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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 November 2001 (15.11.2001)

PCT

(10) International Publication Number
WO 01/85742 A2

(51) International Patent Classification[7]: C07H

(21) International Application Number: PCT/US01/14773

(22) International Filing Date: 7 May 2001 (07.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/201,950 5 May 2000 (05.05.2000) US

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

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **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). **LEE, Helen, H.** [US/US]; 26 Rivendell Drive, Shelton, CT 06484 (US).

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

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, 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 01/85742 A2

(54) Title: HAPLOTYPES OF THE NPY1R GENE

(57) Abstract: Novel genetic variants of the Neuropeptide Y Receptor Y1 (NPY1R) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the NPY1R gene. Compositions and methods for haplotyping and/or genotyping the NPY1R gene in an individual are also disclosed. Polynucleotides defined by the sequence the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE NPY1R GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/201,950 filed May 5, 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 neuropeptide Y receptor Y1 (NPY1R) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

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

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

groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

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

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

One pharmaceutically-important gene for the treatment of cardiovascular diseases and depression is the neuropeptide Y receptor Y1 (NPY1R) gene or its encoded product. Neuropeptide Y (NPY) is the most abundant neuropeptide in the mammalian nervous system and exhibits a diverse range of important physiological activities, including effects on psychomotor activity, food intake, regulation of central endocrine secretion, and potent vasoactive effects on the cardiovascular system (Herzog et al., *Proc. Natl. Acad. Sci. U. S. A* 1992; 89:5794-5798). NPY agonists and antagonists may be useful in treating various conditions, such as anorexia, epilepsy, anxiety, depression, hypertension and heart

failure (Munglani et al., *Drugs* 1996; 52:371-389). NPY1R is one of the receptors of NPY and is a member of the G protein-coupled receptor family (Larhammar et al., *J. Biol. Chem.* 1992; 267:10935-10938). NPY1R has been associated with NPY-evoked vasoconstriction due to its expression in human cultered vascular smooth muscle cells. The presence of the NPY1R protein determines NPY activity in blood vessels and thus plays an important role in the effects of NPY on the cardiovascular system (Qureshi et al., *Regul. Pept.* 1998; 75-76:215-220). Therefore, NPY1R may be involved in the etiology of cardiovascular diseases, including hypertension and heart failure.

NPY1R is also expressed in higher levels in the brain regions of the Flinders Sensitive Line rats (FSL), which are an animal model of depression (Caberlotto et al., *Neurosci Lett.* 1999; 265:191-194). Thus, the authors concluded that NPY1R may be involved in depression.

The neuropeptide Y receptor Y1 gene is located on chromosome 4q31.3-q32 and contains 3 exons that encode a 384 amino acid protein. Reference sequences for the NPY1R gene (Genaissance Reference No. 448189; SEQ ID NO: 1), coding sequence (GenBank Accession No:NM_000909), and protein are shown in Figures 1, 2 and 3, respectively.

Because of the potential for variation in the NPY1R gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the NPY1R 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 NPY1R as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 3 novel polymorphic sites in the NPY1R gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 504 (PS1), 685 (PS2) and 1587 (PS3). The polymorphisms at these sites are thymine or guanine at PS1, guanine or adenine at PS2 and thymine or cytosine at PS3. 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 PS1-3 in the NPY1R gene, which are shown below in Tables 4 and 3, respectively. Each of these NPY1R haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the NPY1R 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 NPY1R 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 and PS3 in both copies of the NPY1R 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 NPY1R polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel NPY1R polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 4 below or has one of the haplotype pairs in Table 3 below.

The invention also provides a method for haplotyping the NPY1R gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the NPY1R gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2 and PS3. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's NPY1R gene is defined by one of the NPY1R haplotypes shown in Table 4, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's NPY1R gene are defined by one of the NPY1R haplotype pairs shown in Table 3 below, or a sub-haplotype pair thereof. The method for establishing the NPY1R 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 NPY1R activity, e.g., cardiovascular diseases and depression.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate NPY1R as a candidate target for treating a specific condition or disease predicted to be associated with NPY1R activity. Determining for a particular population the frequency of one or more of the individual NPY1R haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue NPY1R as a target for treating the specific disease of interest. In particular, if variable NPY1R activity is associated with the disease, then one or more NPY1R haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed NPY1R haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable NPY1R 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 NPY1R haplotype or haplotype pair, apply the information derived from detecting NPY1R haplotypes in an individual to decide whether modulating NPY1R activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting NPY1R to treat a specific condition or disease predicted to be associated with NPY1R activity. For example, detecting which of the NPY1R 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 NPY1R isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular NPY1R 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 NPY1R 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 NPY1R 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 NPY1R haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute NPY1R 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 NPY1R 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 NPY1R haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a NPY1R genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the NPY1R genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the NPY1R genotype or haplotype in a reference population. A higher frequency of the NPY1R genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the NPY1R 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 NPY1R haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for cardiovascular diseases and depression.

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 NPY1R gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at PS1, adenine at PS2 and cytosine at PS3.

A particularly preferred polymorphic variant is an isogene of the NPY1R gene. A NPY1R isogene of the invention comprises thymine or guanine at PS1, guanine or adenine at PS2 and thymine or cytosine at PS3. The invention also provides a collection of NPY1R isogenes, referred to herein as a NPY1R genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a NPY1R cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises cytosine at a position corresponding to nucleotide 507. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a NPY1R isogene defined by haplotypes 1-3.

Polynucleotides complementary to these NPY1R genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the NPY1R gene will be useful in studying the expression and function of NPY1R, and in expressing NPY1R protein for use in screening for candidate drugs to treat diseases related to NPY1R 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 NPY1R for protein structure analysis and drug binding studies.

The present invention also provides nonhuman transgenic animals comprising one of the NPY1R polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the NPY1R isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against NPY1R protein, and for testing the efficacy of therapeutic agents and compounds for cardiovascular diseases and depression in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the NPY1R gene. The computer system comprises a computer 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 NPY1R gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing NPY1R haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the NPY1R gene (Genaissance Reference No. 448189; contiguous lines; 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 reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:21 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).

Figure 2 illustrates a reference sequence for the NPY1R coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the NPY1R protein (contiguous lines; SEQ ID NO:3).

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the NPY1R gene. As described in more detail below, the inventors herein discovered 4 isogenes of the NPY1R gene by characterizing the NPY1R gene found in genomic DNAs isolated from an Index Repository that 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 (22 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents 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		22
	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		17
	Caribbean	7
	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 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 NPY1R isogenes present in the human reference population are defined by haplotypes for 3 polymorphic sites in the NPY1R gene, all of which are believed to be novel. The novel NPY1R polymorphic sites identified by the inventors are referred to as PS1-3 to designate the order in which they are located in the gene (see Table 2 below). Using the genotypes identified in the Index Repository

for PS1-3 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the NPY1R gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the NPY1R gene include those shown in Tables 3 and 4, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether NPY1R is a suitable target for drugs to treat cardiovascular diseases and depression, screening for such drugs and reducing bias in clinical trials of such drugs.

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

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

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

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

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

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

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

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

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

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

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

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

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a

population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

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

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

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

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

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

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

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

Polymorphic site (PS) – A position 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 amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

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

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

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

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

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

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

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

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

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

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

As discussed above, information on the identity of genotypes and haplotypes for the NPY1R 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 NPY1R polymorphisms and haplotypes identified herein.

The compositions comprise at least one NPY1R genotyping oligonucleotide. In one embodiment, a NPY1R 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 preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and

the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a NPY1R polynucleotide, i.e., a NPY1R 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-NPY1R 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 NPY1R gene using the polymorphism information provided herein in conjunction with the known sequence information for the NPY1R 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 NPY1R gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TCCAAAAKTCCTAGG (SEQ ID NO:4) and its complement,
TACCAGCRGATCTTC (SEQ ID NO:5) and its complement, and
TGGCTTCYTCTTTGC (SEQ ID NO:6) and its complement.

A preferred ASO primer for detecting NPY1R gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TTTCTTTCCAAAAKT (SEQ ID NO:7); GCATTGCCTAGGAMT (SEQ ID NO:8);
GAGAAATACCAGCRG (SEQ ID NO:9); AGTGGGGAAGATCYG (SEQ ID NO:10);
TTGCTGTGGCTTCYT (SEQ ID NO:11); and GGAAAGGCAAAGARG (SEQ ID NO:12).

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 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 NPY1R gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

CTTTCCAAAA (SEQ ID NO:13); TTGCCTAGGA (SEQ ID NO:14);
AAATACCAGC (SEQ ID NO:15); GGGGAAGATC (SEQ ID NO:16);
CTGTGGCTTC (SEQ ID NO:17); and AAGGCAAAGA (SEQ ID NO:18).

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.

NPY1R 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 NPY1R 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 NPY1R gene in an individual. As used herein, the terms "NPY1R genotype" and "NPY1R 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 NPY1R 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 NPY1R 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 and PS3 in the two copies to assign a NPY1R 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 PS1-3.

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 NPY1R 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 NPY1R gene fragment is isolated, it must 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 NPY1R 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 and PS3 in that copy to assign a NPY1R haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the NPY1R gene or fragment such as one of the methods described above for preparing NPY1R 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 only provide haplotype information on one of the two NPY1R gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional NPY1R 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 NPY1R gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-3 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the NPY1R haplotypes shown in Table 4. This can be accomplished by identifying, for one or both copies of the individual's NPY1R gene, the phased sequence of nucleotides present at each of PS1-3. The present invention also contemplates that typically only a subset of PS1-3 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* **97**:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* **22**:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* **4**:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a NPY1R 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 and PS3 in each copy of the NPY1R 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 PS1-3 in each copy of the NPY1R gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while

detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the NPY1R gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* **89**:392-396, 1992).

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

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support

subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the NPY1R 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, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

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

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

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. 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 NPY1R haplotype pair is predicted from its NPY1R genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a NPY1R genotype for the individual at two or more NPY1R polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing NPY1R 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 NPY1R haplotype pairs shown in Table 3.

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

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

In one embodiment of this method for predicting a NPY1R haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is

compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*). A preferred process for predicting NPY1R haplotype pairs from NPY1R genotypes is described in U.S. Provisional Application Serial No. 60/198,340 and the corresponding International Application filed April 18, 2001.

The invention also provides a method for determining the frequency of a NPY1R 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 NPY1R 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 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 NPY1R 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 NPY1R genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. 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 reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular NPY1R 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 NPY1R genotype, haplotype or haplotype pair.

Preferably, the NPY1R genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, 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 NPY1R or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a NPY1R genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

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

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the NPY1R 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 NPY1R genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their NPY1R 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 NPY1R 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 NPY1R 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 NPY1R 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 NPY1R genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the NPY1R gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the NPY1R 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 NPY1R genotype or haplotype that is in turn correlated with the clinical response. In a

preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the NPY1R 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 NPY1R 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 and PS3. Similarly, the nucleotide sequence of a variant fragment of the NPY1R 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 NPY1R gene, which is defined by haplotype 4, (or other reported NPY1R sequences) or to portions of the reference sequence (or other reported NPY1R 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 guanine at PS1, adenine at PS2 and cytosine at PS3. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the NPY1R gene which is defined by any one of haplotypes 1-3 shown in Table 4 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the NPY1R 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.

NPY1R isogenes may be isolated using any method that allows separation of the two "copies" of the NPY1R gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides NPY1R genome anthologies, which are collections of NPY1R isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A NPY1R genome anthology may comprise individual NPY1R 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 NPY1R 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, freeze-dried preparations and the like. A preferred NPY1R genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 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 NPY1R protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant NPY1R sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the NPY1R gene will produce NPY1R 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 NPY1R cDNA comprising a nucleotide sequence which is a polymorphic variant of the NPY1R reference coding sequence shown in Figure 2. Thus, the invention also provides NPY1R 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 cytosine at a position corresponding to nucleotide 507. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a NPY1R isogene defined by haplotypes 1-3. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphism described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized NPY1R 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.

As used herein, a polymorphic variant of a NPY1R 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.

In describing the NPY1R 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 NPY1R 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 polynucleotides which are complementary to the sense strand of the NPY1R 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 NPY1R protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the NPY1R 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 NPY1R isogene. Expression of a NPY1R 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 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 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 NPY1R mRNA transcribed from a

particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of NPY1R mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, 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 as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

Effect(s) of the polymorphisms identified herein on expression of NPY1R may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the NPY1R 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 NPY1R 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 NPY1R isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the NPY1R isogene is introduced into a cell in such a way that it recombines with the endogenous NPY1R gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired NPY1R gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the NPY1R isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the NPY1R 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 NPY1R gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One

method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the NPY1R isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human NPY1R isogene and producing human NPY1R protein can be used as biological models for studying diseases related to abnormal NPY1R expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel NPY1R isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel NPY1R isogenes; an antisense oligonucleotide directed against one of the novel NPY1R isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel NPY1R 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 NPY1R 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 NPY1R 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 NPY1R polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

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

EXAMPLE 1

This example illustrates examination of various regions of the NPY1R gene for polymorphic sites.

Amplification of Target Regions

The following target regions were amplified using either the PCR primers represented below or 'tailed' PCR primers, each of which includes a universal sequence forming a noncomplementary 'tail' attached to the 5′ end of each unique sequence in the PCR primer pairs. The universal 'tail' sequence for

the forward PCR primers comprises the sequence 5′-TGTAAAACGACGGCCAGT-3′ (SEQ ID NO:19) and the universal 'tail' sequence for the reverse PCR primers comprises the sequence 5′-AGGAAACAGCTATGACCAT-3′ (SEQ ID NO:20). The nucleotide positions of the first and last nucleotide of the forward and reverse primers for each region amplified are presented below and correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	160-181	complement of 788-765	629 nt
Fragment 2	1100-1123	complement of 1597-1576	498 nt
Fragment 3	1384-1406	complement of 1994-1972	611 nt
Fragment 4	1681-1704	complement of 2295-2276	615 nt
Fragment 5	1995-2017	complement of 2526-2505	532 nt
Fragment 6	2222-2243	complement of 2827-2804	606 nt

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

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

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

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

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/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using either the primer sets represented below with the positions of their first and last nucleotide corresponding to positions in Figure 1, or the appropriate universal 'tail' sequence as a primer.

Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	Tailed Seq.	
Fragment 2	Tailed Seq.	
Fragment 3	1409-1430	complement of 1957-1936
Fragment 4	1726-1747	complement of 2243-2223
Fragment 5	Tailed Seq.	
Fragment 6	Tailed Seq.	

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 NPY1R gene are listed in Table 2 below.

Table 2. Polymorphic Sites Identified in the NPY1R Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position	Reference Allele	Variant Allele	CDS Position	AA Variant
PS1	7794926	504	T	G		
PS2	7794834	685	G	A		
PS3	448212	1587	T	C	507	S169S

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

EXAMPLE 2

This example illustrates analysis of the NPY1R polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 3 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 3, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 3 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 3. Genotypes and Haplotype Pairs Observed for NPY1R Gene

Genotype Number	Polymorphic Sites PS1	PS2	PS3	HAP Pair	
1	T	G	T	4	4
2	T/G	G	T	4	1
3	T	G	T/C	4	3
4	T	G/A	T	4	2

The haplotype pairs shown in Table 3 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 augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family) and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 4 human NPY1R haplotypes shown in Table 4 below.

Table 4. Haplotypes Identified in the NPY1R Gene

HAP No.	HAP ID	Polymorphic Sites PS 1	PS 2	PS 3
1	7815183	G	G	T
2	7815180	T	A	T
3	7815185	T	G	C
4	7815172	T	G	T

Isogenes of the NPY1R gene are identical to the reference sequence at the regions examined except where they vary at the polymorphic site. In particular, a specific NPY1R isogene identified by a haplotype in Table 4 comprises each oligo in the set of oligos shown in Table 5 for that NPY1R haplotype.

TCCAAAATTCCTAGG (SEQ ID NO:22)
TCCAAAAGTCCTAGG (SEQ ID NO:23)
TACCAGCGGATCTTC (SEQ ID NO:24)
TACCAGCAGATCTTC (SEQ ID NO:25)
TGGCTTCTTCTTTGC (SEQ ID NO:26)
TGGCTTCCTCTTTGC (SEQ ID NO:27)

Table 5. Oligonucleotides Comprising Haplotypes for the NPY1R Gene

Haplotype Number	PS1		PS2		PS3	
	SEQ ID NO:22	SEQ ID NO:23	SEQ ID NO:24	SEQ ID NO:25	SEQ ID NO:26	SEQ ID NO:27
1		X	X		X	
2	X			X	X	
3	X		X			X
4	X		X		X	

Polymorphic variants of NPY1R coding sequence isogenes are identical at the regions examined to the reference NPY1R coding sequence, SEQ ID NO:2 (Fig. 2), except where they vary at a polymorphic site. NPY1R coding sequence isogenes are defined by subhaplotypes of the full NPY1R haplotypes shown in Table 4. In particular, a specific NPY1R coding sequence isogene identified by a sub-haplotype of a haplotype in Table 5 comprises each oligo in the set of oligos shown in Table 6 for that NPY1R sub-haplotype.

Table 6. Oligonucleotides Comprising Coding Sequence Sub-Haplotypes for the NPY1R Gene

Coding Sequence Sub-Haplotype Number	PS3	
	SEQ ID NO:26	SEQ ID NO:27
1c	X	
2c	X	
3c		X
4c	X	

Table 7 below shows the percent of chromosomes characterized by a given NPY1R 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 NPY1R haplotype pair is shown in Table 8. In Tables 7 and 8, 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 7 and 8 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and NA = Native American.

Table 7. Frequency of Observed NPY1R Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	7815183	1.22	2.38	0.0	0.0	2.78	0.0
2	7815180	3.05	2.38	2.5	5.0	2.78	0.0
3	7815185	1.22	0.0	5.0	0.0	0.0	0.0
4	7815172	94.51	95.24	92.5	95.0	94.44	100.0

Table 8. Frequency of Observed NPY1R Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
4	4	89.02	90.48	85.0	90.0	88.89	100.0
4	1	2.44	4.76	0.0	0.0	5.56	0.0
4	3	2.44	0.0	10.0	0.0	0.0	0.0
4	2	6.1	4.76	5.0	10.0	5.56	0.0

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

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

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense. All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the neuropeptide Y receptor Y1 (NPY1R) gene of an individual which comprises determining whether the individual has one of the NPY1R haplotypes shown in Table 4 or one of the haplotype pairs shown in Table 3, wherein each of the haplotypes in Table 4 comprises the corresponding set of SEQ ID NOS shown in Table 5.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of polymorphic sites (PS) PS1-3 on at least one copy of the individual's NPY1R gene, wherein PS1-3 have the location and alternative alleles shown in SEQ ID NO:21 (Fig. 1).

3. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-3 on both copies of the individual's NPY1R gene.

4. A method for genotyping the neuropeptide Y receptor Y1 (NPY1R) gene of an individual, comprising determining for the two copies of the NPY1R 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 and PS3, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:21 (Fig. 1).

5. The method of claim 4, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid mixture comprising both copies of the NPY1R 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.

6. The method of claim 4, which comprises determining for the two copies of the NPY1R gene present in the individual the identity of the nucleotide pair at each of PS1-3.

7. A method for haplotyping the neuropeptide Y receptor Y1 (NPY1R) gene of an individual which comprises determining, for one copy of the NPY1R 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 and PS3, wherein the two or more PS have the location and alternative alleles shown in SEQ ID NO:21 (Fig.1).

8. The method of claim 7, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid sample containing only one of the two copies of

the NPY1R gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid molecule 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.

9. A method for predicting a haplotype pair for the neuropeptide Y receptor Y1 (NPY1R) gene of an individual comprising:

(a) identifying a NPY1R 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 and PS3, having the location and alternative alleles shown in SEQ ID NO:21 (Fig.1);

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the data in Table 3; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

10. The method of claim 9, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-3, which have the location and alternative alleles shown in SEQ ID NO:21 (Fig. 1).

11. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the neuropeptide Y receptor Y1 (NPY1R) 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 1-4 shown in Table 4, wherein each of the haplotypes in Table 4 comprises the corresponding set of SEQ ID NOS shown in Table 5, and the haplotype pair is selected from the haplotype pairs shown in Table 3, 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.

12. The method of claim 11, wherein the trait is a clinical response to a drug targeting NPY1R.

13. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the neuropeptide Y receptor Y1 (NPY1R) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2 and PS3, having the location and alternative alleles shown in SEQ ID NO:21 (Fig.1).

14. The composition of claim 13, wherein the genotyping oligonucleotide is an allele-specific

oligonucleotide that specifically hybridizes to an allele of the NPY1R gene at a region containing the polymorphic site.

15. The composition of claim 14, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-6, the complements of SEQ ID NOS:4-6, and SEQ ID NOS:7-12.

16. The composition of claim 13, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.

17. The composition of claim 16, wherein the primer extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:13-18.

18. A kit for genotyping the NPY1R gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS2 and PS3, having the location and alternative alleles shown in SEQ ID NO:21 (Fig.1).

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

(a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for the neuropeptide Y receptor Y1 (NPY1R) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises a NPY1R isogene defined by a haplotype selected from the group consisting of haplotypes 1-4 in Table 4, wherein each of the haplotypes in Table 4 comprises the corresponding set of SEQ ID NOS shown in Table 5, and the fragment comprises at least one novel nucleotide at a polymorphic site (PS) selected from the group consisting of guanine at PS1, adenine at PS2 and cytosine at PS3; and

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

20. The isolated polynucleotide of claim 19, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

21. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 19, wherein the organism expresses a NPY1R protein encoded by the first nucleotide sequence.

22. The recombinant organism of claim 21, which is a nonhuman transgenic animal.

23. The isolated polynucleotide of claim 19, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the NPY1R gene, the fragment comprising one or more polymorphic sites (PS) with the polymorphic allele selected from the group consisting of guanine at PS1, adenine at PS2 and cytosine at PS3.

24. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the NPY1R coding sequence or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises a set of

oligonucleotides selected from the group consisting of the corresponding sets of SEQ ID NOS shown in Table 6.

25. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 24, wherein the organism expresses a neuropeptide Y receptor Y1 (NPY1R) protein encoded by the polymorphic variant sequence.

26. The recombinant organism of claim 25, which is a nonhuman transgenic animal.

27. A computer system for storing and analyzing polymorphism data for the neuropeptide Y receptor Y1 gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 3 and the haplotypes shown in Table 4.

28. A genome anthology for the neuropeptide Y receptor Y1 (NPY1R) gene which comprises NPY1R isogenes defined by any one of haplotypes 1-4 shown in Table 4, wherein each of the haplotypes in Table 4 comprises the corresponding set of SEQ ID NOS shown in Table 5.

POLYMORPHISMS IN THE NPY1R GENE

```
AAATATTCAG CTCATGGTTC TCCCCAAAAA AACTAAAAAG CAGCTAAGCG
CTGGGAACAA ATCTGACTTA TTGCATTTTC TCAGTGGGCC AAAGAAAGGA      100
GGGCCGATTG ACTGCTTTGA CTTTTTAAAG GTCTTCTCTT TGTTCACTTA
TAAAGTGAGG AAAACAAATT CTCGGCACTG GCGTGAGAGT TGAGCGTCAC      200
AAAAGAAAGC AAAAGAAAAT ATTAGTGCCA TTATTGTGGC GAATTTCATG
TTTCCCAGCG AGCCCTTTGA TTCCTGGTTT GGGCTGGCGC TCGAGCTCTC      300
CAGCCGGGTA TGACTTCGGC CACAAGATGG CACTGACCTG CAAACAAAGA
AAAGCACAGT GGCACCGACT TTTTCAAGCC TCGGGAAACT GCCCTGCCTT      400
CCCCGGAGTC GAGGACTGTG GGGATTAGGG CTTCCTTTCC CCTGCGCGGG
AGGTCTGTGT CGAATAATGT GTGGCTTCTG TTGGATTGCT TTTCTTTCCA      500
AAATTCCTAG GCAATGCTTC CCCGAGGTGT GCACCTTTGT GAGGTGTTTG
   G
TGGGGTTGGG GGAGCTTCAG GCGCTACTCG CGGGACGACG TCACGTGATC      600
CGGGATGAGG TGGAGTTCGG CTTTAAGGAG GCGTCTCTTC CTAGCTTCAT
CAATCTTTAG GATCTGAGCA GGAGAAATAC CAGCGGATCT TCCCCACTCT      700
                                    A
GCTCCCTTCC ATTCCCACCC TTCCTTCTTT AATAAGCAGG AGCGAAAAAG
ACAAATTCCA AAGAGGGTAA GTTGCGAGTT TATGCCTTTC CAGAGACTTC      800
TGCGAAATCT CTCATTGACA AGGTGAAGGA TGAGAGGGGA TTTTCACTTT
AAAGAGTTCT GTGAGTCAGA AGTCATTTTG ACTGCCCTCA ATAAAATTAG      900
TAATGCAATT GGTCATTTTC TCTTTACAGA TTGTTCAGTT CAAGGGAATG
AAGAATTCAG AATAATTTTG GTAAATGGAT TCCAATATCG GGAATAAGAA     1000
TAAGCTGAAC AGTTGACCTG CTTTGAAGAA ACATACTGTC CATTTGTCTA
AAATAATCTA TAACAACCAA ACCAATCAAA ATGAATTCAA CATTATTTTC     1100
      [exon 2: 1081..
CCAGGTTGAA AATCATTCAG TCCACTCTAA TTTCTCAGAG AAGAATGCCC
AGCTTCTGGC TTTTGAAAAT GATGATTGTC ATCTGCCCTT GGCCATGATA     1200
TTTACCTTAG CTCTTGCTTA TGGAGCTGTG ATCATTCTTG GTGTCTCTGG
AAACCTGGCC TTGATCATAA TCATCTTGAA ACAAAAGGAG ATGAGAAATG     1300
TTACCAACAT CCTGATTGTG AACCTTTCCT TCTCAGACTT GCTTGTTGCC
ATCATGTGTC TCCCCTTTAC ATTTGTCTAC ACATTAATGG ACCACTGGGT     1400
CTTTGGTGAG GCGATGTGTA AGTTGAATCC TTTTGTGCAA TGTGTTTCAA
TCACTGTGTC CATTTTCTCT CTGGTTCTCA TTGCTGTGGA ACGACATCAG     1500
CTGATAATCA ACCCTCGAGG GTGGAGACCA AATAATAGAC ATGCTTATGT
AGGTATTGCT GTGATTTGGG TCCTTGCTGT GGCTTCTTCT TTGCCTTTCC     1600
                                    C
TGATCTACCA AGTAATGACT GATGAGCCGT TCCAAAATGT AACACTTGAT
GCGTACAAAG ACAAATACGT GTGCTTTGAT CAATTTCCAT CGGACTCTCA     1700
TAGGTTGTCT TATACCACTC TCCTCTTGGT GCTGCAGTAT TTTGGTCCAC
TTTGTTTTAT ATTTATTTGC TACTTCAAGG TAAGAAAACT TTTTTTCTAT     1800
                ..1779]
CATTTCCATT TTTACCTTCT TTACACAGAA TTCCTCATCA AATGAGTGTT
TCTATTTAAA CTTTTTTCTT CCATAGATAT ATATACGCCT AAAAAGGAGA     1900
      [exon 3: 1877..
AACAACATGA TGGACAAGAT GAGAGACAAT AAGTACAGGT CCAGTGAAAC
CAAAAGAATC AATATCATGC TGCTCTCCAT TGTGGTAGCA TTTGCAGTCT     2000
GCTGGCTCCC TCTTACCATC TTTAACACTG TGTTTGATTG GAATCATCAG
ATCATTGCTA CCTGCAACCA CAATCTGTTA TTCCTGCTCT GCCACCTCAC     2100
AGCAATGATA TCCACTTGTG TCAACCCCAT ATTTTATGGG TTCCTGAACA
AAAACTTCCA GAGAGACTTG CAGTTCTTCT TCAACTTTTG TGATTTCCGG     2200
```

FIGURE 1A

```
TCTCGGGATG ATGATTATGA AACAATAGCC ATGTCCACGA TGCACACAGA
TGTTTCCAAA ACTTCTTTGA AGCAAGCAAG CCCAGTCGCA TTTAAAAAAA      2300
TCAACAACAA TGATGATAAT GAAAAAATCT GAAACTACTT ATAGCCTATG
                ..2332]
GTCCCGGATG ACATCTGTTT AAAAACAAGC ACAACCTGCA ACATACTTTG      2400
ATTACCTGTT CTCCCAAGGA ATGGGGTTGA AATCATTTGA AAATGACTAA
GATTTTCTTG TCTTGCTTTT TACTGCTTTT GTTGTAGTTG TCATAATTAC      2500
ATTTGGAACA AAAGGTGTGG GCTTTGGGGT CTTCTGGAAA TAGTTTTGAC
CAGACATCTT TGAAGTGCTT TTTGTGAATT TATGCATATA ATATAAAGAC      2600
TTTTATACTG TACTTATTGG AATGAAATTT CTTTAAAGTA TTACGATGCG
CTGACTTCAG AAGTACCTGC CATCCAATAC GGTCATTAGA TTGGGTCATC      2700
TTGATTAGAT TAGATTAGAT TAGATTGTCA ACAGATTGGG CCATCCTTAC
TTTATGATAG GCATCATTTT AGTGTGTTAC AATAGTAACA GTATGCAAAA      2800
GCAGCATTCA GGAGCCGAAA GATAGTCTTG AAGTCATTCA GAAGTGGTTT
GAGGTTTCTG TTTTTTGGTG GTTTTTGTTT GTTTTTTTTT TTTTTCACCT      2900
TAAGGGAGGC TTTCATTTCC TCCCGACTGA TTGTCACTTA AATCAAAATT
TAAAAATGAA TAAAAAGACA TACTTCTCAG CTGCAAATAT TATGGAGAAT      3000
TGGGCACCCA CAGGAATGAA GAGAGAAAGC AGCTCCCCAA CTTCAAAACC
ATTTTGGTAC CTGACAACAA GAGCATTTTA GAGTAATTAA TTTAATAAAG      3100
TAAATTAGTA TTGCTGCAAA TAGCTAAATT ATATTTATTT GAATTGATGG
TCAAGAGATT TTCCATTTTT TTTACAGACT GTTCAGTGTT TGTCAAGCTT      3200
CTGGTCTAAT ATGTACTCGA AAGACTTTCC GCTTACAATT TGTAGAAACA
CAAATATCGT TTTCCATACA GCAGTGCCTA TATAGTGACT GATTTTAACT      3300
TTCAATGTCC ATCTTTCAAA GGAAGTAACA CCAAGGTACA ATGTTAAAGG
AATATTCACT TTACCTAGCA GGGAAAAATA CACAAAAACT GCAGATACTT      3400
CATATAGCCC ATTTTAACTT GTATAAACTG TGTGACTTGT GGCGTCTTAT
AAATAATGCA CTGTAAAGAT TACTGAATAG TTGTGTCATG TTAATGTGCC      3500
TAATTTCATG TATCTTGTAA TCATGATTGA GCCTCAGAAT CATTTGGAGA
AACTATATTT TAAAGAACAA GACATACTTC AATGTATTAT ACAGATAAAG      3600
TATTACATGT GTTTGATTTT AAAAGGGCGG ACATTTTATT AAAATCAATA
TTGTTTTTGC TTTTTCTGAG GAGTCTCTTT CAGTTTCATT TTTTCTCATC      3700
CCATGACTTC CCTCCGATGG T                                     3721
```

FIGURE 1B

POLYMORPHISMS IN THE CODING SEQUENCE OF NPY1R

```
ATGAATTCAA CATTATTTTC CCAGGTTGAA AATCATTCAG TCCACTCTAA
TTTCTCAGAG AAGAATGCCC AGCTTCTGGC TTTTGAAAAT GATGATTGTC      100
ATCTGCCCTT GGCCATGATA TTTACCTTAG CTCTTGCTTA TGGAGCTGTG
ATCATTCTTG GTGTCTCTGG AAACCTGGCC TTGATCATAA TCATCTTGAA      200
ACAAAAGGAG ATGAGAAATG TTACCAACAT CCTGATTGTG AACCTTTCCT
TCTCAGACTT GCTTGTTGCC ATCATGTGTC TCCCCTTTAC ATTTGTCTAC      300
ACATTAATGG ACCACTGGGT CTTTGGTGAG GCGATGTGTA AGTTGAATCC
TTTTGTGCAA TGTGTTTCAA TCACTGTGTC CATTTTCTCT CTGGTTCTCA      400
TTGCTGTGGA ACGACATCAG CTGATAATCA ACCCTCGAGG GTGGAGACCA
AATAATAGAC ATGCTTATGT AGGTATTGCT GTGATTTGGG TCCTTGCTGT      500
GGCTTCTTCT TTGCCTTTCC TGATCTACCA AGTAATGACT GATGAGCCGT
      C
TCCAAAATGT AACACTTGAT GCGTACAAAG ACAAATACGT GTGCTTTGAT      600
CAATTTCCAT CGGACTCTCA TAGGTTGTCT TATACCACTC TCCTCTTGGT
GCTGCAGTAT TTTGGTCCAC TTTGTTTTAT ATTTATTTGC TACTTCAAGA      700
TATATATACG CCTAAAAAGG AGAAACAACA TGATGGACAA GATGAGAGAC
AATAAGTACA GGTCCAGTGA AACCAAAAGA ATCAATATCA TGCTGCTCTC      800
CATTGTGGTA GCATTTGCAG TCTGCTGGCT CCCTCTTACC ATCTTTAACA
CTGTGTTTGA TTGGAATCAT CAGATCATTG CTACCTGCAA CCACAATCTG      900
TTATTCCTGC TCTGCCACCT CACAGCAATG ATATCCACTT GTGTCAACCC
CATATTTTAT GGGTTCCTGA ACAAAAACTT CCAGAGAGAC TTGCAGTTCT     1000
TCTTCAACTT TTGTGATTTC CGGTCTCGGG ATGATGATTA TGAAACAATA
GCCATGTCCA CGATGCACAC AGATGTTTCC AAAACTTCTT TGAAGCAAGC     1100
AAGCCCAGTC GCATTTAAAA AAATCAACAA CAATGATGAT AATGAAAAAA
TCTGA                                                      1155
```

FIGURE 2

AMINO ACID SEQUENCE OF THE NPY1R PROTEIN

MNSTLFSQVE NHSVHSNFSE KNAQLLAFEN DDCHLPLAMI FTLALAYGAV
IILGVSGNLA LIIIILKQKE MRNVTNILIV NLSFSDLLVA IMCLPFTFVY 100
TLMDHWVFGE AMCKLNPFVQ CVSITVSIFS LVLIAVERHQ LIINPRGWRP
NNRHAYVGIA VIWVLAVASS LPFLIYQVMT DEPFQNVTLD AYKDKYVCFD 200
QFPSDSHRLS YTTLLLVLQY FGPLCFIFIC YFKIYIRLKR RNNMMDKMRD
NKYRSSETKR INIMLLSIVV AFAVCWLPLT IFNTVFDWNH QIIATCNHNL 300
LFLLCHLTAM ISTCVNPIFY GFLNKNFQRD LQFFFNFCDF RSRDDDYETI
AMSTMHTDVS KTSLKQASPV AFKKINNNDD NEKI 384

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Choi, Julie Y.
Kliem, Stefanie E.
Koshy, Beena
Lee, Helen H.

<120> Haplotypes of the NPY1R Gene

<130> MWH-0338PCT NPY1R

<140> To be assigned
<141> 2001-05-07

<150> 60/301,950
<151> 2000-05-05

<160> 27

<170> PatentIn Ver. 2.1

<210> 1
<211> 3721
<212> DNA
<213> Homo sapien

<400> 1
aaatattcag ctcatggttc tccccaaaaa aactaaaaag cagctaagcg ctgggaacaa 60
atctgactta ttgcattttc tcagtgggcc aaagaaagga gggccgattg actgctttga 120
ctttttaaag gtcttctctt tgttcactta taaagtgagg aaaacaaatt ctcggcactg 180
gcgtgagagt tgagcgtcac aaaagaaagc aaaagaaaat attagtgcca ttattgtggc 240
gaatttcatg tttcccagcg agccctttga ttcctggttt gggctggcgc tcgagctctc 300
cagccgggta tgacttcggc cacaagatgg cactgacctg caaacaaaga aaagcacagt 360
ggcaccgact ttttcaagcc tcgggaaact gccctgcctt ccccggagtc gaggactgtg 420
gggattaggg cttcctttcc cctgcgcggg aggtctgtgt cgaataatgt gtggcttctg 480
ttggattgct tttctttcca aaattcctag gcaatgcttc cccgaggtgt gcacctttgt 540
gaggtgtttg tggggttggg ggagcttcag gcgctactcg cgggacgacg tcacgtgatc 600
cgggatgagg tggagttcgg ctttaaggag gcgtctcttc ctagcttcat caatctttag 660
gatctgagca ggagaaatac cagcggatct tccccactct gctcccttcc attcccaccc 720
ttccttcttt aataagcagg agcgaaaaag acaaattcca aagagggtaa gttgcgagtt 780
tatgcctttc cagagacttc tgcgaaatct ctcattgaca aggtgaagga tgagagggga 840
ttttcacttt aaagagttct gtgagtcaga agtcattttg actgccctca ataaaattag 900
taatgcaatt ggtcattttc tctttacaga ttgttcagtt caagggaatg aagaattcag 960
aataattttg gtaaatggat tccaatatcg ggaataagaa taagctgaac agttgacctg 1020
ctttgaagaa acatactgtc catttgtcta aaataatcta taacaaccaa accaatcaaa 1080
atgaattcaa cattattttc ccaggttgaa aatcattcag tccactctaa tttctcagag 1140
aagaatgccc agcttctggc ttttgaaaat gatgattgtc atctgccctt ggccatgata 1200
tttaccttag ctcttgctta tggagctgtg atcattcttg gtgtctctgg aaacctggcc 1260
ttgatcataa tcatcttgaa acaaaaggag atgagaaatg ttaccaacat cctgattgtg 1320
aacctttcct tctcagactt gcttgttgcc atcatgtgtc tccccttcac atttgtctac 1380
acattaatgg accactgggt ctttggtgag gcgatgtgta agttgaatcc ttttgtgcaa 1440
tgtgtttcaa tcactgtgtc cattttctct ctggttctca ttgctgtgga acgacatcag 1500
ctgataatca accctcgagg gtggagacca aataatagac atgcttatgt aggtattgct 1560
gtgatttggg tccttgctgt ggcttcttct ttgcctttcc tgatctacca agtaatgact 1620
gatgagccgt tccaaaatgt aacacttgat gcgtacaaag acaaatacgt gtgctttgat 1680
caatttccat cggactctca taggttgtct tataccactc tcctcttggt gctgcagtat 1740
tttggtccac tttgttttat atttatttgc tacttcaagg taagaaaact ttttttctat 1800
catttccatt tttaccttct ttacacagaa ttcctcatca aatgagtgtt tctatttaaa 1860
ctttttttctt ccatagatat atatacgcct aaaaaggaga aacaacatga tggacaagat 1920
gagagacaat aagtacaggt ccagtgaaac caaaagaatc aatatcatgc tgctctccat 1980
tgtggtagca tttgcagtct gctggctccc tcttaccatc tttaacactg tgtttgattg 2040
gaatcatcag atcattgcta cctgcaacca caatctgtta ttcctgctct gccacctcac 2100
agcaatgata tccacttgtg tcaaccccat attttatggg ttcctgaaca aaaacttcca 2160
gagagacttg cagttcttct tcaacttttg tgatttccgg tctcgggatg atgattatga 2220
aacaatagcc atgtccacga tgcacacaga tgtttccaaa acttctttga agcaagcaag 2280

cccagtcgca tttaaaaaaa tcaacaacaa tgatgataat gaaaaaatct gaaactactt 2340
atagcctatg gtcccggatg acatctgttt aaaaacaagc acaacctgca acatactttg 2400
attacctgtt ctcccaagga atggggttga aatcatttga aaatgactaa gattttcttg 2460
tcttgctttt tactgctttt gttgtagttg tcataattac atttggaaca aaaggtgtgg 2520
gctttggggt cttctggaaa tagttttgac cagacatctt tgaagtgctt tttgtgaatt 2580
tatgcatata atataaagac ttttatactg tacttattgg aatgaaattt ctttaaagta 2640
ttacgatgcg ctgacttcag aagtacctgc catccaatac ggtcattaga ttgggtcatc 2700
ttgattagat tagattagat tagattgtca acagattggg ccatccttac tttatgatag 2760
gcatcatttt agtgtgttac aatagtaaca gtatgcaaaa gcagcattca ggagccgaaa 2820
gatagtcttg aagtcattca gaagtggttt gaggtttctg ttttttggtg gtttttgttt 2880
gttttttttt tttttcacct taagggaggc tttcatttcc tcccgactga ttgtcactta 2940
aatcaaaatt taaaaatgaa taaaaagaca tacttctcag ctgcaaatat tatggagaat 3000
tgggcaccca caggaatgaa gagagaaagc agctccccaa cttcaaaacc attttggtac 3060
ctgacaacaa gagcatttta gagtaattaa tttaataaag taaattagta ttgctgcaaa 3120
tagctaaatt atatttattt gaattgatgg tcaagagatt ttccattttt tttacagact 3180
gttcagtgtt tgtcaagctt ctggtctaat atgtactcga aagactttcc gcttacaatt 3240
tgtagaaaca caaatatcgt tttccataca gcagtgccta tatagtgact gattttaact 3300
ttcaatgtcc atctttcaaa ggaagtaaca ccaaggtaca atgttaaagg aatattcact 3360
ttacctagca gggaaaaata cacaaaaact gcagatactt catatagccc attttaactt 3420
gtataaactg tgtgacttgt ggcgtcttat aaataatgca ctgtaaagat tactgaatag 3480
ttgtgtcatg ttaatgtgcc taatttcatg tatcttgtaa tcatgattga gcctcagaat 3540
catttggaga aactatattt taaagaacaa gacatacttc aatgtattat acagataaag 3600
tattacatgt gtttgatttt aaaagggcgg acattttatt aaaatcaata ttgtttttgc 3660
tttttctgag gagtctcttt cagtttcatt ttttctcatc ccatgacttc cctccgatgg 3720
t 3721

<210> 2
<211> 1155
<212> DNA
<213> Homo sapien

<400> 2
atgaattcaa cattattttc ccaggttgaa aatcattcag tccactctaa tttctcagag 60
aagaatgccc agcttctggc ttttgaaaat gatgattgtc atctgccctt ggccatgata 120
tttaccttag ctcttgctta tggagctgtg atcattcttg gtgtctctgg aaacctggcc 180
ttgatcataa tcatcttgaa acaaaaggag atgagaaatg ttaccaacat cctgattgtg 240
aacctttcct tctcagactt gcttgttgcc atcatgtgtc tccccttttac atttgtctac 300
acattaatgg accactgggt ctttggtgag gcgatgtgta agttgaatcc ttttgtgcaa 360
tgtgtttcaa tcactgtgtc cattttctct ctggttctca ttgctgtgga acgacatcag 420
ctgataatca accctcgagg gtggagacca aataatagac atgcttatgt aggtattgct 480
gtgatttggg tccttgctgt ggcttcttct ttgcctttcc tgatctacca agtaatgact 540
gatgagccgt tccaaaatgt aacacttgat gcgtacaaag acaaatacgt gtgctttgat 600
caatttccat cggactctca taggttgtct tataccactc tcctcttggt gctgcagtat 660
tttggtccac tttgttttat atttatttgc tacttcaaga tatatatacg cctaaaaagg 720
agaaacaaca tgatggacaa gatgagagac aataagtaca ggtccagtga aaccaaaaga 780
atcaatatca tgctgctctc cattgtggta gcatttgcag tctgctggct ccctcttacc 840
atctttaaca ctgtgtttga ttggaatcat cagatcattg ctacctgcaa ccacaatctg 900
ttattcctgc tctgccacct cacagcaatg atatccactt gtgtcaaccc catattttat 960
gggttcctga acaaaaactt ccagagagac ttgcagttct tcttcaactt ttgtgatttc 1020
cggtctcggg atgatgatta tgaaacaata gccatgtcca cgatgcacac agatgtttcc 1080
aaaacttctt tgaagcaagc aagcccagtc gcatttaaaa aaatcaacaa caatgatgat 1140
aatgaaaaaa tctga 1155

<210> 3
<211> 384
<212> PRT
<213> Homo sapien

<400> 3
Met Asn Ser Thr Leu Phe Ser Gln Val Glu Asn His Ser Val His Ser
1 5 10 15

Asn Phe Ser Glu Lys Asn Ala Gln Leu Leu Ala Phe Glu Asn Asp Asp
20 25 30

```
Cys His Leu Pro Leu Ala Met Ile Phe Thr Leu Ala Leu Ala Tyr Gly
         35                  40                  45

Ala Val Ile Ile Leu Gly Val Ser Gly Asn Leu Ala Leu Ile Ile Ile
     50                  55                  60

Ile Leu Lys Gln Lys Glu Met Arg Asn Val Thr Asn Ile Leu Ile Val
 65                  70                  75                  80

Asn Leu Ser Phe Ser Asp Leu Leu Val Ala Ile Met Cys Leu Pro Phe
                 85                  90                  95

Thr Phe Val Tyr Thr Leu Met Asp His Trp Val Phe Gly Glu Ala Met
            100                 105                 110

Cys Lys Leu Asn Pro Phe Val Gln Cys Val Ser Ile Thr Val Ser Ile
        115                 120                 125

Phe Ser Leu Val Leu Ile Ala Val Glu Arg His Gln Leu Ile Ile Asn
    130                 135                 140

Pro Arg Gly Trp Arg Pro Asn Asn Arg His Ala Tyr Val Gly Ile Ala
145                 150                 155                 160

Val Ile Trp Val Leu Ala Val Ala Ser Ser Leu Pro Phe Leu Ile Tyr
                165                 170                 175

Gln Val Met Thr Asp Glu Pro Phe Gln Asn Val Thr Leu Asp Ala Tyr
            180                 185                 190

Lys Asp Lys Tyr Val Cys Phe Asp Gln Phe Pro Ser Asp Ser His Arg
        195                 200                 205

Leu Ser Tyr Thr Thr Leu Leu Leu Val Leu Gln Tyr Phe Gly Pro Leu
    210                 215                 220

Cys Phe Ile Phe Ile Cys Tyr Phe Lys Ile Tyr Ile Arg Leu Lys Arg
225                 230                 235                 240

Arg Asn Asn Met Met Asp Lys Met Arg Asp Asn Lys Tyr Arg Ser Ser
                245                 250                 255

Glu Thr Lys Arg Ile Asn Ile Met Leu Leu Ser Ile Val Val Ala Phe
            260                 265                 270

Ala Val Cys Trp Leu Pro Leu Thr Ile Phe Asn Thr Val Phe Asp Trp
        275                 280                 285

Asn His Gln Ile Ile Ala Thr Cys Asn His Asn Leu Leu Phe Leu Leu
    290                 295                 300

Cys His Leu Thr Ala Met Ile Ser Thr Cys Val Asn Pro Ile Phe Tyr
305                 310                 315                 320

Gly Phe Leu Asn Lys Asn Phe Gln Arg Asp Leu Gln Phe Phe Phe Asn
                325                 330                 335

Phe Cys Asp Phe Arg Ser Arg Asp Asp Asp Tyr Glu Thr Ile Ala Met
            340                 345                 350

Ser Thr Met His Thr Asp Val Ser Lys Thr Ser Leu Lys Gln Ala Ser
        355                 360                 365

Pro Val Ala Phe Lys Lys Ile Asn Asn Asn Asp Asp Asn Glu Lys Ile
    370                 375                 380
```

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

<400> 4
tccaaaaktc ctagg 15

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

<400> 5
taccagcrga tcttc 15

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

<400> 6
tggcttcytc tttgc 15

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

<400> 7
tttctttcca aaakt 15

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

<400> 8
gcattgccta ggamt 15

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

<400> 9
gagaaatacc agcrg 15

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

<400> 10
agtggggaag atcyg 15

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

<400> 11
ttgctgtggc ttcyt 15

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

<400> 12
ggaaaggcaa agarg 15

<210> 13
<211> 10
<212> DNA
<213> Homo sapien

<400> 13
ctttccaaaa 10

<210> 14
<211> 10
<212> DNA
<213> Homo sapien

<400> 14
ttgcctagga 10

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

<400> 15
aaataccagc 10

<210> 16
<211> 10
<212> DNA
<213> Homo sapien

<400> 16
ggggaagatc 10

<210> 17
<211> 10
<212> DNA
<213> Homo sapien

<400> 17
ctgtggcttc 10

<210> 18

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

<400> 18
aaggcaaaga 10

<210> 19
<211> 18
<212> DNA
<213> Homo sapien

<400> 19
tgtaaaacga cggccagt 18

<210> 20
<211> 19
<212> DNA
<213> Homo sapien

<400> 20
aggaaacagc tatgaccat 19

<210> 21
<211> 3721
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (504)
<223> PS1: Polymorphic base T or G

<220>
<221> allele
<222> (685)
<223> PS2: Polymorphic base G or A

<220>
<221> allele
<222> (1587)
<223> PS3: Polymorphic base T or C

<400> 21
aaatattcag ctcatggttc tccccaaaaa aactaaaaag cagctaagcg ctgggaacaa 60
atctgactta ttgcattttc tcagtgggcc aaagaaagga gggccgattg actgctttga 120
ctttttaaag gtcttctctt tgttcactta taaagtgagg aaaacaaatt ctcggcactg 180
gcgtgagagt tgagcgtcac aaaagaaagc aaaagaaaat attagtgcca ttattgtggc 240
gaatttcatg tttcccagcg agccctttga ttcctggttt gggctggcgc tcgagctctc 300
cagccgggta tgacttcggc cacaagatgg cactgacctg caaacaaaga aaagcacagt 360
ggcaccgact ttttcaagcc tcgggaaact gccctgcctt ccccggagtc gaggactgtg 420
gggattaggg cttcctttcc cctgcgcggg aggtctgtgt cgaataatgt gtggcttctg 480
ttggattgct tttctttcca aaaktcctag gcaatgcttc cccgaggtgt gcacctttgt 540
gaggtgtttg tggggttggg ggagcttcag gcgctactcg cgggacgacg tcacgtgatc 600
cgggatgagg tggagttcgg ctttaaggag gcgtctcttc ctagcttcat caatctttag 660
gatctgagca ggagaaatac cagcrgatct tccccactct gctcccttcc attcccaccc 720
ttccttcttt aataagcagg agcgaaaaag acaaattcca aagagggtaa gttgcgagtt 780
tatgcctttc cagagacttc tgcgaaatct ctcattgaca aggtgaagga tgagagggga 840
ttttcacttt aaagagttct gtgagtcaga agtcattttg actgccctca ataaaattag 900
taatgcaatt ggtcattttc tctttacaga ttgttcagtt caagggaatg aagaattcag 960
aataattttg gtaaatggat tccaatatcg ggaataagaa taagctgaac agttgacctg 1020
ctttgaagaa acatactgtc catttgtcta aaataatcta taacaaccaa accaatcaaa 1080
atgaattcaa cattattttc ccaggttgaa aatcattcag tccactctaa tttctcagag 1140
aagaatgccc agcttctggc ttttgaaaat gatgattgtc atctgccctt ggccatgata 1200

tttaccttag ctcttgctta tggagctgtg atcattcttg gtgtctctgg aaacctggcc 1260
ttgatcataa tcatcttgaa acaaaaggag atgagaaatg ttaccaacat cctgattgtg 1320
aacctttcct tctcagactt gcttgttgcc atcatgtgtc tccccctttac atttgtctac 1380
acattaatgg accactgggt ctttggtgag gcgatgtgta agttgaatcc ttttgtgcaa 1440
tgtgtttcaa tcactgtgtc cattttctct ctggttctca ttgctgtgga acgacatcag 1500
ctgataatca accctcgagg gtggagacca aataatagac atgcttatgt aggtattgct 1560
gtgatttggg tccttgctgt ggcttcytct ttgcctttcc tgatctacca agtaatgact 1620
gatgagccgt tccaaaatgt aacacttgat gcgtacaaag acaaatacgt gtgctttgat 1680
caatttccat cggactctca taggttgtct tataccactc tcctcttggt gctgcagtat 1740
tttggtccac tttgttttat atttatttgc tacttcaagg taagaaaact ttttttctat 1800
catttccatt tttaccttct ttacacagaa ttcctcatca aatgagtgtt tctatttaaa 1860
ctttttttctt ccatagatat atatacgcct aaaaaggaga aacaacatga tggacaagat 1920
gagagacaat aagtacaggt ccagtgaaac caaaagaatc aatatcatgc tgctctccat 1980
tgtggtagca tttgcagtct gctggctccc tcttaccatc tttaacactg tgtttgattg 2040
gaatcatcag atcattgcta cctgcaacca caatctgtta ttcctgctct gccacctcac 2100
agcaatgata tccacttgtg tcaaccccat attttatggg ttcctgaaca aaaacttcca 2160
gagagacttg cagttcttct tcaacttttg tgatttccgg tctcgggatg atgattatga 2220
aacaatagcc atgtccacga tgcacacaga tgtttccaaa acttctttga agcaagcaag 2280
cccagtcgca tttaaaaaaa tcaacaacaa tgatgataat gaaaaaatct gaaactactt 2340
atagcctatg gtcccggatg acatctgttt aaaaacaagc acaacctgca acatactttg 2400
attacctgtt ctcccaagga atggggttga aatcatttga aaatgactaa gattttcttg 2460
tcttgctttt tactgctttt gttgtagttg tcataattac atttggaaca aaaggtgtgg 2520
gctttggggt cttctggaaa tagttttgac cagacatctt tgaagtgctt tttgtgaatt 2580
tatgcatata atataaagac ttttatactg tacttattgg aatgaaattt ctttaaagta 2640
ttacgatgcg ctgacttcag aagtacctgc catccaatac ggtcattaga ttgggtcatc 2700
ttgattagat tagattagat tagattgtca acagattggg ccatccttac tttatgatag 2760
gcatcatttt agtgtgttac aatagtaaca gtatgcaaaa gcagcattca ggagccgaaa 2820
gatagtcttg aagtcattca gaagtggttt gaggtttctg ttttttggtg gtttttgttt 2880
gtttttttttt tttttcacct taagggaggc tttcatttcc tcccgactga ttgtcactta 2940
aatcaaaatt taaaaatgaa taaaaagaca tacttctcag ctgcaaatat tatggagaat 3000
tgggcaccca caggaatgaa gagagaaagc agctccccaa cttcaaaacc attttggtac 3060
ctgacaacaa gagcatttta gagtaattaa tttaataaag taaattagta ttgctgcaaa 3120
tagctaaatt atatttattt gaattgatgg tcaagagatt ttccattttt tttacagact 3180
gttcagtgtt tgtcaagctt ctggtctaat atgtactcga aagactttcc gcttacaatt 3240
tgtagaaaca caaatatcgt tttccataca gcagtgccta tatagtgact gattttaact 3300
ttcaatgtcc atctttcaaa ggaagtaaca ccaaggtaca atgttaaagg aatattcact 3360
ttacctagca gggaaaaata cacaaaaact gcagatactt catatagccc attttaactt 3420
gtataaactg tgtgacttgt ggcgtcttat aaataatgca ctgtaaagat tactgaatag 3480
ttgtgtcatg ttaatgtgcc taatttcatg tatcttgtaa tcatgattga gcctcagaat 3540
catttggaga aactatattt taaagaacaa gacatacttc aatgtattat acagataaag 3600
tattacatgt gtttgatttt aaaagggcgg acattttatt aaaatcaata ttgtttttgc 3660
tttttctgag gagtctcttt cagtttcatt ttttctcatc ccatgacttc cctccgatgg 3720
t 3721

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

<400> 22
tccaaaattc ctagg 15

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

<400> 23
tccaaaagtc ctagg 15

<210> 24
<211> 15

<212> DNA
<213> Homo sapien

<400> 24
taccagcgga tcttc 15

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

<400> 25
taccagcaga tcttc 15

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

<400> 26
tggcttcttc tttgc 15

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

<400> 27
tggcttcctc tttgc 15