Title: HAPLOTYPES OF THE MPL GENE

Abstract: Novel single nucleotide polymorphisms in the human myeloproliferative leukemia virus oncogene (MPL) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the MPL gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the MPL gene in an individual are also disclosed. Polynucleotides containing one or more of the MPL polymorphisms disclosed herein are also described.
HAPLOTYPES OF THE MPL GENE

RELATED APPLICATIONS
This application claims the benefit of U.S. Provisional Application Serial No. 60/197,839 filed April 14, 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 myeloproliferative leukemia virus oncogene (MPL) 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.


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 urgent 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 congenital amegakaryocytic thrombocytopenia (CAMT) is the myeloproliferative leukemia virus oncogene (MPL) gene or its encoded product. MPL is the receptor for thrombopoietin (TPO), which is a megakaryocyte growth and differentiation factor capable of raising platelet counts in animals (Columbyova et al., Cancer Res. 1995; 55:3509-3512). MPL is a member of the hematopoietic cytokine receptor family and has
previously been shown to rapidly activate one or more cytoplasmic tyrosine kinases after ligand binding (Sattler et al., *Exp. Hematol.* 1995; 23:1040-1048). MPL is expressed on hematopoietic stem cells, immature hematopoietic progenitor cells, megakaryocytes, and platelets (Miyakawa et al., *J. Biol. Chem.*, Vol. 275, Issue 41, 32214-32219, October 13, 2000). The genetic elimination of either TPO or c-Mpl in mice reduces the levels of all of these cell types by 70-90%, establishing that TPO-induced signaling plays a major physiological role in hematopoiesis in general and megakaryopoiesis in particular. In animal studies, a deficiency of the MPL gene results in amegakaryocytic thrombocytopenia (any bleeding disorder due to the lack of functional platelets), decreased numbers of hematopoietic progenitors, and increased concentrations of circulating TPO with no physical or developmental abnormalities (Ihara et al., *Proc. Natl. Acad. Sci. USA* Vol. 96, Issue 6, 3132-3136, March 16, 1999). CAMT is a rare human disorder presented in infancy and characterized by isolated thrombocytopenia and megakaryocytopenia with no physical anomalies. The cause of CAMT seems to be an intrinsic stem cell defect rather than an abnormality of the bone marrow microenvironment or an inhibitory factor in the patients' plasma (Ballmaier et al., *Blood*, 1 January 2001, Vol. 97, No. 1, pp. 139-146). Serum levels of TPO, the pivotal regulator of megakaryopoiesis but also an important factor for early multipotent hematopoietic progenitors, are highly elevated in patients affected by CAMT. To date, bone marrow transplantation is the only curative therapy for CAMT. Thus, since considerable similarities exist between human CAMT and murine c-mpl deficiency, MPL is a putative candidate gene for CAMT.

The myeloproliferative leukemia virus oncogene gene is located on chromosome 1p34 and contains 12 exons that encode a 635 amino acid protein. A reference sequence for the MPL gene comprises the non-contiguous sequences shown in Figure 1 (GenBank Accession No. U68159.1; SEQ ID NO:1), Figure 2 (GenBank Accession No. U68160.1; SEQ ID NO:2), Figure 3 (GenBank Accession No. U68161.1; SEQ ID NO:3), and Figure 4 (GenBank Accession No. U68162.1; SEQ ID NO:4). Reference sequences for the coding sequence (GenBankAccession No. NM_005373.1) and protein are shown in Figures 5 (SEQ ID NO:5) and 6 (SEQ ID NO:6), respectively.

Muraoka et al., (*Br J Haematol.* 1997;96:287-292) described a patient with CAMT who had a defective response to TPO in megakaryocyte colony formation, decreased numbers of erythroid and myeloid progenitors, and elevated TPO serum levels. Two polymorphisms in the MPL gene of this patient were identified and predicted to result in a complete absence of functional MPL protein (Ihara et al., *supra*). First, a polymorphism of cytosine or thymine at a position corresponding to nucleotide 2537 in Figure 1, which results in an amino acid variation of glutamine or a premature stop codon in the patient at a position corresponding to amino acid 186 in Figure 6. Second, a heterozygous deletion of thymine at position 1,058 in Figure 3 results in a frameshift, leading to a premature truncation. In a 2-year-old Italian boy with CAMT, Tonelli et al. (2000, *Hum. Genet.* 107: 225-233) found additional polymorphisms of cytosine or thymine at positions corresponding to nucleotides 3142 in exon 5 (Figure 1) and 903 in exon 12 (Figure 4). The mutation in exon 5 results in a variation of arginine or
cysteine at a position corresponding to amino acid 257 in Figure 6, which is located in the extracellular domain, 11 amino acids distant from the WSXWS motif conserved in the cytokine-receptor superfamily. The mutation in exon 12 results in a variation of proline or leucine at a position corresponding to nucleotide 635 in Figure 6. This is the last amino acid of the C-terminal intracellular domain, which is responsible for signal transduction. TPO plasma levels were greatly increased in the patient. Other polymorphisms have also been seen in the MPL gene of patients with CAMT.

Because of the potential for variation in the MPL gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the MPL 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 MPL 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 19 novel polymorphic sites in the MPL gene. These polymorphic sites (PS) correspond to the following nucleotide positions in the indicated GenBank Accession Number: 550 (PS1), 590 (PS2), 652 (PS3), 1238 (PS4), 1598 (PS5), 1771 (PS6), 2671 (PS7), 2805 (PS8), 2886 (PS9), 3160 (PS10) and 3595 (PS11) in U68159.1; 168 (PS12), 224 (PS13) and 807 (PS14) in U68160.1; 276 (PS15), 309 (PS16), 475 (PS17), 636 (PS18) and 925 (PS19) in U68162.1. The polymorphisms at these sites are cysteine or adenine at PS1, guanine or adenine at PS2, guanine or adenine at PS3, guanine or thymine at PS4, cysteine or guanine or adenine at PS5, guanine or adenine at PS6, adenine or guanine at PS7, cysteine or thymine at PS8, cysteine or thymine at PS9, adenine or cysteine at PS10, guanine or adenine at PS11, adenine or guanine at PS12, guanine or adenine at PS13, adenine or guanine at PS14, cysteine or thymine at PS15, cysteine or thymine at PS16, cysteine or adenine at PS17, cysteine or guanine at PS18 and cysteine or thymine at PS19. 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-PS19 in the MPL gene, which are shown below in Tables 5 and 4, respectively. PS5 is potentially a tri-allelic site. The reference GenBank Accession No. U68159.1 shows a cysteine at this position, while another GenBank Accession No. AL139289.6, shows a guanine at this position. Our sequencing results did not show the cysteine allele at this position, but rather a guanine or adenine allele. Thus, the three alleles at this site may be due to sequencing errors in GenBank Accession No. U68159.1 or may be a naturally occurring tri-allelic site. Each of these MPL haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the MPL gene that exists in the human population.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the MPL gene in an individual. The genotyping method comprises identifying the
nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19 in both copies of the MPL 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 MPL polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel MPL polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the MPL gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the MPL gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's MPL gene is defined by one of the MPL haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's MPL gene are defined by one of the MPL haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the MPL 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 MPL activity, e.g., congenital amegakaryocytic thrombocytopenia (CAMT).

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

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

In another embodiment, the invention provides a method for identifying an association between a trait and a MPL genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the MPL genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the MPL genotype, haplotype, or haplotype pair in a reference population. A higher frequency of the MPL genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the MPL 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 MPL haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for congenital amegakaryocytic thrombocytopenia (CAMT).

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 MPL gene or a fragment thereof. The reference sequence comprises SEQ ID NOS:1-4 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of adenine at PS1, adenine at PS2, adenine at PS3, thymine at PS4, guanine or adenine at PS5, adenine at PS6, guanine at PS7, thymine at PS8, thymine at PS9, cytosine at PS10, adenine at PS11, guanine at PS12, adenine at PS13, guanine at PS14, thymine at PS15, thymine at PS16, adenine at PS17, guanine at PS18 and thymine at PS19.
A particularly preferred polymorphic variant is an isogene of the MPL gene. A MPL isogene of the invention comprises cytosine or adenine at PS1, guanine or adenine at PS2, guanine or adenine at PS3, guanine or thymine at PS4, cytosine or guanine at PS5, guanine or adenine at PS6, adenine or guanine at PS7, cytosine or thymine at PS8, cytosine or thymine at PS9, adenine or cytosine at PS10, guanine or adenine at PS11, adenine or guanine at PS12, guanine or adenine at PS13, adenine or guanine at PS14, cytosine or thymine at PS15, cytosine or thymine at PS16, cytosine or adenine at PS17, cytosine or guanine at PS18 and cytosine or thymine at PS19. The invention also provides a collection of MPL isogenes, referred to herein as a MPL genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a MPL cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:5 (Fig.5) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 117, adenine at a position corresponding to nucleotide 340, guanine at a position corresponding to nucleotide 690, cytosine at a position corresponding to nucleotide 787, adenine at a position corresponding to nucleotide 962 and adenine at a position corresponding to nucleotide 996. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a MPL isogene defined by haplotypes 1-14.

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

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the MPL protein. The reference amino acid sequence comprises SEQ ID NO:6 (Fig.6) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of asparagine at a position corresponding to amino acid position 39, methionine at a position corresponding to amino acid position 114, leucine at a position corresponding to amino acid position 263 and glutamine at a position corresponding to amino acid position 321. A polymorphic variant of MPL is useful in studying the effect of the variation on the biological activity of MPL as well as on the binding affinity of candidate drugs targeting MPL for the treatment of congenital amegakaryocytic thrombocytopenia (CAMT).

The present invention also provides antibodies that recognize and bind to the above polymorphic MPL protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.
The present invention also provides nonhuman transgenic animals comprising one of the MPL polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the MPL isogenes in vivo, for in vivo screening and testing of drugs targeted against MPL protein, and for testing the efficacy of therapeutic agents and compounds for congenital amegakaryocytic thrombocytopenia (CAMT) in a biological system.

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

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a partial reference sequence for part of the MPL gene (Genbank Accession Number U68159.1; 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: 102 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, W = A or T, and V=A or G or C; WIPO standard ST.25).

Figure 2 illustrates a partial reference sequence for part of the MPL gene (Genbank Accession Number U68160.1; contiguous lines; SEQ ID NO:2), 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: 103 is equivalent to Figure 2, 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 3 illustrates a partial reference sequence for part of the MPL gene (Genbank Accession Number U68161.1; contiguous lines; SEQ ID NO:3), with the start and stop positions of each region of coding sequence indicated with a bracket ([ ] or ]) and the numerical position below the sequence.

Figure 4 illustrates a partial reference sequence for part of the MPL gene (Genbank Accession Number U68162.1; contiguous lines; SEQ ID NO:4), 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: 104 is equivalent to Figure 4, 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 5 illustrates a reference sequence for the MPL coding sequence (contiguous lines; SEQ ID NO:5) 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 6 illustrates a reference sequence for the MPL protein (contiguous lines; SEQ ID NO:6), with the variant amino acid(s) caused by the polymorphism(s) of Figure 5 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the MPL gene. As described in more detail below, the inventors herein discovered 19 isogenes of the MPL gene by characterizing the MPL 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 (21 individuals) (CA), African descent (20 individuals) (AF), Asian (20 individuals) (AS), or Hispanic/Latino (18 individuals) (HL). 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

<table>
<thead>
<tr>
<th>Population Group</th>
<th>Population Subgroup</th>
<th>No. of Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>African descent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sierra Leone</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burma</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>China</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Korea</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Vietnam</td>
<td>4</td>
</tr>
<tr>
<td>Caucasian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>British Isles</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>British Isles/Central</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>British Isles/Eastern</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Central/Eastern</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Eastern</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Central/Mediterranean</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mediterranean</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Scandinavian</td>
<td>2</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caribbean</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Caribbean (Spanish Descent)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Central American (Spanish Descent)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mexican American</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>South American (Spanish Descent)</td>
<td>3</td>
</tr>
</tbody>
</table>

In addition, the Index Repository contains three unrelated indigenous American Indians (AM) (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 MPL isogenses present in the human reference population are defined by haplotypes for 19 polymorphic sites in the MPL gene, all of which are believed to be novel. The MPL polymorphic sites identified by the inventors are referred to as PS1-PS19 to designate the order in which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index Repository for PS1-PS19 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the MPL gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the MPL gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether MPL is a suitable target for drugs to treat congenital amegakaryocytic thrombocytopenia (CAMT), 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 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 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 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 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 MPL 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 MPL polymorphisms and haplotypes identified herein.

The compositions comprise at least one MPL genotyping oligonucleotide. In one embodiment, a MPL 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, deoxiribonucleotides, 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 MPL polynucleotide, i.e., a MPL 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-MPL 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 MPL gene using the polymorphism information provided herein in conjunction with the known sequence information for the MPL 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, W=A or T, V=A or G or C, and B=T or G or C; 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 MPL gene polymorphisms comprises a nucleotide
sequence, listed 5' to 3', selected from the group consisting of:

Accession No.: U68159.1

TATTGTGGMGATTTGT (SEQ ID NO:7) and its complement,
AACCACAGCCAGCAT (SEQ ID NO:8) and its complement,
TATAATCTTAGCAC (SEQ ID NO:9) and its complement,
CCCTGAAATGTTTCT (SEQ ID NO:10) and its complement,
CCTCAAGVGGTCGGCA (SEQ ID NO:11) and its complement,
GAAGAATRTGTCTCCT (SEQ ID NO:12) and its complement,
GTAGAGAGTATGCT (SEQ ID NO:13) and its complement,
ATAGGCAYGCCTAAT (SEQ ID NO:14) and its complement,
GAGATTTYGCAACCAA (SEQ ID NO:15) and its complement,
TGATGGGCTCTTCCCT (SEQ ID NO:16) and its complement,
AGGGCACRGTCGTGC (SEQ ID NO:17) and its complement.

Accession No.: U68160.1

GTCTCTGRGGGCAGGC (SEQ ID NO:18) and its complement,
TCTGGGARAACLCTGG (SEQ ID NO:19) and its complement,
TGCTAGTRIATCTAT (SEQ ID NO:20) and its complement.

Accession No.: U68162.1

CTGCTCCYCCACCCC (SEQ ID NO:21) and its complement,
TGCCAATYCACTGCCC (SEQ ID NO:22) and its complement,
CACCAAMCTGCCTG (SEQ ID NO:23) and its complement,
AGCGCCTGCTTTCCAT (SEQ ID NO:24) and its complement, and
CTCAGCTCYCCAGGTC (SEQ ID NO:25) and its complement.

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

Accession No.: U68159.1

CAAGACCTATTGTGMR (SEQ ID NO:26); ACTTAGACAATCCKC (SEQ ID NO:27);
GTCCTTAAACACACRG (SEQ ID NO:28); CCTACTATGCTGCYG (SEQ ID NO:29);
CCATCTTAAATACRT (SEQ ID NO:30); GTATCTGTGCTAAYG (SEQ ID NO:31);
CAGAGGCCCTGAAKT (SEQ ID NO:32); TCCGGGAGAAACAMT (SEQ ID NO:33);
CTGGGTCTCAGGGVC (SEQ ID NO:34); CCACCAGGGGACBC (SEQ ID NO:35);
CTGGGGAAGAAATRT (SEQ ID NO:36); TGGTTTAGAACAYA (SEQ ID NO:37);
CCTCATTAGTAGARG (SEQ ID NO:38); AAGGTCAGCATCYT (SEQ ID NO:39);
TCAGATATGAGCAAYG (SEQ ID NO:40); CCCCTAAGTACCT (SEQ ID NO:41);
TTATTGAGATTTYG (SEQ ID NO:42); AGGGTATTGTTGCR (SEQ ID NO:43);
CGAACTGATGGMT (SEQ ID NO:44); CCACCAGGGGAGAKC (SEQ ID NO:45);
CACAGCAGGCGACRG (SEQ ID NO:46); TCTGGGGAGCACAYG (SEQ ID NO:47).

Accession No.: U68160.1

GGATTAGCTCTGGRG (SEQ ID NO:48); AATCGGCCTGCCYC (SEQ ID NO:49);
ACCCCATCTGGGARA (SEQ ID NO:50); CCTCTGCCAGTTYT (SEQ ID NO:51);
AGTCTCTGCCTAGTTR (SEQ ID NO:52); ATAAAATAGATAYA (SEQ ID NO:53).

Accession No.: U68162.1
GCCTCAGCTGCCYC (SEQ ID NO:54); GCAATGGGCTGGRG (SEQ ID NO:55);
CCTCCCTGCAATY (SEQ ID NO:56); AGGCATGGCAGTGR (SEQ ID NO:57);
GCCACCCACACMC (SEQ ID NO:58); CAGTACGGCAGCKG (SEQ ID NO:59);
TAATCCAGGCGCTST (SEQ ID NO:60); GAGAGATGAGGAGSA (SEQ ID NO:61);
AGGCTCTCTACCTCYC (SEQ ID NO:62); and GTCCAGGAAGCTGRG (SEQ ID NO:63).

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

Accession No.:U68159.1

GACTATGCTG (SEQ ID NO:64); TAGACAATCC (SEQ ID NO:65);
CTTAACACAC (SEQ ID NO:66); ACTATGCTGC (SEQ ID NO:67);
TTCTATAACT (SEQ ID NO:68); TCTGTGCTAA (SEQ ID NO:69);
AGCACCCTGAA (SEQ ID NO:70); GGGAGAAGAC (SEQ ID NO:71);
GTTCTCGAGG (SEQ ID NO:72); CCATGCAGCAC (SEQ ID NO:73);
GTTGAGAAG (SEQ ID NO:74); TTTAGGAACA (SEQ ID NO:75);
CAAGTACAGA (SEQ ID NO:76); TGTACAGAC (SEQ ID NO:77);
GATATGAGCA (SEQ ID NO:78); CTAAGTAC (SEQ ID NO:79);
ATGGAGATTG (SEQ ID NO:80); GTTTGTGTC (SEQ ID NO:81);
ACCTGATGGG (SEQ ID NO:82); CCGAGGGAG (SEQ ID NO:83);
AGCAGGGCA (SEQ ID NO:84); GGGGCACAC (SEQ ID NO:85);

Accession No.:U68160.1

TTAATCTCTG (SEQ ID NO:86); CAGCGCCTGC (SEQ ID NO:87);
CCATCGAGCA (SEQ ID NO:88); GTTCCGAGTGT (SEQ ID NO:89);
CTCTGCTAGT (SEQ ID NO:90); AACATAGATA (SEQ ID NO:91);

Accession No.:U68162.1

TCACCTGCTCC (SEQ ID NO:92); GTTGGGCTG (SEQ ID NO:93);
CCCCTGCGAAT (SEQ ID NO:94); CATGCGAGT (SEQ ID NO:95);
CACCCACAC (SEQ ID NO:96); TACCGCAG (SEQ ID NO:97);
TTCAGGCGCCT (SEQ ID NO:98); GAGATGAGG (SEQ ID NO:99);
CTCTCACTC (SEQ ID NO:100); and CAGGGACTG (SEQ ID NO:101).

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.

MPL 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 MPL 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 MPL gene in an individual. As used herein, the terms “MPL genotype” and “MPL 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 MPL 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 MPL gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19 in the two copies to assign a MPL 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-PS19.

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 MPL 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 MPL 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 MPL gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19 in that copy to assign a MPL haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the MPL gene or fragment such as one of the methods described above for preparing MPL isogenes, with targeted \textit{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 MPL gene copies present in an individual. If haplotype information is desired for the individual’s other copy, additional MPL 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 MPL gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-PS19 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the MPL haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual’s MPL gene, the phased sequence of nucleotides present at each of PS1-PS19. The present invention also contemplates that typically only a subset of PS1-PS19 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 \textit{PNAS} 97:10483-10488; Rieder MJ et al. 1999 \textit{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, \textit{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 \textit{Genetic Data Analysis II}, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a MPL haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19 in each copy of the MPL 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-PS19 in each copy of the MPL 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 MPL 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 MPL 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.


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 (Ruano et al., Nucl. Acids Res. 17:8392, 1989; Ruano 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 MPL haplotype pair is predicted from its MPL genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a MPL genotype for the individual at two or more MPL polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing MPL 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 MPL haplotype pairs shown in Table 4.

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

\[ n = \frac{\log(1-q)}{\log(1-p)} \]

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

\[ p_{H_1,H_2}(H_1 / H_2) = 2p[H_1]p[H_2] \] if \( H_1 \neq H_2 \)

and

\[ p_{H_1,H_2}(H_1 / H_2) = p[H_1]p[H_2] \] 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 MPL haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., supra).

The invention also provides a method for determining the frequency of a MPL 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 MPL 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 MPL 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 MPL 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 MPL 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 MPL genotype, haplotype, or haplotype pair. Preferably, the MPL genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting MPL 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 MPL 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 MPL 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 MPL genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their MPL 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 MPL 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”.


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 MPL 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 MPL 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 MPL 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 MPL 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 MPL 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 MPL 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 MPL gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19. Similarly, the nucleotide sequence of a variant fragment of the MPL 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 MPL gene (or other reported MPL sequences) or to portions of the reference sequence (or other reported MPL 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 NOS:1-4. The polymorphism is selected from the group consisting of adenine at PS1, adenine at PS2, adenine at PS3, thymine at PS4, guanine or adenine at PS5, adenine at PS6, guanine at PS7, thymine at PS8, thymine at PS9, cytosine at PS10, adenine at PS11, guanine at PS12, adenine at PS13, guanine at PS14, thymine at PS15, thymine at PS16, adenine at PS17, guanine at PS18 and thymine at PS19. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the MPL gene which is defined by any one of haplotypes 1-14 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the MPL 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.
MPL isogenes may be isolated using any method that allows separation of the two "copies" of the MPL 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 MPL genome anthologies, which are collections of MPL 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 MPL genome anthology may comprise individual MPL 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 MPL 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 MPL genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded MPL 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 MPL 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 MPL gene will produce MPL 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 MPL cDNA comprising a nucleotide sequence which is a polymorphic variant of the MPL reference coding sequence shown in Figure 5. Thus, the invention also provides MPL mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO: 5 (Fig. 5), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 117, adenine at a position corresponding to nucleotide 340, guanine at a position corresponding to nucleotide 690, cytosine at a position corresponding to nucleotide 787, adenine at a position corresponding to nucleotide 962 and adenine at a position corresponding to nucleotide 996. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a MPL isogene defined by haplotypes 1-14. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized MPL 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 MPL 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 MPL 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 MPL 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 MPL 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 MPL protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the MPL 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 MPL isogene. Expression of a MPL 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 MPL mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of MPL 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.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference MPL amino acid sequence shown in Figure 6. The location of a variant amino acid in a MPL polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:6 (Fig.6). A MPL protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of asparagine at a position corresponding to amino acid position 39, methionine at a position corresponding to amino acid position 114, leucine at a position corresponding to amino acid position 263 and glutamine at a position corresponding to amino acid position 321. The invention specifically excludes amino acid sequences identical to those previously identified for MPL, including SEQ ID
NO:6, and previously described fragments thereof. MPL protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:6 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a MPL protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

<table>
<thead>
<tr>
<th>Polymorphic Variant Number</th>
<th>Amino Acid Position and Identities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>K V I Q</td>
</tr>
<tr>
<td>2</td>
<td>K V L R</td>
</tr>
<tr>
<td>3</td>
<td>K V L Q</td>
</tr>
<tr>
<td>4</td>
<td>K M I R</td>
</tr>
<tr>
<td>5</td>
<td>K M I Q</td>
</tr>
<tr>
<td>6</td>
<td>K M L R</td>
</tr>
<tr>
<td>7</td>
<td>K M L Q</td>
</tr>
<tr>
<td>8</td>
<td>N V I R</td>
</tr>
<tr>
<td>9</td>
<td>N V I Q</td>
</tr>
<tr>
<td>10</td>
<td>N V L R</td>
</tr>
<tr>
<td>11</td>
<td>N V L Q</td>
</tr>
<tr>
<td>12</td>
<td>N M I R</td>
</tr>
<tr>
<td>13</td>
<td>N M I Q</td>
</tr>
<tr>
<td>14</td>
<td>N M L R</td>
</tr>
<tr>
<td>15</td>
<td>N M L Q</td>
</tr>
</tbody>
</table>

The invention also includes MPL peptide variants, which are any fragments of a MPL protein variant that contain one or more of the amino acid variations shown in Table 2. A MPL peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such MPL peptide variants may be useful as antigens to generate antibodies specific for one of the above MPL isoforms. In addition, the MPL peptide variants may be useful in drug screening assays.

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

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

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

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

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

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

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

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

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

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

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

Effect(s) of the polymorphisms identified herein on expression of MPL may be investigated
by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant
animals, containing a polymorphic variant of the MPL 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 MPL 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 MPL 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
MPL isogene is introduced into a cell in such a way that it recombines with the endogenous MPL gene
present in the cell. Such recombination requires the occurrence of a double recombination event,
thereby resulting in the desired MPL 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 MPL 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 MPL 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 MPL 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 MPL
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 MPL isogene and producing human MPL protein can be used as biological models for studying diseases related to abnormal MPL 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 MPL isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel MPL isogenes; an antisense oligonucleotide directed against one of the novel MPL isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel MPL 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 MPL 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 MPL 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
ethenographic origin, clinical responses, genotypes, and haplotypes for one or more populations).
The MPL 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,
(1989).

EXAMPLE 1

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

Amplification of Target Regions

The following target regions of the MPL gene were amplified using PCR primer pairs. The
primers used for each region are represented below by providing the nucleotide positions of their
initial and final nucleotides, which correspond to positions in the indicated GenBank Accession
Number.
### PCR Primer Pairs

<table>
<thead>
<tr>
<th>GenBank Acc No.</th>
<th>Fragment No.</th>
<th>Forward Primer</th>
<th>Reverse Primer (complement of)</th>
<th>PCR Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>U68159.1</td>
<td>Fragment 1</td>
<td>381-403</td>
<td>985-964</td>
<td>605 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 2</td>
<td>670-692</td>
<td>1253-1232</td>
<td>584 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 3</td>
<td>1014-1034</td>
<td>1521-1498</td>
<td>508 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 4</td>
<td>1488-1509</td>
<td>2025-2003</td>
<td>538 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 5</td>
<td>2153-2177</td>
<td>2672-2650</td>
<td>520 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 6</td>
<td>2345-2366</td>
<td>2844-2822</td>
<td>500 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 7</td>
<td>2843-2864</td>
<td>3535-3511</td>
<td>693 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 8</td>
<td>3218-3241</td>
<td>3855-3833</td>
<td>638 nt</td>
</tr>
<tr>
<td>U68160.1</td>
<td>Fragment 9</td>
<td>56-79</td>
<td>591-569</td>
<td>536 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 10</td>
<td>318-340</td>
<td>1008-986</td>
<td>691 nt</td>
</tr>
<tr>
<td>U68162.1</td>
<td>Fragment 11</td>
<td>159-181</td>
<td>676-654</td>
<td>518 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 12</td>
<td>477-497</td>
<td>1055-1032</td>
<td>579 nt</td>
</tr>
<tr>
<td></td>
<td>Fragment 13</td>
<td>760-782</td>
<td>1406-1385</td>
<td>647 nt</td>
</tr>
</tbody>
</table>

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**: $= 10 \mu$l
- **10 x Advantage 2 Polymerase reaction buffer (Clontech)** $= 1 \mu$l
- **100 ng of human genomic DNA**: $= 1 \mu$l
- **10 mM dNTP**: $= 0.4 \mu$l
- **Advantage 2 Polymerase enzyme mix (Clontech)** $= 0.2 \mu$l
- **Forward Primer (10 \mu M)** $= 0.4 \mu$l
- **Reverse Primer (10 \mu M)** $= 0.4 \mu$l
- **Water**: $= 6.6 \mu$l

**Amplification profile:**
- **97°C - 2 min.** $\{ 1 \text{ cycle} $
- **97°C - 15 sec.**
- **70°C - 45 sec.** $\{ 10 \text{ cycles} $
- **72°C - 45 sec.** $\} 35 \text{ cycles}$

**Sequencing of PCR Products**

The PCR products were purified using a Whatman/Polyfiltronics 100 \mu l 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 \mu l of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in the
indicated GenBank Accession Number. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

**Sequencing Primer Pairs**

<table>
<thead>
<tr>
<th>GenBank Acc No.</th>
<th>Fragment No.</th>
<th>Forward Primer</th>
<th>Reverse Primer (complement of)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U68159.1</td>
<td>Fragment 1</td>
<td>424-443</td>
<td>932-914</td>
</tr>
<tr>
<td></td>
<td>Fragment 2</td>
<td>721-741</td>
<td>1226-1207</td>
</tr>
<tr>
<td></td>
<td>Fragment 3</td>
<td>1068-1087</td>
<td>1494-1475</td>
</tr>
<tr>
<td></td>
<td>Fragment 4</td>
<td>1515-1534</td>
<td>1960-1941</td>
</tr>
<tr>
<td></td>
<td>Fragment 5</td>
<td>2231-2250</td>
<td>2624-2605</td>
</tr>
<tr>
<td></td>
<td>Fragment 6</td>
<td>2413-2431</td>
<td>2819-2800</td>
</tr>
<tr>
<td></td>
<td>Fragment 7</td>
<td>2903-2922</td>
<td>3353-3334</td>
</tr>
<tr>
<td></td>
<td>Fragment 8</td>
<td>3306-3325</td>
<td>3807-3788</td>
</tr>
<tr>
<td>U68160.1</td>
<td>Fragment 9</td>
<td>87-106</td>
<td>547-527</td>
</tr>
<tr>
<td></td>
<td>Fragment 10</td>
<td>426-445</td>
<td>934-915</td>
</tr>
<tr>
<td>U68162.1</td>
<td>Fragment 11</td>
<td>205-224</td>
<td>579-560</td>
</tr>
<tr>
<td></td>
<td>Fragment 12</td>
<td>522-541</td>
<td>1014-995</td>
</tr>
<tr>
<td></td>
<td>Fragment 13</td>
<td>798-816</td>
<td>1333-1312</td>
</tr>
</tbody>
</table>

**Analysis of Sequences for Polymorphic Sites**

Sequences were analyzed for the presence of polymorphisms using the Polphred 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 MPL gene are listed in Table 3 below.

**Table 3. Polymorphic Sites Identified in the MPL Gene**

<table>
<thead>
<tr>
<th>Polymorphic Site Number</th>
<th>PolylIda</th>
<th>Nucleotide Position</th>
<th>Reference Allele</th>
<th>Variant Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS1</td>
<td>7091</td>
<td>550(Acc#U68159.1)</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>PS2</td>
<td>7089</td>
<td>590(Acc#U68159.1)</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>PS3</td>
<td>7088</td>
<td>652(Acc#U68159.1)</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>PS4</td>
<td>7094</td>
<td>1238(Acc#U68159.1)</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>PS5</td>
<td>7097</td>
<td>1598(Acc#U68159.1)</td>
<td>C</td>
<td>G or A</td>
</tr>
<tr>
<td>PS6</td>
<td>7096</td>
<td>1771(Acc#U68159.1)</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>PS7</td>
<td>7101</td>
<td>2671(Acc#U68159.1)</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>PS8</td>
<td>7100</td>
<td>2805(Acc#U68159.1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>PS9</td>
<td>7123</td>
<td>2886(Acc#U68159.1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>PS10</td>
<td>7104</td>
<td>3160(Acc#U68159.1)</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>PS11</td>
<td>7106</td>
<td>3595(Acc#U68159.1)</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>PS12</td>
<td>7109</td>
<td>168(Acc#U68160.1)</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>PS13</td>
<td>7110</td>
<td>224(Acc#U68160.1)</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>PS14</td>
<td>7115</td>
<td>807(Acc#U68160.1)</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>PS15</td>
<td>7116</td>
<td>276(Acc#U68162.1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>PS16</td>
<td>7118</td>
<td>309(Acc#U68162.1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>PS17</td>
<td>7117</td>
<td>475(Acc#U68162.1)</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>PS18</td>
<td>7122</td>
<td>636(Acc#U68162.1)</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>PS19</td>
<td>7121</td>
<td>925(Acc#U68162.1)</td>
<td>C</td>
<td>T</td>
</tr>
</tbody>
</table>

*PolyIda is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.*
EXAMPLE 2

This example illustrates analysis of the MPL 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 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

<table>
<thead>
<tr>
<th>Genotype Number</th>
<th>Polymorphic Sites</th>
<th>HAP PAIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C G G G G G A C C C C C C C 1 1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C/A G/A G G G G A C A G G A C C C C C C C 2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C G G G G G A/G C C A G G A C C C C C C C 1 3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>C G G G G G A C C A G/A G G A/G C C C C C C 1 4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>C G G G G G G/A A C C A G G A C C C C C C 1 5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>C/A G/A G G G G A C A G G A C C C C C C C 1 7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>C G G G G G A C C T A G G A C C C C C C C 1 8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>C G G G G/T G G A C/T C A G G A C C C C C C 1 9</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>C G G G G A/A G A C C A/G G G A C C C C C C 1 10</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>C/A G/A G G G G A C T C A G G A C C C C C C C 1 11</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>C G G G G/A G A C C A G G A C C C C C C C 1 13</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>C/A G G G G A C C A G G/A G - C/T C C C C C 1 14</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>C/A G G/A G G A C C T A G G/A G A C C C C C C 2 8</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>A G A G G G A C C A G A G/A A - - - - - - - 2 12</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>C G G G G G A C C A G G G C C C C C 4 4</td>
<td></td>
</tr>
</tbody>
</table>
The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark’s algorithm (Clark, A.G. 1990 Mol Bio Evol 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample. 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 14 human MPL haplotypes shown in Table 5 below. Haplotypes 14 and 16 disclosed in the previous application (Serial No. 60/197,839) are not shown in the present application. It was discovered that these haplotypes were specific only to the chimpanzee DNA and are not included in the human haplotypes presented below.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Polymorphic Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>PS</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
</tr>
<tr>
<td>11</td>
<td>A</td>
</tr>
<tr>
<td>12</td>
<td>A</td>
</tr>
<tr>
<td>13</td>
<td>C</td>
</tr>
<tr>
<td>14</td>
<td>A</td>
</tr>
</tbody>
</table>

Table 6 below shows the number of chromosomes in unrelated individuals seen for each haplotype arranged by the ethnic background of the subjects in the Index Repository.
Table 6. Frequencies of Haplotypes Observed in the MPL Gene

<table>
<thead>
<tr>
<th>Haplotype Number</th>
<th>AF</th>
<th>AS</th>
<th>CA</th>
<th>HL</th>
<th>AM</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>35</td>
<td>24</td>
<td>24</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7 below shows the number of unrelated subjects assigned to a given haplotype pair arranged by the ethnic background of the subjects in the Index Repository.

Table 7. Frequencies of Haplotype Pairs Observed in the MPL Gene

<table>
<thead>
<tr>
<th>HAP PAIR</th>
<th>AF</th>
<th>AS</th>
<th>CA</th>
<th>HL</th>
<th>AM</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>16</td>
<td>8</td>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

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 myeloproliferative leukemia virus oncogene (MPL) gene of an individual which comprises determining whether the individual has one of the MPL haplotypes shown in Table 5 or one of the haplotype pairs shown in Table 4.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS19 on at least one copy of the individual’s MPL gene.

3. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS19 on both copies of the individual’s MPL gene.

4. A method for genotyping the myeloproliferative leukemia virus oncogene (MPL) gene of an individual, comprising determining for the two copies of the MPL gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19.

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 MPL 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 MPL gene present in the individual the identity of the nucleotide pair at each of PS1-PS19.

7. A method for haplotyping the myeloproliferative leukemia virus oncogene (MPL) gene of an individual which comprises determining, for one copy of the MPL gene present in the individual, the identity of the nucleotide at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19.

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 MPL 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 myeloproliferative leukemia virus oncogene
(MPL) gene of an individual comprising:
(a) identifying a MPL genotype for the individual, wherein the genotype comprises the
nucleotide pair at two or more polymorphic sites selected from the group consisting of
PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15,
PS16, PS17, PS18 and PS19;
(b) enumerating all possible haplotype pairs which are consistent with the genotype;
(c) comparing the possible haplotype pairs to the data in Table 4; 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-PS19.
11. A method for identifying an association between a trait and at least one haplotype or haplotype
pair of the myeloproliferative leukemia virus oncogene (MPL) 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 1-14 shown in Table 5 and the haplotype pair is selected
from the haplotype pairs shown in Table 4, 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 MPL.
13. A composition comprising at least one genotyping oligonucleotide for detecting a
polymorphism in the myeloproliferative leukemia virus oncogene (MPL) gene at a polymorphic
site selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10,
PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19.
14. The composition of claim 13, wherein the genotyping oligonucleotide is an allele-specific
oligonucleotide that specifically hybridizes to an allele of the MPL 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:7-22, the complements of SEQ ID NOS:4-22, and SEQ ID NOS:23-60.

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:61-98.

18. A kit for genotyping the MPL gene of an individual, which comprises a set of oligonucleotides designed to genotype each of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18 and PS19.

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 myeloproliferative leukemia virus oncogene (MPL) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises a MPL isogene defined by a haplotype selected from the group consisting of haplotypes 1-14 in Table 5; 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 MPL 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 MPL gene, the fragment comprising one or more polymorphisms selected from the group consisting of adenine at PS1, adenine at PS2, adenine at PS3, thymine at PS4, guanine or adenine at PS5, adenine at PS6, guanine at PS7, thymine at PS8, thymine at PS9, cytosine at PS10, adenine at PS11, guanine at PS12, adenine at PS13, guanine at PS14, thymine at PS15, thymine at PS16, adenine at PS17, guanine at PS18 and thymine at PS19.

24. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the MPL cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises the coding sequence of a MPL isogene defined by one of the haplotypes shown in Table 5.

25. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 24, wherein the organism expresses a myeloproliferative leukemia virus oncogene
26. The recombinant organism of claim 25, which is a nonhuman transgenic animal.

27. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the MPL protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant is encoded by an isogene defined by one of the haplotypes shown in Table 5.

28. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 27.

29. A method for screening for drugs targeting the isolated polypeptide of claim 27 which comprises contacting the MPL polymorphic variant with a candidate agent and assaying for binding activity.

30. A computer system for storing and analyzing polymorphism data for the myeloproliferative leukemia virus oncogene 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 4 and the haplotypes shown in Table 5.

31. A genome anthology for the myeloproliferative leukemia virus oncogene (MPL) gene which comprises MPL isogenes defined by any one of haplotypes 1-14 shown in Table 5.
POLYMORPHISMS IN THE MPL GENE

CAAGGGGGCCA GGGTAAGGAG TGTGAGCCAT CTCCCAATCTG AGCAAAACAG
ATAAAAGTACC TGAAGACCAT TGCTTCTCCA ACTTTAAGTT ACATTCAAC 100
CATCTGGGGA TTTTGCCAAA ATGCAGGCTC AACCTTAAAT GCCCTGGGCT
GGGGGCTTAA ATCTCGGATT TCTAATAGTG GCCCACCTAA GGCTGGTGCT 200
CCTGTGTCTAC AGACACCATG TTGATGAGCA AGGGAATAGA GGACACCGGT
GAGGCCACCATA ATATAAGGAA GGGATTCCC AGTCCACGGGG AAGATGTTT 300
TATTTCATCGC AAAAGTCACA GAACATATTAA AGTGGAAATA CCCACATT
GGAGAAGGTG TGGGCCGAGA AGTATTTTCT CAGCGAGATT GCTCTCCACGA
AGCATAGCAG AGCATGGGCT TTAATGACGT GGGTCTGAAG GGTTGACACT
CCTGACCTTT ATGTCTTATG ATGTAAATCT TGAACATCTCT TAACTTCTC 500
TGATCTCCCG TTTCTCCTAC TCTATAATGG AAATACAAAG ACTATTGGC
A
GGATGTCTCA AGTGAATGGA TATAAGTGAC TTAACACAGG GCAGCATAGT
A
AAGTGCTTCA TAAAGATTGC TATTTGGATG ATTAATACCC ATTCTATAAT
CGTTCAGACA GATACAGACA CTCTAGTTGC TTTTGTTACA CGTCTCAATAA 700
A
TACACCCACC ATCTCTCTCA CACCAAAAGG CCTGTGTCCT CTCTTTCTAA
TTCTCTCCTC ACTCTCCGGC ACTCTCCCTC CCTGCTGCCC TGGCCCCAGT 800
TGGTCTGCA TGGGCCGAGA AGGGAACGAGG CAGGCGACAG GACCTGGGCA
TGATTCTGAC AGGAACCTGA GGGATCTGCC TGGAGGGGGA TTGGGAGGCC
GCTTCTGCA AAGGAGATGG GCTAAGCCAG GCACACAGTG GCAGAGAGAGA
[EXON 1: 950..]
TGCCCTCTCG GGCCTCTTTC ATGTTCACTT CCTGCTCTCT CTCGCCCTCT 1000
CAAAACCTGG CCAAGCTCAG CAGCCAAGGT GAGGGCACA GACCTGGTAGG
..1028]
ATCACCCTG CCCCCGGGAG AGGAGGCTCT GGAGGCTGAT GACGGGCCTC 1100
CGGAGGGGAG GTAGAGTAAAG AGGCCTCTCT CAGGTCTCCC TCCCTTCTCC
CATAAACATG CCTGAGGAGA CCCAGGCACACTACCACTAGGCT 1200
ATGCTCCTTT GCTGGGCTACA GACAGCAAGC CTCCTGAGAT TTTCTCCCGA
T
[EXON 2: 1201..]
ACATTTGAGA ACCTCAGCTG CTCTCTGGAT AGGAAGAGAG AGGCGCCCGAG 1300
TGGGAGAACG CAGCTGCTGT ATGCTCTCCC GCGGTAGTGT CTGGACTGTA
..1333]
CCCCACTCCC CATGTATCTG TCCCTCTAGT TAGCTGAGTC CCACTCCAGC 1400
AGGTTTCTTG CAGTTCCGAGG GACCACCTCT ATACAGCCAG CTGACGACTC
ACTTTTTGTAGG CATGGTCTTG TGGGTGTATG GTGGGCGCGA GCCCGACGGC 1500
TACATACGAA CTAAATCCAC ATATCCCCAG CAGTGAAGAG AAAAAAGGCA
GTACTAGAACA CAGAAGTGGG GCAAGGGGGC AGGGTGCTGT CCTCAGGCGT
G A
CCGCATGGGTG GCTGCTGGAG AGGGACCTCT TCTATGCAAA CAGGGAGAAG
[EXON 3: 1644..]
CCCCGTGTCTT GCCGCTGAGG TTTTCCAGAGG ATGGAGGACT TTGRAGCCCC 1700
ATACGTCTGCA CAGTTTTCCAG ACCAGGGAGA AGTGCGCTCT TCTCTTCCGC
TGACACCTCTG GGATGAAGAT GTGTCTCTAA ACCAGACCTG GACTCCAGCA
800
A
GTCTCTTTTG TGGACAGTGT AGGTAAGAGC CATCCTCTGT TACCCCTGCC
..1822]
CCCTCCACTT GCTCCCCCCA GTCCGGCTGC CGGAATCAGA CTTGCTCTGG
1900

FIGURE 1A
FIGURE 2

POLYMORPHISMS IN THE MPL GENE (Accession No. U68160.1)

AAATGCAA AACAACGAA CAAATGACA AAGGAAA A CAAAAGAA
ACATCATGGC ATAAAGACAG AGGGAAAGCT CTGGAGGCCCT GTGGAGGA 100
GGAGTTAAAT GATAATCCCT GCAGGCCATC GTTCTTGTAG GATGGAAGGC
CTTGGGATTAA GTCTCTGAGG CAGGCCCTAG TCAATGACTTC TGTTGGGCTG 200
GGTCTTACCTG ACCCCTACGCA GAGAAGAAG AGAAACAAAA

[EXON 7: 209]
TCCAGGACTA CAGACCCAC ATGTTCTCTG CTGGCCA TTTC AAGTCAGA 300
ATGACAGCAT TATTCACATC CTTGGAGGG TGACCAACAG CCGGTAC 400
GTTCCAGACT CCCTTGCCCTG ATCCAGCAGG CTGTATTAGA 500
CTTCTCTCGTG CATGCACACG ACGACCAGCA ACCCCAGTCT ATCTGAC 550
TGTCACCAG GATCCCACCTG AGACCCCTCAG TCACCCCTCAG GTCTGTC 600
CTGAGTCCGCT TCCACACGC AAACCTGCAAGT TGAGGGAAGA TCTCAGTGG

[EXON 8: 555]
GCATCTGGAA TTGGAGTGCG AGCACCACATC GTCTTGGGCA GCACAACAGA
CCTGTTATCA ACTCCGATAC ACAAGAAGAAG GCAAGTAGGA 700

TTGTCAGGCA ACAATGACC AGACACCTCA TACGCACGGG GATCCCTGGG

GTGGCCCATG CCTGCCTAGCA GGAGTTAAGG TGTCATATGA TCCAGTCTC 800

CTTGATATAT CTTGTTATAT CAGATTGCACT TGATCTCTTA GCTCCTCTG

GGCATCTGAT ATTCTCCTGG ACTCTCTATTG AACATTTCTC TCCCTCTTG 900

CTTGGGGAATA TCAGCTCTCC CGATTTCCTT TCCTTCATCT GTGACTGAC

TTTCCTACTC CTCCTCCTCTG GCCTCCTTCTC CCAGGCTCCTC CATAAATG 1000

TGTCCTCTCT ACAGCTCTCA CTTAGGCTCT TTTCTCTCTCTGATAAAATAA

ATATGCTCTCC CCAAGGTGAG TCCTTTATCT CACTGTTTCT AATGTCATCA

CATPACAAT CGAAATTCTT CACATCTGAA CTCCATATTG GAAGAATTTT

AACAAACACA GGCAATTAC CAAATGTTT TTGAAGCTGTCTT CTTCTATGG

TGAATGGAGG ATGAGCTTACT TCACCTTATG GATTTATAGT AAGATAAAA

TATATATCT GAAATACCTAG CACACTGCTT GCATAAGGA AGGTGCTCAA

TATACATAGCT TAAATATATG TTGGTTCTAC TTTATATCTTAT ATTGACAC

CATATATACA ACTATGTACA AATCTATTG GTT

1383
POLYMORPHISMS IN THE MPL GENE

Accession No. U68161.1

| CCTGTCGACTA | ATTAATTTGT | GTGATTATTT | ACTCTGGTGG | ATGCTCTTTCT |
| CATTAGATAT | GTAGAGAGCG | GTGTTTATTT | GCTTTCTCCC | CAGTGGCTAA   |
| CCCATGGTCT | GGAGTATCAC | CACGTCTCAA | ATGTCTTGCT | GCTTTCTGCG |
| TTTGATCATC | AGTATCTTGA | CCCAGAGGCC | TAAAGCAGAT | CTGGCATTCT |
| CTGCCAATCT | AGTGATATTT | CGGTTGCTGT | CTGATGTGCT | CTGAATATAT |
| CTGGTGGTGT | GGTTGTCTTG | CGCCGACCT | GAGATTGCTA | GAACCGCCGG |
| TAGGGGTGCT | GTTGACGGGC | ATCCCGTGCG | AGGTTGATC | TGGGTGTTG |
| TGGTAGACAA | CTAGTGCTTG | TGAGGTGAGG | TCTGTGTCTC | AGGAGGTGGG |
| CTTGTCGCGT | TTCTGTGCGG | CAGAGCTGTA | TCCGGTTAAG | GAGGCTCGT |
| GTAGGCTGCG | CTATCTGTGT | TGCTGGAGAG | CGGTTTTTCT | GCCGGTGGGG |
| CTCTTTTGAA | GAAGCTGGAG | CCAGCTGGGG | AGTGGAGAGT | CGAGAGTTG |
| GCTCAAGAGC | ACCTGCGGCT | ATCGAAGCCC | CGACCGCCGG | CCACCGCAC |
| CTTCTTTGCT | CAGGTGCTGG | AGCCGCCTCT | CGGGGCCCCG | GAGGAGGCCC |

[EXON 9: 614..
TGGAGCTGCG CGCCGCGATCT CGCTACGGTT TACAGCTGCG CGCCAGGCTC 700
AAGCCGCCCAG CTTACCAAGG TCCCTGGAGC TCTGTGCTGG ACCAAACTAG 800
GTTGAGACC GCCACCAGCA CGGTTGAGGC AAGCCCCCGC GCACCAAAAG 773
..773] CCGCACAGCG CTCGCGGAGG GACTGGCGGC CGGTTGCGG AGGGGCGGGG 900
CTCGAGAGGG GGCAGGGNGG GCGGNGGAGA GGGGGGCGCC CTGACCTTGG 1000
GGGGAGACGC TGGCGAGGGG CGGAGCCTCC AGGGGCGGCG AGGGGGCGGG 1100
GCCAGGTGATG GAGGCTGGCTG ATAGGAGGGG GGCTCGCCGC CCGGTGGGGG
[EXON 10: 1028..
CTGCACTCTAG TGCTGGGGCTT CAGCCGCGTC TGCCCCGTGC TCTGTGCTAG 1200
GTGCGAGTTT CCTGACACTA ACAGGTACGG CCACCCGGAC GACAGGACT 1124]

GCCGCGTGAC CAGGTGAGGC CGAACCCGGGT TAAACAGAGCA TTCTTGATTC 1300
GCTGCTGAGT CCCGATGCTCT CTTGCCCGGG CCGTGACCGC CTAGGGGTCC
GCACCTGCTG CAGCTCAGCT CGGTGAATCG CGGGCGCTCC TTAACACTCTA
ACCCGGAGGC TATACCCGCC CCCCTGACCTA CTTGCTGCTCA 1400
CTCCTGAGCC AGGACAGCCT GCCACCCCTC AGCGAACCTG GACACTCCAC
CCATTCAAGA CTCTTGCTCC GCTCCCGCCTA CTAAAGCTAG TCAACCACAG
GGCCCTTTT CTCTAGCCCT ATACAGAGCA GTAGTTGACCA CCACAAAAAT 1500
CTCTTTGCCC CTCCGACTGA CTGAGCTGCA CACACAAATA ATCTTCTCAA
AATGCCGCA CAGAACCTATC AAAATCTTCC TTAAATCTTT CTAATGACTC
CCAGATATA GTCCACGTCT CTTATAGTGCA CAGAGAACCC TCTGGAAAT
CGCGGCGGTC CAGCGAGCG TCCAGCCTGT TCTACCTCTGT GCACAGGCTT 1700
TTCCAAACCTG GTTTTCTCCTA CAACTCACGT TTTACTCTCT AGTGCTTTTG
CACATAAACG CTGGCTGCCC TGAAAATGCC TCCCT 1785

FIGURE 3
POLYMORPHISMS IN THE MPL GENE

(Accession No. U68162.1)

CTGCAGATAAA AGAGGGAGTA CTGTGTCTCA GGCAAGGTAG TACGGGATGA
CATTTGACACA GAGGCCTTCA ATGGAGTGGG GGAAGTAACT CAGTGTGCCA  100
GACAGAACA ACAATGGAGT GTTTGAGGAA ACTGCAGCAC CTTTGGTCTA
ACCTGGCAGC CACGACTCTT CTCCTCTCCC CAGGTGGTTT TCGAAGTGTC  200
TACTCTCTGC CCCAACTTTA TTGGGCTGAC CATGCCTTAA TATCTCCAAG
CTCTATCCAG CGGCCTTCCC GCTCCTCCCC CCCAACTGCT GTCTCCTTCC  300

CTGCCAATTCG ACTGCGATGG CTCAGCTTGC TTTCTTCTCT TCTCTCCCAG

T

GAGACTGAGG CATGCCCTGT GGCCCTCCTC TCCAGACCTG CACCCGCGTC
[EXON 11: 351...]

TAGGCCAGTA CTTAGGGGA ACTGCAGCCC TGAGCCCGCT GAGTGTGCTT
..438]

CCCTCCCTCG TGCCCACACC CAACCGTGC CTTGACTTGG TCTCTGCCCC  500

A

AACAAATACAA CTTGTTCAGG GCTCGGTGCC TACCTAGGTG CATCCACGC
GCACTACCCC AGCAGATCCC CCACCTCCCT CCTCTCTTGA CATGAGACGC
CCCTCTCCAG CAGAGCTCGT TTATATCCAG GCCTGCTCTC CATCTCTCCC  600

G

AGCCCAAGGC CACAGTCTCA GATACCTGTG AGAGAAATGGA ACCCAGGCTC
[EXON 12: 653...]

CTTGAATGCC TCCCCACGTC CTCAGAGAGG ACTCTTGGGC CCGCTGTGTC
CTCCAGGCC CAGATGGCCT AAGCAAGATG GCCGTGCTTC TCGCCTGGGCA  800
CCATGGCCCT GTGGTTCGCG GACCCACAGG CAGTAAGGAG GCTCTCTGGT
ACCCACGACA TTGCAACAGC TTTCTACCTA CCACTAAGCT ATGGGCGAGA
GGCTTGAGGA CAGGCTTCTC ACTCCACCTG CTTGGGACC AGCTAAACCT

T

..907]

TCGAGACTTC TCTGTGAACCT TCCCTACCCCT ACCCCCAAAA CACAAAGCAC
CCAGACCTCA CTCCTAATCCC CTTCTTGCTG CCACTCAAT TAGGCCTCAT
TGACATCTAC TTACTCTACTG CTTAGTCCAA GAAATTAACAG AATCTCTTC
CACAGGGCAG CTTAGTCCTA TAAGCCTACCT TTTACTTTCC TCTCTCTTCT
TTGATCTGACA CGGAGCCTAG AACAGCCTCA CAGAACCCGT ACACTCTCTG
TTACTCTGC GACCTCTTCA AATAGTTCCC CTACTACACT TGACTGTGTA
ACTGCCAGG CAAGTGCCAC CTCAAACCTT ATAATCTCCA GATCCATAG
GACTCTCTTA ATAATGCTGT CTTTTTACAT CTGCTGAAAG TTTGTCAAGG
CTGATACCTC ACTTCTCTCT AATAATTTTT CACTACCTGG TCTGCTCCTC
TTGATGTATT TAGTTAGTTT TTTTTATTTG TTTTGAAGAC GGTCTACCTC
TTCACCGCCA GCTGGAGGGC CTTAGCTTGC AAGAGCTGCA CAGGGCGTG
ATTACAGGGC CACACCAACA CAAAGCTCCT ATTTTTTTTT TTTTTTTTT
TTTTTTTTTT TTTTTTTTTT GAGGAGCCTG CTTGTTGGAC AGCTGAGTGC
AGTCGGGCAAG TCTCGGTCTGA CTGGCAACCTC TGGGCTGCGG GGTCAAGGCA
TCTCTGCTCA GCTCCCAAGC TAGCGGAGGA TACAGGCTC TCCCGACATG
CATTATTTTT TCTATTGATA GGAGAGACGG TGGTTACACCA CGTGGGCCAG
GATGCTGTATG ATAATCTGATG TCAGCTGTCC CTCCTGCTTG CTCCCAARG
TTGCGGATT AGCGGTGTTA CCACCTCGG ACAGCGCGCG ATAAATTTCA
TATTTTTTGT AGAGAGCGGG TTTGCGCATG TCGCCCGGG GGTCTCTGGA
CTCTACTGTT CGGGTGATCC ACCACACTGG GCCTCCTAAG GTCTAGGAT
TACAGGCGAT AGCCAGCCTG CCGCCGCGAG TCTCTAGATA GTTAAAGGAA
TAAACCTAGAT CTAGAATCAC AGCTGGATTC AATCTCTGTCC TCCACATT

FIGURE 4A
ACTAGCTGTG CAACCTGGG CACATAACTT AATGTCTTTG AGCCCTAGTT
TTTTCACTCTG TAAACAGGG ATAATAACAG CACCCCCATAG AGTTGGGACG 2200
AGGATGAGA TAATCTAAGT AAAGCACAGT CCCCAGACA TAGTAAATGA
TTCATATATCG GAACATACGT TTATAATTAT TCTCTTCTAC TCTCT CCTTC 2300
TAGCATTTCT TCCAAATTAT ACAGTCCTTC AAGTATCCAT TTCTTACACAG
TCTCCAATCC CATCTATTTCT CGCCCTTTAC TATAGTTTTG CCAATTCCAAA 2400
GTTCTTATCT CTAGCTCAGA CATCTACTAG AGCAGTGTTA TGCTTTATGC
AACTAACTGT TTACATATCT GTCCCCGTCT ACTAGATTGG GAGGCTCCTTG 2500
AGGGAGAGGA ACATGATTTA TTGGTCCTTT TCCCCACGCA CCTAGACTAG
TGCTTGGGTGC ATGATAGTAG GCCTCTCAATA AAATTTCCTT AATGAATGA 2600
AATGTTTTCT CAGCAAGTTA CTTTATCCAA ATGCGAGCTG TCCTCATCCT
CTAAACCGTT CATAGGCAGC AGTAAATGTT CCAGACAGTT CTGATTATGG 2700
ACTCTCCATTA CACATTCTAA TTCTCAACAG AATGTTTTTG TTGTCTATTGT
TTTTNCCTGGC TCCAAATATA CCCACTGTAT CAGTCAGGGT CTTGGAAAGAA 2800
ATAGATGACA CACTCAAATT GGGTAATATG ATGACAGTGT AATAAACT 2848

FIGURE 4B
### POLYMORPHISMS IN THE CODING SEQUENCE OF MPL

| ATGCCCTCCT | GGGCCCTTTT | CATGGTCACC | TCCTGCCTCC | TCCTGGCCCG |
| TCAAAACCTG | GCCCAAGTCA | GCACCAGAAG | TGCTCTCTTG | CTGGCATCAG |
| ACTCAGAGCC | CCTGAAGTGT | TTCTCCTCGA | CATTTGAGGA | CCTCACTTGC |
| TTCTGGGATG | AGGAAAGGGC | AGCGCCCCAGT | GGGACATACTC | AGCTGCTGTA |
| TGCTTACGCT | CGGGAGAAGGC | CCCCCTGATT | CCCCCTGAGT | TCCCGAGGCA |
| TGCCCACTTT | TGAAACCAGG | TACGTGTCGG | AGTTTCCAGC | CCAGAGGAGA |
| GTGCCTCTCT | GCTTTCCTCG | GCACCTCTGG | GTGAAGAAGT | TGTTCTCTAA |
| CCAGAATCAG | ACTCAGAGG | TCCCTTTTTG | GGACAGTGT | GGCCTGCGG |
| CTCCTCCCCG | TATCATCAAG | GCCATGGTGA | GGAAGACCAC | AGGGGAACCT |
| CAGTATCAGT | GGGAGAGGCG | AGCTCCAGAA | ATCGAGATTT | TCCCTGAGGT |
| CGAAATCTCCG | TATGCGCCCA | GAGATCCCCA | GAATCCATCT | GGTCCCAAGG |
| TCATACTGGT | GATGCGCAAC | GAAACTCTGCT | GCCCTGCTCT | CGAGACGCTC |
| CACTCACTCT | CTGCTCTCGC | CCAAGCTCAG | TGTGCTGAGC | CCCAATGCCC |
| CGTGCAAGAT | GGACCAAGAC | AGACCTCCCC | AAGTAGAGAA | GCTTCTGCTC |
| TGACAGCAGA | GGGTGGAGAG | TGCCCTCTTC | CAGGACTCCA | GCCTGGCAAC |
| TCTACTGCGC | TGCCAGTGGC | CAGCGGACTT | GATGGGATCT | CCTCTGGTGG |
| CTCCTGGGGA | TCCTGTCGCC | TCCCTGTGAC | TGGAGACCTG | CCTGGGAGATG |
| CAGTGGGACT | TGGACTGCAA | TGCTTACTTC | TGGACCTGAA | GAATGGTACC |
| TGCTAATGGC | AGCAACAGGA | CCACTGTCAGG | TCCCAAGGCT | TTCTCTCACA |
| CAGCAGGCGA | CGGTGCTGCC | CCAAGACAGC | GTACCCCCATC | TGGAGAGACT |

### FIGURE 5
ISOFORMS OF THE MPL PROTEIN

MPSWALFMVT SCLLLAPQNL AQVSSQDVSL LASDSEPLKC FSRTFEDLTC
FWDEEEAAPS GTYQLLYAYP REKPRACPLS SQSMPHFGTR YVCQFPDQEE 100
VRLFFPLHLW VKNVFLNQTR TQRVLFVDSV GLPAPPSIIK AMGGSQPGE

QISWEEPAPF ISDFLRYELR YGPRDPKNST GPTVIQLIAT ETCCPALQRP 200
HSASALDQSP CAQPTMPWQD GPKQTSPSRE ASALTAEYGGS CLISGLQPQG
SYWLQLRSEP DGISLGGSGW SWSLPVTVDL PGDAVALGLQ CFTLDLKNVT 300
CQWQQQDHAS SQGFFYHSRA RCCPRDYPPI WENCEEEEKT NPGLQTPQFS

RCHFKSRNDS IIHILVEVTN APGTVHSYLG SPFWIHQAVR LPTPNLHWRE 400
ISSGHLELEW QHPSSWAQQE TCYQLRYTGE GHQDWKVLEP PLGARGGTL
LRPRSRYRLQ LRARLNGPQ YGQPWSSWDP TVETATETA WISLVTALHL 500
VLGLSAVLGL LLLRWQFPAH YRRLRHALWP SLFDLHRVLQ QYLRTDAALS
PPKATVSDTC EEEVEPSLLEI LPKSSERTPL PLCSSQAQMD YRRLQPSCLG 600
TMPLSVCPPM AESGSCCTTH IANHSYLPLS YWQQP 635

FIGURE 6
SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
    Chew, Anne
    Choi, Julie Y.
    Koshy, Beena
    Stephens, J. Claiborne

<120> HAPLOTYPES OF THE MPL GENE

<130> MWH-0173PCT MPL

<140> TBA
<141> 2001-04-16

<150> 60/197,839
<151> 2000-04-14

<160> 104

<170> PatentIn Ver. 2.1

<210> 1
<211> 3946
<212> DNA
<213> Homo sapien

<220>
<221> unsure
<222> (2026)...(2027)
<223> Nucleotide identity unknown

<400> 1
caaggggca gggtaaggag tgtgagcctc ctccaatctg agcaaacaag ataaagtac 60
tgaagacctc ttgcttcctca atcttaagtt acattcaaac acatctggga ttctggcaaa 120
atgcaggttc aaccttaata ggcctgggtt ggggcctaaag atctctcatt tctataaatgt 180
gccccctctaattgct tgtgcttctc cccgttcataag caccacatttt tgtgtagcag aggattaga 240
ggcaagccgtg gagcccccaca ataatggaga gggcttttccc agttccaggg agaatggtttt 300
tattcatgcc caaagtcaca gaactataa agtgaaataa cccaccacattt ggcattgttgg 360
tggggcagac agttatatact cagcagatga cccctccacag agcataagcgac gcatgcctgtgctg 420
ttatgcaagtt ggtttctgaa ggttgaaacct ccttgagcttt atgtcttaagt atgtaatctt 480
gggcaattct tttaaccttctc tggatatcgcg ttccttcatc tcataaatgg aaataacaag 540
actattgtgc ggatttgctta agtgaaatgg aataaaggtc tataacacag gcagctagatg 600
aatggctcag taaaaagtgtgc tatctgtagt ctattatatc atctataaattgc tctagcaca 660
gatacagcgg ctgtggttgcc ttttggtaca cggtaacattc ctaacaccccc atctccctca 720
cacccacacgg cggtgttcctc ccttttccaaa cttctctctcc ctcccttggcc actcccctctc 780
ccttgccccc tgcccccgacct tgtgtgctgaa tggccccagcc aggagcagc acagggacag 840
gacgtgggcc tgttatctgac aggaaccataa gggctggcct cgggggagga tgtgggccca 900
taccttccta acccctctct gtccccattg ggaatgcttt ggttttagttt agcccttaatc 3840
atgctcactgg gaccattgca acaatctctaca gtctaatctca cctctcaatgt attcccccatg 3900
cctgctaacag agtgtatcttt gttaagctcga aatcgtgcat gacctt 3946

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

<400> 2
aaaatgcaa aacaacgaa caaatgaca aagggaaaaa caaaaagaaa acaatcatggc 60
ataaaagcag agqggaaagct ctggagggccg cttcgaagaa gatgttaaat gataatcccc 120
gcaggccccac gtctctagat gatgagagggc cttctgtggg caagggctgat 180
tcaatgacct tgttggtgctgt ggttcttaggt atcccatctg gagaactctc gaagagggag 240
agaaacccct tccagactca cagacccccac atgcctctcg ctgctctacctc aagctacgaa 300
atgagccat cctctctccttg ccagccagcag cttctcctct catctctctcc 360
acctgggtctc cctctctctgtg atcccacagg ctggtgtaagaa cttctctcct ctctctctcc 420
catagttccc accccccactg aacctcgaccc tgtgcgaagcc atcccccactg ctgctctctc 480
tgacgcagttc cgctctgtggg acaaatgctct gtgcacagaga gacatctagc tgtgtctgctc 540
tgcacactcct gtatggtgcccc tccctccaccc acaactgcaac tggagggaga ctctcctgtg 600
gcatctggaa tggagtgccg agcaccctatc gtctctgggc gcccaagaga cctgttatca 660
acatccgactg aggagagacg gcatccagctct gtgaagagtt tggctaagca acaaactgccc 720
acagacccct tcagccaggc gactctgtggg gtgttgcagt ccgtttagca ggaagtgagag 780
tgcatcttgaa tcgcagctttctgt gctatgttat ctatggtttat cagatctacag tctctcctta 840
gtctctcttg ggcatctgtat attctctgctg actctctacttt aacattttct ctctctctttgt 900
ctttggtattc tcaagctctcc cccgatctttt cttctaccctt ctgacttaga tttctctactc 960
tctctctctctgtt ctccagccccct cccacaatgc tgtggctctct atgtctctat 1020
cctaggctct ctctctctctgt tgaatataaa atatgtctctct cccaggtgag tgtgttaact 1080
cctccctcata tcacatgggtc caaagatttc ctttctctttct tcccatatgg 1140
aaaaaat tctcctctctctgtg cggcaatgggg gaaatgctttt ggttttagttt agcccttaatc 1200
tgtgcagttgg tggataagct ctactcatttag gatgttaaat gataatcccc 1260
aaatataggc ataatcagca tggctgttcga ctggtgtaagaa cttctctctct ctctctctct 1320
gttgtctcattttctcctctgtt ggtcgcacata ctatgtacac aclatcattgc 1380
gtt 1383
<221> unsure
<223> Nucleotide identity unknown

<400> 3
cccttgcacta attaattgta gtagattatt acctgttttgg atgtcttttct cattagatat 60
gtaagagacg gttttttatt gtttttctcc cagtgctttaa ccaaatgttt gaggatcacc 120
acagcttata aatgtttgtct gctttctctgc tttgacactat gatgcttgta cccacgcgcc 180
taacggcaaat ctggcactccct ctagcagcatg atgtattatt tgggcatgctgt cttgtggcct 240
tcgtattatat ctgtttcttttg gtgggcttctcag cgcggtgagg gcagctttgtg 300
tagggtgtgc gttgacccggg ttcctctgcg cagttggtttgg tttgattgaga tttgatts 360
tgtagcttctt tgaggttggg tttgcttcttc aggaggtttgg cctggcactgt ttcttcttgttgcg 420
cagagcgtgca tcccgtttaaag gaggctcttttg gttaggggcc tctatctctct gtctgtggaaag 480
cgtgattctcc gccgggtttggc ctcttttggta gaactctccga cccgcttggtgg attcggaagt 540
gcaggatttg gtcgaacacag acgctggacct acgaagccgg cgcacggtgggcc cccggcgcac 600
ccttttttgt ctaggtctgttg cgcgtcccttc cggggccccc gagaggaccc cgggtgctgg 660
cccgcggctct cgctatggcctg tacaactgtgg cgccagggctc caagcggccca caaaccgaag 720
tccctggacca cggcggtggc agacaacttag gggccagacc ccagggggag cgggtgaggc 780
aagcgcgcgc cgcacaccaag gcggcgacag ccgtcgacag gactggccgc cgggtgagag 840
tggggcgcttg ctgcaagagc ggaggagagc ggagnggagc gcggcggggc ctgcaacttcgc 900
gggccagagc tgcgcaggttg cccgcaaggtgc cggggcgcgc gggggcggag cccgagagtg 960
gggctgcttgc gatgagggcg gggctcgcgc ccgggttgggc cgaagtctgta cccctttttgtg 1020
tctcttagctt gcagtagccttg cttgacgacgct tcctctctctg ctgctgctggct cgcgtggcctc 1080
tctggcggtcc tcgctggctg tggcaggttt cctgacacac acaggttacct cccggcgcag 1140
gcagagcttg gcggctggagc cagcggctttgc cagacacggc tgccagcccg ttccagtgcttc 1200
gcctgtgaccc cgcagatcttccg cggcgcgcgc ccgggtggcag ctcctgtctcc gcacactctg 1260
caacgtcacct ctgggactcg ggcgggtccttc ttaaatcaat cgttacaccc cagcggccga 1320
acaccaacaag gcgcgggcctct ctcgctggtcctc cctcggcgcac tagcgcagcc tccacccctcc 1380
agcgcgaactcg cccactcccc ccctcactcca gttctccac gcggtcgctca gttccacccct ctaaacactag 1440
tccaaacatg gcggcctcttt cccccctctac gcaagcagcc ccttgctctcc tggaaagagta 1500
cctcttcgcc cctcggtacta cgtacaccacgc caccgagaata atcttttcaaat cgtcagccac 1560
tgaacactac atctccctcc ttataatcttc gtcattacgc ccagagatata cctgcagagctcttctgctc 1620
ccctagatgac ccgtcgagcttc tctctgaaact cggcgggtggc cgcagcgacac tccggtctcgtc 1680
tctccctctgc gcacgacgcct tccaaacctgc gtttttctCCA caaagtcaggtt tttatctctga 1740
gagtgcctttg ccacataacacgccgtcctgcgtcctgc taagatagcc tccct 1785
<210> 5
<211> 1908
<212> DNA
<213> Homo sapien

<400> 5
atgcctctctt ggccctcttt cagtgtcacc tcttgctccc tcttgcccc ctaaaaacctg 60
gcctcaagtc gcagccacaga gtcgggtgct ggtggcatcag acctgaggcc cctgtaagttg 120
ttcctccgaa cattctgga gcctcacttc cctggggtatg aggaagagcc aagcgcaccgt 180
ggagataac agctgtgtaa tgctctaccg cgggagaagc cccggtcgttg cccctctgtg 240
tccagagacg tgcaggactt tggaggacca taagctgtggc tgtctccagc aaagagagag 300
gttctctctt tccttctgctt gcacaccttg gtgagaagag tgctctcataaa gcagactcgg 360
acctcagagag cccctctgtg tggacatgtg ggtggccagg ctgcctccagg tccctatcaag 420
gccatgtgggt gcagcagccgc aggggaacctt cagatcagct ggtggagag ccctcagagga 480
actagtcggtct cctctgtgag tcgacactcc ccctgcttcctt ccctgccctgaagaag 540
gctcccaaggg tcatacagctt gatgtgacca gcacacttgcct gcctctcgtct gcagaggctc 600
cacctgcctt cgggctctgga ccagctcactc tgtctcagac cccacaggcc ctgggcaagat 660
ggacacaccc agcaactcccc cagtagagag gttgctagct tcgacagcag ggtggaagac 720
tgctctacacct cagctggccac cctcacttgc ctgcagctgcag ccgcacaacc 780
gatgggtctct ccctctgggtgc tccctctgaggt tccctgctccc tccctgctctc tgtggagctg 840
cctcgagagt caagccacact ctgagctgca actccatctct ctgggctgaa gaaatgttacc 900
tgcaataaggcg agcagcagag ctgtgctgctt ctcagagcttc ctcctacactc ccagagacgc 960
cggtctgcgt ccagacagcaag ctgcctctgag tactacagcttc gcagagccgta aaataaaggt 1020
aatcagagcccc acctcagcact ctgcgactcact ctaagatcagc aatgagacacg 1080
attatttacc tctctgtgga ggtgcacacca gcgcggaggt tcggtccagc ccctctgctgc 1140
tccccnttttc ggtccagcctt gcctcgtgctgcct ccctctgctgct ccacacttgc cagagacgag 1200
atctctcagtc gccttgccatg acagaggttg cagcagccct cgggccttgac ccgaccagggt 1260
acctgtttacta aactccgata cacagagagac ggcctctcggc acgagaggttg gctcgagcggc 1320
cctctctggac ccggagagac gcacgctgga cccgagggata gacgctctgca ccggttacagc 1380
tgctccgcca gcgttcagcag ccggccttct cgagttccctgc ggtgagccggt cctgctctca 1440
actaggtttgg agacccgctc ctgctcactct gcagatctgctg ggtggtgctgg cctgctctca 1500
gtctgggccgc ttcagcgctct cttgggctcg tctgtggtctag gttggcgatt tccctgcacact 1560
tcagagacgc gcgagctggcgc cctgctgctag acctgaccac gcctgcctagc 1620
catcactata gcagatcctgag accagcagcc accacagccag gcagatatccgc acgatcctagt 1680
gaagagttcg aggccagcacct cttgagagat cttgctctgag cgtgctctgc gctgagtttttg 1740
ccctctgtgt ctccctccagct ccaagcagtgc tccagacagct cgggacctgc tggagcttgg 1800
accatgcgcccc tcctgtgtttgc cccacccatgc gcctagcagct gcctctgctg taccacccac 1860
attgcacacc actctcactct accacttaacg tattgagcagcg acgctgtga 1908
<table>
<thead>
<tr>
<th>Met</th>
<th>Pro</th>
<th>Ser</th>
<th>Trp</th>
<th>Ala</th>
<th>Leu</th>
<th>Phe</th>
<th>Met</th>
<th>Val</th>
<th>Thr</th>
<th>Ser</th>
<th>Cys</th>
<th>Leu</th>
<th>Leu</th>
<th>Leu</th>
<th>Ala</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Gln</td>
<td>Asn</td>
<td>Leu</td>
<td>Ala</td>
<td>Gln</td>
<td>Val</td>
<td>Ser</td>
<td>Ser</td>
<td>Gln</td>
<td>Asp</td>
<td>Val</td>
<td>Ser</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Asp</td>
<td>Ser</td>
<td>Glu</td>
<td>Pro</td>
<td>Leu</td>
<td>Lys</td>
<td>Cys</td>
<td>Phe</td>
<td>Ser</td>
<td>Arg</td>
<td>Thr</td>
<td>Phe</td>
<td>Glu</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Cys</td>
<td>Phe</td>
<td>Trp</td>
<td>Asp</td>
<td>Glu</td>
<td>Glu</td>
<td>Ala</td>
<td>Ala</td>
<td>Pro</td>
<td>Ser</td>
<td>Gly</td>
<td>Thr</td>
<td>Tyr</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Tyr</td>
<td>Ala</td>
<td>Tyr</td>
<td>Pro</td>
<td>Arg</td>
<td>Glu</td>
<td>Lys</td>
<td>Pro</td>
<td>Arg</td>
<td>Ala</td>
<td>Cys</td>
<td>Pro</td>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Gln</td>
<td>Ser</td>
<td>Met</td>
<td>Pro</td>
<td>His</td>
<td>Phe</td>
<td>Gly</td>
<td>Thr</td>
<td>Arg</td>
<td>Tyr</td>
<td>Val</td>
<td>Cys</td>
<td>Gln</td>
<td>Phe</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85</td>
<td></td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Gln</td>
<td>Glu</td>
<td>Glu</td>
<td>Val</td>
<td>Arg</td>
<td>Leu</td>
<td>Phe</td>
<td>Phe</td>
<td>Pro</td>
<td>Leu</td>
<td>His</td>
<td>Leu</td>
<td>Trp</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Val</td>
<td>Phe</td>
<td>Leu</td>
<td>Asn</td>
<td>Gln</td>
<td>Thr</td>
<td>Arg</td>
<td>Thr</td>
<td>Gln</td>
<td>Arg</td>
<td>Val</td>
<td>Leu</td>
<td>Phe</td>
<td>Val</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>115</td>
<td></td>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Gly</td>
<td>Leu</td>
<td>Pro</td>
<td>Ala</td>
<td>Pro</td>
<td>Ser</td>
<td>Ile</td>
<td>Ile</td>
<td>Lys</td>
<td>Ala</td>
<td>Met</td>
<td>Gly</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>130</td>
<td></td>
<td>135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Gln</td>
<td>Pro</td>
<td>Gly</td>
<td>Leu</td>
<td>Gln</td>
<td>Ile</td>
<td>Ser</td>
<td>Trp</td>
<td>Glu</td>
<td>Glu</td>
<td>Pro</td>
<td>Ala</td>
<td>Pro</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>145</td>
<td></td>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Ser</td>
<td>Asp</td>
<td>Phe</td>
<td>Leu</td>
<td>Arg</td>
<td>Tyr</td>
<td>Glu</td>
<td>Leu</td>
<td>Arg</td>
<td>Tyr</td>
<td>Gly</td>
<td>Pro</td>
<td>Arg</td>
<td>Asp</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>165</td>
<td></td>
<td>170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asn</td>
<td>Ser</td>
<td>Thr</td>
<td>Gly</td>
<td>Pro</td>
<td>Thr</td>
<td>Val</td>
<td>Ile</td>
<td>Gln</td>
<td>Leu</td>
<td>Ile</td>
<td>Ala</td>
<td>Thr</td>
<td>Glu</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>180</td>
<td></td>
<td>185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Cys</td>
<td>Pro</td>
<td>Ala</td>
<td>Leu</td>
<td>Gln</td>
<td>Arg</td>
<td>Pro</td>
<td>His</td>
<td>Ala</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Asp</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>195</td>
<td></td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Pro</td>
<td>Cys</td>
<td>Ala</td>
<td>Gln</td>
<td>Pro</td>
<td>Thr</td>
<td>Met</td>
<td>Pro</td>
<td>Thr</td>
<td>Met</td>
<td>Trp</td>
<td>Gln</td>
<td>Asp</td>
<td>Gly</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>210</td>
<td></td>
<td>215</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Ser</td>
<td>Pro</td>
<td>Ser</td>
<td>Arg</td>
<td>Glu</td>
<td>Ala</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Thr</td>
<td>Ala</td>
<td>Glu</td>
<td>Gly</td>
<td>Gly</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>220</td>
<td></td>
<td>225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Leu</td>
<td>Ile</td>
<td>Ser</td>
<td>Gly</td>
<td>Leu</td>
<td>Gln</td>
<td>Pro</td>
<td>Gly</td>
<td>Asn</td>
<td>Ser</td>
<td>Tyr</td>
<td>Trp</td>
<td>Leu</td>
<td>Gln</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>240</td>
<td></td>
<td>245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7
Arg Ser Glu Pro Asp Gly Ile Ser Leu Gly Gly Ser Trp Gly Ser Trp
  260  265  270
Ser Leu Pro Val Thr Val Asp Leu Pro Gly Asp Ala Val Ala Leu Gly
  275  280  285
Leu Gln Cys Phe Thr Leu Asp Leu Lys Asn Val Thr Cys Gln Trp Gln
  290  295  300
Gln Gln Asp His Ala Ser Gln Gly Phe Tyr His Ser Arg Ala
  305  310  315  320
Arg Cys Cys Pro Arg Asp Arg Tyr Pro Ile Trp Glu Asn Cys Glu Glu
  325  330  335
Glu Glu Lys Thr Asn Pro Gly Leu Gln Thr Pro Gln Phe Ser Arg Cys
  340  345  350
His Phe Lys Ser Arg Asn Ser Ser Ile Ile His Ile Leu Val Glu Val
  355  360  365
Thr Thr Ala Pro Gly Thr Val His Ser Tyr Leu Gly Ser Pro Phe Trp
  370  375  380
Ile His Gln Ala Val Arg Leu Pro Thr Pro Asn Leu His Trp Arg Glu
  385  390  395  400
Ile Ser Ser Gly His Leu Glu Leu Glu Trp Gln His Pro Ser Ser Trp
  405  410  415
Ala Ala Gln Gln Thr Cys Tyr Gln Leu Arg Tyr Thr Gly Glu Gly His
  420  425  430
Gln Asp Trp Lys Val Leu Glu Pro Pro Leu Gly Ala Arg Gly Gly Thr
  435  440  445
Leu Glu Leu Arg Pro Arg Ser Arg Tyr Arg Leu Gln Leu Arg Ala Arg
  450  455  460
Leu Asn Gly Pro Thr Tyr Gln Gly Pro Trp Ser Ser Trp Ser Asp Pro
  465  470  475  480
Thr Arg Val Glu Thr Ala Thr Glu Thr Ala Trp Ile Ser Leu Val Thr
  485  490  495
Ala Leu His Leu Val Leu Gly Leu Ser Ala Val Leu Gly Leu Leu Leu
  500  505  510
Leu Arg Trp Gln Phe Pro Ala His Tyr Arg Arg Leu Arg His Ala Leu
515 520 525

Trp Pro Ser Leu Pro Asp Leu His Arg Val Leu Gly Gln Tyr Leu Arg
530 535 540

Asp Thr Ala Ala Leu Ser Pro Pro Lys Ala Thr Val Ser Asp Thr Cys
545 550 555 560

Glu Glu Val Glu Pro Ser Leu Leu Glu Ile Leu Pro Lys Ser Ser Glu
565 570 575

Arg Thr Pro Leu Pro Leu Cys Ser Ser Gln Ala Gln Met Asp Tyr Arg
580 585 590

Arg Leu Gln Pro Ser Cys Leu Gly Thr Met Pro Leu Ser Val Cys Pro
595 600 605

Pro Met Ala Glu Ser Gly Ser Cys Thr Thr His Ile Ala Asn His
610 615 620

Ser Tyr Leu Pro Leu Ser Tyr Trp Gln Gln Pro
625 630 635

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

<400> 7
tattgtgmgg.attgt

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

<400> 8
aacacacr gc agcat

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

9
<210> 10
<211> 15
<212> DNA
<213> Homo sapien

<400> 10
tataatcrtt agcac

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

<400> 11
cctgaaaagt tttct

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

<400> 12
gaagaatr tg ttcct

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

<400> 13
gtagagartg atgct

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

<400> 14
atgagcaygc ctact
<210> 15
<211> 15
<212> DNA
<213> Homo sapien

<400> 15
gagatttygc aacaa 15

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

<400> 16
tgatgggmtc tccct 15

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

<400> 17
aggcacrgt gctgc 15

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

<400> 18
gtctcrtgrgg caggc 15

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

<400> 19
tctgggaraa ctcgc 15

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

<400> 20
tgctagtrta tctat

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

<400> 21
c tgctccycc acccc

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

<400> 22
tgccaatyca ctc gc

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

<400> 23
caccaacmct gcctg

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

<400> 24
agcgctctc ctc tcat

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

caagactatt gtgmg

acttagacaa tcckc

gtgcattaaca cacrg

cattactatgc tgcyg

ccattctata atcrt
<210> 31
<211> 15
<212> DNA
<213> Homo sapien

<400> 31
gtatctgtgc taayg 15

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

<400> 32
cagagccct gaakt 15

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

<400> 33
ttcgggagaa acamt 15

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

<400> 34
ctggtccctc aggvg 15

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

<400> 35
ccaccatgcg gacbc 15

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

<400> 36
ctgggtgaag aatrt

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

<400> 37
tgtttagga acaya

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

<400> 38
ccccaagtag agarg

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

<400> 39
aagtgacga tacyt

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

<400> 40
tcagatatga gcayg

<210> 41
<211> 15
<212> DNA
<213> Homo sapien
cccctaagta ggcr\textunderscore t

\textcolor{red}{<210> 42 15}
\textcolor{red}{<211> DNA 15}
\textcolor{red}{<212> Homo sapien 15}

\textcolor{red}{<400> 42 ttatggaga ttyg 15}

\textcolor{red}{<210> 43 15}
\textcolor{red}{<211> DNA 15}
\textcolor{red}{<212> Homo sapien 15}

\textcolor{red}{<400> 43 aaggttttgt tcg\textunderscore ra 15}

\textcolor{red}{<210> 44 15}
\textcolor{red}{<211> DNA 15}
\textcolor{red}{<212> Homo sapien 15}

\textcolor{red}{<400> 44 cgaacctgat gg\textunderscore m\textunderscore t 15}

\textcolor{red}{<210> 45 15}
\textcolor{red}{<211> DNA 15}
\textcolor{red}{<212> Homo sapien 15}

\textcolor{red}{<400> 45 ccac\textunderscore gc\textunderscore gagg ag\textunderscore akc 15}

\textcolor{red}{<210> 46 15}
\textcolor{red}{<211> DNA 15}
\textcolor{red}{<212> Homo sapien 15}

\textcolor{red}{<400> 46 cacagcagg gcacr\textunderscore g 15}
<210> 47
<211> 15
<212> DNA
<213> Homo sapien

<400> 47
tctggggcag cacyg 15

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

<400> 48
ggattagct ctgrg 15

<210> 49
<211> 15
<212> DNA
<213> Homo sapien

<400> 49
aactaggcct gccyc 15

<210> 50
<211> 15
<212> DNA
<213> Homo sapien

<400> 50
accccattctg ggara 15

<210> 51
<211> 15
<212> DNA
<213> Homo sapien

<400> 51
cctccccgca gtttyt 15

<210> 52
<211> 15
<212> DNA
<213> Homo sapien

<400> 52
agtcctgcgt agtrt

15

<210> 53
<211> 15
<212> DNA
<213> Homo sapien

<400> 53
ataaaacatag atayaa

15

<210> 54
<211> 15
<212> DNA
<213> Homo sapien

<400> 54
gcctcactgc tcyc
c

15

<210> 55
<211> 15
<212> DNA
<213> Homo sapien

<400> 55
gcagttggtg tggg
g

15

<210> 56
<211> 15
<212> DNA
<213> Homo sapien

<400> 56
cctccctgcc aatyc

15

<210> 57
<211> 15
<212> DNA
<213> Homo sapien

18
agccatggca gtgra

geccaccacc aacmc

cagtaccagg cagkg

taatccagcg cctst

ggagaatga ggasa

agctcctca ctcc
<210> 63
<211> 15
<212> DNA
<213> Homo sapien

<400> 63
gtccagggaa ctgrg 15

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

<400> 64
gactattgtg 10

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

<400> 65
tagacaatcc 10

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

<400> 66
ccttaacacac 10

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

<400> 67
actatgctgctgc 10

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

<400> 68
ttctataatc
10

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

<400> 69
tctgtgctaa
10

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

<400> 70
agccccctgaa
10

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

<400> 71
ggagaaacaca
10

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

<400> 72
ggtcctcagg
10

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

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

<400> 74
ggtgaagaat 10

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

<400> 75
tttaggaaca 10

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

<400> 76
caaagttagaga 10

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

<400> 77
gtcagcatac 10

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

<400> 78
gatatgagca 10
<210> 79
<211> 10
<212> DNA
<213> Homo sapien

<400> 79
cgaagttaggc 10

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

<400> 80
atggagatattt 10

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

<400> 81
gttttggttgc 10

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

<400> 82
acctgatggg 10

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

<400> 83
cggagggagaga 10

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

<400> 84
agcagggcac

<210> 85
<211> 10
<212> DNA
<213> Homo sapien

<400> 85
ggggcagcac

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

<400> 86
ttagtctctg

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

<400> 87
caggcctgcc

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

<400> 88
ccatctggga

<210> 89
<211> 10
<212> DNA
<213> Homo sapien
<210> 90
<211> 10
<212> DNA
<213> Homo sapien

<400> 90
cctgcagtt

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

<400> 91
aacatagata

<210> 92
<211> 10
<212> DNA
<213> Homo sapien

<400> 92
tcactgcctcc

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

<400> 93
gttggggtgg

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

<400> 94
cctgcacaaat

25
<210> 95
<211> 10
<212> DNA
<213> Homo sapien

<400> 95
catggcagtg

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

<400> 96
caccaccaaac

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

<400> 97
taccaggcag

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

<400> 98
tccagcgcct

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

<400> 99
gagatgagga

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

<400> 100
ctcctcaactc

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

<400> 101
cagggaactg

<210> 102
<211> 3946
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (550)
<223> PS1: Polymorphic base C or A

<220>
<221> allele
<222> (590)
<223> PS2: Polymorphic base G or A

<220>
<221> allele
<222> (652)
<223> PS3: Polymorphic base G or A

<220>
<221> allele
<222> (1238)
<223> PS4: Polymorphic base G or T

<220>
<221> allele
<222> (1598)
<223> PS5: Polymorphic base C or G or A

<220>
<221> allele
<222> (1771)
<223> PS6: Polymorphic base G or A

<220>
<221> unsure
<222> (2026)...(2027)
<223> Nucleotide identity unknown

<220>
<221> allele
<222> (2671)
<223> PS7: Polymorphic base A or G

<220>
<221> allele
<222> (2805)
<223> PS8: Polymorphic base C or T

<220>
<221> allele
<222> (3160)
<223> PS9: Polymorphic base C or T

<220>
<221> allele
<222> (3595)
<223> PS10: Polymorphic base A or C

<220>
<221> allele
<222> (3595)
<223> PS11: Polymorphic base G or A

<400> 102
caagggggca gggtaagggag tgtgagccat ctccaatctg agcaacagg ataaaaagtac 60
tgaagaccat tgtctctccaa accttaagtt acattcaac cactttggga ttittgcaa 120
tagcaggttc aactttaata ggcccgggtt ggggctaaag atttctgatt tctataaagt 180
gccacccaa tgtcgtgtcgt cctgtgtctac agaccacatt ttgagtagca aggattaga 240
gccaccccttg gaggccaccc ataatggaga gggcattccc agttcaggag agaatgttgtt 300	tattattcgcc caaagcgcac gaaactattaa agttgaaata ccaccacattt gcgcaggtttg 360
tggggcagag agttattttct cagcagatta gctcocaagag aggcatagcac agcatggctg 420
tttcgaggt ggctgctgaa cttgaaccc cctgtgcttt atgtcttagt atgtaatctt 480
ggcaacaact ttaactcttcct tgtatctcgt tttctctcttc tctataaatgg aataaacaag 540
actattttggt gatatgctta agtagcaga tataaagttgc ttaactccac gcgtatgtgt 600
aagtgctcag taaaagttgc tattctgtatg ctattataac atcttataat crttagcaca 660
gatagcaggg tgtagttgcccc tttggtctca aggtcaata tagaaaaaccctt atctcctcctc 720
caccaaaaag ccctgtgctgt cctctctcaaa ttctctcttc cctctgtgcc acctccccct 780
cctggcccc gttgtctgag tgggccccag agggggcagg acagggacag 840
<210>  103
<211>  1383
<212>  DNA
<213>  Homo sapien

<220>
<221>  allele
<222>  (168)
<223>  PS12: Polymorphic base A or G

<220>
<221>  allele
<222>  (224)
<223>  PS13: Polymorphic base G or A

<220>
<221>  allele
<222>  (807)
<223>  PS14: Polymorphic base A or G

<400>  103
aaaatgcgaaa acaaaacgaa caaatgaaca aaggaaaaa caaaaaagaa acatcatggc 60
ataaaagcac agggaaagct ctggaagccc ctggaagaaa ggaagtttaaat gataatcccct 120
gcaggcccc gcgtcttttag gatgggaagac cttggaggata gtctcttgagg cagggcttgat 180
tcgaatgtct tcgtaggtcgtc gttccttaggt accccctcgtg ggaraactgc gaagagggag 240
agaaaaaaca tccgggacta cagacccccac agtttcttcgg cttcaccactc aagtacagaa 300
atgacagact tatattccacct cttggaggag tcacacagag ccgcgggtacct gttcagcatg 360
acctgggtct ctctttcttg atccacaccagc ctggttaagaa cttctctctct cttctctcca 420
catggtttcc acctccctcgg aatctgaccc tggggccagg atcccccaatc ctgaccctgc 480
tgaccagatgg ctctggttgcg acaatgcctt gtgcacagaa ggaacttaagct tgcctccctg 540
tgacatccct gtagtcgccccc tcccccacccc aaacttgcacag tggagggaga tctccagttg 600
gcatctggaa ttggagttggcg agccacattg aacagtggcctt ctgctgcggca gccaaagaga cctggttatca 660
actccgatac acagagagag gcacatcagca tcgggaagta tgcgtaagca acaatgcccc 720
acagacctca ctagcagggc gatcctctgtc gttgacatgc cctgtaagca ggaagttggaag 780
tgtctatgta gctagtcctctgtcgtagtrtat cttatggttat cagatctaac agtgatctta 840
gtctctctttg ggcattcctg attttctggt ctacctctatc acaatttttcct ttctctttgt 900
ctttggtata tcagtctctcc ccggattttcct tccctctctc ctgactgacat ttctctactc 960
tctctctcttg gccctctcttc tccagctcttc cacaataatgc gttgcctcttt cagagctcctat 1020
cctaggtcttt ttctctctctt gataaaataa atagttcttc cccaggtagct gtcttttaacct 1080
cctatttcct aagtgcctata cattaacact cggaaatttt ccattctcactt tcccataagg 1140
aaagggatatt acaaaacccga ggcaatttaca caaaaatttga gaagactgttt tttctacttgg 1200
tgaatagctg atgaagcatatct caaattatgtc gatttagctt aagagataaa tactatatct 1260
aaatacctg ctacatgctg ggccataagg gaagttgctca ctatacattac ttatatatata 1320
gtttgctac tttatattct atttgcacaca catatataga actatgtaca actatctatg 1380
gtt 1383

<210> 104
<211> 2848
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (276)
<223> PS15: Polymorphic base C or T

<220>
<221> allele
<222> (309)
<223> PS16: Polymorphic base C or T

<220>
<221> allele
<222> (475)
<223> PS17: Polymorphic base C or A

<220>
<221> allele
<222> (636)
<223> PS18: Polymorphic base C or G

<220>
<221> allele
<222> (925)
<223> PS19: Polymorphic base C or T

<220>
<221> unsure
<222> (2755)
<223> Nucleotide identity unknown

<400> 104
cctgcagataa agagggagta cttgttttca ggcaagggag tcagggatga catttgaaca 60
gagggctaca atagaggt gaggaatgaatc cagtgttcca gacagacaac acattgga 120
gttgagggaa actgcaagac ccctttgtgga acctggccac ccactgacttc cctccctac 180
cattgcgtttt ctctttactc taccctcctg ccaactctta ttgctttgac catgtctta 240
tatctcccaag ccctactaccc cggcctcact gcctcyccac ccaactgtgt gctctctctcc 300
cgcatacygc actgcaatggt ctcagctctgct ttctctctct ctccccccca gagaactgaggg 360
catgatctgt ggccctcact tccagacctg caccgggtcc tagcccaagta cctaggac 420
actgcagcc cggagccggg tgggtgtcct cccctcccccc tgcaccaccc caacmetgcg 480
tgggtaactc tccctggcccc aacaatacaaa cctgttcaag gctcctgccc tccagagtgtg 540
ccatcctccc gacactccac agacactacc cagccctctct cctctctctgt cagtcacgcc 600
cctccctcca caggtacttg gcgctctctc cctctctctc cagcccaagcc 660
caagcttcac gataacctgt aattaaggtga accgaccccc tggtaacatcc tcccaagatct 720
ctcagaggtt actcctctcc acgcagctgt cccctcagcc cagatggact acgcagagt 780
gcagccctct tgcgtccgca acatgcctct cttctctctc ccccaagcttg ctggactcag 840
gcctgtctgt aaccccaacc cagagcctca cccaccaccct cctccacactct gcctgaccct 900
gcctgtctct tgggcaactc gctgtttttt gacgtagctcc cttctctctctt gcctctctct 960
tctgtgaact cccctcactc aaccccaaca cacagcctca cccctccactcc 1020
tctctgtctcg cccctcacaat agcgttcttt gtccgactgt ccctctctctca gctctgctgca 1080
taaacacag ctctcttttc cacagcagctc ctctctctct caagctctcc ttttctcccc 1140
tctctctctct tcctgcttta cttatttattc tttttatttt ttttttatttt ttttttatttt 1200
ttaactttgta gactacacctc attagttcccc cttactacact ttgctagtga ctctgccccagg 1260
caaatgtcgc cttccctct ccctctctcta atctcctcctc gcacagctcc 1320
cctctgtttga cgggcacctgc gtgcttgcttt ccgccctctcc cggacgctgctt gctctgctcatactct 1380
ccctctctctc cggacgctgctt tagatatag tatagtgtt gttgtgagctc gagagccagt 1440
ggcaccttcgc tgcctacacc gctgagccagt aatggcgagaga tccctgctctt cctgcaacccctc 1500
caccttggggt cttccagctc cggccctctt gctctcttgtt gctctgctcatactct 1560
cacacacacac acacacacac acacacacac acacacacac acacacacac 1620
ggagcgttcgg atgggctgctc cttcctcctt cccgctctctct ctcttctcctc cctttttttt 1680
tgcttttcttt cttgctccct cggccctctc cggccctctc ccctttttttt ctttcttcttt 1740
gccttccctg gatgcgtgtc cttctctctt cggccctctc tttttcttttt ttttttttttt ttttttttttt 1800
ggcgccttcc cttggtgagt cttctctctt cggccctctc tttttcttttt ttttttttttt 1860
actagtgtgtg gacgatccttt gccatctctc gacttttttt ttttttttctcttttttctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt