Title: HAPLOTYPES OF THE CTLA4 GENE

Abstract: Novel genetic variants of the Cytotoxic T-Lymphocyte-Associated Protein 4 (CTLA4) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the CTLA4 gene. Compositions and methods for haplotyping and/or genotyping the CTLA4 gene in an individual are also disclosed. Polynucleotides defined by the sequence the haplotypes disclosed herein are also described.
RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/206,353 filed May 23, 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 cytotoxic T-lymphocyte-associated protein 4 (CTLA4) 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 unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of autoimmune disorders is the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene or its encoded product. CTLA4 is involved in T-cell activation. It is highly homologous to CD28 and binds to the same ligands, however it is only found on cells that are activated after antigen presentation. Therefore, the CTLA4 gene is involved in both immune and autoimmune responses, and is likely to play a role in numerous T-cell mediated autoimmune disorders (Chistiakov et al., *BMC Genet* 2001; 2:6). CTLA4-deficient mice had higher serum immunoglobulin levels than normal mice and died at about 3-4 weeks of age due to
myocardial failure caused by T-cell blast infiltration (Waterhouse et al., Science 1995; 270:985-988). These results suggest that CTLA4 may play an inhibitory role in regulating lymphocyte expansion.

The cytotoxic T-lymphocyte-associated protein 4 gene is located on chromosome 2q33 and contains 4 exons that encode a 223 amino acid protein. Reference sequences for the CTLA4 gene (Genaissance Reference No. 743670; SEQ ID NO: 1), coding sequence (GenBank Accession No:NM_005214.1), and protein are shown in Figures 1, 2 and 3, respectively.

Several variations have been identified in the CTLA4 gene that may be related to various disorders. A polymorphism of adenine or guanine at nucleotide position 37902 in Figure 1 results in an amino acid variation of threonine or alanine at amino acid position 17 in Figure 3 (HGbase:Snp000000387). Donner et al., (J Clin Endocrinol. Metab 1997; 82:4130-4132) reported that patients with Hashimoto thyroiditis had higher frequencies of the Thr17Ala mutation. However, there was no significant variation in patients with Addison's Disease and control subjects. The Thr17Ala mutation has also been shown to be associated with Grave's Disease in a dataset of white Caucasian subjects (Heward et al., J Clin Endocrinol. Metab 1999; 84:2398-2401). Djilali-Saiah et al. (Gut 1998; 43:187-189) found that in French Caucasian patients with Celiac disease, which is characterized by immunologically mediated intestinal injury following ingestion of gluten, the Thr17Ala mutation was found with greater frequency in patients than in controls. These results suggest that the location of this polymorphism on the gene is critical to the function of the CTLA4 protein.

The association between Thr17Ala polymorphism and insulin dependent diabetes mellitus (IDDM) has also been studied in numerous populations. Associations have been observed several populations including Asian, Mexican-American, and certain Caucasian populations (Awata et al., Diabetes 1998; 47:128-129; Donner et al., J Clin Endocrinol. Metab 1997; 82:143-146; Lee et al., Clin Endocrinol. (Oxf) 2000; 52:153-157; Marron et al., Hum Mol. Genet 1997; 6:1275-1282), while a lack of association has been observed in other Caucasian groups (Owerbach et al., Diabetes 1997; 46:1069-1074). Chistiakov et al. tested whether or not the CTLA4 gene was a susceptibility marker for IDDM in a Russian population. In the case of the codon 17 polymorphism, the alanine allele was associated with the disease. The authors also examined an (AT)n microsatellite marker in the 3' untranslated region of the gene, and examined the transmission of 17 alleles varying from 92 to 130 bp in length. The transmission of three alleles was significantly different for diabetic and non-diabetic offspring (Chistiakov et al., supra). Therefore, the CTLA4 gene is associated with IDDM in a Russian population.

Because of the potential for variation in the CTLA4 gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in the CTLA4 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 CTLA4 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 6 novel polymorphic sites in the CTLA4 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 37535 (PS1), 38038 (PS3), 40867 (PS4), 41547 (PS5), 42460 (PS6) and 42508 (PS7). The polymorphisms at these sites are cytosine or thymine at PS1, adenine or guanine at PS3, thymine or cytosine at PS4, cytosine or thymine at PS5, adenine or guanine at PS6 and adenine or guanine at PS7. In addition, the inventors have determined the identity of the alleles at these sites, as well as at the previously identified site at nucleotide position PS2 (37902), 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 PS1-7 in the CTLA4 gene, which are shown below in Tables 4 and 3, respectively. Each of these CTLA4 haplotypes defines a naturally-occurring isoform (also referred to herein as an “isogene”) of the CTLA4 gene that exists in the human population. The frequency with which each haplotype and haplotype pair occurs within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the CTLA4 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, PS3, PS4, PS5, PS6 and PS7 in both copies of the CTLA4 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 CTLA4 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel CTLA4 polymorphic sites. In a preferred embodiment, the genotyping kit comprises a set of oligonucleotides designed to genotype each of PS1-7. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 4 below or has one of the haplotype pairs in Table 3 below.

The invention also provides a method for haplotyping the CTLA4 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the CTLA4 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7. In another embodiment, the haplotyping method comprises determining whether one copy of the individual’s CTLA4 gene is defined by one of the CTLA4 haplotypes shown in Table 4, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual’s CTLA4 gene are defined by one of the CTLA4 haplotype pairs shown in Table 3 below, or a sub-haplotype pair thereof. The method for establishing the CTLA4 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 CTLA4 activity, e.g., autoimmune disorders.
For example, the haplotyping method can be used by the pharmaceutical research scientist to validate CTLA4 as a candidate target for treating a specific condition or disease predicted to be associated with CTLA4 activity. Determining for a particular population the frequency of one or more of the individual CTLA4 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue CTLA4 as a target for treating the specific disease of interest. In particular, if variable CTLA4 activity is associated with the disease, then one or more CTLA4 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 CTLA4 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable CTLA4 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 CTLA4 haplotype or haplotype pair, apply the information derived from detecting CTLA4 haplotypes in an individual to decide whether modulating CTLA4 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting CTLA4 to treat a specific condition or disease predicted to be associated with CTLA4 activity. For example, detecting which of the CTLA4 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 CTLA4 isoforms present in the disease population. Thus, without requiring any a priori knowledge of the phenotypic effect of any particular CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute CTLA4 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 CTLA4 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 CTLA4 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a CTLA4 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the CTLA4 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the CTLA4 genotype or haplotype in a reference population. A higher frequency of the CTLA4 genotype,
haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the CTLA4 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 CTLA4 haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for autoimmune disorders.

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 CTLA4 gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, guanine at PS3, cytosine at PS4, thymine at PS5, guanine at PS6 and guanine at PS7. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of guanine at PS2.

A particularly preferred polymorphic variant is an isogene of the CTLA4 gene. A CTLA4 isogene of the invention comprises cytosine or thymine at PS1, adenine or guanine at PS2, adenine or guanine at PS3, thymine or cytosine at PS4, cytosine or thymine at PS5, adenine or guanine at PS6 and adenine or guanine at PS7. The invention also provides a collection of CTLA4 isogenes, referred to herein as a CTLA4 genome anthology.

Polynucleotides complementary to these CTLA4 genomic DNA variants are also provided by the invention. It is believed that polymorphic variants of the CTLA4 gene will be useful in studying the expression and function of CTLA4, and in expressing CTLA4 protein for use in screening for candidate drugs to treat diseases related to CTLA4 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 CTLA4 for protein structure analysis and drug binding studies.

The present invention also provides nonhuman transgenic animals comprising one of the CTLA4 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the CTLA4 isogenes in vivo, for in vivo screening and testing of drugs targeted against CTLA4 protein, and for testing the efficacy of therapeutic agents and compounds for autoimmune disorders in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the CTLA4 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 CTLA4 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing CTLA4 haplotypes organized according to their evolutionary relationships.
BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the CTLA4 gene (Genaissance Reference No. 743670; 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:36 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the CTLA4 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the CTLA4 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the CTLA4 gene. As described in more detail below, the inventors herein discovered 8 isogenes of the CTLA4 gene by characterizing the CTLA4 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), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.
<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>20</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>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>British Isles</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>British Isles/Central</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>British Isles/Eastern</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Central/Eastern</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Eastern</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Central/Mediterranean</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mediterranean</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Scandinavian</td>
<td>2</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>Caribbean</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Caribbean (Spanish Descent)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Central American (Spanish Descent)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mexican American</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>South American (Spanish Descent)</td>
<td>4</td>
</tr>
</tbody>
</table>

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The CTLA4 isogenes present in the human reference population are defined by haplotypes for 7 polymorphic sites in the CTLA4 gene, 6 of which are believed to be novel. The CTLA4 polymorphic sites identified by the inventors are referred to as PS1-7 to designate the order in which they are located in the gene (see Table 2 below). Using the genotypes identified in the Index Repository for PS1-7 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the CTLA4 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the CTLA4 gene include those shown in Tables 3 and 4, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether CTLA4 is a suitable target for drugs to treat autoimmune disorders, screening for such drugs and reducing bias in clinical trials of such drugs.

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

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

**Candidate Gene** – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.
Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

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

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

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

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

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

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

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

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

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

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

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

Isogene - One of the isoforms 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 CTLA4 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 CTLA4 polymorphisms and haplotypes identified herein.

The compositions comprise at least one CTLA4 genotyping oligonucleotide. In one embodiment, a CTLA4 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term “oligonucleotide” refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptns, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a CTLA4 polynucleotide, i.e., a CTLA4 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-CTLA4 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 CTLA4 gene using the polymorphism information provided herein in conjunction with the known sequence information for the CTLA4 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 Ruanò et al., 87 Proc. Natl. Acad. Sci. USA 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3' terminal nucleotide, or preferably a 3' penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting CTLA4 gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

AGATCTCTTAAAGTGA (SEQ ID NO:4) and its complement,
CAGTCAARGCGAGTG (SEQ ID NO:5) and its complement,
CAGTCAGTGAGCCAC (SEQ ID NO:6) and its complement,
CTAGAATCGTGGCCA (SEQ ID NO:7) and its complement,
TTTAAATRGTCTGAAT (SEQ ID NO:8) and its complement, and
GCTGTGARCATTCAT (SEQ ID NO:9) and its complement.

A preferred ASO primer for detecting CTLA4 gene polymorphisms comprises a nucleotide
sequence, listed 5’ to 3’, selected from the group consisting of:

TTATCCAGATCCTYA (SEQ ID NO:10); TCATGGCACTTTTRA (SEQ ID NO:11);
TTTCCAGCTCAARG (SEQ ID NO:12); ATAAATCAGTCCCYT (SEQ ID NO:13);
CCATTCACCTGCTYT (SEQ ID NO:14); GCAACAGGTGTCACT (SEQ ID NO:15);
AACGCACTAGAACY (SEQ ID NO:16); TGGCAATGCTACR (SEQ ID NO:17);
ATATAAATTATAATRG (SEQ ID NO:18); TTCTTGATTCAGCYA (SEQ ID NO:19);
CTGTATGCTGTGARC (SEQ ID NO:20); and TTAAAAATGAATGYT (SEQ ID NO:21).

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 CTLA4 gene polymorphisms by
primer extension terminates in a nucleotide sequence, listed 5’ to 3’, selected from the group consisting of:

TCCAGATCCT (SEQ ID NO:22); TGTCAGTTT (SEQ ID NO:23);
CAGCAGTCAG (SEQ ID NO:24); AATCAGTCCT (SEQ ID NO:25);
TTTCCACTGAG (SEQ ID NO:26); ACGGTTGCTA (SEQ ID NO:27);
GCATCTAGAAC (SEQ ID NO:28); CAATGCTAC (SEQ ID NO:29);
AAAATTGTATAT (SEQ ID NO:30); TTGGATTCAG (SEQ ID NO:31);
TATGCTGTGGA (SEQ ID NO:32); and AAAATGAATG (SEQ ID NO:33).

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.

CTLA4 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
cell (polymorphism detection assays, including but not limited to probe hybridization and polymerase
extension assays. Immobilized CTLA4 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 CTLA4 gene in an individual. As used herein, the terms “CTLA4 genotype” and “CTLA4 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 CTLA4 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 CTLA4 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, PS3, PS4, PS5, PS6 and PS7 in the two copies to assign a CTLA4 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 preferred embodiment of the genotyping method, the identity of the nucleotide pair at PS2 is also determined. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-7.

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 CTLA4 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 CTLA4 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 CTLA4 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, PS3, PS4, PS5, PS6 and PS7 in that copy to assign a CTLA4 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the CTLA4 gene or fragment such as one of the methods described above for preparing CTLA4 isogenes, with targeted in vivo cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only
provide haplotype information on one of the two CTLA4 gene copies present in an individual. If haplotype information is desired for the individual’s other copy, additional CTLA4 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 CTLA4 gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at PS2. In a particularly preferred embodiment, the nucleotide at each of PS1-7 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the CTLA4 haplotypes shown in Table 4. This can be accomplished by identifying, for one or both copies of the individual’s CTLA4 gene, the phased sequence of nucleotides present at each of PS1-7. The present invention also contemplates that typically only a subset of PS1-7 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 PNAS 97:10483-10488; Rieder MJ et al. 1999 Nature Genetics 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, Mol. Diag. 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 Genetic Data Analysis II, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a CTLA4 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, PS3, PS4, PS5, PS6 and PS7 in each copy of the CTLA4 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 PS1-7 in each copy of the CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 haplotype pair is predicted from its CTLA4 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a CTLA4 genotype for the individual at two or more CTLA4 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing CTLA4 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 CTLA4 haplotype pairs shown in Table 3.
Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by \( n = \frac{\log(1-q)}{\log(1-p)} \) 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 \( p_{H-H} (H_1 / H_2) = 2p(H_1)p(H_2) \) if \( H_1 \neq H_2 \) and \( p_{H-H} (H_1 / H_2) = p(H_1)p(H_2) \) 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 CTLA4 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is
preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., supra). A preferred process for predicting CTLA4 haplotype pairs from CTLA4 genotypes is described in U.S. Provisional Application Serial No. 60/198,340 and the corresponding International Application filed April 18, 2001.

The invention also provides a method for determining the frequency of a CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 genotype, haplotype or haplotype pair. Preferably, the CTLA4 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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, PS3, PS4, PS5, PS6 and PS7, and may also comprise an additional polymorphism of guanine at PS2. Similarly, the nucleotide sequence of a variant fragment of the
CTLA4 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 CTLA4 gene, which is defined by haplotype 1, (or other reported CTLA4 sequences) or to portions of the reference sequence (or other reported CTLA4 sequences), except for genotyping oligonucleotides as described above.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS1, guanine at PS3, cytosine at PS4, thymine at PS5, guanine at PS6 and guanine at PS7. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the CTLA4 gene which is defined by any one of haplotypes 2-8 shown in Table 4 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the CTLA4 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.

CTLA4 isogenes may be isolated using any method that allows separation of the two “copies” of the CTLA4 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 Ruano et al., Proc. Natl. Acad. Sci. 87:6296-6300, 1990; and allele specific PCR (Ruano et al., 1989, supra; Ruano et al., 1991, supra; Michalatos-Beloin et al., supra).

The invention also provides CTLA4 genome anthologies, which are collections of CTLA4 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 CTLA4 genome anthology may comprise individual CTLA4 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 CTLA4 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 CTLA4 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded CTLA4 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 CTLA4 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 used herein, a polymorphic variant of a CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the CTLA4 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 CTLA4 isogene. Expression of a CTLA4 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 CTLA4 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of CTLA4 mRNA transcribed from a particular isogene.

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

Effect(s) of the polymorphisms identified herein on expression of CTLA4 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 isogene is introduced into a cell in such a way that it recombines with the endogenous CTLA4 gene present in the cell. Such recombination requires the occurrence of a double recombination event,
thereby resulting in the desired CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 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 CTLA4 isogene and producing human CTLA4 protein can be used as biological models for studying diseases related to abnormal CTLA4 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 CTLA4 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel CTLA4 isogenes; an antisense oligonucleotide directed against one of the novel CTLA4 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel CTLA4 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 CTLA4 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 CTLA4 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The CTLA4 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer’s hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

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

EXAMPLES

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

EXAMPLE 1

This example illustrates examination of various regions of the CTLA4 gene for polymorphic sites.

Amplification of Target Regions

The following target regions were amplified using either the PCR primers represented below or 'tailed' PCR primers, each of which includes a universal sequence forming a noncomplementary 'tail' attached to the 5' end of each unique sequence in the PCR primer pairs. The universal 'tail' sequence for the forward PCR primers comprises the sequence 5'-TGTTAAGCAGCTGAGCCAGT-3' (SEQ ID NO:34) and the universal 'tail' sequence for the reverse PCR primers comprises the sequence 5'-AGGAAACAGCTATGACCAT-3' (SEQ ID NO:35). The nucleotide positions of the first and last nucleotide of the forward and reverse primers for each region amplified are presented below and correspond to positions in Figure 1.

<table>
<thead>
<tr>
<th>Fragment No.</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>PCR Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment 1</td>
<td>37286-37308</td>
<td>complement of 37909-37889</td>
<td>624 nt</td>
</tr>
<tr>
<td>Fragment 2</td>
<td>37608-37630</td>
<td>complement of 38295-38273</td>
<td>688 nt</td>
</tr>
<tr>
<td>Fragment 3</td>
<td>37690-37711</td>
<td>complement of 38108-38088</td>
<td>419 nt</td>
</tr>
<tr>
<td>Fragment 4</td>
<td>40310-40333</td>
<td>complement of 40820-40798</td>
<td>511 nt</td>
</tr>
<tr>
<td>Fragment 5</td>
<td>40466-40488</td>
<td>complement of 41010-40988</td>
<td>545 nt</td>
</tr>
<tr>
<td>Fragment 6</td>
<td>41123-41147</td>
<td>complement of 41644-41621</td>
<td>522 nt</td>
</tr>
<tr>
<td>Fragment 7</td>
<td>42294-42315</td>
<td>complement of 42918-42896</td>
<td>625 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 μl |
| 10 x Advantage 2 Polymerase reaction buffer (Clontech) | = 1 μl |
| 100 ng of human genomic DNA | = 1 μl |
| 10 mM dNTP               | = 0.4 μl |
| Advantage 2 Polymerase enzyme mix (Clontech)          | = 0.2 μl |
| Forward Primer (10 μM)  | = 0.4 μl |
| Reverse Primer (10 μM)  | = 0.4 μl |
| Water                   | = 6.6μl |

Amplification profile:
97°C - 2 min. 1 cycle
97°C - 15 sec.
70°C - 45 sec.  
72°C - 45 sec.  
97°C - 15 sec.  
64°C - 45 sec.  
72°C - 45 sec.  
\{  
\}  
35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 μl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 μl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using either the primer sets represented below with the positions of their first and last nucleotide corresponding to positions in Figure 1, or the appropriate universal ‘tail’ sequence as a primer. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

<table>
<thead>
<tr>
<th>Fragment No.</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment 1</td>
<td>37358-37377</td>
<td>complement of 37882-37863</td>
</tr>
<tr>
<td>Fragment 2</td>
<td>37637-37656</td>
<td>complement of 38175-38155</td>
</tr>
<tr>
<td>Fragment 3</td>
<td>Tailed Seq</td>
<td></td>
</tr>
<tr>
<td>Fragment 4</td>
<td>40349-40368</td>
<td>complement of 40789-40769</td>
</tr>
<tr>
<td>Fragment 5</td>
<td>Tailed Seq.</td>
<td></td>
</tr>
<tr>
<td>Fragment 6</td>
<td>41162-41182</td>
<td>complement of 41565-41546</td>
</tr>
<tr>
<td>Fragment 7</td>
<td>42340-42361</td>
<td>complement of 42868-42850</td>
</tr>
</tbody>
</table>

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the CTLA4 gene are listed in Table 2 below.

Table 2. Polymorphic Sites Identified in the CTLA4 Gene

<table>
<thead>
<tr>
<th>Polymorphic Site Number</th>
<th>PolyIda</th>
<th>Nucleotide Position</th>
<th>Reference Allele</th>
<th>Variant Allele</th>
<th>Variant CDS Variant AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS1</td>
<td>743786</td>
<td>37535</td>
<td>C</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>PS2R</td>
<td>743788</td>
<td>37902</td>
<td>A</td>
<td>G</td>
<td>49</td>
</tr>
<tr>
<td>PS3</td>
<td>743794</td>
<td>38038</td>
<td>A</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>PS4</td>
<td>743804</td>
<td>40867</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>PS5</td>
<td>743808</td>
<td>41547</td>
<td>C</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>PS6</td>
<td>743812</td>
<td>42460</td>
<td>A</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>PS7</td>
<td>743814</td>
<td>42508</td>
<td>A</td>
<td>G</td>
<td></td>
</tr>
</tbody>
</table>

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

rPreviously identified in literature

EXAMPLE 2

This example illustrates analysis of the CTLA4 polymorphisms identified in the Index
Repository for human genotypes and haplotypes.

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

<table>
<thead>
<tr>
<th>Genotype Number</th>
<th>PS1</th>
<th>PS2</th>
<th>PS3</th>
<th>PS4</th>
<th>PS5</th>
<th>PS6</th>
<th>PS7</th>
<th>HAP Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>A</td>
<td>G/A</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A/G</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>C/T</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>A/G</td>
<td>G</td>
<td>T/C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>C/T</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A/G</td>
<td>A</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>C/T</td>
<td>G/A</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>A/G</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>2</td>
</tr>
</tbody>
</table>

The haplotype pairs shown in Table 3 were estimated from the unphased genotypes using a computer-implemented extension of Clark’s algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in U.S. Provisional Application Serial No. 60/198,340 entitled “A Method and System for Determining Haplotypes from a Collection of Polymorphisms” and the corresponding International Application filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family) and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 8 human CTLA4 haplotypes shown in Table 4 below.
Table 4. Haplotypes Identified in the CTLA4 Gene

<table>
<thead>
<tr>
<th>HAP No.</th>
<th>HAP ID</th>
<th>Polymorphic Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PS PS PS PS PS PS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1  2  3  4  5  6  7</td>
</tr>
<tr>
<td>1</td>
<td>745030</td>
<td>C   A   A   T   C   A   A</td>
</tr>
<tr>
<td>2</td>
<td>745025</td>
<td>C   A   G   T   C   A   A</td>
</tr>
<tr>
<td>3</td>
<td>745029</td>
<td>C   A   G   T   C   A   G</td>
</tr>
<tr>
<td>4</td>
<td>745031</td>
<td>C   A   G   T   C   G   A</td>
</tr>
<tr>
<td>5</td>
<td>745028</td>
<td>C   A   G   T   T   A   A</td>
</tr>
<tr>
<td>6</td>
<td>745032</td>
<td>C   G   G   C   C   A   A</td>
</tr>
<tr>
<td>7</td>
<td>745026</td>
<td>C   G   G   T   C   A   A</td>
</tr>
<tr>
<td>8</td>
<td>745027</td>
<td>T   A   G   T   C   A   A</td>
</tr>
</tbody>
</table>

Table 5 below shows the percent of chromosomes characterized by a given CTLA4 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given CTLA4 haplotype pair is shown in Table 6. In Tables 5 and 6, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 5 and 6 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and NA = Native American.
Table 5. Frequency of Observed CTLA4 Haplotypes In Unrelated Individuals

<table>
<thead>
<tr>
<th>HAP No.</th>
<th>HAP ID</th>
<th>Total</th>
<th>CA</th>
<th>AF</th>
<th>AS</th>
<th>HL</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>745030</td>
<td>0.61</td>
<td>0.0</td>
<td>2.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>745025</td>
<td>48.78</td>
<td>54.76</td>
<td>50.0</td>
<td>25.0</td>
<td>63.89</td>
<td>66.67</td>
</tr>
<tr>
<td>3</td>
<td>745029</td>
<td>0.61</td>
<td>2.38</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>745031</td>
<td>0.61</td>
<td>0.0</td>
<td>2.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>745028</td>
<td>0.61</td>
<td>0.0</td>
<td>2.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>745032</td>
<td>0.61</td>
<td>0.0</td>
<td>2.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>745026</td>
<td>42.07</td>
<td>38.1</td>
<td>40.0</td>
<td>60.0</td>
<td>30.56</td>
<td>33.33</td>
</tr>
<tr>
<td>8</td>
<td>745027</td>
<td>6.1</td>
<td>4.76</td>
<td>0.0</td>
<td>15.0</td>
<td>5.56</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 6. Frequency of Observed CTLA4 Haplotype Pairs In Unrelated Individuals

<table>
<thead>
<tr>
<th>HAP1</th>
<th>HAP2</th>
<th>Total</th>
<th>CA</th>
<th>AF</th>
<th>AS</th>
<th>HL</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>23.17</td>
<td>23.81</td>
<td>25.0</td>
<td>5.0</td>
<td>38.89</td>
<td>33.33</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>19.51</td>
<td>14.29</td>
<td>25.0</td>
<td>35.0</td>
<td>5.56</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.22</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1.22</td>
<td>4.76</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>1.22</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>1.22</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>6.1</td>
<td>9.52</td>
<td>0.0</td>
<td>10.0</td>
<td>5.56</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1.22</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>6.1</td>
<td>0.0</td>
<td>20.0</td>
<td>5.56</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>39.02</td>
<td>47.62</td>
<td>30.0</td>
<td>30.0</td>
<td>44.44</td>
<td>66.67</td>
</tr>
</tbody>
</table>

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the CTLA4 gene are likely to be similar to the relative frequencies of these CTLA4 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.
As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

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

1. A method for haplotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, which comprises determining which of the CTLA4 haplotypes shown in the table immediately below defines one copy of the individual’s CTLA4 gene, wherein each of the CTLA4 haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

<table>
<thead>
<tr>
<th>Haplotype Numbera</th>
<th>Nucleotide Positionb</th>
<th>SEQ ID No.</th>
<th>Region Examinedc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PS No.b</td>
<td>37535</td>
<td>32786 - 37909</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>37902</td>
<td>37909</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>38038</td>
<td>37608 - 38295</td>
</tr>
<tr>
<td>T</td>
<td>3</td>
<td>40687</td>
<td>40466 - 41010</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>41547</td>
<td>41123 - 41644</td>
</tr>
<tr>
<td>A</td>
<td>6</td>
<td>42460</td>
<td>42294 - 42918</td>
</tr>
<tr>
<td>A</td>
<td>7</td>
<td>42508</td>
<td>42294 - 42918</td>
</tr>
</tbody>
</table>

aAlleles for haplotypes are presented 5' to 3' in each column
bPS = polymorphic site;
cLocation of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
dRegion examined represents the nucleotide positions defining the start and stop positions of the sequenced region.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-7 on the one copy of the individual’s CTLA4 gene.

3. A method for haplotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, which comprises determining which of the CTLA4 haplotype pairs shown in the table immediately below defines both copies of the individual’s CTLA4 gene, wherein each of the CTLA4 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

<table>
<thead>
<tr>
<th>Haplotype Pairsa</th>
<th>Nucleotide Positionb</th>
<th>SEQ ID No.</th>
<th>Region Examinedc</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/2</td>
<td>PS No.b</td>
<td>37535</td>
<td>32786 - 37909</td>
</tr>
<tr>
<td>C/C</td>
<td>1</td>
<td>37902</td>
<td>37909</td>
</tr>
<tr>
<td>A/A</td>
<td>2</td>
<td>38038</td>
<td>37608 - 38295</td>
</tr>
<tr>
<td>G/G</td>
<td>3</td>
<td>40687</td>
<td>40466 - 41010</td>
</tr>
<tr>
<td>T/T</td>
<td>4</td>
<td>41547</td>
<td>41123 - 41644</td>
</tr>
<tr>
<td>C/C</td>
<td>5</td>
<td>42460</td>
<td>42294 - 42918</td>
</tr>
<tr>
<td>A/A</td>
<td>6</td>
<td>42508</td>
<td>42294 - 42918</td>
</tr>
</tbody>
</table>

aHaplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5' to 3' as 1st Nt/2nd Nt in each column, where Nt = nucleotide;
bPS = polymorphic site;
cLocation of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
dRegion examined represents the nucleotide positions defining the start and stop positions of
the sequenced region.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-7 on both copies of the individual’s CTLA4 gene.

5. A method for genotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, comprising determining for the two copies of the CTLA4 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:36.

6. The method of claim 5, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid mixture comprising both copies of the CTLA4 gene, or a fragment thereof, that are present in the individual;
   (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
   (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
   (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
   (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the CTLA4 gene present in the individual the identity of the nucleotide pair at each of PS1-7.

8. A method for haplotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual which comprises determining, for one copy of the CTLA4 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7, wherein the two or more PS have the location and alternative alleles shown in SEQ ID NO:36.

9. The method of claim 8, further comprising determining the identity of the nucleotide at PS2, which has the location and alternative alleles shown in SEQ ID NO:36.

10. The method of claim 8, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid sample containing only one of the two copies of the CTLA4 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 oligonucleotide.
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.

11. A method for predicting a haplotype pair for the cytotoxic T-lymphocyte-associated protein 4
(CTLA4) gene of an individual comprising:
(a) identifying a CTLA4 genotype for the individual, wherein the genotype comprises the
nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of
PS1, PS3, PS4, PS5, PS6 and PS7, having the location and alternative alleles shown in SEQ
ID NO:36;
(b) enumerating all possible haplotype pairs which are consistent with the genotype;
(c) comparing the possible haplotype pairs to the haplotype pair data set forth in the table
immediately below; and
(d) assigning a haplotype pair to the individual that is consistent with the data

<table>
<thead>
<tr>
<th>Haplotype Pairs&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PS No.&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Nucleotide Position&lt;sup&gt;c&lt;/sup&gt;</th>
<th>SEQ ID NO</th>
<th>Region Examined&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/2</td>
<td>C/C</td>
<td>1</td>
<td>37535</td>
<td>36</td>
</tr>
<tr>
<td>2/3</td>
<td>C/C</td>
<td>2</td>
<td>37902</td>
<td>36</td>
</tr>
<tr>
<td>2/5</td>
<td>G/G</td>
<td>3</td>
<td>38038</td>
<td>36</td>
</tr>
<tr>
<td>2/6</td>
<td>G/G</td>
<td>4</td>
<td>40867</td>
<td>36</td>
</tr>
<tr>
<td>2/8</td>
<td>T/T</td>
<td>5</td>
<td>41547</td>
<td>36</td>
</tr>
<tr>
<td>2/4</td>
<td>T/T</td>
<td>6</td>
<td>42460</td>
<td>36</td>
</tr>
<tr>
<td>7/7</td>
<td>A/A</td>
<td>7</td>
<td>42508</td>
<td>36</td>
</tr>
</tbody>
</table>

<sup>a</sup>Haplotype pairs are represented as 1<sup>st</sup> Haplotype/2<sup>nd</sup> Haplotype; with alleles of each haplotype
shown 5' to 3' as 1<sup>st</sup> Nt/2<sup>nd</sup> Nt in each column, where Nt = nucleotide;
<sup>b</sup>PS = polymorphic site;
<sup>c</sup>Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
<sup>d</sup>Region examined represents the nucleotide positions defining the start and stop positions of the
sequenced region.

12. The method of claim 11, wherein the identified genotype of the individual comprises the nucleotide
pair at each of PS1-7, which have the location and alternative alleles shown in SEQ ID NO:36.

13. A method for identifying an association between a trait and at least one haplotype or haplotype pair
of the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) 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 1-8 shown in the table presented immediately below, wherein each of the haplotypes
comprises a set of polymorphisms whose locations and identities are set forth in the table
immediately below:

35
and the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the CTLA4 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

<table>
<thead>
<tr>
<th>Haplotype Pairs&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PS No.&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Nucleotide Position&lt;sup&gt;c&lt;/sup&gt;</th>
<th>SEQ ID NO</th>
<th>Region Examined&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/2 7/7 2/1 2/3 2/5 2/6 2/8 2/4 7/8 2/7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C C/C C/C C/C C/T C/T C/T C/C</td>
<td>1</td>
<td>37535</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>A/A G/G A/A A/A A/G A/A A/G A/G</td>
<td>2</td>
<td>37902</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>G/G G/G G/A G/G G/G G/G G/G G/G</td>
<td>3</td>
<td>38038</td>
<td>36</td>
<td>37608 - 38295</td>
</tr>
<tr>
<td>T/T T/T T/T T/T T/C T/T T/T T/T</td>
<td>4</td>
<td>40867</td>
<td>36</td>
<td>40466 - 41010</td>
</tr>
<tr>
<td>C/C C/C C/C C/C C/C C/C C/C C/C</td>
<td>5</td>
<td>41547</td>
<td>36</td>
<td>41123 - 41644</td>
</tr>
<tr>
<td>A/A A/A A/A A/A A/A A/A A/A A/A</td>
<td>6</td>
<td>42460</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
<tr>
<td>A/A A/A A/A A/G A/A A/A A/A A/A</td>
<td>7</td>
<td>42508</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
</tbody>
</table>

<sup>a</sup>Haplotype pairs are represented as 1<sup>st</sup> Haplotype/2<sup>nd</sup> Haplotype; with alleles of each haplotype shown 5' to 3' as 1<sup>st</sup> N<sub>t</sub>/2<sup>nd</sup> N<sub>t</sub> in each column, where N<sub>t</sub> = nucleotide;

<sup>b</sup>PS = polymorphic site;

<sup>c</sup>Location of PS within the indicated SEQ ID NO, wherein N<sub>t</sub> = nucleotide;

<sup>d</sup>Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region,
the polymorphic site.

17. The composition of claim 16, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-9, the complements of SEQ ID NOS:4-9, and SEQ ID NOS:10-21.

18. The composition of claim 15, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.

19. The composition of claim 18, wherein the primer extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-33.

20. A kit for genotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS3, PS4, PS5, PS6 and PS7, having the location and alternative alleles shown in SEQ ID NO:36.

21. The kit of claim 20, which further comprises oligonucleotides designed to genotype PS2, having the location and alternative alleles shown in SEQ ID NO:36.

22. An isolated polymucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which comprises a cytotoxic T-lymphocyte-associated protein 4 (CTLA4) isogene, wherein the CTLA4 isogene is selected from the group consisting of isogenes 2-8 shown in the table immediately below and wherein each of the isogenes comprises the regions of the SEQ ID NOS shown in the table immediately below and wherein each of the isogenes 2-8 is further defined by the corresponding set of polymorphisms whose locations and polymorphisms are set forth in the table immediately below

<table>
<thead>
<tr>
<th>Isogene Number</th>
<th>PS No.</th>
<th>Nucleotide Position</th>
<th>SEQ ID NO</th>
<th>Region Examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C C C C C C C C</td>
<td>1</td>
<td>37535</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>A A A A A A G G A</td>
<td>2</td>
<td>37902</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>A G G G G G G G G</td>
<td>3</td>
<td>38038</td>
<td>36</td>
<td>37608 - 38295</td>
</tr>
<tr>
<td>T T T T T T T T</td>
<td>4</td>
<td>40867</td>
<td>36</td>
<td>40466 - 41010</td>
</tr>
<tr>
<td>C C C C C T C C C</td>
<td>5</td>
<td>41547</td>
<td>36</td>
<td>41123 - 41644</td>
</tr>
<tr>
<td>A A A G A A A A A</td>
<td>6</td>
<td>42460</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
<tr>
<td>A A G A A A A A A</td>
<td>7</td>
<td>42508</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
</tbody>
</table>

a Alleles for isogenes are presented 5' to 3' in each column
b PS = polymorphic site;

c Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
d Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region;

(b) a second nucleotide sequence which comprises a fragment of the first nucleotide sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at PS1, guanine at PS3, cytosine at PS4, thymine at PS5, guanine at
PS6 and guanine at PS7, wherein the selected polymorphism has the location set forth in the table immediately above; and

(c) a third nucleotide sequence which is complementary to the first or second nucleotide sequence.

23. The isolated polynucleotide of claim 22, which is a DNA molecule and comprises both the first and third nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

24. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 22, wherein the organism expresses a CTLA4 protein encoded by the first nucleotide sequence.

25. The recombinant organism of claim 24, which is a nonhuman transgenic animal.

26. The isolated polynucleotide of claim 22 which consists of the second nucleotide sequence.

27. A computer system for storing and analyzing polymorphism data for the cytotoxic T-lymphocyte-associated protein 4 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 haplotypes set forth in the table immediately below:

<table>
<thead>
<tr>
<th>Haplotype Number \ Haplotype Pair</th>
<th>PS No.</th>
<th>Nucleotide Position</th>
<th>SEQ ID NO</th>
<th>Region Examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/C C/C C/C C/C C/C C/T C/T C/T</td>
<td>1</td>
<td>37535</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>A/A A/A A/A A/G A/A A/G A/G</td>
<td>2</td>
<td>37902</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>A/G G/G G/G G/G G/G G/G G/G</td>
<td>3</td>
<td>38038</td>
<td>36</td>
<td>37608 - 38295</td>
</tr>
<tr>
<td>T/T T/T T/T T/T T/T T/T T/T</td>
<td>4</td>
<td>40867</td>
<td>36</td>
<td>40466 - 41010</td>
</tr>
<tr>
<td>C/C C/C C/T C/T C/C C/C C/C</td>
<td>5</td>
<td>41547</td>
<td>36</td>
<td>41123 - 41644</td>
</tr>
<tr>
<td>A/A A/A A/A A/A A/A A/A A/A</td>
<td>6</td>
<td>42460</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
<tr>
<td>A/A A/G A/A A/A A/A A/A A/A</td>
<td>7</td>
<td>42508</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
</tbody>
</table>

*a Alleles for haplotypes are presented 5' to 3' in each column.

b PS = polymorphic site;

*Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;

*d Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region;

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

<table>
<thead>
<tr>
<th>Haplotype Pairs</th>
<th>PS No.</th>
<th>Nucleotide Position</th>
<th>SEQ ID NO</th>
<th>Region Examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/C C/C C/C C/C C/C C/T C/T C/T</td>
<td>1</td>
<td>37535</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>A/A G/G A/A A/A A/G A/A A/G A/G</td>
<td>2</td>
<td>37902</td>
<td>36</td>
<td>37286 - 37909</td>
</tr>
<tr>
<td>G/G G/G G/A G/G G/G G/G G/G G/G</td>
<td>3</td>
<td>38038</td>
<td>36</td>
<td>37608 - 38295</td>
</tr>
<tr>
<td>T/T T/T T/T T/T T/T T/T T/T</td>
<td>4</td>
<td>40867</td>
<td>36</td>
<td>40466 - 41010</td>
</tr>
<tr>
<td>C/C C/C C/C C/C C/T C/T C/T C/T C/C C/C C/C C/C C/C C/C C/C C/C</td>
<td>5</td>
<td>41547</td>
<td>36</td>
<td>41123 - 41644</td>
</tr>
<tr>
<td>A/A A/A A/A A/A A/A A/G A/A A/A</td>
<td>6</td>
<td>42460</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
<tr>
<td>A/A A/A A/G A/A A/A A/A A/A A/A</td>
<td>7</td>
<td>42508</td>
<td>36</td>
<td>42294 - 42918</td>
</tr>
</tbody>
</table>
Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5’ to 3’ as 1st Nt/2nd Nt in each column, where Nt = nucleotide;

PS = polymorphic site;

Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;

Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region.

28. A genome anthology for the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene which comprises CTLA4 isogenes defined by any one of haplotypes 1–8 set forth in the table shown below:

<table>
<thead>
<tr>
<th>Haplotype Numbera</th>
<th>PS No. b</th>
<th>Nucleotide Positionb</th>
<th>SEQ ID NO</th>
<th>Region Examinedd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

* Alleles for haplotypes are presented 5’ to 3’ in each column
  bPS = polymorphic site;
  cLocation of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
  dRegion examined represents the nucleotide positions defining the start and stop positions of the sequenced region.
| ACTCAAAATTT  | CCCCAGGTTT  | CTCAGTAATC  | TGCTTGTTAT  | TATTTTTTCT  | 33900 |
| AATCAATATT  | AACACATTTA  | ATCTCTGAAT  | CATCTCCTTA  | TTTCCTCCTT  | 34000 |
| TAATCTAGAA  | TGGTTTCTCA  | GCTTTTCCCA  | TTTTTAATCT  | TTTTGTAGTG  | 34100 |
| TAACATTTTTT  | GAGAAGATCTA  | GCCAGATTTA  | TTTTCGAAGAT  | ATGGATTTGTT  | 34200 |
| CTGATATTTT  | CCTCAAGGTT  | AGATTGAGTT  | AAAACATTTT  | TGGCAATTAAC  | 34300 |
| TACACAAAGT  | TGTTGTGCTC  | TTCACAGGTT  | ATTTGAATCTG  | GAGCTAACCTT  | 34400 |
| ATGTCAGAAG  | AGCTGACACAT  | AGTAATCTCT  | ATGTTGGGATC  | ACTGGTTTAA  | 34500 |
| AGGCTAATCT  | GCTGATTTTA  | TCAATTATCT  | TTTCCTATTAT  | GATAATTACCA  | 34600 |
| ATGAATCTTT  | CACCCATCAG  | TGCCTCCTA  | GTTGCCCTAA  | CTAATGCCTC  | 34700 |
| TTTACCTCAG  | ATTTTTTTTT  | ACAACCATTT  | GAATTTTCTT  | GCTATGACAG  | 34800 |
| AATATTAGTTT  | ATAAAGAGCAT  | AATTTAGGAA  | TCTGTAAGAG  | CTGCTTTGGA  | 34900 |
| ATGCTTAAT  | ACAATCCCTA  | TCTCTTCCTTA  | TTTAATACGT  | TTTCCATTCA  | 35000 |
| TTTAACCTACC  | TATTTAACAAG  | ATGTATTAGG  | AATGCTGTTG  | TTGTGCTTGC  | 35100 |
| GCCGTGCAAA  | TTCTTCATCG  | TTGGCCCATAA  | TATGGCCCAA  | TATGCAATTT  | 35200 |
| TCACAGAAGAC  | CACACAGCTT  | CTGCTCCCCAA  | AGCTGCTGCA  | GAGACCTGGA  | 35300 |
| ACTCAATCAT  | CTTCCAGGTT  | ATCAATCCTG  | AATCCCTGCT  | TTTTTGATTAC  | 35400 |
| ATCTCTACTC  | TACCTCCTCA  | CTATGGAGTA  | TATTTTCTAC  | TTGCTTTGAC  | 35500 |
| TGGTTTGATG  | GAGCTGAGCTG  | GCCAGCTGGC  | ACTCAGCCCT  | GTGGCAATTA  | 35600 |
| TCTTTGACCA  | AGATTTGATTT  | GAGAAGATTT  | AATTTTTTTT  | AGGCTAACTT  | 35700 |
| TGTGGCTGAAC  | GCCACACCC  | AGTACTTAGCAT  | TTTCCTATCT  | TATATTCGCA  | 35800 |
| TCAACATTTG  | GCCACCGACT  | GAGCTGATTTC  | TTCCCTAATT  | TTTCCTAATG  | 35900 |
| TGGCTGCTGC  | GCCAGATACT  | CTTTCTGCTG  | TTTCTCTGCTG  | TCTGCTCAAA  | 36000 |
| TACCACTCAG  | TAATGAGCATT  | GCCAGACATT  | CTTTCTGCTG  | GAGCTGCTG  | 36100 |
| TAGGGCCGCTT  | GGACCATGTT  | TTGAGGCTTT  | CTAATATGTT  | TAGCTGCTA  | 36200 |
| TAGAATACGAG  | GCAGTCCTGAC  | AAGCTGCTGT  | GTTCCTCCTC  | GCCACACCC  | 36300 |
**FIGURE 1C**

CTATCTCTCT CTCAGATCCT CTTGGTTAAG AACACATGTA GTTGTTATGA 38800
AGTAGGTACT CAAAGATGAT TTTGATGTAT AATTTTTAAT GGAGAGAAA 38900
TATCTATAA TGCCTTCTTA TTTTACATG TTTTAAATAT GTAAGATTA 38900
AATAAACACT CTTAGAAGTA TTTAATTTTC CTAAAGTAAA TTTTCTCAAA 39000
CCAGTACAG CACGCTACC CGTATCGAAA GTTGAGTGAG ACCGATTTTA 39000
GTGGGTAGTA GTGGTACCTT GCTTGGGAGG AGGGCCACG ATTTGGATTT 39100
TCTTGACCTT AGATCGAATA CAATGAAGAT TATGGATGAC TACCTTGGGT 39100
GAGCTTAAT CACTGTCGAC AGATGTCAT CAAACCACAG TAAAGAACGT 39200
AGGAGATGAT ATGATCTCAG AAAATACTTT AATACAGAGC AGAAGATAGAA 39200
GAACATCTCA AATGGATTA CACACTGCTT GAGATGCTC AGACGACCT 39400
TCTGATAGA AGAGGATGAG GATAGGAAGA GAGCTGTAAG TCTGCTAGAC 39400
TACACGCTGT CAGAATGAC CTTGCTGCTA GAAACATAGA 39500
CCAGATGTCG AGATGTTGAA TGGCATATG GGGCTCATTT GAAGTGTATT 39600
TATTATTTTC CTTGTGTCGA TCGATGTAG AAGATGTTCA 39600
GCAGATGAGT GACATCATGC AACAAAGTTT TTGGATTTTA CTTGTCAGAC 39600
TACATGAGAA GGGGCTCCTT TACCTGAAAG AGGTAATATT GCATGCGAC 39700
CATAATTTCG GATGATGAC AAGCTCTCTT GTATGGAGAC TCTGCTCTCT 39700
TTATCTGCTT GCAGTGTGCA CTTGCTACCA TCTGCTTCTA 39800
GAGAATTAA CCGAGAAGA CAGCACAGG AGTGGGATGT TACAAAGGAG 39800
AGTAAACACT CAGCTGTCCTT CACACTGAG CAGAAAGCAG TCAGTGCCCT 39900
CGCCCTGTGG GCTCTCTCCA TCTACAACCA TCCTCCACTC TACAACATGT 40000
CCAGGAACAG GACGATGCGC TACATGACCA CAGACAGACT CATCCTCAGG 40100
GTGACAGGAG TACGATGCA GCAGAGGCAG TAGGCAAGAC TTTTGACATG 40200
ATACTTTGAT GATATAGGAC AGATTTCTTC TCTAACCTCT GCAAATTTTC 40200
CCTACACATC AATTATGTTT CGGAGCTATA TAGCTCTATG CATCCTATCA 40400
TACACTGAGA AGATGCTACA TCACTCAATT ATCTCTCATT 40500
ATATGTCGAA AAAAACTGGA AGATGGGGGA GATGCTACTT GCCAAGCAATG 40600
ACATATTTAC GGTGAGAGCC AGGTCTCTTG TTTGCTCATAT CAGGTCTTCT 40600
CCTGCCACA CCAATTTGCA GAATCTATTT CTGACGTAAG AAAATATTTT 40700
ATGGAGAATC GTCGGAAACG AGTTGGTGAAG GGAGAGCGA GTGCGGATCT 40800
TGAGAGGGG AAGGGGTAAG TGGATAGATC GTTGAAGGGG GGAAGAAAGG 40800
CCGGTGGGAT GAGAAGTCAA GCGAGAAGGG TCTGCGCTGG GTGTGGCATG 40900
AAGGACATGAC GTGTCATCTG TGGCTCTTGG GCTTCTTCCAT GCTGACATGC 40900

*[@exon 2: 4045]*

CAGCTGCGCC AGCTGCTGCT GGTACTGCGC AGCAGCGAGA GCTACCGCCAG 40600
CTGCTGCGT CTGATGTCAT CTCCAGCGCA AGCCACTGAG GTCCGGGGTA 40600
CAGCTGTCGG GACGGTCCTG AGCCAGCTGA CGTACGCTCG TGGCCGCAAC 40700
TACATGATTG GAATAGTATT GACCCTCTCA GATGTATCCA TCTGACCGGG 40700
CACCCTGCGT GAAATACAG TGACCCTCCTA TATCCAAGGA CTTGAGGGCA 40800
TGGAGACCG ACCTACATCG TGCAAGGTGG AGCTCACTGA CCCACCGCCA 40800
TACTCCGCT GGATAGCCAC CGGAAACCGAG ATTTATGATA TGGGTGACGA 40800

..[4043]..
<table>
<thead>
<tr>
<th>Gene Region</th>
<th>Sequence</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGTCTCAGGG</td>
<td>GAGGCTCAGTCTTTTGTTCAGAGTCATCT</td>
<td>41300</td>
</tr>
<tr>
<td></td>
<td>TTGGTTTTCTTGTCCAGAGTTTCCAGAATGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGATTTTTTACGAGAAGGTGGTTTTGATT</td>
<td>41600</td>
</tr>
<tr>
<td></td>
<td>TAGAGCTGCTCATTCCATTTAGGAGAGGA</td>
<td>41700</td>
</tr>
<tr>
<td></td>
<td>AGATTTTTTACGAGAAGGTGGTTTTGATT</td>
<td>42000</td>
</tr>
<tr>
<td></td>
<td>TAGCTCTGGTCTTCTGGGCTCTCTCTCTCTC</td>
<td>42100</td>
</tr>
<tr>
<td></td>
<td>TAGCTCTGGGTCTTCTGGGCTCTCTCTCTC</td>
<td>42200</td>
</tr>
<tr>
<td></td>
<td>TAGCTCTGGGTCTTCTGGGCTCTCTCTCTC</td>
<td>42300</td>
</tr>
<tr>
<td></td>
<td>TAGCTCTGGGTCTTCTGGGCTCTCTCTCTC</td>
<td>42400</td>
</tr>
<tr>
<td></td>
<td>TAGCTCTGGGTCTTCTGGGCTCTCTCTCTC</td>
<td>42500</td>
</tr>
</tbody>
</table>

**FIGURE 1D**
POLYMORPHISMS IN THE CODING SEQUENCE OF CTLA4

ATGGCTTGCC TTGGATTCTA GCGGCACAAG GCTCAGCTGA ACCTGGCTAC
G
CAGGACCTGG CCCTGCACTC TCCTGTNTTT TTTCTCTTTT ATCCCTGTCT 100
TCTGCAAAGC AATGCACCTG GCCACGCTG CTGTGGTACT GCCAGCAAGC
CGAGGCATCG CCAGCTTGTG TTGAGTATGC GCACTCCAG GCAAAGCCAC 200
TGAGGTCCTG GTGACAGTGC TCGGGCAAGC TGCACGCCAG TGACTGAAG
TCTGTGCCGC AACCTACAG ATGGGGAAAG AGTTGACCTT CCTAGATGAT 300
TCCATCTGCA CGGCACTCTC CATGGGAAAT CAGTGAACC TCACTATCCA
AGGACTGAGG GCCATGGACA GGGAACCTCTA CATCTGGAAG GTGGAGCTCA 400
TGACTCCACC GCCATCTAC CGGCAACTCTA CAGTGGAAAT CAGTGAACC TCACTATCCA
GTAATTGATC CAGAACCCTT GCCAGATTCT GACTTCTCTC TCTGATCCT 500
TGCACTGATG GTTCCGGGGT TAGTTTTTTC TAGCTTTCTC TGTCAAGCTG
TTTTTGAG CAAATGCTA AAGAAGAAAG GCCCTCTTCTC AAGACTGTCG 600
TATGTGAAAA TGCCCAACT AGGCGGCAAG TGTAAGAAAG AATTCCAGCC
TTATTATTAT CCCATCAATT GA 672

FIGURE 2
ISOFORMS OF THE CTLA4 PROTEIN

MACLGFQRHK AQLNLATRTW PCTLLLFFLLF IPVFCAMHV AQPAVVLASS

RGIASFVCEY ASPGKATEVR VTVLRQADSQ VTEVCAATYM MGNEFLFLLD 100
SICTGTSSGN QVNLIQGLR AMDGLYICK VELMPPPYY LGIGNGTQIY
VIDPECPDPS DPLLWILAAV SSSLFFYSFL LTAVALSKML KKRSPITTTGV 200
YVKMPPEFE CEKQFQPYFI PIN 223

FIGURE 3
SEQNCE LISTING

GenartBioge Pharmaceuticals, Inc.
Chew, Anne
Choi, Julie Y.
Messer, Chad

Haplotypes of the CTLA4 Gene

MWH-0537 PCT CTLLA4

TBA
2001-05-23
60/206, 353
2000-05-23
36
PatentIn Ver. 2.1
1
9336
DNA
Homo sapien

1
actcaaaattt ccoccagttt ctcagtaatc tggctttgtat tattttttccc aatcaataat 60
aacacatatt attctgtcatt catgtctcta ttttcctcttt taatctagaa tggttttccaa 120
ggctttccc tttttaatct ttgtgatgt taacttttttt gaagaatcta ggcaggatt 180
)tngctgaat atggattgtt cttgattatttt ctcotatgcttt agatctcgtt aaacaaat 240
tgggataaatt tacaaaagtt tggctgttcc ttctcaggtg attgtatctg gaggtacctt 300
atgtcagaaac acctatcctc agtaatgttc agtggggtac attgggttaaa acagatattc 360
gctagattta tcaattatct tttctctttatt gtaattatt caaagttattt ggggggttatt 420
atgtgaaat atttttttctt accagatcattttt ctttaccccttc attttttttttttcaacatttt 480
gaatatcctt gctatgcag acltagtttcc atagagcatt aataaggaaa tctggtaaga 540
cgtctttttga agttcattttt aaacccctca cttcctttctta gttttaaccttg ccctttcatctt 600
tttactacc tatttaacaa atgttttaaac aatgtctgttt ttggtgtttgc gctgtgca 660
ttttcatctag tgctctcata gttgccatata ctaattgcttc tttctctctt ccaacatactt 720
attggcattt attttcattttg atgccataaa atctctccata acctcatgaa atgctgccatt 780
tcactttttc accgtactctt ccttcttttt attccagttta attggcccttc ttgcttttcta 840
ttgaccctat aataaggaa gcctttttatt tgcgtggcata ctcgtttcataa cggccttttt 900
atccgttta cttggttattatttttttttttttaa atctgagaga ctagtgcgtta 960
ttatctggag cagattggtggc ttggacccac aggtcaacagt taagagtaac ccacgtctag 1020
gatacaatc caagttatcctt ggccttcaaa atctcttttg attcagatag ataaaggttt 1080
cactgtgcctt aatacctgctt tcgcctcactt ctcaggggttc atgggtaaag tcaatcttcatc 1140
tttcaatgcc tgctttttcttt ttttataaa atgggaaaaa ttagttcctag ctggctttctag 1200
gttgtttagt gaattgagaa aaaaaaact tgcggtggcct gttggctcctc gcctgtataatc 1260
<210> 2
<211> 672
<212> DNA
<213> Homo sapien

<400> 2
atggcttgcc ttggtattca ggcggcacaag gctgagctgctgta acgctgtgctg 7080
ccctgcacctc tctgtttttt ttcttctctc atccctgtct tctgcaaaagc aatgcaagtat 120
gcccaagcttg ctgtgtctct gcggcaacagc cggaggcatcg ccagcttttggt gtgtgatgtat 180
gcatccagc gcagcatcatg ttgagtcggg attctcagagc ttgagcgccg tgcagcagcag 240
gtagctgaag tccttgccggc acctacatg atggggaatg gttgacactt cctatgatgat 300
tccatctgca cgggcaacctc cagtggaaat caaagcaacc ttctatccaa tgcagaatggg 360
gcatacgca cgggactctct cattctgcaaat gttggagctca tgtaccaccgc gcaatcactatc 420
cgggcatatgc caacgcaagac ccagatttat gtaattgatc cagaacgctg ccagcattctc 480
gacttcctcc cttccggatct tggacaggtt atgggggtgt tgggggttta tagccttcctc 540
cctccagctgt ttgctttgag caaatgctaa aagaaagaaaga gcccctttaac acacaggggct 600
ttctgaaaaat tcgcccaccagc agagcaagact gttgaaagcg aatttccagct ttatatatatt 660
cccatcaattg a 672

<210> 3
<211> 223
<212> PRT
<213> Homo sapien

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

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

<400> 4
agatccttyaa agtga

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

<400> 5
cagtcaargg cagtg

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

<400> 6
cactgagytg acacc

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

<400> 7
ctagaacygt aggca
8
ttttaatrgc tgaat

9
gctgtagcagttcat

10
ttatccagat cctya

11
tcagttcagttttra

12	tttcagcagt caarg

13
<400> 18
ataaatattt aatrg 15

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

<400> 19
ttotgtacct agcyt 15

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

<400> 20
tgtatgtctg tgarc 15

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

<400> 21	ttaaaatga atgyt 15

<210> 22
<211> 10
<212> DNA
<213> Homo sapien

<400> 22	tccagatcct 10

<210> 23
<211> 10
<212> DNA
<213> Homo sapien

<400> 23
tgttcacttt 10
<210> 24
<211> 10
<212> DNA
<213> Homo sapien

<400> 24
cagcagtc

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

<400> 25
aatcactg

<210> 26
<211> 10
<212> DNA
<213> Homo sapien

<400> 26
tttcactgag

<210> 27
<211> 10
<212> DNA
<213> Homo sapien

<400> 27
acaggtgtca

<210> 28
<211> 10
<212> DNA
<213> Homo sapien

<400> 28
gcactagaac

<210> 29

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

<400> 29
caatgcc tac 10

<210> 30
<211> 10
<212> DNA
<213> Homo sapien

<400> 30
aaat ttaaat 10

<210> 31
<211> 10
<212> DNA
<213> Homo sapien

<400> 31
ttgattcagc 10

<210> 32
<211> 10
<212> DNA
<213> Homo sapien

<400> 32
ttgctgtgta 10

<210> 33
<211> 10
<212> DNA
<213> Homo sapien

<400> 33
aaatgaatg 10

<210> 34
<211> 18
<212> DNA
<213> Homo sapien

11
<400> 34
tgtaaaacga cggccagt 18

<210> 35
<211> 19
<212> DNA
<213> Homo sapien

<400> 35
aggaacagc tattaccat 19

<210> 36
<211> 9336
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (3735)
<223> PS1: polymorphic base C or T

<220>
<221> allele
<222> (4102)
<223> PS2: polymorphic base A or G

<220>
<221> allele
<222> (4238)
<223> PS3: polymorphic base A or G

<220>
<221> allele
<222> (7067)
<223> PS4: polymorphic base T or C

<220>
<221> allele
<222> (7747)
<223> PS5: polymorphic base C or T

<220>
<221> allele
<222> (8660)
<223> PS6: polymorphic base A or G
<220> WO 01/90122
<221> allele
<222> (8708)
<223> PS7: polymorphic base A or G

<400> 36:
actcaaatct cccagagttt ctctagtagg tgcgggtgat tattttttttt aactcaaatct 60
aacacatcct atctgtgcac atgcctctca tttctctctt tacatctagaa tgggttttcca 120
ggctttcccc ttcttaattc tttgtgtgat taacatctttt ggaagaatta gccacagttt 180
ttgctgaaat atggtttggt ctgattatct ctctctgcttt agatctgatct taaaactattt 240
tgccaataac taccctctag tttgcatgtt attatgcagtc gagggtcacc 300
atgtagcaac aacctctcag cagtaatcatc agttggggtct aacttggtttaa agcagatctct 360
gcatttttatca ctatatattt acaacatgtt cttctcttta ttcttaaatg ttcctcttca 420
atattctttt atctttctcct acacagatttc ttttaccagc cttttttttttt acaacatttt 480
gaatctcttct gcagctgcag aatagttttc ataagaagcctaat aatggagatcc tctgcattaa 540
cgctcttttga aagttctatat acaaccccttc tctcttttttca ttttaaatgct ttcctcttca 600
tttactttcctt tttttatataa aagttttttt aagctctgcct tttgtggcatt ggctgtagca 660
ttttactcag tgggctttcct ttttaaatggc ctaatgcttc ttgctgctct ccacatctct 720
attgcacccat attttttact catggcctatct ctttttttttt tttttaaatgct ttctctcttt 780
tcaactttttc aacctgctgtt ctctattttt atctcagtaa atggcacactc ttccctcttca 840
tggacccccat aataaagaag gcctttattt cgacactgct ttctcttttca taggggttattt 900
atatcgtgct taccattatttt cttttttcaata atctctgcagaa ttttagtgccca 960
ttactttttgg gcggacagcgg gcggggctag taacactttgct cttctctctttt gaactttcgcct 1020
gatacaactt cagttatttt gggccttaata aatctttttttcatc tagcatagtgat ataaggttat 1080
cactgtggctc aatactccgt cctgctcactc cctcaggttctt atgggttagag tacacattaacc 1140
ttttacttact ttcttatttttt ttttttaaatgcttaggtttag tttgtgtgtag 1200
ggttgcattg cgacggcacta aacaacaaact tttggcagcgc ggtggcctcct gctgtgaactttc 1260
caggcgcttg tgggacacgca ggggcccagag taggtgctctgct cttctctttttt gcctgtagctttc 1320
tggtcataact ggcggacccct catccctact aaaaatatacc aaaaaattag ctggggcggtgg 1380
tggtggtggcg tctgctcctttt cgggctccgg cgggcttccag cagggagcact gctgtagctttttt 1440
agggaccaga ggtttggcagct ccggcatctct acaccaacctt ggggaacaag 1500
agttgcacaaatt ggtgctcttg cttttttaaatgctagctttctt gctgctcatac 1560
atctttctcaac ctaattttgttt gcttttctca tttttttaaatgcttaggtttatg tttgtgtgtag 1620
acactctattt gattttaagttt tagttagttt cccacacagct cacagcttttt tttttttaaatgct 1680
atttttttagtt gggggccacag aagcgccttgg cttttttaaatgcttaggtttatg tttgtgtgtag 1740
atttttttagtt gggggccacag aagcgccttgg cttttttaaatgcttaggtttatg tttgtgtgtag 1800
atcgcttttt ctctttttttttt tctctttttttt tctctttttttt ggtatctctcttc ttaatgggctc 1860
gcaatctacat cttttttaaatgcttaggtttatg tttgtgtgtag 1920
tgggtgctttt cttttttaaatgcttaggtttatg tttgtgtgtag 1980
ctctcttcctt cttttttaaatgcttaggtttatg tttgtgtgtag 2040
ctcctctttttttt tctctttttttt tctctttttttt ggtatctctcttc ttaatgggctc 2100
ctcctctttttttt tctctttttttt tctctttttttt ggtatctctcttc ttaatgggctc 2160
ctcctctttttttt tctctttttttt tctctttttttt ggtatctctcttc ttaatgggctc 2220
ctcctctttttttt tctctttttttt tctctttttttt ggtatctctcttc ttaatgggctc 2280
tacttcatct cccacagttgag ctcctttttttttt tctctttttttt ggtatctctcttc ttaatgggctc 2340
ctcctctttttttt tctctttttttt tctctttttttt ggtatctctcttc ttaatgggctc 2400
tctcctctttttttt tctctttttttt tctctttttttt ggtatctctcttc ttaatgggctc 2460
gaagctgggct tggaaggttt ctagaatgtg tagcagctgta tagaaacag gcaggtcaga 2520
aaaggtctct gtcctacaca ccaacagtgc acatgtatac atatgtaaca atctgtcagt 2580
tggctacct atgccctaaat accctaaaag taattataat aaatatttaa aaaaaaaag 2640
aagggctctc ctggaggggt tagcagctgta gctagttctct gggagatgag aagaggtata 2700
aaataagata atagggagaa agaagcaagga gcagagcagcg ggttaaggtg gatgacgcc 2760
gaaaggggca cgtggaagga aagcacaatg gaggaagggc gatggtcgg gataaataaa 2820
gtacagccca aagggagggg cattcgggga gttgttcctct aacactccttg tgtgtcgtata 2880
atgtggtggt atggagccag gtcgagccat cccacacctg ttcgcccgttt 2940
acaactgtct gtttgctgtg cagcctgctga aagggccctt aaggtatcac ctattgcttttt 3000
gttttgctcat cattcaaccc taagtgcaaca gatcctgcgg gatattcagcg ctctccgtatg 3060
aatgttctct tctttatatg aatagtgaag aaccttccag aattaaggaatttgcctcata 3120
gtttcaattg ttttggtgta ggtgagttgt gcttcttcag gtttttcatc attttgagcg 3180
tctggtttata ttttactttc atcttttgag attttctgctg aaaaaatttgta attagaaaaaa 3240
aagagtactat cttttatgga cggctcttaat ctctgtagaatt attttggtctgctgattttctttttcttttttttttctt 3300
aggaactctt cttgctctctg tgggtgaggg ctttaaaggacttacctaatgc ctctgccgttct 3360
tttccctttct gccaaacccgct cttttttttg gtttttttttt ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
taaaagagtgg aagagatgtg attagctcgt tataataacttt ttataaagttc agaaagttaa 5400
gaacatctgcc aataggtgtta tcagcagctt aagagttgctc aagggctatc tctgatggc 5460
agatgggtgat tagaggggttg aagagctgtt ctagccgtctt cagaggtgctt gaaagagctg 5520
tcgaacatgttc gtaattccttct tcttcagcacc gaaatcagctt ctagctggtt 5580
tagtttgagc gatttaagta tctagcttcc tatttcctgtt gatatttcaag aaataatttttttt 5640
gaaatcagag ccaagagtcag aatagatctt gggcttttatt gattgtcagt aagagagccttg 5700
ttgcttttcat aagagtttgat gtcagctgctt cagagccttcc cgaagccttg 5760
gactctagtc aacacagtgtt tagtgtagctt tctgctagcac gacagctttgcc gggccctg 5820
tcagttgctt gaaagagatgtag tctagcttag cggagccttgg tctcagcacc 5880
tcagttggagc ttgagagctt tagataagta cagagccttcc aagagagagcttg 5940
tcagttggagc ttgagagctt tagataagta cagagccttcc aagagagagcttg 6000
actcaggtagc aaatatctgtt aagttctatttt cggagagcttt tctgcttagat 6060
gctctctcttc tctacacatc tctcctactt ttacacatgtt ctctctactt ctctctactt 6120
tcagttgctt gaaagagctt aagttctatttt cggagagcttt tctgcttagat 6180
tcagttgctt gaaagagctt aagttctatttt cggagagcttt tctgcttagat 6240
tcagttgctt gaaagagctt aagttctatttt cggagagcttt tctgcttagat 6300
tcagttgctt gaaagagctt aagttctatttt cggagagcttt tctgcttagat 6360
tcagttgctt gaaagagctt aagttctatttt cggagagcttt tctgcttagat 6420
tcagttgctt gaaagagctt aagttctatttt cggagagcttt tctgcttagat 6480
tcagttgctt gaaagagctt aagttctatttt cggagagcttt tctgcttagat 6540
tcgtaggggtg tggagagcgtt aagttctatttt cggagagcttt tctgcttagat 6600
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 6660
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 6720
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 6780
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 6840
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 6900
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 6960
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7020
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7080
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7140
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7200
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7260
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7320
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7380
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7440
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7500
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7560
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7620
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7680
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7740
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7800
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7860
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7920
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 7980
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 8040
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 8100
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 8160
tcctcaggttt ttgtaggagatatgct tctcctatttt cggagagcttt tctgcttagat 8220
ttctttttga cagtcacctc cagacacctc tgccttaaggc cagctttgcc attacaaact 8280
cctcttcttc cctcttcgcc ttctttcttt cctcttcctt cttctctgctc tttctctctc 8340
tctctttcttc cctctgtgtc tctctataac atacacaag atatactcga ttcacaagctc 8400
cctctacccca cctgacagag atgtcctttg ctgttaggttc agcagtgggg atgagaagaata 8460
cagctctccaa acaggttataa taaggtttaaat tattttatatc agctttttcct cttgcagaca 8520
agattgatca attatcatag gctttctggt tgtttttttcct gaaagcttctct caaagtctct 8580
tttcctctac ttcccattccaa ggcaaatgat tgcctattaa cattcacaac acagttaattt 8640
atctaaattt aattttaatrt gctgtaacaaaa aaaaaatttcc tgggtttatat atatctgtat 8700
gctgtgaraac ttcttttttt accagctagg gcaccaaatgt gttctgagtt ctattataggt 8760
tagagttgcgg ttctgtattc actaagttagct attactgtttt cttttggttgttg ttgacagcata 8820
aagaaaaaga gcccttttac acacaggggtct taggtgaaaa tgccccccac agacccagaa 8880
tgtgaaaaagc aatttccagc ttatattttatt cccatcaattt gagaacact tatgaagaagt 8940
agagtccata ttctaaattgcc caagagctga ggcaattctca acattttgggc tatccagctgt 9000
ttttttattgg ttgtgctatt tgggggaatat tctctctctt ttaatatataa gttggattgg 9060
gaacccaat tacgtgtaact cacaattttaa gcacaaggagt agaaagacag agcttggtgatg 9120
tttctgtccac atcaaggctcct cctccagttga aagcatacact tgggattatat atggggtgtg 9180
agcattatgta tgtgggtcaca ggaattaagt taggggaatgg cacagcccaa agaggggaaag 9240
gcagggagcg agggagaaga ctataattgtga cacaaccttat attacgtgatg tagacgtttta 9300
tagccgaat gatctgttcca agtttagatgt tatgccc 9336