



US 20070254289A1

(19) **United States**

(12) **Patent Application Publication**  
**Li et al.**

(10) **Pub. No.: US 2007/0254289 A1**

(43) **Pub. Date: Nov. 1, 2007**

(54) **GENETIC POLYMORPHISMS ASSOCIATED WITH ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF**

**Publication Classification**

(75) Inventors: **Yonghong Li**, Palo Alto, CA (US);  
**Andrew Grupe**, Orinda, CA (US)

(51) **Int. Cl.**  
*G01N 33/53* (2006.01)  
*C07H 21/04* (2006.01)  
*C07K 16/18* (2006.01)  
*C12Q 1/68* (2006.01)

Correspondence Address:  
**CELERA DIAGNOSTICS, LLC**  
**1401 HARBOR DAY PARKWAY**  
**ALAMEDA, CA 94502 (US)**

(52) **U.S. Cl.** ..... **435/6**; 435/7.1; 530/300; 530/387.1;  
530/387.9; 536/22.1; 536/23.5

(73) Assignee: **APPLERA CORPORATION**, Norwalk, CT

(57) **ABSTRACT**

(21) Appl. No.: **11/586,427**

(22) Filed: **Oct. 24, 2006**

**Related U.S. Application Data**

(60) Provisional application No. 60/730,918, filed on Oct. 26, 2005.

The present invention is based on the discovery of genetic polymorphisms that are associated with Alzheimer's Disease. In particular, the present invention relates to nucleic acid molecules containing the polymorphisms, variant proteins encoded by such nucleic acid molecules, reagents for detecting the polymorphic nucleic acid molecules and proteins, and methods of using the nucleic acid and proteins as well as methods of using reagents for their detection.

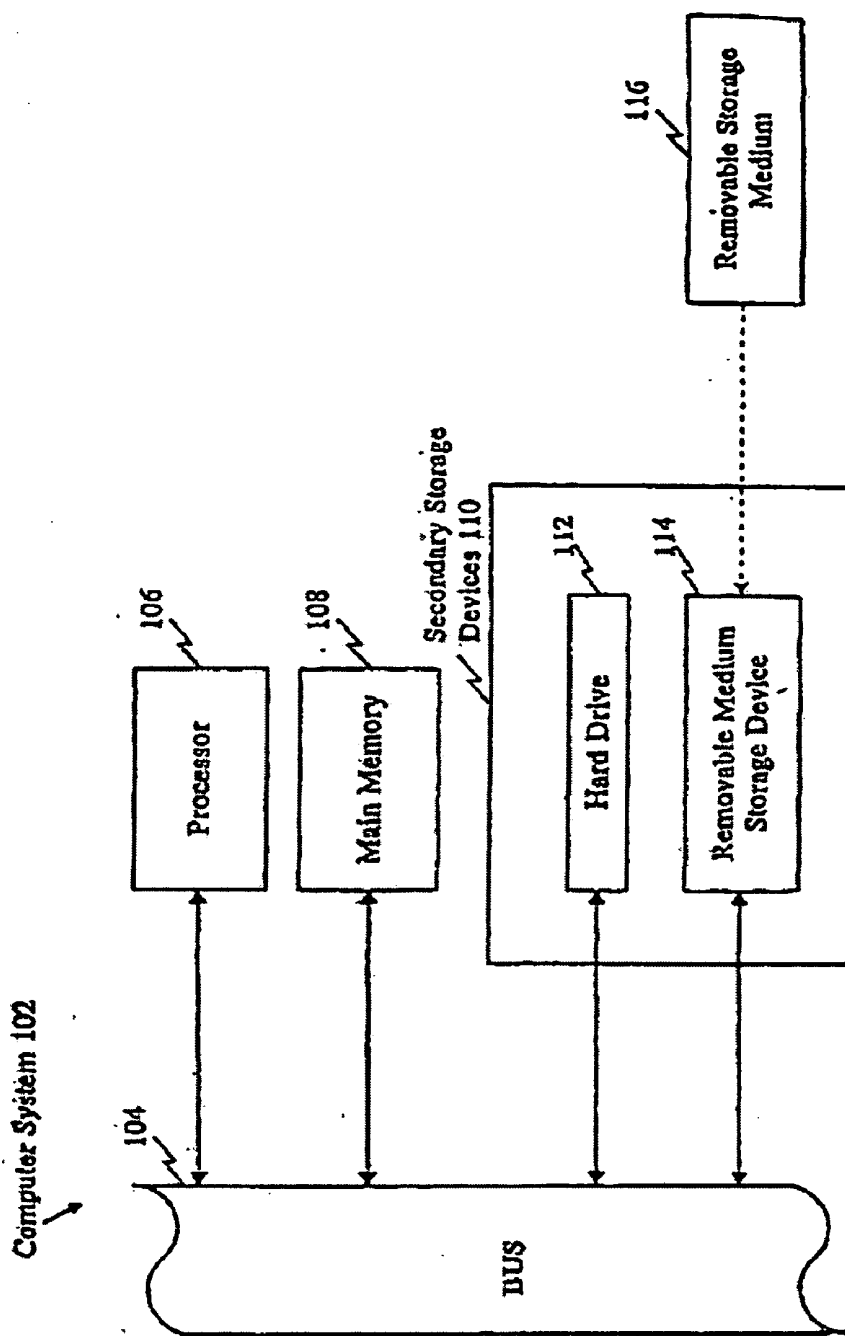


FIGURE 1

Figure 2

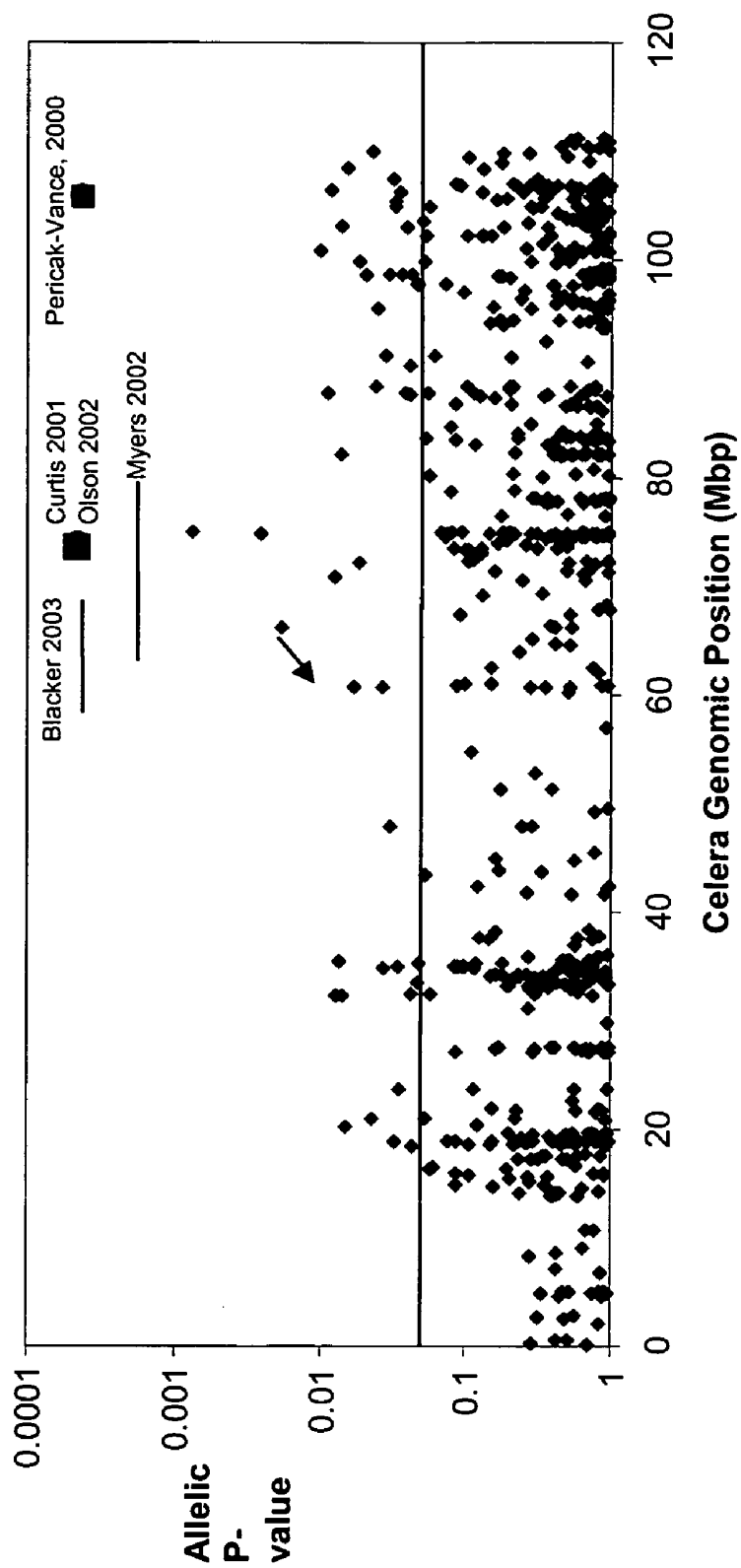


Figure 3

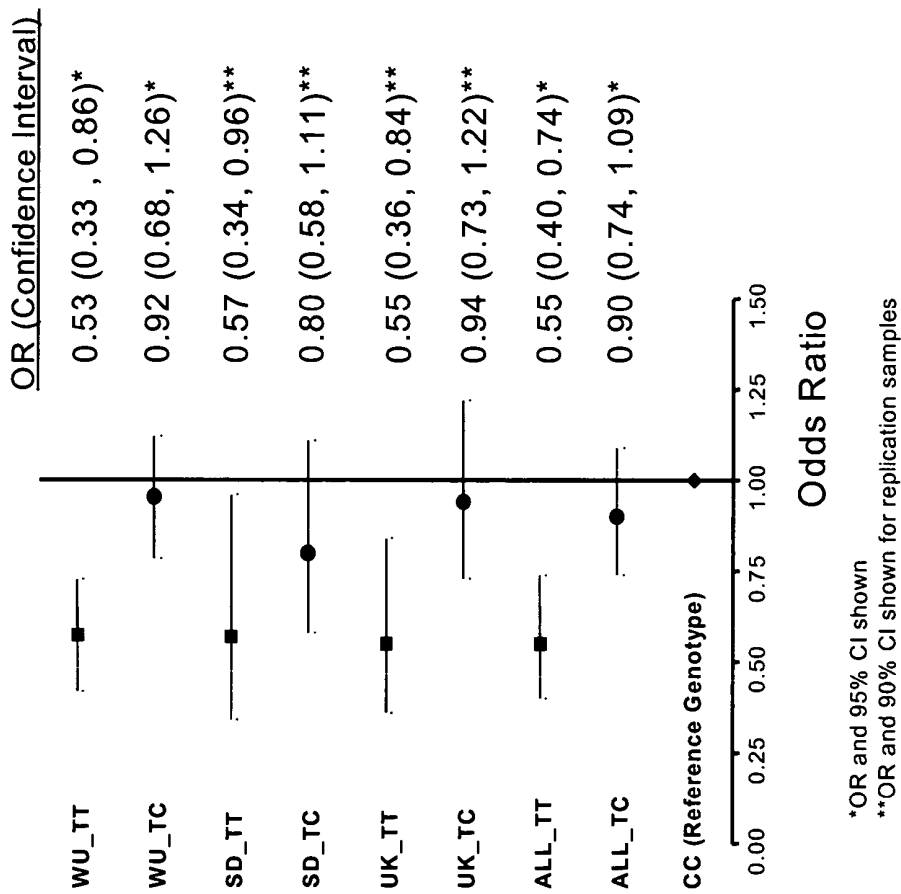
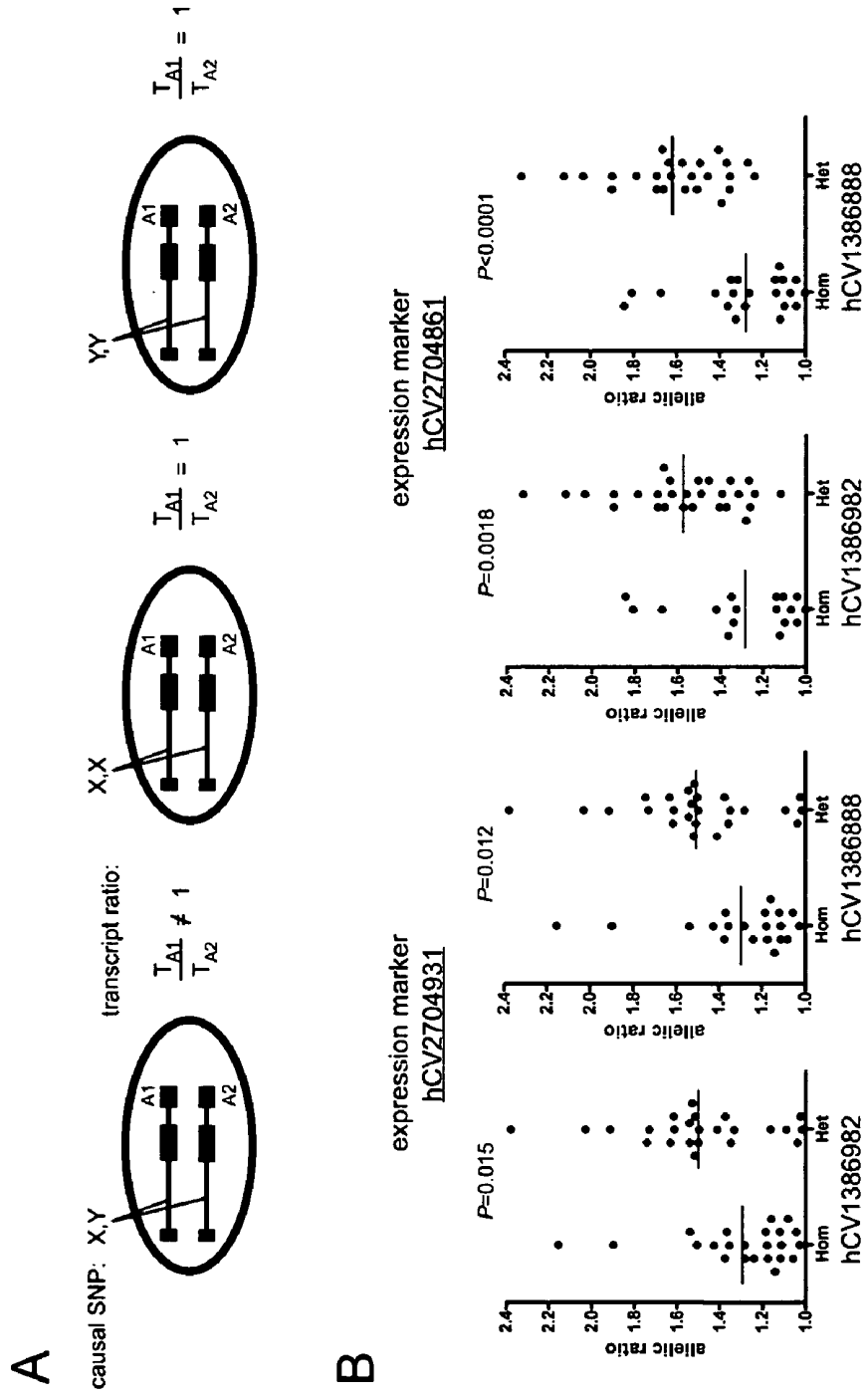


Figure 4.



**GENETIC POLYMORPHISMS ASSOCIATED WITH  
ALZHEIMER'S DISEASE, METHODS OF  
DETECTION AND USES THEREOF**

CROSS-REFERENCE TO RELATED  
APPLICATIONS

[0001] This application claims the benefit of U.S. provisional application Ser. No. 60/730,918, filed on Oct. 26, 2005, the contents of which are hereby incorporated by reference in its entirety into this application.

FIELD OF THE INVENTION

[0002] The present invention is in the field of Alzheimer's Disease diagnosis and therapy. In particular, the present invention relates to specific single nucleotide polymorphisms (SNPs) in the human genome, and their association with Alzheimer's Disease and related pathologies. Based on differences in allele frequencies in the Alzheimer's Disease patient population relative to normal individuals, the naturally-occurring SNPs disclosed herein can be used as targets for the design of diagnostic reagents and the development of therapeutic agents, as well as for disease association and linkage analysis. In particular, the SNPs of the present invention are useful for identifying an individual who is at an increased or decreased risk of developing Alzheimer's Disease and for early detection of the disease, for providing clinically important information for the prevention and/or treatment of Alzheimer's Disease, and for screening and selecting therapeutic agents. The SNPs disclosed herein are also useful for human identification applications. Methods, assays, kits, and reagents for detecting the presence of these polymorphisms and their encoded products are provided.

BACKGROUND OF THE INVENTION

[0003] Neurodegenerative Diseases

[0004] A varied assortment of central nervous system disorders (neurodegenerative diseases) are associated with aging. Neurodegenerative diseases are characterized by a gradual and progressive loss of neural tissue or nerve cells. These diseases, directly or indirectly, affect millions of people worldwide. The number of individuals affected by neurodegenerative diseases is anticipated to grow attendant with the increase in human life expectancy.

[0005] Specific diseases exemplifying this class of disorders include: age-related dementia, such as Alzheimer's Disease (AD); leukodystrophies, such as adrenoleukodystrophy, metachromatic leukodystrophy, Krabbe Disease (globoid cell leukodystrophy), Canavan Disease, Alexander Disease, Pelizaeus-Merzbacher Disease, and the like; and others such as neuronal ceroid lipofuscinoses, amyotrophic lateral sclerosis (ALS, or Lou Gehrig's Disease), Huntington's Disease (HD), dentatorubral-pallidolusian atrophy (DRPLA), stroke and the like.

[0006] Parkinson's Disease affects 1 to 2 percent of people over the age of 50 and 10 to 15% of those over 80. Huntington's Disease and ALS each afflict approximately 30,000 in the United States. Stroke is the leading cause of neurological impairment with half a million new stroke victims surviving each year with some degree of permanent neurological damage.

[0007] Alzheimer's Disease (described in greater detail in the following section) alone affects 20 million people world-

wide. Alzheimer's Disease is the fourth leading cause of death in industrialized societies, afflicting 5-11% of the population over the age of 65 and 30% of those over the age of 85. Alzheimer's Disease is fast becoming the paramount healthcare problem as the world's geriatric population continues to grow.

[0008] Alzheimer's Disease

[0009] Alzheimer's Disease is the most significant and common cause of dementia in developed countries, accounting for 60% or more of all cases of dementia. Alzheimer's Disease is a progressive neurodegenerative disorder characterized clinically by memory loss of subtle onset, followed by a slowly progressive dementia that has a course of several years. Brain pathology of Alzheimer's Disease is characterized by gross, diffuse atrophy of the cerebral cortex with secondary enlargement of the ventricular system. Microscopically, there are neuritic plaques containing A $\beta$  amyloid, silver-staining neurofibrillary tangles in neuronal cytoplasm, and accumulation of A $\beta$  amyloid in arterial walls of cerebral blood vessels. A definite diagnosis of Alzheimer's Disease can only occur at autopsy, where the presence of amyloid plaques and neurofibrillary tangles is confirmed.

[0010] The frequency of Alzheimer's Disease increases with each decade of adult life, reaching 20 to 40 percent of the population over the age of 85. Because more and more people will live into their 80's and 90's, the number of patients is expected to triple over the next 20 years. More than 4 million people suffer from Alzheimer's Disease in the USA, where 800,000 deaths per year are associated with Alzheimer's Disease. It is estimated that the cost of Alzheimer's Disease in the USA is \$80 billion to \$100 billion a year in medical care, personal caretaking and lost productivity. Alzheimer's Disease also puts a heavy emotional toll on family members and caregivers: about 2.7 million people care for Alzheimer's Disease patients in the USA. Alzheimer's Disease patients live for 7 to 10 years after diagnosis and spend an average of 5 years under care either at home or in a nursing home.

[0011] In spite of the high prevalence of Alzheimer's Disease today and its expected prevalence increase in an aging population, there are currently no diagnostic tests available that determine the cause of dementia and adequately differentiate between Alzheimer's Disease and other types of dementias. A diagnostic test that enables physicians to identify Alzheimer's Disease early in the disease process, or identify individuals who are at high risk of developing the disease, will provide the option to intervene at an early stage in the disease process. Early intervention in disease processes does generally result in better treatment results by delaying disease onset or progression compared to later intervention.

[0012] Alzheimer's Disease is presumed to have a genetic component, as evidenced by an increased risk for Alzheimer's Disease among first degree relatives of affected individuals. So far, three genes have been identified in patients with early onset Alzheimer's Disease that lead to the less common, dominantly inherited form of dementia. Mutations in the three genes, beta-amyloid precursor protein (Goate et al., *Nature* 1991, 349:704-706), presenilin 1 (Sherington et al., *Nature* 1995, 375:754-760), and presenilin 2 (Levy-Lahad et al., *Science* 1996, 269:973-977) lead to an increase in the production of long amyloid beta (A $\beta$ 42), the

main component in amyloid plaques. Although early onset Alzheimer's Disease makes up less than 5% of all Alzheimer's Disease cases, the identification of these genes has contributed substantially to the understanding of the disease process.

[0013] Late onset Alzheimer's Disease (LOAD), the much more common form of this dementia, is inherited in a non-Mendelian pattern and involves genetic susceptibility factors and environmental factors. Early genetic studies of Alzheimer's Disease demonstrated association and linkage to the same region on chromosome 19 containing the ApoE gene (Schellenberg et al., *J. Neurogenet.* 1987, 4:97-108, Pericak-Vance et al., *Am. J. Hum. Gen.* 1991, 48:1034-1050). Three common alleles were identified for the ApoE gene,  $\epsilon 2$ ,  $\epsilon 3$ ,  $\epsilon 4$ . The  $\epsilon 4$  allele frequency is increased to 50% in affected individuals vs. 14% in controls (Corder et al., *Science* 1993, 281:921-923). Although there is strong association with the ApoE- $\epsilon 4$  allele, which has been replicated in many studies, most investigators consider the ApoE- $\epsilon 4$  allele to be neither necessary nor sufficient for the development of Alzheimer's Disease. ApoE is considered a major risk factor, but ApoE testing does not provide enough sensitivity and specificity for use as an independent diagnostic test and therefore is not recommended as a diagnostic marker for the prediction of Alzheimer's Disease (National Institute on Aging/Alzheimer's Association Working Group, 1996).

[0014] Genome-wide linkage screens in LOAD patients, duplicated in at least 2 studies, identified regions on four chromosomes, chromosomes 6, 9, 10, and 12 (reviewed by: Myers and Goate, *Curr. Op. Neurol.* 2001, 14:433-440, Lendon and Craddock, *TINS* 2001, 24:557-559), implying that other genetic risk factors besides ApoE must exist. Co-localization of a quantitative trait for A $\epsilon$ 42 and a susceptibility locus for LOAD on chromosome 10, for example, suggests the locus influences LOAD risk through increased levels of the A $\beta$ 42 peptide (Ertekin-Taner, *Science* 2000, 290:2303-2304).

[0015] The majority of the putative LOAD susceptibility loci were identified through linkage studies of affected sib pairs (ASPs) by looking for regions with increased allele sharing. In order to identify the genes and mutations for LOAD, it would be beneficial to conduct association studies, which have relatively better power than linkage studies to detect genes of modest or small effect. Association studies compare unrelated cases to controls and analyze allele frequency differences between affected and unaffected individuals.

[0016] Obviously, there is a definite need for novel diagnostic markers that enable the detection of Alzheimer's Disease at an early stage of the disease. The availability of a genetic test will also provide a non-invasive method to assess an individual's risk for developing Alzheimer's Disease. Furthermore, there is an urgent need for new and improved treatments for Alzheimer's Disease to prevent or significantly delay the onset of the disease, or to reverse or slow down disease progression after onset.

[0017] DAPK1: Association with Neurological Pathology and Use as a Druggable Target

[0018] Death-associated protein kinase 1 (DAPK1) is a  $\text{Ca}^{2+}$ /calmodulin-dependent serine/threonine kinase that

plays a pro-apoptotic role in the programmed cell death cascade. DAPK1 is located on chromosome 9 at the 9q22 locus, and encodes a structurally unique 160-kD protein that carries a kinase domain, a  $\text{Ca}^{2+}$ /calmodulin-binding domain, eight ankyrin repeats, two putative P-loop consensus sites, a cytoskeleton binding domain and a death domain.

[0019] DAPK1 was originally identified for its involvement in interferon gamma-induced cell death, but has subsequently been found to be involved in apoptosis (including neuronal cell apoptosis) caused by other stimuli/insults. A number of studies strongly suggest that DAPK1 could be associated with neurological diseases, and there is strong support for DAPK1 as one of the genetic factors affecting susceptibility to LOAD in particular. The evidence can be summarized as follows: (i) DAPK1 is in the center of a previously reported AD-associated linkage peak, (ii) DAPK1 is highly expressed in the brain, and in adults is largely restricted to the hippocampus and cortex, (iii) DAPK1 is a pro-apoptotic mediator in the programmed cell death pathway, (iv) SNPs in DAPK1 are significantly associated with LOAD (described below) and (v) DAPK1 shows allelic variation at the RNA level, and the SNPs associated with LOAD risk may directly or indirectly modulate this allele-specific gene expression. Importantly, because DAPK1 is an enzyme whose activity is correlated with neuronal cell death, the inhibition of DAPK1 is a very attractive target for drug development.

[0020] DAPK1 is in the center of a linkage peak that has been shown in several studies to be related to AD (e.g. A. Myers et al., "Full genome screen for Alzheimer's disease: stage 11 analysis," *Am. J. Med. Genet.* 114, 235-244 [2002]; D. Blacker et al., "Results of a high-resolution genome screen of 437 Alzheimer's disease families," *Hum. Mol. Genet.* 12, 23-32 [2003]).

[0021] DAPK1 is most abundantly expressed in the brain and lung, and is poorly or not detected in muscle cells or tissue from the stomach, small intestine, testes, etc. (M. Yamamoto et al., "Developmental changes in distribution of death-associated protein kinase mRNAs," *J. Neurosci. Res.* 58, 674-683 [1999]). In embryonic rat brain, DAPK1 mRNA is detected at high levels in the cerebral cortex, cerebellar Purkinje cells, and hippocampus; in some studies of the adult rat brain, DAPK1 mRNA appears to be restricted to the hippocampus (Id.). In other studies of rat brain, however, Western blot analysis detected DAPK1 protein in the hippocampus, cortex and olfactory bulb, and not in the cerebellum, hindbrain or mesencephalon (J. H. Tian et al., " $\text{Ca}^{2+}$ -dependent Phosphorylation of Syntaxin-1A by the Death-associated Protein [DAP] Kinase Regulates Its Interaction with Munc18," *J. Biol. Chem.* 278[28]: 26265-26274 [Jul. 11, 2003]). This expression pattern is extremely relevant to AD, because the hippocampus and cortex are the most severely affected regions in this disease.

[0022] Increased DAPK1 activity or expression has been associated with neuronal cell death, which makes this enzyme an attractive candidate for drug development in the treatment of AD and other neurological pathologies. Exposure of PC12 cells (model cells for study of neuronal cell death) with a cell-permeable ceramide analog,  $\text{C}_2$ -ceramide, resulted in increased DAPK1 activity leading to apoptosis (M. Yamamoto et al., "DAP kinase activity is critical for  $\text{C}_2$ -ceramide-induced apoptosis in PC12 cells," *Eur. J. Bio-*

*chem.* 269:139-147 [2002]; ceramide has been proposed as one regulator of cell-cycle arrest and apoptosis of various cell types, including neuronal). Conversely, neuronal cells lacking DAPK1 are less susceptible to apoptotic insults in cell culture (D. Pelled et al., "Death-associated Protein [DAP] Kinase Plays a Central Role in Ceramide-Induced Apoptosis in Cultured Hippocampal Neurons," *J. Biol. Chem.* 277:1957-1961 [Jan. 18, 2002]), and studies in knockout animal models support the association of reduced DAPK1 activity with increased neuronal protection (H. Schori et al., "Immune-related mechanisms participating in resistance and susceptibility to glutamate toxicity," *Eur. J. Neurosci.* 16, 557-564 [2002]). A study in humans showed that DAPK1 expression and phosphorylation were significantly increased in the epileptic brain when compared with normal (D. C. Henshall et al., "Death-Associated Protein Kinase Expression in Human Temporal Lobe Epilepsy," *Ann. Neurol.* 55:485-494 [2004]). Additionally, a small molecule inhibitor of DAPK1 has been shown to reduce brain injury from hypoxia-ischemia (A. V. Velentza et al., "An aminopyridazine-based inhibitor of a pro-apoptotic protein kinase attenuates hypoxia-ischemia induced acute brain injury," *Bioorg. Med. Chem. Lett.* 13, 3465-3470 [2003]).

[0023] Two polymorphisms and a haplotype in DAPK1 are found to be significantly associated with LOAD (results below). It should be noted that these markers are located in a region of high LD (linkage disequilibrium) that includes only DAPK1 and no other known or predicted genes, therefore supporting a role for DAPK1 in the genetics of LOAD.

[0024] DAPK1 expression shows allelic imbalance, a phenomenon that is seen for genetic risk factors in other complex diseases such as calpain-10 and type 2 diabetes (Y. Horikawa et al., "Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus," *Nat. Genet.* 26, 163-175 [2000]). The difference in allelic expression is controlled by cis-acting elements and may include SNPs within regulatory regions of a particular gene. Our studies show that the genotypes of two DAPK1 intronic LOAD-associated SNPs in particular are significantly associated with DAPK1 allele-specific expression, although they are unlikely to be the sole cis-acting mutations. This association suggests that the SNPs may interact with other unidentified polymorphic cis-acting regulatory factors to influence the level of DAPK1 transcripts. It is also possible that they are in high linkage disequilibrium with other polymorphic cis-acting elements governing DAPK1 transcription. Nevertheless, allele-specific expression of DAPK1 variants does provide a plausible mechanism linking the genetic association with LOAD to a disease-relevant functional outcome, considering that DAPK1 allele-specific expression predicts variation in DAPK1 protein activity and thus neuronal apoptotic potential. Expression of DAPK1 can be induced during neuronal apoptosis (see, for example, M. Yamamoto et al., "Developmental changes in distribution of death-associated protein kinase mRNAs," *J. Neurosci. Res.* 58, 674-683 [1999]). This could potentially mask the effect of allele-specific expression of DAPK1 when samples from patients with different degrees of apoptosis are assayed. Indeed, we observed a large distribution in the DAPK1 transcript level across subjects of various disease severities, probably reflecting an allele-specific effect and induction of

DAPK1 during apoptosis. Therefore, it may be more likely to observe allele-specific expression of DAPK1 in normal brains.

[0025] In conclusion, the expression of DAPK1 in brain cells, the association of DAPK1 activity with neuronal cell death, and the relative immunity of neuronal cells lacking DAPK1 to apoptotic insults, all support the association of DAPK1 with neurological diseases. As an enzyme whose activity is involved in neuronal cell apoptosis, it is an attractive candidate for drug development for the treatment of AD and other neurodegenerative diseases. The genetic association of DAPK1 with LOAD observed in our studies specifically implicates DAPK1 as a potential contributor to AD susceptibility.

[0026] SNPs

[0027] The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution, generating variant forms of progenitor genetic sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 [1986]). A variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. Additionally, the effects of a variant form may be both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. In many cases, both progenitor and variant forms survive and co-exist in a species population. The coexistence of multiple forms of a genetic sequence gives rise to genetic polymorphisms, including SNPs.

[0028] Approximately 90% of all polymorphisms in the human genome are SNPs. SNPs are single base positions in DNA at which different alleles, or alternative nucleotides, exist in a population. The SNP position (interchangeably referred to herein as SNP, SNP site, SNP locus, SNP marker, or marker) is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations). An individual may be homozygous or heterozygous for an allele at each SNP position. A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP is an amino acid coding sequence.

[0029] A SNP may arise from a substitution of one nucleotide for another at the polymorphic site. Substitutions can be transitions or transversions. A transition is the replacement of one purine nucleotide by another purine nucleotide, or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine, or vice versa. A SNP may also be a single base insertion or deletion variant referred to as an "indel" (Weber et al., "Human diallelic insertion/deletion polymorphisms," *Am. J. Hum. Genet.* 71[4]:854-62 [October 2002]).

[0030] A synonymous codon change, or silent mutation/SNP (terms such as "SNP," "polymorphism," "mutation," "mutant," "variation," and "variant" are used herein interchangeably), is one that does not result in a change of amino acid due to the degeneracy of the genetic code. A substitui-

tion that changes a codon coding for one amino acid to a codon coding for a different amino acid (i.e., a non-synonymous codon change) is referred to as a missense mutation. A nonsense mutation results in a type of non-synonymous codon change in which a stop codon is formed, thereby leading to premature termination of a polypeptide chain and a truncated protein. A read-through mutation is another type of non-synonymous codon change that causes the destruction of a stop codon, thereby resulting in an extended polypeptide product. While SNPs can be bi-, tri-, or tetra-allelic, the vast majority of SNPs are bi-allelic, and are thus often referred to as “bi-allelic markers,” or “di-allelic markers.”

[0031] As used herein, references to SNPs and SNP genotypes include individual SNPs and/or haplotypes, which are groups of SNPs that are generally inherited together. Haplotypes can have stronger correlations with diseases or other phenotypic effects compared with individual SNPs, and therefore may provide increased diagnostic accuracy in some cases (Stephens et al., *Science* 293, 489-493 [20 Jul. 2001]).

[0032] Causative SNPs are those SNPs that produce alterations in gene expression or in the expression, structure, and/or function of a gene product, and therefore are most predictive of a possible clinical phenotype. One such class includes SNPs falling within regions of genes encoding a polypeptide product, i.e. cSNPs. These SNPs may result in an alteration of the amino acid sequence of the polypeptide product (i.e., non-synonymous codon changes) and give rise to the expression of a defective or other variant protein. Furthermore, in the case of nonsense mutations, a SNP may lead to premature termination of a polypeptide product. Such variant products can result in a pathological condition, e.g. genetic disease. Examples of genes in which a SNP within a coding sequence causes a genetic disease include sickle cell anemia and cystic fibrosis.

[0033] Causative SNPs do not necessarily occur in coding regions; causative SNPs can occur in, for example, any genetic region that can ultimately affect the expression, structure, and/or activity of the protein encoded by a nucleic acid. Such genetic regions include, for example, those involved in transcription, such as SNPs in transcription factor binding domains, SNPs in promoter regions, in areas involved in transcript processing, such as SNPs at intron-exon boundaries that may cause defective splicing, or SNPs in mRNA processing signal sequences such as polyadenylation signal regions. Some SNPs that are not causative SNPs nevertheless are in close association with, and therefore segregate with, a disease-causing sequence. In this situation, the presence of a SNP correlates with the presence of, or predisposition to, or an increased risk in developing the disease. These SNPs, although not causative, are nonetheless also useful for diagnostics, disease predisposition screening, and other uses.

[0034] An association study of a SNP and a specific disorder involves determining the presence or frequency of the SNP allele in biological samples from individuals with the disorder of interest, such as Alzheimer's Disease, and comparing the information to that of controls (i.e., individuals who do not have the disorder; controls may be also referred to as “healthy” or “normal” individuals) who are preferably of similar age and race. The appropriate selection

of patients and controls is important to the success of SNP association studies. Therefore, a pool of individuals with well-characterized phenotypes is extremely desirable.

[0035] A SNP may be screened in diseased tissue samples or any biological sample obtained from a diseased individual, and compared to control samples, and selected for its increased (or decreased) occurrence in a specific pathological condition, such as pathologies related to Alzheimer's Disease. Once a statistically significant association is established between one or more SNPs and a pathological condition (or other phenotype) of interest, then the regions around the SNPs can optionally be thoroughly screened to identify the causative genetic locus or sequences (e.g., the causative SNP/mutation, gene, regulatory region, etc.) that influences the pathological condition or phenotype. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies).

[0036] Clinical trials have shown that patient response to treatment with pharmaceuticals is often heterogeneous. There is a continuing need to improve pharmaceutical agent design and therapy. In that regard, SNPs can be used to identify patients most suited to therapy with particular pharmaceutical agents (this is often termed “pharmacogenomics”). Similarly, SNPs can be used to exclude patients from certain treatments due to the patient's increased likelihood of developing toxic side effects or his [their] likelihood of not responding to the treatment. Pharmacogenomics can also be used in pharmaceutical research to assist the drug development and selection process (Linder et al., *Clinical Chemistry* 43, 254 [1997]; Marshall, *Nature Biotechnology* 15, 1249 [1997]; International Patent Application WO 97/40462, Spectra Biomedical; and Schafer et al., *Nature Biotechnology* 16, 3 [1998]).

#### SUMMARY OF THE INVENTION

[0037] The present invention relates to the identification of novel SNPs, unique combinations of such SNPs, and haplotypes of SNPs that are associated with Alzheimer's Disease and related neurological pathologies. The polymorphisms disclosed herein are directly useful as targets for the design of diagnostic reagents, as druggable targets in the development of therapeutic agents for use in the treatment and diagnosis of Alzheimer's Disease and also other neurological pathologies.

[0038] Based on the identification of SNPs associated with Alzheimer's Disease, the present invention also provides methods of detecting these variants as well as the design and preparation of detection reagents needed to accomplish this task. The invention specifically provides, for example, novel SNPs in genetic sequences involved in Alzheimer's Disease, isolated nucleic acid molecules (including, for example, DNA and RNA molecules) containing these SNPs, variant proteins encoded by nucleic acid molecules containing such SNPs, antibodies to the encoded variant proteins, computer-based and data storage systems containing the novel SNP information, methods of detecting these SNPs in a test sample, methods of identifying individuals who have an altered (i.e., increased or decreased) risk of developing Alzheimer's Disease based on the presence or absence of one or more particular nucleotides (alleles) at one or more SNP sites disclosed herein or the detection of one or more

encoded variant products (e.g., variant mRNA transcripts or variant proteins), methods of identifying individuals who are more or less likely to respond to a treatment (or more or less likely to experience undesirable side effects from a treatment, etc.), methods of screening for compounds useful in the treatment of a disorder associated with a variant gene/protein, compounds identified by these methods, methods of treating disorders mediated by a variant gene/protein, methods of using the novel SNPs of the present invention for human identification, etc.

[0039] In Tables 1-2, the present invention provides gene information, transcript sequences (SEQ ID NOS:1-5), encoded amino acid sequences (SEQ ID NOS:6-10), genomic sequences (SEQ ID NOS:16-19), transcript-based context sequences (SEQ ID NOS:11-15) and genomic-based context sequences (SEQ ID NOS:20-31) that contain the SNPs of the present invention, and extensive SNP information that includes observed alleles, allele frequencies, populations/ethnic groups in which alleles have been observed, information about the type of SNP and corresponding functional effect, and, for cSNPs, information about the encoded polypeptide product. The transcript sequences (SEQ ID NOS:1-5), amino acid sequences (SEQ ID NOS:6-10), genomic sequences (SEQ ID NOS:16-19), transcript-based SNP context sequences (SEQ ID NOS:11-15), and genomic-based SNP context sequences (SEQ ID NOS:20-31) are also provided in the Sequence Listing.

[0040] In a specific embodiment of the present invention, SNPs which occur naturally in the human genome are provided as isolated nucleic acid molecules. Either individually or in haplotypes, these SNPs are associated with Alzheimer's Disease such that they can have a variety of uses in the diagnosis and/or treatment of Alzheimer's Disease and related pathologies. One aspect of the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence in which at least one nucleotide is a SNP disclosed in Table(s) 3 and/or 4. In an alternative embodiment, a nucleic acid of the invention is an amplified polynucleotide, which is produced by amplification of a SNP-containing nucleic acid template. In another embodiment, the invention provides for a variant protein which is encoded by a nucleic acid molecule containing a SNP disclosed herein.

[0041] In yet another embodiment of the invention, a reagent for detecting a SNP in the context of its naturally-occurring flanking nucleotide sequences (which can be, e.g., either DNA or mRNA) is provided. In particular, such a reagent may be in the form of, for example, a hybridization probe or an amplification primer that is useful in the specific detection of a SNP of interest. In an alternative embodiment, a protein detection reagent is used to detect a variant protein which is encoded by a nucleic acid molecule containing a SNP disclosed herein. A preferred embodiment of a protein detection reagent is an antibody or an antigen-reactive antibody fragment.

[0042] Various embodiments of the invention also provide kits comprising SNP detection reagents, and methods for detecting the SNPs disclosed herein by employing detection reagents. In a specific embodiment, the present invention provides for a method of identifying an individual having an increased or decreased risk of developing Alzheimer's Disease by detecting the presence or absence of one or more

SNP alleles disclosed herein. In another embodiment, a method for diagnosis of Alzheimer's Disease by detecting the presence or absence of one or more SNP alleles disclosed herein is provided.

[0043] The nucleic acid molecules of the invention can be inserted in an expression vector, such as to produce a variant protein in a host cell. Thus, the present invention also provides for a vector comprising a SNP-containing nucleic acid molecule, genetically-engineered host cells containing the vector, and methods for expressing a recombinant variant protein using such host cells.

[0044] In another specific embodiment, the host cells, SNP-containing nucleic acid molecules, and/or variant proteins can be used as targets in a method for screening and identifying therapeutic agents or pharmaceutical compounds useful in the treatment of Alzheimer's Disease and other neurological pathologies.

[0045] One aspect of this invention is a method for treating Alzheimer's Disease in a human subject wherein said subject harbors a SNP, gene, transcript, and/or encoded protein identified in Tables 1-2, which method comprises administering to said subject a therapeutically or prophylactically effective amount of one or more agents counteracting the effects of the disease, such as by inhibiting (or stimulating) the activity of the gene, transcript, and/or encoded protein identified in Tables 1-2.

[0046] Another aspect of this invention is a method for identifying an agent useful in therapeutically or prophylactically treating Alzheimer's Disease in a human subject wherein said subject harbors a SNP, gene, transcript, and/or encoded protein identified in Tables 1-2, which method comprises contacting the gene, transcript, or encoded protein with a candidate agent under conditions suitable to allow formation of a binding complex between the gene, transcript, or encoded protein and the candidate agent and detecting the formation of the binding complex, wherein the presence of the complex identifies said agent.

[0047] Another aspect of this invention is a method for treating Alzheimer's Disease in a human subject, which method comprises:

[0048] (i) determining that said subject harbors a SNP, SNP haplotypes, gene, transcript, and/or encoded protein identified in Tables 1-2, and

[0049] (ii) administering to said subject a therapeutically or prophylactically effective amount of one or more agents counteracting the effects of the disease.

[0050] Many other uses and advantages of the present invention will be apparent to those skilled in the art upon review of the detailed description of the preferred embodiments herein. Solely for clarity of discussion, the invention is described in the sections below by way of non-limiting examples.

[0051] Description of the Files Contained on the CD-R Named CD000003ORD CDR Duplicate Copy 1 and Duplicate Copy 2

[0052] Each of the CD-Rs contains the following text (ASCII) file:

[0053] File SEQLIST\_000003ORD.TXT provides the Sequence Listing. The Sequence Listing provides the tran-

script sequences (SEQ ID NOS:1-5) and protein sequences (SEQ ID NOS:6-10) as shown in Table 1, and genomic sequences (SEQ ID NOS:16-19) as shown in Table 2, for each Alzheimer's Disease-associated gene that contains one or more SNPs of the present invention. Also provided in the Sequence Listing are context sequences flanking each SNP, including both transcript-based context sequences as shown in Table 1 (SEQ ID NOS:11-15) and genomic-based context sequences as shown in Table 2 (SEQ ID NOS:20-31). The context sequences generally provide 100 bp upstream (5') and 100 bp downstream (3') of each SNP, with the SNP in the middle of the context sequence, for a total of 200 bp of context sequence surrounding each SNP.

[0054] File SEQLIST\_CD000003ORD.TXT is 96 KB in size, and was created on Oct. 20, 2006. A computer readable format of the sequence listing is also submitted herein on a separate CDR labeled CRF. The information recorded in the CRF CDR is identical to the sequence listing as provided on the CDR Duplicate Copy 1 and Duplicate Copy 2.

[0055] The material contained on the CD-R labeled CRF is hereby incorporated by reference pursuant to 37 CFR 1.77(b)(4).

[0056] Description of Table 1 and Table 2

[0057] Table 1 and Table 2 (both provided on the CD-R) disclose the SNP and associated gene/transcript/protein information of the present invention. For each gene, Table 1 and Table 2 each provides a header containing gene/transcript/protein information, followed by a transcript and protein sequence (in Table 1) or genomic sequence (in Table 2), and then SNP information regarding each SNP found in that gene/transcript.

[0058] NOTE: SNPs may be included in both Table 1 and Table 2; Table 1 presents the SNPs relative to their transcript sequences and encoded protein sequences, whereas Table 2 presents the SNPs relative to their genomic sequences (in some instances Table 2 may also include, after the last gene sequence, genomic sequences of one or more intergenic regions, as well as SNP context sequences and other SNP information for any SNPs that lie within these intergenic regions). SNPs can readily be cross-referenced between Tables based on their hCV (or, in some instances, hDV) identification numbers.

[0059] The gene/transcript/protein information includes:

- [0060] a gene number (1 through n, where n=the total number of genes in the Table);
- [0061] a Celera hCG and UID internal identification numbers for the gene;
- [0062] a Celera hCT and UID internal identification numbers for the transcript (Table 1 only);
- [0063] a public Genbank accession number (e.g., RefSeq NM number) for the transcript (Table 1 only);
- [0064] a Celera hCP and UID internal identification numbers for the protein encoded by the hCT transcript (Table 1 only);
- [0065] a public Genbank accession number (e.g., RefSeq NP number) for the protein (Table 1 only);
- [0066] an art-known gene symbol;

[0067] an art-known gene/protein name;

[0068] Celera genomic axis position (indicating start nucleotide position-stop nucleotide position);

[0069] the chromosome number of the chromosome on which the gene is located;

[0070] an OMIM (Online Mendelian Inheritance in Man, Johns Hopkins University/NCBI) public reference number for obtaining further information regarding the medical significance of each gene; and

[0071] the alternative gene/protein name(s) and/or symbol(s) in the OMIM entry.

[0072] NOTE: Due to the presence of alternative splice forms, multiple transcript/protein entries can be provided for a single gene entry in Table 1; i.e., for a single Gene Number, multiple entries may be provided in series that differ in their transcript/protein information and sequences.

[0073] Following the gene/transcript/protein information is a transcript sequence and protein sequence (in Table 1), or a genomic sequence (in Table 2), for each gene, as follows:

[0074] transcript sequence (Table 1 only) (corresponding to SEQ ID NOS:1-5 of the Sequence Listing), with SNPs identified by their IUB codes (transcript sequences can include 5' UTR, protein coding, and 3' UTR regions);

[0075] NOTE: If there are differences between the nucleotide sequence of the hCT transcript and the corresponding public transcript sequence identified by the Genbank accession number, the hCT transcript sequence (and encoded protein) is provided, unless the public sequence is a RefSeq transcript sequence identified by an NM number, in which case the RefSeq NM transcript sequence (and encoded protein) is provided. However, whether the hCT transcript or RefSeq NM transcript is used as the transcript sequence, the disclosed SNPs are represented by their IUB codes within the transcript.

[0076] the encoded protein sequence (Table 1 only) (corresponding to SEQ ID NOS:6-10 of the Sequence Listing); and

[0077] the genomic sequence of the gene (Table 2 only), including 6 kb on each side of the gene boundaries (i.e., 6 kb on the 5' side of the gene plus 6 kb on the 3' side of the gene) (corresponding to SEQ ID NOS:16-19 of the Sequence Listing).

[0078] After the last gene sequence, Table 2 may include additional genomic sequences of intergenic regions (in such instances, these sequences are identified by "Intergenic region:" followed by a numerical identification number), as well as SNP context sequences and other SNP information for any SNPs that lie within each intergenic region (and such SNPs are identified as "INTERGENIC" for SNP type).

[0079] NOTE: The transcript, protein, and transcript-based SNP context sequences are provided in both Table 1 and in the Sequence Listing. The genomic and genomic-based SNP context sequences are provided in both Table 2 and in the Sequence Listing. SEQ ID NOS are indicated in Table 1 for each transcript sequence (SEQ ID NOS:1-5), protein sequence (SEQ ID NOS:6-10), and transcript-based SNP context sequence (SEQ ID NOS:11-15), and SEQ ID

NOS are indicated in Table 2 for each genomic sequence (SEQ ID NOS:16-19), and genomic-based SNP context sequence (SEQ ID NOS:20-31).

[0080] The SNP information includes:

[0081] context sequence (taken from the transcript sequence in Table 1, and taken from the genomic sequence in Table 2) with the SNP represented by its IUB code, including 100 bp upstream (5') of the SNP position plus 100 bp downstream (3') of the SNP position (the transcript-based SNP context sequences in Table 1 are provided in the Sequence Listing as SEQ ID NOS:11-15; the genomic-based SNP context sequences in Table 2 are provided in the Sequence Listing as SEQ ID NOS:20-31);

[0082] Celera hCV internal identification number for the SNP (in some instances, an "hDV" number is given instead of an "hCV" number);

[0083] SNP position (position of the SNP within the given transcript sequence [Table 1] or within the given genomic sequence [Table 2]);

[0084] SNP source (may include any combination of one or more of the following five codes, depending on which internal sequencing projects and/or public databases the SNP has been observed in: "Applera"=SNP observed during the re-sequencing of genes and regulatory regions of 39 individuals, "Celera"=SNP observed during shotgun sequencing and assembly of the Celera human genome sequence, "Celera Diagnostics"=SNP observed during re-sequencing of nucleic acid samples from individuals who have Alzheimer's Disease or a related pathology, "dbSNP"=SNP observed in the dbSNP public database, "HGBASE"=SNP observed in the HGBASE public database, "HGMD"=SNP observed in the Human Gene Mutation Database [HGMD] public database, "HapMap"=SNP observed in the International HapMap Project public database, "CSNP"=SNP observed in an internal Applied Biosystems [Foster City, Calif.] database of coding SNPs [cSNPs]);

[0085] NOTE: multiple "Applera" source entries for a single SNP indicate that the same SNP was covered by multiple overlapping amplification products and the re-sequencing results (e.g., observed allele counts) from each of these amplification products is being provided.

[0086] Population/allele/allele count information in the format of [population1(first\_allele,count|second\_allele,count)population2(first\_allele,count|second\_allele,count) total (first\_allele,total count|second\_allele,total count)]. The information in this field includes populations/ethnic groups in which particular SNP alleles have been observed ("cau"=Caucasian, "his"=Hispanic, "chn"=Chinese, and "afr"=African-American, "jpn"=Japanese, "ind"=Indian, "mex"=Mexican, "ain"="American Indian, "cra"=Celera donor, "no\_pop"=no population information available), identified SNP alleles, and observed allele counts (within each population group and total allele counts), where available ["-" in the allele field represents a deletion allele of an insertion/deletion ("indel") polymorphism (in which case the corresponding insertion allele, which may be comprised of one or more nucleotides, is

indicated in the allele field on the opposite side of the "|"); "-" in the count field indicates that allele count information is not available]. For certain SNPs from the public dbSNP database, population/ethnic information is indicated as follows (this population information is publicly available in dbSNP): "HISP1"=human individual DNA (anonymized samples) from 23 individuals of self-described HISPANIC heritage; "PAC1"=human individual DNA (anonymized samples) from 24 individuals of self-described PACIFIC RIM heritage; "CAUC1"=human individual DNA (anonymized samples) from 31 individuals of self-described CAUCASIAN heritage; "AFR1"=human individual DNA (anonymized samples) from 24 individuals of self-described AFRICAN/AFRICAN AMERICAN heritage; "P1"=human individual DNA (anonymized samples) from 102 individuals of self-described heritage; "PA130299515"; "SC\_12\_A"=SANGER 12 DNAs of Asian origin from Coriell cell repositories, 6 of which are male and 6 female; "SC\_12\_C"=SANGER 12 DNAs of Caucasian origin from Coriell cell repositories from the CEPH/UTAH library. Six male and 6 female; "SC\_12\_AA"=SANGER 12 DNAs of African-American origin from Coriell cell repositories 6 of which are male and 6 female; "SC\_95\_C"=SANGER 95 DNAs of Caucasian origin from Coriell cell repositories from the CEPH/UTAH library; and "SC\_12\_CA"=Caucasians-12 DNAs from Coriell cell repositories that are from the CEPH/UTAH library. Six male and 6 female;

[0087] NOTE: For SNPs of "Applera" SNP source, genes/regulatory regions of 39 individuals (20 Caucasians and 19 African Americans) were re-sequenced and, since each SNP position is represented by two chromosomes in each individual (with the exception of SNPs on X and Y chromosomes in males, for which each SNP position is represented by a single chromosome), up to 78 chromosomes were genotyped for each SNP position. Thus, the sum of the African-American ("afr") allele counts is up to 38, the sum of the Caucasian allele counts ("cau") is up to 40, and the total sum of all allele counts is up to 78.

[0088] NOTE: semicolons separate population/allele/count information corresponding to each indicated SNP source; i.e., if four SNP sources are indicated, such as "Celera," "dbSNP," "HGBASE," and "HGMD," then population/allele/count information is provided in four groups which are separated by semicolons and listed in the same order as the listing of SNP sources, with each population/allele/count information group corresponding to the respective SNP source based on order; thus, in this example, the first population/allele/count information group would correspond to the first listed SNP source (Celera) and the third population/allele/count information group separated by semicolons would correspond to the third listed SNP source (HGBASE); if population/allele/count information is not available for any particular SNP source, then a pair of semicolons is still inserted as a place-holder in order to maintain correspondence between the list of SNP sources and the corresponding listing of population/allele/count information.

[0089] SNP type (e.g., location within gene/transcript and/or predicted functional effect) ("MIS-SENSE MUTATION"=SNP causes a change in the encoded

amino acid [i.e., a non-synonymous coding SNP]; “SILENT MUTATION”=SNP does not cause a change in the encoded amino acid [i.e., a synonymous coding SNP]; “STOP CODON MUTATION”=SNP is located in a stop codon; “NONSENSE MUTATION”=SNP creates or destroys a stop codon; “UTR 5”=SNP is located in a 5' UTR of a transcript; “UTR 3”=SNP is located in a 3' UTR of a transcript; “PUTATIVE UTR 5”=SNP is located in a putative 5' UTR; “PUTATIVE UTR 3”=SNP is located in a putative 3' UTR; “DONOR SPLICE SITE”=SNP is located in a donor splice site [5' intron boundary]; “ACCEPTOR SPLICE SITE”=SNP is located in an acceptor splice site [3' intron boundary]; “CODING REGION”=SNP is located in a protein-coding region of the transcript; “EXON”=SNP is located in an exon; “INTRON”=SNP is located in an intron; “hmCS”=SNP is located in a human-mouse conserved segment; “TFBS”=SNP is located in a transcription factor binding site; “UNKNOWN”=SNP type is not defined; “INTERGENIC”=SNP is intergenic, i.e., outside of any gene boundary); and

[0090] Protein coding information (Table 1 only), where relevant, in the format of (protein SEQ ID NO:#, amino acid position, [amino acid-1, codon1][amino acid-2, codon2]). The information in this field includes SEQ ID NO of the encoded protein sequence, position of the amino acid residue within the protein identified by the SEQ ID NO that is encoded by the codon containing the SNP, amino acids (represented by one-letter amino acid codes) that are encoded by the alternative SNP alleles (in the case of stop codons, “X” is used for the one-letter amino acid code), and alternative codons containing the alternative SNP nucleotides which encode the amino acid residues (thus, for example, for missense mutation-type SNPs, at least two different amino acids and at least two different codons are generally indicated; for silent mutation-type SNPs, one amino acid and at least two different codons are generally indicated, etc.). In instances where the SNP is located outside of a protein-coding region (e.g., in a UTR region), “None” is indicated following the protein SEQ ID NO.

[0091] Description of Table 3 and Table 4

[0092] Tables 3 and 4 show the statistical results of the association of the two non-DAPK1 SNPs with Alzheimer's Disease in three sample sets (see Example: Statistical Analysis of SNPs Associated with Alzheimer's Disease, below). The two SNPs are hCV8715115, and hCV1920609.

[0093] Table 3 shows the allelic test analysis of the POMT1 SNP rs2018621 (hCV8715115). Table 4 shows the allelic test analysis of the DFN31 SNP rs2274159 (hCV1920609). See legends for the tables for column headings descriptions.

[0094] Description of Table 5 and Table 6

[0095] Tables 5-6 show the statistical results of the association of DAPK1 SNPs with Alzheimer's Disease in three sample sets (see Example: Statistical Analysis of SNPs Associated with Alzheimer's Disease, below).

[0096] Table 5 shows the number of case and control samples for the various genotypes analyzed for

hCV1386982 and hCV1386888, and the association of each genotype with Alzheimer's Disease. In this table “Allelic P”=p Value, and “Allelic OR”=Odds Ratio.

[0097] Table 6 shows statistical results for the haplotype analysis of DAPK1 markers hCV1386973, hCV1386978, hCV1386982. Definitions for the column headings are as follows: “Sample”=study set, “Haplotype”=haplotype analyzed, “control\_freq”=frequency of haplotype in control samples, “case\_freq”=frequency in cases, “Haplotype P”=p Value measure of association for the haplotype, and “Global”=global pValue for this haplotype (haplotype and global p Values are described in Study Design, below).

[0098] Description of Table 7

[0099] Table 7 provides a list of the sample LD SNPs that are related to and derived from an interrogated SNP. These LD SNPs are provided as an example of the groups of SNPs which can also serve as markers for disease association based on their being in LD with the interrogated SNP. The criteria and process of selecting such LD SNPs, including the calculation of the  $r^2$  value and the  $r^2$  threshold value, are described in Example, below.

[0100] In Table 7, the column labeled “Interrogated SNP” presents each marker as identified by its unique identifier, the hCV number. The column labeled “Interrogated rs” presents the publicly known identifier rs number for the corresponding hCV number. The column labeled “LD SNP” presents the hCV numbers of the LD SNPs that are derived from their corresponding interrogated SNPs. The column labeled “LD SNP rs” presents the publicly known rs number for the corresponding hCV number. The column labeled “Power (T)” presents the level of power where the  $r^2$  threshold is set. For example, when power is set at 51%, the threshold  $r^2$  value calculated therefrom is the minimum  $r^2$  that an LD SNP must have in reference to an interrogated SNP, in order for the LD SNP to be classified as a marker capable of being associated with a disease phenotype at greater than 51% probability. The column labeled “Threshold  $r_T^{2*}$ ” presents the minimum value of  $r^2$  that an LD SNP must meet in reference to an interrogated SNP in order to qualify as an LD SNP. The column labeled “ $r^2$ ” presents the actual  $r^2$  value of the LD SNP in reference to the interrogated SNP to which it is related.

#### DESCRIPTION OF THE FIGURES

[0101] FIG. 1 provides a diagrammatic representation of a computer-based discovery system containing the SNP information of the present invention in computer readable form.

[0102] FIG. 2 shows the association of exploratory markers on chromosome 9 with LOAD. The p Values of 674 SNPs were plotted against their physical positions based on Celera's human genome assembly, version R27. A horizontal line at p Value=0.05 was drawn to distinguish significant markers from non-significant ones. The previously identified linkage peak regions were noted with solid horizontal lines along with references. The arrow points to the DAPK1 SNP hCV1386982.

[0103] FIG. 3 shows the various hCV1386982 genotypes and their association with LOAD risk. The odds ratios (OR) for the minor allele homozygotes (TT) and heterozygotes (TC) are grouped by study set: WU, SD, UK, and the 3 sets combined (All).

[0104] FIG. 4 consists of FIGS. 4A and 4B: DAPK1 allele-specific gene expression stratified by the LOAD-associated markers hCV1386982 or hCV1386888. (A) Allele-specific transcript ratios that are different from one require heterozygosity of the causal variant (X,Y). Markers that show association with allele-specific expression, when stratified by genotype, are expected to be in strong LD with the causal variant or represent the causal variant itself. Markers that are not in LD with the causal variant are unlikely to show association with allele-specific expression. A1, A2: allele-specific expression marker. Note that the alleles (X/Y) of the causal SNP can occur on either allele specific transcript, if the causal SNP is not in perfect LD with the expression marker (A1/A2). (B) The ratio of allele-specific gene expression is shown for cDNA. Two markers, hCV2704931 and hCV2704861, were analyzed to measure allelic gene expression (Hom: homozygotes; Het, heterozygotes). A Mann-Whitney test was performed to assess the association. The relative expression was calculated as  $2^{-\Delta C_t}$  ( $\Delta C_t$  was determined by subtracting the smaller  $C_t$  value of one allele PCR reaction from the larger  $C_t$  value of the other allele PCR reaction).

#### DETAILED DESCRIPTION OF THE INVENTION

[0105] The present invention provides SNPs associated with Alzheimer's Disease, nucleic acid molecules containing these SNPs, methods and reagents for the detection of the SNPs disclosed herein, uses of these SNPs for the development of detection reagents, and assays or kits that utilize such reagents. The AD-associated SNPs disclosed herein are useful for diagnosing, screening for, and evaluating predisposition to Alzheimer's Disease and other neurological pathologies in humans. Furthermore, such SNPs and their encoded products are useful targets for the development of therapeutic agents in treating Alzheimer's Disease and other neurological pathologies.

[0106] A large number of SNPs have been identified from re-sequencing DNA from 39 individuals, and they are indicated as "Applera" SNP source in Tables 1-2. Their allele frequencies, observed in each of the Caucasian and African-American ethnic groups, are provided. Additional SNPs included herein were previously identified during shotgun sequencing and assembly of the human genome, and they are indicated as "Celera" SNP source in Tables 1-2. Furthermore, the information provided in Table 1-2, particularly the allele frequency information obtained from 39 individuals and the identification of the precise position of each SNP within each gene/transcript, allows haplotypes (i.e., groups of SNPs that are co-inherited) to be readily inferred. The present invention encompasses SNP haplotypes, as well as individual SNPs.

[0107] Thus, the present invention provides individual SNPs associated with Alzheimer's Disease, as well as combinations of SNPs and haplotypes in genetic regions associated with Alzheimer's Disease, polymorphic/variant transcript sequences (SEQ ID NOS:1-5) and genomic sequences (SEQ ID NOS:16-19) containing SNPs, encoded amino acid sequences (SEQ ID NOS: 6-10), and both transcript-based SNP context sequences (SEQ ID NOS: 11-15) and genomic-based SNP context sequences (SEQ ID NOS:20-31) (transcript sequences, protein sequences, and transcript-based SNP context sequences are provided in Table 1 and the

Sequence Listing; genomic sequences and genomic-based SNP context sequences are provided in Table 2 and the Sequence Listing), methods of detecting these polymorphisms in a test sample, methods of determining the risk of an individual of having or developing Alzheimer's Disease, methods of screening for compounds useful for treating neurological pathologies such as Alzheimer's Disease associated with a variant gene/protein, compounds identified by these screening methods, methods of using the disclosed SNPs to select a treatment strategy, methods of treating a disorder associated with a variant gene/protein (i.e., therapeutic methods), and methods of using the SNPs of the present invention for human identification.

[0108] The present invention provides novel SNPs associated with Alzheimer's Disease, as well as SNPs that were previously known in the art, but were not previously known to be associated with Alzheimer's Disease. Accordingly, the present invention provides novel compositions and methods based on the novel SNPs disclosed herein, and also provides novel methods of using the known, but previously unassociated, SNPs in methods relating to Alzheimer's Disease (e.g., for diagnosing Alzheimer's Disease). In Tables 1-2, known SNPs are identified based on the public database in which they have been observed, which is indicated as one or more of the following SNP types: "dbSNP"=SNP observed in dbSNP, "HGBASE"=SNP observed in HGBASE, and "HGMD"=SNP observed in the Human Gene Mutation Database (HGMD).

[0109] Particular SNP alleles of the present invention can be associated with either an increased risk of having or developing Alzheimer's Disease, or a decreased risk of having or developing Alzheimer's Disease. SNP alleles that are associated with a decreased risk of having or developing Alzheimer's Disease may be referred to as "protective" alleles, and SNP alleles that are associated with an increased risk of having or developing Alzheimer's Disease may be referred to as "susceptibility" alleles, "risk" alleles, or "risk factors." Thus, whereas certain SNPs (or their encoded products) can be assayed to determine whether an individual possesses a SNP allele that is indicative of an increased risk of having or developing Alzheimer's Disease (i.e., a susceptibility allele), other SNPs (or their encoded products) can be assayed to determine whether an individual possesses a SNP allele that is indicative of a decreased risk of having or developing Alzheimer's Disease (i.e., a protective allele). Similarly, particular SNP alleles of the present invention can be associated with either an increased or decreased likelihood of responding to a particular treatment or therapeutic compound, or an increased or decreased likelihood of experiencing toxic effects from a particular treatment or therapeutic compound. The term "altered" may be used herein to encompass either of these two possibilities (e.g., an increased or a decreased risk/likelihood).

[0110] Those skilled in the art will readily recognize that nucleic acid molecules may be double-stranded molecules and that reference to a particular site on one strand refers, as well, to the corresponding site on a complementary strand. In defining a SNP position, SNP allele, or nucleotide sequence, reference to an adenine, a thymine (uridine), a cytosine, or a guanine at a particular site on one strand of a nucleic acid molecule also defines the thymine (uridine), adenine, guanine, or cytosine (respectively) at the corresponding site on a complementary strand of the nucleic acid

molecule. Thus, reference may be made to either strand in order to refer to a particular SNP position, SNP allele, or nucleotide sequence. Probes and primers, may be designed to hybridize to either strand and SNP genotyping methods disclosed herein may generally target either strand. Throughout the specification, in identifying a SNP position, reference is generally made to the protein-encoding strand, only for the purpose of convenience.

[0111] References to variant peptides, polypeptides, or proteins of the present invention include peptides, polypeptides, proteins, or fragments thereof, that contain at least one amino acid residue that differs from the corresponding amino acid sequence of the art-known peptide/polypeptide/protein (the art-known protein may be interchangeably referred to as the “wild-type,” “reference,” or “normal” protein). Such variant peptides/polypeptides/proteins can result from a codon change caused by a nonsynonymous nucleotide substitution at a protein-coding SNP position (i.e., a missense mutation) disclosed by the present invention. Variant peptides/polypeptides/proteins of the present invention can also result from a nonsense mutation, i.e. a SNP that creates a premature stop codon, a SNP that generates a read-through mutation by abolishing a stop codon, or due to any SNP disclosed by the present invention that otherwise alters the structure, function/activity, or expression of a protein, such as a SNP in a regulatory region (e.g. a promoter or enhancer) or a SNP that leads to alternative or defective splicing, such as a SNP in an intron or a SNP at an exon/intron boundary. As used herein, the terms “polypeptide,” “peptide,” and “protein” are used interchangeably.

#### Isolated Nucleic Acid Molecules and SNP Detection Reagents & Kits

[0112] Tables 1 and 2 provide a variety of information about each SNP of the present invention that is associated with Alzheimer’s Disease, including the transcript sequences (SEQ ID NOS:1-5), genomic sequences (SEQ ID NOS:16-19), and protein sequences (SEQ ID NOS:6-10) of the encoded gene products (with the SNPs indicated by IUB codes in the nucleic acid sequences). In addition, Tables 1 and 2 include SNP context sequences, which generally include 100 nucleotide upstream (5’) plus 100 nucleotides downstream (3’) of each SNP position (SEQ ID NOS:11-15 correspond to transcript-based SNP context sequences disclosed in Table 1, and SEQ ID NOS:20-31 correspond to genomic-based context sequences disclosed in Table 2), the alternative nucleotides (alleles) at each SNP position, and additional information about the variant where relevant, such as SNP type (coding, missense, splice site, UTR, etc.), human populations in which the SNP was observed, observed allele frequencies, information about the encoded protein, etc.

#### [0113] Isolated Nucleic Acid Molecules

[0114] The present invention provides isolated nucleic acid molecules that contain one or more SNPs disclosed Table 1 and/or Table 2. Isolated nucleic acid molecules containing one or more SNPs disclosed in at least one of Tables 1-4 may be interchangeably referred to throughout the present text as “SNP-containing nucleic acid molecules.” Isolated nucleic acid molecules may optionally encode a full-length variant protein or fragment thereof. The isolated nucleic acid molecules of the present invention also include probes and primers (which are described in greater detail

below in the section entitled “SNP Detection Reagents”), which may be used for assaying the disclosed SNPs, and isolated full-length genes, transcripts, cDNA molecules, and fragments thereof, which may be used for such purposes as expressing an encoded protein.

[0115] As used herein, an “isolated nucleic acid molecule” generally is one that contains a SNP of the present invention or one that hybridizes to such molecule such as a nucleic acid with a complementary sequence, and is separated from most other nucleic acids present in the natural source of the nucleic acid molecule. Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule containing a SNP of the present invention, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered “isolated.” Nucleic acid molecules present in non-human transgenic animals, which do not naturally occur in the animal, are also considered “isolated.” For example, recombinant DNA molecules contained in a vector are considered “isolated.” Further examples of “isolated” DNA molecules include recombinant DNA molecules maintained in heterologous host cells, and purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated SNP-containing DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[0116] Generally, an isolated SNP-containing nucleic acid molecule comprises one or more SNP positions disclosed by the present invention with flanking nucleotide sequences on either side of the SNP positions. A flanking sequence can include nucleotide residues that are naturally associated with the SNP site and/or heterologous nucleotide sequences. Preferably the flanking sequence is up to about 500, 300, 100, 60, 50, 30, 25, 20, 15, 10, 8, or 4 nucleotides (or any other length in-between) on either side of a SNP position, or as long as the full-length gene or entire protein-coding sequence (or any portion thereof such as an exon), especially if the SNP-containing nucleic acid molecule is to be used to produce a protein or protein fragment.

[0117] For full-length genes and entire protein-coding sequences, a SNP flanking sequence can be, for example, up to about 5 KB, 4 KB, 3 KB, 2 KB, 1 KB on either side of the SNP. Furthermore, in such instances, the isolated nucleic acid molecule comprises exonic sequences (including protein-coding and/or non-coding exonic sequences), but may also include intronic sequences. Thus, any protein coding sequence may be either contiguous or separated by introns. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences and is of appropriate length such that it can be subjected to the specific manipulations or uses described herein such as recombinant protein expression, preparation of probes and primers for assaying the SNP position, and other uses specific to the SNP-containing nucleic acid sequences.

[0118] An isolated SNP-containing nucleic acid molecule can comprise, for example, a full-length gene or transcript, such as a gene isolated from genomic DNA (e.g., by cloning or polymerase chain reaction [PCR] amplification), a cDNA

molecule, or an mRNA transcript molecule. Polymorphic transcript sequences are provided in Table 1 and in the Sequence Listing (SEQ ID NOS: 1-5), and polymorphic genomic sequences are provided in Table 2 and in the Sequence Listing (SEQ ID NOS:16-19). Furthermore, fragments of such full-length genes and transcripts that contain one or more SNPs disclosed herein are also encompassed by the present invention, and such fragments may be used, for example, to express any part of a protein, such as a particular functional domain or an antigenic epitope.

[0119] Thus, the present invention also encompasses fragments of the nucleic acid sequences provided in Tables 1-2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-5, genomic sequences are provided in Table 2 as SEQ ID NOS:16-19, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:11-15, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:20-31) and their complements. A fragment typically comprises a contiguous nucleotide sequence at least about eight or more nucleotides, more preferably at least about twelve or more nucleotides, and even more preferably at least about sixteen or more nucleotides. Further, a fragment could comprise at least about 18, 20, 22, 25, 30, 40, 50, 60, 80, 100, 150, 200, 250 or 500 nucleotides in length, or any other number in between. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope-bearing regions of a variant peptide or regions of a variant peptide that differ from the normal/wild-type protein, or can be useful as a polynucleotide probe or primer. Such fragments can be isolated using the nucleotide sequences provided in Table 1 and/or Table 2 for the synthesis of a polynucleotide probe. A labeled probe can then be used, for example, to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in amplification reactions, such as for purposes of assaying one or more SNPs sites or for cloning specific regions of a gene.

[0120] An isolated nucleic acid molecule of the present invention further encompasses a SNP-containing polynucleotide that is the product of any one of a variety of nucleic acid amplification methods, which are used to increase the copy numbers of a polynucleotide of interest in a nucleic acid sample. Such amplification methods are well known in the art, and they include but are not limited to, polymerase chain reaction (PCR) (U.S. Pat. Nos. 4,683,195 and 4,683,202; *PCR Technology: Principles and Applications for DNA Amplification*, ed. H. A. Erlich, Freeman Press, New York, N.Y. [1992]), ligase chain reaction (LCR) (Wu and Wallace, *Genomics* 4:560 [1989]; Landegren et al., *Science* 241:1077 [1988]), strand displacement amplification (SDA) (U.S. Pat. Nos. 5,270,184 and 5,422,252), transcription-mediated amplification (TMA) (U.S. Pat. No. 5,399,491), linked linear amplification (LLA) (U.S. Pat. No. 6,027,923), and the like, and isothermal amplification methods such as nucleic acid sequence based amplification (NASBA), and self-sustained sequence replication (Guatelli et al., *Proc. Natl. Acad. Sci. USA* 87: 1874 [1990]). Based on such methodologies, a person skilled in the art can readily design primers in any suitable regions 5' and 3' to a SNP disclosed herein. Such primers may be used to amplify DNA of any length so long that it contains the SNP of interest in its sequence.

[0121] As used herein, an "amplified polynucleotide" of the invention is a SNP-containing nucleic acid molecule whose amount has been increased at least two fold by any nucleic acid amplification method performed in vitro as compared to its starting amount in a test sample. In other preferred embodiments, an amplified polynucleotide is the result of at least ten fold, fifty fold, one hundred fold, one thousand fold, or even ten thousand fold increase as compared to its starting amount in a test sample. In a typical PCR amplification, a polynucleotide of interest is often amplified at least fifty thousand fold in amount over the unamplified genomic DNA, but the precise amount of amplification needed for an assay depends on the sensitivity of the subsequent detection method used.

[0122] Generally, an amplified polynucleotide is at least about 16 nucleotides in length. More typically, an amplified polynucleotide is at least about 20 nucleotides in length. In a preferred embodiment of the invention, an amplified polynucleotide is at least about 30 nucleotides in length. In a more preferred embodiment of the invention, an amplified polynucleotide is at least about 32, 40, 45, 50, or 60 nucleotides in length. In yet another preferred embodiment of the invention, an amplified polynucleotide is at least about 100, 200, 300, 400, or 500 nucleotides in length. While the total length of an amplified polynucleotide of the invention can be as long as an exon, an intron or the entire gene where the SNP of interest resides, an amplified product is typically up to about 1,000 nucleotides in length (although certain amplification methods may generate amplified products greater than 1000 nucleotides in length). More preferably, an amplified polynucleotide is not greater than about 600-700 nucleotides in length. It is understood that irrespective of the length of an amplified polynucleotide, a SNP of interest may be located anywhere along its sequence.

[0123] In a specific embodiment of the invention, the amplified product is at least about 201 nucleotides in length, and comprises one of the transcript-based context sequences or the genomic-based context sequences shown in Tables 1-2. Such a product may have additional sequences on its 5' end or 3' end or both. In another embodiment, the amplified product is about 101 nucleotides in length, and it contains a SNP disclosed herein. Preferably, the SNP is located at the middle of the amplified product (e.g., at position 101 in an amplified product that is 201 nucleotides in length, or at position 51 in an amplified product that is 101 nucleotides in length), or within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, or 20 nucleotides from the middle of the amplified product (however, as indicated above, the SNP of interest may be located anywhere along the length of the amplified product).

[0124] The present invention provides isolated nucleic acid molecules that comprise, consist of, or consist essentially of one or more polynucleotide sequences that contain one or more SNPs disclosed herein, complements thereof, and SNP-containing fragments thereof.

[0125] Accordingly, the present invention provides nucleic acid molecules that consist of any of the nucleotide sequences shown in Table 1 and/or Table 2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-5, genomic sequences are provided in Table 2 as SEQ ID NOS:16-19, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:11-15, and genomic-based SNP context sequences are provided in Table 2 as

SEQ ID NO:20-31), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:6-10). A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

**[0126]** The present invention further provides nucleic acid molecules that consist essentially of any of the nucleotide sequences shown in Table 1 and/or Table 2 (transcript sequences are provided in Table 1 as SEQ ID NOS:1-5, genomic sequences are provided in Table 2 as SEQ ID NOS:16-19, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:11-15, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:20-31), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:6-10). A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleotide residues in the final nucleic acid molecule.

**[0127]** The present invention further provides nucleic acid molecules that comprise any of the nucleotide sequences shown in Table 1 and/or Table 2 or a SNP-containing fragment thereof (transcript sequences are provided in Table 1 as SEQ ID NOS:1-5, genomic sequences are provided in Table 2 as SEQ ID NOS:16-19, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:11-15, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:20-31), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:6-10). A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleotide residues, such as residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have one to a few additional nucleotides or can comprise many more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made and isolated is provided below, and such techniques are well known to those of ordinary skill in the art (*Molecular Cloning: A Laboratory Manual*, Sambrook and Russell, Cold Spring Harbor Press, New York [2000]).

**[0128]** The isolated nucleic acid molecules can encode mature proteins plus additional amino or carboxyl-terminal amino acids or both, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life, or facilitate manipulation of a protein for assay or production. As generally is the case in situ, the additional amino acids may be processed away from the mature protein by cellular enzymes.

**[0129]** Thus, the isolated nucleic acid molecules include, but are not limited to, nucleic acid molecules having a sequence encoding a peptide alone, a sequence encoding a mature peptide and additional coding sequences such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), a sequence encoding a mature peptide with or without additional coding sequences, plus additional non-

coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but untranslated sequences that play a role in, for example, transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding, and/or stability of mRNA. In addition, the nucleic acid molecules may be fused to heterologous marker sequences encoding, for example, a peptide that facilitates purification.

**[0130]** Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA, which may be obtained, for example, by molecular cloning or produced by chemical synthetic techniques or by a combination thereof (*Molecular Cloning: A Laboratory Manual*, Sambrook and Russell, Cold Spring Harbor Press, New York [2000]). Furthermore, isolated nucleic acid molecules, particularly SNP detection reagents such as probes and primers, can also be partially or completely in the form of one or more types of nucleic acid analogs, such as peptide nucleic acid (PNA) (U.S. Pat. Nos. 5,539,082; 5,527,675; 5,623,049; 5,714,331). The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the complementary non-coding strand (anti-sense strand). DNA, RNA, or PNA segments can be assembled, for example, from fragments of the human genome (in the case of DNA or RNA) or single nucleotides, short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic nucleic acid molecule. Nucleic acid molecules can be readily synthesized using the sequences provided herein as a reference; oligonucleotide and PNA oligomer synthesis techniques are well known in the art (see, e.g., Corey, "Peptide nucleic acids: expanding the scope of nucleic acid recognition," *Trends Biotechnol.* 15[6]:224-9 [June 1997], and Hyrup et al., "Peptide nucleic acids [PNA]: synthesis, properties and potential applications," *Bioorg. Med. Chem.* 4[1]:5-23 [January 1996]). Furthermore, large-scale automated oligonucleotide/PNA synthesis (including synthesis on an array or bead surface or other solid support) can readily be accomplished using commercially available nucleic acid synthesizers, such as the Applied Biosystems (Foster City, Calif.) 3900 High-Throughput DNA Synthesizer or Expedite 8909 Nucleic Acid Synthesis System, and the sequence information provided herein.

**[0131]** The present invention encompasses nucleic acid analogs that contain modified, synthetic, or non-naturally occurring nucleotides or structural elements or other alternative/modified nucleic acid chemistries known in the art. Such nucleic acid analogs are useful, for example, as detection reagents (e.g., primers/probes) for detecting one or more SNPs identified in Table 1 and/or Table 2. Furthermore, kits/systems (such as beads, arrays, etc.) that include these analogs are also encompassed by the present invention. For example, PNA oligomers that are based on the polymorphic sequences of the present invention are specifically contemplated. PNA oligomers are analogs of DNA in which the phosphate backbone is replaced with a peptide-like backbone (Lagriffoul et al., *Bioorganic & Medicinal Chemistry Letters* 4:1081-1082 [1994], Petersen et al., *Bioorganic & Medicinal Chemistry Letters* 6:793-796 [1996], Kumar et al., *Organic Letters* 3[9]:1269-1272 [2001], WO96/04000). PNA hybridizes to complementary RNA or DNA with higher affinity and specificity than conventional oligonucleotides and oligonucleotide analogs. The properties of PNA

enable novel molecular biology and biochemistry applications unachievable with traditional oligonucleotides and peptides.

[0132] Additional examples of nucleic acid modifications that improve the binding properties and/or stability of a nucleic acid include the use of base analogs such as inosine, intercalators (U.S. Pat. No. 4,835,263) and the minor groove binders (U.S. Pat. No. 5,801,115). Thus, references herein to nucleic acid molecules, SNP-containing nucleic acid molecules, SNP detection reagents (e.g., probes and primers), oligonucleotides/polynucleotides include PNA oligomers and other nucleic acid analogs. Other examples of nucleic acid analogs and alternative/modified nucleic acid chemistries known in the art are described in *Current Protocols in Nucleic Acid Chemistry*, John Wiley & Sons, New York (2002).

[0133] The present invention further provides nucleic acid molecules that encode fragments of the variant polypeptides disclosed herein as well as nucleic acid molecules that encode obvious variants of such variant polypeptides. Such nucleic acid molecules may be naturally occurring, such as paralogs (different locus) and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, the variants can contain nucleotide substitutions, deletions, inversions and insertions (in addition to the SNPs disclosed in Tables 1-2). Variation can occur in either or both the coding and non-coding regions. The variations can produce conservative and/or non-conservative amino acid substitutions.

[0134] Further variants of the nucleic acid molecules disclosed in Tables 1-2, such as naturally occurring allelic variants (as well as orthologs and paralogs) and synthetic variants produced by mutagenesis techniques, can be identified and/or produced using methods well known in the art. Such further variants can comprise a nucleotide sequence that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with a nucleic acid sequence disclosed in Table 1 and/or Table 2 (or a fragment thereof) and that includes a novel SNP allele disclosed in Table 1 and/or Table 2. Further, variants can comprise a nucleotide sequence that encodes a polypeptide that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with a polypeptide sequence disclosed in Table 1 (or a fragment thereof) and that includes a novel SNP allele disclosed in Table 1 and/or Table 2. Thus, an aspect of the present invention that is specifically contemplated is isolated nucleic acid molecules that have a certain degree of sequence variation compared with the sequences shown in Tables 1-2, but that contain a novel SNP allele disclosed herein. In other words, as long as an isolated nucleic acid molecule contains a novel SNP allele disclosed herein, other portions of the nucleic acid molecule that flank the novel SNP allele can vary to some degree from the specific transcript, genomic, and context sequences shown in Tables 1-2, and can encode a polypeptide that varies to some degree from the specific polypeptide sequences shown in Table 1.

[0135] To determine the percent identity of two amino acid sequences or two nucleotide sequences of two mol-

ecules that share sequence homology, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0136] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm (Computational Molecular Biology, ed. A. M. Lesk, Oxford University Press, New York [1988]; *Biocomputing: Informatics and Genome Projects*, ed. Smith, D. W., Academic Press, New York [1993]; *Computer Analysis of Sequence Data, Part 1*, ed. A. M. Griffin, and H. G. Griffin, Humana Press, New Jersey [1994]; *Sequence Analysis in Molecular Biology*, G. von Heinje, Academic Press [1987]; and *Sequence Analysis Primer*, eds. M. Gribskov and J. M. Devereux, Stockton Press, New York [1991]). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch algorithm (J. Mol. Biol. 48:444-453 [1970]) which has been incorporated into the GAP program in the GCG software package, using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

[0137] In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (J. Devereux et al., *Nucleic Acids Res.* 12[1]:387 [1984]), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 [1989]) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4.

[0138] The nucleotide and amino acid sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul et al. (*J. Mol. Biol.* 215:403-10 [1990]). BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino

acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25[17]:3389-3402 [1997]). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. In addition to BLAST, examples of other search and sequence comparison programs used in the art include, but are not limited to, FASTA (Pearson, *Methods Mol. Biol.* 25, 365-389 [1994]) and KERR (Dufresne et al., *Nat. Biotechnol.* 20[12]:1269-71 [December 2002]). For further information regarding bioinformatics techniques, see *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., New York.

[0139] The present invention further provides non-coding fragments of the nucleic acid molecules disclosed in Table 1 and/or Table 2. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, intronic sequences, 5' untranslated regions (UTRs), 3' untranslated regions, gene modulating sequences and gene termination sequences. Such fragments are useful, for example, in controlling heterologous gene expression and in developing screens to identify gene-modulating agents.

#### [0140] SNP Detection Reagents

[0141] In a specific aspect of the present invention, the SNPs disclosed in Table 1 and/or Table 2, and their associated transcript sequences (provided in Table 1 as SEQ ID NOS:1-5), genomic sequences (provided in Table 2 as SEQ ID NOS:16-19), and context sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:11-15; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:20-31), can be used for the design of SNP detection reagents. As used herein, a "SNP detection reagent" is a reagent that specifically detects a specific target SNP position disclosed herein, and that is preferably specific for a particular nucleotide (allele) of the target SNP position (i.e., the detection reagent preferably can differentiate between different alternative nucleotides at a target SNP position, thereby allowing the identity of the nucleotide present at the target SNP position to be determined). Typically, such detection reagent hybridizes to a target SNP-containing nucleic acid molecule by complementary base-pairing in a sequence specific manner, and discriminates the target variant sequence from other nucleic acid sequences such as an art-known form in a test sample. An example of a detection reagent is a probe that hybridizes to a target nucleic acid containing one or more of the SNPs provided in Table 1 and/or Table 2. In a preferred embodiment, such a probe can differentiate between nucleic acids having a particular nucleotide (allele) at a target SNP position from other nucleic acids that have a different nucleotide at the same target SNP position. In addition, a detection reagent may hybridize to a specific region 5' and/or 3' to a SNP position, particularly a region corresponding to the context sequences provided in Table 1 and/or Table 2 (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:11-15; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:20-31). Another example of a detection reagent is a primer which acts as an initiation point of nucleotide extension along a complementary strand of a target polynucleotide. The SNP sequence information pro-

vided herein is also useful for designing primers, e.g. allele-specific primers, to amplify (e.g., using PCR) any SNP of the present invention.

[0142] In one preferred embodiment of the invention, a SNP detection reagent is an isolated or synthetic DNA or RNA polynucleotide probe or primer or PNA oligomer, or a combination of DNA, RNA and/or PNA, that hybridizes to a segment of a target nucleic acid molecule containing a SNP identified in Table 1 and/or Table 2. A detection reagent in the form of a polynucleotide may optionally contain modified base analogs, intercalators or minor groove binders. Multiple detection reagents such as probes may be, for example, affixed to a solid support (e.g., arrays or beads) or supplied in solution (e.g., probe/primer sets for enzymatic reactions such as PCR, RT-PCR, TaqMan assays, or primer-extension reactions) to form a SNP detection kit.

[0143] A probe or primer typically is a substantially purified oligonucleotide or PNA oligomer. Such oligonucleotide typically comprises a region of complementary nucleotide sequence that hybridizes under stringent conditions to at least about 8, 10, 12, 16, 18, 20, 22, 25, 30, 40, 50, 55, 60, 65, 70, 80, 90, 100, 120 (or any other number in-between) or more consecutive nucleotides in a target nucleic acid molecule. Depending on the particular assay, the consecutive nucleotides can either include the target SNP position, or be a specific region in close enough proximity 5' and/or 3' to the SNP position to carry out the desired assay.

[0144] Other preferred primer and probe sequences can readily be determined using the transcript sequences (SEQ ID NOS:1-5), genomic sequences (SEQ ID NOS:16-19), and SNP context sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:11-15; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:20-31) disclosed in the Sequence Listing and in Tables 1-2. It will be apparent to one of skill in the art that such primers and probes are directly useful as reagents for genotyping the SNPs of the present invention, and can be incorporated into any kit/system format.

[0145] In order to produce a probe or primer specific for a target SNP-containing sequence, the gene/transcript and/or context sequence surrounding the SNP of interest is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene/SNP context sequence, have a GC content within a range suitable for hybridization, lack predicted secondary structure that may interfere with hybridization, and/or possess other desired characteristics or that lack other undesired characteristics.

[0146] A primer or probe of the present invention is typically at least about 8 nucleotides in length. In one embodiment of the invention, a primer or a probe is at least about 10 nucleotides in length. In a preferred embodiment, a primer or a probe is at least about 12 nucleotides in length. In a more preferred embodiment, a primer or probe is at least about 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length. While the maximal length of a probe can be as long as the target sequence to be detected, depending on the type of assay in which it is employed, it is typically less than about 50, 60, 65, or 70 nucleotides in length. In the case of a primer, it is typically less than about 30 nucleotides in length. In a specific preferred embodiment of the invention,

a primer or a probe is within the length of about 18 and about 28 nucleotides. However, in other embodiments, such as nucleic acid arrays and other embodiments in which probes are affixed to a substrate, the probes can be longer, such as on the order of 30-70, 75, 80, 90, 100, or more nucleotides in length (see the section below entitled "SNP Detection Kits and Systems").

[0147] For analyzing SNPs, it may be appropriate to use oligonucleotides specific for alternative SNP alleles. Such oligonucleotides which detect single nucleotide variations in target sequences may be referred to by such terms as "allele-specific oligonucleotides," "allele-specific probes," or "allele-specific primers." The design and use of allele-specific probes for analyzing polymorphisms is described in, e.g., *Mutation Detection: A Practical Approach*, ed. Cotton et al., Oxford University Press [1998]; Saiki et al., *Nature* 324, 163-166 [1986]; Dattagupta, EP235,726; and Saiki, WO 89/11548.

[0148] While the design of each allele-specific primer or probe depends on variables such as the precise composition of the nucleotide sequences flanking a SNP position in a target nucleic acid molecule, and the length of the primer or probe, another factor in the use of primers and probes is the stringency of the condition under which the hybridization between the probe or primer and the target sequence is performed. Higher stringency conditions utilize buffers with lower ionic strength and/or a higher reaction temperature, and tend to require a more perfect match between probe/primer and a target sequence in order to form a stable duplex. If the stringency is too high, however, hybridization may not occur at all. In contrast, lower stringency conditions utilize buffers with higher ionic strength and/or a lower reaction temperature, and permit the formation of stable duplexes with more mismatched bases between a probe/primer and a target sequence. By way of example and not limitation, exemplary conditions for high stringency hybridization conditions using an allele-specific probe are as follows: Prehybridization with a solution containing 5× standard saline phosphate EDTA (SSPE), 0.5% NaDodSO<sub>4</sub> (SDS) at 55° C., and incubating probe with target nucleic acid molecules in the same solution at the same temperature, followed by washing with a solution containing 2×SSPE, and 0.1% SDS at 55° C. or room temperature.

[0149] Moderate stringency hybridization conditions may be used for allele-specific primer extension reactions with a solution containing, e.g., about 50 mM KCl at about 46° C. Alternatively, the reaction may be carried out at an elevated temperature such as 60° C. In another embodiment, a moderately stringent hybridization condition suitable for oligonucleotide ligation assay (OLA) reactions wherein two probes are ligated if they are completely complementary to the target sequence may utilize a solution of about 100 mM KCl at a temperature of 46° C.

[0150] In a hybridization-based assay, allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms (e.g., alternative SNP alleles/nucleotides) in the respective DNA segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant detectable difference in hybridization intensity between alleles, and

preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles or significantly more strongly to one allele. While a probe may be designed to hybridize to a target sequence that contains a SNP site such that the SNP site aligns anywhere along the sequence of the probe, the probe is preferably designed to hybridize to a segment of the target sequence such that the SNP site aligns with a central position of the probe (e.g., a position within the probe that is at least three nucleotides from either end of the probe). This design of probe generally achieves good discrimination in hybridization between different allelic forms.

[0151] In another embodiment, a probe or primer may be designed to hybridize to a segment of target DNA such that the SNP aligns with either the 5'-most end or the 3'-most end of the probe or primer. In a specific preferred embodiment which is particularly suitable for use in an oligonucleotide ligation assay (U.S. Pat. No. 4,988,617), the 3'-most nucleotide of the probe aligns with the SNP position in the target sequence.

[0152] Oligonucleotide probes and primers may be prepared by methods well known in the art. Chemical synthetic methods include, but are not limited to, the phosphotriester method described by Narang et al., *Methods in Enzymology* 68:90 [1979]; the phosphodiester method described by Brown et al., *Methods in Enzymology* 68:109 [1979], the diethylphosphoamidate method described by Beaucage et al., *Tetrahedron Letters* 22:1859 [1981]; and the solid support method described in U.S. Pat. No. 4,458,066.

[0153] Allele-specific probes are often used in pairs (or, less commonly, in sets of 3 or 4, such as if a SNP position is known to have 3 or 4 alleles, respectively, or to assay both strands of a nucleic acid molecule for a target SNP allele), and such pairs may be identical except for a one nucleotide mismatch that represents the allelic variants at the SNP position. Commonly, one member of a pair perfectly matches a reference form of a target sequence that has a more common SNP allele (i.e., the allele that is more frequent in the target population) and the other member of the pair perfectly matches a form of the target sequence that has a less common SNP allele (i.e., the allele that is rarer in the target population). In the case of an array, multiple pairs of probes can be immobilized on the same support for simultaneous analysis of multiple different polymorphisms.

[0154] In one type of PCR-based assay, an allele-specific primer hybridizes to a region on a target nucleic acid molecule that overlaps a SNP position and only primes amplification of an allelic form to which the primer exhibits perfect complementarity (Gibbs, *Nucleic Acid Res.* 17:2427-2448 [1989]). Typically, the primer's 3'-most nucleotide is aligned with and complementary to the SNP position of the target nucleic acid molecule. This primer is used in conjunction with a second primer that hybridizes at a distal site. Amplification proceeds from the two primers, producing a detectable product that indicates which allelic form is present in the test sample. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification or substantially reduces amplification efficiency, so that either no detectable product is formed or it is formed in lower amounts or at a

slower pace. The method generally works most effectively when the mismatch is at the 3'-most position of the oligonucleotide (i.e., the 3'-most position of the oligonucleotide aligns with the target SNP position) because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456). This PCR-based assay can be utilized as part of the TaqMan assay, described below.

**[0155]** In a specific embodiment of the invention, a primer of the invention contains a sequence substantially complementary to a segment of a target SNP-containing nucleic acid molecule except that the primer has a mismatched nucleotide in one of the three nucleotide positions at the 3'-most end of the primer, such that the mismatched nucleotide does not base pair with a particular allele at the SNP site. In a preferred embodiment, the mismatched nucleotide in the primer is the second from the last nucleotide at the 3'-most position of the primer. In a more preferred embodiment, the mismatched nucleotide in the primer is the last nucleotide at the 3'-most position of the primer.

**[0156]** In another embodiment of the invention, a SNP detection reagent of the invention is labeled with a fluorogenic reporter dye that emits a detectable signal. While the preferred reporter dye is a fluorescent dye, any reporter dye that can be attached to a detection reagent such as an oligonucleotide probe or primer is suitable for use in the invention. Such dyes include, but are not limited to, Acridine, AMCA, BODIPY, Cascade Blue, Cy2, Cy3, Cy5, Cy7, Dabcyl, Edans, Eosin, Erythrosin, Fluorescein, 6-Fam, Tet, Joe, Hex, Oregon Green, Rhodamine, Rhodol Green, Tamra, Rox, and Texas Red.

**[0157]** In yet another embodiment of the invention, the detection reagent may be further labeled with a quencher dye such as Tamra, especially when the reagent is used as a self-quenching probe such as a TaqMan (U.S. Pat. Nos. 5,210,015 and 5,538,848) or Molecular Beacon probe (U.S. Pat. Nos. 5,118,801 and 5,312,728), or other stemless or linear beacon probe (Livak et al., *PCR Method Appl.* 4:357-362 [1995]; Tyagi et al., *Nature Biotechnology* 14:303-308 [1996]; Nazarenko et al., *Nucl. Acids Res.* 25:2516-2521 [1997]; U.S. Pat. Nos. 5,866,336 and 6,117,635).

**[0158]** The detection reagents of the invention may also contain other labels, including but not limited to, biotin for streptavidin binding, hapten for antibody binding, and oligonucleotide for binding to another complementary oligonucleotide such as pairs of zipcodes.

**[0159]** The present invention also contemplates reagents that do not contain (or that are complementary to) a SNP nucleotide identified herein but that are used to assay one or more SNPs disclosed herein. For example, primers that flank, but do not hybridize directly to a target SNP position provided herein are useful in primer extension reactions in which the primers hybridize to a region adjacent to the target SNP position (i.e., within one or more nucleotides from the target SNP site). During the primer extension reaction, a primer is typically not able to extend past a target SNP site if a particular nucleotide (allele) is present at that target SNP site, and the primer extension product can be detected in order to determine which SNP allele is present at the target SNP site. For example, particular ddNTPs are typically used in the primer extension reaction to terminate primer extension once a ddNTP is incorporated into the extension product (a primer extension product which includes a ddNTP at

the 3'-most end of the primer extension product, and in which the ddNTP is a nucleotide of a SNP disclosed herein, is a composition that is specifically contemplated by the present invention). Thus, reagents that bind to a nucleic acid molecule in a region adjacent to a SNP site and that are used for assaying the SNP site, even though the bound sequences do not necessarily include the SNP site itself, are also contemplated by the present invention.

**[0160]** SNP Detection Kits and Systems

**[0161]** A person skilled in the art will recognize that, based on the SNP and associated sequence information disclosed herein, detection reagents can be developed and used to assay any SNP of the present invention individually or in combination, and such detection reagents can be readily incorporated into one of the established kit or system formats which are well known in the art. The terms "kits" and "systems", as used herein in the context of SNP detection reagents, are intended to refer to such things as combinations of multiple SNP detection reagents, or one or more SNP detection reagents in combination with one or more other types of elements or components (e.g., other types of biochemical reagents, containers, packages such as packaging intended for commercial sale, substrates to which SNP detection reagents are attached, electronic hardware components, etc.). Accordingly, the present invention further provides SNP detection kits and systems, including but not limited to, packaged probe and primer sets (e.g., TaqMan probe/primer sets), arrays/microarrays of nucleic acid molecules, and beads that contain one or more probes, primers, or other detection reagents for detecting one or more SNPs of the present invention. The kits/systems can optionally include various electronic hardware components; for example, arrays ("DNA chips") and microfluidic systems ("lab-on-a-chip" systems) provided by various manufacturers typically comprise hardware components. Other kits/systems (e.g., probe/primer sets) may not include electronic hardware components, but may be comprised of, for example, one or more SNP detection reagents (along with, optionally, other biochemical reagents) packaged in one or more containers.

**[0162]** In some embodiments, a SNP detection kit typically contains one or more detection reagents and other components (e.g., a buffer, enzymes such as DNA polymerases or ligases, chain extension nucleotides such as deoxynucleotide triphosphates, and in the case of Sanger-type DNA sequencing reactions, chain terminating nucleotides, positive control sequences, negative control sequences, and the like) necessary to carry out an assay or reaction, such as amplification and/or detection of a SNP-containing nucleic acid molecule. A kit may further contain means for determining the amount of a target nucleic acid, and means for comparing the amount with a standard, and can comprise instructions for using the kit to detect the SNP-containing nucleic acid molecule of interest. In one embodiment of the present invention, kits are provided which contain the necessary reagents to carry out one or more assays to detect one or more SNPs disclosed herein. In a preferred embodiment of the present invention, SNP detection kits/systems are in the form of nucleic acid arrays, or compartmentalized kits, including microfluidic/lab-on-a-chip systems.

**[0163]** SNP detection kits/systems may contain, for example, one or more probes, or pairs of probes, that

hybridize to a nucleic acid molecule at or near each target SNP position. Multiple pairs of allele-specific probes may be included in the kit/system to simultaneously assay large numbers of SNPs, at least one of which is a SNP of the present invention. In some kits/systems, the allele-specific probes are immobilized to a substrate such as an array or bead. For example, the same substrate can comprise allele-specific probes for detecting at least 1; 10; 100; 1000; 10,000; 100,000 (or any other number in-between) or substantially all of the SNPs shown in Table 1 and/or Table 2.

[0164] The terms "arrays," "microarrays," and "DNA chips" are used herein interchangeably to refer to an array of distinct polynucleotides affixed to a substrate, such as glass, plastic, paper, nylon or other type of membrane, filter, chip, or any other suitable solid support. The polynucleotides can be synthesized directly on the substrate, or synthesized separate from the substrate and then affixed to the substrate. In one embodiment, the microarray is prepared and used according to the methods described in U.S. Pat. No. 5,837,832 (Chee et al.), PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (*Nat. Biotech.* 14:1675-1680 [1996]) and Schena, M. et al. (*Proc. Natl. Acad. Sci.* 93:10614-10619 [1996]), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et al., U.S. Pat. No. 5,807,522.

[0165] Nucleic acid arrays are reviewed in the following references: Zammattéo et al., "New chips for molecular biology and diagnostics," *Biotechnol. Annu. Rev.* 8:85-101 (2002); Sosnowski et al., "Active microelectronic array system for DNA hybridization, genotyping and pharmacogenomic applications," *Psychiatr. Genet.* 12(4): 181-92 (December 2002); Heller, "DNA microarray technology: devices, systems, and applications," *Annu. Rev. Biomed. Eng.* 4:129-53 (2002); Epub Mar. 22 2002; Kolchinsky et al., "Analysis of SNPs and other genomic variations using gel-based chips," *Hum. Mutat.* 19(4):343-60 (April 2002); and McGall et al., "High-density genechip oligonucleotide probe arrays," *Adv. Biochem. Eng. Biotechnol.* 77:21-42 (2002).

[0166] Any number of probes, such as allele-specific probes, may be implemented in an array, and each probe or pair of probes can hybridize to a different SNP position. In the case of polynucleotide probes, they can be synthesized at designated areas (or synthesized separately and then affixed to designated areas) on a substrate using a light-directed chemical process. Each DNA chip can contain, for example, thousands to millions of individual synthetic polynucleotide probes arranged in a grid-like pattern and miniaturized (e.g., to the size of a dime). Preferably, probes are attached to a solid support in an ordered, addressable array.

[0167] A microarray can be composed of a large number of unique, single-stranded polynucleotides, usually either synthetic antisense polynucleotides or fragments of cDNAs, fixed to a solid support. Typical polynucleotides are preferably about 6-60 nucleotides in length, more preferably about 15-30 nucleotides in length, and most preferably about 18-25 nucleotides in length. For certain types of microarrays or other detection kits/systems, it may be preferable to use oligonucleotides that are only about 7-20 nucleotides in length. In other types of arrays, such as arrays used in conjunction with chemiluminescent detection technology,

preferred probe lengths can be, for example, about 15-80 nucleotides in length, preferably about 50-70 nucleotides in length, more preferably about 55-65 nucleotides in length, and most preferably about 60 nucleotides in length. The microarray or detection kit can contain polynucleotides that cover the known 5' or 3' sequence of a gene/transcript or target SNP site, sequential polynucleotides that cover the full-length sequence of a gene/transcript; or unique polynucleotides selected from particular areas along the length of a target gene/transcript sequence, particularly areas corresponding to one or more SNPs disclosed in Table 1 and/or Table 2. Polynucleotides used in the microarray or detection kit can be specific to a SNP or SNPs of interest (e.g., specific to a particular SNP allele at a target SNP site, or specific to particular SNP alleles at multiple different SNP sites), or specific to a polymorphic gene/transcript or genes/transcripts of interest.

[0168] Hybridization assays based on polynucleotide arrays rely on the differences in hybridization stability of the probes to perfectly matched and mismatched target sequence variants. For SNP genotyping, it is generally preferable that stringency conditions used in hybridization assays are high enough such that nucleic acid molecules that differ from one another at as little as a single SNP position can be differentiated (e.g., typical SNP hybridization assays are designed so that hybridization will occur only if one particular nucleotide is present at a SNP position, but will not occur if an alternative nucleotide is present at that SNP position). Such high stringency conditions may be preferable when using, for example, nucleic acid arrays of allele-specific probes for SNP detection. Such high stringency conditions are described in the preceding section, and are well known to those skilled in the art and can be found in, for example, *Current Protocols in Molecular Biology* 6.3.1-6.3.6, John Wiley & Sons, New York (1989).

[0169] In other embodiments, the arrays are used in conjunction with chemiluminescent detection technology. The following patents and patent applications, which are all hereby incorporated by reference, provide additional information pertaining to chemiluminescent detection: U.S. patent application Ser. Nos. 10/620,332 and 10/620,333 describe chemiluminescent approaches for microarray detection; U.S. Pat. Nos. 6,124,478, 6,107,024, 5,994,073, 5,981,768, 5,871,938, 5,843,681, 5,800,999, and 5,773,628 describe methods and compositions of dioxetane for performing chemiluminescent detection; and U.S. published application US2002/0110828 discloses methods and compositions for microarray controls.

[0170] In one embodiment of the invention, a nucleic acid array can comprise an array of probes of about 15-25 nucleotides in length. In further embodiments, a nucleic acid array can comprise any number of probes, in which at least one probe is capable of detecting one or more SNPs disclosed in Table 1 and/or Table 2, and/or at least one probe comprises a fragment of one of the sequences selected from the group consisting of those disclosed in Table 1, Table 2, the Sequence Listing, and sequences complementary thereto, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 12, 15, 16, 18, 20, more preferably 22, 25, 30, 40, 47, 50, 55, 60, 65, 70, 80, 90, 100, or more consecutive nucleotides (or any other number in-between) and containing (or being complementary to) a novel SNP allele disclosed in Table 1 and/or Table 2. In

some embodiments, the nucleotide complementary to the SNP site is within 5, 4, 3, 2, or 1 nucleotide from the center of the probe, more preferably at the center of said probe.

[0171] A polynucleotide probe can be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more polynucleotides, or any other number which lends itself to the efficient use of commercially available instrumentation.

[0172] Using such arrays or other kits/systems, the present invention provides methods of identifying the SNPs disclosed herein in a test sample. Such methods typically involve incubating a test sample of nucleic acids with an array comprising one or more probes corresponding to at least one SNP position of the present invention, and assaying for binding of a nucleic acid from the test sample with one or more of the probes. Conditions for incubating a SNP detection reagent (or a kit/system that employs one or more such SNP detection reagents) with a test sample vary. Incubation conditions depend on such factors as the format employed in the assay, the detection methods employed, and the type and nature of the detection reagents used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification and array assay formats can readily be adapted to detect the SNPs disclosed herein.

[0173] A SNP detection kit/system of the present invention may include components that are used to prepare nucleic acids from a test sample for the subsequent amplification and/or detection of a SNP-containing nucleic acid molecule. Such sample preparation components can be used to produce nucleic acid extracts (including DNA and/or RNA), proteins or membrane extracts from any bodily fluids (such as blood, serum, plasma, urine, saliva, phlegm, gastric juices, semen, tears, sweat, etc.), skin, hair, cells (especially nucleated cells), biopsies, buccal swabs or tissue specimens. The test samples used in the above-described methods will vary based on such factors as the assay format, nature of the detection method, and the specific tissues, cells or extracts used as the test sample to be assayed. Methods of preparing nucleic acids, proteins, and cell extracts are well known in the art and can be readily adapted to obtain a sample that is compatible with the system utilized. Automated sample preparation systems for extracting nucleic acids from a test sample are commercially available, and examples are Qiagen's BioRobot 9600, Applied Biosystems' PRISM 6700, and Roche Molecular Systems' COBAS AmpliPrep System.

[0174] Another form of kit contemplated by the present invention is a compartmentalized kit. A compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include, for example,

small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allow one to efficiently transfer reagents from one compartment to another compartment such that the test samples and reagents are not cross-contaminated, or from one container to another vessel not included in the kit, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another or to another vessel. Such containers may include, for example, one or more containers which will accept the test sample, one or more containers which contain at least one probe or other SNP detection reagent for detecting one or more SNPs of the present invention, one or more containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and one or more containers which contain the reagents used to reveal the presence of the bound probe or other SNP detection reagents. The kit can optionally further comprise compartments and/or reagents for, for example, nucleic acid amplification or other enzymatic reactions such as primer extension reactions, hybridization, ligation, electrophoresis (preferably capillary electrophoresis), mass spectrometry, and/or laser-induced fluorescence detection. The kit may also include instructions for using the kit. Exemplary compartmentalized kits include microfluidic devices known in the art (see, e.g., Weigl et al., "Lab-on-a-chip for drug development," *Adv. Drug Deliv. Rev.* 24, 55[3]:349-77 [February 2003]). In such microfluidic devices, the containers may be referred to as, for example, microfluidic "compartments," "chambers," or "channels."

[0175] Microfluidic devices, which may also be referred to as "lab-on-a-chip" systems, biomedical micro-electro-mechanical systems (bioMEMS), or multicomponent integrated systems, are exemplary kits/systems of the present invention for analyzing SNPs. Such systems miniaturize and compartmentalize processes such as probe/target hybridization, nucleic acid amplification, and capillary electrophoresis reactions in a single functional device. Such microfluidic devices typically utilize detection reagents in at least one aspect of the system, and such detection reagents may be used to detect one or more SNPs of the present invention. One example of a microfluidic system is disclosed in U.S. Pat. No. 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips. Exemplary microfluidic systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples may be controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage can be used as a means to control the liquid flow at intersections between the micro-machined channels and to change the liquid flow rate for pumping across different sections of the microchip. See, for example, U.S. Pat. Nos. 6,153,073, Dubrow et al., and 6,156,181, Parce et al.

[0176] For genotyping SNPs, an exemplary microfluidic system may integrate, for example, nucleic acid amplification, primer extension, capillary electrophoresis, and a detection method such as laser induced fluorescence detection. In a first step of an exemplary process for using such an exemplary system, nucleic acid samples are amplified, preferably by PCR. Then, the amplification products are subjected to automated primer extension reactions using ddNTPs (specific fluorescence for each ddNTP) and the

appropriate oligonucleotide primers to carry out primer extension reactions which hybridize just upstream of the targeted SNP. Once the extension at the 3' end is completed, the primers are separated from the unincorporated fluorescent ddNTPs by capillary electrophoresis. The separation medium used in capillary electrophoresis can be, for example, polyacrylamide, polyethyleneglycol or dextran. The incorporated ddNTPs in the single nucleotide primer extension products are identified by laser-induced fluorescence detection. Such an exemplary microchip can be used to process, for example, at least 96 to 384 samples, or more, in parallel.

**[0177]** Uses of Nucleic Acid Molecules

**[0178]** The nucleic acid molecules of the present invention have a variety of uses, especially in the diagnosis and treatment of Alzheimer's Disease. For example, the nucleic acid molecules are useful as hybridization probes, such as for genotyping SNPs in messenger RNA, transcript, cDNA, genomic DNA, amplified DNA or other nucleic acid molecules, and for isolating full-length cDNA and genomic clones encoding the variant peptides disclosed in Table 1 as well as their orthologs.

**[0179]** A probe can hybridize to any nucleotide sequence along the entire length of a nucleic acid molecule provided in Table 1 and/or Table 2. Preferably, a probe of the present invention hybridizes to a region of a target sequence that encompasses a SNP position indicated in Table 1 and/or Table 2. More preferably, a probe hybridizes to a SNP-containing target sequence in a sequence-specific manner such that it distinguishes the target sequence from other nucleotide sequences which vary from the target sequence only by which nucleotide is present at the SNP site. Such a probe is particularly useful for detecting the presence of a SNP-containing nucleic acid in a test sample, or for determining which nucleotide (allele) is present at a particular SNP site (i.e., genotyping the SNP site).

**[0180]** A nucleic acid hybridization probe may be used for determining the presence, level, form, and/or distribution of nucleic acid expression. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes specific for the SNPs described herein can be used to assess the presence, expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in gene expression relative to normal levels. In vitro techniques for detection of mRNA include, for example, Northern blot hybridizations and in situ hybridizations. In vitro techniques for detecting DNA include Southern blot hybridizations and in situ hybridizations (Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. [2000]).

**[0181]** Probes can be used as part of a diagnostic test kit for identifying cells or tissues in which a variant protein is expressed, such as by measuring the level of a variant protein-encoding nucleic acid (e.g., mRNA) in a sample of cells from a subject or determining if a polynucleotide contains a SNP of interest.

**[0182]** Thus, the nucleic acid molecules of the invention can be used as hybridization probes to detect the SNPs disclosed herein, thereby determining whether an individual with the polymorphisms is at risk for Alzheimer's Disease or

has developed early stage Alzheimer's Disease. Detection of a SNP associated with a disease phenotype provides a diagnostic tool for an active disease and/or genetic predisposition to the disease.

**[0183]** Furthermore, the nucleic acid molecules of the invention are therefore useful for detecting a gene (gene information is disclosed in Table 2, for example) which contains a SNP disclosed herein and/or products of such genes, such as expressed mRNA transcript molecules (transcript information is disclosed in Table 1, for example), and are thus useful for detecting gene expression. The nucleic acid molecules can optionally be implemented in, for example, an array or kit format for use in detecting gene expression.

**[0184]** The nucleic acid molecules of the invention are also useful as primers to amplify any given region of a nucleic acid molecule, particularly a region containing a SNP identified in Table 1 and/or Table 2.

**[0185]** The nucleic acid molecules of the invention are also useful for constructing recombinant vectors (described in greater detail below). Such vectors include expression vectors that express a portion of, or all of, any of the variant peptide sequences provided in Table 1. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter in situ expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced SNPs.

**[0186]** The nucleic acid molecules of the invention are also useful for expressing antigenic portions of the variant proteins, particularly antigenic portions that contain a variant amino acid sequence (e.g., an amino acid substitution) caused by a SNP disclosed in Table 1 and/or Table 2.

**[0187]** The nucleic acid molecules of the invention are also useful for constructing vectors containing a gene regulatory region of the nucleic acid molecules of the present invention.

**[0188]** The nucleic acid molecules of the invention are also useful for designing ribozymes corresponding to all, or a part, of an mRNA molecule expressed from a SNP-containing nucleic acid molecule described herein.

**[0189]** The nucleic acid molecules of the invention are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and variant peptides.

**[0190]** The nucleic acid molecules of the invention are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and variant peptides. The production of recombinant cells and transgenic animals having nucleic acid molecules which contain the SNPs disclosed in Table 1 and/or Table 2 allow, for example, effective clinical design of treatment compounds and dosage regimens.

**[0191]** The nucleic acid molecules of the invention are also useful in assays for drug screening to identify compounds that, for example, modulate nucleic acid expression.

**[0192]** The nucleic acid molecules of the invention are also useful in gene therapy in patients whose cells have aberrant gene expression. Thus, recombinant cells, which

include a patient's cells that have been engineered ex vivo and returned to the patient, can be introduced into an individual where the recombinant cells produce the desired protein to treat the individual.

**[0193]** SNP Genotyping Methods

**[0194]** The process of determining which specific nucleotide (i.e., allele) is present at each of one or more SNP positions, such as a SNP position in a nucleic acid molecule disclosed in Table 1 and/or Table 2, is referred to as SNP genotyping. The present invention provides methods of SNP genotyping, such as for use in screening for Alzheimer's Disease or related pathologies, or determining predisposition thereto, or determining responsiveness to a form of treatment, or in genome mapping or SNP association analysis, etc.

**[0195]** Nucleic acid samples can be genotyped to determine which allele(s) is/are present at any given genetic region (e.g., SNP position) of interest by methods well known in the art. The neighboring sequence can be used to design SNP detection reagents such as oligonucleotide probes, which may optionally be implemented in a kit format. Exemplary SNP genotyping methods are described in Chen et al., "Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput," *Pharmacogenomics J.* 3(2):77-96 (2003); Kwok et al., "Detection of single nucleotide polymorphisms," *Curr. Issues Mol. Biol.* 5(2):43-60 (April 2003); Shi, "Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes," *Am. J. Pharmacogenomics* 2(3):197-205 (2002); and Kwok, "Methods for genotyping single nucleotide polymorphisms," *Annu. Rev. Genomics Hum. Genet.* 2:235-58 (2001). Exemplary techniques for high-throughput SNP genotyping are described in Marnellos, "High-throughput SNP analysis for genetic association studies," *Curr. Opin. Drug Discov. Devel.* 6(3):317-21 (May 2003). Common SNP genotyping methods include, but are not limited to, TaqMan assays, molecular beacon assays, nucleic acid arrays, allele-specific primer extension, allele-specific PCR, arrayed primer extension, homogeneous primer extension assays, primer extension with detection by mass spectrometry, pyrosequencing, multiplex primer extension sorted on genetic arrays, ligation with rolling circle amplification, homogeneous ligation, OLA (U.S. Pat. No. 4,988,167), multiplex ligation reaction sorted on genetic arrays, restriction-fragment length polymorphism, single base extension-tag assays, and the Invader assay. Such methods may be used in combination with detection mechanisms such as, for example, luminescence or chemiluminescence detection, fluorescence detection, time-resolved fluorescence detection, fluorescence resonance energy transfer, fluorescence polarization, mass spectrometry, and electrical detection.

**[0196]** Various methods for detecting polymorphisms include, but are not limited to, methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers et al., *Science* 230:1242 [1985]; Cotton et al., *PNAS* 85:4397 [1988]; and Saleeba et al., *Meth. Enzymol.* 217:286-295 [1992]), comparison of the electrophoretic mobility of variant and wild type nucleic acid molecules (Orita et al., *PNAS* 86:2766 [1989]; Cotton et al., *Mutat. Res.* 285:125-144 [1993]; and Hayashi et al., *Genet. Anal. Tech. Appl.* 9:73-79 [1992]), and

assaying the movement of polymorphic or wild-type fragments in polyacrylamide gels containing a gradient of denaturant using denaturing gradient gel electrophoresis (DGGE) (Myers et al., *Nature* 313:495 [1985]). Sequence variations at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or chemical cleavage methods.

**[0197]** In a preferred embodiment, SNP genotyping is performed using the TaqMan assay, which is also known as the 5' nuclease assay (U.S. Pat. Nos. 5,210,015 and 5,538,848). The TaqMan assay detects the accumulation of a specific amplified product during PCR. The TaqMan assay utilizes an oligonucleotide probe labeled with a fluorescent reporter dye and a quencher dye. The reporter dye is excited by irradiation at an appropriate wavelength, it transfers energy to the quencher dye in the same probe via a process called fluorescence resonance energy transfer (FRET). When attached to the probe, the excited reporter dye does not emit a signal. The proximity of the quencher dye to the reporter dye in the intact probe maintains a reduced fluorescence for the reporter. The reporter dye and quencher dye may be at the 5'-most and the 3'-most ends, respectively, or vice versa. Alternatively, the reporter dye may be at the 5'- or 3'-most end while the quencher dye is attached to an internal nucleotide, or vice versa. In yet another embodiment, both the reporter and the quencher may be attached to internal nucleotides at a distance from each other such that fluorescence of the reporter is reduced.

**[0198]** During PCR, the 5' nuclease activity of DNA polymerase cleaves the probe, thereby separating the reporter dye and the quencher dye and resulting in increased fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of the reporter dye. The DNA polymerase cleaves the probe between the reporter dye and the quencher dye only if the probe hybridizes to the target SNP-containing template which is amplified during PCR, and the probe is designed to hybridize to the target SNP site only if a particular SNP allele is present.

**[0199]** Preferred TaqMan primer and probe sequences can readily be determined using the SNP and associated nucleic acid sequence information provided herein. A number of computer programs, such as Primer Express (Applied Biosystems, Foster City, Calif.), can be used to rapidly obtain optimal primer/probe sets. It will be apparent to one of skill in the art that such primers and probes for detecting the SNPs of the present invention are useful in diagnostic assays for Alzheimer's Disease and related pathologies, and can be readily incorporated into a kit format. The present invention also includes modifications of the Taqman assay well known in the art such as the use of Molecular Beacon probes (U.S. Pat. Nos. 5,118,801 and 5,312,728) and other variant formats (U.S. Pat. Nos. 5,866,336 and 6,117,635).

**[0200]** Another preferred method for genotyping the SNPs of the present invention is the use of two oligonucleotide probes in an OLA (see, e.g., U.S. Pat. No. 4,988,617). In this method, one probe hybridizes to a segment of a target nucleic acid with its 3'-most end aligned with the SNP site. A second probe hybridizes to an adjacent segment of the target nucleic acid molecule directly 3' to the first probe. The two juxtaposed probes hybridize to the target nucleic acid molecule, and are ligated in the presence of a linking agent

such as a ligase if there is perfect complementarity between the 3' most nucleotide of the first probe with the SNP site. If there is a mismatch, ligation would not occur. After the reaction, the ligated probes are separated from the target nucleic acid molecule, and detected as indicators of the presence of a SNP.

[0201] The following patents, patent applications, and published international patent applications, which are all hereby incorporated by reference, provide additional information pertaining to techniques for carrying out various types of OLA: U.S. Pat. Nos. 6,027,889, 6,268,148, 5,494,810, 5,830,711, and 6,054,564 describe OLA strategies for performing SNP detection; WO 97/31256 and WO 00/56927 describe OLA strategies for performing SNP detection using universal arrays, wherein a zipcode sequence can be introduced into one of the hybridization probes, and the resulting product, or amplified product, hybridized to a universal zipcode array; U.S. application Ser. No. 01/17329 (and Ser. No. 09/584,905) describes OLA (or LDR) followed by PCR, wherein zipcodes are incorporated into OLA probes, and amplified PCR products are determined by electrophoretic or universal zipcode array readout; U.S. application 60/427,818, 60/445,636, and 60/445,494 describe SNPlex methods and software for multiplexed SNP detection using OLA followed by PCR, wherein zipcodes are incorporated into OLA probes, and amplified PCR products are hybridized with a zipchute reagent, and the identity of the SNP determined from electrophoretic readout of the zipchute. In some embodiments, OLA is carried out prior to PCR (or another method of nucleic acid amplification). In other embodiments, PCR (or another method of nucleic acid amplification) is carried out prior to OLA.

[0202] Another method for SNP genotyping is based on mass spectrometry. Mass spectrometry takes advantage of the unique mass of each of the four nucleotides of DNA. SNPs can be unambiguously genotyped by mass spectrometry by measuring the differences in the mass of nucleic acids having alternative SNP alleles. MALDI-TOF (Matrix Assisted Laser Desorption Ionization—Time of Flight) mass spectrometry technology is preferred for extremely precise determinations of molecular mass, such as SNPs. Numerous approaches to SNP analysis have been developed based on mass spectrometry. Preferred mass spectrometry-based methods of SNP genotyping include primer extension assays, which can also be utilized in combination with other approaches, such as traditional gel-based formats and microarrays.

[0203] Typically, the primer extension assay involves designing and annealing a primer to a template PCR amplicon upstream (5') from a target SNP position. A mix of dideoxynucleotide triphosphates (ddNTPs) and/or deoxynucleotide triphosphates (dNTPs) are added to a reaction mixture containing template (e.g., a SNP-containing nucleic acid molecule which has typically been amplified, such as by PCR), primer, and DNA polymerase. Extension of the primer terminates at the first position in the template where a nucleotide complementary to one of the ddNTPs in the mix occurs. The primer can be either immediately adjacent (i.e., the nucleotide at the 3' end of the primer hybridizes to the nucleotide next to the target SNP site) or two or more nucleotides removed from the SNP position. If the primer is several nucleotides removed from the target SNP position, the only limitation is that the template sequence between the

3' end of the primer and the SNP position cannot contain a nucleotide of the same type as the one to be detected, or this will cause premature termination of the extension primer. Alternatively, if all four ddNTPs alone, with no dNTPs, are added to the reaction mixture, the primer will always be extended by only one nucleotide, corresponding to the target SNP position. In this instance, primers are designed to bind one nucleotide upstream from the SNP position (i.e., the nucleotide at the 3' end of the primer hybridizes to the nucleotide that is immediately adjacent to the target SNP site on the 5' side of the target SNP site). Extension by only one nucleotide is preferable, as it minimizes the overall mass of the extended primer, thereby increasing the resolution of mass differences between alternative SNP nucleotides. Furthermore, mass-tagged ddNTPs can be employed in the primer extension reactions in place of unmodified ddNTPs. This increases the mass difference between primers extended with these ddNTPs, thereby providing increased sensitivity and accuracy, and is particularly useful for typing heterozygous base positions. Mass-tagging also alleviates the need for intensive sample-preparation procedures and decreases the necessary resolving power of the mass spectrometer.

[0204] The extended primers can then be purified and analyzed by MALDI-TOF mass spectrometry to determine the identity of the nucleotide present at the target SNP position. In one method of analysis, the products from the primer extension reaction are combined with light absorbing crystals that form a matrix. The matrix is then hit with an energy source such as a laser to ionize and desorb the nucleic acid molecules into the gas-phase. The ionized molecules are then ejected into a flight tube and accelerated down the tube towards a detector. The time between the ionization event, such as a laser pulse, and collision of the molecule with the detector is the time of flight of that molecule. The time of flight is precisely correlated with the mass-to-charge ratio ( $m/z$ ) of the ionized molecule. Ions with smaller  $m/z$  travel down the tube faster than ions with larger  $m/z$  and therefore the lighter ions reach the detector before the heavier ions. The time-of-flight is then converted into a corresponding, and highly precise,  $m/z$ . In this manner, SNPs can be identified based on the slight differences in mass, and the corresponding time of flight differences, inherent in nucleic acid molecules having different nucleotides at a single base position. For further information regarding the use of primer extension assays in conjunction with MALDI-TOF mass spectrometry for SNP genotyping, see, e.g., Wise et al., "A standard protocol for single nucleotide primer extension in the human genome using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry," *Rapid Commun. Mass Spectrom.* 17(11):1195-202 (2003).

[0205] The following references provide further information describing mass spectrometry-based methods for SNP genotyping: Bocker, "SNP and mutation discovery using base-specific cleavage and MALDI-TOF mass spectrometry," *Bioinformatics* 19 Suppl 1:144-153 (July 2003); Storm et al., "MALDI-TOF mass spectrometry-based SNP genotyping," *Methods Mol. Biol.* 212:241-62 (2003); Jurinke et al., "The use of MassARRAY technology for high throughput genotyping," *Adv. Biochem. Eng. Biotechnol.* 77:57-74 (2002); and Jurinke et al., "Automated genotyping using the DNA MassArray technology," *Methods Mol. Biol.* 187:179-92 (2002).

[0206] SNPs can also be scored by direct DNA sequencing. A variety of automated sequencing procedures can be utilized (Biotechniques 19:448 [1995]), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO94/16101; Cohen et al., *Adv. Chromatogr.* 36:127-162 [1996]; and Griffin et al., *Appl. Biochem. Biotechnol.* 38:147-159 [1993]). The nucleic acid sequences of the present invention enable one of ordinary skill in the art to readily design sequencing primers for such automated sequencing procedures. Commercial instrumentation, such as the Applied Biosystems 377, 3100, 3700, 3730, and 3730x1 DNA Analyzers (Foster City, Calif.), is commonly used in the art for automated sequencing.

[0207] Other methods that can be used to genotype the SNPs of the present invention include single-strand conformational polymorphism (SSCP), and denaturing gradient gel electrophoresis (DGGE) (Myers et al., *Nature* 313:495 [1985]). SSCP identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita et al., *Proc. Nat. Acad.* Single-stranded PCR products can be generated by heating or otherwise denaturing double stranded PCR products. Single-stranded nucleic acids may refold or form secondary structures that are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products are related to base-sequence differences at SNP positions. DGGE differentiates SNP alleles based on the different sequence-dependent stabilities and melting properties inherent in polymorphic DNA and the corresponding differences in electrophoretic migration patterns in a denaturing gradient gel ("PCR Technology," *Principles and Applications for DNA Amplification Chapter 7*, ed. Erlich, W.H. Freeman and Co., New York, [1992]).

[0208] Sequence-specific ribozymes (U.S. Pat. No. 5,498, 531) can also be used to score SNPs based on the development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature. If the SNP affects a restriction enzyme cleavage site, the SNP can be identified by alterations in restriction enzyme digestion patterns, and the corresponding changes in nucleic acid fragment lengths determined by gel electrophoresis.

[0209] SNP genotyping can include the steps of, for example, collecting a biological sample from a human subject (e.g., sample of tissues, cells, fluids, secretions, etc.), isolating nucleic acids (e.g., genomic DNA, mRNA or both) from the cells of the sample, contacting the nucleic acids with one or more primers which specifically hybridize to a region of the isolated nucleic acid containing a target SNP under conditions such that hybridization and amplification of the target nucleic acid region occurs, and determining the nucleotide present at the SNP position of interest, or, in some assays, detecting the presence or absence of an amplification product (assays can be designed so that hybridization and/or amplification will only occur if a particular SNP allele is present or absent). In some assays, the size of the amplification product is detected and compared to the length of a control sample; for example, deletions and insertions can be detected by a change in size of the amplified product compared to a normal genotype.

[0210] SNP genotyping is useful for numerous practical applications, as described below. Examples of such appli-

cations include, but are not limited to, SNP-disease association analysis, disease predisposition screening, disease diagnosis, disease prognosis, disease progression monitoring, determining therapeutic strategies based on an individual's genotype ("pharmacogenomics"), developing therapeutic agents based on SNP genotypes associated with a disease or likelihood of responding to a drug, stratifying a patient population for clinical trial for a treatment regimen, predicting the likelihood that an individual will experience toxic side effects from a therapeutic agent, and human identification applications such as forensics.

[0211] Analysis of Genetic Association Between SNPs and Phenotypic Traits

[0212] SNP genotyping for disease diagnosis, disease predisposition screening, disease prognosis, determining drug responsiveness (pharmacogenomics), drug toxicity screening, and other uses described herein, typically relies on initially establishing a genetic association between one or more specific SNPs and the particular phenotypic traits of interest.

[0213] Different study designs may be used for genetic association studies (Modern Epidemiology 609-622, Lippincott Williams & Wilkins [1998]). Observational studies are most frequently carried out in which the response of the patients is not interfered with. The first type of observational study identifies a sample of persons in whom the suspected cause of the disease is present and another sample of persons in whom the suspected cause is absent, and then the frequency of development of disease in the two samples is compared. These sampled populations are called cohorts, and the study is a prospective study. The other type of observational study is case-control or a retrospective study. In typical case-control studies, samples are collected from individuals with the phenotype of interest (cases) such as certain manifestations of a disease, and from individuals without the phenotype (controls) in a population (target population) that conclusions are to be drawn from. Then the possible causes of the disease are investigated retrospectively. As the time and costs of collecting samples in case-control studies are considerably less than those for prospective studies, case-control studies are the more commonly used study design in genetic association studies, at least during the exploration and discovery stage.

[0214] In both types of observational studies, there may be potential confounding factors that should be taken into consideration. Confounding factors are those that are associated with both the real cause(s) of the disease and the disease itself, and they include demographic information such as age, gender, ethnicity as well as environmental factors. When confounding factors are not matched in cases and controls in a study, and are not controlled properly, spurious association results can arise. If potential confounding factors are identified, they should be controlled for by analysis methods explained below.

[0215] In a genetic association study, the cause of interest to be tested is a certain allele or a SNP or a combination of alleles or a haplotype from several SNPs. Thus, tissue specimens (e.g., whole blood) from the sampled individuals may be collected and genomic DNA genotyped for the SNP(s) of interest. In addition to the phenotypic trait of interest, other information such as demographic (e.g., age, gender, ethnicity, etc.), clinical, and environmental informa-

tion that may influence the outcome of the trait can be collected to further characterize and define the sample set. In many cases, these factors are known to be associated with diseases and/or SNP allele frequencies. There are likely gene-environment and/or gene-gene interactions as well. Analysis methods to address gene-environment and gene-gene interactions (for example, the effects of the presence of both susceptibility alleles at two different genes can be greater than the effects of the individual alleles at two genes combined) are discussed below.

[0216] After all the relevant phenotypic and genotypic information has been obtained, statistical analyses are carried out to determine if there is any significant correlation between the presence of an allele or a genotype with the phenotypic characteristics of an individual. Preferably, data inspection and cleaning are first performed before carrying out statistical tests for genetic association. Epidemiological and clinical data of the samples can be summarized by descriptive statistics with tables and graphs. Data validation is preferably performed to check for data completion, inconsistent entries, and outliers. Chi-squared tests and t-tests (Wilcoxon rank-sum tests if distributions are not normal) may then be used to check for significant differences between cases and controls for discrete and continuous variables, respectively. To ensure genotyping quality, Hardy-Weinberg disequilibrium tests can be performed on cases and controls separately. Significant deviation from Hardy-Weinberg equilibrium (HWE) in both cases and controls for individual markers can be indicative of genotyping errors. If HWE is violated in a majority of markers, it is indicative of population substructure that should be further investigated. Moreover, Hardy-Weinberg disequilibrium in cases only can indicate genetic association of the markers with the disease (*Genetic Data Analysis*, Weir B., Sinauer [1990]).

[0217] To test whether an allele of a single SNP is associated with the case or control status of a phenotypic trait, one skilled in the art can compare allele frequencies in cases and controls. Standard chi-squared tests and Fisher exact tests can be carried out on a 2x2 table (2 SNP allelesx2 outcomes in the categorical trait of interest). To test whether genotypes of a SNP are associated, chi-squared tests can be carried out on a 3x2 table (3 genotypesx2 outcomes). Score tests are also carried out for genotypic association to contrast the three genotypic frequencies (major homozygotes, heterozygotes and minor homozygotes) in cases and controls, and to look for trends using 3 different modes of inheritance, namely dominant (with contrast coefficients 2, -1, -1), additive (with contrast coefficients 1, 0, -1) and recessive (with contrast coefficients 1, 1, -2). Odds ratios for minor versus major alleles, and odds ratios for heterozygote and homozygote variants versus the wild type genotypes are calculated with the desired confidence limits, usually 95%.

[0218] In order to control for confounders and to test for interaction and effect modifiers, stratified analyses may be performed using stratified factors that are likely to be confounding, including demographic information such as age, ethnicity, and gender, or an interacting element or effect modifier, such as a known major gene (e.g., APOE for Alzheimer's Disease or HLA genes for autoimmune diseases), or environmental factors such as smoking in lung cancer. Stratified association tests may be carried out using Cochran-Mantel-Haenszel tests that take into account the ordinal nature of genotypes with 0, 1, and 2 variant alleles.

Exact tests by StatXact may also be performed when computationally possible. Another way to adjust for confounding effects and test for interactions is to perform stepwise multiple logistic regression analysis using statistical packages such as SAS or R. Logistic regression is a model-building technique in which the best fitting and most parsimonious model is built to describe the relation between the dichotomous outcome (for instance, getting a certain disease or not) and a set of independent variables (for instance, genotypes of different associated genes, and the associated demographic and environmental factors). The most common model is one in which the logit transformation of the odds ratios is expressed as a linear combination of the variables (main effects) and their cross-product terms (interactions) (Applied Logistic Regression, Hosmer and Lemeshow, Wiley [2000]). To test whether a certain variable or interaction is significantly associated with the outcome, coefficients in the model are first estimated and then tested for statistical significance of their departure from zero.

[0219] In addition to performing association tests one marker at a time, haplotype association analysis may also be performed to study a number of markers that are closely linked together. Haplotype association tests can have better power than genotypic or allelic association tests when the tested markers are not the disease-causing mutations themselves but are in linkage disequilibrium with such mutations. The test will even be more powerful if the disease is indeed caused by a combination of alleles on a haplotype (e.g., APOE is a haplotype formed by 2 SNPs that are very close to each other). In order to perform haplotype association effectively, marker-marker linkage disequilibrium measures, both  $D'$  and  $R^2$ , are typically calculated for the markers within a gene to elucidate the haplotype structure. Recent studies (Daly et al., *Nature Genetics* 29, 232-235 [2001]) in linkage disequilibrium indicate that SNPs within a gene are organized in block pattern, and a high degree of linkage disequilibrium exists within blocks and very little linkage disequilibrium exists between blocks. Haplotype association with the disease status can be performed using such blocks once they have been elucidated.

[0220] Haplotype association tests can be carried out in a similar fashion as the allelic and genotypic association tests. Each haplotype in a gene is analogous to an allele in a multi-allelic marker. One skilled in the art can either compare the haplotype frequencies in cases and controls or test genetic association with different pairs of haplotypes. It has been proposed (Schaid et al., *Am. J. Hum. Genet.* 70, 425-434 [2002]) that score tests can be done on haplotypes using the program "haplo.score." In that method, haplotypes are first inferred by EM algorithm and score tests are carried out with a generalized linear model (GLM) framework that allows the adjustment of other factors.

[0221] An important decision in the performance of genetic association tests is the determination of the significance level at which significant association can be declared when the P value of the tests reaches that level. In an exploratory analysis where positive hits will be followed up in subsequent confirmatory testing, an unadjusted P value<0.1 (a significance level on the lenient side) may be used for generating hypotheses for significant association of a SNP with certain phenotypic characteristics of a disease. It is preferred that a P value<0.05 (a significance level traditionally used in the art) is achieved in order for a SNP to be

considered to have an association with a disease. It is more preferred that a P value < 0.01 (a significance level on the stringent side) is achieved for an association to be declared. When hits are followed up in confirmatory analyses in more samples of the same source or in different samples from different sources, adjustment for multiple testing will be performed as to avoid excess number of hits while maintaining the experiment-wise error rates at 0.05. While there are different methods to adjust for multiple testing to control for different kinds of error rates, a commonly used but rather conservative method is Bonferroni correction to control the experiment-wise or family-wise error rate (Westfall et al., *Multiple comparisons and multiple tests*, SAS Institute [1999]). Permutation tests to control for the false discovery rates, FDR, can be more powerful (Benjamini and Hochberg, *Journal of the Royal Statistical Society Series B* 57, 1289-1300 [1995], *Resampling-based Multiple Testing*, Westfall and Young, Wiley [1993]). Such methods to control for multiplicity would be preferred when the tests are dependent and controlling for false discovery rates is sufficient as opposed to controlling for the experiment-wise error rates.

[0222] In replication studies using samples from different populations after statistically significant markers have been identified in the exploratory stage, meta-analyses can then be performed by combining evidence of different studies (*Modern Epidemiology* 643-673, Lippincott Williams & Wilkins [1998]). If available, association results known in the art for the same SNPs can be included in the meta-analyses.

[0223] Since both genotyping and disease status classification can involve errors, sensitivity analyses may be performed to see how odds ratios and P values would change upon various estimates on genotyping and disease classification error rates.

[0224] It has been well known that subpopulation-based sampling bias between cases and controls can lead to spurious results in case-control association studies (Ewens and Spielman, *Am. J. Hum. Genet.* 62, 450-458 [1995]) when prevalence of the disease is associated with different subpopulation groups. Such bias can also lead to a loss of statistical power in genetic association studies. To detect population stratification, Pritchard and Rosenberg (Pritchard et al., *Am. J. Hum. Gen.* 65:220-228 [1999]) suggested typing markers that are unlinked to the disease and using results of association tests on those markers to determine whether there is any population stratification. When stratification is detected, the genomic control (GC) method as proposed by Devlin and Roeder (Devlin et al., *Biometrics* 55:997-1004 [1999]) can be used to adjust for the inflation of test statistics due to population stratification. GC method is robust to changes in population structure levels as well as being applicable to DNA pooling designs (Devlin et al., *Genet. Epidemiol.* 21:273-284 [2001]).

[0225] While Pritchard's method recommended using 15-20 unlinked microsatellite markers, it suggested using more than 30 biallelic markers to get enough power to detect population stratification. For the GC method, it has been shown (Bacanu et al., *Am. J. Hum. Genet.* 66:1933-1944 [2000]) that about 60-70 biallelic markers are sufficient to estimate the inflation factor for the test statistics due to population stratification. Hence, 70 intergenic SNPs can be

chosen in unlinked regions as indicated in a genome scan (Kehoe et al., *Hum. Mol. Genet.* 8:237-245 [1999]).

[0226] Once individual risk factors, genetic or non-genetic, have been found for the predisposition to disease, the next step is to set up a classification/prediction scheme to predict the category (for instance, disease or no-disease) that an individual will be in depending on his genotypes of associated SNPs and other non-genetic risk factors. Logistic regression for discrete trait and linear regression for continuous trait are standard techniques for such tasks (*Applied Regression Analysis*, Draper and Smith, Wiley [1998]). Moreover, other techniques can also be used for setting up classification. Such techniques include, but are not limited to, MART, CART, neural network, and discriminant analyses that are suitable for use in comparing the performance of different methods (*The Elements of Statistical Learning*, Hastie, Tibshirani & Friedman, Springer [2002]).

[0227] Disease Diagnosis and Predisposition Screening

[0228] Information on association/correlation between genotypes and disease-related phenotypes can be exploited in several ways. For example, in the case of a highly statistically significant association between one or more SNPs with predisposition to a disease for which treatment is available, detection of such a genotype pattern in an individual may justify immediate administration of treatment, or at least the institution of regular monitoring of the individual. Detection of the susceptibility alleles associated with serious disease in a couple contemplating having children may also be valuable to the couple in their reproductive decisions. In the case of a weaker but still statistically significant association between a SNP and a human disease, immediate therapeutic intervention or monitoring may not be justified after detecting the susceptibility allele or SNP. Nevertheless, the subject can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little or no cost to the individual but would confer potential benefits in reducing the risk of developing conditions for which that individual may have an increased risk by virtue of having the susceptibility allele(s).

[0229] The SNPs of the invention may contribute to Alzheimer's Disease in an individual in different ways. Some polymorphisms occur within a protein coding sequence and contribute to disease phenotype by affecting protein structure. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on, for example, replication, transcription, and/or translation. A single SNP may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by multiple SNPs in different genes.

[0230] As used herein, the terms "diagnose," "diagnosis," and "diagnostics" include, but are not limited to any of the following: detection of Alzheimer's Disease that an individual may presently have, predisposition/susceptibility screening (i.e., determining the increased risk of an individual in developing Alzheimer's Disease in the future, or determining whether an individual has a decreased risk of developing Alzheimer's Disease in the future), determining a particular type or subclass of Alzheimer's Disease in an individual known to have Alzheimer's Disease, confirming or reinforcing a previously made diagnosis of Alzheimer's Disease, pharmacogenomic evaluation of an individual to determine which therapeutic strategy that individual is most

likely to positively respond to or to predict whether a patient is likely to respond to a particular treatment, predicting whether a patient is likely to experience toxic effects from a particular treatment or therapeutic compound, and evaluating the future prognosis of an individual having Alzheimer's Disease. Such diagnostic uses are based on the SNPs individually or in a unique combination or SNP haplotypes of the present invention.

[0231] Haplotypes are particularly useful in that, for example, fewer SNPs can be genotyped to determine if a particular genomic region harbors a locus that influences a particular phenotype, such as in linkage disequilibrium-based SNP association analysis.

[0232] Linkage disequilibrium (LD) refers to the co-inheritance of alleles (e.g., alternative nucleotides) at two or more different SNP sites at frequencies greater than would be expected from the separate frequencies of random occurrence of each allele in a given population. The expected frequency of co-occurrence of two alleles that are inherited independently is the frequency of the first allele multiplied by the frequency of the second allele. Alleles that co-occur at expected frequencies are said to be in "linkage equilibrium" In contrast, LD refers to any non-random genetic association between allele(s) at two or more different SNP sites, which is generally due to the physical proximity of the two loci along a chromosome. LD can occur when two or more SNPs sites are in close physical proximity to each other on a given chromosome and therefore alleles at these SNP sites will tend to remain unseparated for multiple generations, with the consequence that a particular nucleotide (allele) at one SNP site will show a non-random association with a particular nucleotide (allele) at another SNP site located nearby. Hence, genotyping one of the SNP sites will give almost the same information as genotyping the other SNP site that is in LD. The physical area of the chromosome that contains SNPs in LD with each other is referred to as an LD block.

[0233] Various degrees of LD can be encountered between two or more SNPs with the result being that some SNPs are more closely associated (i.e., in stronger LD) than others. Furthermore, the physical distance over which LD extends along a chromosome differs between different regions of the genome, and therefore the degree of physical separation between two or more SNP sites necessary for LD to occur can differ between different regions of the genome.

[0234] For diagnostic purposes and similar uses, if a particular SNP site is found to be useful for diagnosing Alzheimer's Disease (e.g., has a significant statistical association with the condition and/or is recognized as a causative polymorphism for the condition), then the skilled artisan would recognize that other SNP sites which are in LD with this SNP site would also be useful for diagnosing the condition. Thus, polymorphisms (e.g., SNPs and/or haplotypes) that are not the actual disease-causing (causative) polymorphisms, but are in LD with such causative polymorphisms, are also useful. In such instances, the genotype of the polymorphism(s) that is/are in LD with the causative polymorphism is predictive of the genotype of the causative polymorphism and, consequently, predictive of the phenotype (e.g., Alzheimer's Disease) that is influenced by the causative SNP(s). Therefore, polymorphic markers that are in LD with causative polymorphisms are useful as diagnos-

tic markers, and are particularly useful when the actual causative polymorphism(s) is/are unknown.

[0235] Examples of polymorphisms that can be in LD with one or more causative polymorphisms (and/or in LD with one or more polymorphisms that have a significant statistical association with a condition) and therefore useful for diagnosing the same condition that the causative/associated SNP(s) is used to diagnose, include, for example, other SNPs in the same gene, protein-coding, or mRNA transcript-coding region as the causative/associated SNP, other SNPs in the same exon or same intron as the causative/associated SNP, other SNPs in the same haplotype block as the causative/associated SNP, other SNPs in the same intergenic region as the causative/associated SNP, SNPs that are outside but near a gene (e.g., within 6 kb on either side, 5' or 3', of a gene boundary) that harbors a causative/associated SNP, etc. Such useful LD SNPs can be selected from among the SNPs disclosed in Tables 1-2, for example.

[0236] Linkage disequilibrium in the human genome is reviewed in: Wall et al., "Haplotype blocks and linkage disequilibrium in the human genome", *Nat Rev Genet.* 2003 August; 4(8):587-97; Garner et al., "On selecting markers for association studies: patterns of linkage disequilibrium between two and three diallelic loci", *Genet Epidemiol.* 2003 January; 24(1):57-67; Ardlie et al., "Patterns of linkage disequilibrium in the human genome", *Nat Rev Genet.* 2002 April; 3(4):299-309 (erratum in *Nat Rev Genet* 2002 July; 3(7):566); and Remm et al., "High-density genotyping and linkage disequilibrium in the human genome using chromosome 22 as a model"; *Curr Opin Chem. Biol.* 2002 February; 6(1):24-30; Haldane JBS (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors. *J Genet* 8:299-309; Mendel, G. (1866) Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereines in Brünn [Proceedings of the Natural History Society of Brünn]; Lewin B (1990) *Genes* IV Oxford University Press, New York, USA; Hartl DL and Clark AG (1989) *Principles of Population Genetics* 2<sup>nd</sup> ed. Sinauer Associates, Inc. Sunderland, Mass., USA; Gillespie JH (2004) *Population Genetics: A Concise Guide.* 2<sup>nd</sup> ed. Johns Hopkins University Press. USA; Lewontin R C (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49:49-67; Hoel PG (1954) *Introduction to Mathematical Statistics* 2<sup>nd</sup> ed. John Wiley & Sons, Inc. New York, USA; Hudson R R (2001) Two-locus sampling distributions and their application. *Genetics* 159:1805-1817; Dempster A P, Laird N M, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc* 39:1-38; Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 12(5):921-927; Tregouet D A, Escolano S, Tiret L, Mallet A, Golmard J L (2004) A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm. *Ann Hum Genet* 68(Pt 2):165-177; Long AD and Langley C H (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research* 9:720-731; Agresti A (1990) *Categorical Data Analysis.* John Wiley & Sons, Inc. New York, USA; Lange K (1997) *Mathematical and Statistical Methods for Genetic Analysis.* Springer-Verlag New York, Inc. New York, USA; The International HapMap Consortium (2003)

[0237] The International HapMap Project. *Nature* 426:789-796; The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299-1320; Thorisson G A, Smith A V, Krishnan L, Stein LD (2005), The International HapMap Project Web Site. *Genome Research* 15:1591-1593; McVean G, Spencer C C A, Chaix R (2005) Perspectives on human genetic variation from the HapMap project. *PLoS Genetics* 1(4):413-418; Hirschhorn J N, Daly M J (2005) Genome-wide association studies for common diseases and complex traits. *Nat Genet* 6:95-108; Schrodin S J (2005) A probabilistic approach to large-scale association scans: a semi-Bayesian method to detect disease-predisposing alleles. *SAGMB* 4(1):31; Wang W Y S, Barratt B J, Clayton D G, Todd J A (2005) Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 6:109-118. Pritchard J K, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 69:1-14.

[0238] As discussed above, one aspect of the present invention is the discovery that SNPs which are in certain LD distance with the interrogated SNP can also be used as valid markers for identifying an increased or decreased risks of having or developing VT. As used herein, the term "interrogated SNP" refers to SNPs that have been found to be associated with an increased or decreased risk of disease using genotyping results and analysis, or other appropriate experimental method as exemplified in the working examples described in this application. As used herein, the term "LD SNP" refers to a SNP that has been characterized as a SNP associating with an increased or decreased risk of diseases due to their being in LD with the "interrogated SNP" under the methods of calculation described in the application. Below, applicants describe the methods of calculation with which one of ordinary skilled in the art may determine if a particular SNP is in LD with an interrogated SNP. The parameter  $r^2$  is commonly used in the genetics art to characterize the extent of linkage disequilibrium between markers (Hudson, 2001). As used herein, the term "in LD with" refers to a particular SNP that is measured at above the threshold of a parameter such as  $r^2$  with an interrogated SNP.

[0239] It is now common place to directly observe genetic variants in a sample of chromosomes obtained from a population. Suppose one has genotype data at two genetic markers located on the same chromosome, for the markers A and B. Further suppose that two alleles segregate at each of these two markers such that alleles  $A_1$  and  $A_2$  can be found at marker A and alleles  $B_1$  and  $B_2$  at marker B. Also assume that these two markers are on a human autosome. If one is to examine a specific individual and find that they are heterozygous at both markers, such that their two-marker genotype is  $A_1A_2B_1B_2$ , then there are two possible configurations: the individual in question could have the alleles  $A_1B_1$  on one chromosome and  $A_2B_2$  on the remaining chromosome; alternatively, the individual could have alleles  $A_1B_2$  on one chromosome and  $A_2B_1$  on the other. The arrangement of alleles on a chromosome is called a haplotype. In this illustration, the individual could have haplotypes  $A_1B_1/A_2B_2$  or  $A_1B_2/A_2B_1$  (see Hartl and Clark (1989) for a more complete description). The concept of linkage equilibrium relates the frequency of haplotypes to the allele frequencies.

[0240] Assume that a sample of individuals is selected from a larger population. Considering the two markers

described above, each having two alleles, there are four possible haplotypes:  $A_1B_1$ ,  $A_1B_2$ ,  $A_2B_1$  and  $A_2B_2$ . Denote the frequencies of these four haplotypes with the following notation.

$$P_{11} = \text{freq}(A_1B_1) \quad (1)$$

$$P_{12} = \text{freq}(A_1B_2) \quad (2)$$

$$P_{21} = \text{freq}(A_2B_1) \quad (3)$$

$$P_{22} = \text{freq}(A_2B_2) \quad (4)$$

The allele frequencies at the two markers are then the sum of different haplotype frequencies, it is straightforward to write down a similar set of equations relating single-marker allele frequencies to two-marker haplotype frequencies:

$$p_1 = \text{freq}(A_1) = P_{11} + P_{12} \quad (5)$$

$$p_2 = \text{freq}(A_2) = P_{21} + P_{22} \quad (6)$$

$$q_1 = \text{freq}(B_1) = P_{11} + P_{21} \quad (7)$$

$$q_2 = \text{freq}(B_2) = P_{12} + P_{22} \quad (8)$$

[0241] Note that the four haplotype frequencies and the allele frequencies at each marker must sum to a frequency of 1.

$$P_{11} + P_{12} + P_{21} + P_{22} = 1 \quad (9)$$

$$p_1 + p_2 = 1 \quad (10)$$

$$q_1 + q_2 = 1 \quad (11)$$

If there is no correlation between the alleles at the two markers, one would expect that the frequency of the haplotypes would be approximately the product of the composite alleles. Therefore,

$$P_{11} \approx p_1 q_1 \quad (12)$$

$$P_{12} \approx p_1 q_2 \quad (13)$$

$$P_{21} \approx p_2 q_1 \quad (14)$$

$$P_{22} \approx p_2 q_2 \quad (15)$$

These approximating equations (12)-(15) represent the concept of linkage equilibrium where there is independent assortment between the two markers—the alleles at the two markers occur together at random. These are represented as approximations because linkage equilibrium and linkage disequilibrium are concepts typically thought of as properties of a sample of chromosomes; and as such they are susceptible to stochastic fluctuations due to the sampling process. Empirically, many pairs of genetic markers will be in linkage equilibrium, but certainly not all pairs.

[0242] Having established the concept of linkage equilibrium above, applicants can now describe the concept of linkage disequilibrium (LD), which is the deviation from linkage equilibrium. Since the frequency of the  $A_1B_1$  haplotype is approximately the product of the allele frequencies for  $A_1$  and  $B_1$  under the assumption of linkage equilibrium as stated mathematically in (12), a simple measure for the amount of departure from linkage equilibrium is the difference in these two quantities, D,

$$D = P_{11} - p_1 q_1 \quad (16)$$

$D=0$  indicates perfect linkage equilibrium. Substantial departures from  $D=0$  indicates LD in the sample of chromosomes examined. Many properties of D are discussed in Lewontin (1964) including the maximum and minimum values that D can take. Mathematically, using basic algebra, it can be shown that D can also be written solely in terms of haplotypes:

$$D = P_{11}P_{22} - P_{12}P_{21} \quad (17)$$

If one transforms  $D$  by squaring it and subsequently dividing by the product of the allele frequencies of  $A_1, A_2, B_1$  and  $B_2$ , the resulting quantity, called  $r^2$ , is equivalent to the square of the Pearson's correlation coefficient commonly used in statistics (e.g. Hoel, 1954).

$$r^2 = \frac{D^2}{p_1 p_2 q_1 q_2} \quad (18)$$

[0243] As with  $D$ , values of  $r^2$  close to 0 indicate linkage equilibrium between the two markers examined in the sample set. As values of  $r^2$  increase, the two markers are said to be in linkage disequilibrium. The range of values that  $r^2$  can take are from 0 to 1.  $r^2=1$  when there is a perfect correlation between the alleles at the two markers.

[0244] In addition, the quantities discussed above are sample-specific. And as such, it is necessary to formulate notation specific to the samples studied. In the approach discussed here, three types of samples are of primary interest: (i) a sample of chromosomes from individuals affected by a disease-related phenotype (cases), (ii) a sample of chromosomes obtained from individuals not affected by the disease-related phenotype (controls), and (iii) a standard sample set used for the construction of haplotypes and calculation pairwise linkage disequilibrium. For the allele frequencies used in the development of the method described below, an additional subscript will be added to denote either the case or control sample sets.

$$p_{1,cs} = \text{freq}(A_1 \text{ in cases}) \quad (19)$$

$$p_{2,cs} = \text{freq}(A_2 \text{ in cases}) \quad (20)$$

$$q_{1,cs} = \text{freq}(B_1 \text{ in cases}) \quad (21)$$

$$q_{2,cs} = \text{freq}(B_2 \text{ in cases}) \quad (22)$$

Similarly,

$$p_{1,ct} = \text{freq}(A_1 \text{ in controls}) \quad (23)$$

$$p_{2,ct} = \text{freq}(A_2 \text{ in controls}) \quad (24)$$

$$q_{1,ct} = \text{freq}(B_1 \text{ in controls}) \quad (25)$$

$$q_{2,ct} = \text{freq}(B_2 \text{ in controls}) \quad (26)$$

[0245] As a well-accepted sample set is necessary for robust linkage disequilibrium calculations, data obtained from the International HapMap project (The International HapMap Consortium 2003, 2005; Thorisson et al, 2005; McVean et al, 2005) can be used for the calculation of pairwise  $r^2$  values. Indeed, the samples genotyped for the International HapMap Project were selected to be representative examples from various human sub-populations with sufficient numbers of chromosomes examined to draw meaningful and robust conclusions from the patterns of genetic variation observed. The International HapMap project website (hapmap.org) contains a description of the project, methods utilized and samples examined. It is useful to examine empirical data to get a sense of the patterns present in such data.

[0246] Haplotype frequencies were explicit arguments in equation (18) above. However, knowing the 2-marker haplotype frequencies requires that phase to be determined for doubly heterozygous samples. When phase is unknown in the data examined, various algorithms can be used to infer phase from the genotype data. This issue was discussed

earlier where the doubly heterozygous individual with a 2-SNP genotype of  $A_1 A_2 B_1 B_2$  could have one of two different sets of chromosomes:  $A_1 B_1 / A_2 B_2$  or  $A_1 B_2 / A_2 B_1$ . One such algorithm to estimate haplotype frequencies is the expectation-maximization (EM) algorithm first formalized by Dempster et al (1977). This algorithm is often used in genetics to infer haplotype frequencies from genotype data (e.g. Excoffier and Slatkin, 1995; Tregouet et al, 2004). It should be noted that for the two-SNP case explored here, EM algorithms have very little error provided that the allele frequencies and sample sizes are not too small. The impact on  $r^2$  values is typically negligible.

[0247] As correlated genetic markers share information, interrogation of SNP markers in LD with a disease-associated SNP marker can also have sufficient power to detect disease association (Long and Langley, 1999). The relationship between the power to directly find disease-associated alleles and the power to indirectly detect disease-association was investigated by Pritchard and Przeworski (2001). In a straight-forward derivation, it can be shown that the power to detect disease association indirectly at a marker locus in linkage disequilibrium with a disease-association locus is approximately the same as the power to detect disease-association directly at the disease-association locus if the sample size is increased by a factor of  $1/r^2$  (the reciprocal of equation 18) at the marker in comparison with the disease-association locus.

[0248] Therefore, if one calculated the power to detect disease-association indirectly with an experiment having  $N$  samples, then equivalent power to directly detect disease-association (at the actual disease-susceptibility locus) would necessitate an experiment using approximately  $r^2 N$  samples. This elementary relationship between power, sample size and linkage disequilibrium can be used to derive an  $r^2$  threshold value useful in determining whether or not genotyping markers in linkage disequilibrium with a SNP marker directly associated with disease status has enough power to indirectly detect disease-association.

[0249] To commence a derivation of the power to detect disease-associated markers through an indirect process, define the effective chromosomal sample size as

$$n = \frac{4N_{cs}N_{ct}}{N_{cs} + N_{ct}}; \quad (27)$$

where  $N_{cs}$  and  $N_{ct}$  are the numbers of diploid cases and controls, respectively. This is necessary to handle situations where the numbers of cases and controls are not equivalent. For equal case and control sample sizes,  $N_{cs}=N_{ct}=N$ , the value of the effective number of chromosomes is simply  $n=2N$ —as expected. Let power be calculated for a significance level  $\alpha$  (such that traditional  $P$ -values below  $\alpha$  will be deemed statistically significant). Define the standard Gaussian distribution function as  $\Phi(\bullet)$ . Mathematically,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{\theta^2}{2}} d\theta \quad (28)$$

Alternatively, the following error function notation (Erf) may also be used,

$$\Phi(x) = \frac{1}{2} \left[ 1 + \operatorname{Erf} \left( \frac{x}{\sqrt{2}} \right) \right] \quad (29)$$

[0250] For example,  $\Phi(1.644854)=0.95$ . The value of  $r^2$  may be derived to yield a pre-specified minimum amount of power to detect disease association through indirect interrogation. Noting that the LD SNP marker could be the one that is carrying the disease-association allele, therefore that this approach constitutes a lower-bound model where all indirect power results are expected to be at least as large as those interrogated.

[0251] Denote by  $\beta$  the error rate for not detecting truly disease-associated markers. Therefore,  $1-\beta$  is the classical definition of statistical power. Substituting the Pritchard-Pzreworski result into the sample size, the power to detect disease association at a significance level of  $\alpha$  is given by the approximation

$$1 - \beta \approx \Phi \left[ \frac{|q_{1,cs} - q_{1,ct}|}{\sqrt{\frac{q_{1,cs}(1 - q_{1,cs}) + q_{1,ct}(1 - q_{1,ct})}{r^2 n}}} - Z_{1-\frac{\alpha}{2}} \right]; \quad (30)$$

where  $Z_u$  is the inverse of the standard normal cumulative distribution evaluated at  $u$  ( $u \in (0,1)$ ).  $Z_u = \Phi^{-1}(u)$ , where  $\Phi(\Phi^{-1}(u)) = \Phi^{-1}(\Phi(u)) = u$ . For example, setting  $\alpha=0.05$ , and therefore  $1-\alpha/2=0.975$ , we obtain  $Z_{0.975}=1.95996$ . Next, setting power equal to a threshold of a minimum power of  $T$ ,

$$T = \Phi \left[ \frac{|q_{1,cs} - q_{1,ct}|}{\sqrt{\frac{q_{1,cs}(1 - q_{1,cs}) + q_{1,ct}(1 - q_{1,ct})}{r^2 n}}} - Z_{1-\frac{\alpha}{2}} \right] \quad (31)$$

and solving for  $r^2$ , the following threshold  $r^2$  is obtained:

$$r_T^2 = \frac{[q_{1,cs}(1 - q_{1,cs}) + q_{1,ct}(1 - q_{1,ct})]}{n(q_{1,cs} - q_{1,ct})^2} \left[ \Phi^{-1}(T) + Z_{1-\frac{\alpha}{2}} \right]^2 \quad (32)$$

Or,

$$r_T^2 = \left( \frac{Z_T + Z_{1-\frac{\alpha}{2}}}{n} \right)^2 \left[ \frac{q_{1,cs} - (q_{1,cs})^2 + q_{1,ct} - (q_{1,ct})^2}{(q_{1,cs} - q_{1,ct})^2} \right] \quad (33)$$

[0252] Suppose that  $r^2$  is calculated between an interrogated SNP and a number of other SNPs with varying levels of LD with the interrogated SNP. The threshold value  $r_T^2$  is the minimum value of linkage disequilibrium between the interrogated SNP and the potential LD SNPs such that the LD SNP still retains a power greater or equal to  $T$  for detecting disease-association. For example, suppose that SNP rs200 is genotyped in a case-control disease-association study and it is found to be associated with a disease

phenotype. Further suppose that the minor allele frequency in 1,000 case chromosomes was found to be 16% in contrast with a minor allele frequency of 10% in 1,000 control chromosomes. Given those measurements one could have predicted, prior to the experiment, that the power to detect disease association at a significance level of 0.05 was quite high—approximately 98% using a test of allelic association. Applying equation (32) one can calculate a minimum value of  $r^2$  to indirectly assess disease association assuming that the minor allele at SNP rs200 is truly disease-predisposing for a threshold level of power. If one sets the threshold level of power to be 80%, then  $r_T^2=0.489$  given the same significance level and chromosome numbers as above. Hence, any SNP with a pairwise  $r^2$  value with rs200 greater than 0.489 is expected to have greater than 80% power to detect the disease association. Further, this is assuming the conservative model where the LD SNP is disease-associated only through linkage disequilibrium with the interrogated SNP rs200.

[0253] The contribution or association of particular SNPs and/or SNP haplotypes with disease phenotypes, such as Alzheimer's Disease, enables the SNPs of the present invention to be used to develop superior diagnostic tests capable of identifying individuals who express a detectable trait, such as Alzheimer's Disease, as the result of a specific genotype, or individuals whose genotype places them at an increased or decreased risk of developing a detectable trait at a subsequent time as compared to individuals who do not have that genotype. As described herein, diagnostics may be based on a single SNP or a group of SNPs. Combined detection of a plurality of SNPs (for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 24, 25, 30, 32, 48, 50, 64, 96, 100, or any other number in-between, or more, of the SNPs provided in Table 1 and/or Table 2) typically increases the probability of an accurate diagnosis. For example, the presence of a single SNP known to correlate with Alzheimer's Disease might indicate a probability of 20% that an individual has or is at risk of developing Alzheimer's Disease, whereas detection of five SNPs, each of which correlates with Alzheimer's Disease, might indicate a probability of 80% that an individual has or is at risk of developing Alzheimer's Disease. To further increase the accuracy of diagnosis or predisposition screening, analysis of the SNPs of the present invention can be combined with that of other polymorphisms or other risk factors of Alzheimer's Disease, such as disease symptoms, pathological characteristics, family history, diet, environmental factors or lifestyle factors.

[0254] It will, of course, be understood by practitioners skilled in the treatment or diagnosis of Alzheimer's Disease that the present invention generally does not intend to provide an absolute identification of individuals who are at risk (or less at risk) of developing Alzheimer's Disease, and/or pathologies related to Alzheimer's Disease, but rather to indicate a certain increased (or decreased) degree or likelihood of developing the disease based on statistically significant association results. However, this information is extremely valuable as it can be used to, for example, initiate preventive treatments or to allow an individual carrying one or more significant SNPs or SNP haplotypes to foresee warning signs such as minor clinical symptoms, or to have regularly scheduled physical exams to monitor for appearance of a condition in order to identify and begin treatment of the condition at an early stage. Particularly with diseases

that are extremely debilitating or fatal if not treated on time, the knowledge of a potential predisposition, even if this predisposition is not absolute, would likely contribute in a very significant manner to treatment efficacy.

[0255] The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a SNP or a SNP pattern associated with an increased or decreased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular polymorphism/mutation, including, for example, methods which enable the analysis of individual chromosomes for haplotyping, family studies, single sperm DNA analysis, or somatic hybrids. The trait analyzed using the diagnostics of the invention may be any detectable trait that is commonly observed in pathologies and disorders related to Alzheimer's Disease.

[0256] Another aspect of the present invention relates to a method of determining whether an individual is at risk (or less at risk) of developing one or more traits or whether an individual expresses one or more traits as a consequence of possessing a particular trait-causing or trait-influencing allele. These methods generally involve obtaining a nucleic acid sample from an individual and assaying the nucleic acid sample to determine which nucleotide(s) is/are present at one or more SNP positions, wherein the assayed nucleotide(s) is/are indicative of an increased or decreased risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular trait-causing or trait-influencing allele.

[0257] In another embodiment, the SNP detection reagents of the present invention are used to determine whether an individual has one or more SNP allele(s) affecting the level (e.g., the concentration of mRNA or protein in a sample, etc.) or pattern (e.g., the kinetics of expression, rate of decomposition, stability profile, Km, Vmax, etc.) of gene expression (collectively, the "gene response" of a cell or bodily fluid). Such a determination can be accomplished by screening for mRNA or protein expression (e.g., by using nucleic acid arrays, RT-PCR, TaqMan assays, or mass spectrometry), identifying genes having altered expression in an individual, genotyping SNPs disclosed in Table 1 and/or Table 2 that could affect the expression of the genes having altered expression (e.g., SNPs that are in and/or around the gene(s) having altered expression, SNPs in regulatory/control regions, SNPs in and/or around other genes that are involved in pathways that could affect the expression of the gene(s) having altered expression, or all SNPs could be genotyped), and correlating SNP genotypes with altered gene expression. In this manner, specific SNP alleles at particular SNP sites can be identified that affect gene expression.

[0258] Pharmacogenomics and Therapeutics/Drug Development

[0259] The present invention provides methods for assessing the pharmacogenomics of a subject harboring particular SNP alleles or haplotypes to a particular therapeutic agent or pharmaceutical compound, or to a class of such compounds. Pharmacogenomics deals with the roles which clinically significant hereditary variations (e.g., SNPs) play in the response to drugs due to altered drug disposition and/or abnormal action in affected persons. See, e.g., Roses, *Nature* 405, 857-865 (2000); Gould Rothberg, *Nature Biotechnol-*

*ogy* 19, 209-211 (2001); Eichelbaum, *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996); and Linder, *Clin. Chem.* 43(2):254-266 (1997). The clinical outcomes of these variations can result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the SNP genotype of an individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. For example, SNPs in drug metabolizing enzymes can affect the activity of these enzymes, which in turn can affect both the intensity and duration of drug action, as well as drug metabolism and clearance.

[0260] The discovery of SNPs in drug metabolizing enzymes, drug transporters, proteins for pharmaceutical agents, and other drug targets has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. SNPs can be expressed in the phenotype of the extensive metabolizer and in the phenotype of the poor metabolizer. Accordingly, SNPs may lead to allelic variants of a protein in which one or more of the protein functions in one population are different from those in another population. SNPs and the encoded variant peptides thus provide targets to ascertain a genetic predisposition that can affect treatment modality. For example, in a ligand-based treatment, SNPs may give rise to amino terminal extracellular domains and/or other ligand-binding regions of a receptor that are more or less active in ligand binding, thereby affecting subsequent protein activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing particular SNP alleles or haplotypes.

[0261] As an alternative to genotyping, specific variant proteins containing variant amino acid sequences encoded by alternative SNP alleles could be identified. Thus, pharmacogenomic characterization of an individual permits the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic uses based on the individual's SNP genotype, thereby enhancing and optimizing the effectiveness of the therapy. Furthermore, the production of recombinant cells and transgenic animals containing particular SNPs/haplotypes allow effective clinical design and testing of treatment compounds and dosage regimens. For example, transgenic animals can be produced that differ only in specific SNP alleles in a gene that is orthologous to a human disease susceptibility gene.

[0262] Pharmacogenomic uses of the SNPs of the present invention provide several significant advantages for patient care, particularly in treating Alzheimer's Disease. Pharmacogenomic characterization of an individual, based on an individual's SNP genotype, can identify those individuals unlikely to respond to treatment with a particular medication and thereby allows physicians to avoid prescribing the ineffective medication to those individuals. On the other hand, SNP genotyping of an individual may enable physicians to select the appropriate medication and dosage regimen that will be most effective based on an individual's SNP genotype. This information increases a physician's confidence in prescribing medications and motivates patients to comply with their drug regimens. Furthermore, pharmacogenomics may identify patients predisposed to toxicity and adverse reactions to particular drugs or drug dosages.

Adverse drug reactions lead to more than 100,000 avoidable deaths per year in the United States alone and therefore represent a significant cause of hospitalization and death, as well as a significant economic burden on the healthcare system (Pfoest et al., *Trends in Biotechnology* [August 2000]). Thus, pharmacogenomics based on the SNPs disclosed herein has the potential to both save lives and reduce healthcare costs substantially.

[0263] Pharmacogenomics in general is discussed further in Rose et al., "Pharmacogenetic analysis of clinically relevant genetic polymorphisms," *Methods Mol. Med.* 85:225-37 (2003). Pharmacogenomics as it relates to Alzheimer's Disease and other neurodegenerative disorders is discussed in Cacabelos, "Pharmacogenomics for the treatment of dementia," *Ann. Med.* 34(5):357-79 (2002); Maimone et al., "Pharmacogenomics of neurodegenerative diseases," *Eur. J. Pharmacol.* 9, 413(1): 11-29 (February 2001), and Poirier, "Apolipoprotein E: a pharmacogenetic target for the treatment of Alzheimer's Disease," *Mol. Diagn.* 4(4):335-41 (December 1999). Pharmacogenomics as it relates to cardiovascular disorders is discussed in Siest et al., "Pharmacogenomics of drugs affecting the cardiovascular system," *Clin. Chem. Lab Med.* 41(4):590-9 (April 2003); Mukherjee et al., "Pharmacogenomics in cardiovascular diseases," *Prog. Cardiovasc. Dis.* 44(6):479-98 (May-June 2002); and Mooser et al., "Cardiovascular pharmacogenetics in the SNP era," *J. Thromb. Haemost.* 1(7):1398-402 (July 2003). Pharmacogenomics as it relates to cancer is discussed in McLeod et al., "Cancer pharmacogenomics: SNPs, chips, and the individual patient," *Cancer Invest.* 21(4):630-40 (2003) and Watters et al., "Cancer pharmacogenomics: current and future applications," *Biochim. Biophys. Acta.* 17; 1603(2):99-111 (March 2003).

[0264] The SNPs of the present invention also can be used to identify novel therapeutic targets for Alzheimer's Disease. For example, genes containing the disease-associated variants ("variant genes") or their products, as well as genes or their products that are directly or indirectly regulated by or interacting with these variant genes or their products, can be targeted for the development of therapeutics that, for example, treat the disease or prevent or delay disease onset. The therapeutics may be composed of, for example, small molecules, proteins, protein fragments or peptides, antibodies, nucleic acids, or their derivatives or mimetics which modulate the functions or levels of the target genes or gene products.

[0265] The SNP-containing nucleic acid molecules disclosed herein, and their complementary nucleic acid molecules, may be used as antisense constructs to control gene expression in cells, tissues, and organisms. Antisense technology is well established in the art and extensively reviewed in *Antisense Drug Technology: Principles, Strategies, and Applications*, ed. Crooke, Marcel Dekker, Inc., New York (2001). An antisense nucleic acid molecule is generally designed to be complementary to a region of mRNA expressed by a gene so that the antisense molecule hybridizes to the mRNA and thereby blocks translation of mRNA into protein. Various classes of antisense oligonucleotides are used in the art, two of which are cleavers and blockers. Cleavers, by binding to target RNAs, activate intracellular nucleases (e.g., RNaseH or RNase L) that cleave the target RNA. Blockers, which also bind to target RNAs, inhibit protein translation through steric hindrance of

ribosomes. Exemplary blockers include peptide nucleic acids, morpholinos, locked nucleic acids, and methylphosphonates (see, e.g., Thompson, *Drug Discovery Today*, 7 [17]: 912-917 [2002]). Antisense oligonucleotides are directly useful as therapeutic agents, and are also useful for determining and validating gene function (e.g., in gene knock-out or knock-down experiments).

[0266] Antisense technology is further reviewed in: Lavery et al., "Antisense and RNAi: powerful tools in drug target discovery and validation," *Curr. Opin. Drug Discov. Devel.* 6(4):561-9 (July 2003); Stephens et al., "Antisense oligonucleotide therapy in cancer," *Curr. Opin. Mol. Ther.* 5(2): 118-22 (April 2003); Kurreck, "Antisense technologies. Improvement through novel chemical modifications," *Eur. J. Biochem.* 270(8): 1628-44 (April 2003); Dias et al., "Antisense oligonucleotides: basic concepts and mechanisms," *Mol. Cancer Ther.* 1(5):347-55 (March 2002); Chen, "Clinical development of antisense oligonucleotides as anticancer therapeutics," *Methods Mol. Med.* 75:621-36 (2003); Wang et al., "Antisense anticancer oligonucleotide therapeutics," *Curr. Cancer Drug Targets* 1(3):177-96 (November 2001); and Bennett, "Efficiency of antisense oligonucleotide drug discovery," *Antisense Nucleic Acid Drug. Dev.* 12(3):215-24 (June 2002).

[0267] The SNPs of the present invention are particularly useful for designing antisense reagents that are specific for particular nucleic acid variants. Based on the SNP information disclosed herein, antisense oligonucleotides can be produced that specifically target mRNA molecules that contain one or more particular SNP nucleotides. In this manner, expression of mRNA molecules that contain one or more undesired polymorphisms (e.g., SNP nucleotides that lead to a defective protein such as an amino acid substitution in a catalytic domain) can be inhibited or completely blocked. Thus, antisense oligonucleotides can be used to specifically bind a particular polymorphic form (e.g., a SNP allele that encodes a defective protein), thereby inhibiting translation of this form, but which do not bind an alternative polymorphic form (e.g., an alternative SNP nucleotide that encodes a protein having normal function).

[0268] Antisense molecules can be used to inactivate mRNA in order to inhibit gene expression and production of defective proteins. Accordingly, these molecules can be used to treat a disorder, such as Alzheimer's Disease, characterized by abnormal or undesired gene expression or expression of certain defective proteins. This technique can involve cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible mRNA regions include, for example, protein-coding regions and particularly protein-coding regions corresponding to catalytic activities, substrate/ligand binding, or other functional activities of a protein.

[0269] The SNPs of the present invention are also useful for designing RNA interference reagents that specifically target nucleic acid molecules having particular SNP variants. RNA interference (RNAi), also referred to as gene silencing, is based on using double-stranded RNA (dsRNA) molecules to turn genes off. When introduced into a cell, dsRNAs are processed by the cell into short fragments (generally about 21, 22, or 23 nucleotides in length) known as small interfering RNAs (siRNAs) which the cell uses in

a sequence-specific manner to recognize and destroy complementary RNAs (Thompson, *Drug Discovery Today*, 7 [17]: 912-917 [2002]). Accordingly, an aspect of the present invention specifically contemplates isolated nucleic acid molecules that are about 18-26 nucleotides in length, preferably 19-25 nucleotides in length, and more preferably 20, 21, 22, or 23 nucleotides in length, and the use of these nucleic acid molecules for RNAi. Because RNAi molecules, including siRNAs, act in a sequence-specific manner, the SNPs of the present invention can be used to design RNAi reagents that recognize and destroy nucleic acid molecules having specific SNP alleles/nucleotides (such as deleterious alleles that lead to the production of defective proteins), while not affecting nucleic acid molecules having alternative SNP alleles (such as alleles that encode proteins having normal function). As with antisense reagents, RNAi reagents may be directly useful as therapeutic agents (e.g., for turning off defective, disease-causing genes), and are also useful for characterizing and validating gene function (e.g., in gene knock-out or knock-down experiments).

[0270] The following references provide a further review of RNAi: Reynolds et al., "Rational siRNA design for RNA interference," *Nat. Biotechnol.* 22(3):326-30 (March 2004); Epub Feb. 1 2004; Chi et al., "Genomewide view of gene silencing by small interfering RNAs," *PNAS* 100(11):6343-6346 (2003); Vickers et al., "Efficient Reduction of Target RNAs by Small Interfering RNA and RNase H-dependent Antisense Agents," *J. Biol. Chem.* 278: 7108-7118 (2003); Agami, "RNAi and related mechanisms and their potential use for therapy," *Curr. Opin. Chem. Biol.* 6(6):829-34 (December 2002); Lavery et al., "Antisense and RNAi: powerful tools in drug target discovery and validation," *Curr. Opin. Drug Discov. Devel.* 6(4):561-9 (July 2003); Shi, "Mammalian RNAi for the masses," *Trends Genet.* 19(1):9-12 (January 2003); Shuey et al., "RNAi: gene-silencing in therapeutic intervention," *Drug Discovery Today* 7(20):1040-1046 (October 2002); McManus et al., *Nat. Rev. Genet.* 3(10):737-47 (October 2002); Xia et al., *Nat. Biotechnol.* 20(10):1006-10 (October 2002); Plasterk et al., *Curr. Opin. Genet. Dev.* 10(5):562-7 (October 2000); Boshier et al., *Nat. Cell Biol.* 2(2):E31-6 (February 2000); and Hunter, *Curr. Biol.* 9(12):R440-2 (Jun. 17 1999).

[0271] A subject suffering from a pathological condition, such as Alzheimer's Disease, ascribed to a SNP may be treated so as to correct the genetic defect (see Kren et al., *Proc. Natl. Acad. Sci. USA* 96:10349-10354 [1999]). Such a subject can be identified by any method that can detect the polymorphism in a biological sample drawn from the subject. Such a genetic defect may be permanently corrected by administering to such a subject a nucleic acid fragment incorporating a repair sequence that supplies the normal/wild-type nucleotide at the position of the SNP. This site-specific repair sequence can encompass an RNA/DNA oligonucleotide that operates to promote endogenous repair of a subject's genomic DNA. The site-specific repair sequence is administered in an appropriate vehicle, such as a complex with polyethylenimine, encapsulated in anionic liposomes, a viral vector such as an adenovirus, or other pharmaceutical composition that promotes intracellular uptake of the administered nucleic acid. A genetic defect leading to an inborn pathology may then be overcome, as the chimeric oligonucleotides induce incorporation of the normal sequence into the subject's genome. Upon incorporation, the normal gene product is expressed, and the replacement is propa-

gated, thereby engendering a permanent repair and therapeutic enhancement of the clinical condition of the subject.

[0272] In cases in which a cSNP results in a variant protein that is ascribed to be the cause of, or a contributing factor to, a pathological condition, a method of treating such a condition can include administering to a subject experiencing the pathology the wild-type/normal cognate of the variant protein. Once administered in an effective dosing regimen, the wild-type cognate provides complementation or remediation of the pathological condition.

[0273] The invention further provides a method for identifying a compound or agent that can be used to treat Alzheimer's Disease. The SNPs disclosed herein are useful as targets for the identification and/or development of therapeutic agents. A method for identifying a therapeutic agent or compound typically includes assaying the ability of the agent or compound to modulate the activity and/or expression of a SNP-containing nucleic acid or the encoded product and thus identifying an agent or a compound that can be used to treat a disorder characterized by undesired activity or expression of the SNP-containing nucleic acid or the encoded product. The assays can be performed in cell-based and cell-free systems. Cell-based assays can include cells naturally expressing the nucleic acid molecules of interest or recombinant cells genetically engineered to express certain nucleic acid molecules.

[0274] Variant gene expression in a Alzheimer's Disease patient can include, for example, either expression of a SNP-containing nucleic acid sequence (for instance, a gene that contains a SNP can be transcribed into an mRNA transcript molecule containing the SNP, which can in turn be translated into a variant protein) or altered expression of a normal/wild-type nucleic acid sequence due to one or more SNPs (for instance, a regulatory/control region can contain a SNP that affects the level or pattern of expression of a normal transcript).

[0275] Assays for variant gene expression can involve direct assays of nucleic acid levels (e.g., mRNA levels), expressed protein levels, or of collateral compounds involved in a signal pathway. Further, the expression of genes that are up- or down-regulated in response to the signal pathway can also be assayed. In this embodiment, the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

[0276] Modulators of variant gene expression can be identified in a method wherein, for example, a cell is contacted with a candidate compound/agent and the expression of mRNA determined. The level of expression of mRNA in the presence of the candidate compound is compared to the level of expression of mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of variant gene expression based on this comparison and be used to treat a disorder such as Alzheimer's Disease that is characterized by variant gene expression (e.g., either expression of a SNP-containing nucleic acid or altered expression of a normal/wild-type nucleic acid molecule due to one or more SNPs that affect expression of the nucleic acid molecule) due to one or more SNPs of the present invention. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When

nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

[0277] The invention further provides methods of treatment, with the SNP or associated nucleic acid domain (e.g., catalytic domain, ligand/substrate-binding domain, regulatory/control region, etc.) or gene, or the encoded mRNA transcript, as a target, using a compound identified through drug screening as a gene modulator to modulate variant nucleic acid expression. Modulation can include either up-regulation (i.e., activation or agonization) or down-regulation (i.e., suppression or antagonization) of nucleic acid expression.

[0278] Expression of mRNA transcripts and encoded proteins, either wild type or variant, may be altered in individuals with a particular SNP allele in a regulatory/control element, such as a promoter or transcription factor binding domain, that regulates expression. In this situation, methods of treatment and compounds can be identified, as discussed herein, that regulate or overcome the variant regulatory/control element, thereby generating normal, or healthy, expression levels of either the wild type or variant protein.

[0279] The SNP-containing nucleic acid molecules of the present invention are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of a variant gene, or encoded product, in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as an indicator for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance, as well as an indicator for toxicities. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

[0280] In another aspect of the present invention, there is provided a pharmaceutical pack comprising a therapeutic agent (e.g., a small molecule drug, antibody, peptide, antisense or RNAi nucleic acid molecule, etc.) and a set of instructions for administration of the therapeutic agent to humans diagnostically tested for one or more SNPs or SNP haplotypes provided by the present invention.

[0281] The SNPs/haplotypes of the present invention are also useful for improving many different aspects of the drug development process. For instance, an aspect of the present invention includes selecting individuals for clinical trials based on their SNP genotype. For example, individuals with SNP genotypes that indicate that they are likely to positively respond to a drug can be included in the trials, whereas those individuals whose SNP genotypes indicate that they are less likely to or would not respond to the drug, or who are at risk for suffering toxic effects or other adverse reactions, can be excluded from the clinical trials. This not only can improve the safety of clinical trials, but also can enhance the chances that the trial will demonstrate statistically significant efficacy. Furthermore, the SNPs of the present invention may explain why certain previously developed drugs performed

poorly in clinical trials and may help identify a subset of the population that would benefit from a drug that had previously performed poorly in clinical trials, thereby "rescuing" previously developed drugs, and enabling the drug to be made available to a particular Alzheimer's Disease patient population that can benefit from it.

[0282] SNPs have many important uses in drug discovery, screening, and development. A high probability exists that, for any gene/protein selected as a potential drug target, variants of that gene/protein will exist in a patient population. Thus, determining the impact of gene/protein variants on the selection and delivery of a therapeutic agent should be an integral aspect of the drug discovery and development process. (Jazwinska, *A Trends Guide to Genetic Variation and Genomic Medicine* S30-S36 [March 2002]).

[0283] Knowledge of variants (e.g., SNPs and any corresponding amino acid polymorphisms) of a particular therapeutic target (e.g., a gene, mRNA transcript, or protein) enables parallel screening of the variants in order to identify therapeutic candidates (e.g., small molecule compounds, antibodies, antisense or RNAi nucleic acid compounds, etc.) that demonstrate efficacy across variants (Rothberg, *Nat. Biotechnol.* 19[3]:209-11 [March 2001]). Such therapeutic candidates would be expected to show equal efficacy across a larger segment of the patient population, thereby leading to a larger potential market for the therapeutic candidate.

[0284] Furthermore, identifying variants of a potential therapeutic target enables the most common form of the target to be used for selection of therapeutic candidates, thereby helping to ensure that the experimental activity that is observed for the selected candidates reflects the real activity expected in the largest proportion of a patient population (Jazwinska, *A Trends Guide to Genetic Variation and Genomic Medicine* S30-S36 [March 2002]).

[0285] Additionally, screening therapeutic candidates against all known variants of a target can enable the early identification of potential toxicities and adverse reactions relating to particular variants. For example, variability in drug absorption, distribution, metabolism and excretion (ADME) caused by, for example, SNPs in therapeutic targets or drug metabolizing genes, can be identified, and this information can be utilized during the drug development process to minimize variability in drug disposition and develop therapeutic agents that are safer across a wider range of a patient population. The SNPs of the present invention, including the variant proteins and encoding polymorphic nucleic acid molecules provided in Tables 1-2, are useful in conjunction with a variety of toxicology methods established in the art, such as those set forth in *Current Protocols in Toxicology*, John Wiley & Sons, Inc., New York.

[0286] Furthermore, therapeutic agents that target any art-known proteins (or nucleic acid molecules, either RNA or DNA) may cross-react with the variant proteins (or polymorphic nucleic acid molecules) disclosed in Table 1, thereby significantly affecting the pharmacokinetic properties of the drug. Consequently, the protein variants and the SNP-containing nucleic acid molecules disclosed in Tables 1-2 are useful in developing, screening, and evaluating therapeutic agents that target corresponding art-known protein forms (or nucleic acid molecules). Additionally, as discussed above, knowledge of all polymorphic forms of a

particular drug target enables the design of therapeutic agents that are effective against most or all such polymorphic forms of the drug target.

[0287] Pharmaceutical Compositions and Administration Thereof

[0288] Any of the Alzheimer's Disease-associated proteins, and encoding nucleic acid molecules, disclosed herein can be used as therapeutic targets (or directly used themselves as therapeutic compounds) for treating Alzheimer's Disease and related pathologies, and the present disclosure enables therapeutic compounds (e.g., small molecules, antibodies, therapeutic proteins, RNAi and antisense molecules, etc.) to be developed that target (or are comprised of) any of these therapeutic targets.

[0289] In general, a therapeutic compound will be administered in a therapeutically effective amount by any of the accepted modes of administration for agents that serve similar utilities. The actual amount of the therapeutic compound of this invention, i.e., the active ingredient, will depend upon numerous factors such as the severity of the disease to be treated, the age and relative health of the subject, the potency of the compound used, the route and form of administration, and other factors.

[0290] Therapeutically effective amounts of therapeutic compounds may range from, for example, approximately 0.01-50 mg per kilogram body weight of the recipient per day; preferably about 0.1-20 mg/kg/day. Thus, as an example, for administration to a 70 kg person, the dosage range would most preferably be about 7 mg to 1.4 g per day.

[0291] In general, therapeutic compounds will be administered as pharmaceutical compositions by any one of the following routes: oral, systemic (e.g., transdermal, intranasal, or by suppository), or parenteral (e.g., intramuscular, intravenous, or subcutaneous) administration. The preferred manner of administration is oral or parenteral using a convenient daily dosage regimen, which can be adjusted according to the degree of affliction. Oral compositions can take the form of tablets, pills, capsules, semisolids, powders, sustained release formulations, solutions, suspensions, elixirs, aerosols, or any other appropriate compositions.

[0292] The choice of formulation depends on various factors such as the mode of drug administration (e.g., for oral administration, formulations in the form of tablets, pills, or capsules are preferred) and the bioavailability of the drug substance. Recently, pharmaceutical formulations have been developed especially for drugs that show poor bioavailability based upon the principle that bioavailability can be increased by increasing the surface area, i.e., decreasing particle size. For example, U.S. Pat. No. 4,107,288 describes a pharmaceutical formulation having particles in the size range from 10 to 1,000 nm in which the active material is supported on a cross-linked matrix of macromolecules. U.S. Pat. No. 5,145,684 describes the production of a pharmaceutical formulation in which the drug substance is pulverized to nanoparticles (average particle size of 400 nm) in the presence of a surface modifier and then dispersed in a liquid medium to give a pharmaceutical formulation that exhibits remarkably high bioavailability.

[0293] Pharmaceutical compositions are comprised of, in general, a therapeutic compound in combination with at least one pharmaceutically acceptable excipient. Acceptable

excipients are non-toxic, aid administration, and do not adversely affect the therapeutic benefit of the therapeutic compound. Such excipients may be any solid, liquid, semi-solid or, in the case of an aerosol composition, gaseous excipient that is generally available to one skilled in the art.

[0294] Solid pharmaceutical excipients include starch, cellulose, talc, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, magnesium stearate, sodium stearate, glycerol monostearate, sodium chloride, dried skim milk and the like. Liquid and semisolid excipients may be selected from glycerol, propylene glycol, water, ethanol and various oils, including those of petroleum, animal, vegetable or synthetic origin, e.g., peanut oil, soybean oil, mineral oil, sesame oil, etc. Preferred liquid carriers, particularly for injectable solutions, include water, saline, aqueous dextrose, and glycols.

[0295] Compressed gases may be used to disperse a compound of this invention in aerosol form. Inert gases suitable for this purpose are nitrogen, carbon dioxide, etc.

[0296] Other suitable pharmaceutical excipients and their formulations are described in *Remington's Pharmaceutical Sciences* 18th edition, ed. E. W. Martin (Mack Publishing Company [1990]).

[0297] The amount of the therapeutic compound in a formulation can vary within the full range employed by those skilled in the art. Typically, the formulation will contain, on a weight percent (wt %) basis, from about 0.01-99.99 wt % of the therapeutic compound based on the total formulation, with the balance being one or more suitable pharmaceutical excipients. Preferably, the compound is present at a level of about 1-80 wt %.

[0298] Therapeutic compounds can be administered alone or in combination with other therapeutic compounds or in combination with one or more other active ingredient(s). For example, an inhibitor or stimulator of a Alzheimer's Disease-associated protein can be administered in combination with another agent that inhibits or stimulates the activity of the same or a different Alzheimer's Disease-associated protein to thereby counteract the affects of Alzheimer's Disease.

[0299] For further information regarding pharmacology, see *Current Protocols in Pharmacology*, John Wiley & Sons, Inc., New York.

[0300] Human Identification Applications

[0301] In addition to their diagnostic and therapeutic uses in Alzheimer's Disease and related pathologies, the SNPs provided by the present invention are also useful as human identification markers for such applications as forensics, paternity testing, and biometrics (see, e.g., Gill, "An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes," *Int. J. Legal Med.* 114-[4-5]:204-10 [2001]). Genetic variations in the nucleic acid sequences between individuals can be used as genetic markers to identify individuals and to associate a biological sample with an individual. Determination of which nucleotides occupy a set of SNP positions in an individual identifies a set of SNP markers that distinguishes the individual. The more SNP positions that are analyzed, the lower the probability that the set of SNPs in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked (i.e., inherited

independently). Thus, preferred sets of SNPs can be selected from among the SNPs disclosed herein, which may include SNPs on different chromosomes, SNPs on different chromosome arms, and/or SNPs that are dispersed over substantial distances along the same chromosome arm.

[0302] Furthermore, among the SNPs disclosed herein, preferred SNPs for use in certain forensic/human identification applications include SNPs located at degenerate codon positions (i.e., the third position in certain codons which can be one of two or more alternative nucleotides and still encode the same amino acid), since these SNPs do not affect the encoded protein. SNPs that do not affect the encoded protein are expected to be under less selective pressure and are therefore expected to be more polymorphic in a population, which is typically an advantage for forensic/human identification applications. However, for certain forensic/human identification applications, such as predicting phenotypic characteristics (e.g., inferring ancestry or inferring one or more physical characteristics of an individual) from a DNA sample, it may be desirable to utilize SNPs that affect the encoded protein.

[0303] For many of the SNPs disclosed in Tables 1-2 (which are identified as "Applera" SNP source), Tables 1-2 provide SNP allele frequencies obtained by re-sequencing the DNA of chromosomes from 39 individuals (Tables 1-2 also provide allele frequency information for "Celera" source SNPs and, where available, public SNPs from dbEST, HGBASE, and/or HGMD). The allele frequencies provided in Tables 1-2 enable these SNPs to be readily used for human identification applications. Although any SNP disclosed in Table 1 and/or Table 2 could be used for human identification, the closer that the frequency of the minor allele at a particular SNP site is to 50%, the greater the ability of that SNP to discriminate between different individuals in a population since it becomes increasingly likely that two randomly selected individuals would have different alleles at that SNP site. Using the SNP allele frequencies provided in Tables 1-2, one of ordinary skill in the art could readily select a subset of SNPs for which the frequency of the minor allele is, for example, at least 1%, 2%, 5%, 10%, 20%, 25%, 30%, 40%, 45%, or 50%, or any other frequency in-between. Thus, since Tables 1-2 provide allele frequencies based on the re-sequencing of the chromosomes from 39 individuals, a subset of SNPs could readily be selected for human identification in which the total allele count of the minor allele at a particular SNP site is, for example, at least 1, 2, 4, 8, 10, 16, 20, 24, 30, 32, 36, 38, 39, 40, or any other number in-between.

[0304] Furthermore, Tables 1-2 also provide population group (interchangeably referred to herein as ethnic or racial groups) information coupled with the extensive allele frequency information. For example, the group of 39 individuals whose DNA was re-sequenced was made-up of 20 Caucasians and 19 African-Americans. This population group information enables further refinement of SNP selection for human identification. For example, preferred SNPs for human identification can be selected from Tables 1-2 that have similar allele frequencies in both the Caucasian and African-American populations; thus, for example, SNPs can be selected that have equally high discriminatory power in both populations. Alternatively, SNPs can be selected for which there is a statistically significant difference in allele frequencies between the Caucasian and African-American

populations (as an extreme example, a particular allele may be observed only in either the Caucasian or the African-American population group but not observed in the other population group); such SNPs are useful, for example, for predicting the race/ethnicity of an unknown perpetrator from a biological sample such as a hair or blood stain recovered at a crime scene. For a discussion of using SNPs to predict ancestry from a DNA sample, including statistical methods, see Frudakis et al., "A Classifier for the SNP-Based Inference of Ancestry," *Journal of Forensic Sciences* 48[4]:771-782 [2003].

[0305] SNPs have numerous advantages over other types of polymorphic markers, such as short tandem repeats (STRs). For example, SNPs can be easily scored and are amenable to automation, making SNPs the markers of choice for large-scale forensic databases. SNPs are found in much greater abundance throughout the genome than repeat polymorphisms. Population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci. SNPs are mutationally more stable than repeat polymorphisms. SNPs are not susceptible to artefacts such as stutter bands that can hinder analysis. Stutter bands are frequently encountered when analyzing repeat polymorphisms, and are particularly troublesome when analyzing samples such as crime scene samples that may contain mixtures of DNA from multiple sources. Another significant advantage of SNP markers over STR markers is the much shorter length of nucleic acid needed to score a SNP. For example, STR markers are generally several hundred base pairs in length. A SNP, on the other hand, comprises a single nucleotide, and generally a short conserved region on either side of the SNP position for primer and/or probe binding. This makes SNPs more amenable to typing in highly degraded or aged biological samples that are frequently encountered in forensic casework in which DNA may be fragmented into short pieces.

[0306] SNPs also are not subject to microvariant and "off-ladder" alleles frequently encountered when analyzing STR loci. Microvariants are deletions or insertions within a repeat unit that change the size of the amplified DNA product so that the amplified product does not migrate at the same rate as reference alleles with normal sized repeat units. When separated by size, such as by electrophoresis on a polyacrylamide gel, microvariants do not align with a reference allelic ladder of standard sized repeat units, but rather migrate between the reference alleles. The reference allelic ladder is used for precise sizing of alleles for allele classification; therefore alleles that do not align with the reference allelic ladder lead to substantial analysis problems. Furthermore, when analyzing multi-allelic repeat polymorphisms, occasionally an allele is found that consists of more or less repeat units than has been previously seen in the population, or more or less repeat alleles than are included in a reference allelic ladder. These alleles will migrate outside the size range of known alleles in a reference allelic ladder, and therefore are referred to as "off-ladder" alleles. In extreme cases, the allele may contain so few or so many repeats that it migrates well out of the range of the reference allelic ladder. In this situation, the allele may not even be observed, or, with multiplex analysis, it may migrate within or close to the size range for another locus, further confounding analysis.

[0307] SNP analysis avoids the problems of microvariants and off-ladder alleles encountered in STR analysis. Importantly, microvariants and off-ladder alleles may provide significant problems, and may be completely missed, when using analysis methods such as oligonucleotide hybridization arrays, which utilize oligonucleotide probes specific for certain known alleles. Furthermore, off-ladder alleles and microvariants encountered with STR analysis, even when correctly typed, may lead to improper statistical analysis, since their frequencies in the population are generally unknown or poorly characterized, and therefore the statistical significance of a matching genotype may be questionable. All these advantages of SNP analysis are considerable in light of the consequences of most DNA identification cases, which may lead to life imprisonment for an individual, or re-association of remains to the family of a deceased individual.

[0308] DNA can be isolated from biological samples such as blood, bone, hair, saliva, or semen, and compared with the DNA from a reference source at particular SNP positions. Multiple SNP markers can be assayed simultaneously in order to increase the power of discrimination and the statistical significance of a matching genotype. For example, oligonucleotide arrays can be used to genotype a large number of SNPs simultaneously. The SNPs provided by the present invention can be assayed in combination with other polymorphic genetic markers, such as other SNPs known in the art or STRs, in order to identify an individual or to associate an individual with a particular biological sample.

[0309] Furthermore, the SNPs provided by the present invention can be genotyped for inclusion in a database of DNA genotypes, for example, a criminal DNA databank such as the FBI's Combined DNA Index System (CODIS) database. A genotype obtained from a biological sample of unknown source can then be queried against the database to find a matching genotype, with the SNPs of the present invention providing nucleotide positions at which to compare the known and unknown DNA sequences for identity. Accordingly, the present invention provides a database comprising novel SNPs or SNP alleles of the present invention (e.g., the database can comprise information indicating which alleles are possessed by individual members of a population at one or more novel SNP sites of the present invention), such as for use in forensics, biometrics, or other human identification applications. Such a database typically comprises a computer-based system in which the SNPs or SNP alleles of the present invention are recorded on a computer readable medium (see the section of the present specification entitled "Computer-Related Embodiments").

[0310] The SNPs of the present invention can also be assayed for use in paternity testing. The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child, with the SNPs of the present invention providing nucleotide positions at which to compare the putative father's and child's DNA sequences for identity. If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the

putative father, it can be concluded, barring experimental error, that the putative father is not the father of the child. If the set of polymorphisms in the child attributable to the father match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match, and a conclusion drawn as to the likelihood that the putative father is the true biological father of the child.

[0311] In addition to paternity testing, SNPs are also useful for other types of kinship testing, such as for verifying familial relationships for immigration purposes, or for cases in which an individual alleges to be related to a deceased individual in order to claim an inheritance from the deceased individual, etc. For further information regarding the utility of SNPs for paternity testing and other types of kinship testing, including methods for statistical analysis, see Krawczak, "Informativity assessment for biallelic single nucleotide polymorphisms," *Electrophoresis* 20(8):1676-81 (June 1999).

[0312] The use of the SNPs of the present invention for human identification further extends to various authentication systems, commonly referred to as biometric systems, which typically convert physical characteristics of humans (or other organisms) into digital data. Biometric systems include various technological devices that measure such unique anatomical or physiological characteristics as finger, thumb, or palm prints; hand geometry; vein patterning on the back of the hand; blood vessel patterning of the retina and color and texture of the iris; facial characteristics; voice patterns; signature and typing dynamics; and DNA. Such physiological measurements can be used to verify identity and, for example, restrict or allow access based on the identification. Examples of applications for biometrics include physical area security, computer and network security, aircraft passenger check-in and boarding, financial transactions, medical records access, government benefit distribution, voting, law enforcement, passports, visas and immigration, prisons, various military applications, and for restricting access to expensive or dangerous items, such as automobiles or guns (see, for example, O'Connor, *Stanford Technology Law Review and U.S. Pat. No. 6,119,096*).

[0313] Groups of SNPs, particularly the SNPs provided by the present invention, can be typed to uniquely identify an individual for biometric applications such as those described above. Such SNP typing can readily be accomplished using, for example, DNA chips/arrays. Preferably, a minimally invasive means for obtaining a DNA sample is utilized. For example, PCR amplification enables sufficient quantities of DNA for analysis to be obtained from buccal swabs or fingerprints, which contain DNA-containing skin cells and oils that are naturally transferred during contact. Further information regarding techniques for using SNPs in forensic/human identification applications can be found in, for example, *Current Protocols in Human Genetics* 14.1-14.7, John Wiley & Sons, New York (2002).

[0314] Variant Proteins, Antibodies, Vectors & Host Cells, & Uses Thereof

[0315] Variant Proteins Encoded by SNP-Containing Nucleic Acid Molecules

[0316] The present invention provides SNP-containing nucleic acid molecules, many of which encode proteins

having variant amino acid sequences as compared to the art-known (i.e., wild-type) proteins. Amino acid sequences encoded by the polymorphic nucleic acid molecules of the present invention are provided as SEQ ID NOS:6-10 in Table 1 and the Sequence Listing. These variants will generally be referred to herein as variant proteins/peptides/polypeptides, or polymorphic proteins/peptides/polypeptides of the present invention. The terms "protein," "peptide," and "polypeptide" are used herein interchangeably.

[0317] A variant protein of the present invention may be encoded by, for example, a nonsynonymous nucleotide substitution at any one of the cSNP positions disclosed herein. In addition, variant proteins may also include proteins whose expression, structure, and/or function is altered by a SNP disclosed herein, such as a SNP that creates or destroys a stop codon, a SNP that affects splicing, and a SNP in control/regulatory elements, e.g. promoters, enhancers, or transcription factor binding domains.

[0318] As used herein, a protein or peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or chemical precursors or other chemicals. The variant proteins of the present invention can be purified to homogeneity or other lower degrees of purity. The level of purification will be based on the intended use. The key feature is that the preparation allows for the desired function of the variant protein, even if in the presence of considerable amounts of other components.

[0319] As used herein, "substantially free of cellular material" includes preparations of the variant protein having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins.

When the variant protein is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

[0320] The language "substantially free of chemical precursors or other chemicals" includes preparations of the variant protein in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the variant protein having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

[0321] An isolated variant protein may be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant host cells), or synthesized using known protein synthesis methods. For example, a nucleic acid molecule containing SNP(s) encoding the variant protein can be cloned into an expression vector, the expression vector introduced into a host cell, and the variant protein expressed in the host cell. The variant protein can then be isolated from the cells by any appropriate purification scheme using standard protein purification techniques. Examples of these techniques are described in detail below (Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [2000]).

[0322] The present invention provides isolated variant proteins that comprise, consist of or consist essentially of amino acid sequences that contain one or more variant amino acids encoded by one or more codons which contain a SNP of the present invention.

[0323] Accordingly, the present invention provides variant proteins that consist of amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein consists of an amino acid sequence when the amino acid sequence is the entire amino acid sequence of the protein.

[0324] The present invention further provides variant proteins that consist essentially of amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues in the final protein.

[0325] The present invention further provides variant proteins that comprise amino acid sequences that contain one or more amino acid polymorphisms (or truncations or extensions due to creation or destruction of a stop codon, respectively) encoded by the SNPs provided in Table 1 and/or Table 2. A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein may contain only the variant amino acid sequence or have additional amino acid residues, such as a contiguous encoded sequence that is naturally associated with it or heterologous amino acid residues. Such a protein can have a few additional amino acid residues or can comprise many more additional amino acids. A brief description of how various types of these proteins can be made and isolated is provided below.

[0326] The variant proteins of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a variant protein operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the variant protein. "Operatively linked" indicates that the coding sequences for the variant protein and the heterologous protein are ligated in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the variant protein. In another embodiment, the fusion protein is encoded by a fusion polynucleotide that is synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see Ausubel et al., *Current Protocols in Molecular Biology* [1992]). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A variant protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the variant protein.

[0327] In many uses, the fusion protein does not affect the activity of the variant protein. The fusion protein can

include, but is not limited to, enzymatic fusion proteins, for example, beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate their purification following recombinant expression. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Fusion proteins are further described in, for example, Terpe, "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems," *Appl. Microbiol. Biotechnol.* 60(5):523-33 (January 2003); Epub Nov. 7 2002; Graddis et al., "Designing proteins that work using recombinant technologies," *Curr. Pharm. Biotechnol.* 3(4):285-97 (December 2002); and Nilsson et al., "Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins," *Protein Expr. Purif.* 11(1): 1-16 (October 1997).

[0328] The present invention also relates to further obvious variants of the variant polypeptides of the present invention, such as naturally-occurring mature forms (e.g., allelic variants), non-naturally occurring recombinantly-derived variants, and orthologs and paralogs of such proteins that share sequence homology. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude those known in the prior art before the present invention.

[0329] Further variants of the variant polypeptides disclosed in Table 1 can comprise an amino acid sequence that shares at least 70-80%, 80-85%, 85-90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with an amino acid sequence disclosed in Table 1 (or a fragment thereof) and that includes a novel amino acid residue (allele) disclosed in Table 1 (which is encoded by a novel SNP allele). Thus, an aspect of the present invention that is specifically contemplated are polypeptides that have a certain degree of sequence variation compared with the polypeptide sequences shown in Table 1, but that contain a novel amino acid residue (allele) encoded by a novel SNP allele disclosed herein. In other words, as long as a polypeptide contains a novel amino acid residue disclosed herein, other portions of the polypeptide that flank the novel amino acid residue can vary to some degree from the polypeptide sequences shown in Table 1.

[0330] Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the amino acid sequences disclosed herein can readily be identified as having complete sequence identity to one of the variant proteins of the present invention as well as being encoded by the same genetic locus as the variant proteins provided herein.

[0331] Orthologs of a variant peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of a variant peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from non-human mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs can be encoded by a nucleic acid sequence that hybridizes to a variant peptide-encoding nucleic acid molecule under moderate to stringent conditions depending on the degree of relatedness of the two organisms yielding the homologous proteins.

[0332] Variant proteins include, but are not limited to, proteins containing deletions, additions and substitutions in the amino acid sequence caused by the SNPs of the present invention. One class of substitutions is conserved amino acid substitutions in which a given amino acid in a polypeptide is substituted for another amino acid of like characteristics. Typical conservative substitutions are replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in, for example, Bowie et al., *Science* 247:1306-1310 (1990).

[0333] Variant proteins can be fully functional or can lack function in one or more activities, e.g. ability to bind another molecule, ability to catalyze a substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variations or variations in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, truncations or extensions, or a substitution, insertion, inversion, or deletion of a critical residue or in a critical region.

[0334] Amino acids that are essential for function of a protein can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham et al., *Science* 244:1081-1085 [1989]), particularly using the amino acid sequence and polymorphism information provided in Table 1. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as enzyme activity or in assays such as an in vitro proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al., *J. Mol. Biol.* 224:899-904 [1992]; de Vos et al., *Science* 255:306-312 [1992]).

[0335] Polypeptides can contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Accordingly, the variant proteins of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (e.g., polyethylene glycol), or in which additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

[0336] Known protein modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amida-

tion, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

[0337] Such protein modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins—Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W.H. Freeman and Company, New York (1993); F. Wold, *Posttranslational Covalent Modification of Proteins* 1-12, ed. B. C. Johnson, Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182: 626-646 (1990); and Rattan et al., *Ann. N.Y. Acad. Sci.* 663:48-62 (1992).

[0338] The present invention further provides fragments of the variant proteins in which the fragments contain one or more amino acid sequence variations (e.g., substitutions, or truncations or extensions due to creation or destruction of a stop codon) encoded by one or more SNPs disclosed herein. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that have been disclosed in the prior art before the present invention.

[0339] As used herein, a fragment may comprise at least about 4, 8, 10, 12, 14, 16, 18, 20, 25, 30, 50, 100 (or any other number in-between) or more contiguous amino acid residues from a variant protein, wherein at least one amino acid residue is affected by a SNP of the present invention, e.g., a variant amino acid residue encoded by a nonsynonymous nucleotide substitution at a cSNP position provided by the present invention. The variant amino acid encoded by a cSNP may occupy any residue position along the sequence of the fragment. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the variant protein or the ability to perform a function, e.g., act as an immunogen. Particularly important fragments are biologically active fragments. Such fragments will typically comprise a domain or motif of a variant protein of the present invention, e.g., active site, transmembrane domain, or ligand/substrate binding domain. Other fragments include, but are not limited to, domain or motif-containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known to those of skill in the art (e.g., PROSITE analysis) (*Current Protocols in Protein Science*, John Wiley & Sons, New York [2002]).

[0340] Uses of Variant Proteins

[0341] The variant proteins of the present invention can be used in a variety of ways, including but not limited to, in assays to determine the biological activity of a variant

protein, such as in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another type of immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the variant protein (or its binding partner) in biological fluids; as a marker for cells or tissues in which it is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); as a target for screening for a therapeutic agent; and as a direct therapeutic agent to be administered into a human subject. Any of the variant proteins disclosed herein may be developed into reagent grade or kit format for commercialization as research products. Methods for performing the uses listed above are well known to those skilled in the art (see, e.g., *Molecular Cloning: A Laboratory Manual*, Sambrook and Russell, Cold Spring Harbor Laboratory Press, New York [2000], and *Methods in Enzymology: Guide to Molecular Cloning Techniques*, eds. S. L. Berger and A. R. Kimmel, Academic Press [1987]).

[0342] In a specific embodiment of the invention, the methods of the present invention include detection of one or more variant proteins disclosed herein. Variant proteins are disclosed in Table 1 and in the Sequence Listing as SEQ ID NOS: 6-10. Detection of such proteins can be accomplished using, for example, antibodies, small molecule compounds, aptamers, ligands/substrates, other proteins or protein fragments, or other protein-binding agents. Preferably, protein detection agents are specific for a variant protein of the present invention and can therefore discriminate between a variant protein of the present invention and the wild-type protein or another variant form. This can generally be accomplished by, for example, selecting or designing detection agents that bind to the region of a protein that differs between the variant and wild-type protein, such as a region of a protein that contains one or more amino acid substitutions that is/are encoded by a non-synonymous cSNP of the present invention, or a region of a protein that follows a nonsense mutation-type SNP that creates a stop codon thereby leading to a shorter polypeptide, or a region of a protein that follows a read-through mutation-type SNP that destroys a stop codon thereby leading to a longer polypeptide in which a portion of the polypeptide is present in one version of the polypeptide but not the other.

[0343] In another specific aspect of the invention, the variant proteins of the present invention are used as targets for diagnosing Alzheimer's Disease or for determining predisposition to Alzheimer's Disease in a human. Accordingly, the invention provides methods for detecting the presence of, or levels of, one or more variant proteins of the present invention in a cell, tissue, or organism. Such methods typically involve contacting a test sample with an agent (e.g., an antibody, small molecule compound, or peptide) capable of interacting with the variant protein such that specific binding of the agent to the variant protein can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an array, for example, an antibody or aptamer array (arrays for protein detection may also be referred to as "protein chips"). The variant protein of interest can be isolated from a test sample and assayed for the presence of a variant amino acid sequence encoded by one or more SNPs disclosed by the present invention. The SNPs may cause changes to the protein and the corresponding protein function/activity, such as through non-synonymous substitutions in protein coding

regions that can lead to amino acid substitutions, deletions, insertions, and/or rearrangements; formation or destruction of stop codons; or alteration of control elements such as promoters. SNPs may also cause inappropriate post-translational modifications.

[0344] One preferred agent for detecting a variant protein in a sample is an antibody capable of selectively binding to a variant form of the protein (antibodies are described in greater detail in the next section). Such samples include, for example, tissues, cells, and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

[0345] In vitro methods for detection of the variant proteins associated with Alzheimer's Disease that are disclosed herein and fragments thereof include, but are not limited to, enzyme linked immunosorbent assays (ELISAs), radioimmunoassays (RIA), Western blots, immunoprecipitations, immunofluorescence, and protein arrays/chips (e.g., arrays of antibodies or aptamers). For further information regarding immunoassays and related protein detection methods, see *Current Protocols in Immunology*, John Wiley & Sons, New York, and Hage, "Immunoassays," *Anal. Chem.* 71(12):294R-304R (Jun. 15 1999).

[0346] Additional analytic methods of detecting amino acid variants include, but are not limited to, altered electrophoretic mobility, altered tryptic peptide digest, altered protein activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, and direct amino acid sequencing.

[0347] Alternatively, variant proteins can be detected in vivo in a subject by introducing into the subject a labeled antibody (or other type of detection reagent) specific for a variant protein. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[0348] Other uses of the variant peptides of the present invention are based on the class or action of the protein. For example, proteins isolated from humans and their mammalian orthologs serve as targets for identifying agents (e.g., small molecule drugs or antibodies) for use in therapeutic applications, particularly for modulating a biological or pathological response in a cell or tissue that expresses the protein. Pharmaceutical agents can be developed that modulate protein activity.

[0349] As an alternative to modulating gene expression, therapeutic compounds can be developed that modulate protein function. For example, many SNPs disclosed herein affect the amino acid sequence of the encoded protein (e.g., non-synonymous cSNPs and nonsense mutation-type SNPs). Such alterations in the encoded amino acid sequence may affect protein function, particularly if such amino acid sequence variations occur in functional protein domains, such as catalytic domains, ATP-binding domains, or ligand/substrate binding domains. It is well established in the art that variant proteins having amino acid sequence variations in functional domains can cause or influence pathological conditions. In such instances, compounds (e.g., small molecule drugs or antibodies) can be developed that target the variant protein and modulate (e.g., up- or down-regulate) protein function/activity.

[0350] The therapeutic methods of the present invention further include methods that target one or more variant

proteins of the present invention. Variant proteins can be targeted using, for example, small molecule compounds, antibodies, aptamers, ligands/substrates, other proteins, or other protein-binding agents. Additionally, the skilled artisan will recognize that the novel protein variants (and polymorphic nucleic acid molecules) disclosed in Table 1 may themselves be directly used as therapeutic agents by acting as competitive inhibitors of corresponding art-known proteins (or nucleic acid molecules such as mRNA molecules).

[0351] The variant proteins of the present invention are particularly useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can utilize cells that naturally express the protein, a biopsy specimen, or cell cultures. In one embodiment, cell-based assays involve recombinant host cells expressing the variant protein. Cell-free assays can be used to detect the ability of a compound to directly bind to a variant protein or to the corresponding SNP-containing nucleic acid fragment that encodes the variant protein.

[0352] A variant protein of the present invention, as well as appropriate fragments thereof, can be used in high-throughput screening assays to test candidate compounds for the ability to bind and/or modulate the activity of the variant protein. These candidate compounds can be further screened against a protein having normal function (e.g., a wild-type/non-variant protein) to further determine the effect of the compound on the protein activity. Furthermore, these compounds can be tested in animal or invertebrate systems to determine in vivo activity/effectiveness. Compounds can be identified that activate (agonists) or inactivate (antagonists) the variant protein, and different compounds can be identified that cause various degrees of activation or inactivation of the variant protein.

[0353] Further, the variant proteins can be used to screen a compound for the ability to stimulate or inhibit interaction between the variant protein and a target molecule that normally interacts with the protein. The target can be a ligand, a substrate or a binding partner that the protein normally interacts with (for example, epinephrine or norepinephrine). Such assays typically include the steps of combining the variant protein with a candidate compound under conditions that allow the variant protein, or fragment thereof, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the variant protein and the target, such as any of the associated effects of signal transduction.

[0354] Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam et al., *Nature* 354:82-84 [1991]; Houghten et al., *Nature* 354:84-86 [1991]) and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang et al., *Cell* 72:767-778 [1993]); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab)<sub>2</sub>, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

[0355] One candidate compound is a soluble fragment of the variant protein that competes for ligand binding. Other candidate compounds include mutant proteins or appropriate fragments containing mutations that affect variant protein function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

[0356] The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) variant protein activity. The assays typically involve an assay of events in the signal transduction pathway that indicate protein activity. Thus, the expression of genes that are up or down-regulated in response to the variant protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the variant protein, or a variant protein target, could also be measured. Any of the biological or biochemical functions mediated by the variant protein can be used as an endpoint assay. These include all of the biochemical or biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art.

[0357] Binding and/or activating compounds can also be screened by using chimeric variant proteins in which an amino terminal extracellular domain or parts thereof, an entire transmembrane domain or subregions, and/or the carboxyl terminal intracellular domain or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate than that which is normally recognized by a variant protein. Accordingly, a different set of signal transduction components is available as an endpoint assay for activation. This allows for assays to be performed in other than the specific host cell from which the variant protein is derived.

[0358] The variant proteins are also useful in competition binding assays in methods designed to discover compounds that interact with the variant protein. Thus, a compound can be exposed to a variant protein under conditions that allow the compound to bind or to otherwise interact with the variant protein. A binding partner, such as ligand, that normally interacts with the variant protein is also added to the mixture. If the test compound interacts with the variant protein or its binding partner, it decreases the amount of complex formed or activity from the variant protein. This type of assay is particularly useful in screening for compounds that interact with specific regions of the variant protein (Hodgson, *Biotechnology*, 10[9], 973-80 [September 1992]).

[0359] To perform cell-free drug screening assays, it is sometimes desirable to immobilize either the variant protein or a fragment thereof, or its target molecule, to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Any method for immobilizing proteins on matrices can be used in drug screening assays. In one embodiment, a fusion protein containing an added domain allows the protein to be bound to a matrix. For example, glutathione-S-transferase/<sup>125</sup>I fusion proteins can be

adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., <sup>35</sup>S-labeled) and a candidate compound, such as a drug candidate, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads can be washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of bound material found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

[0360] Either the variant protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Alternatively, antibodies reactive with the variant protein but which do not interfere with binding of the variant protein to its target molecule can be derivatized to the wells of the plate, and the variant protein trapped in the wells by antibody conjugation. Preparations of the target molecule and a candidate compound are incubated in the variant protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the protein target molecule, or which are reactive with variant protein and compete with the target molecule, and enzyme-linked assays that rely on detecting an enzymatic activity associated with the target molecule.

[0361] Modulators of variant protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the protein pathway, such as Alzheimer's Disease. These methods of treatment typically include the steps of administering the modulators of protein activity in a pharmaceutical composition to a subject in need of such treatment.

[0362] The variant proteins, or fragments thereof, disclosed herein can themselves be directly used to treat a disorder characterized by an absence of, inappropriate, or unwanted expression or activity of the variant protein. Accordingly, methods for treatment include the use of a variant protein disclosed herein or fragments thereof.

[0363] In yet another aspect of the invention, variant proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al., *Cell* 72:223-232 [1993]; Madura et al., *J. Biol. Chem.* 268:12046-12054 [1993]; Bartel et al., *Biotechniques* 14:920-924 [1993]; Iwabuchi et al., *Oncogene* 8:1693-1696 [1993]; and Brent WO94/10300) to identify other proteins that bind to or interact with the variant protein and are involved in variant protein activity. Such variant protein-binding proteins are also likely to be involved in the propagation of signals by the variant proteins or variant protein targets as, for example, elements of a protein-mediated signaling pathway. Alternatively, such variant protein-binding proteins are inhibitors of the variant protein.

[0364] The two-hybrid system is based on the modular nature of most transcription factors, which typically consist of separable DNA-binding and activation domains. Briefly, the assay typically utilizes two different DNA constructs. In

one construct, the gene that codes for a variant protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming a variant protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected, and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein that interacts with the variant protein.

**[0365]** Antibodies Directed to Variant Proteins

**[0366]** The present invention also provides antibodies that selectively bind to the variant proteins disclosed herein and fragments thereof. Such antibodies may be used to quantitatively or qualitatively detect the variant proteins of the present invention. As used herein, an antibody selectively binds a target variant protein when it binds the variant protein and does not significantly bind to non-variant proteins, i.e., the antibody does not significantly bind to normal, wild-type, or art-known proteins that do not contain a variant amino acid sequence due to one or more SNPs of the present invention (variant amino acid sequences may be due to, for example, nonsynonymous cSNPs, nonsense SNPs that create a stop codon, thereby causing a truncation of a polypeptide or SNPs that cause read-through mutations resulting in an extension of a polypeptide).

**[0367]** As used herein, an antibody is defined in terms consistent with that recognized in the art: they are multi-subunit proteins produced by an organism in response to an antigen challenge. The antibodies of the present invention include both monoclonal antibodies and polyclonal antibodies, as well as antigen-reactive proteolytic fragments of such antibodies, such as Fab, F(ab)<sub>2</sub>, and Fv fragments. In addition, an antibody of the present invention further includes any of a variety of engineered antigen-binding molecules such as a chimeric antibody (U.S. Pat. Nos. 4,816,567 and 4,816,397; Morrison et al., *Proc. Natl. Acad. Sci. USA*, 81:6851 [1984]; Neuberger et al., *Nature* 312:604 [1984]), a humanized antibody (U.S. Pat. Nos. 5,693,762; 5,585,089; and 5,565,332), a single-chain Fv (U.S. Pat. No. 4,946,778; Ward et al., *Nature* 334:544 [1989]), a bispecific antibody with two binding specificities (Segal et al., *J. Immunol. Methods* 248:1 [2001]; Carter, *J. Immunol. Methods* 248:7 [2001]), a diabody, a triabody, and a tetrabody (Todorovska et al., *J. Immunol. Methods*, 248:47 [2001]), as well as a Fab conjugate (dimer or trimer), and a minibody.

**[0368]** Many methods are known in the art for generating and/or identifying antibodies to a given target antigen (Harlow, *Antibodies*, Cold Spring Harbor Press, New York [1989]). In general, an isolated peptide (e.g., a variant protein of the present invention) is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit, hamster or mouse. Either a full-length protein, an antigenic peptide fragment (e.g., a peptide fragment containing a region that varies between a variant protein and a

corresponding wild-type protein), or a fusion protein can be used. A protein used as an immunogen may be naturally-occurring, synthetic or recombinantly produced, and may be administered in combination with an adjuvant, including but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substance such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and the like.

**[0369]** Monoclonal antibodies can be produced by hybridoma technology (Kohler and Milstein, *Nature*, 256:495 [1975]), which immortalizes cells secreting a specific monoclonal antibody. The immortalized cell lines can be created in vitro by fusing two different cell types, typically lymphocytes, and tumor cells. The hybridoma cells may be cultivated in vitro or in vivo. Additionally, fully human antibodies can be generated by transgenic animals (He et al., *J. Immunol.*, 169:595 [2002]). Fd phage and Fd phagemid technologies may be used to generate and select recombinant antibodies in vitro (Hoogenboom and Chames, *Immunol. Today* 21:371 [2000]; Liu et al., *J. Mol. Biol.* 315:1063 [2002]). The complementarity-determining regions of an antibody can be identified, and synthetic peptides corresponding to such regions may be used to mediate antigen binding (U.S. Pat. No. 5,637,677).

**[0370]** Antibodies are preferably prepared against regions or discrete fragments of a variant protein containing a variant amino acid sequence as compared to the corresponding wild-type protein (e.g., a region of a variant protein that includes an amino acid encoded by a nonsynonymous cSNP, a region affected by truncation caused by a nonsense SNP that creates a stop codon, or a region resulting from the destruction of a stop codon due to read-through mutation caused by a SNP). Furthermore, preferred regions will include those involved in function/activity and/or protein/binding partner interaction. Such fragments can be selected on a physical property, such as fragments corresponding to regions that are located on the surface of the protein, e.g., hydrophilic regions, or can be selected based on sequence uniqueness, or based on the position of the variant amino acid residue(s) encoded by the SNPs provided by the present invention. An antigenic fragment will typically comprise at least about 8-10 contiguous amino acid residues in which at least one of the amino acid residues is an amino acid affected by a SNP disclosed herein. The antigenic peptide can comprise, however, at least 12, 14, 16, 20, 25, 50, 100 (or any other number in-between) or more amino acid residues, provided that at least one amino acid is affected by a SNP disclosed herein.

**[0371]** Detection of an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody or an antigen-reactive fragment thereof to a detectable substance. Detectable substances include, but are not limited to, various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent

material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

[0372] Antibodies, particularly the use of antibodies as therapeutic agents, are reviewed in: Morgan, "Antibody therapy for Alzheimer's Disease," *Expert Rev. Vaccines*. 2(1):53-9 (February 2003); Ross et al., "Anticancer antibodies," *Am. J. Clin. Pathol.* 119(4):472-85 (April 2003); Goldenberg, "Advancing role of radiolabeled antibodies in the therapy of cancer", *Cancer Immunol. Immunother.* 52(5):281-96 (May 2003); Epub Mar. 11 2003; Ross et al., "Antibody-based therapeutics in oncology," *Expert Rev. Anticancer Ther.* 3(1):107-21 (February 2003); Cao et al., "Bispecific antibody conjugates in therapeutics," *Adv. Drug Deliv. Rev.* 55(2):171-97 (Feb. 10 2003); von Mehren et al., "Monoclonal antibody therapy for cancer," *Annu. Rev. Med.* 54:343-69 (2003); Epub Dec. 3 2001; Hudson et al., "Engineered antibodies," *Nat. Med.* 9(1):129-34 (January 2003); Brekke et al., "Therapeutic antibodies for human diseases at the dawn of the twenty-first century," *Nat. Rev. Drug Discov.* 2(1):52-62 (January 2003) (Erratum in: *Nat. Rev. Drug Discov.* 2[3]:240 [March 2003]); Houdebine, "Antibody manufacture in transgenic animals and comparisons with other systems," *Curr. Opin. Biotechnol.* 13(6):625-9 (December 2002); Andreakos et al., "Monoclonal antibodies in immune and inflammatory diseases," *Curr. Opin. Biotechnol.* 13(6):615-20 (December 2002); Kellermann et al., "Antibody discovery: the use of transgenic mice to generate human monoclonal antibodies for therapeutics," *Curr. Opin. Biotechnol.* 13(6):593-7 (December 2002); Pini et al., "Phage display and colony filter screening for high-throughput selection of antibody libraries," *Comb. Chem. High Throughput Screen* 5(7):503-10 (November 2002); Batra et al., "Pharmacokinetics and biodistribution of genetically engineered antibodies," *Curr. Opin. Biotechnol.* 13(6):603-8 (December 2002); and Tangri et al., "Rationally engineered proteins or antibodies with absent or reduced immunogenicity," *Curr. Med. Chem.* 9(24):2191-9 (December 2002).

#### [0373] Uses of Antibodies

[0374] Antibodies can be used to isolate the variant proteins of the present invention from a natural cell source or from recombinant host cells by standard techniques, such as affinity chromatography or immunoprecipitation. In addition, antibodies are useful for detecting the presence of a variant protein of the present invention in cells or tissues to determine the pattern of expression of the variant protein among various tissues in an organism and over the course of normal development or disease progression. Further, antibodies can be used to detect variant protein in situ, in vitro, in a bodily fluid, or in a cell lysate or supernatant in order to evaluate the amount and pattern of expression. Also, antibodies can be used to assess abnormal tissue distribution, abnormal expression during development, or expression in an abnormal condition, such as Alzheimer's Disease. Additionally, antibody detection of circulating fragments of the full-length variant protein can be used to identify turnover.

[0375] Antibodies to the variant proteins of the present invention are also useful in pharmacogenomic analysis. Thus, antibodies against variant proteins encoded by alternative SNP alleles can be used to identify individuals that require modified treatment modalities.

[0376] Further, antibodies can be used to assess expression of the variant protein in disease states such as in active stages of the disease or in an individual with a predisposition to a disease related to the protein's function, particularly Alzheimer's Disease. Antibodies specific for a variant protein encoded by a SNP-containing nucleic acid molecule of the present invention can be used to assay for the presence of the variant protein, such as to screen for predisposition to Alzheimer's Disease as indicated by the presence of the variant protein.

[0377] Antibodies are also useful as diagnostic tools for evaluating the variant proteins in conjunction with analysis by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays well known in the art.

[0378] Antibodies are also useful for tissue typing. Thus, where a specific variant protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

[0379] Antibodies can also be used to assess aberrant subcellular localization of a variant protein in cells in various tissues. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting the expression level or the presence of variant protein or aberrant tissue distribution or developmental expression of a variant protein, antibodies directed against the variant protein or relevant fragments can be used to monitor therapeutic efficacy.

[0380] The antibodies are also useful for inhibiting variant protein function, for example, by blocking the binding of a variant protein to a binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting a variant protein's function. An antibody can be used, for example, to block or competitively inhibit binding, thus modulating (agonizing or antagonizing) the activity of a variant protein. Antibodies can be prepared against specific variant protein fragments containing sites required for function or against an intact variant protein that is associated with a cell or cell membrane. For in vivo administration, an antibody may be linked with an additional therapeutic payload such as a radionuclide, an enzyme, an immunogenic epitope, or a cytotoxic agent. Suitable cytotoxic agents include, but are not limited to, bacterial toxin such as diphtheria, and plant toxin such as ricin. The in vivo half-life of an antibody or a fragment thereof may be lengthened by pegylation through conjugation to polyethylene glycol (Leong et al., *Cytokine* 16:106 [2001]).

[0381] The invention also encompasses kits for using antibodies, such as kits for detecting the presence of a variant protein in a test sample. An exemplary kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting variant proteins in a biological sample; means for determining the amount, or presence/absence of variant protein in the sample; means for comparing the amount of variant protein in the sample with a standard; and instructions for use.

#### [0382] Vectors and Host Cells

[0383] The present invention also provides vectors containing the SNP-containing nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport a

SNP-containing nucleic acid molecule. When the vector is a nucleic acid molecule, the SNP-containing nucleic acid molecule can be covalently linked to the vector nucleic acid. Such vectors include, but are not limited to, a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, or MAC.

[0384] A vector can be maintained in a host cell as an extrachromosomal element where it replicates and produces additional copies of the SNP-containing nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the SNP-containing nucleic acid molecules when the host cell replicates.

[0385] The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the SNP-containing nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

[0386] Expression vectors typically contain cis-acting regulatory regions that are operably linked in the vector to the SNP-containing nucleic acid molecules such that transcription of the SNP-containing nucleic acid molecules is allowed in a host cell. The SNP-containing nucleic acid molecules can also be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the SNP-containing nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

[0387] The regulatory sequences to which the SNP-containing nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage  $\lambda$ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

[0388] In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

[0389] In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region, a ribosome-binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. A person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors (see, e.g., Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [2000]).

[0390] A variety of expression vectors can be used to express a SNP-containing nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors can also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g., cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2000).

[0391] The regulatory sequence in a vector may provide constitutive expression in one or more host cells (e.g., tissue specific expression) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor, e.g., a hormone or other ligand. A variety of vectors that provide constitutive or inducible expression of a nucleic acid sequence in prokaryotic and eukaryotic host cells are well known to those of ordinary skill in the art.

[0392] A SNP-containing nucleic acid molecule can be inserted into the vector by methodology well-known in the art. Generally, the SNP-containing nucleic acid molecule that will ultimately be expressed is joined to an expression vector by cleaving the SNP-containing nucleic acid molecule and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

[0393] The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial host cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic host cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

[0394] As described herein, it may be desirable to express the variant peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the variant peptides. Fusion vectors can, for example, increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting, for example, as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired variant peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes suitable for such use include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith et al., *Gene* 67:3140 [1988]), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., *Gene* 69:301-315 [1988]) and pET 11d

(Studier et al., *Gene Expression Technology: Methods in Enzymology* 185:60-89 [1990]).

[0395] Recombinant protein expression can be maximized in a bacterial host by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein (S. Gottesman, *Gene Expression Technology: Methods in Enzymology* 185 119-128, Academic Press, San Diego, Calif. [1990]). Alternatively, the sequence of the SNP-containing nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example, *E. coli* (Wada et al., *Nucleic Acids Res.* 20:2111-2118 [1992]).

[0396] The SNP-containing nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast (e.g., *S. cerevisiae*) include pYepSec1 (Baldari, et al., *EMBO J.* 6:229-234 [1987]), pMFa (Kurjan et al., *Cell* 30:933-943 [1982]), pJRY88 (Schultz et al., *Gene* 54:113-123 [1987]), and pYES2 (Invitrogen Corporation, San Diego, Calif.).

[0397] The SNP-containing nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf9 cells) include the pAc series (Smith et al., *Mol. Cell. Biol.* 3:2156-2165 [1983]) and the pVL series (Lucklow et al., *Virology* 170:31-39 [1989]).

[0398] In certain embodiments of the invention, the SNP-containing nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (B. Seed, *Nature* 329:840 [1987]) and pMT2PC (Kaufman et al., *EMBO J.* 6:187-195 [1987]).

[0399] The invention also encompasses vectors in which the SNP-containing nucleic acid molecules described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to the SNP-containing nucleic acid sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

[0400] The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include, for example, prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

[0401] The recombinant host cells can be prepared by introducing the vector constructs described herein into the cells by techniques readily available to persons of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those described in Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2000).

[0402] Host cells can contain more than one vector. Thus, different SNP-containing nucleotide sequences can be intro-

duced in different vectors into the same cell. Similarly, the SNP-containing nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the SNP-containing nucleic acid molecules, such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced, or joined to the nucleic acid molecule vector.

[0403] In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication can occur in host cells that provide functions that complement the defects.

[0404] Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be inserted in the same vector that contains the SNP-containing nucleic acid molecules described herein or may be in a separate vector. Markers include, for example, tetracycline or ampicillin-resistance genes for prokaryotic host cells, and dihydrofolate reductase or neomycin resistance genes for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait can be effective.

[0405] While the mature variant proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these variant proteins using RNA derived from the DNA constructs described herein.

[0406] Where secretion of the variant protein is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as G-protein-coupled receptors (GPCRs), appropriate secretion signals can be incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

[0407] Where the variant protein is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze/thaw, sonication, mechanical disruption, use of lysing agents, and the like. The variant protein can then be recovered and purified by well-known purification methods including, for example, ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

[0408] It is also understood that, depending upon the host cell in which recombinant production of the variant proteins described herein occurs, they can have various glycosylation patterns, or may be non-glycosylated, as when produced in bacteria. In addition, the variant proteins may include an initial modified methionine in some cases as a result of a host-mediated process.

[0409] For further information regarding vectors and host cells, see *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.

[0410] Uses of Vectors and Host Cells, and Transgenic Animals

[0411] Recombinant host cells that express the variant proteins described herein have a variety of uses. For example, the cells are useful for producing a variant protein that can be further purified into a preparation of desired amounts of the variant protein or fragments thereof. Thus, host cells containing expression vectors are useful for variant protein production.

[0412] Host cells are also useful for conducting cell-based assays involving the variant protein or variant protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a variant protein is useful for assaying compounds that stimulate or inhibit variant protein function. Such an ability of a compound to modulate variant protein function may not be apparent from assays of the compound on the native/wild-type protein, or from cell-free assays of the compound. Recombinant host cells are also useful for assaying functional alterations in the variant proteins as compared with a known function.

[0413] Genetically-engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a non-human mammal, for example, a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA containing a SNP of the present invention which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more of its cell types or tissues. Such animals are useful for studying the function of a variant protein in vivo, and identifying and evaluating modulators of variant protein activity. Other examples of transgenic animals include, but are not limited to, non-human primates, sheep, dogs, cows, goats, chickens, and amphibians. Transgenic non-human mammals such as cows and goats can be used to produce variant proteins which can be secreted in the animal's milk and then recovered.

[0414] A transgenic animal can be produced by introducing a SNP-containing nucleic acid molecule into the male pronuclei of a fertilized oocyte, e.g., by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any nucleic acid molecules that contain one or more SNPs of the present invention can potentially be introduced as a transgene into the genome of a non-human animal.

[0415] Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the variant protein in particular cells or tissues.

[0416] Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described in, for example, U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al., and in B. Hogan, *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1986). Similar methods are used for production of other transgenic animals. A transgenic founder ani-

mal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes a non-human animal in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

[0417] In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1 (Lakso et al. *PNAS* 89:6232-6236 [1992]). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman et al., *Science* 251:1351-1355 [1991]). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are generally needed. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected variant protein and the other containing a transgene encoding a recombinase.

[0418] Clones of the non-human transgenic animals described herein can also be produced according to the methods described in, for example, I. Wilmut et al., *Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell (e.g., a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell (e.g., a somatic cell) is isolated.

[0419] Transgenic animals containing recombinant cells that express the variant proteins described herein are useful for conducting the assays described herein in an in vivo context. Accordingly, the various physiological factors that are present in vivo and that could influence ligand or substrate binding, variant protein activation, signal transduction, or other processes or interactions, may not be evident from in vitro cell-free or cell-based assays. Thus, non-human transgenic animals of the present invention may be used to assay in vivo variant protein function as well as the activities of a therapeutic agent or compound that modulates variant protein function/activity or expression. Such animals are also suitable for assessing the effects of null mutations (i.e., mutations that substantially or completely eliminate one or more variant protein functions).

[0420] For further information regarding transgenic animals, see Houdebine, "Antibody manufacture in transgenic animals and comparisons with other systems," *Curr. Opin. Biotechnol.* 13(6):625-9 (December 2002); Petters et al., "Transgenic animals as models for human disease," *Transgenic Res.* 9(4-5):347-51 (2000); discussion 345-6; Wolf et

al., "Use of transgenic animals in understanding molecular mechanisms of toxicity," *J. Pharm. Pharmacol.* 50(6):567-74 (June 1998); Echelard, "Recombinant protein production in transgenic animals," *Curr. Opin. Biotechnol.* 7(5):536-40 (October 1996); Houdebine, "Transgenic animal bioreactors," *Transgenic Res.* 9(4-5):305-20 (2000); Purity et al., "Embryonic stem cells, creating transgenic animals," *Methods Cell Biol.* 57:279-93 (1998); and Robl et al., "Artificial chromosome vectors and expression of complex proteins in transgenic animals," *Theriogenology* 59(1):107-13 (Jan. 1 2003).

#### Computer-Related Embodiments

[0421] The SNPs provided in the present invention may be "provided" in a variety of mediums to facilitate use thereof. As used in this section, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, that contains SNP information of the present invention. Such a manufacture provides the SNP information in a form that allows a skilled artisan to examine the manufacture using means not directly applicable to examining the SNPs or a subset thereof as they exist in nature or in purified form. The SNP information that may be provided in such a form includes any of the SNP information provided by the present invention such as, for example, polymorphic nucleic acid and/or amino acid sequence information such as SEQ ID NOS:1-5, SEQ ID NOS:6-10, SEQ ID NOS:16-19, SEQ ID NOS:11-15, and SEQ ID NOS:20-31; information about observed SNP alleles, alternative codons, populations, allele frequencies, SNP types, and/or affected proteins; or any other information provided by the present invention in Tables 1-2 and/or the Sequence Listing.

[0422] In one application of this embodiment, the SNPs of the present invention can be recorded on a computer readable medium. As used herein, "computer readable medium" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable media can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. One such medium is provided with the present application, namely, the present application contains computer readable medium (CD-R) that has nucleic acid sequences (and encoded protein sequences) containing SNPs provided/recorded thereon in ASCII text format in a Sequence Listing along with accompanying Tables that contain detailed SNP and sequence information (transcript sequences are provided as SEQ ID NOS:1-5, protein sequences are provided as SEQ ID NOS:6-10, genomic sequences are provided as SEQ ID NOS:16-19, transcript-based context sequences are provided as SEQ ID NOS:11-15, and genomic-based context sequences are provided as SEQ ID NOS:20-31).

[0423] As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the SNP information of the present invention.

[0424] A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide/amino acid sequence information of the present invention on computer readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, represented in the form of an ASCII file, or stored in a database application, such as OB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the SNP information of the present invention.

[0425] By providing the SNPs of the present invention in computer readable form, a skilled artisan can routinely access the SNP information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Examples of publicly available computer software include BLAST (Altschul et al., *J. Mol. Biol.* 215:403-410 [1990]) and BLAZE (Brutlag et al., *Comp. Chem.* 17:203-207 [1993]) search algorithms.

[0426] The present invention further provides systems, particularly computer-based systems, which contain the SNP information described herein. Such systems may be designed to store and/or analyze information on, for example, a large number of SNP positions, or information on SNP genotypes from a large number of individuals. The SNP information of the present invention represents a valuable information source. The SNP information of the present invention stored/analyzed in a computer-based system may be used for such computer-intensive applications as determining or analyzing SNP allele frequencies in a population, mapping disease genes, genotype-phenotype association studies, grouping SNPs into haplotypes, correlating SNP haplotypes with response to particular drugs, or for various other bioinformatic, pharmacogenomic, drug development, or human identification/forensic applications.

[0427] As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the SNP information of the present invention. The minimum hardware means of the computer-based systems of the present invention typically comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. Such a system can be changed into a system of the present invention by utilizing the SNP information provided on the CD-R, or a subset thereof, without any experimentation.

[0428] As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein SNPs of the present invention and the neces-

sary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store SNP information of the present invention, or a memory access means which can access manufactures having recorded thereon the SNP information of the present invention.

[0429] As used herein, "search means" refers to one or more programs or algorithms that are implemented on the computer-based system to identify or analyze SNPs in a target sequence based on the SNP information stored within the data storage means. Search means can be used to determine which nucleotide is present at a particular SNP position in the target sequence. As used herein, a "target sequence" can be any DNA sequence containing the SNP position(s) to be searched or queried.

[0430] As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences containing a SNP position in which the sequence(s) is chosen based on a three-dimensional configuration that is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures, and inducible expression elements (protein binding sequences).

[0431] A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. An exemplary format for an output means is a display that depicts the presence or absence of specified nucleotides (alleles) at particular SNP positions of interest. Such presentation can provide a rapid, binary scoring system for many SNPs simultaneously.

[0432] One exemplary embodiment of a computer-based system comprising SNP information of the present invention is provided in FIG. 1. FIG. 1 provides a block diagram of a computer system 102 that can be used to implement the present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable storage medium 116 once inserted in the removable medium storage device 114.

[0433] The SNP information of the present invention may be stored in a well-known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. Software for accessing and processing the SNP information (such as SNP scoring tools, search tools, comparing tools, etc.) preferably resides in main memory 108 during execution.

#### EXAMPLE ONE

##### Statistical Analysis of SNPs Associated with Alzheimer's Disease

[0434] The following example is offered to illustrate, but not limit the claimed invention.

[0435] Results from previous whole-genome screens have identified several chromosomal regions that show linkage with late-onset Alzheimer's disease (LOAD) (M. A. Pericak-Vance et al., *Exp. Gerontol.* 35[9-10]:1343-1352 [2000]; D. Curtis et al., *Ann. Hum. Genet.* 65[Pt. 5]:473-481 [2001]; A. Myers et al., *Am. J. Med. Genet.* 114[2]:235-244 [2002]; J. M. Olson et al., *Am. J. Hum. Genet.* 71[1]:154-161 [2002]; D. Blacker et al., *Hum. Mol. Genet.* 12[1]:23-32 [2003]). To identify genetic variants and genes associated with LOAD on chromosome 9, a scan was done for single nucleotide polymorphisms (SNPs) across the entire chromosome using DNA samples collected from LOAD patients matched with non-demented individuals. Statistical analysis of frequencies of a particular SNP or a combination of SNPs (haplotypes) was done to identify disease-associated variants. To reduce the likelihood of identifying spurious associations, three independently collected LOAD case-control sample sets were employed, and genotyping analyses were carried out in two phases.

##### [0436] Study Design

[0437] Three late-onset Alzheimer's disease case-control sample sets, collected with informed consent/assent from the participating individuals and approvals from the participating institutions, were used in this study. Cases were diagnosed of dementia with the Alzheimer's type according to National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (G. McKhann et al., "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology* 34, 939-44 (1984)), with a minimum age of disease onset of 65 years. Controls were matched cognitively healthy individuals, ascertained through neuropsychological tests and clinical interviews at the age of 65 years or older. The three sample sets are the WU sample set, obtained through the Washington University Alzheimer's Disease Research Center (ADRC) patient registry, the UCSD sample set, obtained from the ADRC of the University of California, San Diego, and the UK sample set, obtained from the Medical Research Council Late Onset AD Genetic Resource that included those from Cardiff University, Wales School of Medicine and from King's College London. The number of cases and controls in each sample set can be found in Table 5 and other tables. These sample sets have been used in our other recent studies and detailed phenotypic information can be found therein (Li, Y. et al. "Association of late-onset Alzheimer's disease with genetic variation in multiple members of the GAPD gene family," *Proc Natl Acad Sci USA* 101, 15688-93 (2004); Li, Y. et al., "Genetic association of the APP binding protein 2 gene (APBB2) with late onset Alzheimer disease," *Hum Mutat* 25, 270-277 (2005)). They all show an expected APOE  $\epsilon$ 4-genotype distribution, and none appears to have evidence of population stratification.

[0438] Genotyping of SNPs was performed by allele-specific real time PCR for individual samples using primers

designed and validated in-house (Germer, S., Holland, M. J. & Higuchi, R. "High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR." *Genome Res* 10, 258-66 (2000)). Overall, the accuracy of the genotyping was better than 99%, as determined by internal comparisons of differentially designed assays for the same marker and comparisons for the same marker across different groups.

[0439] Hardy-Weinberg equilibrium for genotyping was evaluated using an exact test as described by Weir ("Genetic Data Analysis II", Sinauer Associates, Sunderland Mass., 1996, 2nd edition). Tests for allelic association of SNPs with disease status were carried out using the  $\chi^2$  test. Markers with minor allele frequency (MAF) of  $\geq 2\%$  in either cases or controls were analyzed.

[0440] When combining data from different sample sets, association was assessed by meta analysis using the method of Cochran Mantel and Haenzel using sample set as the stratifying variable (Agresti, A. *Categorical Data Analysis*, (John Wiley & Sons, 1990).

[0441] Estimation of haplotype frequencies and tests of association between haplotype and disease status were performed using the R package haplo.stats version 1.1.1 (see <http://cran.us.r-project.org>). This package estimates haplotype frequencies using an EM algorithm with progressive insertion based on the algorithm implemented in the software "snphap" by David Clayton (<http://wwwgene.cimr.cam.ac.uk/clayton/software>). The "global" and haplotype specific tests of association between haplotypes and disease status are computed using a score test as described by Schaid et al. (Schaid, D. J., Rowland, C. M., Tines, D. E., Jacobson, R. M. & Poland, G. A. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70, 425-34 (2002)).

[0442] DNA and total RNA were extracted from peripheral blood mononuclear cells of normal blood donors. A cDNA library was prepared from total RNA using random hexamers and MultiScribe reverse transcriptase (ABI). For DAPK1, two high frequency exonic SNPs, hCV2704861 or hCV2704931, were used to measure allele-specific gene expression. High frequency markers were selected to increase the number of heterozygotes. Allele-specific expression assays were carried out on cDNA samples with the same real time PCR condition as described for genotyping. The same primers were used for genotyping and allele-specific expression assays.

[0443] For the examination of allele-specific expression, 92 individual samples of genomic DNA were first genotyped for two expression markers (hCV2704861 or hCV2704931). 69 donors of Caucasian descent, heterozygous for hCV2704861 or hCV2704931, were then examined for allele-specific gene expression. cDNA was arrayed in quadruplicates or duplicates onto 384-well plates, together with appropriate PCR controls, and were run on an ABI-7900 real time PCR system under standard conditions. Genomic DNA was also arrayed onto the same plate as a control. The relative expression of both alleles for each expression marker was determined by subtracting the smaller Ct value of one allele PCR reaction from the larger Ct value of another allele PCR reaction (dCt). The ratio of two allele-specific transcripts was calculated as  $2^{-dCt}$  (i.e. a one-cycle difference in our real-time PCR based assay results in a

2-fold relative difference). For the statistical analysis, dCt values were obtained as an average of 2 to 4 reactions for each sample and data point (standard error=0.11 for each assay, averaged across all samples). As a control, the dCt values were also obtained with heterozygous genomic DNA, which theoretically should equal zero (actual:  $-0.05$  [ $\pm 0.13$ ] for hCV2704861 and  $-0.02$  [ $\pm 0.13$ ] for hCV2704931; average dCt [ $\pm$ Stdev]).

[0444] For testing the relationship between disease-associated variants and allele-specific expression, the individuals who are heterozygous of the expression markers were then genotyped for the two DAPK1 disease-associated markers. The allele-specific expression level measured by the expression marker was then stratified by the genotype status of the disease-associated markers (homozygote or heterozygote). The Mann-Whitney test was used to assess whether hetero- and homozygosity of the LOAD-associated SNPs are significantly associated with the allele-specific expression ratio. As a control, the same test was performed for the experimentally determined genotype ratio in heterozygous genomic DNA ( $P=0.45-0.95$ ). Statistical significance was calculated separately for each combination of expression marker v.s. disease-associated marker.

[0445] Results

[0446] In the initial "discovery" phase, 674 SNPs were genotyped and analyzed in one sample set (FIG. 2). These SNPs are located in 347 genes or other regions across the chromosome. Forty-seven of them were significantly associated with LOAD in a  $\chi^2$ -test for allelic association ( $p$  Value $<0.05$ ) (FIG. 2). In the second, or "validation" phase, 47 of these markers were genotyped in two other sample sets. Three markers were replicated in the combined validation sample sets (allelic  $p$  Value $<0.05$ , 1 sided). One marker, hCV1386982, was significantly associated with LOAD in each of the three sample sets (Table 5) and showed remarkable consistency in allele frequency within cases and controls among the three sample sets. A meta-analysis of all three sample sets showed an allelic  $P$  value of 0.0006 and an odds ratio (OR) of 0.79 (95% confidence interval [CI]: 0.69-0.90), suggesting a protective nature of the minor allele. Under a recessive model the effect is stronger in minor allele homozygotes compared with major allele homozygotes ( $p$  Value $<0.05$  in three studies), while heterozygous subjects were not at significantly different risk than subjects homozygous for the major allele (FIG. 3). Marker hCV1386982 is located in the DAPK1 gene, which is in the center of the previously identified linkage peak reported by D. Blacker et al. ("Results of a high-resolution genome screen of 437 Alzheimer's Disease families," *Hum. Mol. Genet.* 12, 23-32 [2003]). DAPK1 encodes a pro-apoptotic death-associated protein kinase, and is thus an excellent biological candidate gene for AD.

[0447] Another marker, hCV8715115, in POMT1 encoding protein-O-mannosyltransferase 1, was significant in one of the validation sample sets and the meta analysis of the combined validation samples. However the allele frequency of the marker is relatively low, with 3.1% in controls vs 5.3% in cases. The third marker, hCV1920609, in DFNB31 was not significant in either of the two validation sample sets individually but reached significance when the two were combined. Defects in DFNB31 cause hereditary non-syndromic recessive hearing loss. The POMT1 and DFNB31

markers are 44.2 Mbp and 27.0 Mbp distal from the DAPK1 marker, respectively, and share little LD among them. Thus, DAPK1 became the most outstanding finding from this chromosome scan.

**[0448]** Next the linkage disequilibrium (LD) structure of the region containing the DAPK1 SNP hCV1386982 was examined in HapMap (HapMap public release #16c.1; see <http://www.hapmap.org>). Marker hCV1386982 is within a 58.5 kbp region of high LD from rs913778 to rs888333, encompassed only in the 210.7 Kbp DAPK1 gene without other known or predicted genes. To fine-map this region, 14 tagging SNPs with  $r^2 < 0.8$  were identified. One of the tagging SNPs was hCV1386982 itself. The remaining 12 tagging SNPs were genotyped in the WU sample set (see Table 5 for number of samples in each set). One tagging marker, hCV1386888, with  $r^2$  of 0.64 and 37.3 Kbp away from hCV1386982, was significantly associated with LOAD (p Value=0.020). Haplotype analysis with a three-SNP sliding window method identified a significant haplotype composed of hCV1386973, hCV1386978 and hCV1386982 (Table 6; best individual haplotype AAT p Value=0.0025, global haplotype p Value=0.046). Both the significant individual marker and the significant haplotype markers were then typed in the UK and UCSD samples. Marker hCV1386888 was replicated in the combined analysis of the validation sample sets. The marker was significant in the UCSD sample set and trended to significance in the UK sample set (Table 5). A meta analysis of the marker in all three sample sets showed a significant association (p Value=0.0017). The significant haplotype composed of hCV1386973, hCV1386978 and hCV1386982 was replicated in the combined validation sample sets as well (p Value for the AAT haplotype was 0.0041, global p Value=0.056; Table 6).

**[0449]** Because of the intronic nature of the significant SNPs, we sought to determine whether the disease-associated variants have a direct effect (cis) on DAPK1 transcript levels. This information can be obtained directly by measuring the relative expression level of the allele 1 specific transcript and the allele 2 specific transcript. Because this method measures both allele specific transcripts in the same sample, the measurement is much less impacted by biological variation that occurs when comparing expression levels between different samples. In addition, this method does not require normalization to a housekeeping gene because the transcript level in each sample should be identical for allele 1 and allele 2, unless there is a cis-controlling factor that is present on only one allele/haplotype and changes its relative expression level (e.g. a mutation in a repressor element on allele 1/haplotype 1 leads to an expression ratio that is higher for allele 1 than allele 2). Therefore, allele specific expression analysis is much more sensitive to small relative changes in expression levels, while across sample comparisons include much more noise (biological variation between samples and technical noise from normalizing to a housekeeping gene) and can only detect larger expression differences. We thus tested DAPK1 for allelic expression differences and evaluated whether hCV1386982 or hCV1386888 is associated with the allelic expression differences. The disease-associated markers could not be used directly for this analysis because they are located within a DAPK1 intron; therefore, we used a two-step approach. First, we genotyped two high frequency SNPs that map to the DAPK1 transcript and identified 69 Caucasians, who were heterozygous for at least one of these expression markers (46 for

hCV2704861 and 48 for hCV2704931; minor allele frequency: 0.44 for hCV2704861 and 0.47 for hCV2704931). The differences in the two allele-specific transcripts were as large as ~2 fold.

**[0450]** To determine whether the two disease-associated SNPs, hCV1386982 and hCV1386888, are associated with the observed DAPK1 allele-specific expression, we next genotyped these two SNPs in all 69 heterozygous carriers of hCV2704861 and/or hCV2704931. The two expression SNPs are 72 and 129 Kbp away from hCV1386982, the closer of the two disease-associated markers, and do not reside in the high LD region shared by the two disease-associated markers. If either hCV1386982 or hCV1386888 is the sole causative element, it would be expected that only heterozygous carriers of the disease-associated variant show an allele-specific expression difference while homozygous carriers would not (FIG. 4A). As shown in FIG. 4B, both homozygous and heterozygous carriers of hCV1386982 and hCV1386888 showed allele specific differences in expression. However, the allele specific expression ratio was significantly higher in heterozygous carriers of either hCV1386982 or hCV1386888 compared to homozygous carriers ( $P < 0.05$ , FIG. 4B), while the genomic DNA control showed no such genotype dependent difference. These results indicate that the genotype status of both LOAD-associated SNPs is significantly associated with DAPK1 allele-specific expression.

#### **[0451]** Discussion

**[0452]** The above results show that two SNPs and a haplotype in DAPK1 were found to be significantly associated with LOAD. Marker hCV1386982 was consistently associated with LOAD in three case-control sample sets (allelic p Value < 0.05) that do not show evidence of population stratification. In the three-sample meta analysis this polymorphism confers an allelic OR of 0.79 (95% CI: 0.69 to 0.90), with a stronger effect under a recessive model (OR=0.55; 95% CI: 0.40 to 0.74). The DAPK1 allele identified here is relatively common, with frequencies over 30% in cases or controls, which is consistent with the common disease-common variant hypothesis. Another SNP, hCV1386888, was significantly associated with LOAD in two of the three sample sets and trended to significance in the third sample set. A logistic regression analysis of the two SNPs showed that neither SNP remains significant when adjusted for the other, suggesting that association of both polymorphisms may be due to their relatively high LD with each other and/or another causative SNP. In addition, a haplotype consisting of three SNPs including hCV1386982 showed association with LOAD that was comparable to the association of hCV1386982 alone. It should be noted that these significant markers are located in a region of high LD that includes only DAPK1 and no other known or predicted genes, therefore supporting a role for DAPK1 in the genetics of LOAD.

**[0453]** Our functional study revealed that DAPK1 expression shows allelic imbalance, a phenomenon (Pastinen, T. and Hudson, T. J. Cis-acting regulatory variation in the human genome. *Science* 306, 647-50 (2004)) that is seen for genetic risk factors in other complex diseases such as calpain 10 and type 2 diabetes (Horikawa, Y., Oda, N., Cox, N. J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T. H., Mashima, H., Schwarz, P. E. et al. Genetic

variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.* 26, 163-75 (2000)) and may explain the genetic association of DAPK1 with LOAD. In addition, our study shows that the genotypes of the two intronic LOAD-associated SNPs are significantly associated with DAPK1 allele-specific expression. This association suggests that these SNPs may interact with other unidentified polymorphic cis-acting regulatory factors to influence the level of DAPK1 transcripts. We cannot exclude the possibility that they are in LD with other polymorphic cis-acting elements governing DAPK1 transcription. Regardless of the molecular mechanism of this regulation, considering that DAPK1 allele-specific expression predicts variation in DAPK1 protein/activity and thus neuronal apoptotic potential, allele-specific expression of DAPK1 variants provide a plausible explanation linking the genetic association with LOAD to a disease-relevant functional outcome. This finding can be applied directly to the diagnosis of AD where disease predisposition, confirmation of clinical diagnosis, or disease progression may be assessed based on DAPK1 genotypes. Genotyping can be done by standard methods such as sequencing or other detection methods such as allele-specific real-time PCR with allele specific primers and a common primer that distinguish the genotypes. In addition, as an enzyme whose activity is involved in neuronal cell apoptosis, DAPK1 represents a novel drug target for the treatment of AD and other neurological pathologies. Furthermore, treatment decisions for AD and other diseases may be made based on specific DAPK1 genotype of a patient.

**[0454]** A genotyping kit for the detection of various SNPs including those in DAPK1 may be designed for the diagnosis of AD and treatment response in the management of AD and other neurological pathologies.

**[0455]** Novel medicines including, but not limited to, small molecules, proteins, protein fragments or peptides, antibodies, or nucleotide acids, may be designed to target DAPK1 for the treatment of AD and other neurological pathologies.

#### EXAMPLE TWO

##### Additional SNPs in LD with Alzheimer's Disease-Associated Interrogated SNP Markers

**[0456]** An investigation was conducted to identify SNP markers in linkage disequilibrium (LD) with SNPs which have been found to be associated with Alzheimer's Disease, as shown in Tables 3-6. Briefly, the power threshold (T) was set at 51% for detecting disease association using LD markers. This power threshold is based on equation (31) above, which incorporates allele frequency data from previous disease association studies, the predicted error rate for not detecting truly disease-associated markers, and a significance level of 0.05. Using this power calculation and the sample size, for each interrogated SNP (Table 7) a threshold level of LD, or  $r^2$  value, was derived ( $r_T^2$ , equations (32) and (33)). The threshold value  $r_T^2$  is the minimum value of linkage disequilibrium between the interrogated SNP and its LD SNPs possible such that the non-interrogated SNP still retains a power greater or equal to T for detecting disease-association.

**[0457]** Based on the above methodology, LD SNPs were found for all interrogated SNPs shown in Tables 3-6. LD

SNPs are listed in Table 7, each associated with its respective interrogated SNP. Also shown are the public SNP IDs (rs numbers) for interrogated and LD SNPs, the threshold  $r^2$  value and the power used to determine this, and the  $r^2$  value of linkage disequilibrium between the interrogated SNP and its matching LD SNP. As an example in Table 7, Alzheimer's Disease-associated SNP hCV1920609 was calculated to be in LD with hCV939509 at a  $r_T^2$  value of 0.89, based on a 51% power calculation, thus making hCV939509 a marker which is also associated with Alzheimer's Disease because the value of  $r^2$  between the two SNPs is 1, larger than  $r_T^2$ .

**[0458]** All publications and patents cited in this specification are herein incorporated by reference in their entirety. Various modifications and variations of the described compositions, methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments and certain working examples, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention that are obvious to those skilled in the field of molecular biology, genetics and related fields are intended to be within the scope of the following claims.

TABLE 1

Transcript SNP info and associated gene/protein information	
Gene Number:	1
Celera Gene:	hCG29235 - 83000099300842
Celera Transcript:	hCT1950602 - 83000099300914
Public Transcript	
Accession:	
Celera Protein:	hCP1766647 - 197000064928072
Public Protein	
Accession:	
Gene Symbol:	POMT1
Protein Name:	protein-O-mannosyltransferase 1
Celera Genomic Axis:	GA_x5YUV32W1V9 (18396310 . . . 18438527)
Chromosome:	9
OMIM NUMBER:	607423
OMIM Information:	Walker-Warburg syndrome, 236670 (3); Muscular dystrophy, limb-girdle,/type 2K, 609308 (3)
Transcript Sequence (SEQ ID NO: 1):	
Protein Sequence (SEQ ID NO: 6):	
<u>SNP Information</u>	
Context (SEQ ID NO: 11):	CTGGGCAGCCTCGCCTCTTGGCCTCTGCAGGTGCCTCTGTATGGGAGGCC AGAGTTTCTGTCACTAACTTTTCTAAGCTCACAAATGCTAGAGGTGGGT RCGCTTTTCCACGCAGTGAACATGACTTTTCTTTGAATCTCTGGCAGGT CTGTGTGTCTGTCACTTGTCTCGCCGAGCAGTGGCTTTGCTGGTCATCC
Celera SNP ID:	hCV8715115
Public SNP ID:	rs2018621
SNP in Transcript	SEQ ID NO: 1
Sequence	
SNP Position	1094
Transcript:	

TABLE 1-continued

Transcript SNP info and associated gene/protein information	
SNP Source:	dbSNP; HapMap; ABI_Val; HGBASE
Population(A allele, Count):	caucasian(A, 9 G, 111)
SNP Type:	UTR5
Gene Number:	1
Celera Gene:	hCG29235 - 83000099300842
Celera Transcript:	hCT1959916 - 83000099300937
Public Transcript	
Accession:	
Celera Protein:	hCP1770695 - 197000064928073
Public Protein	
Accession:	
Gene Symbol:	POMT1
Protein Name:	protein-O-mannosyltransferase 1
Celera Genomic Axis:	GA_x5YUV32W1V9 (18396310 . . . 18438527)
Chromosome:	9
OMIM NUMBER:	607423
OMIM Information:	Walker-Warburg syndrome, 236670 (3); Muscular dystrophy, limb-girdle,/type 2K, 609308 (3)
Transcript Sequence (SEQ ID NO: 2):	
Protein Sequence (SEQ ID NO: 7):	
<u>SNP Information</u>	
Context (SEQ ID NO: 12): CTGGGCAGCCTCGCCTCTTGGCCTCGCAGGTGCCTCTGTATGGGAGGCC AGAGTTTCTGTCACCTAACTTTTCTAAGCTCACAAATGCTAGAGTGGGT RCGCTTTCCACGCAGTGAACATGACTTTCTTTGAATCTCTGGCAGGT CTGTGTGTTCTGTCACCTGCTCGCCGAGCAGTGGCTTTGCTGGTCAATC C	
Celera SNP ID:	hCV8715115
Public SNP ID:	rs2018621
SNP in Transcript Sequence	SEQ ID NO: 2
SNP Position	1019
Transcript:	
SNP Source:	dbSNP; HapMap; ABI_Val; HGBASE
Population(A allele, Count):	caucasian(A, 9 G, 111)
SNP Type:	UTR5
Gene Number:	2
Celera Gene:	hCG32519 - 84000314391801
Celera Transcript:	hCT1965717 - 84000314391824
Public Transcript	
Accession:	
Celera Protein:	hCP1781822 - 197000064927550
Public Protein	
Accession:	
Gene Symbol:	DFNB31
Protein Name:	deafness, autosomal recessive 31
Celera Genomic Axis:	GA_x5YUV32W1V9 (1289841 . . . 1413275)
Chromosome:	9
OMIM NUMBER:	607928
OMIM Information:	Deafness, autosomal recessive 31, 607084 (3)

TABLE 1-continued

Transcript SNP info and associated gene/protein information	
Transcript Sequence (SEQ ID NO: 3):	
Protein Sequence (SEQ ID NO: 8):	
<u>SNP Information</u>	
Context (SEQ ID NO: 13): CTCTCCCAGCTCTCGGACAGCGGGCAGACTCTAAGCGAGGACAGTGGTGT GGATGCTGGCGAGGACAGAGGCCAGCGCCCCAGGCCGAGGAAGGCAGTCGG YGTCCACCAAGAGCAGGAGTAGCAAGGAGCTGCCTCGGAACGAGAGGCC ACAGATGGGGCCAACAACCGCCTGGACTTCTGGAGCCCACGTCCACTCT G	
Celera SNP ID:	hCV1920609
Public SNP ID:	rs2274159
SNP in Transcript Sequence	SEQ ID NO: 3
SNP Position	1796
Transcript:	
SNP Source:	dbSNP; Celera; HapMap; ABI_Val; HGBASE
Population(A allele, Count):	caucasian(T, 58 C, 62)
SNP Type:	Missense Mutation
Protein Coding:	SEQ ID NO: 8, at position 400, (V, GTG) (A, GCG)
Gene Number:	2
Celera Gene:	hCG32519 - 84000314391801
Celera Transcript:	hCT1965719 - 84000314391840
Public Transcript	
Accession:	
Celera Protein:	hCP1781858 - 197000064927551
Public Protein	
Accession:	
Gene Symbol:	DFNB31
Protein Name:	deafness, autosomal recessive 31
Celera Genomic Axis:	GA_x5YUV32W1V9 (1289841 . . . 1413275)
Chromosome:	9
OMIM NUMBER:	607928
OMIM Information:	Deafness, autosomal recessive 31, 607084 (3)
Transcript Sequence (SEQ ID NO: 4):	
Protein Sequence (SEQ ID NO: 9):	
<u>SNP Information</u>	
Context (SEQ ID NO: 14): CTCTCCCAGCTCTCGGACAGCGGGCAGACTCTAAGCGAGGACAGTGGTGT GGATGCTGGCGAGGACAGAGGCCAGCGCCCCAGGCCGAGGAAGGCAGTCGG YGTCCACCAAGAGCAGGAGTAGCAAGGAGCTGCCTCGGAACGAGAGGCC ACAGATGGGGCCAACAACCGCCTGGACTTCTGGAGCCCACGTCCACTCT G	
Celera SNP ID:	hCV1920609
Public SNP ID:	rs2274159
SNP in Transcript Sequence	SEQ ID NO: 4
SNP Position	1717
Transcript:	
SNP Source:	dbSNP; Celera; HapMap; ABI_Val; HGBASE



TABLE 2-continued

Genomic SNP info and associated gene information	
Celera SNP ID:	hcv939509
Public SNP ID:	rs717916
SNP in Genomic	SEQ ID NO: 17
Sequence:	
SNP Position Genomic:	21949
Related Interrogated	hcv1920609 (Power = .51)
SNP:	
SNP Source:	dbSNP; HapMap; HGBASE
Population(A allele, Count):	caucasian(T, 58 C, 62)
SNP Type:	INTRON
Context (SEQ ID NO: 23): CCATGCCAGTCTCTCTGGAAACCCCTATCCTAGCGCCTTGGCAGAGTC TGCAGGACGGTGGGAGCTGATGAAATCCCCTTGCAACAACAGCCCCAAT SACATGTTTTAAATGCAAACCGGCTGGACTTTAGGAGGCAAAGGCCAC GTGGACATCGCCTCCCTCCGTCAGAGAAGACTGTGAGACGACTGCCCCAGG A	
Celera SNP ID:	hcv1920602
Public SNP ID:	rs4979377
SNP in Genomic	SEQ ID NO: 17
Sequence:	
SNP Position Genomic:	5016
Related Interrogated	hcv1920609 (Power = .51)
SNP:	
SNP Source:	dbSNP; Celera; HapMap; HGBASE
Population(A allele, Count):	caucasian(C, 58 G, 62)
SNP Type:	HUMAN-MOUSE SYNTENIC REGION; INTRON
Context (SEQ ID NO: 24): AACCTAGGGTTTCAAACCTCAGGGGCCCGGCTGTTATACACATAGTGAAGC AGCGTGGTTGACTGGTGGGCTGGAGGGCATGTGCCACCACCTCAGCTCA SGCAGGGGCCATGCAGAGTTACAGACCAGGTGTGGTGAATTTCTCTTTT TCTCTGGTAAGCTAGAAATTAATCTCTCTCTCTCTCTCTCTCTCTTTT T	
Celera SNP ID:	hcv32122697
Public SNP ID:	rs4979380
SNP in Genomic	SEQ ID NO: 17
Sequence:	
SNP Position Genomic:	12903
Related Interrogated	hcv1920609 (Power = .51)
SNP:	
SNP Source:	dbSNP; HapMap; ABI_Val; HGBASE
Population(A allele, Count):	caucasian(G, 58 C, 62)
SNP Type:	INTRON
Gene Number:	3
Celera Gene:	hcg27735 - 83000098893244
Gene Symbol:	DAPK1
Protein Name:	death-associated protein kinase 1
Celera Genomic Axis:	GA_x5YUV32VUMT (19054067 . . . 19283800)
Chromosome:	9
OMIM NUMBER:	600831
OMIM Information:	
Genomic Sequence (SEQ ID NO: 18):	
<u>SNP Information</u>	
Context (SEQ ID NO: 25): GAAATCAGTCTGCCAAGATAAAATATAACTGATAATAGGCTTATCAG TGTGACCACAGTAAATAATCCATTATTATATACATCATAACACAGCGTA	

TABLE 2-continued

Genomic SNP info and associated gene information	
RTTGCTCACTTACCTGTCTCTTAAGGGAACCTGCATCTAATTATCTTTTA ACTTTTAGGGCTGACATATACTTGACCTATAATTGTTGAATTAGCCAACT C	
Celera SNP ID:	hcv1386888
Public SNP ID:	rs4877365
SNP in Genomic	SEQ ID NO: 18
Sequence:	
SNP Position Genomic:	52962
SNP Source:	dbSNP; Celera; HapMap; ABI_Val; HGBASE; CDX_Alzheimer
Population(A allele, Count):	caucasian(G, 90 A, 30)
SNP Type:	INTRON
Context (SEQ ID NO: 26): TCTTTGCGCTTGTGGATCAGCCAAGGCCACCTTTTAAAGAGTTCCTTTGTT CACTGGGGAGGATTTCTGTCATGCCAAGGGCTGGGAATAAAAAGTCACAA RTAGAAAGACTTGTCTCAGCATTGGCAACTCCCATCCCTCAGTGGTCTCT TGAGTTTTGGTAAATCTAGACTGATGAATATAGCAGCTAATGATCATAGT A	
Celera SNP ID:	hcv1386973
Public SNP ID:	rs11141899
SNP in Genomic	SEQ ID NO: 18
Sequence:	
SNP Position Genomic:	84240
SNP Source:	dbSNP; Celera; HapMap; CDX_Alzheimer
Population(A allele, Count):	caucasian(A, 89 G, 29)
SNP Type:	INTRON
Context (SEQ ID NO: 27): AAAGCAAGCTGAGCAGATGTTTTCCTGTAGTGTGAGTGTACTTCATTA TCTGTGCTAATAATATAAGAATAAAAAGCAAGCGCAGTGGACTCCTAT RTTTGCTCTTATGTTTTGTACTGTCAATCAATCATGCTCTCTCAGGCGC GTGTTTCTCTGGGGATAGTTTTATCTTGTCAAGTGTGTTTTGTTCTTGG A	
Celera SNP ID:	hcv1386978
Public SNP ID:	rs3128521
SNP in Genomic	SEQ ID NO: 18
Sequence:	
SNP Position Genomic:	88531
SNP Source:	dbSNP; Celera; HapMap; ABI_Val; HGBASE; CDX_Alzheimer
Population(A allele, Count):	caucasian(A, 74 G, 46)
SNP Type:	INTRON
Context (SEQ ID NO: 28): AGGGATGAGGCGCGAGAAGCTACCCACAGGCCATAGAACATTTCCATTTT TTTTACCTCATTTGCTGGTGTCTGATATTTTTCGGTGGCAGCCTTCTTC YGATTTTAAATGTAACCTCATGCTCATGTAGAAAGTTTGAACAACATAG AAAACCTTGAAGAGAAAATAGGACTCTCTAGTAATCCCATCCCCGAAA G	
Celera SNP ID:	hcv1386982
Public SNP ID:	rs4878104
SNP in Genomic	SEQ ID NO: 18
Sequence:	
SNP Position Genomic:	90260
SNP Source:	dbSNP; Celera; HapMap; ABI_Val; HGBASE; CDX_Alzheimer
Population(A allele, Count):	caucasian(C, 82 T, 38)
SNP Type:	INTRON; REPEATS
Gene Number:	4
Celera Gene:	hcg32521 - 84000314391890
Gene Symbol:	AKNA

TABLE 2-continued

Genomic SNP info and associated gene information	
Protein Name:	AT-hook transcription factor
Celera Genomic Axis:	GA_x5YUV32W1V9 (1223622 . . .1306270)
Chromosome:	9
OMIM NUMBER:	
OMIM Information:	
Genomic Sequence (SEQ ID NO: 19):	
<u>SNP Information</u>	
Context (SEQ ID NO: 29): GCCCTCTGCATCCTCCTCCTCCTGGCTCACCAGTTTGGTGGCCCATCTG TGGGCTCTCGTTCCGAGGCAGCTCCTTGTCTACTCTGCTCTTGGTGGAC RCCGACTGCCTTCCTCGGCCTGGGGCGCTGGCCTCTGCCTCGCCAGCATC CACACCACTGTCCTCGTTAGAGTCTGCCCGCTGTCCGAGAGCTGGGAGA G	
Celera SNP ID:	hCV1920609
Public SNP ID:	rs2274159
SNP in Genomic	SEQ ID NO: 19
Sequence:	
SNP Position Genomic:	78109
SNP Source:	dbSNP; Celera; HapMap; ABI_Val; HGBASE
Population(AAllele, Count):	caucasian(A, 58 G, 62)
SNP Type:	MISSENSE MUTATION; HUMAN- MOUSE SYNTENIC REGION
Context (SEQ ID NO: 30): CCATTCCTCCAGTCTCTCTGGAAACCCCTATCCTAGCGCCTCTGGCAGAGTC TGCAGGACGGTGGGAGCTGATGAAATCCCTTGCAACAACAGCCCCAAT SACATGTTTTAAATGCAAAACCGGCTGGACTTTAGGAGGCAAAGGCCCCAC	

TABLE 2-continued

Genomic SNP info and associated gene information	
GTGGACATCGCCTCCCTCCTGCAGAAGACTGTGAGACGACTGCCCCAGG A	
Celera SNP ID:	hCV1920602
Public SNP ID:	rs4979377
SNP in Genomic	SEQ ID NO: 19
Sequence:	
SNP Position Genomic:	71235
Related Interrogated	hCV1920609 (Power = .51)
SNP:	
SNP Source:	dbSNP; Celera; HapMap; HGBASE
Population(AAllele, Count):	caucasian(C, 58 G, 62)
SNP Type:	HUMAN-MOUSE SYNTENIC REGION; INTRON
Context (SEQ ID NO: 31): AACCTAGGGTTTCAAACCTCAGGGGCCGGCTGTTATACACATAGTGAAGC AGCGTGGGTTGACTGGTGGGCTGGAGGCATGTGCCACCAGCTCAGCTCA SGCAGGGGCCATGCAGAGGTTAGACAGGTTGGTGGAAATTTCTCTTTT TCTCTTGGTAAGCTAGAAATTAATACTTTCTCTCTCTCTCTCTCTTTT T	
Celera SNP ID:	hCV32122697
Public SNP ID:	rs4979380
SNP in Genomic	SEQ ID NO: 19
Sequence:	
SNP Position Genomic:	79122
Related Interrogated	
SNP:	hCV1920609 (Power = .51)
SNP Source:	dbSNP; HapMap; ABI_Val; HGBASE
Population(AAllele, Count):	caucasian(G, 58 C, 62)
SNP Type:	INTRON

[0460]

TABLE 3

Allelic tests of the POMT1 SNP rs2018621 (hCV8715115)													
Sample	Case*					Control*					Allelic P	Allelic OR	Power
	AA	AG	GG	SUM	MAF	AA	AG	GG	SUM	MAF			
WU**	2	42	347	391	5.9	0	25	330	355	3.5	0.03239	1.71 (1.04:2.82)	
SD***	0	17	171	188	4.5	1	20	337	358	3.1	0.11016	1.49 (0.87:2.57)	0.57
UK1***	2	32	317	351	5.1	0	22	367	389	2.8	0.0114	1.86 (1.18:2.92)	0.70
SD+UK1***	2	49	488	539	4.9	1	42	704	747	2.9	0.0054	1.70 (1.21:2.41)	0.88
All****	4	91	835	930	5.3	1	67	1034	1102	3.1	0.0009	1.71 (1.24:2.35)	

\*Counts of genotype 11, 12 and 22 and minor allele frequency (MAF)  
 \*\*Exploratory sample set, 2 sided P-value and OR (95% CI) are shown  
 \*\*\*Replication sample set, 1 sided P-value and OR (90% CI) are shown  
 \*\*\*\*2-sided P-value, Cochran Mantel and Haenzel test, using sample set as the stratifying variable

[0461]

TABLE 4

Allelic tests of the DFNB31 SNP rs2274159 (hCV1920609)													
Sample	Case*					Control*					Allelic P	Allelic OR	Power
	AA	AG	GG	SUM	MAF	AA	AG	GG	SUM	MAF			
WU**	117	212	90	419	53.2	84	193	99	376	48.0	0.038	1.23 (1.01:1.50)	
SD***	54	101	46	201	52.0	86	192	99	377	48.3	0.11	1.16 (0.91:1.48)	0.51

TABLE 4-continued

Sample	Case*					Control*					Allelic P	Allelic OR	Power
	AA	AG	GG	SUM	MAF	AA	AG	GG	SUM	MAF			
UK1***	77	191	80	348	49.6	79	197	108	384	46.2	0.10	1.14 (0.93:1.40)	0.63
SD+UK1***	131	292	126	549	50.5	165	389	207	761	47.2	0.040	1.15 (0.98:1.35)	0.83
All****	248	504	216	968	51.7	249	582	306	1137	47.5	0.0077	1.18 (1.05:1.34)	

\*Counts of genotype 11, 12 and 22 and minor allele frequency (MAF)  
 \*\*Exploratory sample set, 2 sided P-value and OR (95% CI) are shown  
 \*\*\*Replication sample set, 1 sided P-value and OR (90% CI) are shown  
 \*\*\*\*2-sided P-value, Cochran Mantel and Haenzel test, using sample set as the stratifying variable

[0462]

TABLE 5

Allelic association of the DAPK1 SNP rs4878104 (hCV1386982) and rs4877365 (hCV1386888) with LOAD case control samples														
Marker	Sample	Case*					Control *					Allelic P	Allelic OR	Power
		11****	12****	22****	SUM	MAF	11****	12****	22****	SUM	MAF			
rs4878104	WU**	36	181	170	387	32.7	55	159	138	352	38.2	0.027	0.79 (0.63:0.97)	
	SD***	17	84	79	180	32.8	46	161	121	328	38.6	0.033	0.78 (0.62:0.97)	0.54
	UK***	30	169	148	347	33.0	55	181	149	385	37.8	0.028	0.81 (0.68:0.97)	0.70
	SD+UK***	47	253	227	527	32.9	101	342	270	713	38.1	0.0042	0.80 (0.69:0.92)	0.88
	WU+SD+UK	83	434	397	914	32.8	156	501	408	1065	38.2	0.0006	0.79 (0.69:0.90)	
rs4877365	WU**	24	148	203	375	26.1	41	141	170	352	31.7	0.020	0.76 (0.61:0.96)	
	SD***	11	80	89	180	28.3	35	151	144	330	33.5	0.045	0.79 (0.62:0.99)	0.59
	UK***	18	144	186	348	25.9	30	165	191	386	29.2	0.080	0.85 (0.70:1.03)	0.76
	SD+UK***	29	224	275	528	26.7	65	316	335	716	31.2	0.015	0.82 (0.71:0.95)	0.91
	WU+SD+UK	53	372	478	903	26.5	106	457	505	1068	31.3	0.0017	0.80 (0.69:0.92)	

\*Counts of genotype 11, 12 and 22 and minor allele frequency (MAF)  
 \*\*Exploratory sample set, 2 sided P-value and OR (95% CI) are shown  
 \*\*\*Replication sample set, 1 sided P-value and OR (90% CI) are shown  
 \*\*\*\*11, 12 and 22 denote TT, TC and CC for rs4878104  
 11, 12 and 22 denote AA, AG and GG for rs4877365

[0463]

TABLE 6

Haplotype analysis of DAPK1 SNPs					
Sample	Haplotype*	control_freq	case_freq	Haplotype P**	Global P
WU	AAT	0.320	0.243	0.0025	0.046
	AGC	0.304	0.323	0.40	
	GAC	0.155	0.152	0.75	
	AAC	0.142	0.171	0.26	
	GAT	0.062	0.084	0.23	
SD	GAT	0.017	0.027	0.15	0.52
	GGC	0.017	0.027	0.15	
	AAT	0.328	0.272	0.03	
	AGC	0.294	0.306	0.34	
	GAC	0.143	0.163	0.84	
UK	AAC	0.161	0.189	0.13	0.067
	GAT	0.053	0.056	0.47	
	GGC	0.017	0.015	0.54	
	AAT	0.280	0.257	0.13	
	AGC	0.336	0.322	0.75	
	GAC	0.121	0.164	0.99	
	AAC	0.147	0.173	0.05	
	GAT	0.097	0.073	0.95	
	GGC	0.018	0.010	0.88	

TABLE 6-continued

Haplotype analysis of DAPK1 SNPs					
Sample	Haplotype*	control_freq	case_freq	Haplotype P**	Global P
SD+UK	AAT	0.307	0.262	0.0041	0.056
	AGC	0.312	0.320	0.36	
	GAC	0.140	0.164	0.97	
	AAC	0.150	0.175	0.029	
	GAT	0.073	0.067	0.83	
All	GGC	0.017	0.012	0.75	0.0087
	AAT	0.307	0.254	0.00021	
	AGC	0.312	0.321	0.53	
	GAC	0.140	0.159	0.05	
	AAC	0.150	0.173	0.06	
	GAT	0.073	0.074	0.80	
	GGC	0.017	0.018	0.57	

\*rs11141899, rs3128521, rs4878104 (=hCV1386973, hCV1386978, hCV1386982)  
 \*\*2 sided for WU and All, 1 sided for SD and UK

[0464]

TABLE 7

LD SNPs						
Interrogated SNP	Interrogated rs	LD SNP	LD SNP rs	Power (T)	Threshold $r_r^2$	$r^2$
hCV1920609	rs2274159	hCV1920602	rs4979377	0.51	0.89	1
hCV1920609	rs2274159	hCV32122697	rs4979380	0.51	0.89	1
hCV1920609	rs2274159	hCV939509	rs717916	0.51	0.89	1

[0465]

## SEQUENCE LISTING

&lt;160&gt; NUMBER OF SEQ ID NOS: 31

&lt;210&gt; SEQ ID NO 1

&lt;211&gt; LENGTH: 3591

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Homo sapien

&lt;400&gt; SEQUENCE: 1

```

cccgcaggct cggatgaatcg aacgttgagc agggcgggtg gtggtgcgga gtgccgagcg      60
gcctcacccc caaccgtcgg cccagtcgga cgttccgag cgttgcgag gagccgggag      120
cggctctgtg tggactcggg gaaacgcggg gcgtctgcct gagcccgtt ttctacaaga      180
tgtggggatt tttgaagcgc cctgtagtgg tgacggctga catcaacttg agccttgtgg      240
ccctgactgg gatgggggta ctgagccggc tgtggcgact cacctaccg cgggctgtgg      300
tttttgacga agtataattat gggcagtaca tctcttttta catgaaacaa atcttcttct      360
tggatgacag tggcccgcca tttggccaca tgggtgctgg cttgggaggt tatttaggag      420
gattcgatgg caattttttg tggaacagaa ttggagcaga atacagtagc aacgtgcctg      480
tgtgggcctt gcgcctgctg ccagcactcg cgggggcctt gtcggtcctc atggcctacc      540
agatagtgtt ggagctccac tttttctcatt gtgccccat gggagctgct ctgttgatgc      600
ttatcgagaa tgctctcctc actcagtcga ggctaagct tttggaatca gtgttaatat      660
ttttcaatct attggccgtg ttgtctctacc tgaagtctt caactgcaa aagcacagcc      720
ctttttctct gagctgggtg ttctggctaa cactgacagg ggtcgctgtg tcctgtgcag      780
tgggcatcaa gtacatgggt gtgttcacgt acgtgctcgt gctgggtgtt gcagctgtcc      840
atgcctggca cctgcttggg gaccagactt tgtccaatgt aggtgctgat gtccagtgtc      900
gcatgaggcc ggcctgtatg gggcagatgc ggatgtcaca gggggtactt ggtgaaaaga      960
ctccaatcct caatgtttta gaagcaggca ggctgggca gcctgcctc ttggcctctg     1020
cagggtgcctc tgtatgggag gccagagttt ctgtcactaa ctttttctaa gctcacaatg     1080
tctagagggt ggtgcgcttt tccacgcagt ggaacatgac ttttcttga atctctggca     1140
ggtctgtgtg ttctgtcact tgctgcccag agcagtggct ttgctgggtc tcccggctct     1200
cctgtactta ctgttcttct acgtccactt gattctagtc ttccgctctg gccccacga     1260
ccaaatcatg tccagtcctc tccaggccag ctttagagga ggactagctc ggatcaccca     1320
gggtcagcca ctggagggtg cctttgggtc ccaggtcact ctgaggaacg tctttgggaa     1380

```

-continued

---

acctgtgccc	tgctggcttc	attccocacca	ggacacctac	cccatgatat	atgagaacgg	1440
ccgaggcagc	tcccaccagc	aacaggtgac	ctgttaccoc	ttcaaagacg	tcaataactg	1500
gtggattgta	aaggatccca	ggaggcacca	gctggtggtg	agcagccctc	cgagacctgt	1560
gaggcacggg	gacatggtgc	agctggtcca	cgcatgacc	accgctccc	tgaacacgca	1620
tgatgttgca	gccccctga	gccccattc	acaggaggtc	tctgctaca	ttgactataa	1680
catctccatg	cccgccaga	acctctggag	actggaatt	gtgaacagag	gatctgacac	1740
agacgtctgg	aagaccatcc	tctcagaggt	ccgctttgtg	cacgtgaaca	cttccgctgt	1800
cttaaagctg	agcggggctc	acctccctga	ctgggggtat	cggcaactgg	agatcgtcgg	1860
ggagaagctg	tcccggggct	accacgggag	cacggtgtgg	aacgtggagg	agcaccgata	1920
cggcgcgagc	caggagcaga	gggagcggga	acgggagctg	cactcacctg	cgcaggtgga	1980
cgtcagcagg	aacctcagct	tcatggcgag	attctcggag	ctgcagtgga	ggatgctggc	2040
gctgagaagt	gatgactcgg	aacacaagta	cagctccagc	ccactggagt	gggtcaccct	2100
ggacaccaat	attgcctact	ggctgcaccc	caggaccagc	gctcagatcc	acctacttgg	2160
aaacatagtg	atctgggttt	cgggcagcct	cgctctggcc	atctacgcc	tgctgtcctt	2220
gtggtacctg	ctccgacggc	gaagaaatgt	ccatgacctc	cctcaggatg	cctggctcgg	2280
ctgggtgctg	gctggggcgc	tggtgcccgg	tggtgggca	gtgaactacc	tcccgttctt	2340
cctgatggag	aagacactct	tctctacca	ctacctgccc	gactcacct	tccaaatcct	2400
tctgtccct	tggtcctgc	agcacatcag	cgaccacctg	tgagggtccc	agctccagag	2460
gagcatcttc	agcgcctcgg	tggtggcctg	gtactcctcc	gcgtgccacg	tgtccaacac	2520
gctgcgccca	ctcacctacg	gggacaagtc	actctcgcca	catgaactca	aggcccttcg	2580
ctggaaagac	agctgggaca	tcttgatccg	aaaacactag	aacaagagtg	tgccaagaa	2640
caccctgtct	ggggtcggga	cgaggttgaa	gggtcttgg	caatgtacgt	aatgagcagg	2700
gtggggccca	cgctgggagc	acacgggctg	ggctgagcag	ggcctctagt	ggaacacatg	2760
gggttctcat	tgaaaagctc	tctgatgagc	acctcctttt	tgcaaaagt	aattttttct	2820
cgacaataaa	gatattccgt	gtcttacc	ctgaactaag	acacagggag	tatttcagag	2880
gccagcgtag	gagtcacgca	caacgaaaag	ccgagaaccc	agggccagca	gtggagcctc	2940
agcagaccag	ggcctggctc	ttgctaattg	ctgcagggtg	gagtttgatc	tgccagaccc	3000
gatcctcctt	catgaacacc	cagcaacctg	agcaagtccc	ggcctgccc	tcagcgagcc	3060
cggcagcgt	cccgggacag	ctcagtgttg	gagggccacc	tgaaccacga	gccagggctg	3120
gggcttgcat	gtcattgtct	atgacagcgt	caagactggc	ccttggcacc	gtgctgtgtg	3180
gaaaccctcc	cctctgagac	tccactgaga	cgtggctgag	tgaaatcttc	ctcgtcagtg	3240
gtcaaggtgt	gtcatccata	cagctccatg	cctttgtctt	ttttaaatgt	aattaaaaaa	3300
ggaaccaact	ggcgtttgta	gttggtgttt	gttatttatt	gattctcct	ccaaaggggc	3360
ctaagctcga	caaaaccgtt	acagttgtca	gaaccagcc	acagggtca	gcccccttc	3420
cccacgtcac	gtctgcatca	ctactgtggg	tgagcctgga	cggacggggg	ctggggctgc	3480
ccgtggcagc	ggcaagggat	gctttccaga	gacagccacc	acgagggagg	gaggatcacc	3540
ccaggcaacc	cagacacggg	tgttccatg	tgaggctgtg	agctccacat	a	3591

&lt;210&gt; SEQ ID NO 2

-continued

&lt;211&gt; LENGTH: 3521

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Homo sapien

&lt;400&gt; SEQUENCE: 2

```

cgagccccga agtaagaact ccgtggcgcg tgcagcccta gggaggaagg gcgtgctgcg      60
ggtgcagtct caggggcgctc tgcctgagcc cgcttttcta caagatgtgg ggatttttga      120
agcgcctctg agtgggtgacg gctgacatca acttgagcct tgtggccctg actgggatgg      180
ggttactgag ccggctgtgg cgactcacct acccgcgggc tgtggttttt gacgaagtat      240
attatgggca gtacatctct ttttacatga aacaaatctt cttcttgatg gacagtgggc      300
cgccatttgg ccacatggtg ctggccttgg gaggttattt aggaggattc gatggcaatt      360
ttttgtggaa cagaattgga gcagaataca gtagcaacgt gcctgtgtgg tccttgcgcc      420
tgctgccagc actcgcgggg gccttgtcgg tccccatggc ctaccagata gtgttgagc      480
tccacttttc tcattgtgcc gccatgggag ctgctctgtt gatgcttacc gagaatgctc      540
tcatacactca gtaaggcta atgcttttgg aatcagtgtt aatatttttc aatctattgg      600
ccgtgttctc ctacctgaag ttcttcaact gccaaaagca cagccctttt tctctgagct      660
ggtggttctg gtaaacactg acaggggtcg cttgttctcg tgcagtgggc atcaagtaca      720
tgggtgtgtt cacgtacgtg ctcgtgctgg gtgttgacgc tgtccatgcc tggcacctgc      780
ttggagacca gactttgtcc aatgtaggtg ctgatgtcca gtgctgcatg aggccggcct      840
gtatggggca gatgcggatg tcacaggggg tacttggtga aaagactcca atcctcaatg      900
ttttagaagc aggcaggcct gggcagcctc gcctcttggc ctctgcaggt gcctctgtat      960
gggaggccag agtttctgtc actaaccttt tctaagctca caatgtctag agtggggtgc     1020
gcttttccac gcagtggaac atgacttttc tttgaatctc tggcaggctc gtgtgttctg     1080
tcacttctc gcccagagca tggctttgct ggtcatcccc gtcgtcctgt acttactgtt     1140
cttctacgtc cacttgattc tagtcttccg ctctgggccc cacgacaaa tcatgtccag     1200
tgccttccag gccagcttag agggaggact agctcggatc acccagggtc agccactgga     1260
ggtggccttt ggtcccagg tcactctgag gaacgtcttt gggaaacctg tgccctgctg     1320
gcttcattcc caccaggaca cctaccccat gatgtaagat atgagaacgg ccgaggcagc     1380
tcccaccagc aacaggtgac ctgttaccct tcaaaagacg tcaataactg gtggattgta     1440
aaggatccca ggaggcacca gctgggtggg agcagccctc cgagacctgt gaggcacggg     1500
gacatggtgc agctggtcca cggcatgacc acccgctccc tgaacacgca tgatgttgca     1560
gccccctga gccccattc acaggaggtc tcctgctaca ttgactataa catctccatg     1620
ccgcccaga acctctggag actggaaatt gtgaacagag gatctgacac agacgtctgg     1680
aagaccatcc tctcagaggt ccgctttgtg cacgtgaaca cttccgctgt cttaaagctg     1740
agcggggctc acctcctga ctgggggtat cggcaactgg agatcgtcgg ggagaagctg     1800
tccccgggct accacgggag cacgggtgtg aacgtggagg agcaccgata cggcgcgagc     1860
caggagcaga gggagcggga acgggagctg cactcacctg cgcagggtga cgtcagcagg     1920
aacctcagct tcattggcag attctcggag ctgcagtgga ggatgctggc gctgagaagt     1980
gatgactcgg aacacaagta cagctccagc ccaactggagt gggtcaccct ggacaccaat     2040
attgcctact ggctgcaccc caggaccagc gctcagatcc acctacttgg aaacatagtg     2100

```

-continued

---

```

atctggggttt cgggcagcct cgctctggcc atctacgcc tgcgtgcctt gtgggtacctg 2160
ctccgacggc gaagaaatgt coactgacct cctcaggatg cctggctgcg ctgggtgctg 2220
gctggggcgc tgtgtgccg tggtctggca gtgaactacc tcccgttcct cctgatggag 2280
aagacactct tcctctacca ctacctgcc gcactcacct tccaaatcct tctgctccct 2340
gtggtcctgc agcacatcag cgaccacctg tgcaggctcc agctccagag gagcatcttc 2400
agcgccctgg tgggtgctg gtaactctcc gcgtgccacg tgtccaaacac gctgcgcca 2460
ctcacctacg gggacaagtc actctcgcca catgaactca aggcctctcg ctggaaagac 2520
agctgggaca tcttgatccg aaaacactag aacaagatg tggcaaagaa caccctgct 2580
ggggtcggga cgaggttgaa ggtcttggc caatgtactg aatgagcagg gtgggcccc 2640
cgctgggagg acacgggctg ggctgagcag gccctctagt ggaacacatg ggggtctcat 2700
tgaaaagctc tctgatgagc acctcctttt gtgcaaagt aatttttct cgacaataaa 2760
gatattccgt gtcttcacc cctgaactaa acacaggag tatttcagag gccagcgtag 2820
gagtcacga caacgaaaag ccgagaaacc agggccagca gtggagcctc agcagaccag 2880
ggcctgttcc ttgctaattg ctgcagggtg gagtttgatc tggcagacc gatcctcctt 2940
catgaacacc cagcaacctg agcaagtccc ggccctgcc tcagcgagcc cggcaggcgt 3000
cccgggacag ctactgttg gagggccacc tgaaccacga gccaggcctg gggcttgcac 3060
gtcattgtct atgacagcgt caagactggc ccttggcacc gtgctgtgtg gaaacctcc 3120
cctctgagac tccactgaga cgtggctgag tgaatcttc ctgctcagtg gtcaaggtgt 3180
gtcatccata cagctccatg cctttgtctt ttttaaatgt aattaaanaa ggaaccaact 3240
ggcgtttgta gttggtggtt gttatttatt gattctccct ccaaggggc ctaagctcga 3300
caaaaccgtt acagttgtca gaaccacgac acagggtcga gccccttcc cccacgtcac 3360
gtctgcatca ctactgtggg tgagcctgga cggacggggg ctggggctgc ccgtggcagc 3420
ggcaagggat gctttccaga gacagccacc acgcaggagg gaggatcacc ccaggcaacc 3480
cagacacggg tgttcacatg tgaggctgtg agctccacat a 3521

```

```

<210> SEQ ID NO 3
<211> LENGTH: 2849
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

```

```

<400> SEQUENCE: 3

```

```

aaactcctag gcagcagagg cccacgacta cttcctcctg agtgccgttc agtggcctgt 60
gtccaggctc tgaagggtc caagaagctg gtgctgtctg tgactcagc agggcgcac 120
cctgggggct acgtcaccaa ccacatctac acctgggtgg acccgcaggg ccgagcacc 180
tccccacct cgggcctgcc ccagccccac ggtggtgccc tgaggcagca ggagggtgac 240
cggaggagca cctgcacct cctgcaagga ggggatgaga aaaagggtgaa cctggtgctg 300
ggggacggcc ggtccctggg cctcacgac cgtgggggag ctgagtacgg ccttggcatt 360
tacatcactg gcgtggaccc aggtctctgaa gcagaaggca gcgggctcaa ggttggggac 420
cagattctag aagtgaatgg cgggagcttt ctcaacatcc tacacgacga ggctgtcagg 480
ctgcttaagt catctcggca cctcatcctg acagtgaagg acgtcgggag gctgccccat 540
gcccgcacca ctgtggacga gaccaagtgg atcgccagtt cccggatcag ggagaccatg 600

```

-continued

---

gcgaactcgg cagggtttct tggcgatctc acaacagaag gaataaaca gccaggattt	660
tacaagggcc cagccggctc ccaggtgacc ctgagcagcc tggggaacca gacacgagtg	720
ctgctggagg agcaggctcg gcacctgctg aacgagcagg aacacgccac catggcctac	780
tacctggatg agtaccgtgg cggcagcgtc tctgtggagg ccctcgtcat ggccctgttc	840
aagctgtca acaccacgc caagttctca ctctctctg aggtgagagg caccatttcc	900
ccgaagacc tagaacgctt cgaccacctg gtgctgaggc gtgagattga gtccatgaag	960
gcgaggcagc cccagggccc cggggctggg gacacctact ccattggtctc ctacagtgc	1020
acgggttcat ccacaggcag ccacggcacc tccaccaccg tcagctcggc caggaacact	1080
ctggacctgg aggaaactgg cgaggctgtc cagggcaata tcaacgccct cccagatgtg	1140
tccgtggatg atgtcagatc cacctccagc gggctgtcaa gcttcaagcc actgcctcgc	1200
ccaccacctc tggccaagg caacgacctc ccactaggcc agccaaggaa gctggggaga	1260
gaggacctcc agccacctc tcctatgctt tcctgctcgg gcactgtctt ctcggtcca	1320
cagaaccgca gcccgccagc gggcaccgca cccaccaccg ggacctctc tgcacaggac	1380
ttgcctctt ccccatctc tgcctcctc tcccctgcca accccagctc caagaggccg	1440
ctggacgccc atctggcctt ggtcaaccaa caccacctc gcccttccc acgggtccag	1500
tcacccccgc acctgaaaag cccctctgca gaggccacag tggctggggg ctgcctcttg	1560
ccccatcac cctctggcca cccagaccag acaggcaca accagcactt tgcctggtg	1620
gaggctccac gccccgacag cgagccagac gtcaatgaag tgagggcgtt gccccagacg	1680
cgacacgct ctacgtctc ccagctctc gacagcgggc agactctaag cgaggacagt	1740
ggtgtggatg ctggcgaggc agaggccagc gccccaggcc gaggaaggca gtcggtgtcc	1800
accaagagca ggagtagcaa ggagctgctt cggaacgaga ggcccacaga tggggccaac	1860
aaaccgctg gacttctgga gccacgtcc actctggtcc gttgtaagaa aagtgcggcc	1920
accctgggca tcgccatcga gggctggcgc aacaccgccc agcccctgcc taggattgtc	1980
actattcaga gaggcgctc agctcacaac tgtggcagc tcaaggtggg ccactgtgatt	2040
ctggaagtga atggctgac gcttcggggc aaggagcacc gggaggccgc ccgcattatc	2100
gccgaggcct tcaagactaa ggacctgac tacattgact ttctggtcac tgagttcaat	2160
gtgatgctct agaggccaag gcctgagggc ctcccaccac tgcccagccc ctggtcccag	2220
tccctttcca ccgttggtt catcaagctc cttgcggggg tggggctgca tggccagggt	2280
ggcaggaaga catccccct ccatcccagc ccactggacc agaactggga gaggaagaga	2340
gcaggacaag gcagacagaa ggtcaggtca ggaactggtg ctgtactggg tacacagtag	2400
gcgccagga caagtgggtt gcaagacagg aagaaaggaa aaggaagggc agagtgctgg	2460
tttctccagg ttgggtggg ggcactgctg tccccctcc agctaggacc cagcccatcc	2520
ccagatgctt gacctttgt ccaaagtgag gtcactcgag aattcatgga cacggcccc	2580
agtcagggg catcttgcaa gaccttagt gccacaaata agcatcgagc acctccccat	2640
tcacacccc attcctcctg gctcctatc cccatggtg tttattattt atttccctcc	2700
ccatgcccct ggggaccca agggcccagc ttccctctgc acccccagcc tatcccagag	2760
gccttgacg tgaccagcag gtgcattgta tttatataca gagcttatga ctttaatttt	2820
tcaataaaga aatctgaaca aggttagac	2849

-continued

---

<210> SEQ ID NO 4  
 <211> LENGTH: 2770  
 <212> TYPE: DNA  
 <213> ORGANISM: Homo sapien

<400> SEQUENCE: 4

```

tagtattaaa cattttcaaa gttacttgcc aacatctaga aagataccag gttttctata      60
aaaaaaaaa ctggatttct ggatgcttct taaaaatcag gaagtctggc agcctgagcc      120
cacatcggct ggagctgagc cgcacctgcg agttgcatct gggatctcca gttcaccggc      180
ccctaagctc ctgagggttg ccctgaccct gaggttgccct gtcaatcacc attttctccc      240
tccactcctt gtgttacctg cctgttcctg cggggttggc aacaactcag gagcccacct      300
cgggtggttt tggaggtgcc gtgcacactg ctgattggga ggtgggacgc tgccagtctg      360
tccggagttt cctttacccc tgagtagccc ccagactgaa ctggcagcga gtggaggcca      420
cgatgcatgg ttctcttgaa gctttgctct tcctgcccc agtcaccctg tccttgcccc      480
acgcccattt gatctgctca aatgcacaac tggagatgtg tgtctttccc cacaggtttc      540
ttggcgatct cacaacagaa ggaataaaca agccaggatt ttacaagggc ccagccggct      600
cccagggtgac cctgagcagc ctggggaacc agacacgagt gctgctggag gagcaggctc      660
ggcacctgct gaacgagcag gaacacgcca ccatggccta ctacctggat gagtaccgtg      720
gcggcagcgt ctctgtggag gccctcgtca tggccctggt caagctgctc aacaccccag      780
ccaagttctc actcctctct gaggtgagag gcaccatttc cccgcaagac ctagaacgct      840
togaccacct ggtgctgagg cgtgagattg agtccatgaa ggcgcggcag cccccaggcc      900
ccggggcttg ggacacctac tccatggtct cctacagtga cacgggttca tccacaggca      960
gccacggcac ctccaccacc gtcagctcgg ccaggaacac tctggacctg gaggaaactg     1020
gcgaggctgt ccagggcaat atcaacgccc tcccagatgt gtccgtggat gatgtcagat     1080
ccacctccca ggggctgtca agcttcaagc cactgcctcg cccaccacct ctggcccaag     1140
gcaacgacct cccactaggg cagccaagga agctggggag agaggacctc cagccacctt     1200
cctccatgcc ttctgctcgt ggcactgtct tctcggtctc acagaaccgc agcccgccag     1260
cgggcaccgc acccaccaca gggacctcct ctgcacagga cttgccctct tccccatct     1320
atgcctccgt ctccccctgc aaccccagct ccaagaggcc gctggacgcc catctggccc     1380
tggtaaacca acacccatc ggcccttcc cacgggtcca gtcaccccgc cacctgaaaa     1440
gcccctctgc agaggccaca gtggctgggg gctgccttct gccccatca ccctctggcc     1500
accagacca gacaggcaca aaccagcact ttgtcatggt ggaggccac cgcctcgaca     1560
gcgagccaga cgtcaatgaa gtgagggcgc tgccccagac gcgcacagcc tctacgctct     1620
cccagctctc ggacagcggg cagactctaa gcgagacag tgggtgggat gctggcgagg     1680
cagaggccag cgcctcaggc cgaggaaggg agtcggtgct caccaagagc aggagtagca     1740
aggagctgcc tcggaacgag agggccacag atggggccaa caaacgcctt ggacttctgg     1800
agcccacgct cactctggtc cgtgtgaaga aaagtgcggc caccctgggc atcgccatcg     1860
aggggtggcg caacaccgct cagcccctgc ctaggattgt cactattcag agaggcggt     1920
cagctcacia ctgtgggagc ctcaaggtgg gccacgtgat totggaagtg aatgggctga     1980
cgcttcgggg caaggagcac cgggaggccc cccgatttat cgcgagggcc ttcaagacta     2040

```

-continued

---

```

aggaccgtga ctacattgac tttctggta ctgagttcaa tgtgatgctc tagaggccaa 2100
ggcctgaggg cctcccacca ctgcccagcc cctgggtcca gtcctttcc accgttggt 2160
tcataaagct cttgcgggg ttggggctgc atggccaggg tggcaggaag acatcccccc 2220
tccatcccag cccactggac cagaactggg agaggaagag agcaggacaa ggcagacaga 2280
aggtcaggtc aggaactggg gctgtactgg gtacacagta ggcgcccagg acaagtgggt 2340
tgcaagacag gaagaaagga aaaggaaggg cagagtgtct gtttctccag gttgggttg 2400
gggactgct gtccccctc cagctaggac ccagccatc cccagatgcc tgagccttg 2460
tccaaagtga ggtcactcga gaattcatgg acacggcccc cagtcagggg gcactctgca 2520
agaccttag tgccacaaat aagcatcagc cacctcccca ttcaacccc cattctcct 2580
ggctccttat cccccatgg ttttattatt ttttccctc cccatgcccc tggggacccc 2640
aaggccccag cttccctctg cccccccagc ctatcccaga ggccttgca gtgaccagca 2700
gtgtcattgt atttatatac agagcttatg actttaattt ttcaataaag aaatctgaac 2760
aaggttagac 2770

```

```

<210> SEQ ID NO 5
<211> LENGTH: 4100
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

```

```

<400> SEQUENCE: 5

```

```

aacgagacgg cgtcggaact cgctggaaga ctcgcccggg gtggctcaac agttcgaaa 60
gtccgggctt gcggccgggg ggcgcagagg cgcggaggct ccccgggcgg ctgcatcccc 120
gggctgcccc ggacatgccc cgcacccctt gtcggagccc cgcgcccgtt caccggctgc 180
cactccccgg agtgctctgg gactagagcg ccctccgccc caacttctga ctttggggtc 240
ccccagccct agaggaacgg gcgcccggct gggggagccc cgggcatggg gtgacggggc 300
agtcaggta cttttgctgg gcccccacac tccagccacc cgtcgtctcc gtgccctcca 360
gaatgtagct gctgaccccc cgcccccgg actccagcag cagtctccgc cacctccggg 420
ctccagcagc caactcttgt gtctctcagg aaccgcccga cgtctccgcg tccccggat 480
tccaggttcg cgacccccgc cgtcttcgcg cctccgggac tctagcagcg gtctccgca 540
ccccctagga tccagccact gtaccctgct tccccaaacc caggctccag ccgtctctc 600
cgcgctcccg gggatccagc tgtcgcgcca gtaccgca ggacagcggc accgcccaca 660
ccccgcgcgc tcagagcccc gccgggcctc ggcgtggaga tgaacgcgcc gctggacggc 720
ctgtcggta gctcgtcctc caccggctcg ctgggctcgg cggccggggc gggcgcgcc 780
gggggcgcgg ggtgcgggtt actgtctgcc aacgtgcgcc agctgcacca agcgtgacc 840
gcgctgctga gcgagcgga gcgggagcag ttcacccact gcctgaacgc ttaccacgcg 900
cgccgcaacg tcttcgacct ggtcgcaccc ctgcgcgtgc tgctggacag tccggtcaag 960
cggcgcctgc tgccatgct togtctggtc atccccgct ccgaccagct gctcttcgac 1020
caatacacgg ccgagggcct ctacctgcc gccaccacc cctacaggca gccgcctgg 1080
ggcgcccccg acagcgggg gccaggggag gtgcgcctgg tgagtttgcg gcgtgccaag 1140
gcccacgagg gcttgggctt cagcatccgt gggggctcgg agcacggcgt gggcatctac 1200
gtgtctctgg tgaaccagg ctctctagct gagaaggaag gactgcgggt cggggaccag 1260

```

-continued

---

attctgcgcg tcaacgacaa atccctggcc cgggtgaccc acgcggaggc cgtcaaggct	1320
ctgaagggct ccaagaagct ggtgctgtct gtgtactcag cagggcgcat ccctgggggc	1380
tacgtcacca accacatcta cacctgggtg gacccgcagg gccgcagcat ctccccaccc	1440
tcgggcctgc cccagcccca cgggtggtgc ctgaggcagc aggaggggtga ccggaggagc	1500
accctgcacc tcctgcaagg aggggatgag aaaaagggtga acctggtgct gggggacggc	1560
cggtcctctg gcctcacgat ccgtggggga gctgagtacg gccttggcat ttacatcact	1620
ggcgtggacc caggctctga agcagaaggc agcgggctca aggttgggga ccagattcta	1680
gaagtgaatg ggcggagctt tctcaacatc ctacacgacg aggctgtcag gctgcttaag	1740
tcctctcggc acctcatcct gacagtgaag gacgtcggga ggctgcccc tgcctccacc	1800
actgtggacg agaccaagtg gatcgcaggt tcccggatca gggagaccat ggcgaactcg	1860
gcagggtttc ttggcgatct cacaacagaa ggaataaaca agccaggatt ttacaagggc	1920
ccagccggct cccaggtgac cctgagcagc ctggggaacc agacacgagt gctgctggag	1980
gagcaggctc ggcacctgct gaacagagcag gaacacgcca ccatggccta ctacctggat	2040
gagtaccgtg gcggcagcgt ctctgtggag gcctctgta tggccctggt caagctgctc	2100
aacaccacg ccaagtctc actcctctct gaggtgagag gcaccatttc cccgcaagac	2160
ctagaacgct tcgaccacct ggtgctgagg cgtgagattg agtccatgaa ggcgcggcag	2220
ccccaggcc cggggctctg ggacacctac tccatggtct cctacagtga cacgggttca	2280
tccacaggca gccacggcac ctccaccacc gtcagctcgg ccaggaacac tctggacctg	2340
gaggaaactg gcgaggctgt ccagggcaat atcaacgccc tcccagatgt gtcctggat	2400
gatgtcagat ccacctcca ggggctgtca agcttcaagc cactgcctcg cccaccacct	2460
ctggcccaag gcaacgacct cccactaggc cagccaagga agctggggag agaggacctc	2520
cagccacctt cctccatgac ttctgtctcg ggcactgtct tctcggctcc acagaaccgc	2580
agcccgcag cgggacccgc acccaccoca gggacctcct ctgcacagga cttgccctct	2640
tccccatct atgcctcctg ctccccctgc aaccccagct ccaagaggcc gctggacgcc	2700
catctggccc tggtaacca acacccatc ggccccttc cacgggtcca gtcacccccg	2760
cacctgaaaa gccctctctg agaggccaca gtggctgggg gctgcctct gccccatca	2820
ccctctggcc acccagacca gacaggcaca aaccagcact ttgtcatggt ggaggctcac	2880
cgccccgaca gcgagccaga cgtcaatgaa gtgagggcgc tgccccagac gcgcacagcc	2940
tctacgctct cccagctctc ggacagcggg cagactctaa gcgaggacag tgggtggat	3000
gctggcgagg cagaggccag gcccccaggc cgaggaaggc agtccgtgct caccaagagc	3060
aggagtagca aggagctgac tcggaacgag aggccacag atggggccaa caaacgcct	3120
ggacttctgg agcccacgac cactctggct cgtgtgaaga aaagtgcggc caccctgggc	3180
atcgccatcg aggggtggcc caacaccgc cagcccctgc ctaggattgt cactattcag	3240
agaggcggt cagctcacia ctgtgggcag ctcaagggtg gccacgtgat tctggaagtg	3300
aatgggtga cgcttcgggg caaggagcac cgggagggcc cccgattat cgcgaggcc	3360
ttcaagacta aggacctgta ctacattgac tttctggta ctgagttcaa tgtgatgctc	3420
tagaggccaa ggcctgaggg cctccccaca ctgcccagcc cctggtccca gtcctttcc	3480
accgttggtc tcataagct ccttgcgggg ttggggctgc atggccaggg tggcaggaag	3540

-continued

```

acatcccccc tccatcccag cccactggac cagaactggg agaggaagag agcaggacaa 3600
ggcagacaga aggtcaggtc aggaactggt gctgtactgg gtacacagta ggcgcccagg 3660
acaagtgggt tgcaagacag gaagaaagga aaaggaaggg cagagtgctg gtttctccag 3720
gttgggttg gggcactgct gtccccctc cagctaggac ccagcccatc cccagatgcc 3780
tgagcctttg tccaaagtga ggtcactoga gaattcatgg acacggcccc cagtcagggg 3840
gcactttgca agacctttag tgccacaaat aagcatcgag cacctcccca ttcacacccc 3900
cattcctcct ggctccttat cccccatggt gtttattatt tatttccctc cccatgcccc 3960
tggggacccc aaggccccag cttccctctg cacccccagc ctatcccaga ggccttgca 4020
gtgaccagca gtgtcattgt atttatatac agagcttatg actttaattt ttcaataaag 4080
aatctgaac aaggttagac 4100

```

```

<210> SEQ ID NO 6
<211> LENGTH: 450
<212> TYPE: PRT
<213> ORGANISM: Homo sapien

```

```

<400> SEQUENCE: 6

```

```

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

```

-continued

Leu Ala Leu Arg Ser Asp Asp Ser Glu His Lys Tyr Ser Ser Ser Pro  
 260 265 270  
 Leu Glu Trp Val Thr Leu Asp Thr Asn Ile Ala Tyr Trp Leu His Pro  
 275 280 285  
 Arg Thr Ser Ala Gln Ile His Leu Leu Gly Asn Ile Val Ile Trp Val  
 290 295 300  
 Ser Gly Ser Leu Ala Leu Ala Ile Tyr Ala Leu Leu Ser Leu Trp Tyr  
 305 310 315 320  
 Leu Leu Arg Arg Arg Arg Asn Val His Asp Leu Pro Gln Asp Ala Trp  
 325 330 335  
 Leu Arg Trp Val Leu Ala Gly Ala Leu Cys Ala Gly Gly Trp Ala Val  
 340 345 350  
 Asn Tyr Leu Pro Phe Phe Leu Met Glu Lys Thr Leu Phe Leu Tyr His  
 355 360 365  
 Tyr Leu Pro Ala Leu Thr Phe Gln Ile Leu Leu Leu Pro Val Val Leu  
 370 375 380  
 Gln His Ile Ser Asp His Leu Cys Arg Ser Gln Leu Gln Arg Ser Ile  
 385 390 395 400  
 Phe Ser Ala Leu Val Val Ala Trp Tyr Ser Ser Ala Cys His Val Ser  
 405 410 415  
 Asn Thr Leu Arg Pro Leu Thr Tyr Gly Asp Lys Ser Leu Ser Pro His  
 420 425 430  
 Glu Leu Lys Ala Leu Arg Trp Lys Asp Ser Trp Asp Ile Leu Ile Arg  
 435 440 445  
 Lys His  
 450

<210> SEQ ID NO 7  
 <211> LENGTH: 348  
 <212> TYPE: PRT  
 <213> ORGANISM: Homo sapien  
 <400> SEQUENCE: 7

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

-continued

Asp Ser Glu His Lys Tyr Ser Ser Ser Pro Leu Glu Trp Val Thr Leu  
 165 170 175  
 Asp Thr Asn Ile Ala Tyr Trp Leu His Pro Arg Thr Ser Ala Gln Ile  
 180 185 190  
 His Leu Leu Gly Asn Ile Val Ile Trp Val Ser Gly Ser Leu Ala Leu  
 195 200 205  
 Ala Ile Tyr Ala Leu Leu Ser Leu Trp Tyr Leu Leu Arg Arg Arg Arg  
 210 215 220  
 Asn Val His Asp Leu Pro Gln Asp Ala Trp Leu Arg Trp Val Leu Ala  
 225 230 235 240  
 Gly Ala Leu Cys Ala Gly Gly Trp Ala Val Asn Tyr Leu Pro Phe Phe  
 245 250 255  
 Leu Met Glu Lys Thr Leu Phe Leu Tyr His Tyr Leu Pro Ala Leu Thr  
 260 265 270  
 Phe Gln Ile Leu Leu Leu Pro Val Val Leu Gln His Ile Ser Asp His  
 275 280 285  
 Leu Cys Arg Ser Gln Leu Gln Arg Ser Ile Phe Ser Ala Leu Val Val  
 290 295 300  
 Ala Trp Tyr Ser Ser Ala Cys His Val Ser Asn Thr Leu Arg Pro Leu  
 305 310 315 320  
 Thr Tyr Gly Asp Lys Ser Leu Ser Pro His Glu Leu Lys Ala Leu Arg  
 325 330 335  
 Trp Lys Asp Ser Trp Asp Ile Leu Ile Arg Lys His  
 340 345

<210> SEQ ID NO 8  
 <211> LENGTH: 524  
 <212> TYPE: PRT  
 <213> ORGANISM: Homo sapien

<400> SEQUENCE: 8

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

-continued

165					170					175					
Ala	Leu	Pro	Asp	Val	Ser	Val	Asp	Asp	Val	Arg	Ser	Thr	Ser	Gln	Gly
	180							185					190		
Leu	Ser	Ser	Phe	Lys	Pro	Leu	Pro	Arg	Pro	Pro	Pro	Leu	Ala	Gln	Gly
	195						200					205			
Asn	Asp	Leu	Pro	Leu	Gly	Gln	Pro	Arg	Lys	Leu	Gly	Arg	Glu	Asp	Leu
	210					215					220				
Gln	Pro	Pro	Ser	Ser	Met	Pro	Ser	Cys	Ser	Gly	Thr	Val	Phe	Ser	Ala
	225				230					235					240
Pro	Gln	Asn	Arg	Ser	Pro	Pro	Ala	Gly	Thr	Ala	Pro	Thr	Pro	Gly	Thr
			245					250						255	
Ser	Ser	Ala	Gln	Asp	Leu	Pro	Ser	Ser	Pro	Ile	Tyr	Ala	Ser	Val	Ser
		260						265					270		
Pro	Ala	Asn	Pro	Ser	Ser	Lys	Arg	Pro	Leu	Asp	Ala	His	Leu	Ala	Leu
		275					280					285			
Val	Asn	Gln	His	Pro	Ile	Gly	Pro	Phe	Pro	Arg	Val	Gln	Ser	Pro	Pro
	290					295					300				
His	Leu	Lys	Ser	Pro	Ser	Ala	Glu	Ala	Thr	Val	Ala	Gly	Gly	Cys	Leu
	305					310					315				320
Leu	Pro	Pro	Ser	Pro	Ser	Gly	His	Pro	Asp	Gln	Thr	Gly	Thr	Asn	Gln
			325						330					335	
His	Phe	Val	Met	Val	Glu	Val	His	Arg	Pro	Asp	Ser	Glu	Pro	Asp	Val
		340						345					350		
Asn	Glu	Val	Arg	Ala	Leu	Pro	Gln	Thr	Arg	Thr	Ala	Ser	Thr	Leu	Ser
		355					360					365			
Gln	Leu	Ser	Asp	Ser	Gly	Gln	Thr	Leu	Ser	Glu	Asp	Ser	Gly	Val	Asp
	370					375					380				
Ala	Gly	Glu	Ala	Glu	Ala	Ser	Ala	Pro	Gly	Arg	Gly	Arg	Gln	Ser	Val
	385					390					395				400
Ser	Thr	Lys	Ser	Arg	Ser	Ser	Lys	Glu	Leu	Pro	Arg	Asn	Glu	Arg	Pro
			405					410					415		
Thr	Asp	Gly	Ala	Asn	Lys	Pro	Pro	Gly	Leu	Leu	Glu	Pro	Thr	Ser	Thr
		420						425					430		
Leu	Val	Arg	Val	Lys	Lys	Ser	Ala	Ala	Thr	Leu	Gly	Ile	Ala	Ile	Glu
		435					440					445			
Gly	Gly	Ala	Asn	Thr	Arg	Gln	Pro	Leu	Pro	Arg	Ile	Val	Thr	Ile	Gln
	450					455					460				
Arg	Gly	Gly	Ser	Ala	His	Asn	Cys	Gly	Gln	Leu	Lys	Val	Gly	His	Val
	465					470					475				480
Ile	Leu	Glu	Val	Asn	Gly	Leu	Thr	Leu	Arg	Gly	Lys	Glu	His	Arg	Glu
			485					490					495		
Ala	Ala	Arg	Ile	Ile	Ala	Glu	Ala	Phe	Lys	Thr	Lys	Asp	Arg	Asp	Tyr
			500					505					510		
Ile	Asp	Phe	Leu	Val	Thr	Glu	Phe	Asn	Val	Met	Leu				
		515					520								

<210> SEQ ID NO 9  
 <211> LENGTH: 556  
 <212> TYPE: PRT  
 <213> ORGANISM: Homo sapien  
 <400> SEQUENCE: 9

-continued

---

Met His Gly Ser Leu Glu Ala Leu Leu Phe Leu Pro Gln Val Thr Leu  
1 5 10 15

Ser Leu Ala His Ala His Leu Ile Cys Ser Asn Ala Gln Leu Glu Met  
20 25 30

Cys Val Phe Pro His Arg Phe Leu Gly Asp Leu Thr Thr Glu Gly Ile  
35 40 45

Asn Lys Pro Gly Phe Tyr Lys Gly Pro Ala Gly Ser Gln Val Thr Leu  
50 55 60

Ser Ser Leu Gly Asn Gln Thr Arg Val Leu Leu Glu Glu Gln Ala Arg  
65 70 75 80

His Leu Leu Asn Glu Gln Glu His Ala Thr Met Ala Tyr Tyr Leu Asp  
85 90 95

Glu Tyr Arg Gly Gly Ser Val Ser Val Glu Ala Leu Val Met Ala Leu  
100 105 110

Phe Lys Leu Leu Asn Thr His Ala Lys Phe Ser Leu Leu Ser Glu Val  
115 120 125

Arg Gly Thr Ile Ser Pro Gln Asp Leu Glu Arg Phe Asp His Leu Val  
130 135 140

Leu Arg Arg Glu Ile Glu Ser Met Lys Ala Arg Gln Pro Pro Gly Pro  
145 150 155 160

Gly Ala Gly Asp Thr Tyr Ser Met Val Ser Tyr Ser Asp Thr Gly Ser  
165 170 175

Ser Thr Gly Ser His Gly Thr Ser Thr Thr Val Ser Ser Ala Arg Asn  
180 185 190

Thr Leu Asp Leu Glu Glu Thr Gly Glu Ala Val Gln Gly Asn Ile Asn  
195 200 205

Ala Leu Pro Asp Val Ser Val Asp Asp Val Arg Ser Thr Ser Gln Gly  
210 215 220

Leu Ser Ser Phe Lys Pro Leu Pro Arg Pro Pro Pro Leu Ala Gln Gly  
225 230 235 240

Asn Asp Leu Pro Leu Gly Gln Pro Arg Lys Leu Gly Arg Glu Asp Leu  
245 250 255

Gln Pro Pro Ser Ser Met Pro Ser Cys Ser Gly Thr Val Phe Ser Ala  
260 265 270

Pro Gln Asn Arg Ser Pro Pro Ala Gly Thr Ala Pro Thr Pro Gly Thr  
275 280 285

Ser Ser Ala Gln Asp Leu Pro Ser Ser Pro Ile Tyr Ala Ser Val Ser  
290 295 300

Pro Ala Asn Pro Ser Ser Lys Arg Pro Leu Asp Ala His Leu Ala Leu  
305 310 315 320

Val Asn Gln His Pro Ile Gly Pro Phe Pro Arg Val Gln Ser Pro Pro  
325 330 335

His Leu Lys Ser Pro Ser Ala Glu Ala Thr Val Ala Gly Gly Cys Leu  
340 345 350

Leu Pro Pro Ser Pro Ser Gly His Pro Asp Gln Thr Gly Thr Asn Gln  
355 360 365

His Phe Val Met Val Glu Val His Arg Pro Asp Ser Glu Pro Asp Val  
370 375 380

Asn Glu Val Arg Ala Leu Pro Gln Thr Arg Thr Ala Ser Thr Leu Ser  
385 390 395 400

Gln Leu Ser Asp Ser Gly Gln Thr Leu Ser Glu Asp Ser Gly Val Asp



-continued

---

Lys Gly Ser Lys Lys Leu Val Leu Ser Val Tyr Ser Ala Gly Arg Ile  
 210 215 220  
 Pro Gly Gly Tyr Val Thr Asn His Ile Tyr Thr Trp Val Asp Pro Gln  
 225 230 235 240  
 Gly Arg Ser Ile Ser Pro Pro Ser Gly Leu Pro Gln Pro His Gly Gly  
 245 250 255  
 Ala Leu Arg Gln Gln Glu Gly Asp Arg Arg Ser Thr Leu His Leu Leu  
 260 265 270  
 Gln Gly Gly Asp Glu Lys Lys Val Asn Leu Val Leu Gly Asp Gly Arg  
 275 280 285  
 Ser Leu Gly Leu Thr Ile Arg Gly Gly Ala Glu Tyr Gly Leu Gly Ile  
 290 295 300  
 Tyr Ile Thr Gly Val Asp Pro Gly Ser Glu Ala Glu Gly Ser Gly Leu  
 305 310 315 320  
 Lys Val Gly Asp Gln Ile Leu Glu Val Asn Gly Arg Ser Phe Leu Asn  
 325 330 335  
 Ile Leu His Asp Glu Ala Val Arg Leu Leu Lys Ser Ser Arg His Leu  
 340 345 350  
 Ile Leu Thr Val Lys Asp Val Gly Arg Leu Pro His Ala Arg Thr Thr  
 355 360 365  
 Val Asp Glu Thr Lys Trp Ile Ala Ser Ser Arg Ile Arg Glu Thr Met  
 370 375 380  
 Ala Asn Ser Ala Gly Phe Leu Gly Asp Leu Thr Thr Glu Gly Ile Asn  
 385 390 395 400  
 Lys Pro Gly Phe Tyr Lys Gly Pro Ala Gly Ser Gln Val Thr Leu Ser  
 405 410 415  
 Ser Leu Gly Asn Gln Thr Arg Val Leu Leu Glu Glu Gln Ala Arg His  
 420 425 430  
 Leu Leu Asn Glu Gln Glu His Ala Thr Met Ala Tyr Tyr Leu Asp Glu  
 435 440 445  
 Tyr Arg Gly Gly Ser Val Ser Val Glu Ala Leu Val Met Ala Leu Phe  
 450 455 460  
 Lys Leu Leu Asn Thr His Ala Lys Phe Ser Leu Leu Ser Glu Val Arg  
 465 470 475 480  
 Gly Thr Ile Ser Pro Gln Asp Leu Glu Arg Phe Asp His Leu Val Leu  
 485 490 495  
 Arg Arg Glu Ile Glu Ser Met Lys Ala Arg Gln Pro Pro Gly Pro Gly  
 500 505 510  
 Ala Gly Asp Thr Tyr Ser Met Val Ser Tyr Ser Asp Thr Gly Ser Ser  
 515 520 525  
 Thr Gly Ser His Gly Thr Ser Thr Thr Val Ser Ser Ala Arg Asn Thr  
 530 535 540  
 Leu Asp Leu Glu Glu Thr Gly Glu Ala Val Gln Gly Asn Ile Asn Ala  
 545 550 555 560  
 Leu Pro Asp Val Ser Val Asp Asp Val Arg Ser Thr Ser Gln Gly Leu  
 565 570 575  
 Ser Ser Phe Lys Pro Leu Pro Arg Pro Pro Pro Leu Ala Gln Gly Asn  
 580 585 590  
 Asp Leu Pro Leu Gly Gln Pro Arg Lys Leu Gly Arg Glu Asp Leu Gln  
 595 600 605  
 Pro Pro Ser Ser Met Pro Ser Cys Ser Gly Thr Val Phe Ser Ala Pro

-continued

610			615			620									
Gln	Asn	Arg	Ser	Pro	Pro	Ala	Gly	Thr	Ala	Pro	Thr	Pro	Gly	Thr	Ser
625					630					635					640
Ser	Ala	Gln	Asp	Leu	Pro	Ser	Ser	Pro	Ile	Tyr	Ala	Ser	Val	Ser	Pro
			645						650					655	
Ala	Asn	Pro	Ser	Ser	Lys	Arg	Pro	Leu	Asp	Ala	His	Leu	Ala	Leu	Val
			660					665						670	
Asn	Gln	His	Pro	Ile	Gly	Pro	Phe	Pro	Arg	Val	Gln	Ser	Pro	Pro	His
		675					680					685			
Leu	Lys	Ser	Pro	Ser	Ala	Glu	Ala	Thr	Val	Ala	Gly	Gly	Cys	Leu	Leu
	690					695					700				
Pro	Pro	Ser	Pro	Ser	Gly	His	Pro	Asp	Gln	Thr	Gly	Thr	Asn	Gln	His
	705				710					715					720
Phe	Val	Met	Val	Glu	Val	His	Arg	Pro	Asp	Ser	Glu	Pro	Asp	Val	Asn
				725					730					735	
Glu	Val	Arg	Ala	Leu	Pro	Gln	Thr	Arg	Thr	Ala	Ser	Thr	Leu	Ser	Gln
			740					745						750	
Leu	Ser	Asp	Ser	Gly	Gln	Thr	Leu	Ser	Glu	Asp	Ser	Gly	Val	Asp	Ala
		755					760					765			
Gly	Glu	Ala	Glu	Ala	Ser	Ala	Pro	Gly	Arg	Gly	Arg	Gln	Ser	Val	Ser
	770					775					780				
Thr	Lys	Ser	Arg	Ser	Ser	Lys	Glu	Leu	Pro	Arg	Asn	Glu	Arg	Pro	Thr
	785				790					795					800
Asp	Gly	Ala	Asn	Lys	Pro	Pro	Gly	Leu	Leu	Glu	Pro	Thr	Ser	Thr	Leu
			805					810						815	
Val	Arg	Val	Lys	Lys	Ser	Ala	Ala	Thr	Leu	Gly	Ile	Ala	Ile	Glu	Gly
			820					825						830	
Gly	Ala	Asn	Thr	Arg	Gln	Pro	Leu	Pro	Arg	Ile	Val	Thr	Ile	Gln	Arg
		835					840						845		
Gly	Gly	Ser	Ala	His	Asn	Cys	Gly	Gln	Leu	Lys	Val	Gly	His	Val	Ile
	850				855						860				
Leu	Glu	Val	Asn	Gly	Leu	Thr	Leu	Arg	Gly	Lys	Glu	His	Arg	Glu	Ala
	865				870					875					880
Ala	Arg	Ile	Ile	Ala	Glu	Ala	Phe	Lys	Thr	Lys	Asp	Arg	Asp	Tyr	Ile
			885					890						895	
Asp	Phe	Leu	Val	Thr	Glu	Phe	Asn	Val	Met	Leu					
		900						905							

<210> SEQ ID NO 11  
 <211> LENGTH: 201  
 <212> TYPE: DNA  
 <213> ORGANISM: Homo sapien

<400> SEQUENCE: 11  
 ctgggcagcc tcgcctcttg gcctctgcag gtcctctgt atgggaggcc agagtttctg 60  
 tcactaactt ttctaaact cacaatgtct agaggtgggt rcgcttttcc acgcagtgga 120  
 acatgacttt tctttgaatc tctggcaggt ctgtgtgttc tgtcacttgc tcgcccagac 180  
 agtggctttg ctggctatec c 201

<210> SEQ ID NO 12  
 <211> LENGTH: 201  
 <212> TYPE: DNA

-continued

---

```

<213> ORGANISM: Homo sapien
<400> SEQUENCE: 12
ctgggcagcc tcgcctcttg gcctctgcag gtgcctctgt atgggaggcc agagtttctg    60
tcactaactt tttctaagct cacaatgtct agaggtgggt rcgcttttcc acgcagtgga    120
acatgacttt tctttgaatc tctggcaggt ctgtgtgttc tgtcacttgc tcgcccagac    180
agtggctttg ctggtcaccc c                                201

<210> SEQ ID NO 13
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
<400> SEQUENCE: 13
ctctcccagc tctcggacag cgggcagact ctaagcgagg acagtgggtg ggatgctggc    60
gaggcagagg ccagcgcccc aggccgagga aggcagtcgg ygtccaccaa gacaggagt    120
agcaaggagc tgctcggaa cgagaggccc acagatgggg ccaacaaacc gcctggactt    180
ctggagccca cgtccactct g                                201

<210> SEQ ID NO 14
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
<400> SEQUENCE: 14
ctctcccagc tctcggacag cgggcagact ctaagcgagg acagtgggtg ggatgctggc    60
gaggcagagg ccagcgcccc aggccgagga aggcagtcgg ygtccaccaa gacaggagt    120
agcaaggagc tgctcggaa cgagaggccc acagatgggg ccaacaaacc gcctggactt    180
ctggagccca cgtccactct g                                201

<210> SEQ ID NO 15
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
<400> SEQUENCE: 15
ctctcccagc tctcggacag cgggcagact ctaagcgagg acagtgggtg ggatgctggc    60
gaggcagagg ccagcgcccc aggccgagga aggcagtcgg ygtccaccaa gacaggagt    120
agcaaggagc tgctcggaa cgagaggccc acagatgggg ccaacaaacc gcctggactt    180
ctggagccca cgtccactct g                                201

<210> SEQ ID NO 16
<211> LENGTH: 10000
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
<400> SEQUENCE: 16
tgtacagtcc acccttggtta catgtgggag attcctgaga cccctctgtg tttacaaatg    60
agtgaaacag acctgcctga cccatggtgc cctgtgaggt taacaccatc gctgcagacg    120
ccatgaggta tccccacgac tacacaggta cttgtgggcc tgaggagcgc tctgcaagcc    180
gggcccccat ggtctgcaag ttgctccagc tgcacccct ctgctcgtct gcctccctc    240

```

-continued

tgccctttgc ctgtcctcca tcacagcatt ccaactctgt accttcattt ccagagaaaag	300
gggtttttgt agattttcct cccattccat cttttctgtt cttattttca ctctcctggc	360
ttcatttatt tctgttttta gtacgtttcg ttttctacc atgtctgtca catagaccaa	420
aacatcagga ctgtgtgaaa aagatagctg gacagcaccg accatggagc ggaactggaa	480
ggaggaaggc ggtgcaggtc ctcagcagac gctgtcccgc tegtgccaca ggcagattcc	540
agcagcctct ggtacagag ccttctccaa aaagcaaac ccagagacag agcctcacgg	600
gacaagagtc gcctgtgtct atgtagtgtt aggaatgaat ggcagcagat tattcttttg	660
agcatattta gaaagtggaa gccttaaaact agtagctttg tttctatggg gtttttaggg	720
ggttggggaa attgctataa ttgtataaag cagtgtttgc caaagtgttc tggaaataga	780
gctgctgaga tgtgcctgc aaaaacaagg gttccctggg caggtccttg ggggagtgtc	840
gtccccaggg tggcctgctg gggagggtag tcaccacaca ggctgagcgg gcctaggttg	900
agagtctgt tgaactctgc acacggcact taggcttttt gaccgcagac gttttttttt	960
tccacgtagc acctaatata catccctggg aactaactag tggtcagtgg aacacattta	1020
gaagcatggt gtacattttg gttccccat ccccttttct gaaaagtga gactagaaca	1080
atactcggca cagttctggc tcacgcctag gttcttaggc cttttgtctg gggcactgca	1140
tcgtgttggc ccagaccctg agacggccag tgtcgtttc tgcacagcag gacagccatg	1200
cccactgta cagggcctca gcctccctct cttgctgtgc gtctcctga gctctgtcct	1260
ggaggtgcca ggtgtgtcag tgacagccac cccctcacc ttctgcgttc tgggctgtct	1320
taggtctgtg gttgttttcc atccaagaga tccacggacc gtcccgagtg agcgcctctg	1380
gccgcaggcc cttccttct tgtgctctc ccatgtgcta ggctttgact ccatttctgt	1440
ggtgacacgt tgtgtattct caaccctgc ctttggtttt ttcagcccta gctctgctag	1500
cccagtggtg aagccctctg gatcagcagt taacatgggc tctgtgcagg gacactacgt	1560
gcaacaggta gaagatggct ttccagacc ttcagccctg gacacttagg cccgtctcca	1620
agcgccaaaa gagaagggag tgtccaacct atctgagcgc cccctgtcta aagacagcca	1680
tgccctagcg gttagagcac agaaccggga gctgggccgc cgcgtccct cctggcgaag	1740
gcatttctcc tcttgacggg cgtctggatc cttccttctc ctgcagagaa gggaggatgc	1800
tgattcctct gaggaggtgg tgcgtgaaag tggaccagcc ctggcaggtc ggggctgagg	1860
cgccctggga gactcgggta ggtttggggg tcggctttag gagccggggg gcgctcgtg	1920
gttctagtca gtgggtgtgc tcatctttgt cttggctggg tgggagcaca ggtgaaatta	1980
agttttccca ttttataatg aaaaggacat cgcagtgtct tttctgcac gactattct	2040
cttctctatt ctctaaaacg tgcctgggct cagagcttaa gccctcaga cgtgggtttt	2100
taatcagtca ctgtgcctgg ctggtagaga tgggaagtga ttgctctcat tggttcccct	2160
aggggtcatg aaagtgtgtt ggaatgagct tctaaagctt tctccagag ttcagacctg	2220
gtggggggca gatgcctgg ctgctctgtg acttccccac cccactcccc gtgtagagcc	2280
acagaccagg ggagccctga ctaggaatgg ggccccaggg aaaggttaca gagtgggaat	2340
ggtggggggg ttggctgaat cgcggccaga gatctgtggg ggccaagtct gcctagggcc	2400
ccagcagaca gtcgggggct ccgagacgtg cagggggcac ttggatctac ttgaaggggc	2460
ctttttgccc tgtggggcag gaggtgcact gatttgaaaa ggttcttcca gttggccttg	2520

-continued

---

ggaccagaac	tgettccgcc	ctgggatgtg	ccatcgcgga	attctttgtg	gaggaccgg	2580
tgtcttcctg	cctctccgaa	tcactgcact	cagggcttat	ggggaggag	cagagacggc	2640
actcgcgcca	ctttccttac	taggagggg	gtttctaagt	ctggtccttc	aaccaaggga	2700
ctgacaactt	aggggagctt	tccttgcaac	tcaggagcag	gaagggagt	gtgggaacta	2760
tgtatccaaa	cacgtttcag	cgtcagggtc	actttgtttt	tccattcat	ctgtccaaag	2820
gcaaaacaac	gagtggatga	gaaaccagc	ctgggagccg	tgaagctgca	ggaggcccc	2880
tcggctgcct	cccagatgaa	gcgaaccgga	gcgatcaagc	ctcgggctgt	caaagtggag	2940
gagagtaagg	cctgacagt	cctggctgcc	acctcgcctc	tccctactga	ggacggtgcc	3000
gccatgctgc	ctcgacacag	ccgacactcg	ggagcctcac	cagatccacc	gtccaaatgc	3060
gtggcccaga	ctgagagacc	tccctcctct	ccactcccga	aagctccgtt	gtcaaccagc	3120
ttgcaccctg	ggatatatgg	cattgacccg	cttgctttga	tacgaaacaa	aaaagcagac	3180
gactccttca	tcccatctgc	tccctaccgt	actgtggagt	gacgcctcct	gtgcagtgca	3240
gatttgccct	cctgcctcc	tccctgtcct	gccgcgcagc	caggggcctt	tctcagcagt	3300
gcttccggcc	cagccgcccc	tccctaggca	cagtgatttg	gcagcagggt	cattttactt	3360
tgaggctttt	tgttttaaaa	tgtagccaag	gtttttacaa	aggggaaag	aaaagaaaac	3420
aaaaacgcaa	gctccatgtg	tatagctgaa	cttttatatg	tttcttgcca	gcccctccgc	3480
tcccttccat	ctctagcctc	tgctcctgtt	agtttgatac	gtcactgcag	taccttaaga	3540
ggtgactctt	aagaatgcat	cccctcctga	ttcctcagct	ggttcaccct	tgaggttatt	3600
tgcaaaaaga	aaaggagggt	cttgagggca	ccgattgcga	gcattctggt	gcctggctcc	3660
ccgcctggga	agcagatggg	tgctcagagc	agcaggcagg	ttgggggagg	gggggggtca	3720
tagttgggtt	ccagctcctg	gcttgatgag	cccagggcgc	ttacaggcag	cccataaggt	3780
tgatgacagt	tttagcatga	gaatcacaca	gggtccctgt	cctgggctcc	tctaaagcca	3840
gtggatgtgc	tggccaccag	agacaaatca	tgagatggc	tgctggtggc	tcccagggtg	3900
gcccagatgg	ggtgagctga	catactacag	gccccacca	ggccccgtgg	gctctgcttc	3960
tggggctcca	taccctgccc	tgcaagggtg	ctgtgttttt	cacacatttc	ttccctgaa	4020
gccttctgta	acctgtcatt	ttccttccct	cctcttcccg	agcctgctgc	ttctctgga	4080
cctgtctcca	cctcccacac	agctcatcgt	gaacaccact	tggtgatgga	gggagtggac	4140
ccgtgtgtgg	tcccgaagt	aggccactgg	gagtttgtcc	tttctcctct	ttgcttccct	4200
cccagcagca	gaccaggtt	gtcaggacag	gagggcctga	gctaagcagt	aggcatcagt	4260
ctcgtttgtc	ttcagacggc	ggggcagggt	ccagggtgag	gotgggtgga	gggctgacca	4320
aggccaaga	ggcctgcgca	gcctccggga	gggcagcttc	tccagccaga	ggcttgtgtg	4380
agccatcgtg	tgctgggctt	gtttttaagt	aagaaacaag	gaaatcactc	cagattctgt	4440
cattccaagg	aaagggaagg	ggacagttca	ggtttctcag	ctgttcttag	gggtcactga	4500
gcgtctacct	cctcctccag	aggaggtctg	ctcagaacac	ctagaggagg	gggcccggga	4560
tgcaaccccc	accagaggct	gccttcagcg	tctcacgggt	gcaggacagc	gctcaggctt	4620
gggctctaag	ctctgtgtct	agtgtagaac	atggggaagg	agcatcttag	gaactgctga	4680
agtaacttct	tactgtctct	acaattctaa	ggaagcggga	gaacggcctc	ctaccaacag	4740
cgcccccccc	agagctgcct	gggaaagggc	agttttactg	aaagggtcct	tactgttcac	4800

-continued

---

ctgcatcttt cagcagctcc cctcctgccc tcacctggtc ttttccctct ttatcccaag	4860
cctttatgct tgagtccctt ccccaggggc tgcccaccgc acagtccag gcattcccta	4920
cctgagcttc ttgtctgctt ttccttctcc cactgcaagc ggctgcttgt ggggcctggg	4980
atgagccctc tctgtcccca cggccctcc ttgccaagcc attcctgggt gagttcaggc	5040
ctgcgggagc cacacattca tctccacctg gacacttgag cgcgatggcc agaccctcc	5100
cacctgatgc ggtgggtgct gtgatttgc aaaagaaagc cttctggatg ctgttaagat	5160
gtacccttca ggtgaacctg gtatcagacc cacagtactt gctgtttgag aaaaaataaa	5220
aacaaaaagg tcacctgttc tccagccctt ttctcttacc tggattttcc ttcttttctc	5280
ctccccacc ccaataaaaa aaacaaaaaa cactagaatt tatttatatg tattgatgtt	5340
gtaggcttag gtgaaaaaaaa aagaagtaaa tgtttcactg ctctatttat atataatgtc	5400
tgaattaatt ctgtgcagga aaggccagga aattgcatgt gaagttcggg gcagtcacca	5460
cctgtgtgtg acctgagctg cagtctcttc gctgagatgc aggttttaaa tgagacttgg	5520
gggctgagg gcaggcctca ggcctcccag cgcaccaacc cctccttggg ctaatgaaat	5580
gcagttctta gtgcagagat gttttaaggc gcaatatac tcttctcttc ccgtggtttt	5640
agagccaagc tcaaggtag aggcagtagg gtcttatttt gttttcaaac ccccatcctc	5700
agagcgcaga tacatgcaga ggcttctgcc aggataccac ggggccttag tgggaacagg	5760
tggagaccag cacttccctt tctgctgct gaggtaggga ttgggggggc agaaccact	5820
cacttttgcc tgttaaagtt gccctcctga cgctggcagc tctgcttgg tcaactggga	5880
tgcggctcgt tgcacgcca ccagtggcct tgcggtattg tocaccatcc actagagtgg	5940
gatgaagtcc agagtgtggg tatacatctc agatgccat ctaccactg gggacttcaa	6000
tgccagctgc atttggtttg gttttcttaa ctggtggctt ctccccacag cgtttttgt	6060
tttttttaa acattcatat gtgtttcaaa cttggaattc atagacactc tggctctagg	6120
ttccttaagg gggaaaacaa aagatgactt tatttcacat tcaagaaat cagttcagtt	6180
ccaaagctgt ggtccttcca gccacttcta gggacactgg ggaacctgt taaacgttga	6240
catcagtct ctccagccgt gctgtcacc tcctatcttc tggatctgcc ttcgggatgg	6300
tcagtgcag cttctggaag ctgagcacac acaggtgcac agccatgctg tggctctggc	6360
tgctacggca gcatggcagc tctggtggag ccttctccct tgccatttgg tccccctgtg	6420
ccaagtagct gcaggctgcc cctcaaatct tcatttgtcc cttttcactt cctgcagaac	6480
aagcctgggt tagagggtct gctggaaatg gcctttgaag accaaggata ccaggatgtg	6540
tgcactctgt cgtgttctgt gatgaatggg aaacgtaggc ttccagaaag ccagctctct	6600
tctgaaatgt gacggaccta agcaggaagt catccaggac aggagtggct cagtgttggg	6660
gatggacgct gtcgcccagc catgctccac cagggccacc aatgtgtagt tggctggtgg	6720
tctcgggca tgtgagacct gctcttctc gtttccacc cacttgggtg cctccaggat	6780
ggtagtggca cctcagagc cccatcttca gcatgttctg aagcctcaga gtggaaattc	6840
ctgctaaggc tctgtgtgga cgcctttctc ccgtgatcta aaggggacac tgtactcaag	6900
cttttgacct catgcttctg ttagtaaaaa aggatttggg ggttttgtt gtttctgag	6960
agggttgtgt tttgtttttg tttccttttg tttatgtttt ggccttctct ctttctctt	7020
ccatgtagac cagatatttg aaagggcaga cgatggctag aggtgtaatg tgcagcttgt	7080

-continued

ttataccgga	ttttggaaa	cttaccttg	atgggaaatc	gaatcgtgga	ttcaccaggc	7140
cggctctggc	acactcacc	tcgccccttc	cctccggctc	agtaacctatt	gtttctcctt	7200
tcaaatatgt	gattgtacta	gctctttcca	tatgaaagaa	ttctccttat	ttaaataaaa	7260
aaagtttaaa	aagataaagc	tgaacctgct	ttctcagggt	gtacatccct	cgtttcttgg	7320
tcatttgacc	agagagacct	acagatcaaa	gaattctaata	tttacagagt	tcacacggaa	7380
attaggagtg	gcacggctga	aatgccttaa	cagaagctgt	gtgggtagag	aagggattga	7440
aaccctggc	taagaagcat	cttccaaatg	cggcagcatc	aagcactcca	ctctcttccc	7500
acggatggca	tctggttgg	cacagtgcct	gcctgtggcc	ccgtcacttg	ccgtgagaaa	7560
cgagctggtc	agaggaggta	agcatttgat	gtggcgtgtt	gtgtgtcac	ttctagcgg	7620
gcttttgag	gaagtggagg	aggattatgc	tgtgctttgg	ggtggggagg	gagcaagcca	7680
ggcctgggg	agcgaagggt	aaaaggtaag	aatgggctga	ggcagcagtt	ctcggttttg	7740
ccattgcctc	actgacctgg	gaaagacgt	ttcaccacct	gagtcagtct	ccttaatcat	7800
aaaacacaga	ttgtaatggt	atcgacctca	tggcgtagt	tggattcaaa	gtcaccagta	7860
gcttttagct	tagtgcttag	aaccagctaa	gaaaacttat	aaattacccc	caagaaataa	7920
aacgctgttt	ccatgaggaa	ccccagtccc	tcaacttctc	tctaagcctt	ttgactatgg	7980
aagagagttg	gggacctggc	catgggcatc	tcagctttaa	ctagccacac	tactgcccct	8040
gatttactgc	tccttaggga	caggtggctc	acaaaatctc	aaacggagac	cagcccctgaa	8100
agaggctagt	cctaaggttg	gggcttgca	agatggaggc	tgtaagtccct	gaggcaccac	8160
cactgttagc	atggcatgaa	ccaaagctta	acagcccagc	cccactccct	cctggggccc	8220
aggaaccct	cctgcccac	tgtaaataca	gttgctgca	ggctctggcc	acaaaccag	8280
gtgagcccca	gctccttgcc	tctctgtggg	atcttgggtg	agtcacataa	cccctcaagg	8340
gtaagccata	aagaagtaac	cacgaaattc	taactgggtg	catctaatac	tatggtgccc	8400
aacatggaag	aaaagtcttg	caggcagcca	agaaccataa	tgaaggttat	taaacctgca	8460
gtgtgcccct	ttggtctgta	tagattaaca	tacaaataga	ccaggagatc	aagaaggagc	8520
gaggctgaga	agagtgatcc	atcagcctgg	ttgcgttgct	gtcagttacc	ctgctctggg	8580
tgccgactgt	tcagggaaaa	catgggaagg	ggtgcattga	cccacccaaa	agtaccatgt	8640
ctgctccaaa	gcaggccttt	gggaaggctc	ttgacctca	gcagaggggg	ctggagaggc	8700
tgatggatga	ggaggagccc	tgggaatgga	ggaagaagcc	tccagtcttc	tgccacggag	8760
tttaggggag	atgcccccc	tggagcctct	gatggtcatg	cagcccccc	tgatgtccag	8820
gtttttttg	ttttgtttt	tgttttgaga	cggtctcact	ttgtagtcca	ggctggagtg	8880
cagtgggtgag	accaaggctc	actgcagcct	cgacctctct	gcccccaagt	atctcctgcc	8940
tcagcctccc	ttgtagctgg	gactgcaggc	agacgccaag	gggccccgct	aatttttgta	9000
tttttttgta	gagacgggg	ctcgtctgtg	tgcccaggct	ctcctggact	caagcgggct	9060
cacggatcct	cccacctcgg	ccacccaaag	tgctgggggt	acaggcgtga	gccaccgccc	9120
ccggccctag	gtccgtttta	aatgccctta	ccctcagcca	cttcaccgct	agatgtatgt	9180
gcaagcagga	tcgtggcatc	gatgtttttc	ggcaaggaaa	ctaggagcgc	ccgcgggcgt	9240
gggggcaagt	gcctgtggtc	tcagctgctc	tgaggccga	ggaggaggga	tcgcttcggg	9300
cagagaggtc	gaggctgcag	caagccgtgg	tcctgccact	gtactccagc	ctgggcgaca	9360

-continued

---

```

gagcgagacc ccggtgtctta aaaaaaaaaa taaggtcgga cgcagtggtc cagcctgtga 9420
atcccagcac tttgggaggg cgaggtgggc ggatcacttc aggtcaggag ttcgagacca 9480
gcctggccaa catggtgaaa ccctgtctct actacaaata caaaaattaa cggggcgtgg 9540
tggcgtgcgc cttcaatctc agctactcgg gaggtgagg caggagaatc gcttgaacct 9600
gggaagcggg ggttgcggtg agccgagatc gcgccattgc actccagcct gggcaacaag 9660
agcgaaaact cgtctcaaaa aaaaagaaaa gaaaagaaaa gaaaaaaact aggagcaaac 9720
tggcgcgccg gcgagaacgc cgttgtgcag acgcagaatt aagatcttca cgccagatgg 9780
agtgaaaaag caaggagcag aagatgggtg ggagcccact gcgcttttta aaatctgcac 9840
gtctgcttag agtacacaga aatcctggca cacaaaacag tagcaccoga gagagggggag 9900
gtcgcgcgct ttcacacctg cctctctccc gggaccggcg ccctgcgccg ctgcgcacgc 9960
gcgtcacctc ctcccagtgc gcacgcgcgc tgctgcccgt 10000

```

```

<210> SEQ ID NO 17
<211> LENGTH: 10000
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

```

```

<400> SEQUENCE: 17

```

```

ctccagcctg ggcacaacag taagacgctt gtcaaaaaa aaaaaaaaaa aatccacag 60
gagtctgagg gaaagaggag gcaggtgagg atttccagag tgaaccaaga agaaatgatg 120
aagcctggct agggaagtag agaaccaagt tccagcctga gtgaccagga ataagcacct 180
tcctttctct gagcctttga acctgcggct taatgagatt gtctctgagg ttatcagaga 240
ttatcttcta atgctaacc tcctggattc tttattccat ttgcccacc aggcagaaat 300
gccatgcaga caggaggctg cagagcacag gtgtaagc acatgggctt tgtgggcaaa 360
cagagaactc catgttctcc cttggaagct gtgtggcctt gggcaagctg cttaacctct 420
ctgaatccca gattccacat ctgtaacatg gaccaataat tattttactg ggttcttggg 480
atgacggatg agactatgtg ccaaaagtcc ttggccccag tgcccagctg gcacagaatg 540
actgttcaca aataaatggc cttgcctttt ccgtgaaacg ccctacaacc tgctccctgc 600
aggactgagt ttaatttttc tcatccacag ctccacaat atcaacccca ggtctagcga 660
tgagtctctg acacaaaaag ggagaagcag ttttcccaag tggagggagg tgctgacatt 720
tccctgtgac cctcttcctc gcgaccctta aaccagttcc agggctcact gaccagcta 780
gaaactctgt cctcactctg agtagttaag gagcagaaca aacagcccct tctcagccaa 840
ctgagataag ggttatgagg aggggtggc catattgagg aaaacacata cataaattca 900
aattcatgct cttttcaaa agtcaaaaca accgtgcacg tgccaggaag ggactgctaa 960
cagcaaaact atactgaaca cctattatgt gtcaaatacc ttataaaagc ctttccatgg 1020
actcagatct cattatntaa tcctcactat aatccaagga gggaaaaatt atgatgccca 1080
ttttatttta taggtgagga agctggtttg aaccaagctg tttgaatctg gacagtataa 1140
tccttttttt tttttttttt tggagatgga gtctcgtctc gttgccagc ctggagggca 1200
gtggcgcgat ctcagctcag tcgaaagctc gcctcccggg ttcagccat tctcctgcct 1260
cagcctcctg agtagctgca attacaggca cctgccacca cggccagcta attttttttt 1320
tttttttgta ttttttagtag agacaggggtt tcaactgtgtt agccaggatg gtctcgatct 1380

```

-continued

---

cctgacctcg	tgatccacc	gcctcggcct	cccaaagtgc	tgggattaca	ggcgtgagcc	1440
accgtgccc	gcctcgaca	gtgtgatctt	aatcactttg	caactgaaac	ttgcaatagg	1500
taaaaatggtt	ggaacaata	gcagtgtccc	gtaacaggtg	atcattgaaa	gaaattgtga	1560
tagttccact	caaaggaata	atattagctg	ccgttaacag	tcatgctatt	gaagaatctc	1620
aaataatgta	aaaattttgt	tatgttaagt	ggacaagcca	gttataaaat	aatatatatc	1680
tagtttgagt	ccacttaaaa	aaaaaacagt	aaaatatgga	ggaagaggag	gaggaagaaa	1740
ggaaggagac	aggataggaa	aaagaaagag	agagaaagca	aaagaaagag	aaagacatgc	1800
tttaaaatgt	ttacaacagc	ccatgcgccc	tggctcacac	ctgtaatccc	agcattttgg	1860
gaggctgagg	cgggaggatc	acctgaggtc	aggagtctga	gaccagcctg	gtcaacatgg	1920
cgaaacccc	tctctactaa	aaatacaaaa	attagctggg	tgtggtgacg	tgacacctga	1980
atcccagcta	ctcgggagac	tgaggcagga	gaactgcctc	aacctgggag	gcagagggtg	2040
cagtgagcca	agatcgtgcc	ccccactccc	agcctaggcg	atgacagtga	aaactgtctca	2100
attaaaaaaa	aaaaaagttt	acaatcatta	tcactaggta	gtagaattat	aggcaagttt	2160
ttttgatagg	gaagtttacc	atttttcctt	ttgggtattt	ttcaaatata	acagttgctt	2220
taaaaataag	aatgaaatga	acatgtacta	ttcttaaaaa	caaaacaaac	gaataccacc	2280
tcacctctgt	gtgggtcttc	cttctgaact	tgggcacctc	cgtcttcagt	tgccccctct	2340
gtgaaaggcg	aaatgtatcg	ttgggttctt	tgaggccctt	tacagctctg	acatcctata	2400
acattctgta	acccccacac	cttctcctcc	tctgtttaa	ctctaataca	atatgttgtt	2460
ttctgcaata	cagccattat	gcaacatgtg	ggaacacatg	aaaatcacta	aaaggaaag	2520
aagcaatcag	cgcccagcct	cagccacttt	ttgcatttgg	aggcatggca	tgtgttgctg	2580
gattggatcc	tcataaataa	ttaggcattt	aacagatgaa	gaaagggtca	gagaggttac	2640
acaacttgcc	caaagtcaca	cagcatgtta	aggcacagtt	gggtctgtac	ccacaaataa	2700
agcattatcc	tccagatgaa	tccttagtgg	gggactgagt	gtgcctctta	gaccctgaga	2760
cctcccctcc	tcagtctctg	tccttgctg	tggccccaa	accccgagga	gactgacagc	2820
tggacatccc	actctggccc	tgatcccagg	ctccttcccc	agaaaaagct	gagctggcag	2880
tgggggggct	ggggggagaa	tgctggcaag	gacctagggg	ctggtgtggg	agcggcactg	2940
ggcgcagaag	gtctcaactg	gcaatggagc	tggagtctgc	ctgacatttg	gcaaggaagc	3000
ccagagccaa	gatggaacag	caagctgggg	tgacagggg	agggaggagt	gagaatgtga	3060
gactttgggg	gcaggaagcc	tgggtagatc	agccctttgc	tgttttgtt	cagaacttct	3120
ctgatctct	catctacaga	cagaggcttc	tccttccctc	aggcatgcat	gtgtaaagta	3180
acaagataat	gtgtttgtac	gcaggacaca	ctcagaaaat	gaggagcttc	tgttagcctc	3240
ctggctcccc	tggccaactg	cgcacctct	tccctatcac	ccacctgac	ttacccccca	3300
atccctggc	acaatgacaa	atcttagggc	accatcctgt	caggtttttc	ccctgcttaa	3360
cgaccatccg	tggctcctgt	cactttcagg	ttaaaggcca	gacattttgg	cctggaattc	3420
aggctcatac	caacctgttc	caaagtcagg	ggctgctgca	ctgcataact	cacaccgtac	3480
actacaaaag	ccatctcata	aagggtttct	tcccaaatcg	catactttag	gggagctgca	3540
gtggggcgca	ggccttctgg	cctggctcca	cttggctctg	ccaaaagcc	cccaactcca	3600
ggctgccaag	cagaagcaag	gctggtcttc	catcccccaa	caccttcac	ctgcagctgt	3660

-continued

ctttaaaaaa aaaaaaaaaa aaaaaaagta agaaaaagcc ctgatttgty gtggttacta	3720
atlttgatgg tgttaatac cttactgaa attggccatg ccaatttcaa gctattgata	3780
tcaactgaatg ctgcactggg aagagatggg cacaaatccc tctcccagagc tgacaccagc	3840
acaccctgca caccccaag taaggatgcy ctaacaccag tgagaagctg tgtgactctc	3900
gtcaagtggc tttacctctc tgaacctcag gctttctatc tgtacaagag ggacaacaat	3960
aacacccact tgcacaacat cacgtaactg atataatcta gatgatcaac aacaatgttt	4020
tgagcaccta ctctgtgtag ggctctgtac taggtagaca ctagggaata acaatggtga	4080
caaaacagat aagagctcct gccctcccaa acctcacttc ttgcagtaaa agccctggca	4140
cagcggctag aatgtcgcac tcaataaata ttacctatga tattagctaa taataccaat	4200
ggattatttg ttgtcattat tcttctcttt ctttgagcct cagtttctctc atttatttag	4260
aatgtggact gagcaaaagt ttcttatgat gccacctcac tcaagcctta aaatggtggt	4320
aaagtagatg ggcgcggtgg ctacgcctg taatcccagc acttcaggag gccgaggtgg	4380
gaggatcatg aggtcaggaa atcgagacca tcccggccaa catggtgaaa ccccatctct	4440
actaaaaata caaaaattag ccagggtgag tggcacacgc ctgtgatccc agctactcag	4500
gaggctgagc caggagaatc gcttgaacct ggccggcaga tgttcccgtg agccgagatc	4560
gcccactgca actccagcct ggagacagag cttagactcca tctcaaaaa aaaaagaaaa	4620
aaaaaatggt ggtaaagtgg gcattaccaa cctcattttg tggaaggaa aactgaggca	4680
ctcagaactg aatccaggct ggggatctaa actcaggtct tcttactcct ggtctccaga	4740
tcattcccac ctccaggagc tgaggaaaca taaaaaggac caggaattcg actggggatg	4800
tgaatgcttt aaaaatactt ttccttctc acgccccac ggggggaaac tgaagtgtcc	4860
atgttgtgca ccaatttcca aaacaccctc tcagtatgac ggccctgtag ctgggaccat	4920
tcccagctc tctggaaac cctatcctag cgcctctggc agagtctgca ggacggtggg	4980
agctgatgaa atccccctgc acaaacagc cccaatcaca tgttttaaat gcaaaccggc	5040
tggactttag gaggcaagg ccccacgtg acatcgcctc cctcctgag aagaactgtg	5100
agacgactgc cccaggaata aaccagagt cccaaaacac cataggttcc gctcccctgt	5160
ttgttagtaa gtaaacaga cttaggccct ctctaacaga tgggcagcca ttgttttatt	5220
aacaaaaacag ccccaataa aaaccacca attccagaga aggcgcctgg aaaagtctt	5280
tgggaagtcc ctgtatgat ctaacttag tctttcatgc ttcagttcca gccatttct	5340
tctcctctg tcttagagga caggaacaat gtctctttc actgcagcct gtctcagctt	5400
ctctctgctc caacatagcc caggtttggt atcagcaatt tccctctcg actatctcag	5460
aagcaggtct ttaaacatag gtaatgatg actccaatcc gtcatttctt tcaacaaaag	5520
cccgggaaag gagctttgga agtgtcgtct aaggtcacgc agcaagcaga ctctggttct	5580
gcctgtgcca gccccatcac actggggaga gccagctctc gcgctcctc gctgtgcacc	5640
cttgggccag tgcgcccc tctctggacc tcattctcct ccgaggtgt accaagaggg	5700
aactggactt atttttgag cccgttcagt tctaccatt gtcaggctg cctaccgac	5760
actggccctt gttctacct ccccccgag aaaaaagag gaaaagaaac ttgcgttagg	5820
cgctcgggaa gcgcgcgcgg ctccgagggc cgcaggtctg cgcgattgt tgcgggtga	5880
acgaatctca cacaaatccc agacagaaag cctcacctcc ccagtaacca cagcaaccg	5940

-continued

---

ggccacgcct ccagcccaaa cacagcccga ggccccgcca aggcccaccc ccgagctccc	6000
ccgggacccc tcccctccct gccaccagcc ccagctggaa agagaaaaag ccaacgcccc	6060
cattgtgcag aggtggagac tgaggccgga ggggtggagac tgaggaaccc cgcggccggc	6120
acggtttgga ctggggccca taccgcccgt ccagggtccc ctcccattcc cggagcggtg	6180
atgcgtctcc tccccctgct gcgtgggggt tccacgtgga aaagaccgaa agcagcgcgcg	6240
tcccagtgga acggagccca ctggcgctgt ccaactctccc tcccattct cccgacggag	6300
aaactgagtc ccaacgaggg acagggacca ccccgacgc cccgcgagcc cctccccgag	6360
tctcacctgc tggcgccgc cgagcgcggg tggtttctct gagttgagcg gcaaaaaag	6420
gcctccaatc cggggcgggg gcggggcggt ggctccttg cataatccct gaggcccgc	6480
cgcgacagca gggggcgccc gggggggggc ggggcctctg ggagcgcacc aatcaccgaa	6540
agggcggcgc cacggcccgc ccccaggaac tcggacatag ggagcgcaaa ccctgctgca	6600
ggatgaaaat ctacagacct ggaggctgtg tgtgcctact gcaatcagcc ccaccgcaga	6660
cccgagatc gaggcccggc ctgocggagg taaggcattt acgcaagatc gcacagagtg	6720
gctcacgcct gtaatcccag cactttggga ggccggcag atcacctgag gtcaggagt	6780
cgagactagc ctggccaaca tggtgaccgt cccccaccc ccaccagtct ctactaaaac	6840
tacaaaaatt agctgggctg ggtggcgc atctgtaatc ccagctactt gggaggctga	6900
ggcacgagaa tcgcttgaac ttgggaggcg gaggttgcgg tgagccaaga ttgcgtcact	6960
gcactccagc ctggcgaca gagcaaggct ccatcaaaaa aaaaaagaaa gaaaaagaa	7020
aaaagaaaa aaaaagatcg cacagagcca cagtgaggag gaaatcacia acttccatgg	7080
ctgaaacggc ctgcttagta acagtaataa taaacttttc ttgatcacia attatgggtc	7140
agacctgttg gaagctcttc tgctgttcag tcgtctccta tgaggtaggt aggtattatc	7200
accaccgccc ccaagtatat tcttgtccaa ctccccatca ggcaagcatg gaaaccagc	7260
ccgaagacgg acaagaactt ttccaaggtc ccgctgtgag tgcagcacag acatcatacc	7320
ttgtccacca cacacacacc ttgaaggatg cacggtgtcc aactcacctg ggctctttgc	7380
cccagctctg ccacttgctg ccagtgagcc tcaggcaaat gggattcata cccacatctt	7440
atgcggttgt tgagagtaaa taaataaaca taatggcagc cccgtcagaa ataatcagt	7500
ggtactaatt taaagagcac cttgaccgac cctattaaca caagagggtt aataagtcct	7560
gatgcaagaa gagacttgcc caaggctcct catccagctg catccaaagt tgaatgcttt	7620
ggaatttttc ctccaggtag aaagcacttg ccagttcccc aggttgctctg ctactcccta	7680
gacacatatt tagtagtaag tttagcagca gcttctaccc actaagggcc tgctctgtga	7740
aaggggcaaa catttttagct aatccttctg ggctctggcc ctggaatgct agttaggaa	7800
gggccaagg tgtttttcca cacaggctcat ggaagtgaag tgtttttctg ggtttgagca	7860
gcatggcggc cctgactctt gaacagaatg gggaaaccga ggcttggtgc attgtgtaac	7920
ctgcccagca ttctccagca agggagaagc agaccctgag actccattcc ccccaacccc	7980
cactttggca tcatgagtaa aaatgtcagc gtcacaggta catgtggcca ggatggacct	8040
ctgacaaacc tcccccaact ctcccagtac ttgaagctgc cttagcctac cacactgggc	8100
ccttcaaaag acaccaggac aggtgcagga aagagcagac ccaggcttaa ggcagctgtg	8160
ggtccaaact tgtcctcctt agctgtgtga ctttgacaa ctaactggac ctttctaaga	8220

-continued

```

ccacttgcaa aggaatgttt ttaggttaaa agtttactg ggaggggaac ccatacaactc 8280
aatgggctac agcctcagag tcagggccag gaattccaga agcaacaagt ccagaaactc 8340
attcctgtcc tactccttcc aagtacagac atggggctgt gactaacagc tgctgggggt 8400
ggcagctcca ttgttagttt cctatggcta ctgtaacaaa ttaaacggc acaaatacct 8460
tctcttacag agggcagaag tctgaaatga gtctctgggc ttcaggtcca ggtgacagta 8520
gggcagcttt ctctctggagg ctcttgaggg gagaacccac tgccttgccct ttctctgctt 8580
ctagtggctg tctacattcc tcagctcgtg gtcccttcct tcatogtcac agcacatcac 8640
tgcagcaacc tgtttccatc atcacacacc ttgtcttctt ctgttaacat ctttctctgc 8700
ttcccattta taagagcctg gtgattatac tgggtccacc coactgtcaag actttaatca 8760
tgtccgcaaa atcccatttg ccatataaag taacattcca aagtccagc gatttgaatc 8820
tgggtatctt tgggagccat tatttagcct accaaacca ttttatagat ggggaactag 8880
aggctcagag aggtaaaatt acttgcccat gaccacaaa gctttggagt tgaatgccag 8940
gagatctgag gcctaacgaa gccctcctgc tgctagggtg gagggtcatt ccctgcagag 9000
ggtcaacagc accctgtcag gggatggcag ggggtccctc tgaaacacca gctgcctcag 9060
gggaatccat ggctcaggtc tcaggtggtc tgcccaggct ggaagagaa caaggacagc 9120
caagaacagg atctgggaca agagcatgga accaactgtc cctaaccac ctctggcaga 9180
actctggaca cagaactgga caaagcagga tttctcaaac tcttccatct tccagacaaa 9240
tattgaggac agaaggaata agcactgggg ccgggactga ggccgtacag tagaccatgt 9300
ttgcaacctt ttctctcgag cccatcccat atgttttgct ctgacacccg ctgtcctagg 9360
ctactggaac agcttgacc ccataggtag agagcaggaa gtgattggga atttatggcc 9420
tctctctgta ggctttgtct aaagactgac aggcacaaa gtccagccct ctggccttga 9480
gtggagacaa ctctaggcct agtgtacgcc ctgggaccc cctgggatca ggcagaagct 9540
gccttctgca gactttgcct gaaatctc acctccttgg ctctctccc ttccctgtcc 9600
tgcttcccca ctctctgtct ggtctgccct gagggcaact cttgacgatt gattgggtac 9660
aaatccttat ctccaaggca gcttctggaa aggaatgggc tgcactctgag atttaaaaca 9720
ttcatgcaca ttgtgcatto tgcctcataa tatacttgtg gggtttaatc taaaaatatt 9780
atctgtaaat agtcaagttt tgccttcaag ccttcaatta agaggctcca agatgtgagt 9840
gtggctcaag caccctgca cgccccaccg tgtatcccaa ttggagccat gcgggcagtc 9900
gggggtggtg tctgtggtgc tctgactgcc cccacctgtc ctgccaccc ctecccacat 9960
ggaacccca gccaccagg taaaccacac agcacgctga 10000

```

```

<210> SEQ ID NO 18
<211> LENGTH: 10000
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

```

```

<400> SEQUENCE: 18

```

```

ttttaaccag aatttacaat accataaagt tgcattttat gtgattttgt acacacacac 60
actcaactat aacgatgaat gaagtgtat ggtggatag tatagtgtct ctatgcccct 120
ttttattatt cttggagcat aatcttactg ctttgtatct ccaaggaggt tatagtgtct 180
gagccttggc tttcatgaca tatttttggc tactgccacg atttatttga tgccaaatc 240

```

-continued

ttggtcagta gtatggcttc caattattgt attccttaat ggaaaagagc tttctgttat	300
accaggaact gcataatgct caagattcct gaaatgtatt ttggttacia gtaaagacca	360
cctgtatatac aggatgttat tgtagttaat tcattacact aaaaagcagt taagccatgc	420
gcgttggctc acgcctgtaa tcccagcact ttgggaggcc aaggcaggtg gattacctga	480
ggttgggagt tcgagaccag cctgactcac atggagaaac cccatctcta ctaaaaatac	540
aaaattagcc aggctggtag gcgctgcct gtaatcccag ctactcggga agctgaggca	600
ggagaatcac ttgaacctgg gagtgaggg ttgccgtgag ccaagatcgc gccattgccc	660
tccagcctgg gcaacaggga ggaaactcca tctcaaaaaa aaaaaaaaaa aaaagcaatt	720
aatccagaat taatattcc ctgtgtgcct agtcccacac tctgaactt tttgagttat	780
tctaagacac aaatttcttt tcttttcttt tttttttttt tttttgagac agagtttcgc	840
ttttgttgcc caggctggag tgcaatggcg caatcttggc tcaccacaac ctccgcctcc	900
tgggttcaag cgattctcct gcctcagcct tcccagtag ttgggattac agacatgcac	960
caccatgccc ggctaatttt gtatttttag tagagacagg gtttctccat gttggtcagg	1020
ctggtctcga actcccagcc tcaggtgatc cgcaccctt ggcctccca agtgccggga	1080
ttacaggcat gagccactgc gccagccct ctaagacaca aatttctaaa agacatagac	1140
ttctaagtct caaggagcct acaagtttct taaaagctta tgatgtggcc tacaatcaaa	1200
tgcctaatac aattactcca aagcaagaa acatttttaa aagatgtatt taagatataa	1260
ctttttaaaa agatgtattt aagttataac tttgtaatag atgaaataat ctgttaatac	1320
ttgaaaaatt aatcacacgt gcaaaaagctt ggataggatt gtaggggtgt tttttaagc	1380
ttaagagggt gaggtagact aaaaggagg gagtgagggt aagggtggca gaagagatgg	1440
gcaatgcatt tcaactgtt ttgcctttct ttgccccct ggattagggg attggttgta	1500
gtagggcagt tccacttct gactttattc taagaaatgt aagcaggaaa ttaaaaaaaaa	1560
aaccacaagta gatccttttt taagaggagt ctaaagaaat acaacatgtg atcgcagtgc	1620
ttctacatca taaatccctg ggtctttttc ttggtttttt tttttttttt ttgagatgga	1680
gtctcactct gttgcccagg ctggagtgca gtggcgtgat ctccgctcac tgcaagctct	1740
gcctcccggg ttcacgccat tctcctgcct cagcctcctg agtagctggg actacagggg	1800
cccaccacca cgctctgta attttttttt tgtattttta gtagagtcgg ggtttcatcc	1860
tgtagccag gatggtctcg atctcctgac ctcatgatcc acccgccttg gcctctcaaa	1920
gtgctgggat tacaggcgtg aaccaccacg cccagtcctt tccctgtgtc tttttcactg	1980
ccatccttgt tactcaataa ataattaagt gaggcctggt gtggtggctc atgcctgtaa	2040
tcccacact ttgggagacc aaagcaggat gatcgcttgg ttaagacca tccctgggtaa	2100
cacagagaga cctcatttct acagaaaata aaaaattagc caggcatggt ggtacatagc	2160
tatagtccca gctacttgag aggctgagga gaaaggattg cttgagcctg ggaggttgag	2220
gctacagtgg gcggtgatca tgtcactgca ctccagcttg ggtgacagag tgagagacat	2280
tgtcaaaaac atttttttta aaaaaaaaaa agagagagaa aatagtagaa aatgcttaat	2340
caacacatta aattgcacc atattttgat atggttgtca aaaaagagta tatttgaca	2400
tattaagtct atataaacag aaggaccaca ttttataggt gtgtacattt agagctcttg	2460
gtatgtgaat acagcattga atggatcaaa gattattgtt gtctttgcaa atcttcaaac	2520

-continued

tgacagaacc ttttgaaaa atttatggca agtttgtctt tttttaaag caatgtttct	2580
ataaatccta ccaaagtgct atgaataagg ctgtaaatca ctttactggc ttgagtaatg	2640
tttatttcat gaaacctgac caggaaacct agaggcttcc actgatattg aatcttctgt	2700
aatatttaac acctatccaa cccaaaaaga tgaattttga tactgactaa aaatattttt	2760
accacaattt ttgctgaaa gaaaataaat agaaatcaat ggtgtaagt tggatggctc	2820
gaaagagggt cattagtcac cagaagagaa aaaaattaga agcactagaa aataaatcaa	2880
tgtggatata agttcagatt ttatctgtaa ttatgattgg tacatccatt tctaattgat	2940
gaataatgct gaccgcttct ttctgggtga cttttttctg tcttgcttga tgtttatttg	3000
acaaaacaga aagaaactct tttgaaatgac ttattctcct ttttaattat taggttggtg	3060
tgaaagtgat tgtggttttt gccattattt tggggccaac taatatattc aaatattctt	3120
acaaggaatg tagaatctgg tagccagctg cataaaatct caatagtcaa tgagtcctgt	3180
aggaagaaaa gtgtacatgt tatacaaaa ctaaatttgc acatgctagc aaaatgagtg	3240
acacatacac taccatgagg cacctccatg acaattcagt taggaacctg attgacattc	3300
tgagccttga gttttgcaaa tgctgtttag tgttaccat cttttgcttc aggtagtgc	3360
agaacctaaa tgaatttggc ccaggcataa aagaaattag gctaaaggca aaagtaaac	3420
tgtaatgcca gatcactggc cagtgacttg taagtgccta tttgcacact gcctgttttt	3480
acaaagaatt taatgcatc aataccacag ggctcaatac atgaaaattg ggggaaaaaa	3540
acctgagcca gagagaaagt aggactaggg aggacaggca aaaaacacag gcgatggcgg	3600
cttaaacctg ggatcaaaaa ctctagttc agatggccag gcaggtaatg tgaagcccct	3660
catctttaaa accttgactc acaaaagggc tgagagactg gagaaatctt gggagaaggt	3720
atctgagaga tactgcaaat cactcacagc tcagcaggaa gtgagtaatg tgcatttgct	3780
attttgttt tgaatattca tttctgctct gaaaaagaag aaaaatggga tgaaggaaaa	3840
tcataggatg tgttaagcat agataaggaa gtttagggacc taccagacca agaagacagc	3900
aatcttcat gatgcgatac atactgtatg ccagaagcat aatgaggcac tccatatgca	3960
ggaagtctta attcaccacc cgttgatgat gccattctg gctccacctg ttctctgccc	4020
aagaaacaaa cgaacaaaac aaaaccctt agctgctcat cagatcaact ggcaagctcc	4080
agctactggc caggcattct tgccccctg tatttgctg ggctcatatg ctactgttcc	4140
aaggttctgc cactagtact ctctttttc aatcctgcac accctgtcct taggtggctc	4200
catccccctc tatgcctgca ggccccctg aatatacacc aggttcactt tctccccatc	4260
cctcaggtac ccatcaccca tctcccagac atctcctgga cataatctag atgccccatg	4320
ggtaactcaa tcttcataat cctaagactg acctcatctt ctctcccaa actacctgta	4380
tttcagccta catggcagca agatgtacct aatcactcaa ccaaaacttg ggggtgcacc	4440
cggactgtgt ccttttcctc aggtaccaag gtctcaggtc ctgtgatttc tacttcctta	4500
aagtatctca ttagttttct ataccacca tctgtgctaa attagtatta attaactatc	4560
tttagagtga tacagccaca tgctgagcat gatggacaag gacctcaag ctctaccttg	4620
gttccccctc ccatctcatc tgagttctgg gcatgatgga agagactctt ctctctcca	4680
ctcttgctc ttgccattac ctgctctgg tcgtgattac ctacagatgc ctgaatgagt	4740
catagtcttt ctcatatcca ggattgtacc ctgtgtgtt tctctgctc agatgcccctc	4800

-continued

---

ctttccaacc cccactgtcc aataaccac caccacaatt gtccaccatt ttatcctgct	4860
tactottatt ctcttttcaa atttttcatg agatgtcagc ttctotagga aacttccctg	4920
aaatcccagc ctccctcca ccctcaccct ccactggccc ttatctgaat taagcactcc	4980
ccctgcttgt tttctctgct ttctgagget cctttctcgt tatgctgtca ctcaggggtc	5040
cagagggcag agatggaatg ctcaaattag gacaattcag ggaggtgtgg gagaggtgca	5100
gggggaccat aagggatgga gcaggaaccc aaggtagtaa tagcagagct ggcaccaacc	5160
ctaggctcaa aaagacaaag gagagagagt ggttaccaaa acccagaaaag agaaaacccg	5220
gtaaaagacg ggacctggtg aggagcagca acctgtgttt ggctgtggac tgtggcaatc	5280
ttgcagggag gagagcagag gaatgactat cttggcttca ctctccttcc ttctccttca	5340
tctcatgtca ggtgccccat tggctaaacc taccagcaac cagaggacaa ggaatgcatt	5400
aaaataatct tacggctcag agagtagatt agatgtagag tgacaaatgg attatattta	5460
gcacaatcgc aggtctattt gttgatccgt ctccctcatc agactgagtt acctgaggac	5520
agagaatggt tctcttataat ctgtatccct aacatttggg ggacactaga tactcaagaa	5580
atgcttatgg acgggatcag agaagatggt aaattgggca ttagatggag gtttgtgctt	5640
tctgagcttg ccacatgctt tctggcgaat ccaaggcaaa atgagaaaac gaatgagctc	5700
tgtacttttg tctggtaaaa gaaaatatga tttggtgaga agaaatgaat acttttagca	5760
ctcagtctg ggaggaactg gaatcagtca tgtgactgcc tctatcagag gctctcaact	5820
gogtaacgga cagaatcttt aatggtgaaa ccacaatgca gaaacattct tttttttttt	5880
tttttttttt tttttgagat ggagtcttgc tctgtcacc caggctggagt gcagtggcac	5940
catctcggtc cactgcaagc tccgctccc gggttcacc cttctcctg cctcagcctc	6000
ccaagtagct gggactacag gcacctgcca ccacacctgg ctaatttttt gtatttttag	6060
tagaaatggg gtttcaccgt gttagctagg atgctcttga tctcctgacc tcatgatctg	6120
cccacctcgg cctcccaaaag tgggaaacat tcttaaaaat attagccctt gataaaacca	6180
aagccatggt acaaaaaagc aaccaaaagg tggaaatgag ataatatggt ccactttgag	6240
ctgagaacat cttgccagaa aatgattgtc tcagacttca aatctgacct cctacataaa	6300
ccttcagtgc aatcctccct gtattatctc ttctctcatc agccaaaaga tggaaaggat	6360
aaactgagga acttacgctg gaggagcatg gattcattag atggcagaaa cctgggtttc	6420
tgaatgacta ctgtggagct gatctctgct tatgcccctg aaacctttgt gctgagagca	6480
agtaacacat ttttaatgtg ttaaacctct gggatttgtt gggctttata gcaatggtta	6540
ttttacccta atacgactgt aaaataatgt tctaaaatct acagtcgggt ttgtttaaaa	6600
tattatcgcc atcattatta taatccatag tccatatcat aattattttt atagaagtcc	6660
ttctcaactg gcactccta atttgtggct tcttatattg ttgccttcta tgtgtgagat	6720
gcctctctga gtcctttttg ctagggacat gatagaaaat aaaatatcca acattctggt	6780
gtattttata atgtacaaag tctctgttca tgcattatct ctttagattc tcccaacaaa	6840
cccaaggaat atgtaagggt ctataatcat cttcatttca caggtgatga aactgagacg	6900
tttacatgta aagtggaaat aaccacttat ttatttattt atttattaaa taaaaacccc	6960
gctaatttat ttaaagatta attgagagaa ttatttcatc tgtctggaac ctggtatgtg	7020
ttaagagaaa agttttctta atttttcttt ttaagaagat aatctattgc tacataacaa	7080

-continued

---

accactctga aataatgact taaaacatta aaagtcattg attttgctca cagaacttga	7140
actggaagat ataaacaact gggagtgagg gtggggctgg acagctggct tctctgggca	7200
tctctcgttc tttctgtggt cttgtcacat gtctctacat ggcagcccta gggtaatcat	7260
gacagctgaa ggctttagag tgggtggagg gggagagggg gggagggaaa gagagagaga	7320
gagagagaga gaaagagaag caagcaggca gaagctctaa tgccttatga cctaaccttg	7380
gaaataaatt acactgtgcc actgccacta ccttttctta gaagcaagca gatagggttg	7440
gtgaatattt attgagaggg gaattagact ccacctcttg attgggggag tgtcaaagaa	7500
tttgtgaata tgtttttaa tgaccacatt tcccctttct aaaatgcttg tctaaattgg	7560
tgtacacaaa aagctcgggg gttagtact catccaaggt tattaagctg ttaaattggt	7620
ccaccaatac tagatctaatt tttgtgactc caaatttcat gctcttttcc actactccag	7680
ataatctccg gtgaacatca ttgtccatta gggtagttaa gacattctag ctgccaaata	7740
gattcagtcc ttttaattgc ctcggacaat atatttttta ttttgcaaaa aataaatatt	7800
accatatagg cagtcatcca taaacacott ccaagtttat cccaacatc tgtgttttaa	7860
atTTTTtctc ccaacagatg tattcatttg cattttccca gatgaatcac tttttgttg	7920
ttgttgtgtg tatttttgac tgtatttctc acctctcaag acccttttgt attatcaagc	7980
ccactgctta agctaatacc atctcccaat ttggggactt ccccttgatt cctgtgtgcc	8040
cttctgtggt gtcattaatg caatgttatt taagacgagg atttggctgg gcgcagtggc	8100
tcacgcctgt aatcccagca ctgtgggagg ccgagccggg aggatcacct gaggtcagga	8160
gttcaagacc agcctgacca acatggtgaa accacgtctc tactaaaaat acaaaattag	8220
ccgggcatag tcacatgcct gtaatcccag ctactgggta gcctgaggca ggagaatcgc	8280
ttgaaccggg gaggcggagg cggagtttgc agtgagccaa gatttcacaa ctgcactcca	8340
gtctgggcca caagagcga aaccogtctc aaaaaaaaa aaaagactag gatttgacat	8400
aaggcctgag gggatttctt ttgttttgtt ttgccttgtt ttcaagaggc caaaatcttc	8460
acagttgaaa atttctgttg aaccacagag atttgaacca actcagttta gaaagcctgg	8520
ggatttgaac aacggtatgg atcggaaatc tcttcatctg tcagttttca tcattctagg	8580
cagtaaaata gatttccctt taggagcttt tcaccgtttg gggttctcca gcagtgggat	8640
gtggggaatc aacccttctt cgtctccacc caaacattag gtggggagca ggggtgggaa	8700
gtagagaaag tggatagagg tctccagtgg atatgggac tttgtgtaga ccagcacagt	8760
cctcagaaat ctcatgcaag caacataggt actgttatat tttctagtgg ccacctttta	8820
aaaagtaaac aggtgaggcc gggcgcggtg gctcacgcct gtaatcccag cactttggga	8880
ggcccaggcg ggcggatcac gaggtcaaga gatggagacc atcctggtcg acacggtgaa	8940
acccogtctc tactaaaaat acaaaaatta gctgggcatg gtgacgcgcg actgtagtcc	9000
tagctactgg ggaggccgag gcaggagaat cacttgaacc ctggagggtg aggttgccac	9060
gctccactac actccagcct ggcgacagag tgagactccg tctcaaaaa aagaaagtaa	9120
acaggtgaaa ttaattttta taatatattt tgtttaaccc aacgtatcca aaatactatc	9180
atttgaagt gtaatgaata taaaaatatt catgagatat ttttattct catatccata	9240
ctgtottgga ctctaattgt tattttacac ttacagcaca attaatttgg gactagctac	9300
attcagctc aacaatagcc aatagcatat gggatagcgc aaataaactc tgcgtctctg	9360

-continued

---

```

ttgcttcttt gggctctgga gacctcaacc ctttcttcag attgcaaacc ttcttgccct 9420
caagcctcgg ctccaacacc agtcocggcag aggaaccacg totaatgagg tacgctccct 9480
tcctgccatt ctctattcca ttaacctggt tcgtggtaaa cgtaggactg atcctccaaa 9540
attaccttat taattagctt acatatttat tatctatctg tcccaccaga atgcagggtt 9600
ccggaaggca gggatttaaa aaaatctggt ttgttctatg tgattttccc ataccaagca 9660
ccgtgcccg cacaagctgg gatcccagta cacatctcgg gacggaagaa ccgtgtttcc 9720
ctagaacca gtcagagggc agcttagcaa tgtgtcacag gtggggcgcc cgcgttccgg 9780
gcgagcgac tggctcccc gccggcgtgg gtgtggggcg agtgggtgtg tgcgggggtg 9840
gocggtaga gocgcccag gagcccggag cgcggagctg ggaggagcag cgagcggccg 9900
gcagaaccg cagcggcgg ctggcagggc agctcggagg tgggtgggccc gcgcccag 9960
cccgttgca ggttccccat tggccgcctg ccggccgccc 10000

```

&lt;210&gt; SEQ ID NO 19

&lt;211&gt; LENGTH: 10000

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Homo sapien

&lt;400&gt; SEQUENCE: 19

```

catctgcagt cccaggatcc tgggacaggg cttatgaaca caaccactgt agtcagctca 60
cctgatccac agcctggcac ccccactgtc tggctaggga gcctcgaatg ggtcccaagg 120
ccaccctgct cctcagttac atcatctgca tagtagtggt ggttgtagg aattcaggag 180
ctgcagcata agggctcggc aggtcctaag tgcacagtaa atgccagtga ttcttaagag 240
tctgagctcc cattgtagag gcaagtaagc tgaggttcag agacagaaaa tgacttgccc 300
aagatcacc agctgggaag tgacagtgcc agggttggag ccctgggtga gctggttcca 360
caggccagag ctattctgct cctctccctg gaagacctcc caccctgtcc ccatgcctct 420
gcttctcct caccccaatt ccccgtgccc ttctaggata agtgtgagcc actggagaag 480
cagcacgaga aggagaggaa acaggaggag ggggaatcct agcaggacac agccttgat 540
caggacagag acttgggggc catcctgccc ctccaaccg acatgtgtac ttcagctttt 600
tccctcactt gcatcaataa agcttctgtg ttggaacag ctaagctggt agtcatttgt 660
tcattcattc attctgaagc ctgcccctgag ttcaaccctg gctgggcaact ggataagagc 720
ctgtccctgg tgggaagagc cccacacgtg ccaggatgga ggatgacagg caccgagtgc 780
tgtgggagcc cagcactgtg ggaagacatt ttcttctgca gggcgaaaat ctgggaaggc 840
ttcctagaag gggcatctga accaatctag aaagatgagg agcataatac aacctgctgt 900
tctgtgtaca atgtgtgatg atcaaatcgg ggtcatttcc accacctcaa acatttacca 960
tttctttgtg ttgggaacat tatcacacag ttccccataa atatgtacat tatgtgtcca 1020
tttaaaacaa tacaagcaaa caaatgtgaa ttctaaact gagactactg taaggccttt 1080
ctcctaatag caagattgta ataaaactc aactagtcat aaaaaaaaa ccttttctgt 1140
ttctgctgca agcaaggaaa tgagaggatg cttttgagta accttgaggc agaagaccct 1200
aaacaggagt ggaggtgtga tattgtttgg ctctgtgtcc ccaccgaaat tccacatgga 1260
atcttcagcc ccatgtgttg gaggcgggct ctgctgggag gtgattgtat cacggggctg 1320
gtttctaata gtttagcccc acctcccagg tgctgtttct cacaaggctc cattgtttaa 1380

```

-continued

aagtgtgtag cacctccccc ttcgctctgt ctctcctgct gtcttgtgaa gaagggactt	1440
gcttcctctt tgccttccac catgattgta agtttctga ggcctcccag ccatgcttcc	1500
tctgaagcct gcggaactgt gagtcaatta aacctctttt tttcataaac taccagttct	1560
caggtagttc tttatagcaa tgtgagaaca gacaaataca aggtggatgg ggtggggaga	1620
gttcagcatt gtgggtcccta tcacaaggac cttattccag tattaccatg gccatcgggg	1680
gaggcacggt gtggggacca ggctggaggc aggaaggcct gagtgaaggc tgctgtgggc	1740
aactaggcaa gaggcgggtg cctgggcccag gacagggcag ggcagggaaag ggctggggtg	1800
atgcagaggg cttggggccc aggccagagg ggctgggtgt ctccatggac gtgtcctcag	1860
gggtctcacc tgggagacc cagaagaagga gacaacacgt aagggtgagg ctggctccctg	1920
gaaatcagtt gacatcttgc aaatgtggag ccctgaggtg gagccaagcc cgaggcacag	1980
cccactcagt ggcacaagcc acgaagggcc ggcctcacc agacctgggg gctcatgatt	2040
gacagcctct tgttacgta ggaagcagc actcctcccg tggcagaaaa gttgggcctg	2100
agttgttoca tctattccgt atgtgcagct gcgtacacca tacaggtgta taagaccatg	2160
agcctgaggc tctgacaccc cagggctcag tcctgtgtcc tcaggcacca ctccccttg	2220
tacctcctgt cccaacagcc aagctccaca gggccaagac tacagcatcc tttgctcagc	2280
acccaaacca ctgctgtgcc cacagcacgg aggctgggga caccgagctg atgtcctcag	2340
aaggcaagt gacaggacag cctctggggc cccacagtgg cctggaactg aaggctcaga	2400
acctaaaaga agtttcccca tttcaatggt actttcagta agacaccatt ttcaaccata	2460
ctgaggagga tccgctgaaa aatgaaacag aaatgagtcg tgatgggagc ggcagggaga	2520
agcaagggag acgagaagt gggaaacatg aaggaaaagc cacgtgagga agaaaccaga	2580
ggtcaagaga aaaagaatca tggaggtaga ggaagcaaaa aacacacata acaagaatg	2640
tggactttg agtcaaaact atgtgagtc aaaccaggc tctctccca accagtttg	2700
gcagatggcc cagtggaacc tcaactcct catcagtaaa aagggggcag agtgagggtc	2760
ctgagagcta gtacagggac tgtgtgaagt agacaatgcc cagtgttag ctgaagaatc	2820
agggtccagc tgggtctccc taaacagcag ctgctgttca ctgttgaag gcgctctgga	2880
aggccaggcg cgggtgctca tgcttgaat cccagcactg tgggaggccg aggtgggctg	2940
atcacctgag gtagggagtt cgagaccagc ctgaccaacg tggagaaacc ccatctctcc	3000
taaaaataca aaattagcca ggcgtggtag cacatacctg taatcccagc gactcgggag	3060
gctgaggcaa gagaattgct tgaaccagc aggggaggtt gtggtgagcc aagatcgagc	3120
cattgcactc cagcctgggc aacaagagca aaatggcgaa actccatctc cgagaaaaaa	3180
aaaaaaaaaga atactttctg aaagtattta ttcatacaaa taaagacttg acccataagg	3240
taggaacgca aatgggccac ggaatcactc attccacagt atacaccgag tgccctttaa	3300
gtgctgggca ctgctccagg attggggca tattggtgaa agagaagcaa gctgcctgct	3360
cagatggcag ggaatgggga aaaacagga gacagtttcc tgtttgagat gttgggagat	3420
gcttcagata gtatatttac tggaaataga cattcaactt ggatgtccct ttttgaaaat	3480
gtgctcgt ccaggctgg gttggggccc cattgaactt tggtctgac acagctggtg	3540
ccacactcag tggaaactgaa tccatgtttg ccttcaccgg goatccttca ccccaactct	3600
ccccgccaca acatacatcc catgccagcc tggggaccct caaaggtgct tcatcattag	3660

-continued

---

gtttgtggct gggteccact gaagtaagtc ttggcactca gagggatagg aattgaaatga	3720
agacatgaga ttctctctcg ggagggcctct ctaggaaatc tgtggactca cacgtttact	3780
aatgttgctg cagccccgca cccaccttgg ccttgggcag ccatactcta gggcctttgt	3840
aacctctcca tgtgaggaac tcaaattaga cctgggtttg gaggcgggtgc tccgagctgg	3900
cctttggggg aggttttctg cgaggcattt cccaagtgtc ggcaggattg tgtcacagac	3960
acagagtaaa cttttgctgg gctccaagtg accgcccata gtttattata aagggtgactg	4020
caccctgcag ccaccagcac tgcctggctc cacgtgcctc ctggtctcag tatggcgctg	4080
tcctgggttc ttacagtctc gagcctccta cctctgctgg aagcccagat cccattgtgt	4140
gccaacctag taccggtgcc catcaccaac gccaccctgg accgggtgag tgctgggct	4200
agccctgtcc tgagcacatg ggcagctgcc tccctctctc gggcttcctt ttacctgctg	4260
gctgtggtcg caccctcact cccagctctg cctttttctc ttctgggtcc ccagggtgaa	4320
attctacca gccaccggga ctctggaggc accccctgcc tccaaacaca gaagcctcac	4380
tgacaggtcc ttcacggagg acggttctgt gctgggcctg gaggggctgc ctggggggca	4440
atgactgatc ctacaggtga gctcctgcac gcgactgcc caccaggggc ctcatctccc	4500
catctgcaaa atcagggaga gatctgcctg agtctctctc cagctgacag tcaaagattc	4560
agcatcaagc ccccatcacc agctcccccc ttctccccag atcactggca agtggtttta	4620
tatcgcatcg gctttctgaa acgaggagta caataagtcg gttcaggaga tccaagcaac	4680
cttcttttac tttaccccc acaagacaga ggacacgatc tttctcagag agtaccagac	4740
ccggtgagag cccccattcc aatgcacccc catctcagct tctggccaga agacctgagc	4800
aagtccctcc ttcttctctg ccttggcctt cccatgggtg gaaccgggag ggttgcttt	4860
aatctccacc agactcttgc cccgggactg tgatgggcga ttggccactt ctctctgata	4920
acattactgt tttctctccg ccttctggtt gactttagcc agaaccagtg cttctataac	4980
tccagttacc tgaatgtcca gcgggagaat gggaccgtct ccagatacgg tgagggccag	5040
ccctcaggca ggagggttca ccgtgggaac agggcaggcc agcataaggt gggggctgga	5100
tgtagagccc tggaggcttt gggcacagag aaataaccac taacattttt gagctcttac	5160
cacgtgtca gaaaaaatcc ctaagaagac actgagagaa ttagatgagg aaacataaga	5220
acagagacct caaatagttt cccaaggctc acacagctta taattagaac tagaattgga	5280
actccaggct ggcttcagat ctgcctctct ctacacccct ctttaagatc ctttgcaaac	5340
caatggtaga agcctgtatg ttggagaggt ggtacctca actatgtccc ccatcaccgc	5400
agagggtgca catggcaggg agctgatgga gctgaactga catcatttag catcccagac	5460
ctcctctctg ggctcattt tctcctctg taaaacgggg agaaaggccc tgacagccac	5520
agtctgtgtg aggtcctga gatctcatgt acagaaagt cttggcgtgg agctgggcac	5580
gcagcagggg ctgggcacac ggtggcccaa aggagaccgc ggccttact gatgggcttt	5640
gtggccccgg acacacctag gactctctc ctgtaagaca ggcaccattg tgccatccca	5700
tgttctcacc cagaggctct tttctcttc cagagggagg ccgagaacat gttgctcacc	5760
tgctgttct tagggacacc aagacctga tgtttggtc ctacctggac gatgagaaga	5820
actgggggct gctttctat ggtaggcatg cttagcagcc ccaaactcat gccctctca	5880
ggcctcacc cccattcacc caccctggg ctggccccta gaacccagc cctccctggc	5940

-continued

---

ctcccgcg	gccccaccat	gtcccacg	agtctccttg	ctcccctgc	agctgacaag	6000
ccagagacga	ccaaggagca	actgggagag	ttctacgaag	ctctogactg	cttgtgcatt	6060
cccaggtcag	atgtcatgta	caccgactgg	aaaaaggtaa	acgcaaggga	ttggacagtg	6120
cccaccttgt	ccatggccca	acttgggcag	ccccagaggg	ccagagcagg	aaagctgcca	6180
ggcaaggctg	cacagctagg	cagatcttct	gcttttaggc	acctgcctca	ctgtagggac	6240
agctgagctc	tacagaggcc	caggggtggg	ggatgagagc	ccaggaggga	gaagtccctg	6300
tgaaccagg	gaggacctga	aagctaacag	gagggaacag	cgtgagccac	ggggttgggg	6360
gattggcaat	tggaggggac	gtaatgcggg	gagttaccac	ctacagacgc	gtcccaaacc	6420
ccaggctttc	acccccacct	cactcccgc	tcatttttaa	taccogtgc	gtggggaatt	6480
gataactgtg	ttttcaatgt	caccacact	gcagcacggc	cacagtcacc	atccccgattt	6540
ttgctacaaa	tgaaaattac	tgtataatga	gctccttaac	acttttcttt	aaacctgtgt	6600
ttggaagact	tgtgttgggtg	tggccctgtg	ccctaatacc	tgtgaaatca	cagcaccgat	6660
gagctgggtc	caatttttaa	aatatatata	tgcagtactt	ccatgactat	tcaaagaaaa	6720
acaattcctt	ccatttgcca	cctgagatga	ccaccaggga	tgtgaactac	ctcctgcccc	6780
atccccagcc	ccaggatcct	gggacagggc	ttatgaacgc	aacctgtga	gtcagctcac	6840
ttgatccaca	gcctggcacc	tccactgtct	ggctagggag	cctcgaatgg	gtcccaaggc	6900
cacctgctc	ctcagttaca	tcactctgat	agtagtgggtg	gttgtgagga	atccaggagc	6960
tgcagcataa	gggcccctgca	ggtactatgt	gctcagtaaa	tgccagtggg	tcttaagggt	7020
ctgagctccc	atggtagagc	caagtaagct	gaggttcaga	gacagaaaat	gacttgccca	7080
agatcaccca	gctgggaaat	gacagtgcca	gggttggagc	cctggttgag	ctggttccac	7140
aggccagagc	tcattctgcc	ctctcccgg	aagacctccc	accctgtccc	catgcctctg	7200
cttctccctc	acccaatcc	cccgtgcct	tctaggataa	gtgtgagcca	ctggagaagc	7260
agcacgagaa	ggagagga	caggaggagg	gggaatccta	gcaggacaca	gccttggatc	7320
aggacagaga	cttgggggccc	atcctgcccc	tccaaccgca	catgtgtacc	tcagcttttt	7380
ccctcacttg	catcaataaa	gcttctgtgt	ttggaacagc	taagctgtta	gtcatttgtt	7440
cattcattca	ttctgaagc	tgcctgagt	tcaaccctgg	ctgggcaactg	gataagacgc	7500
tgtccctggg	gggaagagcc	ccacacgtgc	caggatggag	gatgacaggc	accgagtgtc	7560
gtgggagccc	agcactgtgg	gaagacatct	tcttctgcag	ggcgaaaatc	tgggaaggct	7620
tcctagaagg	ggcatctgaa	ccaatctaga	aagatgagga	gcataatata	acctgtgttt	7680
ctgtgtacaa	tgtgtgatga	tcaaatcggg	gtcatttcca	ccacctcaaa	catttaccat	7740
ttctttgtgt	tgggaacatt	atcacacagt	tcccataaa	tatgtacatt	atgtgtccat	7800
ttaaaacaat	acaagcaaac	aatgtgaat	ttctaaactg	agactactgt	aaggcctttc	7860
tcctaatagc	aagattgtaa	taaaacttca	actagtcata	aaaaaaaaatc	cttttctgtt	7920
tctgctgcaa	gcaaggaaat	gagaggatgc	atgtgagtaa	ccttgaggca	gaagacccta	7980
aacaggagtg	gaggtgtgat	attgtttggc	tctgtcccga	cccgaatctc	acatggaatt	8040
ttcagcccca	tgtgttggag	gcgggtcctg	ctgggagggtg	attgtatcac	ggggctgggt	8100
totaatgggt	tagccccacc	tcccagggtc	tgtttctcac	aaggtctcat	tgtttaaaag	8160
tgtgtagcac	ctcccccttc	gctctgtctc	tcctgtctgc	ttgtgaagaa	gggacttgtc	8220

-continued

tctcttttgc cttccaccat gattgtaagt ttcttgaggc ctcccagcca tgettectct	8280
gaagcctgcg gaactgtgag tcaattaaac ctcttttttt cataaactac ccagtctcag	8340
gtagtctttt atagcaatgt gagaacagac aaatacaagg tggatggggg ggggagagtt	8400
caggcatgtg gtgcctatca caaggacctt attccagtat taccatggcc atcgggggag	8460
gcacgttggt gggaccaggc tggaggcagg aaggcctgag tgaaggctgc tgtgggcaac	8520
taggcaagag gcggttgctt gggccaggac agggcagggc agggaagggc tgggtgatg	8580
cagagggctt gggcccagg ccagaggggc tgggtgtctc catggactg tcttagggt	8640
gctcacctgg gagaccaga agaaggagat aacacgtaag ggtgaggctg gtcctggaa	8700
atcagctgac atcctgcaaa tgtggagccc tgaggtgag ccaagcccga ggcacagccc	8760
actcagtgcc acaagccaca aaggccggc ctcaccccaa cctgggtgct cgtgattgac	8820
agcctcttgt tacgtgaggg aagcagcacc ccttccgtgg cagaaaagt ggcctgaat	8880
tgtttgagca attccactg tgcagctgcg tataccatcc agggtgatag gaccatgagc	8940
ccgagccctt gatgtcctgg ggtctagtcc tgtgtccaca ggcaccactc ccctttgtac	9000
ctcctgtccc aacagccaag ctccacaggg ccaagactac agcatccttt gctcagcacc	9060
caaacacag ctgtgcccac agcaccgagg ctggggacac cgagctgatg tcctcagaag	9120
gcaagtggac aggacagcct ctggggctcc acaatgtcct ggaactgaag gctcagaacc	9180
tacaggaagt ttgtccattt cagtgttact ttcaagtgaga caccatttcc aaccatactc	9240
aggaggatcc gctgaaaaat gaaacagaaa tgagtcatga tgggcagagc agggagaagc	9300
aagggagacg agaagtgggg aacatggaag aaaaagccac gtgaggaaga aaccagaggt	9360
caagagaaaa agaataatgg aggtagagga agcaaaaaac acacataaca aagaatgtgg	9420
actttgggtc agaccaatgt gagtttaaat ccaggttctc tcccacaacca gtttgggag	9480
atggcccatt ggaaccccac tctcctcatc agtaaaaggg gggcagagtg agggctctga	9540
gggttagcac atggactgtg tgaagcagac aatgccagct gottagcacc ggaatcaggg	9600
tctagcaggt gctcccataa tggcaggtgc tgtttgctat tgaagtact ggaaggctgg	9660
gogccatggc tcacgcctgt aatcccagca atttgggagg ctgaggtggg tggatcacct	9720
gaggctggga gtgcaagatc agcccagcca acatggagaa accctgtctc tactaaaaat	9780
acaaaattag ccagacgtgg tgggtcatgc ctgtaatccc agctactcgg gaggtgagg	9840
caagagaatc acttgaaccc gggaggcggg ggttgcatg agccgagatt ccaccactgt	9900
acccagcctt gggcaaccag agcaaaactc agtctcaaaa aaaaaaaaaa agaaaaagaa	9960
aaaagaaaaa aaaaagaaag aaaggagcac tggagagtcc	10000

&lt;210&gt; SEQ ID NO 20

&lt;211&gt; LENGTH: 201

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Homo sapien

&lt;400&gt; SEQUENCE: 20

ctgggcagcc tcgcctcttg gcctctgcag gtgcctctgt atgggaggcc agagtttctg	60
tcactaactt tttctaagct cacaatgtct agagtgggg rcgcttttcc acgcagtgga	120
acatgacttt tctttgaaatc totggcaggt ctgtgtgttc tgtcactgac tcgcccagagc	180
agtggtcttg ctggctcatcc c	201

-continued

---

```

<210> SEQ ID NO 21
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

<400> SEQUENCE: 21
gccctctgca tcctcctcct cctggctcac cggtttggtg gcccacatcg tgggcctctc   60
gttccgaggc agctccttgc tactcctgct cttggtggac rccgactgcc ttctcgggcc   120
tggggcgctg gcctctgcct cgcacagatc cacaccactg tcctcogtta gagtctgccc   180
gctgtccgag agctgggaga g                                     201

<210> SEQ ID NO 22
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

<400> SEQUENCE: 22
ctgtggtcct aagaaaagga cgtcagtgga aagagagagt gaccocagaat ctctagaagg   60
ttaaggcaag cagagccctc aggataagc tggaaaccag ygctcctggg agggaatgca   120
ggcaccgag aacttggctg ggggaactgt tacatcttca tttccattaa cctctaactg   180
aaaggccacg tttcctttca t                                     201

<210> SEQ ID NO 23
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

<400> SEQUENCE: 23
ccattcccag tctctctgga aaccctatc ctagcgcctc tggcagagtc tgcaggacgg   60
tgggagctga tgaatcccc ttgcaacaaa cagcccacat sacatgtttt aaatgcaaac   120
cggctggact ttaggaggca aaggcccac gtggacatcg cctccctcct gcagaagaac   180
tgtgagacga ctgccccagg a                                     201

<210> SEQ ID NO 24
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

<400> SEQUENCE: 24
aacctagggt ttcaactca ggggcccgcc tgttatacac atagtgaagc agcgtggggt   60
gactggtggg ctggagggca tgtgccacc actcagctca sgcaggggccc atgcagaggt   120
tcagaccagg tgtggtgaa tttctctttt tctcttgta agctagaaat tactactttc   180
tctctctctc tctctttttt t                                     201

<210> SEQ ID NO 25
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

<400> SEQUENCE: 25
gaaatcagtc tgccaagata aataattata actgataata ggcttatcag tgtgaccaca   60
gtaaataatt ccattattat atacatcata ccacagcgta rttgctcact tacctgtctc   120

```

## -continued

---

```
ttaaggaac ctgcatctaa ttatctttta acttttaggg ctgacatata cttgacctat 180
aattgttgaa ttagccaact c 201
```

```
<210> SEQ ID NO 26
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
```

```
<400> SEQUENCE: 26
```

```
tcttttgcc tttgtgatca gcccaaggcca ccttttaaga gttctttggt cactggggag 60
gattctgtca tgccaagggc tgggaataaa aaagtcacaa rtgaaaagac ttgctcagca 120
ttggcaactc ccacccctca gtggttctct tgagttttgg taaatctaga ctgatgaata 180
tagcagctaa tgatcatagt a 201
```

```
<210> SEQ ID NO 27
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
```

```
<400> SEQUENCE: 27
```

```
aaaggcaagc tgagcagatg ttttctctga gtgagtgtgt tacttcatta tctgtgctaa 60
taattataag aataaaaaag aaagcgcagt ggactcctat rtttgctctt atgttttgta 120
ctgtcatcta aatcatgtct cttcaggcgc gtgtttgctt ctggggatag ttttatcttg 180
tcaagtgtgt ttgttcttgg a 201
```

```
<210> SEQ ID NO 28
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
```

```
<400> SEQUENCE: 28
```

```
agggatgagg cgcgagaagc taccacagc ccatagaaca tttccatttc ttttacctca 60
tttgctggtg ctctgatatt tttcgggtgc agcctctctc ygattttaaa tgtaactcat 120
gctcatgtag aaagtgtgaa aacaacatag aaaactttga aagagaaaat aggactctct 180
agtaatccca tccccgaaa g 201
```

```
<210> SEQ ID NO 29
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
```

```
<400> SEQUENCE: 29
```

```
gccctctgca tctctctcct cctgggtcac cggtttggtg gccccatctg tgggcctctc 60
gttccgaggc agctccttgc tactcctgct cttgggtggac rccgactgcc ttctctggcc 120
tggggcgctg gcctctgcct cgccagcatc cacaccactg tctctgctta gactctgccc 180
gctgtccgag agctgggaga g 201
```

```
<210> SEQ ID NO 30
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien
```

```
<400> SEQUENCE: 30
```

-continued

---

```

ccattcccag tctctctgga aaccctatc ctagegcctc tggcagagtc tgcaggacgg      60
tgggagctga tgaatcccc ttgcaacaaa cagccccaat sacatgtttt aaatgcaaac      120
cggttgact ttaggaggca aaggccccac gtggacatcg cctccctcct gcagaagaac      180
tgtgagacga ctgccccagg a                                                201

```

```

<210> SEQ ID NO 31
<211> LENGTH: 201
<212> TYPE: DNA
<213> ORGANISM: Homo sapien

```

```

<400> SEQUENCE: 31

```

```

aacctagggt ttcaaaacta ggggccggc tgttatacac atagtgaagc agcgtgggtt      60
gactggtggg ctggagggca tgtgccacc actcagctca sgcaggggcc atgcagaggt      120
tcagaccagg tgtggtgaa tttctctttt tctcttgta agctagaat tactactttc      180
tctctctctc tctctttttt t                                                201

```

---

What is claimed is:

1. A method of identifying an individual having an altered risk for developing Alzheimer's Disease, comprising detecting the presence or absence of one or more alleles of a SNP in said individual's nucleic acids, wherein the SNP is selected from any one of the nucleotide sequences of SEQ ID NOS:11-15, and 20-31, and the presence or absence of the alleles are correlated with an individual's risk for developing Alzheimer's Disease.

2. The method of claim 1, wherein the SNP is selected from the group of nucleotide sequences consisting of SEQ ID NOS of 11, 13, 20, 21, 25, 26, 27, and 28.

3. The method of claim 1, comprising contacting the nucleic acids of the individual with a detection reagent, and determining which nucleotide is present or absent at a specified SNP position.

4. The method of claim 3, in which the nucleic acids are prepared from blood cells of the individual.

5. The method of claim 3, in which the detection reagent is a polynucleotide probe.

6. The method of claim 5, in which the 3' end of the probe hybridizes to the SNP in the nucleic acid.

7. The method of claim 5, in which the probe is labeled with a reporter dye.

8. A method of detecting a single nucleotide polymorphism (SNP) in a nucleic acid molecule, comprising contacting a test sample with a reagent which specifically hybridizes to a SNP in any one of the nucleotide sequences of SEQ ID NOS:1-5 and 11-31 under stringent hybridization conditions, and detecting the formation of a hybridized duplex.

9. The method of claim 8 in which detection is carried out by a process selected from the group consisting of: allele-specific probe hybridization, allele-specific primer extension, allele-specific amplification, sequencing, 5' nuclease digestion, molecular beacon assay, oligonucleotide ligation assay, size analysis, and single-stranded conformation polymorphism.

10. A method of identifying an individual having an altered risk for developing Alzheimer's Disease, comprising detecting the presence or absence of a polymorphism in said

individual's nucleic acids, wherein said polymorphism is in linkage disequilibrium with a SNP in any one of the nucleotide sequences of SEQ ID NOS:11-15 and 20-31, and the presence or absence of the polymorphism is correlated with an individual's risk for developing Alzheimer's Disease.

11. The method of claim 10, wherein the SNP is selected from the group of nucleotide sequences consisting of SEQ ID NOS of 11, 13, 20, 21, 25, 26, 27, and 28.

12. A reagent for detecting a variant protein encoded by a SNP-containing nucleic acid molecule, said reagent selectively binds to the variant protein as compared to a protein encoded by another nucleic acid molecule, wherein the SNP-containing nucleic acid molecule contains a SNP in any one of the nucleotide sequences of SEQ ID NOS:11-15 and 20-31.

13. The reagent of claim 12, wherein said reagent is an antibody, an antibody fragment, an aptamer, a peptide, a ligand, or a small molecule compound.

14. The reagent of claim 13, wherein said reagent is labeled with a reporter dye or an imaging agent.

15. A kit comprising the reagent of claim 12 and a buffer.

16. An amplified polynucleotide containing a single nucleotide polymorphism (SNP) selected from any one of the nucleotide sequences of SEQ ID NOS:1-5 and 11-31, or a complement thereof, wherein the amplified polynucleotide is between about 16 and about 1,000 nucleotides in length.

17. The amplified polynucleotide of claim 16 in which the nucleotide sequence comprises any one of the nucleotide sequences of SEQ ID NOS:1-5 and 11-31.

18. An isolated polynucleotide which specifically hybridizes to a nucleic acid molecule containing a single nucleotide polymorphism (SNP) in any one of the nucleotide sequences in SEQ ID NOS:1-5 and 11-31.

19. The polynucleotide of claim 18 which is 8-70 nucleotides in length.

20. The polynucleotide of claim 18 which is an allele-specific probe.

**21.** The polynucleotide of claim 18 which is an allele-specific primer.

**22.** A kit for detecting a single nucleotide polymorphism (SNP) in a nucleic acid, comprising the polynucleotide of claim 18, a buffer, and an enzyme.

**23.** An isolated polypeptide comprising of an amino acid sequence selected from the group of amino acid sequences consisting of SEQ ID NOS: 6-10.

**24.** The isolated polypeptide of claim 23 wherein the polypeptide has the amino acid sequence selected from the group of amino acid sequences consisting of SEQ ID NOS: 6-10.

**25.** An antibody that selectively binds to a polypeptide of claim 23.

**26.** A method for identifying an agent useful in therapeutically or prophylactically treating Alzheimer's Disease, comprising contacting the polypeptide of claim 24 with a candidate agent under conditions suitable to allow formation of a binding complex between the polypeptide and the candidate agent, and detecting the formation of the binding complex, wherein the presence of the complex identifies said agent.

\* \* \* \* \*

专利名称(译)	与阿尔茨海默氏病相关的遗传多态性，检测方法及其用途		
公开(公告)号	<a href="#">US20070254289A1</a>	公开(公告)日	2007-11-01
申请号	US11/586427	申请日	2006-10-24
[标]申请(专利权)人(译)	阿普里拉股份有限公司		
申请(专利权)人(译)	Applera公司		
当前申请(专利权)人(译)	Applera公司		
[标]发明人	LI YONGHONG GRUPE ANDREW		
发明人	LI, YONGHONG GRUPE, ANDREW		
IPC分类号	G01N33/53 C07H21/04 C07K16/18 C12Q1/68		
CPC分类号	C12Q1/6883 C12Q2600/172 C12Q2600/158 C12Q2600/156 A61P25/28		
优先权	60/730918 2005-10-26 US		
其他公开文献	US7695911		
外部链接	<a href="#">Espacenet</a>	<a href="#">USPTO</a>	

摘要(译)

本发明基于与阿尔茨海默氏病相关的遗传多态性的发现。特别地，本发明涉及含有多态性的核酸分子，由这种核酸分子编码的变体蛋白，用于检测多态性核酸分子和蛋白质的试剂，以及使用该核酸和蛋白质的方法以及使用方法。用于检测的试剂。

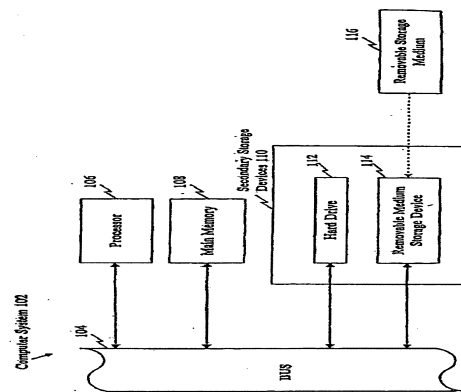


FIGURE 1