	C	A	G	T	G	C	C/T	16	1
	44	C	T	A	C	A	G	T	G	C	C	4	4
	45	C	T	G	C	A	G	T	G	C/T	C	22	15

Table 4 (Part2). Genotypes and Haplotype Pairs Observed for PCDH2 Gene
Genotype Polymorphic Sites

	Number	PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20	HAP	Pair
5	1	C	G	G	C	A	G	G	T	G	G	21	21
	2	T	G	G	C	A	A	G	T	G	G	16	16
	3	C	G	G/C	C	A	G	G/A	T	C/G	G	8	28
	4	C/T	G	G	C	A	G/A	G	T	C/G	G	8	17
	5	C/T	G	G	C	A	G/A	G	T	G/C	G	21	10
10	6	C/T	G	G	C	A	G	G	T	G/C	G	21	2
	7	C	G	C/G	C	A	G	A/G	T	G	G	28	9
	8	T	G	G	C/T	A	G	G	T	G/C	G	19	5
	9	C/T	G	G	C/T	A	G	G	T	G	G	21	6
	10	C	G	G	C	A	G	G	T	G/C	G	21	12
15	11	C/T	G	G	C	A	G/A	G	T	G	G	21	1
	12	C	G	G	C	A	G	G	T	G	G	21	9
	13	C	G	G	C	A	G	G	T	C/G	G	8	9
	14	C/T	G	G	C	A	G	G	T	C	G	8	3
	15	T	G	G	C	A	A/G	G	T	G	G	16	19
20	16	C	G	G	C	A	G	G	T	C/G	G	8	31
	17	C	G	G	C	A	G	G	T	G/C	G	21	24
	18	T/C	G	G	C	A	A/G	G	T	G/C	G	16	12
	19	C	G	G	C	A	G	G	T	C/G	G	8	14
	20	C/T	G	G	C	A	G	G	T	C/G	G	8	19
25	21	C	G	G/C	C	A	G	G	T/C	G	G	21	29
	22	C	G	G/C	C	A	G	G	T	G	G	21	30
	23	C/T	G	G/C	C	A	G	G	T	G/C	G	13	25
	24	C/T	G	G	C	A	G	G	T	G	G	21	7
	25	C/T	G	G	C	A	G	G	T	C	G	8	18
30	26	C	G	G/C	C	A	G	G/A	T	G	G	21	28
	27	C	G	G/C	C	A	G	G	T	G	G	13	27
	28	C	G	C	C	A	G	G	T	G	G/A	30	11
	29	T/C	G	G	C	A	A/G	G	T	G	G	16	13
	30	T	G	G	C/T	A/G	G	G	T	G	G	19	23
35	31	C	G	G	C	A	G	G	T	C/G	G	8	13
	32	C/T	G	G	C	A	G	G	T	G	G	13	26
	33	C	G/A	G	C	A	G	G	T	G	G	21	20
	34	C	G	G	C	A	G	G	T	G/C	G	21	8
	35	C	G	G/C	C	A	G	G	T	C/G	G	8	30
40	36	C/T	G	G	C	A	G/A	G	T	G	G	21	16
	37	C/T	G	G	C	A	G/A	G	T	C/G	G	8	16
	38	C	G	G	C	A	G	G	T	G	G	21	13
	39	T/C	G	G	C	A/G	G	G	T	G	G	4	15
	40	T	G	G	C	A	G	G	T	G	G	4	22
45	41	T/C	G	G	C	A	G/A	G	T	G/C	G	8	1
	42	C/T	G	G	C	A	G/A	G	T	G	G	16	9
	43	T	G	G	C	A	A	G	T	G	G	16	1
	44	T	G	G	C	A	G	G	T	G	G	4	4
	45	C/T	G	G	C	A/G	G	G	T	G	G	22	15

Table 4 (Part3). Genotypes and Haplotype Pairs Observed for PCDH2 Gene

Genotype	Polymorphic Sites					
Number	PS21	PS22	PS23	PS24	HAP	Pair
5	1	C	A	G	C	21
	2	C	A	G	C	16
	3	C	A	G	C	8
	4	C/T	A/G	G	C	8
10	5	C	A/G	G	C	21
	6	C	A	G	C	21
	7	C	A	G	C	28
	8	C	A	G	C	19
15	9	C	A	G	C	21
	10	C	A/G	G	C	21
	11	C	A	G	C	21
	12	C	A	G	C	21
20	13	C	A	G	C	8
	14	C	A	G	C	8
	15	C	A	G	C	16
	16	C	A	G	C	8
25	17	C	A	G	C	21
	18	C	A/G	G	C	16
	19	C	A/G	G	C	8
	20	C	A	G	C	8
30	21	C	A	G/A	C	21
	22	C	A	G	C	21
	23	C	A	G	C	13
	24	C	A	G	C	21
35	25	C	A	G	C	8
	26	C	A	G	C	21
	27	C	A	G	C	13
	28	C	A	G	C	30
40	29	C	A	G	C	16
	30	C	A	G	C	19
	31	C	A	G	C	8
	32	C	A	G	C/T	13
45	33	C	A	G	C	21
	34	C	A	G	C	21
	35	C	A	G	C	8
	36	C	A	G	C	21
50	37	C	A	G	C	8
	38	C	A	G	C	21
	39	C	A	G	C	4
	40	C	A	G	C	4
45	41	C	A	G	C	8
	42	C	A	G	C	16
	43	C	A	G	C	16
	44	C	A	G	C	4
50	45	C	A	G	C	22
						15

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 U.S. Provisional

Application Serial No. 60/198,340 entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms" and the corresponding International Application 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 our 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 31 human PCDH2 haplotypes shown in Table 5 below.

A PCDH2 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.

15 Table 5 (Part 1). Haplotypes Identified in the PCDH2 Gene

Haplotype Number ^a										PS Number ^b	PS Position ^c	SEQ ID NO. ^d	Region Examined ^e
1	2	3	4	5	6	7	8	9	10				
C	C	C	C	C	C	C	C	C	C	1	2027/30	124/125	1904-5118
C	T	T	T	T	T	T	T	T	T	2	2137/150	124/125	1904-5118
G	A	A	A	A	G	G	G	G	G	3	2160/270	124/125	1904-5118
C	C	C	C	C	A	A	C	C	C	4	2393/390	124/125	1904-5118
A	A	A	A	A	A	A	A	A	A	5	2427/510	124/125	1904-5118
G	G	G	G	G	G	G	C	C	C	6	3096/630	124/125	1904-5118
T	T	T	T	T	T	T	T	T	T	7	3253/750	124/125	1904-5118
G	A	G	G	G	G	G	G	G	G	8	3258/870	124/125	1904-5118
C	C	C	C	C	T	C	C	C	C	9	3375/990	124/125	1904-5118
T	C	C	C	C	C	C	C	C	C	10	4629/1110	124/125	1904-5118
T	T	T	T	T	T	T	C	C	T	11	5076/1230	124/125	1904-5118
G	G	G	G	G	G	G	G	G	G	12	11834/1350	124/125	11517-12060
G	G	G	G	G	G	G	G	G	G	13	22396/1470	124/125	22354-22775
C	C	C	C	T	T	C	C	C	C	14	22670/1590	124/125	22354-22775
A	A	A	A	A	A	A	A	A	A	15	28129/1710	124/125	27912-30306
A	G	G	G	G	G	G	G	G	A	16	28318/1830	124/125	27912-30306
G	G	G	G	G	G	G	G	G	G	17	28423/1950	124/125	27912-30306
T	T	T	T	T	T	T	T	T	T	18	28479/2070	124/125	27912-30306
G	C	C	G	C	G	G	C	G	C	19	28743/2190	124/125	27912-30306
G	G	G	G	G	G	G	G	G	G	20	28980/2310	124/125	27912-30306
C	C	C	C	C	C	C	C	C	C	21	29407/2430	124/125	27912-30306
A	A	A	A	A	A	A	A	A	G	22	29500/2550	124/125	27912-30306
G	G	G	G	G	G	G	G	G	G	23	29614/2670	124/125	27912-30306
C	C	C	C	C	C	C	C	C	C	24	29796/2790	124/125	27912-30306

Table 5 (Part 2). Haplotypes Identified in the PCDH2 Gene

Haplotype Number ^a												PS Number ^b	PS Position ^b	SEQ ID NO: ^d	Region Examined ^c
11	12	13	14	15	16	17	18	19	20						
C	C	C	C	C	C	C	C	C	C		1	2027/30	124/125	1904-5118	
T	T	T	T	T	T	T	T	T	T		2	2137/150	124/125	1904-5118	
G	G	G	G	G	G	G	G	G	G		3	2160/270	124/125	1904-5118	
C	C	C	C	C	C	C	C	C	C		4	2393/390	124/125	1904-5118	
A	A	A	A	A	A	A	A	A	A		5	2427/510	124/125	1904-5118	
G	G	G	G	G	G	G	G	G	G		6	3096/630	124/125	1904-5118	
T	T	T	T	T	T	T	T	T	T		7	3253/750	124/125	1904-5118	
G	G	G	G	G	G	G	G	G	G		8	3258/870	124/125	1904-5118	
C	C	C	C	C	C	C	C	C	C		9	3375/990	124/125	1904-5118	
C	C	C	C	C	C	C	C	C	T		10	4629/1110	124/125	1904-5118	
C	C	C	C	C	T	T	T	T	C		11	5076/1230	124/125	1904-5118	
G	G	G	G	G	G	G	G	G	A		12	11834/1350	124/125	11517-12060	
C	G	G	G	G	G	G	G	G	G		13	22396/1470	124/125	22354-22775	
C	C	C	C	C	C	C	C	C	C		14	22670/1590	124/125	22354-22775	
A	A	A	A	G	A	A	A	A	A		15	28129/1710	124/125	27912-30306	
G	G	G	G	A	A	G	G	G	G		16	28318/1830	124/125	27912-30306	
G	G	G	G	G	G	G	G	G	G		17	28423/1950	124/125	27912-30306	
T	T	T	T	T	T	T	T	T	T		18	28479/2070	124/125	27912-30306	
G	C	G	G	G	G	G	C	G	G		19	28743/2190	124/125	27912-30306	
A	G	G	G	G	G	G	G	G	G		20	28980/2310	124/125	27912-30306	
C	C	C	C	C	C	T	C	C	C		21	29407/2430	124/125	27912-30306	
A	G	A	G	A	A	G	A	A	A		22	29500/2550	124/125	27912-30306	
G	G	G	G	G	G	G	G	G	G		23	29614/2670	124/125	27912-30306	
C	C	C	C	C	C	C	C	C	C		24	29796/2790	124/125	27912-30306	

Table 5 (Part 3). Haplotypes Identified in the PCDH2 Gene

Haplotype Number ^a													PS Number ^b	PS Position ^b	SEQ ID NO. ^d	Region Examined ^c
21	22	23	24	25	26	27	28	29	30	31						
C	C	C	C	T	T	T	T	T	T	T	1	2027/30	124/125	1904-5118		
T	T	T	T	T	T	T	T	T	T	T	2	2137/150	124/125	1904-5118		
G	G	G	G	A	A	G	G	G	G	G	3	2160/270	124/125	1904-5118		
C	C	C	C	C	C	C	C	C	C	C	4	2393/390	124/125	1904-5118		
A	A	A	G	A	A	A	A	A	A	A	5	2427/510	124/125	1904-5118		
G	G	G	C	G	G	G	G	G	G	G	6	3096/630	124/125	1904-5118		
T	T	T	T	T	C	T	T	T	T	T	7	3253/750	124/125	1904-5118		
G	G	G	G	G	G	G	G	G	G	G	8	3258/870	124/125	1904-5118		
C	T	T	C	C	C	C	C	C	C	C	9	3375/990	124/125	1904-5118		
T	C	C	C	C	C	C	C	C	C	C	10	4629/1110	124/125	1904-5118		
C	T	T	C	T	T	C	C	C	C	C	11	5076/1230	124/125	1904-5118		
G	G	G	G	G	G	G	G	G	G	G	12	11834/1350	124/125	11517-12060		
G	G	G	G	C	G	C	C	C	C	G	13	22396/1470	124/125	22354-22775		
C	C	T	C	C	C	C	C	C	C	C	14	22670/1590	124/125	22354-22775		
A	A	G	A	A	A	A	A	A	A	A	15	28129/1710	124/125	27912-30306		
G	G	G	G	G	G	G	G	G	G	G	16	28318/1830	124/125	27912-30306		
G	G	G	G	G	G	G	A	G	G	G	17	28423/1950	124/125	27912-30306		
T	T	T	T	T	T	T	C	T	T	T	18	28479/2070	124/125	27912-30306		
G	G	G	C	C	G	G	G	G	G	G	19	28743/2190	124/125	27912-30306		
G	G	G	G	G	G	G	G	G	G	G	20	28980/2310	124/125	27912-30306		
C	C	C	C	C	C	C	C	C	C	C	21	29407/2430	124/125	27912-30306		
A	A	A	A	A	A	A	A	A	A	A	22	29500/2550	124/125	27912-30306		
G	G	G	G	G	G	G	G	A	G	G	23	29614/2670	124/125	27912-30306		
C	C	C	C	T	C	C	C	C	C	C	24	29796/2790	124/125	27912-30306		

^aAlleles for PCDH2 haplotypes are presented 5' to 3' in each column^bPS = polymorphic site;5 ^cPosition 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;10 ^d1st 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, PS1-24, to facilitate electronic searching of the haplotypes;15 ^eRegion examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

SEQ ID NO:124 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:125 is a modified version of SEQ ID NO:124 that shows the context sequence of each polymorphic site 1-24 in a uniform format to facilitate electronic searching of the PCDH2 haplotypes. For each polymorphic site, SEQ ID NO:125 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.

20 Table 6 below shows the percent of chromosomes characterized by a given PCDH2 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 PCDH2 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 NA = Native American.

5 Table 6. Frequency of Observed PCDH2 Haplotypes In Unrelated Individuals

	HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
	1	3438604	0.61	2.38	0.0	0.0	0.0	0.0
	2	3438598	0.61	0.0	0.0	0.0	0.0	16.67
10	3	3438592	0.61	0.0	0.0	0.0	2.78	0.0
	4	3438591	1.22	0.0	5.0	0.0	0.0	0.0
	5	3438599	0.61	0.0	2.5	0.0	0.0	0.0
	6	3438600	0.61	0.0	2.5	0.0	0.0	0.0
	7	3438606	0.61	0.0	2.5	0.0	0.0	0.0
15	8	3438583	12.8	14.29	10.0	0.0	27.78	16.67
	9	3438589	1.83	2.38	2.5	0.0	2.78	0.0
	10	3438597	0.61	0.0	0.0	0.0	0.0	16.67
	11	3438612	0.61	0.0	2.5	0.0	0.0	0.0
	12	3438590	1.22	0.0	0.0	2.5	2.78	0.0
20	13	3438585	8.54	7.14	12.5	12.5	2.78	0.0
	14	3438610	0.61	0.0	0.0	0.0	2.78	0.0
	15	3438595	0.61	0.0	2.5	0.0	0.0	0.0
	16	3438584	12.2	14.29	10.0	10.0	13.89	16.67
	17	3438593	0.61	0.0	0.0	0.0	2.78	0.0
25	18	3438608	0.61	0.0	2.5	0.0	0.0	0.0
	19	3438586	3.05	0.0	10.0	0.0	2.78	0.0
	20	3438609	0.61	2.38	0.0	0.0	0.0	0.0
	21	3438581	42.07	50.0	12.5	72.5	33.33	33.33
	22	3438594	0.61	0.0	2.5	0.0	0.0	0.0
30	23	3438603	0.61	0.0	2.5	0.0	0.0	0.0
	24	3438607	0.61	0.0	0.0	2.5	0.0	0.0
	25	3438611	0.61	2.38	0.0	0.0	0.0	0.0
	26	3438605	0.61	0.0	2.5	0.0	0.0	0.0
	27	3438601	0.61	0.0	2.5	0.0	0.0	0.0
35	28	3438588	1.83	0.0	5.0	0.0	2.78	0.0
	29	3438602	0.61	0.0	0.0	0.0	2.78	0.0
	30	3438587	2.44	4.76	5.0	0.0	0.0	0.0
	31	3438596	0.61	0.0	2.5	0.0	0.0	0.0

Table 7. Frequency of Observed PCDH2 Haplotype Pairs In Unrelated Individuals

	HAP1	HAP2	Total	CA	AF	AS	HL	AM
5	21	21	23.17	23.81	0.0	50.0	22.22	0.0
	16	16	1.22	0.0	5.0	0.0	0.0	0.0
	8	28	1.22	0.0	0.0	0.0	5.56	0.0
	8	17	1.22	0.0	0.0	0.0	5.56	0.0
	21	10	1.22	0.0	0.0	0.0	0.0	33.33
10	21	2	1.22	0.0	0.0	0.0	0.0	33.33
	28	9	1.22	0.0	5.0	0.0	0.0	0.0
	19	5	1.22	0.0	5.0	0.0	0.0	0.0
	21	6	1.22	0.0	5.0	0.0	0.0	0.0
	21	12	1.22	0.0	0.0	5.0	0.0	0.0
15	21	1	1.22	4.76	0.0	0.0	0.0	0.0
	21	9	1.22	4.76	0.0	0.0	0.0	0.0
	8	9	1.22	0.0	0.0	0.0	5.56	0.0
	8	3	1.22	0.0	0.0	0.0	5.56	0.0
	16	19	2.44	0.0	10.0	0.0	0.0	0.0
20	8	31	1.22	0.0	5.0	0.0	0.0	0.0
	21	24	1.22	0.0	0.0	5.0	0.0	0.0
	16	12	1.22	0.0	0.0	0.0	5.56	0.0
	8	14	1.22	0.0	0.0	0.0	5.56	0.0
	8	19	1.22	0.0	0.0	0.0	5.56	0.0
25	21	29	1.22	0.0	0.0	0.0	5.56	0.0
	21	30	1.22	4.76	0.0	0.0	0.0	0.0
	13	25	1.22	4.76	0.0	0.0	0.0	0.0
	21	7	1.22	0.0	5.0	0.0	0.0	0.0
	8	18	1.22	0.0	5.0	0.0	0.0	0.0
30	21	28	1.22	0.0	5.0	0.0	0.0	0.0
	13	27	1.22	0.0	5.0	0.0	0.0	0.0
	30	11	1.22	0.0	5.0	0.0	0.0	0.0
	16	13	3.66	9.52	0.0	5.0	0.0	0.0
	19	23	1.22	0.0	5.0	0.0	0.0	0.0
35	8	13	2.44	0.0	5.0	0.0	5.56	0.0
	13	26	1.22	0.0	5.0	0.0	0.0	0.0
	21	20	1.22	4.76	0.0	0.0	0.0	0.0
	21	8	6.1	19.05	0.0	0.0	5.56	0.0
	8	30	2.44	4.76	5.0	0.0	0.0	0.0
40	21	16	9.76	14.29	0.0	15.0	11.11	0.0
	8	16	4.88	4.76	0.0	0.0	11.11	33.33
	21	13	7.32	0.0	10.0	20.0	0.0	0.0
	4	15	1.22	0.0	5.0	0.0	0.0	0.0
	4	22	1.22	0.0	5.0	0.0	0.0	0.0

45

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 PCDH2 gene are likely to be similar to the relative frequencies of these PCDH2 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 10 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 15 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 20 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 protocadherin 2 (cadherin-like 2) (PCDH2) gene of an individual, which comprises determining which of the PCDH2 haplotypes shown in the table immediately below defines one copy of the individual's PCDH2 gene, wherein each of the PCDH2 haplotypes comprises a set of polymorphisms whose locations in SEQ ID NO:124 and identities are set forth in the table immediately below:

Haplotype Number ^a															PS Number ^b	PS Position ^c
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1	2027
C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137
G	A	A	A	A	G	G	G	G	G	G	G	G	G	G	3	2160
C	C	C	C	C	A	A	C	C	C	C	C	C	C	C	4	2393
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	5	2427
G	G	G	G	G	G	G	C	C	C	G	G	G	G	G	6	3096
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	7	3253
G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	9	3375
T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	10	4629
T	T	T	T	T	T	C	C	T	C	C	C	C	C	C	11	5076
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	12	11834
G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	13	22396
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	14	22670
A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	15	28129
A	G	G	G	G	G	G	G	G	A	G	G	G	G	G	16	28318
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	17	28423
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	18	28479
G	C	C	G	C	G	G	C	G	C	G	C	G	G	G	19	28743
G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	20	28980
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407
A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	22	29500
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	23	29614
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	24	29796

Haplotype Number ^a															PS Number ^b	PS Position ^c	
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	1	2027
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137
G	G	G	G	G	G	G	G	A	A	G	G	G	G	G	G	3	2160
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	4	2393
A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	5	2427
G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	6	3096
T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	7	3253
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258
C	C	C	C	C	C	T	T	C	C	C	C	C	C	C	C	9	3375
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	C	10	4629
T	T	T	T	C	C	T	T	C	T	T	C	C	C	C	C	11	5076
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	12	11834
G	G	G	G	G	G	G	G	G	C	G	C	C	C	C	G	13	22396
C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	14	22670
A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	15	28129
A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	16	28318
G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	17	28423
T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	18	28479
G	G	C	G	G	G	G	G	C	C	G	G	G	G	G	G	19	28743
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	20	28980
C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407
A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	22	29500
G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	23	29614
C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	24	29796

^aAlleles for haplotypes are presented 5' to 3' in each column^bPS = polymorphic site;^cPosition of PS within SEQ ID NO:124.

5

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS24 on the one copy of the individual's PCDH2 gene.
3. A method for haplotyping the protocadherin 2 (cadherin-like 2) (PCDH2) gene of an individual, which comprises determining which of the PCDH2 haplotype pairs shown in the table immediately below defines both copies of the individual's PCDH2 gene, wherein each of the PCDH2 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations in SEQ ID NO:124 and identities are set forth in the table immediately below:
- 10

15

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/21	16/16	8/28	8/17	21/10	21/2	28/9	19/5	21/6	21/12	21/1			
C/C	C/C	C/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	2	2137	
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/A	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	G/G	C/G	C/G	G/C	G/G	G/C	G/G	G/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/T	C/C	C/C	C/C	T/C	T/C	C/C	C/C	T/C	T/C	T/T	10	4629	
C/C	T/T	C/C	C/T	C/T	C/T	C/C	T/T	C/T	C/C	C/T	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/C	G/G	G/G	G/G	C/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	A/A	G/G	G/A	G/A	G/G	G/G	G/G	G/G	G/G	G/A	16	28318	
G/G	G/G	G/A	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/G	G/G	C/G	C/G	G/C	G/C	G/G	G/C	G/G	G/C	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A	A/A	A/G	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/9	8/9	8/3	16/19	8/31	21/24	16/12	8/14	8/19	21/29	21/30			
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/T	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	5	2427	
G/C	C/C	C/G	G/G	C/G	G/C	G/G	C/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/C	C/C	C/C	C/C	C/C	T/C	C/C	C/C	C/C	T/C	T/C	10	4629	
C/C	C/C	C/T	T/T	C/C	C/C	T/C	C/C	C/T	C/C	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C	G/C	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	G/G	G/G	A/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	18	28479	
G/G	C/G	C/C	G/G	C/G	G/C	G/C	C/G	C/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/G	A/G	A/G	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
13/25	21/7	8/18	21/28	13/27	30/11	16/13	19/23	8/13	13/26	21/20			
C/T	C/C	C/C	C/T	C/T	T/C	C/C	C/C	C/C	C/T	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	2160	
C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	G/G	C/G	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	9	3375	
C/C	T/C	C/C	T/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	10	4629	
C/T	C/T	C/C	C/C	C/C	T/C	T/T	C/C	C/T	C/C	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	12	11834	
G/C	G/G	G/G	G/C	G/C	C/C	G/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	15	28129	
G/G	G/G	G/G	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/C	G/G	C/C	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	24	29796	
Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/8	8/30	21/16	8/16	21/13	4/15	4/22	8/1	16/9	16/1	4/4	22/15		
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	1	2027
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/C	T/T	T/T	2	2137
G/G	G/G	G/G	G/G	G/G	A/G	A/G	G/G	G/G	G/G	A/A	G/G	3	2160
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G	6	3096
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	9	3375
T/C	C/C	T/C	C/C	T/C	C/C	C/C	T/C	C/C	C/T	C/C	C/C	10	4629
C/C	C/C	C/T	C/T	C/C	T/C	T/T	T/C	C/T	T/T	T/T	C/T	11	5076
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834
G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	13	22396
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/G	15	28129
G/G	G/G	G/A	G/A	G/G	G/G	G/G	G/A	G/A	A/A	G/G	G/G	16	28318
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/G	G/G	G/G	G/G	19	28743
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796

^aHaplotype 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;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:124.

5

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS24 on both copies of the individual's PCDH2 gene.
5. A method for genotyping the protocadherin 2 (cadherin-like 2) (PCDH2) gene of an individual, comprising determining for the two copies of the PCDH2 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, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:124.
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 PCDH2 gene, or a fragment thereof, that are present in the individual;
 - (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
 - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
 - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
 - (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
7. The method of claim 5, which comprises determining for the two copies of the PCDH2 gene present in the individual the identity of the nucleotide pair at each of PS1-PS24.
8. A method for haplotyping the protocadherin 2 (cadherin-like 2) (PCDH2) gene of an individual which comprises determining, for one copy of the PCDH2 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, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:124.
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 PCDH2 gene, or a fragment thereof, that is present in the individual;
 - (b) amplifying from the nucleic acid molecule a target region containing the selected

- 5 polymorphic site;
- (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
 - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
 - 10 (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
10. A method for predicting a haplotype pair for the protocadherin 2 (cadherin-like 2) (PCDH2) gene of an individual comprising:
- (a) identifying a PCDH2 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, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:124;
 - 5 (b) enumerating all possible haplotype pairs which are consistent with the genotype;
 - (c) comparing the possible haplotype pairs to the haplotype pair data set forth in the table immediately below; and
 - 10 (d) assigning a haplotype pair to the individual that is consistent with the data

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/21	16/16	8/28	8/17	21/10	21/2	28/9	19/5	21/6	21/12	21/1			
C/C	C/C	C/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	2	2137	
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/A	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	G/G	C/G	C/G	G/C	G/G	G/C	G/G	G/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/T	C/C	C/C	T/C	T/C	C/C	C/C	T/C	T/C	T/T	T/T	10	4629	
C/C	T/T	C/C	C/T	C/T	C/C	T/T	C/T	C/C	C/T	C/T	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/C	G/G	G/G	G/G	C/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	A/A	G/G	G/A	G/A	G/G	G/G	G/G	G/G	G/G	G/A	16	28318	
G/G	G/G	G/A	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/G	G/G	C/G	C/G	G/C	G/C	G/G	G/C	G/G	G/C	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A	A/A	A/G	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/9	8/9	8/3	16/19	8/31	21/24	16/12	8/14	8/19	21/29	21/30			
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/T	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	5	2427	
G/C	C/C	C/G	G/G	C/G	G/C	G/G	C/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/C	C/C	C/C	C/C	C/C	T/C	C/C	C/C	C/C	T/C	T/C	10	4629	
C/C	C/C	C/T	T/T	C/C	C/C	T/C	C/C	C/T	C/C	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C	G/C	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	G/G	G/G	A/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	18	28479	
G/G	C/G	C/C	G/G	C/G	G/C	G/C	C/G	C/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
13/25	21/7	8/18	21/28	13/27	30/11	16/13	19/23	8/13	13/26	21/20			
C/T	C/C	C/C	C/T	C/T	T/C	C/C	C/C	C/C	C/T	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	2160	
C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	G/G	C/G	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	9	3375	
C/C	T/C	C/C	T/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	10	4629	
C/T	C/T	C/C	C/C	C/C	C/C	T/C	T/T	C/C	C/T	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	12	11834	
G/C	G/G	G/G	G/C	G/C	C/C	G/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	15	28129	
G/G	G/G	G/G	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/C	G/G	C/C	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	24	29796	

Haplotype Pair Number ^a													PS Number ^b	PS Position ^c
21/8	8/30	21/16	8/16	21/13	4/15	4/22	8/1	16/9	16/1	4/4	22/15			
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/C	T/T	T/T	2	2137	
G/G	G/G	G/G	G/G	G/G	A/G	A/G	G/G	G/G	G/G	A/A	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	9	3375	
T/C	C/C	T/C	C/C	T/C	C/C	C/C	T/C	C/C	C/T	C/C	C/C	10	4629	
C/C	C/C	C/T	C/T	C/C	T/C	T/T	T/C	C/T	T/T	T/T	C/T	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/G	15	28129	
G/G	G/G	G/A	G/A	G/G	G/G	G/G	G/A	G/A	A/A	G/G	G/G	16	28318	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/G	G/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

^aHaplotype 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;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:124.

11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS24, which have the location and alternative alleles shown in SEQ ID NO:124.
- 5 12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the protocadherin 2 (cadherin-like 2) (PCDH2) 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 1-31 shown in the table presented immediately below, wherein each of the haplotypes comprises a set of polymorphisms whose locations in SEQ ID NO:124 and identities are set forth in the table immediately below:
- 10

Haplotype Number ^a															PS Number ^b	PS Position ^c
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1	2027
C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137
G	A	A	A	A	G	G	G	G	G	G	G	G	G	G	3	2160
C	C	C	C	C	A	A	C	C	C	C	C	C	C	C	4	2393
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	5	2427
G	G	G	G	G	G	C	C	C	G	G	G	G	G	G	6	3096
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	7	3253
G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	9	3375
T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	10	4629
T	T	T	T	T	T	C	C	T	C	C	C	C	C	C	11	5076
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	12	11834
G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	13	22396
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	14	22670
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	15	28129
A	G	G	G	G	G	G	G	G	A	G	G	G	G	G	16	28318
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	17	28423
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	18	28479
G	C	C	G	C	G	C	G	C	G	C	G	G	G	G	19	28743
G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	20	28980
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407
A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	22	29500
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	23	29614
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	24	29796

Haplotype Number ^a															PS Number ^b	PS Position ^c	
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	1	2027
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137	
G	G	G	G	G	G	G	G	G	A	A	G	G	G	G	3	2160	
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	4	2393	
A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	5	2427	
G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	6	3096	
T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	7	3253	
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258	
C	C	C	C	C	C	T	T	C	C	C	C	C	C	C	9	3375	
C	C	C	T	T	C	C	C	C	C	C	C	C	C	C	10	4629	
T	T	T	C	C	T	T	C	T	T	C	C	C	C	C	11	5076	
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	11834	
G	G	G	G	G	G	G	G	G	C	G	C	C	C	C	13	22396	
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	14	22670	
A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	15	28129	
A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	16	28318	
G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	17	28423	
T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	18	28479	
G	G	C	G	G	G	G	G	C	C	G	G	G	G	G	19	28743	
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	20	28980	
C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407	
A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	22	29500	
G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	23	29614	
C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	24	29796	

15

^aAlleles for haplotypes are presented 5' to 3' in each column^bPS = polymorphic site;^cPosition of PS in SEQ ID NO:124;

20

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the PCDH2 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations in SEQ ID NO:124 and identities are set forth in the table immediately below:

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/21	16/16	8/28	8/17	21/10	21/2	28/9	19/5	21/6	21/12	21/1			
C/C	C/C	C/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	2	2137	
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/A	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	G/G	C/G	C/G	G/C	G/G	G/C	G/G	G/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/T	C/C	C/C	C/C	T/C	T/C	C/C	C/C	T/C	T/C	T/T	10	4629	
C/C	T/T	C/C	C/T	C/T	C/T	C/C	T/T	C/T	C/C	C/T	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/C	G/G	G/G	G/G	C/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	A/A	G/G	G/A	G/A	G/G	G/G	G/G	G/G	G/G	G/A	16	28318	
G/G	G/G	G/A	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/G	G/G	C/G	C/G	G/C	G/C	G/G	G/C	G/G	G/C	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A	A/A	A/G	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/9	8/9	8/3	16/19	8/31	21/24	16/12	8/14	8/19	21/29	21/30			
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/T	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	5	2427	
G/C	C/C	C/G	G/G	C/G	G/C	G/G	C/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/C	C/C	C/C	C/C	T/C	C/C	C/C	C/C	C/C	T/C	T/C	10	4629	
C/C	C/C	C/T	T/T	C/C	C/C	T/C	C/C	C/T	C/C	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C	G/C	G/C	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	G/G	G/G	A/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	18	28479	
G/G	C/G	C/C	G/G	C/G	G/C	G/G	C/G	C/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	
Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
13/25	21/7	8/18	21/28	13/27	30/11	16/13	19/23	8/13	13/26	21/20			
C/T	C/C	C/C	C/T	C/T	T/C	C/C	C/C	C/C	C/T	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	2160	
C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	C/G	G/G	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	9	3375	
C/C	T/C	C/C	T/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	10	4629	
C/T	C/T	C/C	C/C	C/C	C/C	T/C	T/T	C/C	C/T	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	12	11834	
G/C	G/G	G/G	G/C	G/C	C/C	G/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/G	A/A	A/A	15	28129	
G/G	G/G	G/G	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/C	G/G	C/C	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	24	29796	

Haplotype Pair Number ^a													PS Number ^b	PS Position ^c
21/8	8/30	21/16	8/16	21/13	4/15	4/22	8/1	16/9	16/1	4/4	22/15			
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/C	T/T	T/T	2	2137	
G/G	G/G	G/G	G/G	G/G	A/G	A/G	G/G	G/G	G/G	A/A	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	9	3375	
T/C	C/C	T/C	C/C	T/C	C/C	C/C	T/C	C/C	C/T	C/C	C/C	10	4629	
C/C	C/C	C/T	C/T	C/C	T/C	T/T	T/C	C/T	T/T	T/T	C/T	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/G	15	28129	
G/G	G/G	G/A	G/A	G/G	G/G	G/G	G/A	G/A	A/A	G/G	G/G	16	28318	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/G	G/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

^aHaplotype 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;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:124;

30

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 PCDH2.
14. An isolated genotyping oligonucleotide for detecting a polymorphism in the protocadherin 2 (cadherin-like 2) (PCDH2) 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, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:124.
15. The genotyping oligonucleotide of claim 14, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the PCDH2 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:4-27, the complements of SEQ ID NOS:4-27, and SEQ ID NOS:28-75.
17. The genotyping oligonucleotide of claim 14, which is a primer-extension oligonucleotide.

18. The the primer extension oligonucleotide of claim 17, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:76-123.
19. A kit for genotyping the protocadherin 2 (cadherin-like 2) (PCDH2) gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23 and PS24, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:124.
20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
- (a) a first nucleotide sequence which comprises a protocadherin 2 (cadherin-like 2) (PCDH2) isogene, wherein the PCDH2 isogene is selected from the group consisting of isogenes 1-20 and 22 - 31 shown in the table immediately below and wherein each of the isogenes comprises the regions of the SEQ ID NOS shown in the table immediately below and wherein each of the isogenes 1- 20 and 22 - 31 is further defined by the corresponding set of polymorphisms whose locations in SEQ ID NO:124 and identities are set forth in the table immediately below

Isogene Number ^a															PS Number ^b	PS Position ^c	Region Examined ^d
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1	2027	1904-5118
C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137	1904-5118
G	A	A	A	A	G	G	G	G	G	G	G	G	G	G	3	2160	1904-5118
C	C	C	C	C	A	A	C	C	C	C	C	C	C	C	4	2393	1904-5118
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	5	2427	1904-5118
G	G	G	G	G	G	G	C	C	C	G	G	G	G	G	6	3096	1904-5118
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	7	3253	1904-5118
G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258	1904-5118
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	9	3375	1904-5118
T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	10	4629	1904-5118
T	T	T	T	T	T	C	C	T	C	C	C	C	C	C	11	5076	1904-5118
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	12	11834	11517-12060
G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	13	22396	22354-22775
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	14	22670	22354-22775
A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	15	28129	27912-30306
A	G	G	G	G	G	G	G	G	A	G	G	G	G	G	16	28318	27912-30306
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	17	28423	27912-30306
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	18	28479	27912-30306
G	C	C	G	C	G	G	C	G	C	G	C	G	G	G	19	28743	27912-30306
G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	20	28980	27912-30306
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407	27912-30306
A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	22	29500	27912-30306
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	23	29614	27912-30306
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	24	29796	27912-30306

Isogene Number ^a															PS Number ^b	PS Position ^c	Region Examined ^d
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	1	2027	1904-5118
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137	1904-5118
G	G	G	G	G	G	G	G	A	A	G	G	G	G	G	3	2160	1904-5118
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	4	2393	1904-5118
A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	5	2427	1904-5118
G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	6	3096	1904-5118
T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	7	3253	1904-5118
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258	1904-5118
C	C	C	C	C	C	T	T	C	C	C	C	C	C	C	9	3375	1904-5118
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	10	4629	1904-5118
T	T	T	T	C	C	T	T	C	T	T	C	C	C	C	11	5076	1904-5118
G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	12	11834	11517-12060
G	G	G	G	G	G	G	G	G	C	G	C	C	C	C	13	22396	22354-22775
C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	14	22670	22354-22775
A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	15	28129	27912-30306
A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	16	28318	27912-30306
G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	17	28423	27912-30306
T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	18	28479	27912-30306
G	G	C	G	G	G	G	G	C	C	G	G	G	G	G	19	28743	27912-30306
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	20	28980	27912-30306
C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407	27912-30306
A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	22	29500	27912-30306
G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	23	29614	27912-30306
C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	24	29796	27912-30306

^aAlleles for isogenes are presented 5' to 3' in each column^bPS = polymorphic site;^cPosition of PS in SEQ ID NO:124;^dRegion examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:124 of the sequenced region;

- (b) a second nucleotide sequence which comprises a fragment of the first nucleotide sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at PS1, cytosine at PS2, adenine at PS3, adenine at PS4, guanine at PS5, cytosine at PS6, cytosine at PS7, adenine at PS8, thymine at PS9, cytosine at PS10, thymine at PS11, adenine at PS12, cytosine at PS13, thymine at PS14, guanine at PS15, adenine at PS16, adenine at PS17, cytosine at PS18, cytosine at PS19, adenine at PS20, thymine at PS21, guanine at PS22, adenine at PS23 and thymine at PS24, wherein the selected polymorphism has the location set forth in the table immediately above; and
- (c) a third nucleotide sequence which is complementary to the first or second nucleotide sequence.
21. The isolated polynucleotide of claim 20, which is a DNA molecule and comprises both the first and third 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 a PCDH2 protein encoded by the first nucleotide sequence.
23. The recombinant organism of claim 22, which is a nonhuman transgenic animal.
24. The isolated polynucleotide of claim 20 which consists of the second nucleotide sequence.
25. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
 - (a) a coding sequence for a protocadherin 2 (cadherin-like 2) (PCDH2) isogene wherein the coding sequence is defined by a haplotype selected from the group consisting of 2c, 3c-20c, and 22c-31c shown in the table immediately below and wherein the coding sequence comprises SEQ ID NO:2 except at each of the polymorphic sites which have the locations and polymorphisms set forth in the table immediately below:

Coding Sequence Subhaplotype Number ^a								PS Number ^b	PS Position ^c
2c	3c-6c, 11c-14c, 16c-19c, 25c-26c, 28c-31c	7c, 22c 23c	8c, 9c, 10c, 24c	15c	20c	27c			
G	G	G	C	G	G	G	6	582	
T	T	T	T	T	T	C	7	739	
A	G	G	G	G	G	G	8	744	
C	C	T	C	C	C	C	9	861	
C	C	C	C	C	T	C	10	2115	
G	G	G	G	G	A	G	12	2479	
A	A	A	A	G	A	A	15	2646	

^aAlleles for coding sequence haplotypes are presented 5' to 3' in each column; the numerical portion of the coding sequence haplotype number represents the number of the parent full PCDH2 haplotype;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:2;

and

- (b) a fragment of the coding sequence, wherein the fragment comprises at least one polymorphism in SEQ ID NO:2 selected from the group consisting of cytosine at a position corresponding to nucleotide 582, cytosine at a position corresponding to nucleotide 739, adenine at a position corresponding to nucleotide 744, thymine at a position corresponding to nucleotide 861, cytosine at a position corresponding to nucleotide 2115, adenine at a position corresponding to nucleotide 2479 and guanine at a position corresponding to nucleotide 2646.
26. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 25, wherein the organism expresses a protocadherin 2 (cadherin-like 2) (PCDH2) protein encoded by the polymorphic variant sequence.
27. The recombinant organism of claim 26, which is a nonhuman transgenic animal.
28. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a

reference sequence for the protocadherin 2 (cadherin-like 2) (PCDH2) protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of proline at a position corresponding to amino acid position 247 and serine at a position corresponding to amino acid position 827.

29. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 28.
30. A method for screening for drugs targeting the isolated polypeptide of claim 28 which comprises contacting the PCDH2 polymorphic variant with a candidate agent and assaying for binding activity.
31. A computer system for storing and analyzing polymorphism data for the protocadherin 2 (cadherin-like 2) 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;

5 wherein the polymorphism data comprises the haplotypes set forth in the table immediately below:

Haplotype Number ^a															PS Number ^b	PS Position ^c
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1	2027
C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137
G	A	A	A	A	G	G	G	G	G	G	G	G	G	G	3	2160
C	C	C	C	C	A	A	C	C	C	C	C	C	C	C	4	2393
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	5	2427
G	G	G	G	G	G	C	C	C	G	G	G	G	G	G	6	3096
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	7	3253
G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	9	3375
T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	10	4629
T	T	T	T	T	T	C	C	T	C	C	C	C	C	C	11	5076
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	12	11834
G	G	G	G	G	G	G	G	G	C	G	G	G	G	G	13	22396
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	14	22670
A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	15	28129
A	G	G	G	G	G	G	G	A	G	G	G	G	G	G	16	28318
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	17	28423
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	18	28479
G	C	C	G	C	G	C	G	C	G	C	G	G	G	G	19	28743
G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	20	28980
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407
A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	22	29500
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	23	29614
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	24	29796

Haplotype Number ^a																PS Number ^b	PS Position ^c
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	1	2027
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137
G	G	G	G	G	G	G	G	A	A	G	G	G	G	G	G	3	2160
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	4	2393
A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	5	2427
G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	6	3096
T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	7	3253
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258
C	C	C	C	C	C	T	T	C	C	C	C	C	C	C	C	9	3375
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	C	10	4629
T	T	T	T	C	C	T	T	C	T	T	C	C	C	C	C	11	5076
G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	12	11834
G	G	G	G	G	G	G	G	G	C	G	C	C	C	C	G	13	22396
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	14	22670
A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	15	28129
A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	16	28318
G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	17	28423
T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	18	28479
G	G	C	G	G	G	G	G	C	C	G	G	G	G	G	G	19	28743
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	20	28980
C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407
A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	22	29500
G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	23	29614
C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	24	29796

^aAlleles for haplotypes are presented 5' to 3' in each column

10

^bPS = polymorphic site;^cPosition of PS in SEQ ID NO:124;

and the haplotype pairs set forth in the table immediately below:

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/21	16/16	8/28	8/17	21/10	21/2	28/9	19/5	21/6	21/12	21/1			
C/C	C/C	C/T	C/C	C/C	C/C	T/C	C/C	C/C	C/C	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	2	2137	
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/A	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	G/G	C/G	C/G	G/C	G/G	G/C	G/G	G/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/T	C/C	C/C	C/C	T/C	T/C	C/C	C/C	T/C	T/C	T/T	10	4629	
C/C	T/T	C/C	C/T	C/T	C/T	C/C	T/T	C/T	C/C	C/T	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/C	G/G	G/G	G/G	C/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/T	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	A/A	G/G	G/A	G/A	G/G	G/G	G/G	G/G	G/G	G/A	16	28318	
G/G	G/G	G/A	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/G	G/G	C/G	C/G	G/C	G/G	G/C	G/C	G/G	G/C	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A	A/A	A/G	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	
Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/9	8/9	8/3	16/19	8/31	21/24	16/12	8/14	8/19	21/29	21/30			
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/T	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	2160	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	5	2427	
G/C	C/C	C/G	G/G	C/G	G/C	G/G	C/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	9	3375	
T/C	C/C	C/C	C/C	C/C	T/C	C/C	C/C	C/C	T/C	T/C	10	4629	
C/C	C/C	C/T	T/T	C/C	C/C	T/C	C/C	C/T	C/C	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C	G/C	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	15	28129	
G/G	G/G	G/G	A/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	18	28479	
G/G	C/G	C/C	G/G	C/G	G/C	G/C	C/G	G/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796	

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
13/25	21/7	8/18	21/28	13/27	30/11	16/13	19/23	8/13	13/26	21/20			
C/T	C/C	C/C	C/T	C/T	T/C	C/C	C/C	C/C	C/T	C/C	1	2027	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	2	2137	
G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	3	2160	
C/C	C/A	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427	
G/G	G/G	C/G	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	6	3096	
T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T	T/T	7	3253	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258	
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	9	3375	
C/C	T/C	C/C	T/C	C/C	C/C	C/C	C/C	C/C	C/C	T/T	10	4629	
C/T	C/T	C/C	C/C	C/C	T/C	T/T	C/C	C/T	C/C	C/C	11	5076	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/A	12	11834	
G/C	G/G	G/G	G/C	G/C	C/C	G/G	G/G	G/G	G/G	G/G	13	22396	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	14	22670	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	15	28129	
G/G	G/G	G/G	G/G	G/G	G/G	A/G	G/G	G/G	G/G	G/G	16	28318	
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423	
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479	
G/C	G/G	C/C	G/G	G/G	G/G	G/G	G/G	C/G	G/G	G/G	19	28743	
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	20	28980	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407	
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500	
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614	
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	24	29796	

15

Haplotype Pair Number ^a												PS Number ^b	PS Position ^c
21/8	8/30	21/16	8/16	21/13	4/15	4/22	8/1	16/9	16/1	4/4	22/15		
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	1	2027
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/C	T/T	T/T	2	2137
G/G	G/G	G/G	G/G	A/G	A/G	G/G	G/G	G/G	G/G	A/A	G/G	3	2160
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	4	2393
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	5	2427
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G	6	3096
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	7	3253
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	3258
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/T	9	3375
T/C	C/C	T/C	C/C	T/C	C/C	C/C	T/C	C/C	C/T	C/C	C/C	10	4629
C/C	C/C	C/T	C/C	T/C	T/T	T/C	T/C	T/T	T/T	C/T	C/T	11	5076
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	12	11834
G/G	G/C	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	13	22396
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	14	22670
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/G	15	28129
G/G	G/G	G/A	G/A	G/G	G/G	G/G	G/A	G/A	A/A	G/G	G/G	16	28318
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	28423
T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	18	28479
G/C	C/G	G/G	C/G	G/G	G/G	G/G	G/C	G/G	G/G	G/G	G/G	19	28743
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	28980
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	21	29407
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	22	29500
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	29614
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	24	29796

^aHaplotype 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;

^bLocation of PS in SEQ ID NO:124.

20

32. A genome anthology for the protocadherin 2 (cadherin-like 2) (PCDH2) gene which comprises PCDH2 isogenes defined by any one of haplotypes 1-31 set forth in the table shown below:

Haplotype Number ^a															PS Number ^b	PS Position ^c
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1	2027
C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137
G	A	A	A	A	G	G	G	G	G	G	G	G	G	G	3	2160
C	C	C	C	C	A	A	C	C	C	C	C	C	C	C	4	2393
A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	5	2427
G	G	G	G	G	G	G	C	C	C	G	G	G	G	G	6	3096
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	7	3253
G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	9	3375
T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	10	4629
T	T	T	T	T	T	T	C	C	T	C	C	C	C	C	11	5076
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	12	11834
G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	13	22396
C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	14	22670
A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	15	28129
A	G	G	G	G	G	G	G	G	A	G	G	G	G	G	16	28318
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	17	28423
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	18	28479
G	C	C	G	C	G	G	C	G	C	G	C	G	G	G	19	28743
G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	20	28980
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407
A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	22	29500
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	23	29614
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	24	29796

Haplotype Number ^a															PS Number ^b	PS Position ^b	
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	1	2027
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2	2137
G	G	G	G	G	G	G	G	A	A	G	G	G	G	G	G	3	2160
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	4	2393
A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	5	2427
G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	6	3096
T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	7	3253
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	8	3258
C	C	C	C	C	T	T	C	C	C	C	C	C	C	C	C	9	3375
C	C	C	T	T	C	C	C	C	C	C	C	C	C	C	C	10	4629
T	T	T	C	C	T	T	C	T	T	C	C	C	C	C	C	11	5076
G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	12	11834
G	G	G	G	G	G	G	G	C	G	C	C	C	C	G	G	13	22396
C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	14	22670
A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	15	28129
A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	16	28318
G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	17	28423
T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	T	18	28479
G	G	C	G	G	G	G	C	C	G	G	G	G	G	G	G	19	28743
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	20	28980
C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	21	29407
A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	22	29500
G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	23	29614
C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	24	29796

^aAlleles for haplotypes are presented 5' to 3' in each column

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:124.

1/14

POLYMORPHISMS IN THE PCDH2 GENE

CACATGGGCC	CGTAGGGAGA	GCTTTAGAAC	TAGTGCAGCT	TAATTGAGCA	
GGGGGAAGCC	TTCTGCAGAC	AGTTGAGCCT	AGCGGAGGTT	GGTTGACACG	100
GAAGGGGCTT	GATTTTGAG	AAGGTGAGAC	TTAAATTGAG	GATTGCCAAG	
GATTGGTATG	TTAATCCAAG	GGCCGTTTAT	GCGAATTAAG	GACCAGATT	200
AAAAATTGGG	GAACCCAAGT	TAAAATGGTT	TGGGAACGTC	TAACCTTCGG	
TTATTGTTGG	GGGAATGGCT	AACCCAAAGG	GGTCTTATTG	GTTCATGAAG	300
AGGTCTTAAG	AAAACATATGA	GCTGGGTGCA	GTAGCTCAGG	CCTGTAGTCC	
CATCTACTTA	GAAGTCTGAG	CGGGGAGGAT	CCCTTGAGCC	CAAGGAGCTG	400
GAAGTTACAG	TGAGCTATGA	TTATTCCACT	GTACTCCAGC	CTGGATGACA	
GAGCGAGATC	CTGCCCTCAGA	AACAAAAAAC	AAAAAAATAAA	CAACAAACAAA	500
AACAAACAACA	ACGAAAATCT	CAGAAAACCA	CAGGCCTTAT	TATCAAGAAA	
ATGACAGGCC	TTATTATTGA	AAAAACTACT	TACATTTTTA	AAAGCTGATT	600
TTCAAAAACC	AGATTGTGAA	TTAAGTGATG	CTGAATAGGG	GAAAAAAAGA	
CTGATAAACCC	AGTGTATCTC	TGAATAACCC	ACTAGATCAG	TGCTTGTCAA	700
ACTTTAATGT	TTTTCAAAT	CACCTGATG	TTTGTAAA	ATGTAAGTCA	
TGTCTCAGTA	GGTCTAGAGG	AGGGCATTTC	TGCAATTCTA	ACAAGCTCCC	800
CGGTGATGCC	TAGACCATTG	GTCCTCACAC	TTTGAGTAGC	AAGTGATTGG	
GGTATACTCT	TTATTCTAAA	TTCCAATGT	ATTCTATCTT	CTAATGCCTA	900
TCAGCCCCCT	TGCTGGTCTG	GCTAATTATC	AAACTTGAGC	ATTAAAAACT	
GGGGGTTTT	AGAGATTAAT	TGAAGTGT	AAGACATGGA	AATATAGGTC	1000
ACATATGGCT	TTTCATCTCT	TCCGGAAATA	TCCTATATAT	CCTTACTTCT	
TTTCTACCTA	CCTAACCTAT	GTGCCTACCC	CTTCACTTAT	GTTCAAGGTT	1100
CTATTATACT	TCAAAAGGAC	TCTATGCAAT	CAAAGTGATG	TCCAAGACAA	
ATATACAGAA	ATAGTATGAG	GCTCTAACAC	TTTACATCTA	ATGGAGTGGT	1200
TTGCATCATT	AATTGAGGCT	TACTCTGTG	CAGGCACCTT	ACCTGCCTTC	
TTCATTTAAT	AGGATAGGAG	CAGGAACACT	TAGAGGTTCA	TTTCTTGTTC	1300
AAGGATATAC	ATCTCATCTC	TGGCTTATAT	CACAGTGGAC	TCTTTTATT	
ACTATCCTAA	TCATTATAAT	TTTGTGTTG	GACAACAATT	CTAAATGCAG	1400
AAAAGTTAAG	GCTCTTATT	CTGTGACTAA	AGGCTTTGTA	TTTCTAAAG	
ATTTGTGGAG	CTTGCTGCC	ATTATTCAA	TAATTGTTATA	AAGATCCTAG	1500
GTGGTAAATA	GGCCTGGCTT	TTTTTCTTT	TGAAATCTTC	TCCGCTTCT	
TCTTCATTAG	CAGTGCAACT	TCTTGTTGCA	TATGGTACCC	AAGATCTTGG	1600
AAAAAGGATT	GATCAAGTAA	AGGGAAACGA	TAGAGAGAGT	GTGCTCTACA	
ACTGAGATCA	TGCTCCTTGT	GAAAGACAGG	GTAGGGAAAT	TTCTTATTGC	1700
TCCCTGGTTC	TTAAGAAATG	AGAGCTAAC	ATGACATCCA	GTCTTAAAG	
AAATAAAATC	TGAATTGTAC	TTTTAAATT	TATAAGCCAG	AGTTAAACCA	1800
ATAGATTGTA	ACTCCTGCTA	ATTTGATAGT	AGCTCCAGAT	AATGAGGAAG	
TGACAGGGAA	TGCTAATTTC	ATTTGGTAAG	AGATAAAACT	GAAATTATTA	1900
ATGTCTCCAC	GGAAATGCTA	AATGCCTCCT	TACAAGTAGG	GTCCGGCTAA	
TTGTCTGTCT	TCCAACAAAGC	CAGATTGTT	GGTGTGTTTC	CAGACATGT	2000
TGTGTAGTTT	TCGGTTGGT	TCGATCCCCT	TTCTTGTGAT	CAAAGAAAGT	
		T			
GATTCAAATG	TTTAATGAGT	CTTGATTGAG	ATTGGAGACT	TGCAGAACGG	2100
CTAGCCTCAC	AGCCCCAAGG	CTGGCTTCTC	CTAAGGTAGG	TTTCCTATGC	
		C			
ACCGATACTG	GCAAGGGCGCT	TTGGCTGGAA	ACTCTGGAAG	GAAGCCAAAG	2200
	A				
GAAAGTGAAG	TTCCCTGGCGC	TAGCGCGTGT	CCGTCTCAGA	GAGCCCGGCG	
CTAGCTCATT	CTTCGTGCA	TTATTGGCTG	GGACTCTGTG	TGCCGCTGTC	2300
GGCCAATGAA	GACGCTGGAG	ATCGGGCCCG	TGCCCGTCCC	CTTTCTGCGC	
CCCGGGATGA	GGCAGAGACT	GAACAGCCGG	CGAGCAAATC	AACGGCATCC	2400
	A				
AGAAAAGCCAT	GTCGGACTCG	GCGCCAGCG	CCCAAGCGCT	AACCCGCTGA	
	G				
AAGTTTCTCA	GCGAAATCTC	AGGGACGATC	TGGACCCCGC	TGAGAGGAAC	2500
TGCTTTGAG	TGAGATGGTC	CCAGAGGCCT	GGAGGAGCGG	ACTGGTAAGC	
[exon 1: 2515..]					

FIGURE 1A

	2/14	
ACCGGGAGGG TAGTGGGAGT TTTGCTCTG CTGGTGCCT TGAACAAGGC	2600	
TTCCACGGTC ATTCACTATG AGATCCCGA GAAAGAGAG AAGGGTTTCG	2700	
CTGTGGCAA CGTGGTCGCG AACCTGGTT TGATCTCG TAGCCTCTCA	2800	
GCCCCGAGGT TCCGGGTGGT GTCTGGAGCT AGCCGAAGAT TCTTGAGGT	2900	
GAACCGGGAG ACCGGAGAGA TGTTTGAA CGACCGTCTG GATCGAGAGG	3000	
AGCTGTGTGG GACACTGCC C TCTTGCAGT TAACTCTGGA GTTGGTAGTG	3100	
GAGAACCCGC TGGAGCTGTT CAGCGTGGAA GTGGTGATCC AGGACATCAA	C	
CGACAACAAT CCTGCTTCC CTACCCAGGA AATGAAATTG GAGATTAGCG	3200	
AGGCCGTGGC TCCGGGAGC CGCTTCCGC TCGAGAGCGC GCACGATCCC	3300	
GATGTGGGAA GCAACTCTT ACAAAACCTAT GAGCTGAGCC GAAATGAATA		
CTTTCGCGCTT CGCGTGCAGA CGCGGGAGGA CAGCACCAAG TACGCGGAGC		
	C	
TGGTGTGGGA GCGCGCCCTG GACCGAGAAC GGGAGCCTAG TCTCCAGTTA		
GTGCTGACGG CGTTGGACGG AGGGACCCA GCTCTCTCCG CCAGCCTGCC		
TATTACACATC AAGGTGCTGG ACGCGAATGA CAATGCGCCT GTCTCAACC		
AGTCCTTGTA CGGGCGCGC GTCCTGGAGG ATGCACCCCTC CGGCACCGCG		
	A	
GTGGTACAAG TCCTTGCAAC GGATCTGGAT GAAGGCCCCA ACGGTGAAAT		
TATTTACTCC TTGGCAGCC ACAACCGCGC CGGCGTGCAG CAACTATTG	3400	
	T	
CCTTAGACCT TGTAACCGGG ATGCTGACAA TCAAGGGTCG GCTGGACTTC	3500	
GAGGACACCA AACTCCATGA GATTTACATC CAGGCCAAAG ACAAGGGCGC		
CAATCCCGAA GGAGCACATT GCAAAGTGTGTT GTGGAGGTT GTGGATGTGA		
ATGACAACGC CCCGGAGATC ACAGTCACCT CCGTGTACAG CCCAGTACCC		
GAGGATGCCCT CTCTGGGAC TGTATCGCT TTGCTCAGTG TGACTGACCT		
GGATGCTGGC GAGAACGGGC TGGTGACCTG CGAAGTTCCA CGGGTCTCC	3700	
CTTTCAGCCT TACCTCTTCC CTCAAGAATT ACTTCACCTT GAAAACCACT		
GCAGACCTGG ATCGGGAGAC TGTGCCAGAA TACAACCTCA GCATCACCAG		
CCGAGACGCC GGAACCCCTT CCCTCTCAGC CCTTACAATA GTGCGTGTTC		
AAGTGTCCGA CATCAATGAC AACCTCCAC AATCTTCTCA ATCTTCCCTAC	3900	
GACGTTTACA TTGAAGAAAA CAACCTCCCC GGGGCTCCAA TACTAACCT		
AAGTGTCTGG GACCCGACG CCCCAGAA TGCTCGGCTT FCTTTCTTC		
TCTTGGAGCA AGGAGCTGAA ACCGGGCTAG TGGTGCCTA TTTACAATA		
AATCGTGACA ATGGCATAGT GTCATCCTTA GTGCCCTAG ACTATGAGGA	4100	
TCGGCGGGAA TTTGAATTAA CAGCTCATAT CAGCGATGGG GGCACCCCGG		
TCCTAGCCAC CAACATCAGC GTGAACATAT TTGTCACTGA TCGCAATGAC		
AATGCCCTCC AGGTCCCTATA TCCTCGGCCA GGTGGGAGCT CGGTGGAGAT		
GCTGCCTCGA GGTACCTCAG CTGGCCACCT AGTGTACCG GTGGTAGGCT	4300	
GGGACGCGGA TGCAGGGCAC AATGCCTGGC TCTCCTACAG TCTCTGGGA		
TCCCCTAAC AGAGCCTTT TGCCATAGGG CTGCACACTG GTCAAATCAG		
TACTGCCGT CCAGTCCAAG ACACAGATT ACCCAGGCAG ACTCTCACGG		
TCTTGATCAA AGACAATGGG GAGCCTCGC TCTCACCAC TGCTACCCCTC	4500	
ACTGTGTCAAG TAACCGAGGA CTCTCTGAA GCCCGAGCGC AGTCCCTCTC		
TGGCTCTGCC CCCCGGGAGC AGAAAAAAA TCTCACCTT TATCTACTTC		
TTTCCCTAAT CCTGGTTCT GTGGGGTTTG TGGTCACAGT GTTCGGAGTA	4600	
	C	
ATCATATTCA AAGTTTACAA GTGGAAGCAG TCTAGAGACC TATACCGAGC	4700	
CCCGGTGAGC TCACGTGAC GAAACACAGG GCCCTCCTTG CACGCGGACG		
CCGTGCGGGG AGGCCTGATG TCGCCGCACC TTTACCATCA GGTGTATCTC	4800	
ACCACGGACT CCCGCCCGG CGACCGCTG CTGAAGAAC CTGGTGCAGC		
CAGTCCACTG GCCAGCCGCC AGAACACCGT GCGGAGCTGT GATCCGGTGT		
TCTATAGGCA GGTGTTGGGT GCAGAGAGCG CCCCTCCCGG ACAGGTAAGG	4900	
	..4944]	
TTTAGCAAGT CATGCTTGAC CCTGTTAGTG CTTTTTATT CCTACATCAT	5000	
ATTGAGGAAG GAATGGAGCT GTTTTTAG TGATGAAGAT GTTTCTGG		
TGATGCATTC ACACTTCAA CTGGCCCTTC CTAGATCAA GTTAGTGCCT	5100	
	T	
TTGTGAGATG GTGGCCTGCC AGAGTGTGGT TTGTGGTCCC ATTCAGGGG		
GAAAGATACTT GACTCATCTG TGGACCTAAT TCACATCCTC AGCACTCTT	5200	

FIGURE 1B

3/14

TGCTATCACACTAAACCAAT	CTTGCTAAGG GATGGTTAAG	CTAAAACACA	
AGATCTCAGCGATCAGAGTT	TAGCTTGTATCAT	TAGGAATAAG	5300
CTGCTGGATA CCTCTAACCA	GTGGCAGCTT	CTAGGAATAAC	5400
TCATCCCTCC ACCTTTCAAG	TGATTGTGAC	ATTGTATTA	5500
CTTTTGATA ATTTCTT	GTTCATACAG	ATCGTGTAC	5600
ATAATTTTT ATGAATGAAA	TGAAATTCCA	GGCATCCTT	5700
TCAAGCATTCTACTGGAAAT	GATGTGCACC	CTGCTTACAA	5800
CCTTATAACA GTAAATGCATG	TTTGCTAGAA	AATTCAAGAA	5900
GTATTTAAAA AATTAAAAC	ATAATCTCCA	AAACCAAAGA	6000
TAATTATAAA AGTAATAATT	TATTCAAAAA	ATAAAATCTAG	6100
AAAACATAAA GTAGCCAGAC	TCAGTGGTGT	GCACCTGTAG	6200
TCAGTGCCTG AGGTGGGGAGG	ATTGCTGAG	CCCAGGAGTT	6300
GTGCACTATG CTGATCAATT	GATCAATCCA	CTGTC TGAC	6400
ATGAATACAG TGACCTCTT	GGAGGGCAAG	ACCATCAGGT	6500
GGGGTGAAGT GGCCCACGTT	GGAAAGTGGAG	AAGGTAAAAA	6600
ATCCAGTAGT GGGATGACAT	CTGTGAATAG	CCACTGCAC	6700
CAATATAGGG AAACCATGTC	TCTTTAACAA	TAACAACAAAC	6800
ATCCCCAGAAA CTACAAAAGG	AGAGTCCTT	TGGTGCCTCC	6900
CCTGTCCTTC CAGCCTTATT	TTTCTTAAGT	ATATGCACAA	7000
AGATAAAATTC ATATCCTTAG	ACAGGTAAAG	CATTCAATTAA	7100
TATGCAAGGA TACTATCAA	GGCATGGTAT	CCATGCAAGG	7200
GCCTTGGCCC TGGAGAGAAC	CCTATACATA	CTCTCAACTG	7300
ACTGTTATTG TGTACTCAG	TGCATCATTG	CTATCAACTC	7400
CAGTTAGTCA AATGAGGTTC	TACCATTTAC	CAACTAGGAG	7500
AGTTACTCAA TCTCTTTCT	AAGCCTCTTC	CTCATATGCA	7600
AATAAGTGT TTACAAGATT	CATGCATGAT	ATAATGTATG	7700
AGCATGGTGC CTGGCATATC	ATAACTGTTA	AACAATTATT	7800
TAGCATTG GGAGGCTGAG	GCAGGGGAT	TGCTTGAGGC	7900
AGACCAGCCT GAGCGACATA	GTGAGATCCT	GTCTCTACAA	8000
AAATGTGTT TCATTAGCTG	GGTGTGTTGG	CATGCACCTG	8100
TACTTGGGAG GCTGAAGTGG	GAGAATTGCT	TGAGCCTGGG	
TACAGTCAGC TGTGATTGCA	CCACTACACT	TCAACCATGG	
AGACCCTGTC TCAAAAAAAA	AAATTATCAG	CTATTACTAT	
ATTAGTTCT CACTCACCTA	AAATCTAAA	CACACCTTAG	
ATGAGAACAA CCAAAATGA	CAAAGTAGAA	GCACATATGA	
AAAGCATGAA AGCCAGCAAG	AAATAACTGC	CGCTCTTCTG	
AAAGCAACTG GCATTTCCC	TAGAACATG	TGGTGCACA	
TTAAAATCAC CTAGAGACCT	TTTAAAAATT	CTGAAACCCA	
CGACAAATGA AAGCACAGTC	TCTGGGGTG	GGACATAGGC	
GAAGGTCCCC ACTTGATCCT	AATGTGCAGA	CAAATTGAA	
TTGGGAGGCT GAGGTGGGTG	GATCATAAGA	TCAGGAGTT	
TGGCAATAT GGTGAAACCC	CGTCTCTACT	AAAAATGCAA	
GGCATGGTGG CGCATAACCTG	TAGTCCCAGC	TACTCTAGAG	
AAGCATCAGT GAGGGAGAAT	CAGTGAACCC	AGGATGTGGA	
AGCCCAGATC ACACCACTGC	ACTCCAGCCT	GGGCGACAGA	
ACCTCAAAAA CAAAAGAAA	GAAAGAAAAA	AGAAAAGCAG	
CCTACCAATC CATCATTTAG	CATATTGTG	AAAGTAAAGTG	
ATTTGCAGAA CCAGAACATCA	GCTTTATTGG	CCAGGCCAA	
AAACATGGTA AGGTCTCCCA	GCCTTAAAGT	ATTACATAGC	
AANNNNNNNN NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	
NNNNNNNNNN NNCAGGGAGC	CCGAGCCTCC	TCCTTGTGT	
AGAGGATGGA CGGATCTTG	CGCAGCGTAC	CTCTGACTAT	
AGATGCTGCA GATTGTGGTG	GGGGTTCGAG	ACTCCGGTTG	
GGATGCCAAC ACATCTCTGC	ATGTGTTGT	CCTAGACGAG	
CCCCAGCTGT GCTGCACCCA	CGGCCAGACT	GGGAACACTC	
CGTCTCCCTC GCTCTGCTCC	TCCTGGCTCC	TTGGTCACCA	
CGTGGATGCT GATGCAGGCC	ACAATGCGTG	GCTCTCCTAC	
CACAGTCCAC AGCCCCAGGA	CTGTTCTCG	TGTCTACACA	
GTGCGCACAG CCCGGCCTT	ACTGGAGGAT	GACTCTGACA	
	CCCAGCAGGT	CCAGCAGGT	

FIGURE 1C

4/14

GGTGGTCCTG	GTGAGGGACA	ATGGTGACCC	TTCACTCTCC	TCCACAGCCA	8200
CAGTGCTGCT	GGTCTGGAG	GATGAGGACC	CTGAGGAAT	GCCCAAATCC	8300
AGTGACTTCC	TCATACACCC	TCCTGAGCGT	TCAGACCTTA	CCCTTACCT	8400
CATTGTGGCT	CTAGCGACCG	TCAGTCTCTT	ATCCCTAGTC	ACCTTCACCT	8500
TTCTGTCAAGC	GAAGTGCCTT	CAGGGAAACG	CAGACGGGGA	CGGGGGTGGGA	8600
GGGCAGTGCT	GCAGGCGCCA	GGACTCACCC	TCCCCGGACT	TCTATAAGCA	8700
GTCCAGCCCC	AACCTGCAGG	TGAGCTCGGA	CGGCACGCTC	AAGTACATGG	8800
AGGTGACGCT	GCGGGCCACA	GAETCGCAGA	GCCACTGCTA	CAGGACGTGC	8900
TTTTCACCGG	CCTCGGACGG	CAGTGACTTC	ACTTTCTAA	GACCCCTCAG	9000
CGTTCAGCAG	CCCACAGCTC	TGGCGCTGGA	GCCTGACGCC	ATCCGGTCCC	9100
GCTCTAAATAC	GCTGCGGGAG	CGGAGCCAGG	TGAGGGGCTC	GGCGCCGCC	9200
CGGGCGACCC	CTGGGGGCGG	CACTGGAGAA	GCCGCCCGTC	CTCATAAAGG	9300
ATTGAACTTG	CATCCACTCC	TCTCCGGCCG	GCCTGGTCGC	TGGCTGCGCT	9400
CCACCCGATT	CTCGGGATCA	TTGGACCGTT	TGCGCGAAAC	CAGAGTGGCC	9500
GATTAAGGGA	TGGGGCTCCG	AGCACCGGGG	GTGGTGGCGA	CTGTGGGCAG	9600
GGGGAGGTGG	GACCGACCCC	CACCCCTACA	CTCAAAAAAAAG	GCCGGGGCCT	9700
CCTTCGAGCT	TCCGGTGAAT	TTCGGGGAT	TTCCGGGGGT	GTCGGGGGTC	9800
CCGGGAGGAG	GCAGTCACAG	ATCCACCCCT	GCAGGCCAGCC	TCCTAGGC	9900
CGGCTCCGGC	ACGCTTCGCC	GGTCTGTAGA	TTTCCTCTTC	GATTTCTCCC	10000
CAGCTCCAG	CATCTGTGAC	TTCACTGT	CCCTCCCTAT	CCCCGCATCA	10100
CCCAACCGCA	CCTGTC	GGACTTAGT	GTGCGCGC	GGCTCATGCG	10200
TGTCCTCCCT	GCTGCCACC	CCCACGCC	ACACAAGTTG	CACGGGCTCG	10300
CCACGCCCG	CCAACACGTG	CGCGGACGCA	CGCACGCACT	CCTCGCACGT	10400
GGGCTTACGC	GAATACCAGC	TTTCACTGCC	ACTCGCTCGC	GGCCAGATT	10500
ACAGGGCTGT	TCCGGTCCAC	TCGCAGCTCC	CCTATGCC	TCCCTCCGCC	10600
GGGCTCAGGA	GTACTCGTAG	CTGATTATGC	GCCTGCTGAGG	GTCGGCAGATC	10700
GGGGCCGCC	AGGACCAAGG	GAGGACTCCG	GAGCCTCCTC	TCACCTCTCC	10800
CACCTGCGCC	CCGGGCTGGG	CCGGGTCGCG	TGGGGGGCGG	CCTGAGCGAG	10900
GCGCGGGGCC	AGGAGCGCTG	GAGCAGCT	CGCTCTAAAGT	GCCGGGGCGG	11000
CAGGACTCTA	CGATCCTTGG	GCCAGAGGTC	CGGATGGTCC	CGGGACTCCG	11100
TCTCAAGGGT	CGCGGACCCC	TCAACCCAGA	AGCCTCGAGC	AGGCAGGACAG	
GCAGAGCTGC	CCAGTGGCCG	AGGCGCGGCA	GGGCTCCCGC	TTGGGGCGAGT	
GAGTGAGCCT	CTATAGGACA	GCAGGACTGG	GACTCCAGT	GCACCAAGC	
CCCCTCCTC	CCGCAAGGAAAG	TGAAAGCTC	GAGCCTCCTG	GCTCTCAACC	
CCCAGAGATA	CAGGCTTTTG	GCGCCGTCGT	GATCACAATG	TGCCCAGCGA	
TCTAGGTCA	GAGATTGG	GGTGACAAA	CTATCTGACA	CTCTAACAAAG	
TCCTGTCTCC	TCTGGCAGAT	GGAAAGCTAT	AGGCTCTGCC	AGATGCCAGG	
GTGCCCTAT	GTGTGAGGAA	ACTACAATAG	TAAAAAACAC	AAGTTCTCC	
AACTCCAGGA	GCTTTTATT	AAAATATATC	AATGCTAAC	TCTGCTCCTA	
GGACTGTATT	TTGAAACACC	CCCAGGTGAT	TTTGATAGCT	GATTGAGAGA	
AACTTACTAT	ATAACTCCTT	TGAGAACCTC	ATCTCATTG	CTCTTCCCAC	
CATTGCTGTT	GGCTAGGTAC	TAACATGCCT	CTCTTATAAC	AGCTTCACAG	
AGGTCAAGTG	ACTGCTCAA	GTTCACAGAG	CAAGTAAGAG	AGATTCTAAC	
CCCTGTCTAA	CTCCAGAAATG	TGTGCTTTA	ATTCTTGGC	ACTTGGAACT	
TTAAAAGCTT	GAGGACAGGA	GAGGGGAGTT	GCCTCTGCTG	GGATTTGCT	
TCTGCTGGGA	TGGGGCAAGG	GTGGGGTCC	TTCCCTTTA	GGACCTTACA	
TGTGGGGAAA	GTCTTCTGTG	GCTCCTCATT	TCTGAGCAGT	CCCCGCAGCG	
CAAACATTGGC	CAGTTATCCT	TTTGGAGATT	GAGTCCCCC	AGCTCCGTT	
CCTCCTATCA	CAATCACTGC	ATTTCATGT	AGATTCTGCT	GTGTCTAAGA	
ATACAGTGGC	TGAGGGCTGG	CCATCCCTGT	GCCCTTCTCC	ATGGCAGCCC	
CAGAATGGTG	CTGGTGACTC	CCGATACACC	TGGAATGCTA	GGTTCTGGG	
TTCTGCTCAT	ATCACTGCCA	CCTGTGAGCC	TTGAGTGTGAGC	CACTGGCAGT	
TCCGGAGCTT	CCTTCTCCCT	GGAGTAATCT	GAGATACTCA	CACTCTCTGC	
TTCTGGGAGC	AAGTATGAGT	ATTAATTACC	AGCCCACCCC	AGAACCAAGTG	
AGGAGGTGGC	TCTAGGAGTG	CTCCATGAGA	GTGTGTGATG	GGATAACACA	
CCCCCTGGAA	AGACCAGAAG	GGACTCAGGA	ATGAAGTGGC	TGGCCAGAGC	
CCCACACCTT	CAGCTAGGTG	GGAGATGGCT	ACACATCAGC	CCCTTGGGAG	
CCCTGGAGAC	TTAGTTGGCC	CTGCTTGGAG	GCTGTGGGAG	CTGGATCCCT	
CCCCGCTGCA	TCCCTTCCAT	TTTCTCCGT	CTCAGACAGA	GCAGCCTTGT	

FIGURE 1D

5/14

TCTCTTCTCC	TTAGTCACAG	ACCATTGTCT	GGCACGGAGT	TCTAGGGGTG		
AGAAAGTGTCC	CGGGACTTGG	ATGCCCGCA	AAGGCCAAT	CTGGCATGAC	11200	
TCCTAAATTA	ATAATGTATT	TAGCTGTGGG	AAGAGATTCT	TGCAAGCCAA		
GGGCCAGAG	AAGATGTCCC	TGTGAATGTG	TCACTGCACA	ACCTGGCACC	11300	
AAAAGGGTTA	CCAAGAACAG	CAGCCATCTT	GCTGCAGAGG	ATGCTTGTT		
CCCAGCTGAG	GAGTTGAATA	AATTCAATTCT	AGGGCTGGTA	GAATTCTCAT	11400	
TGAAAAGCCT	CCTTGCCAC	TTAGGGGGC	TTTGTCTGCA	CCTCTTCCCC		
CAGTTCCACA	GAAGATGCCT	TCAGTCCTG	AATTTTGGCT	CAGGAGTTCT	11500	
GACTCTGGGG	GCAGGGAGGA	AGGGGCCATT	TCTTAGGAA	AGGAGTCTCA		
GCTTGCTCAC	TGTGGTCAGA	TGAAATGTGA	TTTATCTCTT	GGTTTCTGGT	11600	
ACCTCAGACC	TCTGAGACCT	GAGGTGTATT	TTGTCTTGG	AGATGAGTCC		
ACCCCTGCC	CCTCTTAGTC	CGTTTTATT	TCTGTGCCA	CTCCCCTCCT	11700	
TCTCTCCCGG	CCCATCCCTA	GGGGCTCGGG	TGACATTCTA	ACTTCTCACG		
GGTACTCAGC	CCCTTTCCAT	CTGTTTCTC	CACAGCAAGC	CCCGCCCAAC	11800	
[exon 2: 11786..]						
ACGGACTGGC	GTTCCTCTCA	GGCCCAGAGA	CCCGGCACCA	GCGGGTAGGT		
		A ..11844]				
GAATGATTCT	CCAGCCCCACC	CTCTTCTCTG	CGGCATTTC	TCACGGATGA	11900	
CGTGAGAGCA	GATGGGGGGAG	GGCCCACCAT	TTGCTACACA	TGGCTTCTCC		
CTCAGTTGA	GATCCCAGGG	AGGTCTTGGT	GTGCGGGGGG	CTGGCACACAA	12000	
GACCCCGGAA	GGAAAGAGGCG	ACTGCCCTGA	CTGTTCAGGA	AGCTCAATT		
ACATGTTGC	CCCTTCCCTC	CTCGCCACGA	CCGGCACCTT	TTCCTATCCC	12100	
CTGAGGGCAC	TGTGGAACCA	AAGGATGGTC	TTAAGCTGGT	CTCTGGGTGA		
AAGCGAGGCT	TTCTATGCC	ATGTACTGCC	TAACCCCTC	CCCTGAGTTG	12200	
AGCTGGGCTC	CATTATGACC	TGGGGCTCAG	GCAGAAAAGC	ATTGACCGG		
AGGAGGGCGGT	CCGCACTCAG	CGCCTCC	AGAGCCTCCA	GAGCCGAGGC	12300	
TGACTGCAGC	CTGGGGAGAG	TGGAGGCAGC	ACAGCTGGAG	TGGAFACTGA		
GGAAAGGACTG	GGGGGGGCAT	GGAGCAGGCC	CCCTCTCCG	GCCCCTCCTC	12400	
CCACTGTCC	CTGCCCCCTAC	TTGCTCTGCT	CTCTGTCTGT	GGGGTCTCCG		
TGTCTCTGCC	CCTTTTCTT	GAGTTCCGT	CTTGCCACT	TTCTCTTACC	12500	
TCTCAGTCCT	TCCCTCAGTC	TCTATCTCGC	TTTGCATCT	CTGCCTCTCC		
CTCTCTTCTC	CATCTCTGTC	TTAGCTCCG	TCTTGATTGC	TGCACCTCAG	12600	
CCTCAGTCCC	TTTCATTGTC	CTACTTGCAT	TGATCTGTG	CGCCCACTGT		
GCCTGCCATG	GCATAAGTGC	TCAATAATG	TGGAGTGAAT	AACTACACGG	12700	
GACCCTTAGT	CTCTTCTCC	TTCTCTATCT	CTGCCTCCCT	GTCCTTGTCC		
TGGACCTCTT	TTCTGTTCT	CCTTCACCG	TTTCTAGCG	CCTTGTGTT	12800	
TTCTATCCCC	AGCCTCTATG	TTTCTCTGTC	TCTCACTATT	TCTGCTTCC		
TCTCTGTTCT	TTGCTCTCT	TTGTCTCTGT	CTCTGTATAT	CTTTCTTGT	12900	
CTCTGTCTCT	GGTCTCTGTG	CCTGTTGTCT	TTTCTGCCT	CTTTCTCTGT		
TACTCTTCT	TTATCTCTCT	TTTTCTTTA	TCTTGCTTT	TTTCTCTGAG	13000	
CCTCTGTCTC	TGTCTCTCAT	TTTTTGTC	TTGTGGCAA	GCCAGCACAC		
ACAACCCCCC	ACCCACCACC	CACCAACACC	CCGCTCTCTT	CCTCCCTGC	13100	
CCCTCACACA	CTGAGCCTTT	GATCGCAGCT	CTCCACCAGA	CACCCCTCCA		
TCCAAGCAGC	CCCAGCTGCT	CATTCAATC	TGGTATGAAT	TCCTGCTGAG	13200	
ACAGGAACCC	CCTGCGGGCT	GAAGGGGAGG	GAAACTTCAG	CAGAAAGACC		
TTCAGTTGGT	CTGAGCAGAG	TGGGATAGGC	TCTGCGCCAG	GCCTCCAGT	13300	
TAGAAGTCAG	GAGCCTGGAG	GAGTCTCAGA	GCCCAGGGAG	AGAACCGAGA		
GCTTGGGCA	CCTTACCTA	CCTCCTCAAT	GGTTGAGTAC	TTGCTGTCCA	13400	
CATTTGCCAA	GTTGCTACAG	ATGCTGAGCT	TCCAAGAGTT	ATCTTCCCC		
CATCCTGCAA	CCAACCAAAC	CCTTGTGCC	ACAAGGACCC	AGGAGCCCTT	13500	
GGTGTATGGC	AGAGATTCCA	GCTTCTGGGC	ATGCACAGTC	CTGTCACCCA		
TTTTAGGAAC	GCAAAGCGCT	TCTTAAAGG	CCCCTAAAAAA	GCAGCAGATC	13600	
AAAAGGCTTG	GGCTGCCCTT	GCCCTTCCCT	TGACCCAGC	TGTTGTCC		
CTCCCTGCTG	GCCTTGCCAA	CCTTCTCATA	AGTTATCCAT	TAAGTCATTA	13700	
ATGTATTCTAT	TCGTTCATTT	ATTCAACAAA	TATTTATCGA	GCATCTACTA		
TGAGCAGGGC	CCTGTGCTAA	ACATTGGGCT	ATAGGAGTGA	ACAAGTAGAT	13800	
GTGATCCCAG	AACTCATAGG	CTTCCAGACC	AGCAGAGGAG	ACTGAAAATT		
TGCAACAAGT	AAACACTAAA	AAAAAAAAAA	AAGGCTGGGC	ACTGTGGTTC	13900	
ACGCCTGTAA	TCCCAGCACT	TTGGGAGGCC	AAGGTGGGTA	GATCACTTGA		

FIGURE 1E

6/14

GGTCAGGGGT	TTGAGACCAG	CCTGGCAGCC	AACATGGTGA	AACCCCACATCT	14000
CGATTAaaaa	TACAAAAAATT	AGCCAGGC	GGTGGCAGGC	ACCTGTAATC	14100
CCAGCGACTT	AGGAGGCTGA	GGCAGAACTG	CTTGAACCCG	GAAGGCAGAG	
GTTGCAGTGA	GCTGAGATCA	CGAAAAAAA	ATAATCTAGC	CACAAATCAC	
AGTAAGTTCT	GAGAGGCAGA	GAACAATGTG	AGTGTAATGG	GGGGGGGAAG	14200
ATCAGAGAAG	GCTTCTAGGA	GGAGGTGACA	TTGAGAAGTT	CTAGGCCATT	
TATGTTCCCT	CTACCTACCA	CCACCCCCAGG	CCATACACTG	GCTTTGAAGC	14300
AGAATTCAAC	ATTGAACCTG	GAAGCCCCAC	CAACTGCCTC	TCATGTGTCC	
TCTGGGTGA	GCACCTTACC	CCTGCCTCAA	CTCCCCATTCC	ATCAAATGAG	14400
AGGCTTAGTG	GCCTGGAATA	CATGTGGAG	CTCTTGGAGA	TATGGAGGAG	
AAGGTGCGGA	ACCTCTCTCT	CTCTCCTCTC	TCTGCTTCCT	TAGTTAACTT	14500
GTGGAGGATG	CAACAAACCT	TTTTTTTTT	TTTTTTTAGA	CAGAGTCTTG	
CTCTGTTGCC	CAAGCTGGAG	TGCAGTGGTG	CGATCTTGGC	TCACTGCAAC	14600
CTCTGCCTGC	CAGGTTCAAG	CGATTCTCCT	GCCTCAGCCT	CCCGAGTAGC	
TGGGACAGCA	GGTGTGCACC	ACCACACCCA	GCTCATTTTT	GTATTTTAG	14700
TAGAGATGGG	TTTCGCCACG	TTGGCCAGGC	TGGTTTCAA	CTCCCCGACC	
TCAACTGATC	CACTGCTTC	AGCTTCCAA	AGTGCTAGAA	TTACAGGTGT	14800
GATGCCCCC	GGCCACAACA	AACATTTTG	ATTCAAGC	TCCAAAGTGA	
AATAAGCGTT	AGGATCTAGT	CCAGTAAC	CAACTCTCT	CCTTCATTCA	14900
TTCAACAAAT	GGTCAGTGAG	CACTTTCTGT	TGCCAGGCA	GTGTTCTCGA	
TGTGGGAGGC	CCCTGCCCTC	AAGGAGTTA	CAGTGCACTG	AAGGAGACAA	15000
ATATTGACCA	AATAATTACA	AAAATAATG	TGAGACTGAA	ACTGTATAA	
GTGCTAGGTA	GAATTGTATC	TGGTGGTGTG	GGCGTATAAT	AGGGATTGAA	15100
TTTAGGGAGC	AGGAGGACAT	CCTGGAAATG	AAAGTTGTGCT	GAGATCTGAA	
AGACAATTAC	AAATTAACCA	GCTAAAGAAA	AGAAGGAAGA	GCATTCCAGA	15200
TGGTCAGGCA	TACCAAGCTTC	AAAGCAGGGC	TGGTGGTGTG	TTCTTCAGTA	
AACTTGGTTC	AAGATCAAGC	CAGCTCTGGG	TCACACTGCC	TACACAGAAG	15300
GAAGAGCATT	CCAAATGGGA	AAAGCCTATG	CAAAGCCTT	GTGGTGGAGGC	
CTCCTGGGAT	CAAGGGAATG	GCAGGGAGTT	TTCTCTTGC	TGGCACTGGA	15400
GTGAGGGGAT	GAAGAGGAGA	GGTTCCCTTT	ATCCAGTCTA	ACCCTGGTTC	
CAGCCTTCTG	TTGATGCCCT	GCCTCACTCC	CTCCCCACCA	TCTTGGCCAC	15500
TGCCCTCCAG	CTGGTCTGG	TCAGACACAC	CAGCTTCAA	GCAGGGCTAG	
TATTGAGTTC	TTCAAGTAAAC	TTGGTCAAG	TTCAAGTCAG	CACTGGGTCA	15600
CACTGCCTAG	ACAGAAGGAA	GAAGACCTGG	CCAGGTGTGG	TGGCTCACGC	
CTGTAATCCC	AGCACTTTGG	GAGGCTGAGG	TGGGTGGATC	ACCTGAGGTC	15700
AGGAGTTGA	GACCAGCCTG	GCCAACATGA	TGAAACCCCA	TCTCTAGTAA	
AAATACAAAA	TATTAGCCAG	GTGTGGTGGT	GCACACCTGT	AGTCCCAGCT	15800
ACTCAGGAGG	CTGAGGCAGG	GGAAATCGCTT	GAACCCAGGA	GGCGGGAGGTT	
GCAGTGANCT	GAGATCACAC	CACTGCACTC	CAGTCTGGGT	GACAGAGCGA	15900
GACTCCATCA	GGAAAGAAAG	AAAGAAAAAG	AGAGAGAGG	AGGGAGGGAG	
GGAAAGGAAGG	AAGGAAGGAA	GGAAGGAAGG	AAGGAAGGAA	GGAAGGAAGG	16000
AAGAAAAGAA	AGAAAAAAGGG	AGAAAAAATG	AAGAAGACTT	ACATTCTGAA	
GTTCCTGCTT	GGCACATGCT	TCTCCTCCCC	ACCACTATCC	CTTCTCAGGT	16100
CATCCTTGG	GTGTCTGATC	CCAATAGCTG	TTGTCTCAAG	CTCTGAGCCC	
AGCAAACCAT	TTCCCCCTTC	TTAGGCTGTA	ACCCAGGCC	TGCCCTGCAG	16200
CTGTCCCCAG	CCTCTGCACA	AAGAGTCTCC	ATTGGTCCC	TAGACTGTT	
TCTGATGGCT	CCACACTACC	ATCCCCTCTC	TGAGAGACAG	TATCCCTGCT	16300
CTCTCTCAGT	TTGGGCAGTC	ATTCAACAAA	CAAATAGCAA	CTTAATT	
TTCCACTTAT	AAAATAGTAC	ATGCTCATTA	TAGAAACATG	AAAATAGAA	16400
AAAAAAATTAA	AAGGAAAACC	ACCCATCATT	TTACAATCTA	GGGAGAACCA	
CCACCAACTA	CAGTTAATA	TGAAACATT	CAAATATGTA	AAAAAGTAGA	16500
GAGAATGGTG	TCATGAACCT	GTATGATACC	ACTATCCAGC	TTCAACTAAT	
GCCTTATCTT	GTTCACCTA	TATCCCTACC	CTTATCCTGT	CCTTGGATT	16600
TTTGAAAGCA	AATCTCAGAC	ATCATATAAT	TTCATCTTGG	TCTCCACCAT	
CTTTAACAGA	TGACTTTTTT	TTTTTTGGAG	ACAGAGTCTC		16700
ACTCTCTTGC	CCAGGCTGTG	GCACAATCTC	AGCTCACTGC	AGCCTTCGCC	
TCCCAGGGTTC	AAGCAATTCT	CATGCTCAG	CCTCCTAAGT	AGCTAGGATT	16800
ACAGGTGTGC	ACCACCACAC	ACAGCTAATA	TTTTGTATT	TCAGTACAAA	
CAGGGTTTCG	CCATGTTGGC	CAGGCTGGTC	TTGAACTCCT	GGCCTCAAGT	16900

FIGURE 1F

7/14

GATCCACCCCT CCTCGGCCTC CCAAAATGTT GGGATTACAG GTGTGAGCCA	17000
CCTTGCCTGG CCAGATGATT CTTTCATAAG GTCCACATT TATATTGAG	
TGAGTGTCTC TTAAGTATCT TAATGCTCTT TTAATGTAAA AGACTTCCC	17100
TCCATCTTCC ATTTTGCAA TTTATTGTT GAATCCCTGC CTTCATAT	
ATATCTTCT AACACTTTCT TTGTGTAATC AAAGAACATG CATGAGCTTC	
ATTTTTATTT TTATTTATTT ATTTATTAT TTATTTATTT ATTATTTAT	17200
TGATACGTAG CCTTGCTCTG TCACCCAGGC TGGAAGTCAG TGCGCAATC	
TCGGCTCACT GCAAGCTCCG CCTCCCAGGT TCACGCCATG CTCTGCCTC	17300
AGCCTCCAGA ATAGCTGGGA CTACAGGCAG CCACTACCAC GCCCGGCTAA	
TTATTTGTA TTTTAGTAG AGACGGGTT TCACCGTGT AGCCAGGATG	17400
GTCTCGATCT CCGACCTCG TGATCCGCC GCCTCGGCCT CCCAAAGTGC	
TGGGATTACA GGCCTGAGCC ACCGCCCTG GCCGAGCTTC ATTTAAAAAA	17500
AATCTCATTC ACCTAAATAA GTTGTTCACA AACTTGTAC ACTTTCATGT	
GACACTTTAT TCACATATTA AGAACTGTTA TTCCCAGTC TACGGTACAT	17600
TTCCCTACTAG TTTTTAAAAA ATAGCAACTG AGGCCATACT GTCCAACAGA	
ATTATAGCTT TTTCTCTCTT TGCAAGTGT CATGAGAATT TCCCCATGTC	17700
TTTCAAAATT CTTCCCAAGT CATTGAGA GTAACTCCTC TTATGAATAT	
ACATATTATT TTACAGAATA AGTCCTCATA TGAATATACA TATTATTTT	17800
CTAATGCTAA TGGGCTTTG CTACATTAGA AACACATACAC ATTCAATTAC	
AATTTTTTTT TTTGAGACA GTCTCGCTCT GTCTCCAGGC TGGGGTGCAG	17900
TGGCGCCATC TCGGCTCACT GCAAGCTCCA CCTCCTGGGT TCAAGCAATT	
CTCCTGCCTC AGCCTCCTGA GTAGCTGGGA CTACAGGCAC GCGCCACCAC	18000
GCCCCAGCTAA TTTTGTTATT TTTAGTAGAG ACAGGGTTTC ACCATGTTGA	
CCAGGATGGT CTCAATCTCT TGACCTCGT ATCCGCCTGC CTCAGCCTCC	18100
CTAAGTGTG GGATTACAGG TGGGAGCCAC CATCCCCAGC CTCATTACA	
TTTTAACACA ATAAATTCA GGGTGTGTC AGGGTGACTT CCTAGATTTC	18200
TCAGTTTTT GAGCATGATG TAGGGAGTAT TTATTATAG TCCAGTCTAT	
GGGATATTCC CTTATACACA CACACACACA CACACACACA CACACACACA	18300
CACCCCAAAC TCAATAGGGC AAGAACATA TTCATCATCT CTTAAATCCT	
AGGTCTGTG ACAGGCCACT GCTTGGAAA TAGTTGACTA AATGTAGTCC	18400
ATTTCTTCA TTTTACTTT TCACTATTCC CCAAATCCTG GAAGAGTCCC	
TCATATCTGC TGCTGGGCT CCAAGGCCCT CAAGCTGAAG CCCAGTACGT	18500
TGTTGTGCAT AAGATCATAG GCCCTGGAAT CATATTAGGC TGGCTTCAG	
GTTGCAACTC TACCAGTTCC AGCTGTGTGA CTCTGGTATA GTCTCTAAC	18600
CTCTCTGAGC CCTGTTTCA TCTGTTGAA AATATAGATA ATCACAAACAC	
TTATCTGCAG GGTGATTCCG AGGATAAAAA AGACAAATAT ATTACCCAGT	18700
CAGCTTAGAA TAGGAAGCTC TCAGTAAATG GTAAAAAAG AGGTCTCTCT	
CCCTCTGCTC ATCTCTTACC CAGCTTCACA TAATTGGCAG CCCACCCACC	18800
TGTTTGGCCC TCAACCTTCA ACCATTCCC AGGACGCCCT CTTACACTCC	
TGATCATCAT GGTTCCAACC CCACTGTTCC ACTCAGCTTT GTTCCCTCAA	18900
CACCACTGCT CCCTGTGACA GGTCACTCTC CTAACCTCTG GCATCTGGTC	
CCGTTGTCTC CCTGACAACC CGCATGCTCT CCTCCCTGCA ACCCCCCCG	19000
CTTGCCTGCT CTCCCTACTT TATTCCATT AGCCCCCTTC ACCTGGGCT	
GAGAACACCT GCCCTTGACC CTGCACCCCTC AGCCAGGCC ACAGAGCTCA	19100
GTCGGGCCGG AAGTAAGGAC CCCAGATATT CAGTTGAGGA ATTTAACATT	
AATACTATAA TATAGAATCC ACCAGCAGAT TTTCNNNNNN NNNNNNNNNNN	19200
NNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNATGGGC	
CTGAAAGTAA GACCCCCAGA TATCTCAGGG AGAACCTCCA ACAGNATTAC	19300
TACATTCGAG TATCCACAAG GCTGACNGTC ATTAACGTT CGAAGTGTAA	
TATTGTCGGC TCCTTTAATG GATGAGATTA TGAGCGAGCA GAGACCTCAT	19400
CCGGTTGTCA CGTGATTCTG GTCATGGTAT ATGGAGCGAG CTCTCATCCT	
GATTATCGGT TTATCGACAT GGAAAGTTGG AGTGAATGCT GGCCAGTAGT	19500
AGACTAATAT GTCCCAGAAC CAGGATAACTC TAGCTATTCT CGGATGGTTA	
GATTGAGATT ACACACTTTG GTCGGATTAC TTCCCGCAGC TTCAAAGTGA	19600
GACTCACCTG ATGTGGGCAC GTCGCAGTAT CAGTGTGTC GAGCTGTAA	
GTGCGGTAAA GAGAGTGTCC CTGATGTGTC GCCTGTATAA ATAGAGTTT	19700
GTAAAGCCAC CTTGACGCAT TTGAGGCAA CTACGTCTTC GCGGAGATAC	
TTTGAGATAG GCAGCAACCC TTCATGCAAT TTAGTATTTC CCTGAGGATC	19800
CTTGCCTGTA TCTGTTTTA CATCAGTGAT TACAAAGCGG TGATGTTCTG	

FIGURE 1G

8/14

TCATTCCCTTC TACATGTATT TGCTGGTATT CTTCTGAAAAA TTACCTGGAT	19900
GATTTCTTCA GATTAAATT CACTGTCTT TCCTTGGGA AGCCTGGACT	20000
GGCTGAGCTG CCTAACCTG ACTCTCTGTC TTTTTTTTT TTTTGACAGG	
GAGTCTAGCT CTGTTGCCAG GCTGGAGTGC AGTGGCAACC TTCACCTCCT	
GGGTTCAAGC GATTCTCCTG CCTCAGCCTC CCAAGTAGCT GGGATTACAG	20100
GCGTGTGCCA CCATGCCCGG TTAATTTTT TTTTTAATA TCTATTAG	
TTGAGACAGG GTTCAACCAT GTTGGTCAGA ATGGTCTCGA TCTCCGTACC	20200
TCGTGGTCTG CCCGCCCTG CCTCTCTCA TACCCCTCTGG TAGCCCTGA	
CACAGCCCCT GGCAAGTATC ACAATTGCAA TTACTCTATT GTGTAATTAT	20300
TTAAAATCAG CCTCTCAGTG CCCACCAGA GCACCGTAAA GATGGACAGT	
TTCTATCATA CTCACAGCCA CAACCCCCAGC ACCTGGCTCT GTGTCGGTA	20400
CATAGAAATT GCTCAAGAAA GAATTGTTGG AGGGGCGCGG TGGCTCACGC	
CTGTAATTCC AGCACTTGG GAAGCGGAGG CAGGTGGATC ATGAGGTCAG	20500
GAGTCGAAA CCAACCTGGC CAATATGGT AAACCCCCATC TCTACTAAAAA	
ATACAAAAAT TCGCTGGCA TGGTGGCATG TGTGCACTTG TCGTCCCAGC	20600
TGCTCAAGAG GCTGAGGAG GAGAACACT TGAACCTGGG AGGCAGAGGT	
TGCAGTGAGC CGAGATCGCG CCACTGTACT CCAGCCTGGG TGACAGAGCG	20700
AGACTCCAGC TCAAAAAAAA AAAAAAGAAA AAAAAAGAAA GAAATAATTAA	
TTGAATCAAT GGAAACAGAA TTACAACCTCT TCCCACCTTTT GGGAAAGGAGA	20800
ATTGAGATTG CTGCTTCCC CTTCAACCCCT AGCTTATGT TTGTTGTGAT	
GGTATAGAGG TCACACATGG CAGCCTGTG TCTGTTCTTA GGCTGAGTTC	20900
ATCTACTTAG GGACGGGGAA TCCCAGATTG GGCAAAACCA AAAGCAGGGAA	
CAGACCTTGG AAAAATTGTA AAGCAGTTCT TGTTGTGCT CACCCACCAT	21000
GACAAATAT GCACACACAC AACGCAACAC ACACACAGAC ATTTCATGC	
CTTCAAGGCC TACCCCTACAG CCTTCCCCT GGTGCCAAC CCTTCTTCTT	21100
ACCTTACAGT CACTTAACCTG TCTCTGCTGG TCTCTCCCA CTCATTAGG	
CAACAAATAT TTATTGAAAA ACTTCTCTGA GCCAGATGGT GCCAAACAGT	21200
TACCTAAAGG GCAGACTGTG TGTGTGCCAG GGCTGTTCC CGCCAACACT	
CCCCTGCAAA TTGAAATAAT TTCATCCTTG GAAATTCAAA AAAATCATGA	21300
AAATTGTCAC TGTGGGAAAA TTCCAAGTAG AGCTGAACCT TCTAAGAACG	
AGAGAGTTCT TCTTATGGTT TAGTATTTTT TAAATTATG AATCATTCA	21400
TGTTTTTCA ACACTCGGAG TTTCTAAAAG TCTCACTTAG GTCCAAGTGC	
TAGGCTTGT GCTAGGTGCT TCAGTAGGAA GCAGGTGGAG TCGCTGCCTC	21500
ACAGAACCCA GTGTGGTGA GGAACAGACA GGCACATACAA CAGCTGCAGT	
GTGACTAGTG CCATGTGGGG CAGCCGCTGG GATGGGAGTA CAGTGGAGGC	21600
ACCTGCCAG TCTGAGTGGA TCTCCTCTGA TATATTCTAT TCGTGTCACTC	
ATGGCAAATG TTGGGGACT GGCATTCTAG GGAACACCAT CTGCCAGGA	21700
TTCACAGCAA GAGGGAACTT GAGGGATAGG AAAGTGCACC TTGGAAAAGG	
TTTGATGATA GAGTGTGTTGA GGGCGGGGG TGGGGATTCT TGTAAAATAG	21800
GAGGGCAGG TTCTCTATG GCCGTGGATT TTACTCTGAG GGCTTAGGAA	
GCCATTGAAT TTAGAAATT TCTTCTCCCT GCTCCAGGGT CTCTGGGGC	21900
CTCCTACATC TCCCCCTAGG TACCTCCTAG GTCCCCACTT TTTCTCTAGC	
TCTGGAACAT TCTCTTCCAT TTCCCACCTT CACAGTCCTC TGGAGAAGGT	22000
CTGATCTCCC TCACATATGAC AGGAAGCCAG GCTCTCTCAT GGTGGGTGGG	
GGAATGCACT ATGTTCAATG CATTGGACCA GCCTGGCCAA CATGGTGAAA	22100
CCCCGTCTGT ACTAAAAATA CAAAAATTAG CCTGGCACAG TGGCAGGTGC	
CTGTCATCCC AGCTACTTGG GAGACTGAGG CAGGAGAACATC GCTTGAACCC	22200
AGGAGGTGGA TGTGCAATG AGCCAAGATC GCGCCACTGC ACTCCAGCCT	
GGATGACAGA GTAAGACCCCT GTCTAAACAA AAAAGAAAAAA AGCATCGGAG	22300
GCAGCAAAGA GCTGGTTTGA GGGACTGACT TGTGGGATTC TGGCTTCTGA	
AGGATTGTAG AAGTGCCTCC TACCTGCTG AGAGAAACAG GTCTGGGCA	22400
C	
TGGGGTAGGG TTAGGGTACT AGGTTGGGA GCCCTGGGAG AGGACAGGAG	
GGGCATGAGC TGTGCCGGCC TGGGAGTCTG TGCTCACCAT CCTACTCTCT	22500
CCCCAGCTCC CAAATGGCG ATGACACCGG CACCTGGCCC AACAAACCGAT	
[Exon 3: 22507..	
TTGACACAGA GATGCTGCAA GCCATGATCT TGGCGTCCGC CAGTGGTAAG	22600
..22595]	
TGGTGTCACT GTGTGTATGG AAGAGTGGGA GACCTGGGGT TCTGGGTGC	

FIGURE 1H

9/14

ATCTCACAGC	CACCATGCC	ACGGACTGGA	TGTCAAACCT	GTGTAGTTTC	22700
T					
TCCAGATCTT	TCGGCAGGTC	TGAAAGGACC	CACAATTCCA	AACATAAAGC	22800
CTGGAATTGT	GGCTAAGGAA	CAGCAGAGGG	GTTGGGGTC	CTGGGATGCC	
TGGAGGAGAG	CGAACAAAGGA	AAAGACTCAT	GGAGGGAATA	GTGGTTACAA	
GTCTAGCTCT	GGATGGGAC	AGTGTAGCTC	AGGTCTTAGC	TCTGCTACTA	22900
TCCTTGGGAC	TTGGATCGAC	TTGCTCAATC	TCTCTAAACC	TCAGTTCCCT	
CAGCCTAGAA	AGTGGGGACA	GGGACCCAA	AGGGTTGTTG	TAGAGATTAA	23000
ATGAGATGAT	ACCACAAAGC	ATAGAGTTCT	GGGCCTGGCG	CTTGGAAAGCC	
CTCAAGCAAT	GAAAGTGGGT	GTAGAAATCC	CCAGCCGAGA	GAACACCTCC	23100
TCTTTATGCG	AGGCTCCTCT	TTTGCTGCC	CTAACTCCAG	AGTAGGATTG	
TGGTTTCCC	ATAAGGTTGA	CTAAGGGCTT	CCTTGTAAT	AGAGATTGCG	23200
CTAGTGGTGG	TTGTCCTGTA	AGAGTCACTA	GGGCCAGAG	CAGGAGAAAGA	
AGAATATCAT	TTGTCCTTAA	GAGCACAGCC	TAAGCTGGG	GTGGTGGCTC	23300
ACGCCTGTAA	TCCCCAGCACT	TTGGGAAGCT	GAGGCAGGAG	GATCATGAGG	
TCAGGAGTTC	GAAACCGGCC	TGGCCAACAT	AGTGAACACCC	TGTCTCTACT	23400
AAAAATACAA	AAATTAGCTG	GGCATGGTGG	TGCGTGCCTG	TAGTCCCAGC	
TACTTGGGAG	GCTGAGGCAG	GAGAACCGCT	TGAACCTGGG	AGGTGGCTGT	23500
GGTGAGCAGA	AAATCGCACC	ACTGCACTCC	AGCCTGGGCA	ACAGTCTCGC	
TCTGTCTCAA	AAAAAAAAAA	AAAAAAAAAG	AGCACAGGCT	TTAGAGGCAG	23600
GCCAATCTGG	ATTCAAATCC	TGGCACCTGG	CACCACCACT	TAGCAATGAG	
TCCTTAGGTA	AGTATTAAA	CCCCCTCGGT	TTCACTTACT	ATTAATGGGC	23700
ACAGTATTAA	CGGATCTCAT	TGTGGTGT	CCAGGGCCAA	ATGCAAGTCC	
CTCAGCACAG	GATTGGCAGA	GAGTAAGGGC	ACAATATATT	ATTATCTTG	23800
CTGACCCAAA	CCCGTTTTT	ACTGGGTAGA	AAGCAGAATT	AGAATAATGC	
CTATTAATAA	AGACTAGCTT	CTGGAGCAGC	AAATCCTGGG	CTTATAAGGA	23900
GGCTGGCAGA	GGATCAAGGC	ATTGCCCTAT	ATCATGAACT	GATAGCCCTG	
CCCTCCAGCA	TGTCTGGAGG	ACTGGTGGGT	AGAGAACCCAG	GTGAAATCAC	24000
AAGAAGCACT	GTCATCACAC	CTGGGCACAT	ACTAAACAAA	CTTTAGGGC	
CTCCTGTCAA	TGAATCCTCT	CAATAGCTCT	GCAAGGCAGG	TCTGATTTCT	24100
CACACTCGAC	AGATGAGAGA	ACCGAGAAGG	CACTTGCCCC	AGGGTCCCTCA	
GTTTGAGCT	GCCAAAGCTGA	GGTTCTAAC	TGGCTCAACT	CCTAAGTTA	24200
TGCTCTTCT	ACTATAGGGA	CCATGGCTGC	CTTGGATCC	AGCCTCGGCT	
TCTCTATTAC	CTGAGCAGGA	TGAGAGGCTG	TCCCTTCCCT	CGAGCTCTGC	24300
TTTATTCTTC	CAGATCAGGG	TTGCCAGATA	AAATACAGGA	TGCCCAGTTA	
CAGTTGAATG	TCAGATAAAC	AGCAAGTACT	ATTCAGCAT	AAGTCAGTCT	24400
CAAATGTGTC	ATGAGACATA	ATGTACTAAA	AAAAAAAATG	CTCATTGTT	
ACCTGAAATT	CAAATTTAAC	TGGGAGCCCT	GTACTTTAT	TTACTAAATC	24500
TGGCAACTCT	ACCCCAGATG	TCTGTGAGGT	TAAGTGGGGC	CAGGCCTACA	
GCTGACGGAA	GGACAGAGAG	AGAGGTGGCA	GGGACTGCTG	GCCTCCTGAG	24600
GCAGAGCTGT	CCCAGGTCTG	GTGGGGCTAT	GATTCCAGAG	AGGCCAGAGA	
CTGAGTATGA	AAGTGGCAGG	CGGCTGGTC	TGAGGAGATG	CCAAGTTGGC	24700
CTCTTGAGGG	AAATAAACAG	GTATATTAG	CCTGTTGTGG	CCTTGCGCC	
TGAGGCCAGG	AAGCAGCTT	TTAGCCTAAA	TCCAGATGTT	AAAAACAGAA	24800
ATGAAATCAG	TATTATGGC	CCCAAACCCCT	CCAAGCAAGT	CATGCAGCTC	
GTTCCCTGT	CAAGGCCTCC	CACCTTGGC	CCACACAGGG	CCTGACCCCTC	24900
GTCTAACGCT	GCGCCCTGGG	GAACGGACCC	TGGGGTGGGA	GGTGGTGGGT	
CAGGCCCTGC	TCTCACTTTC	ACACCCGCTT	CCTAGCCCTG	AAACCAAGAGA	25000
GGTTCCGTAA	GTCCAGGCCA	GCCAGGCCTG	TGGGGCTGCT	GAGAGGGGTT	
AAGTAAGAGG	GAGAGTAAAA	ACCGACACTT	GGGAGGAGGC	TTTTAAAATA	25100
AACATCTGCG	GGGAGGGATG	CTCTCAAGGA	GGCTGCGGTT	TGCAGCTCAG	
CCAGCTGTGT	TCCAGCTAAT	CAGCCTCCTT	GGGCTACTCT	GGAGTTGCTG	25200
CCTTGGCCCT	GGGATGGGGT	ATGTTAAGAG	GACAGAGGGGA	TGTCAGGGAG	
CTGGGGGCTG	AGTTCCCTG	AGTAGAGGCT	GGCACAGGAG	AGAAGGCATC	25300
ACCCCCACCT	CGTCCAGGCC	AGAAGATGTC	CAGCTTCTTG	AGGCTCTCCT	
AAGTCTGCCT	CTCCTGGGAC	CAAGAGAAA	TCCCGGTCT	TGACCAAGGT	25400
GGGCCTTGGG	GGGAGTGGGG	TGTAGAGGGA	GGGGCACTGG	AGAACTGACT	
CTACAGAAAG	TCAAAGCTGG	CAATCCAAC	TCTTCCCTC	AGATTATAG	25500
ATGGGAAAAC	AGGCTTGAGC	CACGCAGAGA	CTTGACCAAG	CTCACACAGT	

FIGURE II

10/14	
TCCTTAGTGG CAGAGCAGAG CAAATATTCT CTTCTTTA CATTCTGGAT	25600
TTCCATATCT TCTCTCCCTC CTGGTCAGC CCAACCTCAG GGCACCCCCC	
ACGAGGTGGG CGGGGGATGG CTTTGTCACT TGACCCACT CGGGGTGCTA	25700
CTCAGAGATC TTGGGTGCAC ATAAGACGTG GGTGGGCCGA GCTTCTAGCT	
ACTCCTGTAG GCCCTTCCTG TCATTCTGTC TCTGCCTCCC TTCTCCCTGC	25800
TTCTCCGTGT TCCTCCTCAT TCTTTCTGT GTGCAGGGAG ACTACACCCC	
CCACCCCGCT CTTCTCTGG CGCCTCTGAG GTCCCCCTC TAGCCCTAA	25900
ATCACTCTGG AATCCTGGCT CTTGAAGCC AGATCTGGC CCCCCTCCCC	
TACCCCTTCC ATTCCCAGGC TGGGAAAGGC TGAAGAGGCT GACCGCTGGA	26000
CGGGAGGGGG CGGGTCGGTG CGGGATCTGG CTTCCTTTG GAGTTAATTA	
GGGAAAACAG AGAAAATGTCA GCGGAATGAA AGGGCTGGGG GTGGGGGCCA	26100
GCTGGGGTAG GAGAGGAGGA AGTGGGCAGC TGCTCCCTCC CACTCAACCC	
CTCTCCCCCG CCCCCCAGAAA GCTCTCAGCT CCGGGGATTG GCGACATGAA	26200
ATGGGGGCTG TAGAACACCTG AGCGCTGGTG CGTGAAGAGA AAAACCGAGG	
CGCATCCCAG CTCTCCCTC CCGTGTGCC TCCCTCTCTT ATGCCGGCTT	26300
GAAAATATT CCTGTCTCTC TATTCCTCAA TCCCTGGTTG ATGTCCCAGG	
ATTACTCAGC CTCTCCGGCT TTAGTCACTC TCGCTACCCG CTCCCAGGGT	26400
CCAGGGTGGG GGCAGAGGGG GGCTGGGAGA GAAGCTCTAC TGCCAGCTGG	
GCCTGGGCTG GCCTGGGCAT CCCTGAGGTT TAACTGTCT CCTAACCCACA	26500
GAGGATCTCA GGGCCTCCAG CAGCGAGCCC CAATGAGTCA AACTCTGTT	
TCCTCCTCTC CCACCCCGA CCCAGTCAGG GCAGGTGAGG GGTAGAGGTG	26600
ATGGCAGTGCT GGATGTGACG GTGTTGATGA TGATTGAC AGGATGAAGC	
ACCATCTCAT TTAGTCCTCA CAGCAGCCTT CACAGTGCCT ACAAAATCAGC	26700
TGGCAATTCC GAGAGGCTGC ATTCTAAACA AGTTCTGGG TGATGCTGAG	
CCAGGGCCAG AGTGTGGACT TCTCTGGGCC CCAGTTCTT CTTCTGTACA	26800
GTGGGACGTT GGACTGGAGG TGCTGTCTGA TGTCACCTAG CTGTGGCACT	
CTGAGCCTGT GCCTAAAGTG TCCCTGAGAT GTCTAGTCCC CGAGATCATC	26900
ATCTCCTCAG CTTCATAGAG CCGAGCTCTT CTCCATCTTC TCTCTACCTC	
CCATTCACTC AGAACAGGGG TAAGGTCAA GCTGCTGGTG GTGATGGTGA	27000
CTGACTGTCC CTTCCAGCAT GCGCTTAGGT ACACTTGGGC CTGAATGCC	
CTCTAACAAA TGCTACCGGG TATGGCCTTG GTCCCTCTAA CACTGGTTC	27100
CCTCATCTGT AAAATGAGGG TCATACACAA TAGCTGGCTG GGCTGTATG	
TAGAGGTTAA AAGTGATTGT GCATGTGAAG CATCCAGCAG AGTGCCTGGC	27200
ACACAGTAGG TGCTCAAATG CTGTTTGAA ATACAAAAAT TAGCTGGGCT	
AGTGGTGTGC ACCTGTAATC TCAGCTACTC AGGAGGTTGA GGCAGGAGAA	27300
TCACATTGAGC CCAGGAGGCA GAGGTTGCAG TGAGCCGGGA TCGCGCCACT	
GCACTCCAGG CTGGGCGACA GAGCAGGACT CCATCTAAA AAAAAAAA	27400
AAAAAAATGCT GTTTGAAAT GGAGGCTTGG AAGAGCACTC TTCACCCCCA	
CCCCCACACAC TTACTAACCG AACTACCGA TCTCTACTCG TGCCAGGCCT	27500
TGCTTGGCAA AGGCTAGGGG CATGTAAAGC CATGGTTCA TGGCTGCTGC	
CCTCCAGGAG CCCATGGTCT AGTGTGGAG TCAGAGGCTC CCTTGAGAAA	27600
GCCAGGGCAA GGAACGTGAGA GCCCCTGTCA CAGCCCTGAG AGAAATACCA	
GCAGATGTGT TTTGAGCACT TACATCTACC AGGCACTATT TTACGTACCT	27700
GACATACATT TTCTTATTTA GCATTCACTA AAACCAAGAG AGGTGGTTAC	
CATTATCATC CCCATTTGC AGATGAGAAA ACTGAAGTGG CATAAGGAGG	27800
TTAGGTAGAC TTGCCAGGA TCACAGTGTAG TAAGGAAAG AGCTAGGAAT	
CAAACCTAGA CTTCTCACT CCAGAGCCTC TTAACCTCTA CCCTCAACTC	27900
TTGTGAAGAG AGACTACCTT GGTGACCCCT ATATTCCCAG TGCTCAGCGT	
GGTCAAGGCC CAGGGTGTG TATAGGCATT CATTAACTGC TGGGGATATA	28000
AGACAGTGAC TGTGAGGAC CCTAAGTTA GCTCCCACCT GATCTCCTC	
TGTCTCTGCA GAAGCTGCTG ATGGGAGCTC CACCCCTGGGA GGGGGTGCCTG	28100
[Exon 4: 28062..	
GCACCATGGG ATTGAGCGCC CGCTACGGAC CCCAGTTCAC CCTGCAGCAC	
G	
GTGCCCGACT ACCGCCAGAA TGTCTACATC CCAGGCAGCA ATGCCACACT	28200
GACCAACGCA GCTGGCAAGC GGGATGGCAA GGCCCGAGCA GGTGGCAATG	
GCAACAAGAA GAAGTCGGGC AAGAAGGAGA AGAAGTAACA TGGAGGCCAG	28300
..28288]	

FIGURE 1J

11/14

GCCAAGAGCC ACAGGGCGGC CTCTCCCCAA CCAGCCCAGC TTCTCCTTAC	
A	
CTGCACCCAG GCCTCAGAGT TTCAGGGCTA ACCCCCCAGAA TACTGGTAGG	28400
GGCCAAGGCC ATGCTCCCT TGGGAAACAG AAACAAGTGC CCAGTCAGCA	
A	
CCTACCCCTT OCCCCCCAGG GGGTTGAATA TGCAAAAGCA GTTCCGCTGG	28500
C	
GAACCCCCAT CCAATCAACT GCTGTACCCA TGGGGGTAGT GGGGTTACTG	28600
TAGACACCAA GAACCATTG CCACACCCCG TTAGTTACA GCTGAACTCC	
TCCATCTTCC AAATCAATCA GGCCCATCCA TCCCATGCCT CCCTCCCTCC	
CACCCCACTC CAACAGTTCC TCTTCCCAGA GTAAGGTGGT TGGGGTGTG	28700
AAGTACCAAG TAACCTACAA GCCTCCTAGT TCTGAAAAGT TGGAAAGGGCA	
C	
TCATGACCTC TTGGCCTCTC CTTTGATTCT CAATCTTCCC CCAAAGCATG	28800
GTGGTGTGCC AGCCCCTTCA CCTCCTTCCA GAGCCCAAGA TCAATGCTCA	
AGTTTTGGAG GACATGATCA CCATCCCCAT GGTAATGATG CTTGCTGGAT	28900
TTAGGGAGGG CATTTGCTA CCAAGCCTCT TCCCAACGCC CTGGGGACCA	
GTCTTCTGTT TTGTTTTCA TTGTTTGACG TTTCACACTGC ATGCCTTGAC	29000
A	
TTCCCCCACC TCCTCCTCAA ACAAGAGACT CCACTGCATG TTCCAAGACA	
GTATGGGTG GTAAGATAAG GAAGGGAAGT GTGTGGATGT GGATGGTGGG	29100
GGCATGGACA AAGCTTGACA CATCAAGTTA TCAAGGCCCTT GGAGGAGGCT	
CTGTATGTCC TCAGGGGACT GACAACATCC TCCAGATTCC AGCCATAAAC	29200
CAATAACTAG GCTGGACCCCT TCCCACATACA TAATAGGGCT CAGCCCAGGC	
AGCCAGCTT GGGCTGAGCT AACAGGACCA ATGGATTAAA CTGGCATTTC	29300
AGTCCAAGGA AGCTCGAAGC AGGTTAGGA CCAGGTCCCC TTGAGAGGTC	
AGAGGGGCCT CTGTGGGTGC TGGGTACTCC AGAGGTGCCA CTGGTGGAAAG	29400
GGTCAGCGGA GCCCCAGCAG GAAGGGTGGG CCAGCCAGGC CATTCTTAGT	
T	
CCCTGGGTTG GGGAGGCAGG GAGCTAGGGC AGGGACCAAA TGAACAGAAA	29500
G	
GTCTCAGGCC AGGATGGGGC TTCTTCAACA GGGCCCCCTGC CCTCCTGAAG	29600
CCTCAGTCCT TCACCTTGCC AGGTGCCGTT TCTCTTCCGT GAAGGCCACT	
GCCCAGGTCC CCAGTGCAGCC CCCTAGTGGC CATAGCCTGG TTAAAGTTCC	
A	
CCAGTGCCTC CTTGTGCATA GACCTTCTTC TCCCACCCCC TTCTGCCCC	29700
GGGTCCCCGG CCATCCAGCG GGGCTGCCAG AGAACCCCCAG ACCTGCCCTT	
ACAGTAGTGT AGCGCCCCCT CCCTTTTCG GCTGGTGTAG AATAGCCAGT	29800
T	
AGTGTAGTGC GGTGTGCTTT TACGTGATGG CGGGTGGGCA GCGGGCGCG	29900
GGCTCCCGCG AGCCGTCTGT CCTTGATCTG CCCCGGGCGG CCCGTGTTGT	
GTGGTGTGCT GTGTCCACGC GCTAAGGCAG CCCCTCCCC CGTACTGACT	
TCTCCTATAA GCGCTTCTCT TCGCATAAGTC ACGTAGCTCC CACCCCCACCC	30000
TCTTCCTGTG TCTCACGCAA GTTTTATACT CTAATATTTA TATGGCTTT	
TTTCTTCGAC AAAAAAATAA TAAAACGTTT CTTCTGAAAA GCTGAACGTT	30100
TCTGTATAAG CGATGGAAGC TCCTGGCATG TGTGCATGAA GTGATGAGCT	
GAGGTGGGTG CTGGAAGAAG GGCGGAATCG GGAGGCCACT CTGTGTGATT	30200
GCGCGTCTAG ATGTTCCGA ATTGCGTGTG TGTGTGTGAC TGTG	30244

12/14

POLYMORPHISMS IN THE CODING SEQUENCE OF PCDH2

ATGGTCCCAG AGGCCTGGAG GAGCGGACTG GTAAGCACCG GGAGGGTAGT	
GGGAGTTTG CTCTGCTTG GTGCCTGAA CAAGGCTTC ACAGGTCAATT	100
ACTATGAGAT CCCGGAGGAA AGAGAGAAGG GTTTCGCTGT GGGCAACGTG	
GTCGCGAACCC TTGGTTGGA TCTCGGTAGC CTCTCAGCCC GCAGGTTCCG	200
GGTGGTGTCT GGAGCTAGCC GAAGATCTT TGAGGTGAAC CGGGAGACCG	
GAGAGATGTT TGTGAACGAC CGTCTGGATC GAGAGGAGCT GTGTGGGACA	300
CTGCCCTCTT GCACGTAAAC TCTGGAGTTG GTAGTGGAGA ACCCGCTGGA	
GCTGTTCAAGC GTGGAAGTGG TGATCCAGGA CATCAACGAC AACAAATCCTG	400
CTTCCCAC CCAGGAAATG AAATTGGAGA TTAGCGAGGC CGTGGCTCCG	
GGGACGCGCT TTCCGCTCGA GAGCGCGCAC GATCCCGATG TGGGAAGCAA	500
CTCTTACAA ACCTATGAGC TGAGCCAAA TGAATACTTT GCGCTTCGCG	
TGCAGACGCG GGAGGACAGC ACCAAGTACG CGGAGCTGGT GTTGGAGCGC	600
C	
GCCCTGGACC GAGAACGGGA GCCTAGTCTC CAGTTAGTGC TGACGGCGTT	
GGACGGAGGG ACCCCAGCTC TCTCCGCCAG CCTGCCTATT CACATCAAGG	700
TGCTGGACGC GAATGACAAT GCGCCTGTCT TCAACCAGTC CTTGTACCGG	
C A	
GCGCGCGTCC TGGAGGATGC ACCCTCCGGC ACGCGCGTGG TACAAGTCCT	800
TGCAACGGAT CTGGATGAAG GCCCCAACGG TGAAATTATT TACTCCTTCG	
GCAGCCACAA CGCGGCCGGC GTGCGGCAAC TATTGCCTT AGACCTTGTA	900
T	
ACGGGGATGC TGACAATCAA GGGTCGGCTG GACTTCGAGG ACACCAAAC	
CCATGAGATT TACATCCAGG CCAAAGACAA GGGCGCCAAT CCCGAAGGAG	1000
CACATTGCAA AGTGTGTTGGT GAGGTTGTGG ATGTGAATGA CAACGCCCG	
GAGATCACAG TCACCTCCGT GTACAGCCCA GTACCCGAGG ATGCCCTCT	1100
GGGGACTGTC ATCGCTTTGC TCAGTGTGAC TGACCTGGAT GCTGGCGAGA	
ACGGGCTGGT GACCTGCGAA GTTCCACCGG GTCTCCCTT CAGCCTTA	1200
TCTTCCCTCA AGAATTACTT CACTTGAAA ACCAGTGCAG ACCTGGATCG	
GGAGACTGTG CCAGAACATA ACCTCAGCAT CACCGCCCGA GACGCCGGAA	1300
CCCCTCCCT CTCAGGCCCT ACAATAGTGC GTGTTCAAGT GTCCGACATC	
AATGACAACC CTCCACAATC TTCTCAATCT TCCTACGACG TTTACATTGA	1400
AGAAAACAAC CTCCCCGGGG CTCCAATACT AAACCTAAGT GTCTGGGACC	
CCGACGCCCG GCAGAACATGCT CGGCTTCTT TCTTCTCTT GGAGCAAGGA	1500
GCTGAAACCG GGCTAGTGGG TCGCTATTTC ACAATAAATC GTGACAATGG	
CATAGTGTCA TCCTTAGTGC CCCTAGACTA TGAGGATCGG CGGGAATTG	1600
AATTAACAGC TCATATCAGC GATGGGGCA CCCCCTGCCT AGCCACCAAC	
ATCAGCGTGA ACATATTGT CACTGATCGC AATGACAATG CCCCCCAGGT	1700
CCTATATCCT CGGCCAGGTG GGAGCTCGGT GGAGATGCTG CCTCGAGGTA	
CCTCAGCTGG CCACCTAGTG TCACGGGTGG TAGGCTGGGA CGCGGATGCA	1800
GGGCACAATG CCTGGCTCTC CTACAGTCTC TTGGGATCCC CTAACCAGAG	
CCTTTTGCC ATAGGGCTGC ACACGGTCA AATCAGTACT GCCCGTCCAG	1900
TCCAAGACAC AGATTCAACCC AGGCAGACTC TCACGGTCTT GATCAAAGAC	
AATGGGGAGC CTTCGCTCTC CACCACTGCT ACCCTCACTG TGTCAGTAAC	2000
CGAGGACTCT CCTGAAGCCC GAGCCGAGTT CCCCTCTGGC TCTGCCCGCC	
GGGAGCAGAA AAAAATCTC ACCTTTATC TACTTCTTTC CCTAATCCTG	2100
GTTTCTGTGG GGTTTGTGGT CACAGTGTTC GGAGTAATCA TATTCAAAGT	
C	
TTACAAGTGG AAGCAGTCTA GAGACCTATA CCGAGCCCCG GTGAGCTCAC	2200
TGTACCGAAC ACCAGGGCCC TCCTTGACAG CGGACGCCGT CGGGGGAGGC	
CTGATGTGCGC CGCACCTTA CCATCAGGTG TATCTCACCA CGGACTCCCG	2300
CCCCGGCGAC CCGCTGCTGA AGAAACCTGG TGCGAGCCAGT CCACTGGCCA	
GCCGCCAGAA CACGCTGCGG AGCTGTGATC CGGTGTTCTA TAGGCAGGTG	2400
TTGGGTGCAG AGAGCGCCCC TCCCAGACAG CAAGCCCCGC CCAACACGGA	
CTGGCGTTTC TCTCAGGCCAG AGAGACCCGG CACCAGCGGC TCCCAAATG	2500
A	
GCGATGACAC CGGCACCTGG CCCAACAAAC AGTTTGACAC AGAGATGCTG	
CAAGCCATGA TCTTGGCGTC CGCCAGTGAA GCTGCTGATG GGAGCTCCAC	2600

FIGURE 2A

13/14

CCTGGGAGGG GGTGCCGGCA CCATGGGATT GAGCGCCCGC TACGGACCCC G	
AGTTCACCCCT GCAGCACGTG CCCGACTACC GCCAGAATGT CTACATCCA	2700
GGCAGCAATG CCACACTGAC CAACGCAGCT GGCAAGCGGG ATGGCAAGGC	2800
CCCAGCAGGT GGCAATGGCA ACAAGAAGAA GTCTGGCAAG AAGGAGAAGA	
AGTAA	2805

14/14

ISOFORMS OF THE PCDH2 PROTEIN

MVPEAWRSGL VSTGRVVGVVL	LLL GALNKAS TVIHYEIPEE REKGFAVG	100
VANLGLDLGS LSARRFRVVS	GASRRFFEVN RETGEMFVND RLDREELCGT	
LPSCTVTLEL VVENPLELFS	VEVVIQDIND NNPAFPTQEM KLEISEAVAP	
GTRFPLESAH DPDVGNSNLQ	TYELSRNEYF ALRVQTREDS TKYAEVLER	200
ALDREREPSL QLVLTALDGG	TPALSASLPI HIKVLDANDN APVFNQSLYR	
P		
ARVLEDAPSG TRVVQVLATD	LDEGPNGEII YSF GSHNRAG VRQLFALDLV	300
TGMLTIKGRL DFEDTKLHEI	YIQAKDKGAN PEGAHCKVLV EVVDVNDNAP	
EITVTSVYSP VPEDAPLGT	IALLSVTLD AGENGLVTCE VPPGLPFSLT	400
SSLKNYFTLK TSADLDRET	PEYNLSITAR DAGTPSLSAL TIVRVQVSDI	
NDNPPQSSQS SYDVYIEENN	LPGAPILNLS VWDPDAPQNA RLSFFLLEQG	500
AETGLVGRYF TINRDNGIVS	SLVPLDYEDR REFELTAHIS DGGTPVLATN	
ISVNIFVTDR NDNAPOVLYP	RPGGSSVEMI PRGTSAGHLV SRVVGWDADA	600
GHNAWLSYSL LGSPNQSLFA	IGLHTGQIST ARPVQDTDSP RQTLTVLIKD	
NGEPSLSTTA TLT VSVTEDS	PEARAEFPSG SAPREQKKNL TFYLLLSLIL	700
VSVGFVVTVF GVIIFKVKW	KQSRDLYRAP VSSLYRTPGP SLHADAVRG	
LMSPHLYHQV YLT TDSRPGD	PLLKKPGAA PLASRQNTLR SCDPVFYRQV	800
LGAESAPPQ QAPPNTDWRF	SQAQRPGTSG SQNGDDTGTW PNNQFDTEML	
S		
QAMILASASE AADGSSTLGG	GAGTMGLSAR YGPQFTLQHV PDYRQNRYIP	900
GSNATLTNAA GKRDGKAPAG	GNGNKKKSGK KEKK	934

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Kliem, Stefanie E.
Koshy, Beena
Tanguay, Debra A

<120> Haplotypes of the PCDH2 Gene

<130> MWH-0774PCT PCDH2

<140> TBA
<141> 2001-06-06

<150> 60/209,564
<151> 2000-06-06

<160> 125

<170> PatentIn Ver. 2.1

<210> 1
<211> 30244
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (7653)..(7712)
<223> Nucleotide identity unknown

<220>
<221> unsure
<222> (19185)..(19244)
<223> Nucleotide identity unknown

<220>
<221> unsure
<222> (19295)
<223> Nucleotide identity unknown

<220>
<221> unsure
<222> (19327)
<223> Nucleotide identity unknown

<400> 1
cacatgggcc cgtagggaga gctttagaac tagtgcagct taattgagca gggggaaagcc 60
ttctgcagac agttgagcct agcgagggtt gttgacacg gaaggggctt gatTTTggg 120
aaggtagac ttaaatttgg gattccaag gatttgtatg ttaatccaag ggccgtttat 180
gcgaattaag gaccagatt gaaaatttggg gaacccaagt taaaatggtt tgggaacgtg 240
taacottcgg ttattgttgg gggaatggct aacccaaagg ggttatttgg gttcatgaag 300
aggtottaag aaaactatga gctgggtgca gtatctcagg cctgtatcc catctactta 360
gaagtcgtgag gcgggaggat cccttgagcc caaggagctg gaagttacag tgagctatga 420
ttattccact gtactccagc ctggatgaca gagcgagatc ctgcctcaga aacaaaaacc 480
aaaaaataaa caacaacaaa aacaacaaca acgaaaaatct cagaaaaacta caggccttat 540
tatcaagaaa atgacaggcc ttattattga gaaaactact tacatTTTA aaagctgatt 600
ttcaaaaaacc agattgtgaa ttaagtgtatg ctgaataggg gaaaaaaaaaga ctgataaacc 660
agtgtatctc tgaataaccc actagatcag tgctgtcaa acttaatgt gtttcaaat 720
caccctgtatg ttttgtaaa atgtaaatca tgtctcagta ggtctagagg agggcattc 780
tgcatttcta acaagctccc cggtgatgcc tagaccattt gtcctcacac tttgagtgc 840

aagtgattgg ggtatactct ttattctaaa ttccaaatgt attctatctt ctaatgccta 900
 tcagccctct tgctggctg gctaattatc aaacttgagc atttaaaaact gggggtttc 960
 agagattaat tgaagtgtt aagacatgga aatataggtc acatatggct tttcatctct 1020
 tccggaaaata tcctatatac ccttacttct ttcttaccta cctaaccat gtgcctaccc 1080
 ctcaacttat gttcaagggtt ctattataact tcaaaaggac tctatgcaat caaagtgtatg 1140
 tccaagacaa atatacagaa atagatgag gctctaaccac tttacatcta atggagtggt 1200
 ttgcatttattt aattgaggct tactctgtgc cagggactt acctgccttc ttcatttat 1260
 aggataggag cagggacttct tagaggttca ttcttgc tttttttttt aaggatatac atctcatctc 1320
 tggcttatac cacagtggac tcttttattt actatcctaa tcattataat tttgttttgtt 1380
 gacaacaattt ctaaatcgac aaaagttaaag gctcttttattt ctgtgactaa aggctttgtt 1440
 tttttttttttt aatttggggat ctttgcgtcc attatttcaa taattttata aagatccctag 1500
 gtggtaaaata ggcctggctt tttttttttt tgaaatcttc tccgtcttct tcttcatttag 1560
 cagtgcaact tcttgcgttca tatggtaccc aagatcttgg gaaaaggatt gatcaagtaa 1620
 agggaaacga tagagagagt gtgctctaca actgagatca tgctccttgc gaaagacagg 1680
 gtagggaaaat ttcttatttgc tccctgggtt ttaagaaaat agagctaaaca atgacatccca 1740
 gttttttttt aataaaaaatc tgaattgtac tttaaattt tataagccag agttaaacca 1800
 atagattgtt actccctgttca atttgatagt agctccagat aatgaggaag tgacaggaa 1860
 tgctaatttc atttggtaag agataaaaact gaaatttata atgtctccac gaaaatgcta 1920
 aatgcctcct tacaagtagg gtccggctaa ttgtctgtct tccaaacaagc cagatttgtt 1980
 ggtttttttt cagacaatgt tttttttttt tttttttttt tttttttttt ttcttgcgtat 2040
 caaagaaaatg gattcaaatg tttaatgagt cttgatttgg attggagact tgccagaacgg 2100
 cttagcctcac agccccaaagg ctggctttcc ctaaggttagg tttccatgc accgatactg 2160
 gcaaggcgct ttggctggaa actctggaaacttggc gaaagccaaag gaaagtgaag ttccctggcgc 2220
 tagcgcgtgt cttgtccaga gagccccggc ctagcttattt cttcgtgcag ttattggctg 2280
 ggactctgtg tgccgcgtgtc ggccaaatgaa gacgctggag atcggggcccc tgcccgtccc 2340
 ctttctgcgc cccggatgaa ggcagagact gaacagccgg cgagcaaaatc aacggcatcc 2400
 agaaagccat gtcggactcg ggcggccagcg cccaaaggctt aaccggctgtaa aagtttctca 2460
 gcgaaatctc aggggacgttc tgacccggc tgagaggaac tgcttttgc tgagatggc 2520
 ccagaggcct ggaggagccg actggtaagc accggggagg tagtgggagt tttgcctctg 2580
 ctttgcgcct tgaacaaggc ttccacggc attcaactatg agatccccga gaaaagagag 2640
 aagggtttcg ctgtggccaa cgtggcgtcg aaccttgggt tggatctcg tagcctctca 2700
 gcccccgagg tccgggttgtt gtctggagct agccgaagat tctttggatg ttttgcgtt 2760
 accggagaga tttttgtgaa cgaccgtctg gatcgagagg agctgtgtgg gacactgcc 2820
 tcttgcactg taactctgtt gttggtagt gagaacccgc tggagctgtt cagcgtggaa 2880
 gtggatgtcc aggacatcaa cgacaacaat cctgttttcc ctacccagga aatgaaattt 2940
 gagattagcg aggccgtggc tccggggacg cgctttccgc tcgagagcgcc gcacgatccc 3000
 gatgtggaa gcaactctt acaaaccat gagctgagcc gaaatgaata ctttgcgtt 3060
 cgcgtgcaga cgcggggagga cagcaccaag tacgcggagc tggtgttggc ggcgcctctg 3120
 gaccgagaac gggagccctag tctcaggtt aatgtgttgc gttgttgcgg cggtggacgg agggacccca 3180
 gctctctccg ccaggcctggc tattcacatc aagggtctgg acgogaatga caatgcgcct 3240
 gtcttcaacc agtccctgtt ccggcgccgc gtcctggagg atgcacccttc cggcacgcgc 3300
 gtggatacaag tccttgcac ggtatctggat gaaggccccca acggtaaat tatttactcc 3360
 ttccggcagcc acaaccgcgc cggcggtcg aactattcg ctttagaccc tptaaccggg 3420
 atgctgacaa tcaagggtcg gctggacttc gaggacacca aactccatga gatttacatc 3480
 caggccaaag acaagggcgc caatcccgaa ggagcacatt gcaaaagtgtt ggtggagtt 3540
 gtggatgttca atgacaacgc cccggagatc acagtccacct ccgtgtacag cccagttaccc 3600
 gaggatgccc ctctgggac tgcattcgat ttgtctgtt gactgtaccc ggtatgttgc 3660
 gagaacgggc tgggtgacccg cgaagttcca ccgggtctcc ctttgcgtt tacttctcc 3720
 ctcaagaattt acttcaattt gaaaaccatg gcagacctgg atcggggagac ttttgcgtt 3780
 tacaacccatca gcatcaccgc ccggagacgc ggaacccctt ccctctcagc ctttacaata 3840
 gtgcgtgttc aagtgtccga catcaatgac aaccctccac aatcttctca atcttccatc 3900
 gacgtttaca ttgaagaaaa caacccccc gggctccaa tactaaaccc aagtgttgc 3960
 gaccccgacg ccccgccagaa tgctggctt tctttcttcc tcttggagca aggagctgaa 4020
 accggggctag tgggtcgcta tttcacaata aatcggtaca atggcatagt gtcataccctt 4080
 gtggcccttag actatgagga tcggcgaa tttgaattaa cagctatcat cagcgtatgg 4140
 ggcacccccc tcctagccac caacatcgcg gtggactt ttttgcgtt tacttctcc 4200
 aatgcctttttt aggtccatca tcctcgccaa ggtggggagct cgggtggagat gctgcctcga 4260
 ggtacccatg ctggccaccc agtgcacgg gtggtaggtt gggacgcggc tgccaggcgc 4320
 aatgcctggc tcttgcacag tcttggaa tcccctaaacc agggcccttt tgccataggg 4380
 ctgcacactg gtccaaatcg tactgcccgt ccagtccaaag acacagattc acccaggcag 4440
 actctcactgg tcttgcgttca agacaatggg gagccttcgc tctccaccac tgctaccctc 4500

gtgaggggaca atggtgaccc ttcactctcc tccacagccca cagtgtctgc ggttctggag 8220
 gatgaggacc ctgagggaaat gcccaaattcc agtgacttcc tcatacaccc tcctgagcgt 8280
 tcagaccta ccctttacact catttgtggct cttagcgaccg tcagtctctt atcccstagtc 8340
 accttcaccc ttctgtcagc gaagtgcctt cagggaaacg cagacgggga cgggggttgg 8400
 gggcagtgct gcaggcgcca ggactcaccc tccccggact tctataagca gtccagcccc 8460
 aacctgcagg tgagctcgga cggcacgctc aagtacatgg aggtgacgct gcggccacaca 8520
 gactcgcaga gccactgcta caggacgtgc ttttacccgg ccttggacgg cagtgaacctc 8580
 actttctaa gaccctctag cggttacgacg cccacagetc tggcgttgg 8640
 atccgttccc gctctaatac gctgcggag cggagccagg tgaggggctc 8700
 cggggcggaccctt cttggggcg 8760
 catccactcc tctccggccg gtttggctcg tggctgcgtt ccacccgatt ctgggatca 8820
 ttggaccgtt tgcgcgaaac cagagtggcc gattaaggga tggggctccg 8880
 gtgggtggcga ctgtgggcga 8940
 gccggggcct cttcgagct tccggtaat ttccggcgat ttccggcggt 9000
 cggggaggag gcagtcaacag atccacccctt gcagccagcc tcctaggcgc 9060
 acgcttcgccc gttctgtaga tttcttcttc gatttctccc cagtccttc 9120
 ttcaactgtta ccctccctat ccccgcatca cccaaccgca cctgtctcg 9180
 gtgcgcgcgg ggctcatgcg tgcctccctt gtcggccacc cccacggccc 9240
 cacgggtctcg ccacgccccg ccaacacgtg cggggacgca cgcacgcact 9300
 gggcttaacgc gaataccagc tttcaactgcc actcgctcgc ggccagattc 9360
 tccggccacac tccggccatcc cctatgccc tccctccggc 9420
 ctgatttatgc ggcgcgttggg gtcccgatgc gggccggccc aggaccaggc 9480
 gagccttcctc tcaaccttcc cacctgcgc 9540
 cctgagcgag ggcggggggc aggagcgctg gagcgacttc 9600
 caggactcta cgatccttgg gccagaggc 9660
 cggcgacccca tcaacccaga agcctcgacg aggccgacag gcagagctgc 9720
 aggccggca gggctccgcg ttggcgagt gagtggactt ctataggaca 9780
 gactccatgt gcaccagcg 9840
 gctctaacc cccagagata caggcttttgc ggcggctgt 9900
 tctagggtca gagatttgg 9960
 tctggcagat gaaaatgtat aggctctgcc agatgccagg 10020
 actacaatag taaaaaaacac aagtttctcc aactccagga 10080
 aatgcctaacc tctgtctcta ggactgtatt ttgaaacacc 10140
 gattgagaga aacttactat ataactccctt tgagaacccctc 10200
 cattgtgtt ggcttaggtac taacatgcctt ctcttataac agtctcacag 10260
 acttgctcaa gttcacagag caagtaagag agattctaacc 10320
 tggctttta atttcttgcg acttggactt taaaagctt 10380
 gcctctgtg gattttgc tctgtgg 10440
 ggaccttaca tggggggaaa gtcttctgtg gtccttcatt tctgagcagt 10500
 caaacttggc cagttatctt tttggagatt 10560
 caatcaactgc atttccatgt agattctgt 10620
 cccatccctgt gccccttc 10680
 tggaaatgtca gtttcttggg ttctgtctcat 10740
 cactggcagt tccggagctt ccttctccctt 10800
 ttctgggagc aagtatgagt attaattacc 10860
 tcttagggatgt ctccatgaga gtgtgtatg 10920
 ggactcagga atgaagtggc tggccagagc 10980
 acacatcagc cccttgggag ccctggagac 11040
 ctggatccctt ccccgctgca tcccttcattt 11100
 tctcttc 11160
 cgggacttgg atgccccca aaggccaaat 11220
 tagctgtggg aagagatct tgcaagccaa 11280
 tcactgcaca acctggcacc 11340
 atgctttgtt cccagctgag gagttgaata 11400
 tgaaaagccctt cttttggccac 11460
 gaagatgcctt ctagtgcctt aattttggctt 11520
 agggggccattt tcttttaggaa 11580
 tttatcttgcgtt acctcgaccc 11640
 agatgagtc accccctgccc 11700
 tctctcccg 11760
 cccttccat ctttttc 11820

ggcccagaga cccggcacca gcgggttaggt gactgattct ccagccccacc ctcttctctg 11880
 cggcatttc tcacggatga cgtgagagca gatgggggag ggcccaccat ttgctacaca 11940
 tggctctcc ctcagttta gatcccagg aggtcttggt gtgcgggggg ctggcacaca 12000
 gaccggaa ggaagaggcg actgccccta ctgttcagga agctcaatic acatgctgc 12060
 cccttccctc ctcgcccacga ccggcacctt ttcctatccc ctgagggcac tggaaacca 12120
 aaggatggtc ttaagcttgt ctctgggtga aagcgaggct ttctatgcc atgtactgcc 12180
 taacccttc ccctgagttt agctgggtc cattatgacc tgggctcag gcagaaaagc 12240
 atttgcacgg aggaggcggt ccgcactcag cgccctctcc agaggctcca gagccgaggc 12300
 tgactgcagc ctggggagag tggaggcagc acagctggag gtgaaactga ggaaggactg 12360
 ggcggggcat ggagcaggcc ccctctccg gcccctccctc ccactgtcct ctgcccctac 12420
 ttgtctctgt ctctgtctgt ggggtctccg tgtctctgcc ccttttctt gagtttccgt 12480
 ctttgcact ttctcttacc ttcagtcct tccctcagtc tctatctcgc tttgcaatct 12540
 ctgcctctcc ctctcttc ttcatctgtc ttagcttccg tcttattgtc tgcacccctac 12600
 cctcagtccc tttcatttgt ctacttgcatt tgatctgtgc cgcccaactgt gcctgcatg 12660
 gcataagtgc tcaataaaatg tggagttagt aactacacgg gacccttagt ctcttctcc 12720
 ttcttatct ctgcctccct gtccttgc tggacctt ttcttgc tcttcacccg 12780
 tttcttagcg ctttgcgtc ttctatcccc agcctctatg ttttctgtc tctcactatt 12840
 tctgcttcc tctctgttct ttgtctctt ttgtctctgt ctctgtatat ctttcttgc 12900
 ctctgtctct ggtctctgtc ccttgcgtt ttttctgc ttttctctgt tactcttct 12960
 ttatctctct ttttctttt ttcgttctt ttctctgtc ccttgcgtc tgcgtctcat 13020
 tttttgtgc ttgtgggcaa gccagcacac acaacccccc acccaccacc caccacacc 13080
 ccgctctctt cttccctgc ccctcacaca ctgagcctt gatgcagct ctccaccaga 13140
 cacccctccca tccaagcagc cccagctgct catttcaatc tggatgaaat tcctgctgag 13200
 acaggaaccc cctgcgggct gaaggggagg gaaacttcag cagaaagacc ttcaagggtt 13260
 ctgagcagag tggatagtc tctgcgcag gcctccctg tagaagtcag gagcctggag 13320
 gagtcctaga gcccaggag agaagcagga gcttgggca ctttacccta cttccctcaat 13380
 ggtttagtac ttgtctgtca catttgc ttttgc ttttgc ttttgc ttttgc ttttgc 13440
 atcttcccc catcctgca ccaacccaaac ctttgc ttttgc ttttgc ttttgc ttttgc 13500
 ggttatggc agagattca gcttctggc atgcacagtc ctgtcaccata tttttaggaac 13560
 gcaaagcgct tcttagaagg cccctaaaaa gcagcagatc aaaaggctt ggctgcctt 13620
 gcccctccct tgacccccc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 13680
 agttatccat taagtcat ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 13740
 gcatctacta tgagcaggc ctttgc ttttgc ttttgc ttttgc ttttgc ttttgc 13800
 gtgatcccag aactcatagg cttccagacc agcagaggag actgaaaatt tgcaacaagt 13860
 aacactaaaa aaaaaaaaaa aaggctggc actgtggtc acgcctgtaa tcccaagcact 13920
 ttggaggcc aaggtggta gatcacttgc ggtcagggtt ttgagaccag ctttgc 13980
 aacatgtga aaccccatct cgataaaaaa tacaaaaatt agccaggcgt ggtggcaggc 14040
 acctgtatcc ccagcgtactt aggaggctga ggcagaactg ctttgc 14100
 gttcgttgc gcttgc gatcacttgc cgatcttgc ataatcttgc cacaatcact agtaagtct 14160
 gagaggcaga gaacaatgtt agttaatgg gggggggaaat atcagagaag gcttcttaga 14220
 ggaggtgaca ttgagaagg ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14280
 ccatacactg gcttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14340
 tcatgtgtcc tctgggttgc gcaaccttacc ctttgc ttttgc ttttgc ttttgc ttttgc 14400
 aggcttagtgc ctttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14460
 acctctctt ctctcccttc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14520
 ttttttttt ttttttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14580
 cgatcttggc tcaacttgc ctttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14640
 cccgagtagc tgggacagca ggttgcacc accacacccca gcttgc ttttgc ttttgc 14700
 tagagatggg ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14760
 cacctgttcc agcttccca agtgcgttgc ttacaggttgc gatgcggggcc gggccacaaca 14820
 aacattttg attcagcgttgc tccaaatgtt aataagcgtt aggatctgtt ccagtaactt 14880
 caacttctt ctttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 14940
 gtgttcttc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15000
 atatttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15060
 gaatttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15120
 ctttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15180
 agaagaaga gcatcgttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15240
 ttcttc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15300
 gaagagcatt ctttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15360
 caaggaaatg gcatcgttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15420
 gtttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc ttttgc 15480

ctccccacca tcttgcccac tgccctccag cctggtctgg tcagacacac cagcttcaaa 15540
 gcaggcgtag tattgagttc ttcatgtaaac ttgggtcaag ttcaagtcag cactgggtca 15600
 cactgcctag acagaaggaa gaagacctgg ccaggtgtgg tggctcacgc ctgtaatccc 15660
 agcactttgg gaggctgagg tgggtggatc acctgagggtc aggagttga gaccagcctg 15720
 gccaacatga taaaacccca tctctagtaa aaatacataaa tattagccag gtgtgggt 15780
 gcacacctgt agtcccgact actcaggagg ctgaggcagg ggaatcgctt gaacctcagga 15840
 ggccggaggtt gcagtggact gagatcacac cactgcactc cagtcgtgg gacagagcga 15900
 gactccatca gaaaaaaaag aaagaaaaag agagagaggg agggagggg ggaaggaagg 15960
 aaggaaggaa ggaaggaagg aaggaaggaa ggaaggaagg aagaaaaagaa agaaaaaagg 16020
 agaaaaaaatg aagaagactt acattctgaa gttcctgctt ggcacatgct tctcctcccc 16080
 accactatcc cttctcaggat catccttgg gtgtctgatc ccaatagctg ttgtctcaag 16140
 ctctgagccc agcaaaccat ttcccccttc tttaggctgtt acccaggccc tgccctgcag 16200
 ctgtccccag cctctgcaca aagagtctcc atttggtccc tagactgttc tctgatggct 16260
 ccacactacc atccctccctc tgagagacag tatccctgct ctctctcagg ttggcagtc 16320
 attcaacaaa caaatagcaa cttaaatttt ttccacttat aaaatagtagtac atgctcatta 16380
 tagaaacatg aaaaatagaa aaaaaattaa aaggaaaaacc acccatcatt ttacaatcta 16440
 gggagaacca ccaccaacta cagtttaata taaaacattt caaatatgtt caaaagttaga 16500
 gagaatggtg tcatgaacact gtatgatacc actatccagc ttcaactaat gccttatctt 16560
 gtttacacca tattccctacc cttatcctgt cttggattc ttttgaagca aatctcagac 16620
 atcatataat ttcatcttgg ttcacccat cttaaacaga tgactttttt tttttttttt 16680
 tttttggag acagagtctc actctcttgc ccaggctgtg gcacaatctc agctcaetgc 16740
 agcctcgcgc tcccgggttc aagcaatttct catgctcagg cctcttaagt agctaggatt 16800
 acagggtgtc accaccacac acagactataa ttttgtattt tcagtagaaaa cagggtttcg 16860
 ccatgttggc caggctggc ttgaactctt ggcctcaagt gatccacccct cctcggccctc 16920
 caaaaatgtt gggattacag gtgtgagcca ctttgcgggg ccagatgattt cttcataaag 16980
 gtccacattt tatatttgag tgagtgtctc ttaagtatct taatgtctt ttaatgtaaa 17040
 agactcccc tccatcttcc atttttgcatttttgaatccctgc cttttcatat 17100
 atatcttctt aaactttct ttgtgtatc aaagaacatg catgagcttc atttttattt 17160
 ttattttttt attttttat ttattttattt attttttat tgatacgttag ctttgetctg 17220
 tcaccaggc tggactgcag tggcgcaatc tcggctcaact gcaagctccg cctcccaagg 17280
 tcacccatg ctcctgcctc agcctccaga atagctggga ctacaggcgc ccactaccac 17340
 gcccggctaa ttatttgtt ttttttagtag agacggggtt tcaccgtgtt agccaggatg 17400
 gtctcgatct ctcgtacccctcg tgatccgccc gcctcggcct cccaaagtgc tgggattaca 17460
 ggcgtgagcc acccgccctg gccgagcttc attttaaaaaa aatctcattt acctaaataa 17520
 gttgtcaca aacttgtcac actttcatgt gacactttat tcacatattt agaactgtt 17580
 ttcccaggc tacggtagat ttccctacttag tttttttttt atagaacttggg 17640
 gtccaaacaga attatagctt ttttcttctt tgcaatgttat catgagaattt tccccatgtc 17700
 tttcaaaaattt cttcccaagt catttgcaga gtaactccctc ttatgaatattt acatattttt 17760
 ttacagaata agtcctcata tgaatataca tattttttt ctaatgtctt tgggctttt 17820
 ctacatttgcataaaacatcatttac aattttttttt tttttagagaca gtctcgctct 17880
 gtctccaggc tgggggtgcag tggcgccatc tcggctcaact gcaagctcca cctcctgggt 17940
 tcaagcaattt ctcctgcctc agcctccatg gtagctggga ctacaggcgc ggcaccac 18000
 gcccggctaa ttatgttattttttttttagtagag acggggttt accatgttgc ccaggatgg 18060
 ctcaatctctt tgacctctcg atccgcctgc ctcagcctcc ctaatgtctg ggattacagg 18120
 tgggagccac catccccccatc ctcatttaca tttaacaca attaaatttca ggggtttgtc 18180
 aggggtgactt cctagatttcc tcagttttt gaggatgtt tagggatgtt ttttattttt 18240
 tccagtttat gggatattcc cttatataca cacacacaca cacacacaca cacacacaca 18300
 cacccaaac tcaatagggc aagaaccata ttcatcatct cttaaatccctt aggtcctgtc 18360
 acaggccact gttggaaaaa tagttgacta aatgttagtcc attttttccat tttttttttt 18420
 tcactattcc ccaaatccctg gaagagtccc tcataatctgc tgctggggctt ccaaggccctc 18480
 caagctgaag cccagtagct tggtgtcat aagatcatag gcccggaaat catatttaggc 18540
 tggctttcag gttgcaactc taccaggatcc agctgtgtga ctctggatata gtctctcaac 18600
 ctctctgagc ctcgtttca tctgttgaa aatatacgat atcacaacac ttatctgcag 18660
 ggtgattccg aggataaaaaaa agacaaaatat attaccatgt cagcttagaa taggaaggttc 18720
 tcagtaaatg gttaaaaaaag aggtctcttcc cctctgtctc atcttttacc cagcttcaca 18780
 taattggcag cccacccacc tggggcccttcaacccttca accatcccaggccctc 18840
 cttacactcc tgcattatcat gtttccaaacc ccactgttcc actcagctt gttccctcaa 18900
 caccactgtt ccctgtgaca ggtcatccctc ctaacctctg gcatctggtc ccgttgcctc 18960
 cctgacaacc cccatgtctt ctcctctgca acccccgcccg cttgcctgtt ctcctactt 19020
 tattccattt agcccccccttcc acctggggctt gagaacacactt ggccttgacc ctgcaccctc 19080
 agccaggcccccc acagagctca gtcggccgg aagtaaggac cccagatattt cagttgagga 19140

atttaacatt aatactataa tatagaatcc accagcagat tttcnnnnnn nnnnnnnnnn 19200
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnatggc ctgaaagtaa 19260
 gaccccccaga tatctcaggc acgaactcca acagnattac tacattcgag tatccacaag 19320
 gctgacngtc attaacgtt cgaagtgtt tattgtcgcc tccttaatg gatgagatta 19380
 tgagcagca gagacctcat ccgggtgtca cgtgattcgt gtcattgtat atggagcag 19440
 ctctcatcct gattatcggt ttatcgacat ggaaagttgg agtgaatgtc ggccagtagt 19500
 agactaatat gtcccagaaac caggatactc tagctattct cgatggta gattgagatt 19560
 acacactttg gtcggattac ttccccgcgc ttcaaagtga gactcacctg atgtggcac 19620
 gtcgcagtt cagtgtatgcc gagtctgtt gtcggtaaa gagagtgtcc ctgatgtgc 19680
 gcctgtataa atagagttt gtaaagccac cttgacgcatttggcataa ctacgtctc 19740
 gcggagatac tttgagatag gcagcaaccc ttcatgcatttggcataa cctgaggatc 19800
 cttgcctgta tctgttttta catcagtat tacaaggcgg tgatgtctg tcattccctc 19860
 tacatgtatt tgctggattt cttctgaaaa ttacctggat gatttcttca gatttaaatt 19920
 cactgtctt tccttgggaa agcctggact ggctgagctg cctaaccctg actctctgtc 19980
 tttttttttt ttttgacagg gagtctagct ctgttgcag gctggagtgc agtggcaacc 20040
 ttcacccctt ggggtcaagc gatttcctg cctcagccctc ccaagtagct gggattacag 20100
 gcgtgtgcca ccatgccccgg ttaattttt ttttttaata tctattttag ttgagacagg 20160
 gtttacccat gttggtcaga atggctcga tctcctgacc tcgtggctg cccgccccctg 20220
 cctcttttca taccctctgg tagccccgtc cacagcccat ggcaagtatc acaattgca 20280
 ttactctatt gtgttaattat taaaatcag cctctcagtg cccaccatgtc gcaccgtaaa 20340
 gatggacagt ttctatcata ctcacagccca caaccccaac acctggctct gtgtctggt 20400
 catagaaatt gctcaagaaaa gaattttgg agggggcggg tggctcacgc ctgtaatcc 20460
 agcactttgg gaagcggagg caggtggatc atgaggtcag gagttcgaaa ccaacctggc 20520
 caatatggt aaaccccatc tctactaaaa atacaaaaat tcgctggca tggggcatg 20580
 tgtgcacttgc tcgtccccagc tgctcaagag gctgaggcag gagaatcact tgaacctggg 20640
 aggccaggg tgcagtgcgc cgagatcgcc ccactgtact ccagccctggg tgacagagcg 20700
 agactccagc tcaaaaaaaaaaaaaaaa aaaaagaaaa gaaataatttca ttgaatcaat 20760
 ggaaacagaa ttacaactct tcccactttt gggaggaga attgagattc ctgccttccc 20820
 cttcaaccctt agctttatgt ttgttgtat ggtatagagg tcacacatgg cagccttg 20880
 tctgttctta gctgagttc atctacttag ggacggggaa tcccagatttgc ggcaaaacca 20940
 aaagcagggc cagaccttgg aaaaattgtt aagcagtttgc tggttgtgc caccacccat 21000
 gacaaaaatgc acacacacac aacgcaacac acacacagac attttcatgc cttcaaggcc 21060
 taccctacag ctttcccat ggtccaaac ctttcttcc accttacatgt cacttaactg 21120
 tctctgtctgg tctcttccca ctcattttagg caacaaatatttgc ttattgaaaaa acttctctga 21180
 gccagatggc gccaaacagt tacctaaagg gcagactgtc tggtgtccag ggctgtttcc 21240
 cgccaaacact cccctgcgggaa ttgaaataat ttcatcctg gaaattcaaa aaaatcatgt 21300
 aaatttgtcac tggggggaaa ttccaaatgt agctgaaccc tctaagaagc agagagttct 21360
 tcttatggtt tagtattttt taaaattatgt aatcatttca tggttttca acactcgagg 21420
 ttctaaaatg tctacttag gtcctaaatgtc taggttttgc tgtaggtgc tcagtaggaa 21480
 gcaggtggag tcgctgcctc acagaagccca gtgtggggaa ggaacagaca ggcactacaa 21540
 cagctgcagt gtgacttagt ccattgtggg cagccgctgg gatggggatg cagtgaggc 21600
 acctggccag tctgagttggc tctcctctgc tatattttat tggtgtcatc atggcaaatg 21660
 ttggggact ggcattctgtc ggaacaccat ctgcccaggat ttcacagca gggggactt 21720
 gagggatagg aaagtgcacc ttggaaaagg ttgtatgata gagtgtttgc ggggggggg 21780
 tggggattct tgtaaaaatgt gggggcagg ttcttctatg gccgtggatt ttactctg 21840
 ggcttaggaa gccattgaat tttagaaatt tcttctccct gctccagggt ctcttgggc 21900
 ctccctacatc tcccccttagg taccccttag gtccttgcatttgc tctggaaatcat 21960
 tctctccat tccccacccat cacatgcctc tggagaaggctc tgatctccc tcactatgac 22020
 aggaagccag gctctctcat ggtgggtggg ggaatgcact atgtcaatg cattggacca 22080
 gcctggccaa catggtaaaa cccctctgt actaaaaata caaaaatttgc cctggcacag 22140
 tggcagggtgc ctgtcatccc agctacttgg gagactgagg caggagaatc gcttgaaccc 22200
 aggaggtggc tggtgcatttgc agccaaatgc ggcggactgc actccagccctt ggtgacaga 22260
 gtaagaccct gtctaaaaca aaaagaaaaa agcatcgagg gcagcaaaaga gctgggttgc 22320
 gggactgact tggggatttgc tggcttctgc aggattgtat ggtgcctcc taccttgc 22380
 agagaaacag gtcttggcataa tggggtaggg ttgggtact aggtttggg gcccctgggag 22440
 aggacaggag gggcatgagc tggccggcc tggggatgtc tgctcaccat cctactctct 22500
 ccccaactcc caaaaatggcg atgacaccgg cacctggccca aacaaccatgttgcacacaga 22560
 gatgtcgcaa gccatgtatct tggcttgcgc cagtggtat ggtgtatgg gttgtatgg 22620
 aagagtggcataa gacctggggat tctggggatc atctcacatc caccatgccc acggactgg 22680
 tgtcaaacctt gtgttagtttgc tccagatctt tcggcaggatc tggaaaggacc cacaatttca 22740
 aacataaaatgc ctggaaatgtt ggtcaaggaa cagcagaggatc gttggggatgc ctgggatg 22800

tggaggagag cgaacaaga aaagactcat ggagggataa gtggttacaa gtctagctct 22860
 ggaatggac agttagctc aggtcctagc tctgctacta tccttggac ttggatcgac 22920
 ttgctcaatc tctctaaacc tcagtttcct cagcctagaa agtggggaca gggacccaa 22980
 agggtgttg tagagattaa atgagatgat accacaaagc atagagtct gggcctgcg 23040
 ctggaaagcc ctaagcaat gaaagtgggt gttagaaatcc ccagccgaga gaacacctcc 23100
 tctttatgcg aggctcctt tttgtgccc ctaactccag agtaggattc tggttttccc 23160
 ataagggtga ctaagggctt ccttgtaat agagattcg ctagtgggtt ttgtccctga 23220
 agagtcacta gggcccagag caggagaaga agaatatcat ttgtccttaa ggcacagcc 23280
 taagctgggc gtgggtggctc acgcctgtaa tcccagact ttggaaagct gaggcaggag 23340
 gatcatgagg tcaggagttc gaaaccggcc tggccaacat agtggaaaccc tgcctctact 23400
 aaaaatacaa aaattagctg ggcattgggg tgcgtgcctg tagtcccagc tacttggag 23460
 gctgaggcag gagaatcgt tgaacctggg aggtggctgt ggtgagcaga aatcgacc 23520
 actgcactcc agcctggca acagtctcg tctgtctcaa aaaaaaaaaa aaaaaaaaaaag 23580
 agcacaggct ttagaggcag gccaatctgg attcaaattcc tggcacctgg caccaccact 23640
 tagcaatgag tccttaggtt agttattaaa cccccctcggt ttcaactact attaatggc 23700
 acagtattaa cggatctcat tgggtgtt ccaggccaa atgcaagtcc ctgcacac 23760
 gattggcaga gagtaaggc acaatatatt attatctttt ctgacccaaa cccgttttt 23820
 actggtaga aagcagaatt agaataatgc ctattaataa agactagttt ctggagcagc 23880
 aaatctggg cttataagga ggctggcaga ggtcaaggc attgccctat atcatgaact 23940
 gatagccctg ccctccagca tgcctggagg actgggggtt agagaaccag gtgaaatcac 24000
 aagaagcact gtcacacac ctggcacat actaaacaaa ctttagggc ctccgtca 24060
 tgaatctct caatagctt gcaaggcagg tctgattttt cacactcgac agatgagaga 24120
 accgagaagg cacttgcctt cgggtcctca gttttgagct gccaagctga ggttctaacc 24180
 tggctcaact ctaagttt tgcctttt actataggga ccatggctgc ctttggatcc 24240
 agcctcggt tctctattac ctgagcagga tgagaggctg tccttccct cgagctctgc 24300
 tttattttt cagatcaggg ttgcagata aaatacagga tgcccaagttt cagttgaatg 24360
 tcagataaac agcaagttt atttcagcat aagtctgtt caaatgttgc atgagacata 24420
 atgtactaaa aaaaaaaaaatg ctcattgttt acctgaaattt caaatttaac tgggagccct 24480
 gtacttttat ttactaaatc tggcaactctt accccagatg tctgtgaggt taagtgggc 24540
 caggcctaca gctgacggaa ggcacagagag agaggtggca gggactgctg gcctcctgag 24600
 gcagagctgt cccaggtctg gtggggctat gattccagag aggccagaga ctgagtatga 24660
 aagtggcagg cggctgggtc tgaggagatg ccaagttggc ctcttgaggg aaataaacag 24720
 gtatattttt ctttgtgtt ctttgcctt tgaggccagg aagcagctt ttagcctaaa 24780
 tccagatgtt aaaaacagaa atgaaatcag tattttatggc cccaaaccctt ccaagcaagt 24840
 catgcagctc gttcccctgt caaggcctcc caccttggc ccacacaggg cctgaccctc 24900
 gtcttaagcct ggccttggg gaacggaccc tgggggttgg ggtgggtgggt caggccctgc 24960
 tctcaatttc acaccccgctt cctagccctg aaaccagaga ggttccctgaa gtccagccca 25020
 gccaggcctg tggggctgtt gagaggggtt aagtaagaggag gagagtaaaa accgacactt 25080
 gggaggaggc ttttaaaaata aacatctgctt gggagggatg ctctcaagga ggctgcgtt 25140
 tgcagctcag ctagctgtt tccagctaat cagccttctt gggctactct ggagtgtctg 25200
 ctttgcctt gggatgggtt atgttaagag gacagaggga tgcaggaggag ctgggggctg 25260
 agttttccctg agtagaggtt ggcacaggag agaaggcatc acccccacctt cgtccagggcc 25320
 agaagatgtc cagttctgtt aggctctctt aagtctgtt ctcctggac caagagaaaa 25380
 tcccgttctt tgacccaagt gggcttggg gggagtggtt tgcaggaggag gggggactgg 25440
 agaactgact ctacagaaaatc tcaaagctgg caatccaact tctttccctc agattttatag 25500
 atggggaaaac aggcttgagc cacgcagaga ctgcaccaag ctcacacagt tcttttagtgg 25560
 cagagcagag caaatattttt ctttctttt cattctggat ttccatatct tcttcccttc 25620
 ctgggtccagc ccaacccctt ggcacccccc acgaggtggg cgggggatgg ctttgcgtt 25680
 tgcaccactt cgggggtgtt ctcagagatc ttgggtgcac ataagacgtt ggtggccga 25740
 gcttcttagctt actccgtctt gccccttctt tcattctgtc tctgcctccc ttctccctgc 25800
 ttctccgtgt tccctcttcat tctttctgtt tgcaggaggag actacaccccc ccaccccgct 25860
 ctttctctgg cgcctctgtt gtccttccctt tagccccctaa atcactctgg aatcctgtct 25920
 ctttgaagcc agatctggc ccccttcccc tacccttcc attcccaggc tggggaaaggc 25980
 tgaagaggctt gaccgctgtt cgggggggggg cgggtcggtt ggcgtctgg cttccctttt 26040
 gagtttaatta gggaaaacag agaaatgtca gcgaaatgaa agggctgggg gtggggggca 26100
 gctgggttag gagaggaggaa agtgggcagc tgctccctcc cactcaaccc ctctccccc 26160
 ccccccagaaaa gctctcagttt cggggattt ggcacatgaa atgggggctg tagaaacactg 26220
 agcgctgggtt cgtgaagaga aaaaccgggg cgcattccgg ctctccctc cctgtgtcccc 26280
 tcctctctt atgcccggctt gaaaatattt cctgtctctc tattttctaa tccctgggtt 26340
 atgtccccagg attactcagc ctctccggctt ttagtacttc tcgcgtacccg ctcccagggt 26400
 ccagggttggc ggcagaggagaa ggctgggatc tgccagctgg gcctggctg 26460

gcctggcat ccctgagggtt ttaactgtct cctaaccaca gaggatctca gggcctccag 26520
 cagcgagccc caatgagtca aactctgtt tcctcccttc ccaccccccga cccagtcagg 26580
 gcaggtgagg ggttagaggtg atggcatgct ggatgtgacg gtgttgcata tgattgcac 26640
 aggatgaagc accatctcat ttatgcctca cagcagccct cacagtgcgt acaaatcagc 26700
 tggcaattcc gagaggctgc attctaaaca agttccctggg tgatgctgag ccagggccag 26760
 agtgtggact tctctggcc ccagtttctt cttctgtaca gtgggacgtt ggactggagg 26820
 tgctgtctga tgcacccat ctgtggcact ctgagccctgt gcctaaagtgc tccctgagat 26880
 gtctagtccc cgagatcatc atctccctc cttcatagag ccgagcttctt ctccatcttc 26940
 tctctacctc ccattcactc agaacaggga taaggtccaa gctgctggg tgatggta 27000
 ctgactgtcc cticccagcat gcgcttaggt acacttgggc ctgaatgccc ctctaacaaa 27060
 tgctaccggg tatggccttg gtccttctaa cacttgggtc cctcatctgt aaaatgaggg 27120
 tcataaccaca tagctggctg ggctgtttag tagaggttaa aagtgattgt gcatgtgaag 27180
 catccagcag agtgcctggc acacagttag tgctcaaattg ctgtttitgaa atacaaaaat 27240
 tagctggct agtgggtgtgc acctgttaatc tcagctactc aggaggttga ggcaggagaa 27300
 tcacttgagc ccaggaggca gaggttgcag tgagccggga tcgcgccact gcactccagg 27360
 ctggcgaca gaggcaggact ccacatctaaa aaaaaaaaaaaa aaaaaaatgtt gtttgaat 27420
 ggaggcttgg aagagcactc ttccccccca ccccacacac ttactaacgg aactaccgaa 27480
 tctctactcg tgccaggcct tgctggcaa aggcttaggg catgtaaagc catggttca 27540
 tggctgctgc cttccaggag cccatggctc agtggggag tcagaggctc ctttgagaaa 27600
 gccaggccaa ggaactgaga gcccgtgtca cagccctgtg agaaatacca gcgaatgtgt 27660
 tttgagact tacatctacc aggcactatt ttacgtacct gacatacatt ttcttattta 27720
 gcattcaact aaaccagaag aggtggttac cattatcatc cccatttgc agatgagaaa 27780
 actgaagtgg cataaggagg ttagttagac ttggccagga tcacagttag gtaaggaaagg 27840
 agcttaggaat caaacctaga ctttctact ccagagcctc ttaacctcta ccctcaactc 27900
 ttgtgaagag agactacccct ggtgaccctt atattccct tgctcagcgt ggtcaaggcc 27960
 cagggtctg tataggcatt cattaactgc tggggatata agacagtgc tggagggac 28020
 cctaagtta gctcccaccc tgccttccctc tgcctctgc gaagctgctg atggagctc 28080
 caccctggg ggggggtccgc gcaccatggg attgagcgcc cgctacggc cccagttcac 28140
 cctgcagcac gtgcccact accggccagaa tgcctacatc ccaggcagca atgcccacact 28200
 gaccaacgcgca gctggcaagc gggatggcaa ggccccagca ggtggcaatg gcaacaagaa 28260
 gaagtggc aagaaggaga agaagtaaca tggaggccag gccaagagcc acaggccggc 28320
 ctctcccaa ccagcccgcc ttctcttac ctgcacccag gcctcagagt ttccaggctca 28380
 acccccagaa tactggtagg ggccaaggcc atgcctccct tggaaacag aaacaagtgc 28440
 ccagtcagca ctttccctt ccccccagg gggttgaata tgccaaagca gttccgttgg 28500
 gaacccttcat ccaatcaact gctgtaccca tgggggttagt ggggttactg tagacaccaa 28560
 gaaccatgg ccacaccccg ttttagttaca gctgaactcc tccatcttcc aaatcaatca 28620
 ggcctatcca tcccattgcct ccctccccc cacccttcc tcaacatcc tctttcccg 28680
 gtaaggttgt tgggggttgt aagtaccaag taacctacaa gcctcctagt tctgaaaagt 28740
 tggaaaggcata tcatgaccc tggccctctc ctttgattt caatcttccc ccaaagcatg 28800
 gtttgggtcc agcccttca ctccttccca gagcccaaga tcaatgctca agtttggag 28860
 gacatgatca ccatccccat ggtactgtat cttgctgtat ttagggaggg cattttgtca 28920
 ccaagcctct tcccaacgcct ctggggacca gtctctgtt ttgttttca ttgtttgacg 28980
 ttccactgc atgccttgac tttcccccacc tcctcttccaa acaagagact ccactgcatt 29040
 ttccaagacata gtaggggtg gtaagataag gaagggaaagt gtgtggatgt ggatgggtgg 29100
 ggcattggaca aagcttgaca catcaagtt tcaaggcctt ggaggaggct ctgtatgtcc 29160
 tcagggact gacaacatcc tccagatcc agccataaac caataacttag gctggacct 29220
 tcccaactaca taatagggtc cagcccaggc agccagctt gggctgagct aacaggacca 29280
 atggattaaa ctggcatttc agtccaaggaa agtctcaagc aggttttagga ccaggtcccc 29340
 ttgagaggc ttagggggct ctgtgggtgc tgggtactcc agaggtgcca ctgggtggaaag 29400
 ggtcagcggc gccccagcag gaaggggtggg ccagccaggc cattcttagt ccctgggttg 29460
 gggaggcagg gagctaggc agggacccaa tgaacagaaa gtctcagccc aggtggggc 29520
 ttcttcaaca gggcccttc ctccttccaa ctcctgttcc tcaccccttcc aggtgcgtt 29580
 tctctccgtt gaaaggccact gcccaggatcc ccagtgccct ccctagtgcc catagccctgg 29640
 tttaaagttcc ccagtgccct cttgtgcata gaccccttc tcccaaaaaa ttctgccccct 29700
 gggttccccgg ccattccageg gggctgcccag agaaccctt acctgcccctt acagtagtgt 29760
 agcggcccccctt ccctctttcg gctgggtgttag aatagccagt agttagtgc ggtgtgcctt 29820
 tacgtatgg cgggtggca gcccggccgg ggtcccgcc agccgtctgt cttgtatctg 29880
 cccggccgg cccgtgttgtt gttttgtgtt gtgtccacgc gctaaggcga ccccccctcccc 29940
 cgtactgact tctcctataaa gcgcttctct tcgcataatgc acgtagctcc caccaccccc 30000
 ttctcctgtg tctcacaagca gttttataact ctaatattta tatggctttt tttcttcgac 30060
 aaaaaaaaaaaa taaaacgtttt cttctgaaaaa gctgaacgtt tctgtataag cgatggaaagc 30120

tcctggcatg tgtgcataaa gtgatgagct gaggtgggtg ctggaaaag ggcggaatcg 30180
 ggaggccact ctgtgtcatt gcgcgtctag atgtttccga attgcgtgtg tgtgtgtac 30240
 tgtg 30244

<210> 2
 <211> 2805
 <212> DNA
 <213> Homo sapiens

<400> 2
 atggtcccag aggccctggag gagcggactg gtaagcaccg ggagggtagt gggagtttg 60
 cttctgttg gtgccttggaa caaggcttcc acggtcattc actatgagat cccggaggaa 120
 agagagaagg gtttcgttgtt gggcaacgtg gtcgcgaacc ttgggttggaa tctcggttagc 180
 ctctcagccc gcagggttccg ggtgggttct ggagcttagcc gaagattctt tgaggtgaac 240
 cgggagaccc gagagatgtt tgtgaacgac cgtctggatc gagaggagct gtgtgggaca 300
 ctgcctctt gcactgtAAC tctggagttt gtatggaga acccgcttggaa gctgttcagc 360
 gtggaaagtgg tgatccagga catcaacgac aacaatctt cttcccttcc 360
 aaattggaga ttagcgaggc cgtggctccg gggacgcgtt ttccgctcga gagcgcgcac 480
 gatcccgatg tgggaagcaa ctcttacaa acctatgagc tgagccgaaa tgaataactt 540
 gcgctcgcg tgcagacgcg ggaggacacg accaagtacg cggagcttgtt gttggagcgc 600
 gcccggacc gagaacggga gcctagtctc cagttatgtc tgacggcggtt ggacggaggg 660
 accccagctc tctccggccag cctgccttatt cacatcaagg tgctggacgc gaatgacaat 720
 gcgctgtct tcaaccaggc cttgtaccgg ggcgcgttcc tggaggatgc accctccggc 780
 acgcgcgtgg tacaagtctt tgcaacggat ctggatgaag gccccaaacgg taaaattatt 840
 tactccttcg gcagccacaa ccgcgcggc gtgcggcaac tattcccttcc 840
 accgggatgc tgacaatcaa gggtcggctg gacttcgagg acaccaaact ccatgagatt 960
 tacatccagg ccaaagacaa gggcgcacat cccgaaggag cacattgcaaa agtgttggtg 1020
 gaggttgtgg atgtgaatga caaccccccg gagatcacat tcacctccgt gtacagccca 1080
 gtacccgagg atgccccttctt ggggactgtc atcgctttgc tcagtgtac tgacctggat 1140
 gctggcgaga acgggctgtt gacccgcgaa gttccacccgg gtctccctt cagccttact 1200
 tcttcctca agaattactt cacttggaa accagtgcag acctggatgc ggagactgtg 1260
 ccagaataca acctcagcat caccggccgaa gacgcggaa ccccttccctt ctcagccctt 1320
 acaatagtgc gtgttcaagt gtccgacatc aatgacaacc ctccacaatc ttctcaatct 1380
 tcctacgacg tttacatgtt agaaaaacaac ctcccccggg ctccaaatact aaacctaagt 1440
 gtctgggacc ccgacgcccc gcagaatgtc cggctttttt tcttccttcc 1500
 gctgaaaccg ggctagtggg tcgttatttc acaataaaatc gtgacaatgg catagtgtca 1560
 tccttagtgc ccctagacta tgaggatcgg cggaaatttg aattaacacgc tcataatcagc 1620
 gatggggca ccccggtctt agccaccaac atcagcgtga acatatttg cactgatcgc 1680
 aatgacaatg ccccccaggc cctatatcct cggccagggtt ggagctcggt ggagatgtg 1740
 cctcgaggtt cctcagctgg ccacccatgt tcacgggtgg taggctggaa cgccggatgca 1800
 gggcacaatg cttggctctc ctacagtctc ttggatccc ctaaccagag ccttttgc 1860
 atagggtgc acactggtca aatcgtact gcccgtccag tccaagacac agattcaccc 1920
 aggcagactc tcacggtctt gatcaaagac aatggggagc cttcgcttc caccactgt 1980
 accctcaactg tgcgttacac cgaggactct cctgaagccc gagccgagtt cccctctggc 2040
 tctgcccccc gggagcagaa aaaaaatctc accttttac tacttcttc cctaattctt 2100
 gtttctgtgg gtttctgtgtt cacagtgttc ggagtaatca tattcaaaatg ttacaatgtt 2160
 aagcaatcta gagacctata ccgagccccg gtgagctcac tgcgttccac accagggccc 2220
 tccttgcacg cggacgccccgtt gggggggggc ctgtatgtcgc cgcaccccttta ccatcagggt 2280
 tatctcacca cggactcccccc cccggccgac ccgtgtctga agaaacctgg tgcagccagt 2340
 ccactggcca gcccggcagaa cacgtgtccgg agctgtgttcc cgggttcttca taggcagggt 2400
 ttgggtgcag agagcgcggccccc tccggacag caagccccgc ccaacacggc ctggcggtttc 2460
 tctcaggcccc agagacccgg caccagccgc tcccaaaatg gcgatgacac cggcacctgg 2520
 cccaaacaacc agtttgcacac agagatgtc caagccatgt tcttggcggtc cgccagtgaa 2580
 gctgtgtatg ggagctccac cctgggggggg ggtggccggca ccatgggatt gagcgcggccgc 2640
 tacggacccc agtttgcacac gcagcacgtt cccgactacc gccagaatgtt ctacatccca 2700
 ggcagcaatg ccacactgac caacgcagct ggcaaggcggg atggcaaggc cccagcaggt 2760
 ggcaatggca acaagaagaa gtcggcaag aaggagaaga agtaa 2805

<210> 3

<211> 934
 <212> PRT
 <213> Homo sapiens

<400> 3
 Met Val Pro Glu Ala Trp Arg Ser Gly Leu Val Ser Thr Gly Arg Val
 1 5 10 15
 Val Gly Val Leu Leu Leu Gly Ala Leu Asn Lys Ala Ser Thr Val
 20 25 30
 Ile His Tyr Glu Ile Pro Glu Glu Arg Glu Lys Gly Phe Ala Val Gly
 35 40 45
 Asn Val Val Ala Asn Leu Gly Leu Asp Leu Gly Ser Leu Ser Ala Arg
 50 55 60
 Arg Phe Arg Val Val Ser Gly Ala Ser Arg Arg Phe Phe Glu Val Asn
 65 70 75 80
 Arg Glu Thr Gly Glu Met Phe Val Asn Asp Arg Leu Asp Arg Glu Glu
 85 90 95
 Leu Cys Gly Thr Leu Pro Ser Cys Thr Val Thr Leu Glu Leu Val Val
 100 105 110
 Glu Asn Pro Leu Glu Leu Phe Ser Val Glu Val Val Ile Gln Asp Ile
 115 120 125
 Asn Asp Asn Asn Pro Ala Phe Pro Thr Gln Glu Met Lys Leu Glu Ile
 130 135 140
 Ser Glu Ala Val Ala Pro Gly Thr Arg Phe Pro Leu Glu Ser Ala His
 145 150 155 160
 Asp Pro Asp Val Gly Ser Asn Ser Leu Gln Thr Tyr Glu Leu Ser Arg
 165 170 175
 Asn Glu Tyr Phe Ala Leu Arg Val Gln Thr Arg Glu Asp Ser Thr Lys
 180 185 190
 Tyr Ala Glu Leu Val Leu Glu Arg Ala Leu Asp Arg Glu Arg Glu Pro
 195 200 205
 Ser Leu Gln Leu Val Leu Thr Ala Leu Asp Gly Gly Thr Pro Ala Leu
 210 215 220
 Ser Ala Ser Leu Pro Ile His Ile Lys Val Leu Asp Ala Asn Asp Asn
 225 230 235 240
 Ala Pro Val Phe Asn Gln Ser Leu Tyr Arg Ala Arg Val Leu Glu Asp
 245 250 255
 Ala Pro Ser Gly Thr Arg Val Val Gln Val Leu Ala Thr Asp Leu Asp
 260 265 270
 Glu Gly Pro Asn Gly Glu Ile Ile Tyr Ser Phe Gly Ser His Asn Arg
 275 280 285
 Ala Gly Val Arg Gln Leu Phe Ala Leu Asp Leu Val Thr Gly Met Leu
 290 295 300

Thr Ile Lys Gly Arg Leu Asp Phe Glu Asp Thr Lys Leu His Glu Ile
 305 310 315 320
 Tyr Ile Gln Ala Lys Asp Lys Gly Ala Asn Pro Glu Gly Ala His Cys
 325 330 335
 Lys Val Leu Val Glu Val Val Asp Val Asn Asp Asn Ala Pro Glu Ile
 340 345 350
 Thr Val Thr Ser Val Tyr Ser Pro Val Pro Glu Asp Ala Pro Leu Gly
 355 360 365
 Thr Val Ile Ala Leu Leu Ser Val Thr Asp Leu Asp Ala Gly Glu Asn
 370 375 380
 Gly Leu Val Thr Cys Glu Val Pro Pro Gly Leu Pro Phe Ser Leu Thr
 385 390 395 400
 Ser Ser Leu Lys Asn Tyr Phe Thr Leu Lys Thr Ser Ala Asp Leu Asp
 405 410 415
 Arg Glu Thr Val Pro Glu Tyr Asn Leu Ser Ile Thr Ala Arg Asp Ala
 420 425 430
 Gly Thr Pro Ser Leu Ser Ala Leu Thr Ile Val Arg Val Gln Val Ser
 435 440 445
 Asp Ile Asn Asp Asn Pro Pro Gln Ser Ser Gln Ser Ser Tyr Asp Val
 450 455 460
 Tyr Ile Glu Glu Asn Asn Leu Pro Gly Ala Pro Ile Leu Asn Leu Ser
 465 470 475 480
 Val Trp Asp Pro Asp Ala Pro Gln Asn Ala Arg Leu Ser Phe Phe Leu
 485 490 495
 Leu Glu Gln Gly Ala Glu Thr Gly Leu Val Gly Arg Tyr Phe Thr Ile
 500 505 510
 Asn Arg Asp Asn Gly Ile Val Ser Ser Leu Val Pro Leu Asp Tyr Glu
 515 520 525
 Asp Arg Arg Glu Phe Glu Leu Thr Ala His Ile Ser Asp Gly Gly Thr
 530 535 540
 Pro Val Leu Ala Thr Asn Ile Ser Val Asn Ile Phe Val Thr Asp Arg
 545 550 555 560
 Asn Asp Asn Ala Pro Gln Val Leu Tyr Pro Arg Pro Gly Gly Ser Ser
 565 570 575
 Val Glu Met Leu Pro Arg Gly Thr Ser Ala Gly His Leu Val Ser Arg
 580 585 590
 Val Val Gly Trp Asp Ala Asp Ala Gly His Asn Ala Trp Leu Ser Tyr
 595 600 605
 Ser Leu Leu Gly Ser Pro Asn Gln Ser Leu Phe Ala Ile Gly Leu His
 610 615 620

Thr Gly Gln Ile Ser Thr Ala Arg Pro Val Gln Asp Thr Asp Ser Pro
 625 630 635 640
 Arg Gln Thr Leu Thr Val Leu Ile Lys Asp Asn Gly Glu Pro Ser Leu
 645 650 655
 Ser Thr Thr Ala Thr Leu Thr Val Ser Val Thr Glu Asp Ser Pro Glu
 660 665 670
 Ala Arg Ala Glu Phe Pro Ser Gly Ser Ala Pro Arg Glu Gln Lys Lys
 675 680 685
 Asn Leu Thr Phe Tyr Leu Leu Ser Leu Ile Leu Val Ser Val Gly
 690 695 700
 Phe Val Val Thr Val Phe Gly Val Ile Ile Phe Lys Val Tyr Lys Trp
 705 710 715 720
 Lys Gln Ser Arg Asp Leu Tyr Arg Ala Pro Val Ser Ser Leu Tyr Arg
 725 730 735
 Thr Pro Gly Pro Ser Leu His Ala Asp Ala Val Arg Gly Gly Leu Met
 740 745 750
 Ser Pro His Leu Tyr His Gln Val Tyr Leu Thr Thr Asp Ser Arg Pro
 755 760 765
 Gly Asp Pro Leu Leu Lys Lys Pro Gly Ala Ala Ser Pro Leu Ala Ser
 770 775 780
 Arg Gln Asn Thr Leu Arg Ser Cys Asp Pro Val Phe Tyr Arg Gln Val
 785 790 795 800
 Leu Gly Ala Glu Ser Ala Pro Pro Gly Gln Gln Ala Pro Pro Asn Thr
 805 810 815
 Asp Trp Arg Phe Ser Gln Ala Gln Arg Pro Gly Thr Ser Gly Ser Gln
 820 825 830
 Asn Gly Asp Asp Thr Gly Thr Trp Pro Asn Asn Gln Phe Asp Thr Glu
 835 840 845
 Met Leu Gln Ala Met Ile Leu Ala Ser Ala Ser Glu Ala Ala Asp Gly
 850 855 860
 Ser Ser Thr Leu Gly Gly Ala Gly Thr Met Gly Leu Ser Ala Arg
 865 870 875 880
 Tyr Gly Pro Gln Phe Thr Leu Gln His Val Pro Asp Tyr Arg Gln Asn
 885 890 895
 Val Tyr Ile Pro Gly Ser Asn Ala Thr Leu Thr Asn Ala Ala Gly Lys
 900 905 910
 Arg Asp Gly Lys Ala Pro Ala Gly Gly Asn Gly Asn Lys Lys Lys Ser
 915 920 925
 Gly Lys Lys Glu Lys Lys
 930

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

<400> 4
ttcgatcycc tttct

15

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

<400> 5
cctaaggyag gtttc

15

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

<400> 6
cgatactrgc aaggc

15

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

<400> 7
aaatcaamgg catcc

15

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

<400> 8
ggcgccccrgc gccca

15

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

<400> 9
agtacgcsga gctgg

15

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

<400> 10
caaccagycc ttgtta

15

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

<400> 11
agtccctrta ccggg

15

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

<400> 12
gccacaaycg cgccg

15

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

<400> 13
tggggttygt ggtca

15

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

<400> 14
aactggcyct tccta

15

<210> 15
<211> 15
<212> DNA
<213> Homo sapiens

<400> 15
gagaccrcgc accag

15

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

<400> 16
aggccttsgg catgg

15

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

<400> 17 ccatgccyac ggact	15	
 <td><210> 18 <211> 15 <212> DNA <213> Homo sapiens</td>	<210> 18 <211> 15 <212> DNA <213> Homo sapiens	
 <td><400> 18 gctacggrrcc ccagt</td> <td style="text-align: right;">15</td>	<400> 18 gctacggrrcc ccagt	15
 <td><210> 19 <211> 15 <212> DNA <213> Homo sapiens</td>	<210> 19 <211> 15 <212> DNA <213> Homo sapiens	
 <td><400> 19 acagggcrgc ctctc</td> <td style="text-align: right;">15</td>	<400> 19 acagggcrgc ctctc	15
 <td><210> 20 <211> 15 <212> DNA <213> Homo sapiens</td>	<210> 20 <211> 15 <212> DNA <213> Homo sapiens	
 <td><400> 20 ccccttgrga aacag</td> <td style="text-align: right;">15</td>	<400> 20 ccccttgrga aacag	15
 <td><210> 21 <211> 15 <212> DNA <213> Homo sapiens</td>	<210> 21 <211> 15 <212> DNA <213> Homo sapiens	
 <td><400> 21 ggttgaayat gcaaa</td> <td style="text-align: right;">15</td>	<400> 21 ggttgaayat gcaaa	15
 <td><210> 22 <211> 15 <212> DNA <213> Homo sapiens</td>	<210> 22 <211> 15 <212> DNA <213> Homo sapiens	
 <td><400> 22 aaagttgsaa gggca</td> <td style="text-align: right;">15</td>	<400> 22 aaagttgsaa gggca	15
 <td><210> 23 <211> 15 <212> DNA <213> Homo sapiens</td>	<210> 23 <211> 15 <212> DNA <213> Homo sapiens	
 <td><400> 23 gttgtacrtt tccac</td> <td style="text-align: right;">15</td>	<400> 23 gttgtacrtt tccac	15
 <td><210> 24 <211> 15 <212> DNA</td>	<210> 24 <211> 15 <212> DNA	

<213> Homo sapiens
<400> 24
gggtcagygg agccc 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapiens
<400> 25
aacagaargt ctcag 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapiens
<400> 26
gtccccartg cgccc 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens
<400> 27
agaatagyca gtagt 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens
<400> 28
gtttggttcg atcyc 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapiens
<400> 29
atcacaagaa aggrrg 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapiens
<400> 30
gctttcccta aggya 15

<210> 31

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

<400> 31
gcataggaaa cctrcc 15

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

<400> 32
atgcaccgat actrcc 15

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

<400> 33
caaagcgcct tgcya 15

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

<400> 34
gcgagcaaat caamgc 15

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

<400> 35
ctttctggat gcckt 15

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

<400> 36
ggactcggcg cccrg 15

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

<400> 37
agcgcttggg cgcyg 15

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

<400> 38
gcaccaagta cgcs

15

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

<400> 39
ccaacaccag ctcs

15

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

<400> 40
tgtcttcaac cagyc

15

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

<400> 41
gccccgtaca aggrc

15

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

<400> 42
tcaaccagtc cttrt

15

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

<400> 43
cgcgccggc gtaya

15

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

<400> 44

tcggcagcca caayc	15
<210> 45	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 45	
gcacgccggc gcgrt	15
<210> 46	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 46	
tttctgtggg gtttg	15
<210> 47	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 47	
acactgtgac cacra	15
<210> 48	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 48	
actttcaact ggcyc	15
<210> 49	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 49	
ttgatctagg aagrg	15
<210> 50	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 50	
ggcccgaga cccrg	15
<210> 51	
<211> 15	
<212> DNA	
<213> Homo sapiens	

<400> 51	
tacccgctgg tgcyg	15
<210> 52	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 52	
agaaaacaggt cttsq	15
<210> 53	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 53	
ccttaccccat gccsa	15
<210> 54	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 54	
cagccaccat gccya	15
<210> 55	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 55	
acatccagtc cgtrg	15
<210> 56	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 56	
gcgcggccta cggrc	15
<210> 57	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 57	
gggtgaactg gggyc	15
<210> 58	
<211> 15	

<212> DNA
<213> Homo sapiens

<400> 58
agagccacag ggcrg

15

<210> 59
<211> 15
<212> DNA
<213> Homo sapiens

<400> 59
gttggggaga ggcyg

15

<210> 60
<211> 15
<212> DNA
<213> Homo sapiens

<400> 60
catgctcccc ttgrg

15

<210> 61
<211> 15
<212> DNA
<213> Homo sapiens

<400> 61
ttgttttgt ttcyc

15

<210> 62
<211> 15
<212> DNA
<213> Homo sapiens

<400> 62
ccaggggtt gaaya

15

<210> 63
<211> 15
<212> DNA
<213> Homo sapiens

<400> 63
actgcttttg catrt

15

<210> 64
<211> 15
<212> DNA
<213> Homo sapiens

<400> 64
ttctgaaaag ttgsa

15

<210> 65
<211> 15
<212> DNA
<213> Homo sapiens

<400> 65
tcatgatgcc cttsc

15

<210> 66
<211> 15
<212> DNA
<213> Homo sapiens

<400> 66
ttcattgttt gacrt

15

<210> 67
<211> 15
<212> DNA
<213> Homo sapiens

<400> 67
catgcagtgg aaayg

15

<210> 68
<211> 15
<212> DNA
<213> Homo sapiens

<400> 68
gtggaagggt cagyg

15

<210> 69
<211> 15
<212> DNA
<213> Homo sapiens

<400> 69
ctgctggggc tccrc

15

<210> 70
<211> 15
<212> DNA
<213> Homo sapiens

<400> 70
caaatgaaca gaarg

15

<210> 71
<211> 15
<212> DNA
<213> Homo sapiens

<400> 71
cctgggctga gacyt

15

<210> 72
<211> 15
<212> DNA
<213> Homo sapiens

<400> 72
gcccaggtcc ccatt

15

<210> 73
<211> 15
<212> DNA
<213> Homo sapiens

<400> 73
actagggggc gcayt

15

<210> 74
<211> 15
<212> DNA
<213> Homo sapiens

<400> 74
tggttagaa tagyc

15

<210> 75
<211> 15
<212> DNA
<213> Homo sapiens

<400> 75
cactacacta ctgrc

15

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

<400> 76
tggtcgatc

10

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

<400> 77
acaagaaaagg

10

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

<400> 78
ttcccttaagg 10

<210> 79
<211> 10
<212> DNA
<213> Homo sapiens

<400> 79
taggaaacct 10

<210> 80
<211> 10
<212> DNA
<213> Homo sapiens

<400> 80
caccgatact 10

<210> 81
<211> 10
<212> DNA
<213> Homo sapiens

<400> 81
agcgcccttgc 10

<210> 82
<211> 10
<212> DNA
<213> Homo sapiens

<400> 82
agcaaatcaa 10

<210> 83
<211> 10
<212> DNA
<213> Homo sapiens

<400> 83
tctggatgcc 10

<210> 84
<211> 10
<212> DNA
<213> Homo sapiens

<400> 84
ctcgccgccc 10

<210> 85
<211> 10
<212> DNA

<213> Homo sapiens

<400> 85
gcttgggcgc

10

<210> 86
<211> 10
<212> DNA
<213> Homo sapiens

<400> 86
ccaagtacgc

10

<210> 87
<211> 10
<212> DNA
<213> Homo sapiens

<400> 87
acaccagctc

10

<210> 88
<211> 10
<212> DNA
<213> Homo sapiens

<400> 88
cttcaaccag

10

<210> 89
<211> 10
<212> DNA
<213> Homo sapiens

<400> 89
cggtacaagg

10

<210> 90
<211> 10
<212> DNA
<213> Homo sapiens

<400> 90
accagtcctt

10

<210> 91
<211> 10
<212> DNA
<213> Homo sapiens

<400> 91
gcgccccggta

10

<210> 92

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

<400> 92
gcagccacaa

10

<210> 93
<211> 10
<212> DNA
<213> Homo sapiens

<400> 93
cgccggcgcg

10

<210> 94
<211> 10
<212> DNA
<213> Homo sapiens

<400> 94
ctgtggggtt

10

<210> 95
<211> 10
<212> DNA
<213> Homo sapiens

<400> 95
ctgtgaccac

10

<210> 96
<211> 10
<212> DNA
<213> Homo sapiens

<400> 96
ttcaactggc

10

<210> 97
<211> 10
<212> DNA
<213> Homo sapiens

<400> 97
atcttaggaag

10

<210> 98
<211> 10
<212> DNA
<213> Homo sapiens

<400> 98
ccagagaccc

10

<210> 99
<211> 10
<212> DNA
<213> Homo sapiens

<400> 99
ccgctgggtgc

10

<210> 100
<211> 10
<212> DNA
<213> Homo sapiens

<400> 100
aacaggtctt

10

<210> 101
<211> 10
<212> DNA
<213> Homo sapiens

<400> 101
accccatgcc

10

<210> 102
<211> 10
<212> DNA
<213> Homo sapiens

<400> 102
ccaccatgcc

10

<210> 103
<211> 10
<212> DNA
<213> Homo sapiens

<400> 103
tccagtccgt

10

<210> 104
<211> 10
<212> DNA
<213> Homo sapiens

<400> 104
cccgctacgg

10

<210> 105
<211> 10
<212> DNA
<213> Homo sapiens

<400> 105

tgaactgggg 10

<210> 106
<211> 10
<212> DNA
<213> Homo sapiens

<400> 106
gccacagggc 10

<210> 107
<211> 10
<212> DNA
<213> Homo sapiens

<400> 107
ggggagaggc 10

<210> 108
<211> 10
<212> DNA
<213> Homo sapiens

<400> 108
gctcccttg 10

<210> 109
<211> 10
<212> DNA
<213> Homo sapiens

<400> 109
tttctgttgc 10

<210> 110
<211> 10
<212> DNA
<213> Homo sapiens

<400> 110
gggggttgaa 10

<210> 111
<211> 10
<212> DNA
<213> Homo sapiens

<400> 111
gcttttgcatt 10

<210> 112
<211> 10
<212> DNA
<213> Homo sapiens

<400> 112 tgaaaagttg	10
<210> 113 <211> 10 <212> DNA <213> Homo sapiens	
<400> 113 tgatgccctt	10
<210> 114 <211> 10 <212> DNA <213> Homo sapiens	
<400> 114 attgtttgac	10
<210> 115 <211> 10 <212> DNA <213> Homo sapiens	
<400> 115 gcagtggaaa	10
<210> 116 <211> 10 <212> DNA <213> Homo sapiens	
<400> 116 gaagggtcag	10
<210> 117 <211> 10 <212> DNA <213> Homo sapiens	
<400> 117 ctggggctcc	10
<210> 118 <211> 10 <212> DNA <213> Homo sapiens	
<400> 118 atgaacagaa	10
<210> 119 <211> 10	

<212> DNA
<213> Homo sapiens

<400> 119
gggctgagac

10

<210> 120
<211> 10
<212> DNA
<213> Homo sapiens

<400> 120
caggtccccca

10

<210> 121
<211> 10
<212> DNA
<213> Homo sapiens

<400> 121
agggggcgca

10

<210> 122
<211> 10
<212> DNA
<213> Homo sapiens

<400> 122
tgtagaatag

10

<210> 123
<211> 10
<212> DNA
<213> Homo sapiens

<400> 123
tacactactg

10

<210> 124
<211> 30244
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (2027)
<223> PS1: Polymorphic base C or T

<220>
<221> allele
<222> (2137)
<223> PS2: Polymorphic base T or C

<220>
<221> allele
<222> (2160)

<223> PS3: Polymorphic base G or A
<220>
<221> allele
<222> (2393)
<223> PS4: Polymorphic base C or A
<220>
<221> allele
<222> (2427)
<223> PS5: Polymorphic base A or G
<220>
<221> allele
<222> (3096)
<223> PS6: Polymorphic base G or C
<220>
<221> allele
<222> (3253)
<223> PS7: Polymorphic base T or C
<220>
<221> allele
<222> (3258)
<223> PS8: Polymorphic base G or A
<220>
<221> allele
<222> (3375)
<223> PS9: Polymorphic base C or T
<220>
<221> allele
<222> (4629)
<223> PS10: Polymorphic base T or C
<220>
<221> allele
<222> (5076)
<223> PS11: Polymorphic base C or T
<220>
<221> unsure
<222> (7653)..(7712)
<223> Nucleotide identity unknown
<220>
<221> allele
<222> (11834)
<223> PS12: Polymorphic base G or A
<220>
<221> unsure
<222> (19185)..(19244)
<223> Nucleotide identity unknown
<220>
<221> unsure
<222> (19295)
<223> Nucleotide identity unknown

<220>
<221> unsure
<222> (19327)
<223> Nucleotide identity unknown

<220>
<221> allele
<222> (22396)
<223> PS13: Polymorphic base G or C

<220>
<221> allele
<222> (22670)
<223> PS14: Polymorphic base C or T

<220>
<221> allele
<222> (28129)
<223> PS15: Polymorphic base A or G

<220>
<221> allele
<222> (28318)
<223> PS16: Polymorphic base G or A

<220>
<221> allele
<222> (28423)
<223> PS17: Polymorphic base G or A

<220>
<221> allele
<222> (28479)
<223> PS18: Polymorphic base T or C

<220>
<221> allele
<222> (28743)
<223> PS19: Polymorphic base G or C

<220>
<221> allele
<222> (28980)
<223> PS20: Polymorphic base G or A

<220>
<221> allele
<222> (29407)
<223> PS21: Polymorphic base C or T

<220>
<221> allele
<222> (29500)
<223> PS22: Polymorphic base A or G

<220>
<221> allele
<222> (29614)
<223> PS23: Polymorphic base G or A

<220>
 <221> allele
 <222> (29796)
 <223> PS24: Polymorphic base C or T

<400> 124
 cacatgggcc cgtagggaga gctttagaac tagtgcagct taattgagca gggggaaagcc 60
 ttctgcagac agttgagct agcggagggt ggttgacacg gaaggggctt gatttttggg 120
 aaggtgagac ttaaatttgg gattgccaag gattggtatg ttaatccaag ggccgtttat 180
 gcgaattaag gaccagattt gaaaattggg gaacccaagt taaaatggtt tgggaacgtg 240
 taaccttcgg ttattgttgg gggaatggct aacccaaagg ggtcttattt gttcatgaag 300
 aggtctaag aaaactatga gctgggtgca gtagctcagg cctgtatgtcc catctactta 360
 gaagtctgag gcggggaggt cccttgagcc caaggagctg gaagttacag tgagctatga 420
 ttattccact gtactccagc ctggatgaca gagcgagatc ctgcctcaga aacaaaaacc 480
 aaaaaataaaa caacaacaaa aacaacaaca acgaaaatct cagaaaaacta caggccttat 540
 tatcaagaaaa atgacaggcc ttattattga gaaaactact tacatttttta aaagctgatt 600
 ttcaaaaacc agattgtcaa ttaagtgtatg ctgaataggg gaaaaaaaaaga ctgataaacc 660
 agtgtatctc tgaataaacc actagatcag tgcttgtaa actttaatgt gtttcaaat 720
 caccctgatg ttttgttaaa atgtaagtca tgtctcagta ggtctagagg agggcatttc 780
 tgcatttcta acaagctccc cggtgatgcc tagaccattt gtcctcacac tttgagtagc 840
 aagtgattgg ggtatactct ttattctaaa ttccaaatgt atttcatctt ctaatgccta 900
 tcagccctct tgctggctg gctaatttac aaacttgagc atttaaaact gggggtttc 960
 agagattaaat tgaagtgtta aagacatgga aatataggc acatatggct tttcatctct 1020
 tccggaaata tcctatataat ccttacttct tttctaccta cctaaccat tgcctaccc 1080
 cttcacttat gttcaaggtt ctattatact tcaaaaggac tctatgcaat caaagtgtatg 1140
 tccaaagacaa atatacagaa atagatcag gctctaaccat tttacatcta atggagtggt 1200
 ttgcattcatt aatttggatct tactctgtgc caggcactt acctgccttc ttcattaat 1260
 aggataggag caggaactct tagaggttca tttcttggc aaggatatac atctcatctc 1320
 tggcttatat cacagtggac tcttttattt actatcctaa tcattataat tttgttgg 1380
 gacaacaatt ctaaatgcag aaaagttaaat gctcttttattt ctgtgactaa aggcttggta 1440
 ttttctaaag atttgtggag ctttgctgcc attatttcaaa taattttata aagatccctag 1500
 gtggtaaataa ggcctggctt ttttctttt taaaatcttc tccgtcttct tcttcattag 1560
 cagtgcaact tcttggca tatggtaccc aagatcttgg gaaaaggatt gatcaagtaa 1620
 agggaaacga tagagagagt gtgcctaca actgagatca tgctccctgt gaaagacagg 1680
 gtagggaaat ttcttatttgc tccctggc ttaagaaaatg agagctaaaca atgacatcca 1740
 gttttaaaag aaataaaaatc tgaattgtac tttaaattt tataagccag agttaaacca 1800
 atagattgtt actcctgtca atttggatgt agctccagat aatggaggaag tgacaggaa 1860
 tgctaatttc atttggtaag agataaaaact gaaatttata atgtctccac gaaaatgcta 1920
 aatgcctcct tacaagtagg gtccggctaa ttgtctgtct tccaaacaacg cagattttgtt 1980
 ggttttttc cagacaatgt tttgtatgtt tcgggtttgg tgcgtccct ttcttggat 2040
 caaagaaagt gattcaaatg tttaatggatg ttgatgggg attggagact tgcagaacgg 2100
 ctagcctcac agccccaaagg ctggctttcc ctaaggyagg tttctatgc accgatactr 2160
 gcaaggcgct ttggctggaa actctggaaag gaagccaaag gaaagtgaag ttccctggcgc 2220
 tagcgcgtgt ctcgtcaga gagccggcg ctgttcattt cttcgatgcag ttatggctg 2280
 ggactctgtg tgccgctgtc ggccaaatgaa gacgctggag atcggggcccc tgcccgtccc 2340
 ctttctgcgc cccggatga ggcagagact gaacagccgg cgagcaaaatc aamggcatcc 2400
 agaaagccat gtcggactcg ggcggccgc cccaaaggctt aaccggctga aagtttctca 2460
 gcgaaatctc agggacgatc tggacccgc tgagaggaac tgcttttggat tgagatgtc 2520
 ccagaggcct ggaggagcgg actggtaagc accggggagg tagtgggagt ttgcgttctg 2580
 cttggctgtc tgaacaaggc ttccacggc attcaatgt agatcccgaa gaaaagagag 2640
 aagggtttcg ctgtggccaa cgtgtcgcc aaccttgggtt tggatctcg tagcctctca 2700
 gcccgcaggc tccgggtgtt gtctggagct agccgaagat tctttggatgtt gaaacgggag 2760
 accggagaga tttttgtgaa cgaccgtctg gatcgagagg agctgtgtgg tttgcgttctg 2820
 tcttgcactg taactctgtt gttggatgtt gagaacccgc tggagctgtt cagcgtgaa 2880
 gtggatgtcc aggacatcaa cgacaacaaat cctgtttcc ctacccaggaa aatgaaattt 2940
 gagatagcg aggccgtgca tccggggacg cgcttccgc tcgagagcgc gcacgatccc 3000
 gatgtggaa gcaactctt acaaaccat gatcgagcc gaaatgaata ctttgcgttctg 3060
 cgcgtgcaga cgcggggagga cagcaccaag tacgcsgagc tggatggatgaa ggcgcctcg 3120
 gaccgagaac gggagccctag tctcagttt gatcgacgg cgatggacgg agggacccca 3180
 gctctctccg ccagcctgcc tattcacatc aaggtgctgg acgcaatga caatgcgcct 3240
 gtcttcaacc agyccttrta ccggcgcgc gtcctggagg atgcaccctc cggcacgcgc 3300

gtggtacaag tccttgcaac ggatctggat gaaggcccca acggtaaaat tatttactcc 3360
 ttcgccagcc acaaycgcgc cggcgtcgaa caactattcg ccttagacct tgtaaccggg 3420
 atgctgacaa tcaagggtcg gctggacttc gaggacacca aactccatga gatttacatc 3480
 caggccaaag acaaggcgcc caatccccaa ggagcacatt gcaaagtgtt ggtggaggtt 3540
 gtggatgtga atgacaacgc cccggagatc acagtcaacct ccgtgtacag cccagtagccc 3600
 gaggatgccc ctctggggac tgtcatcgct ttgctcagtg tgactgaccc ggtatgtggc 3660
 gagaacgggc tggtgacccg cgaagttcca cggggctcc ctttcagccct tacttcttcc 3720
 ctcaagaatt acttcacttt gaaaaccagt gcagacctgg atcgggagac tggccagaa 3780
 tacaaccta gcaccccgcc cggagacgcc ggaacccctt ccctctcagc ccttacaata 3840
 gtgcgtgttc aagtgtccga catcaatgac aaccctccac aatcttctca atcttcttac 3900
 gacgtttaca ttgaagaaaa caacccccc ggggctccaa tactaaacct aagtgtctgg 3960
 gaccccgacg ccccgccagaa tgctcggtt tctttcttc tcttggagca aggagctgaa 4020
 accgggctag tgggtcgcta tttcacaata aatcggtgaca atggcatagt gtcatcctta 4080
 gtgcgccttag actatgagga tcggcggaa tttgaattaa cagctcatat cagcgtatggg 4140
 ggcaccccg tccatggcac caacatcagc gtgaacatat ttgtcaactga tcgcaatgac 4200
 aatgcgcgcgc aggtcctata tcctcggtt ggtggagct cggtggagat gctgcctcg 4260
 ggtacccctcg ctggccaccc agtgcacccg gtggtaggct gggacgcggg tgcaaggccac 4320
 aatgcctggc tctcttacag tctttggaa tcccttaacc agggcttt tgccataggg 4380
 ctgcacactg gtcaaatcag tactgcccgt ccagtccaaag acacagattc acccaggccag 4440
 actctcacccg tcttgatcaa agacaatggg gagccttcgc tctccaccac tgctaccctc 4500
 actgtgtcgag taaccggagga ctctctgaa gcccggccg agttccctc tggctctgccc 4560
 ccccgccggc agaaaaaaaaa ttcacacccctt tatctacttc tttccctaat cctggtttct 4620
 gtgggggttyg tggtcacagt gttcgagta atcatattca aagtttacaa gtggaaaggag 4680
 tctagagacc tataaccggc cccggtgagc tcactgttacc gaacaccagg gccctccctg 4740
 cacgcggacg ccgtgcgggg aggccctgatg tcgcccacc tttaccatca ggtgtatctc 4800
 accacggact cccggcccccgg cgaccggctg ctgaagaaac ctggcgcagc cagtccactg 4860
 gccagccgccc agaacacgcgt gggagctgt gatccgggt tctataggca ggtgttggt 4920
 gcagagagcg cccctccccc acaggtaagg tttagcaagt catgcttgc cctgttagtg 4980
 cttttttatt ctcacatcat attgaggaag gaatggagct gtttttttag ttagtgaagat 5040
 gttttctgg ttagtcattt acacccctaa ctggcycttc cttagatccaa gtttagtgcct 5100
 ttgtgagatg gtggcctgcc agagtgttgt ttgtgttccc atttcagggg gaagataactt 5160
 gactcatctg tggacctaatt tcacatccctc agcactctt tgctatcaca actaaccat 5220
 cttgctaagg gatggtaaag ctaaaacaca agatctcagc gatcagagtt tagcttgta 5280
 tcatttacat taggaaataag ctgctggata cctcttaaccctt gttggcagctt ctaggaataac 5340
 aaaaactacc tcattccctcc accttcaag tgattgtgac atttgttatta aaactaata 5400
 ctttttgcata attttccctt gtttatacag atcgtgttacc tcattctcag ataattttt 5460
 atgaatgaaa tggaaatccca ggcacccctt aaattttatt tcaagcattt tactggaaat 5520
 gatgtgcacc ctgcttacaa aatatttttta ctttataaca gtaatgcatg tttgcttagaa 5580
 aattcagaaaa atacagaaaaa gtatttaaaa aattttttactt ataatctccaa aacccaaaga 5640
 taaccatttt taatttataaa agtaataattt tatttcaaaaa ataaatcttag gcaacccaaag 5700
 aaaacataaaa gtagccagac tcagttgtgt gcacccctgtt ttcttgcgtc tcagtgctg 5760
 aggtgggagg attgcttgcg cccaggatgtt ctgggctgtg gtgcactatg ctgtcaatt 5820
 gatcaatccca ctgtctgcac taaattcagc atgaatacag tgaccccttgg ggagggcaag 5880
 accatcaggt tgcctaaaga ggggtgaact gcccacgtt ggaagtggag aaggtaaaaa 5940
 ctgctgtgtt atccagtagt gggatgacat ctgtgaatag ccactgcact ccagccctagg 6000
 caatataggg aaaccatgtc tcttttacaa taacaacaac aacaacaaca atcccagaaaa 6060
 ctacaaaagg agagtctttt tggtgcctcc agtgttagtc cctgtccctt cagccttatt 6120
 tttcttaagt atatgcacaa tggaaaaggat agataaaattt atatccttag acaggtaaaag 6180
 cattcattaa ttcaagggtgg ttttgcacccaa tactatccaa ggcacccctt ccatgcaccc 6240
 tgactgcaag gccttgcctt tggagagaac ctttatacata ctctcaactg taggagaaac 6300
 actgtttttt tggtactcag tggcatcattt ctatcaactt ttggattttgg cagtttagtca 6360
 aatgagggttc taccattttac caactaggag tctgtggccca agttactcaa tctcccttct 6420
 aaggcccttcc ctcataatgca aaaaggaaat aataagtgtt ttacaagattt catgcacat 6480
 ataatgtatg caaagtgtttt agcatgggtc ctggcatatc ataaactgtt aacaattttt 6540
 agccagctcc tagcattttt ggaggctgag gcaggccgat tgcttgaggc cagcagtcc 6600
 agaccagccct gagcgacata gtggatctt gtccttacaa aaaaaaaaaaa aatgtttttt 6660
 tcatttagctg ggtgtgttgg catgcacccctg tggtagccatc tacttggggag gctgaagtgg 6720
 gagaattgtt tgagccctgg aggttgaagc tacagtccacg tggatgtca ccactacact 6780
 tcaaccatgg caatagatgtt agaccctgtc tcaaaaaaaaaaa aaattatcag ctattactat 6840
 aattttttt attagttctt cactcaccta aaatctccaa caccccttag aaatacacat 6900
 atgagaacaa caaaaatgaa caaagttagaa gcacatatga aaaggcttaag aaagcatgaa 6960

agccagcaag aaataactgc cgctcttctg gtcagcctgg aaagcaactg gcattttccc 7020
 tagacaatg tttcaaaact tggttgcaca ttaaaatcac ctagagacct tttaaaaatt 7080
 ctgaaaccca gcccacacccc cgacaaaatga aagcacagtc tctgggggtg ggacataggc 7140
 atcattttt gaaggcccc acttgatcct aatgtgcaga caaatttcaa aaccacagct 7200
 ttggaggct gaggtgggtg gatcataaga tcaggagtt gagaccagcc tggccaatat 7260
 ggtgaaacccc cgtctctact aaaaatgcaaa acattagctg ggcattgggtg cgcataacctg 7320
 tagtcccage tactcttagag gctgaggcg aagcatca gaggagaat cagtgaaccc 7380
 aggatgtgga agtttcagtg agcccagatc acaccactgc actccagctt gggcgacaga 7440
 gcgagactcc acctcaaaaaa caaaaagaaa gaaagaaaaa agaaaagcag tactctggaa 7500
 cctaccaatc catcattag catatttgc aagtaaaatgt ttattctggt atttgcagaa 7560
 ccagaatcca gctttatgg ccaggccaa ttttttaaa aaacatggta aggtctccca 7620
 gccttaaagt attacatagc atagctctaa aannnnnnnn nnnnnnnnnn nnnnnnnnnn 7680
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nncaggagc ccgagcctcc tcctttgtgt 7740
 atgtcaaccc agaggatgga cggatcttgc cgcagcgtac ctctgactat gaattgtgc 7800
 agatgctgca gattgtggt ggggttcgag actccgggtg tccacccatt ggatgccaac 7860
 acatctctgc atgtgttgtt cctagacgag aatgataatg ccccaagctgt gctgcaccca 7920
 cggccagact gggAACACTC agccccccag cgtctccctc gctctgctcc tcctggctcc 7980
 ttggtcacca aggtgacagc cgtggatgct gatgcaggcc acaatgcgtg gctctctac 8040
 tcactgtgc cacagtccac agccccagga ctgttcctcg tgtctacaca cactggtgag 8100
 gtgcgcacag cccggccctt actggaggat gactctgaca cccagcagggt ggtggctctg 8160
 gtgagggaca atggtgaccc ttcaactctcc tccacagccca cagtgtgtgt ggttctggag 8220
 gatgaggacc ctgagggaaat gcccaaattcc agtgacttcc tcatacaccc tcctgagcgt 8280
 tcagaccccta ccctttacact cattgtggct cttagcagcc tcagtctt atccctagtc 8340
 accttcaccc ttctgtcagc gaagtgcctt cagggaaacg cagacggggg cgggggttgg 8400
 gggcagtgt gcaggcgcca ggactcaccc tccccggact tctataagca gtccagcccc 8460
 aacctgcagg tgagctcgga cggcacgctc aagtacatgg aggtgacgct gcggccacaca 8520
 gactcgcaga gccactgcta caggacgtgc ttttacccgg cctccggacgg cagtgacttc 8580
 actttctaa gaccctcag cgttcagcag cccacagctc tggcgttgg gcctgacgcc 8640
 atccgtccc gctctaatac gctgcgggag cggagccagg tgaggggtc ggcggccccc 8700
 cggggaccc ctggggggcgg cactggagaa gccggccgtc ctatataagggtt attgaacttg 8760
 catccactcc tctccggccg gcttggtcgc tggctgctgt ccacccgatt ctgggatca 8820
 ttggaccgtt tgcgcgaaac cagagtggcc gattaaggga tggggttccg agcaccgggg 8880
 gtggggcga ctgtggcga gggggagggtgg gaccgacccc caccctata ctaaaaaaaaag 8940
 gccggggcct cttcgagct tccgggtaat ttccggcgat ttccgggggt gtcgggggtc 9000
 cggggaggag gcagtcaacatccacccct gcagccagcc tccttagggc cggctccggc 9060
 acgctcgcc ggtctgtaga ttcccttcc tattttctcc cagctccctag catctgtgac 9120
 ttcaactgtta ccctccctat ccccgcatca cccaaacccca cctgtctgctc ggacttaggt 9180
 gtgcgcgcgg ggctcatgct tgcctccct gctggccacc cccacggccc acacaagttg 9240
 cacggctcg ccacgcccccc ccaacacgtg cgggacgcgc cgcacgcact cctcgacagt 9300
 gggctacgc gaataaccagg tttcaactgccc actcgctcgc gcccagattc acaggccctgt 9360
 tccggccac tcgcagctcc cctatgcgc tccctccggc gggctcagga gtactcgtag 9420
 ctgattatgc ggcgcgtgagg gtcccgatc gcccggccccc aggaccaggc gaggactccg 9480
 gagccctcc tCACCTCTCC cacctgcgc cgggctggg cggggtcgcc tggggggccgg 9540
 cctgagcggag ggcggggggc aggagcgtg gagcgtactgc cgctctaagt gcccggccgg 9600
 caggactcta cgatccttgg gccagaggctc cggatggtcc cgggactccg tctcaagggt 9660
 cggcggccccc tcaacccaga agcctcgagc aggccggacag gcagagctgc ccagtggccg 9720
 aggccggca gggctcccg ttggggcagtt gagttagct ctagaggaca gcaggactgg 9780
 gactccatgt gcaccaggccg ccccttcctc cccgcaggaa tggaaaggccctc gagcgccttg 9840
 gctctcaacc cccagagata caggcttttgc gcccgtgt gatcacaatgt tgcccgacgca 9900
 tcttaggtca gagatttggaa ggtgacccaa ctatctgaca ctctaaacaa tccctgttcc 9960
 tctggcagat gaaaagctat aggctctgccc agatgccagg gtgccttat gtgtgaggaa 10020
 actacaatag taaaaaacac aagtttctcc aactccaggaa gcttttatttcaaaaatatac 10080
 aatgcctaacc tctgtccctt ggactgttatt ttgaaacacc cccaggtgtat tttgtatagct 10140
 gattgagaga aacttactat ataactccctt tgagaacccctt atctcatttgc tcttcccac 10200
 cattgtgtt ggcttaggtac taacatgcct ctcttataac agcttcacag aggtcaagt 10260
 acttgcctaa gttcacagag caagtaagag agatcttcaac ccctgtctaa ctccagaatg 10320
 tttgttttta atttcttggc acttggaaact taaaagctt gaggacaggaa gagggggagtt 10380
 gcctctgctg ggattttgc tctgtggga tggggcaagg gtgggggttcc ttccctttta 10440
 ggaccttaca tttggggaaa gtcttctgtg gtccttccatt tctgagcgtt ccccgacgc 10500
 caaacttggc cagttatcct tttggagatt gagttcccccc agctccgtt cctctatca 10560
 caatcactgc atttccatgt agattctgtc gtgtctaaga atacagtggc tgagggttgg 10620

ccatccctgt gcccattctcc atggcagcccc cagaatggtg ctggtgactc ccgatacacc 10680
 tggaaatgcta gttttctggg ttctgctcat atcaactgcga cctgtgagcc ttgagtgagc 10740
 cactggcagt tccggagctt ccttctccct ggagtaatct gagatactca cactctctgc 10800
 ttctgggagc aagtatgagt attaattacc agccacccccc agaaccagtg aggagggtggc 10860
 tcttaggatgt ctccatgaga gtgtgtatg ggataacaca ccccttggaa agaccagaag 10920
 ggactcgagga atgaagtggc tggccagagc cccacaccc cagcttaggtg ggagatggct 10980
 acacatcagc cccttgggag ccctggagac ttatgtggcc ctgcttggag gctgtggag 11040
 ctggatccct ccccgctgca tccctccat ttttctccgt ctcagacaga gcagccttgt 11100
 tctctctcc ttatgtcacag accattgtct ggcacggagt tctaggggtg agaagtgtcc 11160
 cgggacttgg atgcccccca aaggcccaat ctggcatgac tcctaaatta ataatgtatt 11220
 tagctgtggg aagagattct tgcaagccaa gggcccagag aagatgtccc tgtgaatgtg 11280
 tcactgcaca acctggcacc aaaagggtta ccaagaacag cagccatctt gctgcagagg 11340
 atgctttgtt cccagctgag gaggtaata aattcattct agggctggta gaattctcat 11400
 tgaaaagcct ccttgcac ttttagggggc tttgtctgca ccttctccccc cagttccaca 11460
 gaagatgcct tcaatccctt aattttggct caggagttct gactctgggg gcagggagga 11520
 agggggcatt tcttttaggaa aggagtctca gcttgcacac tttgtgtcaga tgaaatgtga 11580
 tttatctctt gttttctggt acctcagacc tctgagacct gaggtgtatt ttgtcttgg 11640
 agatgatgcc accccctgccc cctttagtc cgttttattc tctgtgccc ctccctccct 11700
 tctctcccg cccatcccta ggggctcggt tgacattcta acttctcacc ggtactcagc 11760
 cccttccat ctgtttctc cacagcaagc cccgccccaa acggactggc gtttctctca 11820
 ggcccagaga cccrgcaccga cgggtaggt gactgattct ccagccacc ctcttctctg 11880
 cggcattttc tcacggatga cgtgagagca gatggggggag ggccaccat ttgctacaca 11940
 tggctctcc ctcagtttgta gatcccaggg aggtcttggt gtgggggggg ctggcacaca 12000
 gaccggaa ggaagaggcg actgcccctga ctgttcagga agctcaattt acatgcttgc 12060
 cccttccctc ctgcgccacga ccggcacctt ttcctatccc ctgagggcac tttggaacca 12120
 aaggatggtc ttaagctggt ctctgggtga aagcgaggct ttctatgccc atgtactgccc 12180
 taacccttc ccctgagttt agctgggctc cattatgacc tgggctcag gcagaaaagc 12240
 atttggccgg aggaggcggt ccgcactcag cgccctccccc agggctcca gagccgaggc 12300
 tgactgcagc ctggggagag tggaggcagc acagctggag gtggaaactga ggaaggactg 12360
 ggcggggcat ggagcaggccc cccttctccg gcccctccctt ccactgtcct ctgcccctac 12420
 ttgtctctgt ctctgtctgt ggggtctccg tttgtctgtc ctttttctt gagtttccgt 12480
 ctttgcact ttctcttacc ttcctgtcct tccctcagtc tctatctcgc tttgcaatct 12540
 ctgcctctcc ctctcttctc catctctgtc ttagcttccg ttttgcatttc tgacacccat 12600
 cctcagtc tttcattttgtt ctacttgcatt tttatctgtc ctttttcttgc ctttttctt 12660
 gcataagtgc tcaataaaatg tggagtgtt aactacacgg gacccttagt cttttctcc 12720
 ttctctatct ctgcctccct gtccttgc tttttttttt cttttttttt ctttttcccg 12780
 tttttttttt ctgtgtgtt tttttttttt tttttttttt tttttttttt tttttttttt 12840
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12900
 ctctgtctct ggtctctgtt cttttttttt tttttttttt tttttttttt tttttttttt 12960
 ttatctctct tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13020
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13080
 ccgcctctctt cttccctctc ccctcacaca ctgagccctt gatcgacgt ctccaccaga 13140
 caccctccca tccaaaggcgc cccagctgtt catttcaatc tggatgttgc tttttttttt 13200
 acagggaaaccc cctgcgggtt gaaagggggg gaaacttcag cagaaagacc tttcgttgg 13260
 ctgagcagag tggataggc tctgcgcaggc ccctcccaactt tagaagtgc gggctggag 13320
 gagtctcaga gcccaggaggg agaaggcaggaa gtttggggca ctttaccctt cttccctcaat 13380
 ggtttagtac ttgtgttca catttgc tttttttttt tttttttttt tttttttttt 13440
 atcttcccccc catcctgcaaa ccaacccaaac ctttgc tttttttttt tttttttttt 13500
 ggttatggc agagatttca gtttgc tttttttttt tttttttttt tttttttttt 13560
 gcaaaaggct tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 13620
 gccccttccctt tgacccccc tttttttttt tttttttttt tttttttttt tttttttttt 13680
 agttatccat taagtcattt atgtattcat tttttttttt tttttttttt tttttttttt 13740
 gcatctacta tgagcagggtt cttttttttt tttttttttt tttttttttt tttttttttt 13800
 gtgatccctt aactcatagg tttttttttt tttttttttt tttttttttt tttttttttt 13860
 aacacttaaa aaaaaaaaaa aaggctgggc actgtgggtt acgcctgtt tttttttttt 13920
 ttggggaggcc aagggtgggtt gatcacttgc gtttgc tttttttttt tttttttttt 13980
 aacatggtga aaccccatctt cttttttttt tttttttttt tttttttttt tttttttttt 14040
 acctgttaccc cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 14100
 gtttgcgttca gtttgcgttca cttttttttt tttttttttt tttttttttt tttttttttt 14160
 gagaggcaga gaacaaatgtt gtttgcgttca cttttttttt tttttttttt tttttttttt 14220
 ggaggtgaca ttgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt 14280

ccatacactg gctttgaagc agaattcacc attgaacctg gaagccccac caactgcctc 14340
 tcatgtgtcc tctgggtgta gcacccattacc cctgcctcaa ctcccatcc atcaaatgag 14400
 aggcttagtg gccttggaaa catgtggag ctcttggaga tatggaggag aagggtcgaa 14460
 acctctctt ctctcctctc tctgcttcct tagttactt gtggaggatg caacaaacct 14520
 tttttttttt tttttttttaga cagagtcttgc ctctgttgc caagctggag tgcaactgt 14580
 cgatcttggc tcactgcaac ctctgcctgc caggttcaag cgattcttgc gcctcaggct 14640
 cccgagtagc tgggacaca ggtgtgcacc accacaccca gctcattttt gtattttag 14700
 tagagatggg tttcgccacg ttggccaggc tggttcaaa ctccccgacc tcaactgatc 14760
 cacctgttc agcttcccaa agtgcataa ttacaggtgt gatgccttcc ggccacaaca 14820
 aacattttt attcagcggc tccaaagtga aataagcggtt aggatctgtt ccagtaactt 14880
 caactctct ctttcattca ttcaacaaat ggtcagtgtt gacttctgt tggccaggca 14940
 gtgttctcga tgtgggaggc ccctgcctc aaggagtta cagtgcactg aaggagacaa 15000
 atattgacca aataattaca aaaataatg tgagactgaa actgtcataa tgcttagt 15060
 gaatttgtatc tggtgtgtgt ggcgtataat agggatttga ttttagggagc aggaggacat 15120
 cctggaaatg aagttgtgtt gagatctgaa agacaattac aaattaacca gctaaagaaa 15180
 agaagaaga gcatattccaga tggtcaggca taccagctc aaagcaggc tggtgttag 15240
 ttcttcagta aacttggttc aagatcaagc cagctctgg tcacactgccc tacacagaag 15300
 gaagagcatt ccaaattggaa aaaggctatg caaaaggcctt gtggtgaggc ctccctggat 15360
 caagggaaatg gcaggaggatt ttcttttgc tggactgtt gtagggggat gaagaggaga 15420
 ggttcctttt atccagtcata accctggttc cagccttc ttagtgcctt gcctcactcc 15480
 ctccccacca tcttgcccac tgccctccag cctggtctgg tcagacacac cagcttcaaa 15540
 gcagggtctag tattgagttc ttctgtaaac ttgggttcaag ttcaagtcag cactgggtca 15600
 cactgccttag acagaaggaa gaagacctgg ccaggtgtt gggctcacgc ctgtatccc 15660
 agcactttgg gaggctgagg tgggtggatc acctgaggc aggagtttga gaccagctg 15720
 gccaacatga taaaaaccca tctctagtaa aaatacaaaaa tattagccag tggtgttgg 15780
 gcacacctgt agtcccagct actcaggagg ctgaggcagg ggaatcgctt gaaacccagga 15840
 ggcggagggtt gcagttagctt gagatcacac cactgcactc cagtcgttgc gacagagcga 15900
 gactccatca gaaaaaaaag aaagaaaaag agagagaggg agggagggag ggaaggaagg 15960
 aaggaaggaa ggaaggaaagg aaggaaggaa ggaaggaaagg aagaaaaagaa agaaaaaagg 16020
 agaaaaaaatg aagaagactt acattctgaa gttcctgtt ggcacatgtct tctcctcccc 16080
 accactatcc cttctcagtt catctttgg gtgtctgtt ccaatagctt ttgtctcaag 16140
 ctctgagccc agcaaaccat ttcccccttc ttaggtctgtt acccaggccc tgccctgcag 16200
 ctgtccccag cctctgcaca aagagtctcc atttggtccc tagactgttc tctgtatgt 16260
 ccacactacc atccctccctc tgagagacag tatccctgtt ctctctcagtt ttgggcagtc 16320
 attcaacaaa caaatagcaa cttatttttt ttccactttt aaaaatgtac atgctatcca 16380
 tagaaacatg aaaaatggaa aaaaatattaa aaggaaaaacc acccatcatt ttacaatcta 16440
 gggagaaccca ccaccaacta cagtttaata taaaacattt caaatatgtt caaaagtata 16500
 gagaatgggtg tcatgaacct gtatgatacc actatccagc ttcaactaat gccttatctt 16560
 gtttacaccta tatccctacc cttatccgtt ctttgatc ttttgaagca aatctcagac 16620
 atcatataat ttcatcttgg ttcacccat ctttaacaga tgactttttt tttttttttt 16680
 tttttggag acagagtctc actctttgtt ccaggctgtt gcacaatctc agctcactgc 16740
 agcctcgcc tcccggttgc aagcaattctt catgcttcag cctcttaagt agcttaggatt 16800
 acaggtgtgc accaccacac acagtaataa ttttgtattt tcagtagaaaa cagggttgc 16860
 ccatgttggc caggctggc ttgaactcctt ggcctcaagt gatccaccct cctcgccctc 16920
 caaaaatgtt gggattacag gtgtgagcca ctttgcggccg ccagatgatt ctttcataag 16980
 gtccacattt tatattttagt tgagtgtctc ttaagtatct taatgtctt ttaatgtaaa 17040
 agactcccc tccatcttcc atttttgcatttttttgaatccctgc cttttcatat 17100
 atatcttctt aacttttctt ttgttaatc aaagaacatg catgagctt atttttat 17160
 ttattttttt attttttttt ttattttttt attttttttt tgatcacgtt ctttgctctg 17220
 tcaccctaggc tggactgcag tggcgaatc tcggctcaat gcaagctccg cctccctagg 17280
 tcacccatg ctcctgcctc agccctccaga atagctggga ctacaggcgc ccactaccac 17340
 gccccgctaa ttatttgtt ttttttagtag agacggggtt tcaccgttgc agccaggatg 17400
 gtctcgatct ctcgtacccctg tgatccgccc gcctcgccct cccaaagtgc tgggattaca 17460
 ggcgtgagcc acccgccctg gccgagcttc attttaaaaa aatctcattt acctaaataa 17520
 gttgttcaca aacttgcac actttcatgt gacactttt tcacatatta agaactgtt 17580
 ttcccagtca tacggtagat ttccacttag tttttttttttaatgcaactg aggcataact 17640
 gtccaacaga attatagctt ttttcttctt tgcaactgtt catgagaatt tccccatgtc 17700
 tttcaaaattt cttcccaagt cattgcaga gtaactccctc ttatgaaat ttttgcata 17760
 ttacagaata agtcctcata tgaatataca tattttttt ctaatgttca tgggcctttg 17820
 ctacattaga aaacatacac attcatttac aattttttt ttttgcata gtcctcgctt 17880
 gtctccaggc tgggggtgcag tggcgccatc tcggctcaat gcaagctcca cttccctgg 17940

tcctgcctc agccctccta gtagctggga ctacaggcac gcgccaccac 18000
ttttgtatt ttttagtagag acggggttt accatgttga ccaggatgg 18060
tgacctct tgacctcgta atccgcctgc ctcagcctcc ctaagtctg ggattacagg 18120
tggagccac catccccagc ctcatttaca ttttaacaca attaaattca ggtgttg 18180
agggtgactt cctagattt tcagttttt gagcatgtg tagggagtat ttattatag 18240
tccagtctat gggatattcc ttatacaca cacacacaca cacacacaca 18300
cacccaaacc tcaataggc aagaaccata ttcatcatct cttaaatcct aggtctgtc 18360
acaggccact gctggaaaa tagtgtact aatgttagtco atttcttcca ttttacttt 18420
tcactattcc ccaaattcctg gaagagtccc tcataatctgc tgctgggct ccaaggcctc 18480
caagctgaag cccagtagt tggtgtcat aagatcatag gccctgaaat catattaggc 18540
tggcttcag gttcaactc taccaggcc agctgtgtga ctctgttata gtctctcaac 18600
ctctctgagc cctgttttca tctgttgaa aatatagata atcacacac ttatctgcag 18660
ggtgattccg aggataaaaaa agacaaaatattt attacccagt cagcttagaa taggaagctc 18720
tcagtaaatg tttaaaaaaaag aggtctctct ccctctgtc atcicttacc cagcttcaca 18780
taattggcag cccacccacc tggttgccc tcaaccttca accattccc aggacgcctc 18840
cttacactcc tgatcatcat ggttccaaacc ccaactgttcc actcagctt gttccctcaa 18900
caccactgtc ccctgtgaca ggtcatctc ctaacctctg gcatctggc ccgtgtctc 18960
cctgacaacc cgcattgtct cctccctgca acccccgccg cttgcctgct ctccctactt 19020
tattccatt agcccccttc acctgggct gagaacacctt ggccttgacc ctgcaccctc 19080
agccaggccc acagagctca gtcggggccg aagtaaggac cccagatatt cagttgagga 19140
atthaacatt aatactataa tatagaatcc accagcagat tttcnnnnnn nnnnnnnnnn 19200
nnnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnatggc ctgaaagtaa 19260
gaccccccaga tatctcaggg acgaactcca acagnattac tacattcgag tatccacaag 19320
gctgacngtc attaacgttt cgaagtgta tattgtcgcc tccttaatg gatgagatta 19380
tgagcgagca gagacctcat ccgttgcata cgtgattcgt gtcatgtat atggagcgag 19440
ctctcatctt gattatcggt ttatcgacat ggaaaagttgg agtgaatgct gcccagtagt 19500
agactaatat gtcccagaac caggataactc tagctattct cgatgttta gattgagatt 19560
acacactttg gtccgattac ttcccgcagc ttcaaaagtga gactcacctg atgtggcac 19620
gtcgcagttt cagtgtatgcc gagtctgtt a gatgttctg tcatgttctg 19680
gcctgtataa atagagttt gtaaaggccac cttgacgcattt gaggccaaatcctt 19740
gcggagatac tttagatag gca gcaaccc ttcatgcaat tttagtattt cctgaggatc 19800
cttgcctgta tctgttttta catcagtgtat tacaaagcgg tcatgttctg tcatgttctg 19860
tacatgtatt tgctggattt cttctgaaaa ttacctggat gattcttca gatttaaatt 19920
cactgtctt tccttgggaa agcctggact ggctgagctg ctaaccctg actctctgtc 19980
ttttttttt tttgacagg gagtctagct ctgttgcag gctggagtgc agtggcaacc 20040
ttcacctctt gggtaaagc cttctcttgc ctcagcctc gcaagtagct gggattacag 20100
gcgtgtgcca ccatgccccg gattcttgc ttaattttt ttttttaata tctattttttag ttgagacagg 20160
gtttcaccat gttgtcaga atggctcga tctcctgacc tcgtgtctg cccgccccctg 20220
cctctcttca taccctctgg tagccctga cacagcccat ggcaagtatc acaattgca 20280
ttactctatt gtgtattt gtaaaatcag cctctcagtg cccaccatga gcaccgtaaa 20340
gatggacagt ttctatcata ctcacagcca caaccccccaga acctggctct gtgtctggta 20400
catagaaatt gctcaagaaa gaattttgg agggggcgccg tggctcacgc ctgttaattcc 20460
agcactttgg gaagcggagg cagggtggatc atgaggtcag gagttcgaaa ccaacctggc 20520
caatatggtg aaaccccccatttactttttt ttttttttttgc tgcgtggca tgggtggcatg 20580
tgtcacttg tcgtccccagc tgctcaagag gctgaggcag gagaatcact tgaacctggg 20640
aggcagaggt tgcaagtgc cggatcgcc caactgtact ccagctgggg tggtgtcc 20700
agactccagc taaaaaaaaa aaaaaaaaaaaga aaaaaaaaaaa gaaataatta ttgaatcaat 20760
ggaaacagaa ttacaactct tcccactttt gggaggaga attgagattc ctgcttcccc 20820
cttcaaccct agctttatgt ttgttgcattt ggtatagagg tcacacatgg cagcctgtg 20880
tctgttctt ggctgagttt atctacttag ggacggggaa tcccagattt ggcäääacca 20940
aaagcagggg cagaccctgg aaaaattgtt aagcagtctt tgggtgtct caccacccat 21000
gacaaaatattt gacacacacac aacgcaacac acacacacac attttcatgc ctcaaggcc 21060
taccctacag ctttcccat ggtgccaaac ctttcttctt accttacagt cacttaactg 21120
tctctgtctt tcttcttccca ctcattttagg caacaaaatattt ttattgaaaaa acttctctga 21180
gccagatggt gccaaacagt tacctaaagg gcagactgtg tgcgtccag ggctgtttcc 21240
cgccaaacact ccctgtccaa ttgaaataat ttcatcttgc gaaattcaaa aaaatcatga 21300
aaattgtcac tggggaaaaa ttccaagtag agtgaacct tctaagaagc agagagttct 21360
tcttattggtt tagtattttt ttttttttttca aatcatttca tgggtgtct acactcgag 21420
tttctaaag tctcaacttag gtccaaagtgc taggtttgt gctaggtgct tcagtaggaa 21480
gcaggtggag tcgtgtccctt acagaagcca gtgtgttggaa ggaacagaca ggcactacaa 21540
cqctqcaqt qtqactaqttt ccatttttttttccatgtqqqq cagccqctqq qatqqqaqta cagttqqqq 21600

acctggccag tctgagtgga tctcctctga tatattttat tcgtgtcatc atggcaaatg 21660
 ttggggact ggcattctag ggaacaccat ctgcccagga ttcacagcaa gagggaaacct 21720
 gagggatagg aaagtgcacc ttggaaaagg tttgatgata gagtgttga gggcgggggg 21780
 tggggattct tgtaaaatag gaggggcagg ttcttctatg gccgtggatt ttactctgag 21840
 ggcttaggaa gccattgaat tttagaaatt tcttctccct gctccagggt ctcttgggc 21900
 ctccatcac tcccccttag tacctcctag gtccccactt tttctctagc tctggaacat 21960
 tctctccat ttcccacctt cacagtctc tggagaaggt ctgatctccc tcactatgac 22020
 aggaagccag gctctctcat ggtgggtggg ggaatgcact atgttcaatg cattggacca 22080
 gcctggccaa catggtgaaa ccccgctctgt actaaaaata caaaaattag cctggcacag 22140
 tggcaggtgc ctgtcatccc agctacttgg gagactgagg caggagaatc gcttgaaccc 22200
 aggaggtgga tgttgcaatg agccaagatc ggcgccactgc actccagcct ggatgacaga 22260
 gtaagaccct gtctaaaaca aaaagaaaaa agcatcgagg gcagcaaaaga gctggttiqa 22320
 gggactgact tgtgggattc tggctctga aggattgtag aagtgcctcc taccttgcgt 22380
 agagaaacag gtcttsggca tggggtaggg ttagggtaact aggttggga gccctggag 22440
 aggacaggag gggcatgagc tggccggcc tgggagttcg tgctcaccat cctactct 22500
 cccagctcc caaaaatggcg atgacaccgg cacctggccc aacaaccagt ttgacacaga 22560
 gatgctgcaa gccatgatct tggcgccgc cagtggttaag tgggtcgat gtgtgtatgg 22620
 aagagtggga gacctggggt tctgggtgc atctcacagc caccatgccc acggactgga 22680
 tgtcaaacct gtgttagttc tccagatctt tcggcagggtc taaaaggacc cacaattcca 22740
 aacataaaagc ctggaaattgt ggctaaggaa cagcagaggg gttgggggtc ctggatgcc 22800
 tggaggagag cgaacaagga aaagactcat ggagggaaaata gtggttacaa gtctagct 22860
 ggaatgggac agtgttagctc aggtccttagc tctgctacta tccttggac ttggatcgac 22920
 ttgctcaatc tctctaaacc tcagttcct cagcctagaa agtggggaca gggaccccaa 22980
 agggttgtt tagagattaa atgagatgat accacaaaagc atagagttct gggcctggcg 23040
 ctggaaagcc ctcaagcaat gaaagtgggt gtagaaaatcc ccagccgaga gaacaccc 23100
 tctttatgcg aggctctct tttgtgccc ctaactccag agtaggattc tgggtttccc 23160
 ataaggtga ctaagggttt cctttgtaat agagattcgg ctatgtgttgg ttgtccctga 23220
 agagtcacta gggcccagag caggagaaga agaatatcat ttgtccttaa gagcacagcc 23280
 taagctggc gtgggtggctc acgcctgtaa tcccagact ttgggaagct gaggcaggag 23340
 gatcatgagg tcaggagttc gaaaccggcc tggccaacat agtggaaaccc tgcgtctact 23400
 aaaaatacaa aaattagctg ggcattgtgg tgcgtgcctg tagtcccagc tacttggag 23460
 gctgaggcag gagaatcgc tgaacctggg aggtggctgt ggtgagcaga aaatcgcc 23520
 actgcactcc agcctggca acagtcgc tctgtctcaa aaaaaaaaaa aaaaaaaaaaag 23580
 agcacaggct ttagaggcag gccaatctgg attcaaattcc tggcacctgg caccaccact 23640
 tagcaatgag tccttaggtt agttattaaa cccctcggt ttcacttact attaatggc 23700
 acagtattaa cggatctcat tgggtgtta ccaggccaa atgcaagtcc ctcagcacag 23760
 gattggcaga gagaatggc acaatatatt attatcttgc tgcacccaaa cccgttttt 23820
 actgggtaga aagcagaatt agaataatgc ctattaataa agactagctt ctggagcagc 23880
 aaatcctggg cttataagga ggctggcaga ggtcaaggc attggccat atcatgaact 23940
 gatagccctg ccctccagca tggctggagg actgggtggt agagaaccag gtgaaatcac 24000
 aagaagcaact gtcatcacac ctggccacat actaaacaaa cttagggc ctcctgtcaa 24060
 tgaattctct caatagctc gcaaggcagg tctgatttt cacactcgac agatgagaga 24120
 accgagaagg cacttgcctt cgggtcctca gttttagtgc gccaagctga ggttctaacc 24180
 tggctcaact cctaagttt tgctctttt actataggga ccattggctgc ctttggatcc 24240
 agcctcggt tctctattac ctgagcagga tgagggctg tccttctct cgagctctgc 24300
 tttattctt cagatcaggg ttgccagata aaatacagga tgccagttt cagttgaatg 24360
 tcagataaac agcaagttact atttcagcat aagtcaagtct caaatgttgc atgagacata 24420
 atgtactaaa aaaaaaaaaatg ctcatgttt acctgaaatt caaatthaac tgggagccct 24480
 gtactttat ttactaaatc tggcaactct acccccagatg tctgtgaggt taagtggggc 24540
 caggcctaca gctgacggaa ggacagagag agaggtggca gggactgctg gcctcctgag 24600
 gcagagctgt cccaggctgt gtggggctat gattccagag aggcagaga ctgagtatga 24660
 aagtggcagg cggctgggtc tgaggagatg ccaagttggc ctcttgaggg aaataaacag 24720
 gtatatttag cctgtgtgg ctttgcggcc tgaggccagg aagcagctt ttagcctaaa 24780
 tccagatgtt aaaaacagaa atgaaatcag tattttatgca cccaaaccct ccaagcaagt 24840
 catgcagctc gttccctgt caaggcctcc caccttgc ccacacaggc cctgaccctc 24900
 gtcttaaggct gcggccctggg gaacggaccc tgggggtgga ggtgggtggc caggccctgc 24960
 tctcaatttc acaccccgctt cctagccctg aaaccagaga gtttctgaa gtccagccca 25020
 gccaggcctg tggggctgtc gagaggggtt aagtaagagg gagagtaaaa accgacactt 25080
 gggaggaggg tttaaaata aacatctgcg gggagggatg ctctcaagga ggctcggtt 25140
 tgcagctcag ccagctgtgt tccagctaat cagcctccctt gggctactct ggagttgtc 25200
 cttggccctt gggatggggatg atgtaagag gacagaggaa tgcaggaggg 25260

agtttccctg agtagaggct ggcacaggag agaaggcatc acccccaccc cgtccaggcc 25320
 agaagatgtc cagtttttg aggctctcct aagtctgcct ctccctggac caagagaaaa 25380
 tcccggtctt tgaccaagggt gggcccttggg gggagtgggg ttagaggga gggggactgg 25440
 agaactgact ctacagaaaag tcaaagctgg caatccaact tcttcccttc agatttatag 25500
 atggggaaaac aggcttgagc cacgcagaga ctgaccaag ctcacacagt tccttagtgg 25560
 cagagcagag caaatattct ctttcttta cattctggat ttccatatct tctctccctc 25620
 ctgggtccagc ccaacctcag ggcacccccc acgggtggg cggggatgg ctgttgcact 25680
 tgcacccact cgggggtgcta ctcagagatc ttgggtgcac ataagacgtg ggtggggccga 25740
 gcttctagct actccgtcag gcccttcctg tcattctgtc tctgcctccc ttctccctgc 25800
 ttctccgtgt tcctcctcat tctttctgt gtgcagggag actacaccccc ccaccccgct 25860
 ctttctctgg cgcctctgag gtccccctc tagccccctaa atcactctgg aatcctggct 25920
 ctggtaagcc agatctgggc cccctccccc taccctttcc attcccaaggc tggaaaggc 25980
 tgaagaggct gaccgcttggg cgggaggggg cgggtcggtg gcggatctgg ctgccttttg 26040
 gagttaaatta gggaaaacacag agaaatgtca gcggaatgaa agggctgggg gtggggggca 26100
 gctgggttag gagaggagga agtgggcagc tgccctccctc cactcaaccc ctctccccc 26160
 ccccccagaaaa gctctcagct cgggggattt ggcacatgaa atgggggctg tagaaacccctg 26220
 agcgctggtg cgtgaagaga aaaaccgggg cgcatccccc ctctcccttc ccgtgtgccc 26280
 tcctcctctt atgcccggctt gaaaatattt cctgtctctc tatttctcaa tccctgggtt 26340
 atgtcccagg attactcagc ctctccggct ttagtcaactc tcgcatacccg ctcccaagggt 26400
 ccagggtggg ggcagaggga ggctggaga gaagctctac tgccagctgg gcctgggtg 26460
 gcctggcat ccctgaggtt ttaactgtct cctaaccaca gaggatctca gggcctccag 26520
 cagcgagccc caatgagtc aactcttggt tcctcctctc ccaccccccga cccagtcagg 26580
 gcaggtgagg ggttagagggt atggcatgct ggtatgtacg gtgtgatga tgtattgcac 26640
 aggatgaagc accatctcat ttagtcctca cagcagccct cacagtgcgt acaaattcagc 26700
 tggcaattcc gagaggctgc attctaaaca agttcctggg tgatgtctgag ccagggccag 26760
 agtgtggact tctctggggcc ccagttctt ctctgtaca gtggacgtt ggactggagg 26820
 tgctgtctga tgcacccatg ctgtggact ctgagcctgt gcctaaagtgc tccctgagat 26880
 gtcttagtccc cgagatcatc atctccctcag ctcatagag ccgagctt ctccatcttc 26940
 tctctacctc ccattcactc agaacaggga taaggccaa gctgtgggtg gtgtgggtga 27000
 ctgactgtcc cttccagcat ggccttaggt acacttgggc ctgaatgccc ctctaacaaa 27060
 tgctaccggg tatggccttgc tgccttctaa cacttggtc cctcatctgt aaaatgaggg 27120
 tcataccaca tagctggctg ggctgttatg tagaggttaa aagtattgt gcatgtgaag 27180
 catccagcag agtgcctggc acacagtagg tgctcaaattt ctgttttggaa atacaaaaat 27240
 tagctgggtc agtgggtgtc acctgttaatc tcagctactc aggaggttga ggcaggagaa 27300
 tcacttgagc ccaggaggca gaggttgcag tgagccggga tcgcgccact gcactccagg 27360
 ctggcgaca gacggact ccattttaaa aaaaaaaaaaaa aaaaaatgtt gtttgaaat 27420
 ggaggcttgg aagagcactc ttccatccca ccccacacac ttactaacgg aactaccgaa 27480
 tctctactcg tgccaggcct tgcttggca aggctagggg catgtaaagc catggttca 27540
 tggctgtgc cttccaggag cccatggctc agtgtggag tcagaggcgc ctttgagaaaa 27600
 gcccaggca ggaactgaga gcccgtgtca cagcccttag agaaataccca gcgaatgtgt 27660
 tttagcact tacatctacc aggcattttt ttacgtaccc gacatacatt ttcttatttt 27720
 gcattcacta aaaccagaag aggtggttac cattatcatc cccatggc agatgagaaaa 27780
 actgaagtgg cataaggagg ttaggttagac ttgcccaggaa tcacagttagt gtaagggaaagg 27840
 agcttaggaat caaacctaga ctttctcact ccagaggcctc ttaacctctaa ccctcaactc 27900
 ttgtgaagag agactaccc tggacccctt atattccctt tgctcagctg ggtcaaggcc 27960
 cagggtgctg tataggcatt cattaactgc tggggatata agacagtgc tggtgaggac 28020
 cctaaatcttta gctccacccat gatcttcctc tgcctctgc gaaacctgtg atgggagctc 28080
 caccctggg ggggggtgccc gcaccatggg attgagccgc cgctacggc cccagttcac 28140
 cctgcagcac gtggccgact accgcccagaa tgcctacatc ccaggcagca atgccacact 28200
 gaccaacgca gctggcaagc gggatggca gggcccccagca ggtggcaatg gcaacaagaa 28260
 gaagtcgggc aagaaggaga agaagtaaca tggaggcoag gccaagagcc acagggcrgc 28320
 ctctcccaa ccagccccagc ttctccttac ctgcacccag gcctcagatg ttcagggtca 28380
 accccccagaa tactggtagg ggccaaaggcc atgcctccct tgraaacag aaacaagtgc 28440
 ccagtcagca cttccatccctt ccccccagg gggttgaaya tgcaaaagca gttccgtgg 28500
 gaaccccccattt ccaatcaact gctgtacccca tggggtagt ggggttactg tagacaccaa 28560
 gaaccatgg ccacaccccg ttaggttaca gctgaactcc tccatcttcc aaatcaatca 28620
 gggccatcca tcccatgcct ccctcccttcc caccctccatc caacagtcc tctttccca 28680
 gtaagggtgtt tgggggtgtt aagtaccaag taacctacaa gcctcctagt tctgaaaagt 28740
 tgsaagggca tcatgacccctc ttggcctctc ctgttattt caatctccccc ccaaagcatg 28800
 gtttgggtgcc agcccccattca cttccatccca gagcccaaga tcaatgtca agttttggag 28860
 gacatgatca ccatccccat ggtactgtatg ctgttgcgtt ttagggaggg cattttgtca 28920

ccaaggctct tcccaacgcc ctggggacca gtcttctgtt ttgttttca ttgtttgacr 28980
 tttccactgc atgccttgcac ttccccccacc tcctcctcaaa acaagagact ccactgcatg 29040
 ttccaagaca gtaggggtt gtaagataag gaagggaagt gtgtggatgt ggatgggtgg 29100
 ggcatggaca aagcttgaca catcaagtta tcaaggccctt ggaggaggct ctgtatgtcc 29160
 tcaggggact gacaacatcc tccagattcc agccataaac caataactag gctggacct 29220
 tcccactaca taataggct cagcccaggc agccagctt gggctgagct aacaggacca 29280
 atggattaaa ctggcatttc agtccaaggc agtctcgaaagc aggtttagga ccaggtcccc 29340
 ttgagaggc agaggggctt ctgtgggtgc tgggtactcc agaggtgcca ctgggtggaaag 29400
 ggtcagyggc gccccagcag gaagggtggg ccagccaggc cattcttagt ccctgggttg 29460
 gggaggcagg gagcttagggc agggaccaa tgaacagaar gtctcagccc aggatggggc 29520
 ttcttcaaca gggcccctgc cctcctgaag cctcagtcct tcacccgttcc aggtgccgtt 29580
 tctctccgtt gaaggccact gcccaggtcc ccartgcgcc cccttagtggc catagccctgg 29640
 ttaaagtcc ccagtgcctc cttgtgcata gaccttcttc tcccacccccc ttctgcccct 29700
 gggccccggg ccatccagcg gggctgcccag agaacccccac acctgcccctt acagtagtgt 29760
 agcgcacccctt ccctctttcg gctgggttag aatagycagt agttagtgtc ggtgtgcctt 29820
 tacgtatgg cgggtgggca gcgggcggcg ggctccgcgc agccgtctgt cttgtatctg 29880
 cccgcggcgcc cccgtgttgtt gttttgtctt gtgtccacgc gctaaggcga ccccccctcccc 29940
 cgtaactgact tctcttataaa gcgcctcttc tcgcataagtc acgttagctcc caccacccccc 30000
 tcttcctgtt tctcacgc aaatataact ctaatattt tatggctttt tttcttcgac 30060
 aaaaaataaa taaaacgtt cttctgaaaa gctgaacgtt tctgtataag cgatggaaagc 30120
 tcctggcatg tgtgcattt gatgtgagct gaggtgggtt ctggaaagaag ggcggaaatcg 30180
 ggaggccact ctgtgtcatt gcgcgtctag atgtttccga attgcgtgtg tgtgtgtac 30240
 tgtg 30244

<210> 125
 <211> 2820
 <212> DNA
 <213> Homo sapiens

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

<220>
 <221> misc_feature
 <222> (61)..(120)
 <223> nucleotides represent sequence between PS1 and PS2

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

<220>
 <221> misc_feature
 <222> (181)..(240)
 <223> nucleotides represent sequence between PS2 and PS3

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

<220>
 <221> misc_feature
 <222> (301)..(360)
 <223> nucleotides represent sequence between PS3 and PS4

<220>

```
<221> allele
<222> (390)
<223> PS4: Polymorphic base C or A

<220>
<221> misc_feature
<222> (421)..(480)
<223> nucleotides represent sequence between PS4 and PS5

<220>
<221> allele
<222> (510)
<223> PS5: Polymorphic base A or G

<220>
<221> misc_feature
<222> (541)..(600)
<223> nucleotides represent sequence between PS5 and PS6

<220>
<221> allele
<222> (630)
<223> PS6: Polymorphic base G or C

<220>
<221> misc_feature
<222> (661)..(720)
<223> nucleotides represent sequence between PS6 and PS7

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

<220>
<221> misc_feature
<222> (781)..(840)
<223> nucleotides represent sequence between PS7 and PS8

<220>
<221> allele
<222> (870)
<223> PS8: Polymorphic base G or A

<220>
<221> misc_feature
<222> (901)..(960)
<223> nucleotides represent sequence between PS8 and PS9

<220>
<221> allele
<222> (990)
<223> PS9: Polymorphic base C or T

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> nucleotides represent sequence between PS9 and PS10

<220>
<221> allele
```

```
<222> (1110)
<223> PS10: Polymorphic base T or C

<220>
<221> misc_feature
<222> (1141)..(1200)
<223> nucleotides represent sequence between PS10 and PS11

<220>
<221> allele
<222> (1230)
<223> PS11: Polymorphic base C or T

<220>
<221> misc_feature
<222> (1261)..(1320)
<223> nucleotides represent sequence between PS11 and PS12

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

<220>
<221> misc_feature
<222> (1381)..(1440)
<223> nucleotides represent sequence between PS12 and PS13

<220>
<221> allele
<222> (1470)
<223> PS13: Polymorphic base G or C

<220>
<221> misc_feature
<222> (1501)..(1560)
<223> nucleotides represent sequence between PS13 and PS14

<220>
<221> allele
<222> (1590)
<223> PS14: Polymorphic base C or T

<220>
<221> misc_feature
<222> (1621)..(1680)
<223> nucleotides represent sequence between PS14 and PS15

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

<220>
<221> misc_feature
<222> (1741)..(1800)
<223> nucleotides represent sequence between PS15 and PS16

<220>
<221> allele
<222> (1830)
```

<223> PS16: Polymorphic base G or A

<220>

<221> misc_feature

<222> (1861)..(1920)

<223> nucleotides represent sequence between PS16 and PS17

<220>

<221> allele

<222> (1950)

<223> PS17: Polymorphic base G or A

<220>

<221> misc_feature

<222> (1981)..(2040)

<223> nucleotides represent sequence between PS17 and PS18

<220>

<221> allele

<222> (2070)

<223> PS18: Polymorphic base T or C

<220>

<221> misc_feature

<222> (2101)..(2160)

<223> nucleotides represent sequence between PS18 and PS19

<220>

<221> allele

<222> (2190)

<223> PS19: Polymorphic base G or C

<220>

<221> misc_feature

<222> (2221)..(2280)

<223> nucleotides represent sequence between PS19 and PS20

<220>

<221> allele

<222> (2310)

<223> PS20: Polymorphic base G or A

<220>

<221> misc_feature

<222> (2341)..(2400)

<223> nucleotides represent sequence between PS20 and PS21

<220>

<221> allele

<222> (2430)

<223> PS21: Polymorphic base C or T

<220>

<221> misc_feature

<222> (2461)..(2520)

<223> nucleotides represent sequence between PS21 and PS22

<220>

<221> allele

<222> (2550)

<223> PS22: Polymorphic base A or G

<220>
 <221> misc_feature
 <222> (2581)..(2640)
 <223> nucleotides represent sequence between PS22 and PS23

<220>
 <221> allele
 <222> (2670)
 <223> PS23: Polymorphic base G or A

<220>
 <221> misc_feature
 <222> (2701)..(2760)
 <223> nucleotides represent sequence between PS23 and PS24

<220>
 <221> allele
 <222> (2790)
 <223> PS24: Polymorphic base C or T

<400> 125
 tgttgttag tttcggtt ggttcgatcy ccttcgttgt gatcaaagaa agtgattcaa 60
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
 cacagccccca aggctggctt tccctaaggy aggttcctta tgcaccgata ctggcaaggc 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 ctaaggtagg ttccatatgc accgatactr gcaaggcgct ttggctggaa actctggaaag 300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 agagactgaa cagccggcga gcaaataam ggcatacaga aagccatgtc ggactcggcg 420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
 tccagaaaagc catgtcgac tcggcgccr gcgcacaagc gctaaccgc tgaaagttc 540
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
 cagacgcggg aggacagcac caagatgcgs gagctgggt tggagcgcgc cctggaccga 660
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
 cgaatgacaa tgcgcctgtc ttcaaccagy ccttgcattc ggcgcgcgc ctggaggatg 780
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
 gacaatgcgc ctgtcttcaa ccagtccctt taccggcgc gcgtccgtt gatatgcaccc 900
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
 gaaattattt actccattcg cagccacaay cgccgcggcg tgcgcact attcgcccta 1020
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
 ctcccttaa tcctggttc tgtgggtt gtggtcacag tgccgttgc aatcatattc 1140
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
 ctggatgc attcacactt tcaactggcy cttccatgtat caaagttagt gcctttgtga 1260
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
 actggcgtt ctctcaggcc cagagacccr gcaccagcgg gtaggtgact gattctccag 1380
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
 ctccatgtt gctgagagaa acaggtctt ggcattgggt agggtaggg tactagttt 1500
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
 tctgggtgc atctcacagc caccatgcy acggacttga tgtcaaacct gtgttagttc 1620
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680
 ggcaccatgg gattgagcgc ccgcgtacggr ccccagttca ccctgcagca cgtccccac 1740
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1800
 catggaggcc aggccaagag ccacaggcr gcctctcccc aaccagccca gcttctcctt 1860
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1920
 tggtaggggc caaggccatg ctcccttgc gaaacagaaa caagtccca gtcagcacct 1980
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040
 acctacccct tccccccctt ggggttgaay atgcaaaagc agttccgtg ggaacccca 2100
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2160
 cctacaagcc tccttagttct gaaaagttgs aaggcatca tgactcttg gcctctcctt 2220
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2280
 gtcttctgtt ttgttttca ttgtttgacr ttccactgc atgccttgc ttccccacc 2340

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2400
tccagagggtg ccactggttgg aagggtcagy ggagccccag caggaagggt gggccagcca 2460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2520
gagcttagggc agggaccaaa tgaacagaar gtctcagccc aggatgggc ttcttcaaca 2580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2640
ttccgtgaag gccactgccc aggtccccar tgcgccccct agtggccata gcctggtaa 2700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2760
ccctccctct ttcggctggt gtagaatagy cagtagtgta gtgcgggttg ctttacgtg 2820