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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 March 2006 (02.03.2006)

PCT

(10) International Publication Number
WO 2006/022639 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number:
PCT/US2004/024001

(22) International Filing Date: 21 July 2004 (21.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **APPLERA CORPORATION** [US/US]; 1401 Harbor Bay Parkway, Alameda, CA 94502-7070 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CARGILL, Michele** [US/US]; c/o Celera Diagnostics, LLC, 1401 Harbor Bay Parkway, Alameda, CA 94502-7070 (US). **GRUPE, Andrew** [DE/US]; c/o Celera Diagnostics, LLC, 1401 Harbor Bay Parkway, Alameda, CA 94502-7070 (US). **LI, Yonghong** [US/US]; c/o Celera Diagnostics, LLC, 1401 Harbor Bay Parkway, Alameda, CA 94502-7070 (US).

(74) Agents: **SEPERACK, Peter** et al.; Celera Diagnostics, LLC, 1401 Harbor Bay Parkway, Alameda, CA 94502-7070 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2006/022639 A1

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

(57) Abstract: 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.

CL001496PCT

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

FIELD OF THE INVENTION

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

Neurodegenerative Diseases

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.

Specific diseases exemplifying this class of disorders include age-related dementia, such as Alzheimer's disease, leukodystrophies, such as adrenoleukodystrophy, metachromatic, leukodystrophy, Krabbe Disease (globoid cell leukodystrophy), Canavan Disease, Alexander Disease, Pelizaeus-Merzbacher Disease, and the like, neuronal ceroid lipofuscinoses, stroke, and the like.

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.

Alzheimer's disease (described in greater detail in the following section) alone affects 20 million people worldwide. 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.

Alzheimer's Disease

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 Aβ amyloid, silver-staining neurofibrillary tangles in neuronal cytoplasm, and accumulation of Aβ 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.

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.

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.

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 (Sherrington 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 (Aß42), 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.

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, ε2, ε3, ε4. The ε4 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-ε4 allele, which has been replicated in many studies, most investigators consider the ApoE-ε4 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).

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 Aβ42 and a susceptibility locus for LOAD on chromosome 10, suggests that the locus influences LOAD risk through increased levels of the Aβ42 peptide (Ertekin-Taner *Science* 2000, 290:2303-2304).

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.

Thus, 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 also 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.

GAPDH and Treatment of Neurodegenerative Diseases

Available treatments for neurodegenerative diseases do not provide an effective and long-term treatment. Various treatments that are used with little or no success include monaamine oxidase inhibitors, anti-apoptotics, anti-inflammatory drugs, anti-oxidants, anti-amyloid and neurotropic factors either alone or in combination. The best present therapy is to provide comfort and emotional support for the victim and the victim's closest relatives.

While the molecular basis for some neurodegenerative disease are known, the underlying mechanisms of action for most have not been made clear. However, neuronal death underlies the symptoms of many, if not all, human neurological disorders and there is evidence that a common component of the neuronal death is apoptosis.

The monoamine oxidase inhibitor *R*-(-)-deprenyl or Selegiline was developed for use in treating Parkinson's disease. Selegiline, it is believed, acts to protect neurons or glias from programmed cell death by inhibition of apoptosis. Paradoxically, CGP 3466, a structural analog of *R*-(-)-deprenyl, exhibits little monoamine oxidase inhibiting activity, but is a potent neuroprotective agent. The putative molecular target responsible for mediating the antiapoptotic, neuroprotective effects of Deprenyl and CGP 3466 now is identified as glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Carlile, *et al., Mol. Pharmacol.* 57:2-12 (2000) have reported that deprenyl-like compounds specifically bind to GAPDH and prevent both the apoptotic increases in GAPDH levels and nuclear accumulation of GAPDH.

The anti-dementia drug tetrahydroaminoacridine (THA), whose generic name is tacrine, has been shown to effectively attenuate the "age-induced" apoptosis of cerebellar and

cerebrocortical neurons in culture by suppressing the overexpression of a particulate-bound 38-kDa protein as well as GAPDH mRNA levels during the neuronal death pathway (Sunaga, *et al. Neuroscience Let.* 200:133-136 (1995), Ishitani, *et al., J. Pharmacol. Exp. Ther.* 278:447-454 (1996)). THA has been approved for clinical use in the treatment of Alzheimer's Disease (AD). THA is believed to reverse cholinergic deficits in the CNS of patients via an inhibitory activity on cholinesterase. However, other cholinesterase inhibitors have failed to elicit any significant improvement in AD symptoms (Bartus *et al., Science,* 217:408-417 (1982), and, therefore, the therapeutic effects of THA may involve pharmacological actions other than cholinesterase inhibition.

A second anti-dementia drug, ONO-1603 [(S)-1-[N-(4-chloro-benzyl)succinamoyl]pyrrolidine-2-carbaldehyde], which inhibits prolyl endopeptidase (EC 3.4.21.26), robustly suppresses overexpression of GAPDH mRNA and accumulation of GAPDH protein in particulate fraction (*i.e.* the nucleus) of cultured neurons being subjected to "age-induced" apoptosis (Katsube *et al., J. Pharmacol. Exp. Ther.* 288:6-13 (1999), Katsube and Ishitani, *CNS Drug Review* 6:21-34 (2000).

Previously, the sole function of GAPDH was thought to be as a housekeeping enzyme in the glycolytic pathway. However, in keeping with its proposed role in neuronal apoptosis, GAPDH mRNA and GAPDH protein is found to be up-regulated in the particulate fraction of cell extracts during age-induced apoptosis of mature cerebellar and cerebrocortical neurons and ara-C-induced apoptosis of cultured cerebellar neurons. GAPDH mRNA overexpression is reversed by antisense GAPDH oligonucleotides in each of these cellular assays and apoptosis is delayed concomitantly. The up-regulation of GAPDH mRNA and the increase in GAPDH protein content in the apoptotic cell appears to be a general phenomenon in neuronal cells undergoing apoptosis.

Two CAG-associated neurodegenerative diseases, HD and DRPLA, are known to involve GADPH binding to the polyglutamine domains in the huntingtin protein and DRPLA protein, respectively. GAPDH specifically binds to the carboxy terminal of the β-amyloid precursor protein, which itself, as well as the carboxy terminal fragments thereof, are involved in neuronal loss in Alzheimer's disease. In this regard, a monoclonal antibody raised against amyloid plaques from an Alzheimer's patient's brain was found to cross react with GAPDH. Given that GAPDH has various functions including roles in glycolysis and apoptosis, it is an excellent candidate protein for involvement in the neurodegenerative process. Accordingly, inhibition of GAPDH using therapeutics that antagonize or promote

GAPDH and its related protein's activities is an attractive means for treating the effects of neurodegenerative disease. In addition, determining which alleles of GAPDH and its related proteins GAPDS and p-GAPD are present in an individual will aid in determining how that individual will respond to treatment with specific therapeutic agents.

SNPs

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.

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, or SNP locus) 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.

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* 2002 Oct;71(4):854-62).

A synonymous codon change, or silent mutation/SNP (the terms "SNP" and "mutation" 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 substitution 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 the SNPs are bi-allelic, and are thus often referred to as "bi-allelic markers", or "di-allelic markers".

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 July 2001).

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.

Causative SNPs do not necessarily have to 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.

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.

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 SNP(s) and a pathological condition (or other phenotype) of interest, then the region around the SNP can optionally be thoroughly screened to identify the causative genetic locus/sequence(s) (e.g., 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).

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 treatment due to the patient's increased likelihood of developing toxic side effects or 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. (1997), *Clinical Chemistry*, 43, 254; Marshall (1997), *Nature Biotechnology*, 15, 1249; International Patent Application WO 97/40462, Spectra Biomedical; and Schafer et al. (1998), *Nature Biotechnology*, 16, 3).

SUMMARY OF THE INVENTION

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 pathologies. The polymorphisms disclosed herein are directly useful as targets for the design of diagnostic reagents and the development of therapeutic agents for use in the diagnosis and treatment of Alzheimer's disease and related pathologies.

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 novel SNPs in genetic sequences involved in Alzheimer's disease, 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 of a SNP disclosed herein or its encoded product, methods of identifying individuals who are more or less likely to respond to a treatment, 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, and methods of using the novel SNPs of the present invention for human identification.

In Tables 1-2, the present invention provides gene information, transcript sequences (SEQ ID NOS:1-433), encoded amino acid sequences (SEQ ID NOS:434-866), genomic sequences (SEQ ID NOS:6752-7071), transcript-based context sequences (SEQ ID NOS:867-6751) and genomic-based context sequences (SEQ ID NOS:7072-54,769) 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:1-433), amino acid sequences (SEQ ID NOS:434-866), genomic sequences (SEQ ID NOS:6752-7071), transcript-based SNP context sequences (SEQ ID NOS: 867-6751), and genomic-based SNP context sequences (SEQ ID NOS:7072-54,769) are also provided in the Sequence Listing.

In a specific embodiment of the present invention, naturally-occurring SNPs in the human genome are provided. 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. 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 Tables 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 that is encoded by a nucleic acid molecule containing a SNP disclosed herein.

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.

Also provided in the invention are 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 a SNP allele disclosed herein. In another embodiment, a method for diagnosis of Alzheimer's disease by detecting the presence or absence of a SNP allele disclosed herein is provided.

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. 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.

An aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH, GAPDS, and/or p-GAPD) gene (hCG2005673, hCG23577, and hCG40445 for example, identifies a GAPDH, a GAPDS or a p-GAPD gene respectively that is disclosed in Tables 1-2 along with associated transcript, protein, and genomic sequences and SNP information), which method comprises administering to said human subject a

therapeutically or prophylactically effective amount of one or more agents counteracting the neurodegenerative effects of the disease.

Another aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH, GAPDS, and/or p-GAPD) gene, which method comprises administering to said patient a therapeutically or prophylactically effective amount of one or more neuroprotective agents.

Another aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH, GAPDS, and/or p-GAPD) gene, which method comprises administering to said patient a therapeutically or prophylactically effective amount of one or more anti-apoptotic agents.

Another aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH, GAPDS, and/or p-GAPD) gene, which method comprises administering to said patient a therapeutically or prophylactically effective amount of one or more agents which inhibit the activity of GAPDH.

Another aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH, GAPDS, and/or p-GAPD) gene, which method comprises administering to said patient a therapeutically or prophylactically effective amount of one or more agents which bind to and thereby inhibit the activity of GAPDH, GAPDS, and/or p-GAPD in particular wherein the disease is selected from adrenoleukodystrophy, Alexander Disease, Alzheimer's disease, amyotrophic lateral sclerosis, Canavan Disease, cerebellar degeneration, cerebral ischemias, glaucoma, Krabbe Disease, metachromatic leukodystrophy, multiple sclerosis, neuronal ceroid lipofuscinoses, Parkinson's disease, Pelizaeus-Merzbacher Disease, retinitis pigmentosa, stroke, neurodegenerative disease caused by traumatic injury.

Another aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant GAPDH gene comprising a polynucleotide sequence selected from the group consisting of the genomic sequence of SEQ ID NO:6795, the transcript sequences of SEQ ID NOS:125-127, and nucleic acid sequences that encode a polypeptide comprising an amino acid sequence of SEQ ID NOS:558-560. A further aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant GAPDS gene comprising a

polynucleotide sequence selected from the group consisting of the genomic sequence of SEQ ID NO:6828, the transcript sequence of SEQ ID NO:212, and nucleic acid sequences that encode a polypeptide comprising an amino acid sequence of SEQ ID NO:645. A further aspect of this invention is a method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant p-GAPD gene comprising a polynucleotide sequence selected from the group consisting of the genomic sequence of SEQ ID NO:6889, the transcript sequence of SEQ ID NO:413, and nucleic acid sequences that encode a polypeptide comprising an amino acid sequence of SEQ ID NO:846.

Another aspect of this invention is a method for identifying an agent useful in therapeutically or prophylactically treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH GAPDS, and/or p-GAPD) gene, which method comprises contacting GAPDH GAPDS, and/or p-GAPD with a candidate agent under conditions suitable to allow formation of a binding complex between the GAPDH GAPDS, and/or p-GAPD and the candidate agent and detecting the formation of the binding complex, wherein the presence of the complex identifies said agent.

Another aspect of this invention is a method for treating neurodegenerative disease in a human subject, which method comprises:

(i) determining that said human subject harbors a mutant glyceraldehyde-3-phospate dehydrogenase (GAPDH, GAPDS, and/or p-GAPD) gene, and

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

Another aspect of this invention is a method for determining if an individual will respond to a certain course of treatment for a neurodegenerative disease, which method comprises:

(i) determining the genotype of said individual for specific alleles of GAPDH, GAPDS, and/or p-GAPD;

(ii) correlating the genotype of said individual with predetermined responses of each genotype to a therapeutic agent.

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.

CL001496PCT

DESCRIPTION OF THE FILES CONTAINED ON THE CD-R NAMED CL001496CDR

The CD-Rs named CRF and COPY1 - SEQUENCE LISTING PART, COPY 2 – SEQUENCE LISTING PART and COPY 3 – SEQUENCE LISTING PART contain the following text (ASCII) files:

File SEQLIST_1496.txt provides the Sequence Listing. The Sequence Listing provides the transcript sequences (SEQ ID NOS:1-433) and protein sequences (SEQ ID NOS:434-866) as shown in Table 1, and genomic sequences (SEQ ID NOS:6752-7071) 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:867-6751) and genomic-based context sequences as shown in Table 2 (SEQ ID NOS:7072-54,769). The context sequences generally provide 100bp upstream (5') and 100bp downstream (3') of each SNP, with the SNP in the middle of the context sequence, for a total of 200bp of context sequence surrounding each SNP. File SEQLIST_1496.txt is 54,809 KB in size.

The CD-Rs named COPY 1 – TABLES PART, COPY 2 – TABLES PART and COPY 3 – TABLES PART contain the following text (ASCII) files:

1) File TABLE1_1496.txt provides Table 1. File TABLE1_1496.txt is 5,513 KB in size.

2) File TABLE2_1496.txt provides Table 2. File TABLE2_1496.txt is 48,391 KB in size.

3) File TABLE3_1496.txt provides Table 3. File TABLE3_1496.txt is 57 KB in size.

4) File TABLE4_1496.txt provides Table 4. File TABLE4_1496.txt is 106 KB in size.

The material contained on the CD-Rs labeled CRF, COPY1 - SEQUENCE LISTING PART, COPY 2 – SEQUENCE LISTING PART, COPY 3 – SEQUENCE LISTING PART, COPY 1 – TABLES PART, COPY 2 – TABLES PART and COPY 3 – TABLES PART are hereby incorporated by reference pursuant to 37 CFR 1.77(b)(4).

DESCRIPTION OF TABLE 1 AND TABLE 2

Table 1 and Table 2 (both provided on the CD-R TABLES PART) disclose the SNP and associated gene/transcript/protein information of the present invention. For each gene,

CL001496PCT

Table 1 and Table 2 each provide 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.

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.

The gene/transcript/protein information includes:

- a gene number (1 through n, where n = the total number of genes in the Table)
- a Celera hCG and UID internal identification numbers for the gene
- a Celera hCT and UID internal identification numbers for the transcript (Table 1 only)
- a public Genbank accession number (e.g., RefSeq NM number) for the transcript (Table 1 only)
- a Celera hCP and UID internal identification numbers for the protein encoded by the hCT transcript (Table 1 only)
- a public Genbank accession number (e.g., RefSeq NP number) for the protein (Table 1 only)
- an art-known gene symbol
- an art-known gene/protein name
- Celera genomic axis position (indicating start nucleotide position-stop nucleotide position)
- the chromosome number of the chromosome on which the gene is located
- 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
- alternative gene/protein name(s) and/or symbol(s) in the OMIM entry

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.

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:

- transcript sequence (Table 1 only) (corresponding to SEQ ID NOS:1-433 of the Sequence Listing), with SNPs identified by their IUB codes (transcript sequences can include 5' UTR, protein coding, and 3' UTR regions). (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.)

- the encoded protein sequence (Table 1 only) (corresponding to SEQ ID NOS:434-866 of the Sequence Listing)

- the genomic sequence of the gene (Table 2 only), including 6kb on each side of the gene boundaries (i.e., 6kb on the 5' side of the gene plus 6kb on the 3' side of the gene) (corresponding to SEQ ID NOS:6752-7071 of the Sequence Listing).

After the last gene sequence, Table 2 may include additional genomic sequences of intergenic regions (in such instances, these sequences are identified as "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).

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:1-433), protein sequence (SEQ ID NOS:434-866), and transcript-based SNP context sequence (SEQ ID NOS:867-6751), and SEQ ID NOS are indicated in Table 2 for each genomic sequence (SEQ ID NOS:6752-7071), and genomic-based SNP context sequence (SEQ ID NOS:7072-54,769).

The SNP information includes:

- 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:867-6751; the genomic-based SNP context sequences in Table 2 are provided in the Sequence Listing as SEQ ID NOS:7072-54,769).

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

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

- 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, "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) (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)

- Population/allele/allele count information in the format of [population1(allele1,count|allele2,count) population2(allele1,count|allele2,count) total (allele1,total count|allele2,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].

CL001496PCT

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.

(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)

- 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]

- 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.

DESCRIPTION OF TABLE 3 AND TABLE 4

Tables 3 and 4 (both provided on the CD-RTABLES PART) provide a list of a subset of SNPs from Table 1 (in the case of Table 3) or Table 2 (in the case of Table 4) for which the SNP source falls into one of the following three categories: 1) SNPs for which the SNP source is only "Applera" and none other, 2) SNPs for which the SNP source is only "Celera Diagnostics" and none other, and 3) SNPs for which the SNP source is both "Applera" and "Celera Diagnostics" but none other.

These SNPs have not been observed in any of the public databases (dbSNP, HGBASE, and HGMD), and were also not observed during shotgun sequencing and assembly of the Celera human genome sequence (i.e., "Celera" SNP source). Tables 3 and 4 provide the hCV identification number (or hDV identification number for SNPs having "Celera Diagnostics" SNP source) and the SEQ ID NO of the context sequence for each of these SNPs.

DESCRIPTION OF TABLE 5

Table 5 provides sequences (SEQ ID NOS:54,770-55,342) of primers that have been synthesized and used in the laboratory to carry out allele-specific PCR reactions in order to assay the SNPs disclosed in Tables 6-7 during the course of Alzheimer's disease association studies.

Table 5 provides the following:

- the column labeled "hCV" provides an hCV identification number for each SNP position

- the column labeled "Allele 1" designates which allele at the SNP position identified by the hCV identification number is targeted by the allele-specific primers (the allele-specific primers are shown as "Sequence A" and "Sequence B" in each row)

- the column labeled "Sequence A" provides an allele-specific primer that is specific for an allele at the SNP position identified by the hCV identification number

- the column labeled "Sequence B" provides an allele-specific primer that is specific for the other allele at the SNP position identified by the hCV identification number

- the column labeled "Sequence C" provides a common primer that is used in conjunction with each of the allele-specific primers ("Sequence A" and "Sequence B") and which hybridizes at a site away from the SNP position.

All primer sequences are given in the 5' to 3' direction.

The allele given in the "Allele1" column matches the 3' nucleotide of either "Sequence A" or "Sequence B". Whichever of these two sequences matches the allele is the allele-specific primer that is specific for the allele given in the "Allele1" column, and the other sequence (the sequence which does not match the allele given in the "Allele1" column) will be specific for the other allele at the SNP position identified by the hCV identification number (this other allele is not given in Table 5 but is indicated in Tables 1 and/or 2 for the SNP position identified by the hCV identification number).

DESCRIPTION OF TABLE 6 AND TABLE 7

Tables 6-7 provide results of statistical analyses for SNPs disclosed in Tables 1-5 (SNPs can be cross-referenced between Tables based on their hCV or hDV identification numbers). The statistical results shown in Tables 6-7 provide support for the association of these SNPs with Alzheimer's disease. For example, the statistical results provided in Tables 6-7 show that the association of these SNPs with Alzheimer's disease is supported by p-values < 0.05 in at least one of three genotypic association tests and/or an allelic association test. Moreover, Table 6 provides results of markers that are significant in at least two independently collected sample sets, which further verifies the association of these SNPs with Alzheimer's disease. Table 7 lists an additional set of markers that have shown significant association in one sample set ($p<0.05$) and remain significant ($p<0.01$) after all genotyped sample sets, including the initial set, are analyzed together.

CL001496PCT

Description of column headings for Tables 6 & 7

Table 6 & 7 column heading	Definition
Marker	Identification number for the SNP that is tested
Sample Set	Sample Set used in the analysis (1, 2, or 3)
Strata	Indicates if the analysis of the dataset was based on a substratum such as ApoE4 genotype, gender, or age of disease onset (strata are described below)
Adjust	describes the parameters that were used to adjust the p-values derived by Cochran Mantel Haenszel test [no adjustments (none), presence or absence of ApoE4 allele (apoe4), gender (male), age of disease onset in cases and age at mental exam in controls (age_ge75), sample set (source)]
Allelic p-value	result of the asymptotic chi square test for allelic association or the allelic p-value of the stratified analysis with Cochran Mantel Haenszel test (Cochran Mantel Haenszel test was used when 'Adjust' is different from 'none')
Additive p-value	result of the Armitage trendtest for additive genotypic association or the additive p-value of the stratified analysis with Cochran Mantel Haenszel test with ordered scores (Cochran Mantel Haenszel test was used when 'Adjust' is different from 'none')
Dominant p-value	result of the asymptotic chi square test for dominant genotypic association or the dominant p-value of the stratified analysis with Cochran Mantel Haenszel test (Cochran Mantel Haenszel test was used when 'Adjust' is different from 'none')
Recessive p-value	result of the asymptotic chi square test for recessive genotypic association or the recessive p-value of the stratified analysis with Cochran Mantel Haenszel test (Cochran Mantel Haenszel test was used when 'Adjust' is different from 'none')
OR-allelic	allelic odds ratio
OR-allelic 95%CI	95% confidence interval of the allelic odds ratio
OR-dominant	dominant odds ratio
OR-dominant 95%CI	95% confidence interval of the dominant odds ratio
OR-recessive	recessive odds ratio
OR-recessive 95%CI	95% confidence interval of the recessive odds ratio
Allele 1	Polymorphic nucleotide of the tested SNP for which allele frequencies are being reported
Case Allele 1 Freq	Allele frequency of allele 1 in cases
Control Allele 1 Freq	Allele frequency of allele 1 in controls
Case Samples	Count of case individuals that were analyzed
Control Samples	Count of control individuals that were analyzed

Definition of entries in the "Strata" column (Tables 6 & 7) for stratification-based analyses:

Strata	Definition
apoe4=0	no Apo E4 allele present
apoe4=1	at least one Apo E4 allele present

CL001496PCT

age_ge75=0	age at disease onset is less than 75 years of age (controls are <75 years old at mental exam)
age_ge75=1	age at disease onset is 75 years of age or older (controls are >=75 years old at mental exam)
male=0	only female
male=1	only male
ALL	all individuals

NOTE: SNPs can be cross-referenced between Tables 1-7 based on the hCV (or hDV) identification number of each SNP. However, five of the SNPs that are included in Tables 1-7 possess two different identification numbers, as follows:

-- hCV12126867 is equivalent to hCV27398082

-- hCV2981216 is equivalent to hCV26956511

-- hCV8227677 is equivalent to hCV26838632

-- hCV8856240 is equivalent to hCV26740731

-- hDV68530963 is equivalent to hCV27939864

DESCRIPTION OF THE FIGURE

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

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides SNPs associated with Alzheimer's disease, nucleic acid molecules containing 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 Alzheimer's disease-associated SNPs disclosed herein are useful for diagnosing, screening for, and evaluating predisposition to Alzheimer's disease and related pathologies in humans. Furthermore, such SNPs and their encoded products are useful targets for the development of therapeutic agents.

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.

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:1-433) and genomic sequences (SEQ ID NOS:6752-7071) containing SNPs, encoded amino acid sequences (SEQ ID NOS: 434-866), and both transcript-based SNP context sequences (SEQ ID NOS: 867-6751) and genomic-based SNP context sequences (SEQ ID NOS:7072-54,769) (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 disorders associated with a variant gene/protein such as Alzheimer's disease, 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.

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, etc.). 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). Novel SNPs for which the SNP source is only "Applera" and none other, i.e., those that have not been observed in any public databases and which were also not observed during shotgun sequencing and assembly of the Celera human genome sequence (i.e., "Celera" SNP source), are indicated in Tables 3-4.

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 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).

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.

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

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:1-433), genomic sequences (SEQ ID NOS:6752-7071), and protein sequences (SEQ ID NOS:434-866) 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:867-6751 correspond to transcript-based SNP context sequences disclosed in Table 1, and SEQ ID NOS:7072-54,769 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.

Isolated Nucleic Acid Molecules

The present invention provides isolated nucleic acid molecules that contain one or more SNPs disclosed Table 1 and/or Table 2. Preferred isolated nucleic acid molecules contain one or more SNPs identified in Table 3 and/or Table 4. 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.

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.

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.

For full-length genes and entire protein-coding sequences, a SNP flanking sequence can be, for example, up to about 5KB, 4KB, 3KB, 2KB, 1KB 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.

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 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: 1-433), and polymorphic genomic sequences are provided in Table 2 and in the Sequence Listing (SEQ ID NOS:6752-7071). 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.

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:1-433, genomic sequences are provided in Table 2 as SEQ ID NOS:6752-7071, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:867-6751, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:7072-54,769) and their complements. A fragment typically comprises a contiguous nucleotide sequence at least about 8 or more nucleotides, more preferably at least about 12 or more nucleotides, and even more preferably at least about 16 or more nucleotides. Further, a fragment could comprise at least about 18, 20, 22, 25, 30, 40, 50, 60, 100, 250 or 500 (or any other number in-between) nucleotides in length. 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.

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. Patent Nos. 4,683,195; and 4,683,202; *PCR Technology: Principles and Applications for DNA Amplification*, ed. H.A. Erlich, Freeman Press, NY, NY, 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. Patent Nos. 5,270,184; and 5,422,252), transcription-mediated amplification (TMA) (U.S. Patent No. 5,399,491), linked linear amplification (LLA) (U.S. Patent 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.

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.

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, or 300 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 no greater than 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

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.

In a specific embodiment of the invention, the amplified product is at least about 201 nucleotides in length, 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).

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.

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:1-433, genomic sequences are provided in Table 2 as SEQ ID NOS:6752-7071, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:867-6751, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:7072-54,769), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:434-866). A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

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:1-433, genomic sequences are provided in Table 2 as SEQ ID NOS:6752-7071, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:867-6751, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:7072-54,769), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:434-866). 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.

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:1-433, genomic sequences are provided in Table 2 as SEQ ID NOS:6752-7071, transcript-based SNP context sequences are provided in Table 1 as SEQ ID NO:867-6751, and genomic-based SNP context sequences are provided in Table 2 as SEQ ID NO:7072-54,769), or any nucleic acid molecule that encodes any of the variant proteins provided in Table 1 (SEQ ID NOS:434-866). 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 (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY).

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.

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.

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 (Sambrook and Russell, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY). 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. Patent 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.* 1997 Jun;15(6):224-9, and Hyrup et al., "Peptide nucleic acids (PNA): synthesis, properties and potential applications", *Bioorg Med Chem.* 1996 Jan;4(1):5-23). 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, CA) 3900 High-Throughput DNA Synthesizer or Expedite 8909 Nucleic Acid Synthesis System, and the sequence information provided herein.

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.

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. Patent No. 4,835,263) and the minor groove binders (U.S. Patent 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, N.Y. (2002).

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.

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, the present invention specifically contemplates isolated nucleic acid molecule 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.

To determine the percent identity of two amino acid sequences or two nucleotide sequences of two molecules 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.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M 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 Blossom 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.

In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *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.

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* 2002 Dec;20(12):1269-71).

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.

SNP Detection Reagents

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:1-433), genomic sequences (provided in Table 2 as SEQ ID NOS:6752-7071), and context sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:867-6751; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:7072-54,769), 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:867-6751; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:7072-54,769). 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 provided herein is also useful for designing primers, e.g. allele-specific primers, to amplify (e.g., using PCR) any SNP of the present invention. Preferred sets of primers for allele-specific amplification reactions, which have been synthesized and used in the laboratory to assay SNPs, are provided in Table 5 as SEQ ID NOS:54,770-55,342.

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.

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, 60, 100 (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.

Other preferred primer and probe sequences can readily be determined using the transcript sequences (SEQ ID NOS:1-433), genomic sequences (SEQ ID NOS:6752-7071), and SNP context sequences (transcript-based context sequences are provided in Table 1 as SEQ ID NOS:867-6751; genomic-based context sequences are provided in Table 2 as SEQ ID NOS:7072-54,769) 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.

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.

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").

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.

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 5X standard saline phosphate EDTA (SSPE), 0.5% $NaDodSO_4$ (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 2X SSPE, and 0.1%SDS at 55°C or room temperature.

Moderate stringency hybridization conditions may be used for allele-specific primer extension reactions with a solution containing, e.g., about 50mM 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 100mM KCl at a temperature of 46°C.

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.

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. Patent No. 4,988,617), the 3'most nucleotide of the probe aligns with the SNP position in the target sequence.

Oligonucleotide probes and primers may be prepared by methods well known in the art. Chemical synthetic methods include, but are limited to, the phosphotriester method described by Narang *et al.*, 1979, Methods in Enzymology 68:90; the phosphodiester method described by Brown *et al.*, 1979, Methods in Enzymology 68:109, the diethylphosphoamidate method described by Beaucage *et al.*, 1981, *Tetrahedron Letters* 22:1859; and the solid support method described in U.S. Patent No. 4,458,066.

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.

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, 1989, *Nucleic Acid Res.* 17 2427-2448). Preferred sets of primers for allele-specific amplification reactions, which have been synthesized and used in the laboratory to assay SNPs, are provided in Table 5 as SEQ

ID NOS:54,770-55,342. 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.

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.

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.

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. Patent Nos. 5,210,015 and 5,538,848) or Molecular Beacon probe (U.S. Patent Nos. 5,118,801 and 5,312,728), or other stemless or linear beacon probe (Livak *et al.*, 1995, PCR Method Appl. 4:357-362; Tyagi *et al.*, 1996, Nature Biotechnology 14: 303-308;

Nazarenko *et al.*, 1997, Nucl. Acids Res. 25:2516-2521; U.S. Patent Nos. 5,866,336 and 6,117,635).

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.

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 readily 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 corresponds to a SNP disclosed herein, is a composition that is encompassed by the present invention). Thus, reagents that bind to a nucleic acid molecule in a region adjacent to a SNP site, even though the bound sequences do not necessarily include the SNP site itself, are also encompassed by the present invention.

SNP Detection Kits and Systems

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.

An exemplary kit/system of the present invention can be based on the primer sequences disclosed in Table 5. Preferred sets of primers for allele-specific amplification reactions, which have been synthesized and used in the laboratory to assay SNPs, are provided in Table 5 as SEQ ID NOS:54,770-55,342. It will be apparent to one of skill in the art that these primers disclosed in Table 5 for detecting 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/system format. For example, for a particular target SNP position identified by an hCV identification number in Table 5, the two corresponding allele-specific primers (identified in Table 5 as "Sequence A" and "Sequence B") and the common primer (identified in Table 5 as "Sequence C"), which have been used and validated in the laboratory, can all three be readily packaged into a kit format along with, optionally, other biochemical reagents, such as reagents for carrying out allele-specific amplification reactions (e.g., enzymes, dNTPs, buffer, etc.).

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.

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.

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. Patent No. 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; *Nat. Biotech.* 14: 1675-1680) and Schena, M. et al. (1996; *Proc. Natl. Acad. Sci.* 93: 10614-10619), 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. Patent No. 5,807,522.

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

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.

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.

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*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

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 applications 10/620332 and 10/620333 describe chemiluminescent approaches for microarray detection; U.S. Patent Nos. 6124478, 6107024, 5994073, 5981768, 5871938, 5843681, 5800999, and 5773628 describe methods and compositions of dioxetane for performing chemiluminescent detection; and U.S. published application US2002/0110828 discloses methods and compositions for microarray controls.

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.

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.

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.

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.

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 fluorescent 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.* 2003 Feb 24;55(3):349-77). In such microfluidic devices, the containers may be referred to as, for example, microfluidic "compartments", "chambers", or "channels".

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. Patent 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. Patent Nos. 6,153,073, Dubrow *et al.*, and 6,156,181, Parce *et al.*

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.

USES OF NUCLEIC ACID MOLECULES

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.

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).

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, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY).

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

SNP Genotyping Methods

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.

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.* 2003;3(2):77-96; Kwok et al., "Detection of single nucleotide polymorphisms", *Curr Issues Mol Biol.* 2003 Apr;5(2):43-60; Shi, "Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes", *Am J Pharmacogenomics.* 2002;2(3):197-205; and Kwok, "Methods for genotyping single nucleotide polymorphisms", *Annu Rev Genomics Hum Genet* 2001;2:235-58. Exemplary techniques for high-throughput SNP genotyping are described in Marnellos, "High-throughput SNP analysis for genetic association studies", *Curr Opin Drug Discov Devel.* 2003 May;6(3):317-21. 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. Patent 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.

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.

In a preferred embodiment, SNP genotyping is performed using the TaqMan assay, which is also known as the 5' nuclease assay (U.S. Patent 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. Alternativey, 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.

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.

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, CA), 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. Patent Nos. 5,118,801 and 5,312,728) and other variant formats (U.S. Patent Nos. 5,866,336 and 6,117,635).

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. Patent 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.

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. Patent Nos. 6027889, 6268148, 5494810, 5830711, and 6054564 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 zip code array; U.S. application US01/17329 (and 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. applications 60/427818, 60/445636, and 60/445494 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.

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.

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.

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.* 2003;17(11):1195-202.

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.* 2003 Jul;19 Suppl 1:I44-I53; Storm et al., "MALDI-TOF mass spectrometry-based SNP genotyping", *Methods Mol Biol.* 2003;212:241-62; Jurinke et al., "The use of MassARRAY technology for high throughput genotyping", *Adv Biochem Eng Biotechnol.* 2002;77:57-74; and Jurinke et al., "Automated genotyping using the DNA MassArray technology", *Methods Mol Biol.* 2002;187:179-92.

SNPs can also be scored by direct DNA sequencing. A variety of automated sequencing procedures can be utilized ((1995) *Biotechniques 19*:448), 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 3730xl DNA Analyzers (Foster City, CA), is commonly used in the art for automated sequencing.

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 (Erlich, ed., PCR Technology, *Principles and Applications for DNA Amplification*, W.H. Freeman and Co, New York, 1992, Chapter 7).

Sequence-specific ribozymes (U.S.Patent 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

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.

SNP genotyping is useful for numerous practical applications, as described below. Examples of such applications 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.

Analysis of Genetic Association Between SNPs and Phenotypic Traits

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.

Different study designs may be used for genetic association studies (*Modern Epidemiology*, Lippincott Williams & Wilkins (1998), 609-622). 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.

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.

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 information 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.

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)).

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 alleles x 2 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 genotypes x 2 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%.

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.

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.

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.

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 (*Multiple comparisons and multiple tests*, Westfall et al, 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.

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*, Lippincott Williams & Wilkins, 1998, 643-673). If available, association results known in the art for the same SNPs can be included in the meta-analyses.

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.

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.* 1999, 65:220-228) 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* 1999, 55:997-1004) 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. Epidem.* 20001, 21:273-284).

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.* 2000,

66:1933-1944) 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.* 1999, 8:237-245).

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)).

Disease Diagnosis and Predisposition Screening

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 a family 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 alleles.

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 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.

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 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.

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 (LD)-based SNP association analysis. 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 occurrence of each allele in a given population. The expected frequency of 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 a different 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.

For diagnostic purposes, if a particular SNP site is found to be useful for diagnosing Alzheimer's disease, 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. 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.

For diagnostic applications, 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). Thus, polymorphic markers that are in LD with causative polymorphisms are useful as diagnostic markers, and are particularly useful when the actual causative polymorphism(s) is/are unknown.

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, 25, 30, 50, 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 family history, diet, environmental factors or lifestyle factors.

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, in certain circumstances, be used to 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. 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.

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 mutation, including 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.

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.

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 (i.e., the concentration of mRNA or protein in a sample, etc.) or pattern (i.e., 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 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.

Pharmacogenomics and Therapeutics/Drug Development

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 Biotechnology* 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.

Pharmacogenomics in general is discussed further in Rose et al., "Pharmacogenetic analysis of clinically relevant genetic polymorphisms", *Methods Mol Med.* 2003;85:225-37. Pharmacogenomics as it relates to Alzheimer's disease and other neurodegenerative disorders is discussed in Cacabelos, "Pharmacogenomics for the treatment of dementia", *Ann Med.* 2002;34(5):357-79, Maimone et al., "Pharmacogenomics of neurodegenerative diseases", *Eur J Pharmacol.* 2001 Feb 9;413(1):11-29, and Poirier, "Apolipoprotein E: a pharmacogenetic target for the treatment of Alzheimer's disease", *Mol Diagn.* 1999 Dec;4(4):335-41. Pharmacogenomics as it relates to cardiovascular disorders is discussed in Siest et al., "Pharmacogenomics of drugs affecting the cardiovascular system", *Clin Chem Lab Med.* 2003 Apr;41(4):590-9, Mukherjee et al., "Pharmacogenomics in cardiovascular diseases", *Prog Cardiovasc Dis.* 2002 May-Jun;44(6):479-98, and Mooser et al., "Cardiovascular pharmacogenetics in the SNP era", *J Thromb Haemost.* 2003 Jul;1(7):1398-402.

Pharmacogenomics as it relates to cancer is discussed in McLeod et al., "Cancer pharmacogenomics: SNPs, chips, and the individual patient", *Cancer Invest.* 2003;21(4):630-40 and Watters et al., "Cancer pharmacogenomics: current and future applications", *Biochim Biophys Acta.* 2003 Mar 17;1603(2):99-111.

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.

As an alternative to genotyping, specific variant peptides 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.

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, identifies 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 (Pfost *et. al., Trends in Biotechnology*, Aug. 2000.). Thus, pharmacogenomics based on the SNPs disclosed herein has the potential to both save lives and reduce healthcare costs substantially.

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 or their products, as well as genes or their products that are directly or indirectly regulated by or interacting with these disease-associated SNP-containing 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.

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*, Crooke (ed.), 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).

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.* 2003

Jul;6(4):561-9; Stephens et al., "Antisense oligonucleotide therapy in cancer", *Curr Opin Mol Ther*. 2003 Apr;5(2):118-22; Kurreck, "Antisense technologies. Improvement through novel chemical modifications", *Eur J Biochem*. 2003 Apr;270(8):1628-44; Dias et al., "Antisense oligonucleotides: basic concepts and mechanisms", *Mol Cancer Ther*. 2002 Mar;1(5):347-55; Chen, "Clinical development of antisense oligonucleotides as anti-cancer therapeutics", *Methods Mol Med*. 2003;75:621-36; Wang et al., "Antisense anticancer oligonucleotide therapeutics", *Curr Cancer Drug Targets*. 2001 Nov;1(3):177-96; and Bennett, "Efficiency of antisense oligonucleotide drug discovery", *Antisense Nucleic Acid Drug Dev*. 2002 Jun;12(3):215-24.

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 do not bind an alternative polymorphic form (e.g., an alternative SNP nucleotide that encodes a protein having normal function).

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 encoding regions and particularly protein encoding regions corresponding to catalytic activities, substrate/ligand binding, or other functional activities of a protein.

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 21-22bp 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)). Thus, 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).

The following references provide a further review of RNAi: Agami, "RNAi and related mechanisms and their potential use for therapy", *Curr Opin Chem Biol.* 2002 Dec;6(6):829-34; Lavery et al., "Antisense and RNAi: powerful tools in drug target discovery and validation", *Curr Opin Drug Discov Devel.* 2003 Jul;6(4):561-9; Shi, "Mammalian RNAi for the masses", *Trends Genet* 2003 Jan;19(1):9-12), Shuey *et al.*, "RNAi: gene-silencing in therapeutic intervention", *Drug Discovery Today* 2002 Oct;7(20):1040-1046; McManus *et al.*, *Nat Rev Genet* 2002 Oct;3(10):737-47; Xia *et al.*, *Nat Biotechnol* 2002 Oct;20(10):1006-10; Plasterk *et al.*, *Curr Opin Genet Dev* 2000 Oct;10(5):562-7; Bosher *et al.*, *Nat Cell Biol* 2000 Feb;2(2):E31-6; and Hunter, *Curr Biol* 1999 Jun 17;9(12):R440-2).

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 wild-type nucleotide at the position of the SNP. This site-specific repair sequence encompasses 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 wild type sequence into the subject's genome. Upon incorporation, the wild type gene product is expressed, and

the replacement is propagated, thereby engendering a permanent repair and therapeutic enhancement of the clinical condition of the subject.

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 cognate of the variant protein. Once administered in an effective dosing regimen, the wild type cognate provides complementation or remediation of the pathological condition.

The invention further provides a method for identifying a compound or agent that can be used to treat Alzheimer's disease. Variant gene expression in a Alzheimer's disease patient can include, 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 wild-type nucleic acid sequence (for instance, a regulatory/control region can contain a SNP that affects the level or pattern of expression of a wild-type transcript). 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 the nucleic acid 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. 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 nucleic acid molecules.

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 the 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.

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 abnormal gene expression 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.

The invention further provides methods of treatment, with the SNP or associated nucleic acid domain (e.g., regulatory/control region) 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.

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 are 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.

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 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.

In another aspect of the present invention, there is provided a pharmaceutical pack comprising a therapeutic agent (e.g., a small molecule drug, an antibody, a peptide, 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.

The SNPs/haplotypes of the present invention are also useful for improving many different aspects of the drug development process. For example, individuals can be selected for clinical trials based on their SNP genotype. Individuals with SNP genotypes that indicate that they are most likely to respond to the drug can be included in the trials and those individuals whose SNP genotypes indicate that they are less likely to or would not respond to

the drug, or suffer adverse reactions, can be eliminated from the clinical trials. This not only improves the safety of clinical trials, but also will 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.

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*, 2002 Mar; S30-S36).

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* 2001 Mar;19(3):209-11). 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.

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*, 2002 Mar; S30-S36).

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.

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.

Administration and Pharmaceutical Compositions

In this section, GAPDH inhibitors are described as an exemplary class of therapeutic compounds that target an Alzheimer's disease-associated protein, or encoding nucleic acid molecule, disclosed herein. However, one of skill in the art will recognize that any of the Alzheimer's disease-associated proteins, and encoding nucleic acid molecules, disclosed herein are useful as therapeutic targets for treating Alzheimer's disease, and that the present disclosure enables therapeutic compounds to be developed that target any of these other therapeutic targets.

In general, GAPDH inhibitors (or other therapeutic compounds that target an Alzheimer's disease-associated protein, or encoding nucleic acid molecule, disclosed herein) 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 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.

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

In general, the GAPDH inhibitors 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.

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.

The compositions are comprised of in general, a GAPDH inhibitor 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 GAPDH inhibitors. Such excipient 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.

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.

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.

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

The amount of the 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 GAPDH inhibitor 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 %.

The GAPDH inhibitors can be administered alone or in combination with other inhibitors or in combination with one or more other active ingredient(s). For example, GAPDH inhibitor can be administered in combination with another neuroprotective agent.

Human Identification Applications

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.* 2001;114(4-5):204-10). 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.

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 forensics/human identification applications, such as predicting phenotypic characteristics (e.g., inferring ancestry) from a DNA sample, it may be desirable to utilize SNPs that affect the encoded protein.

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.

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* 2003; 48(4):771-782.

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 mutationaly 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.

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.

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.

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.

Furthermore, the SNPs provided by the present invention can be typed for inclusion in a database of DNA genotypes, for example, a criminal DNA databank. 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.

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.

The use of the SNPs of the present invention for human identification further extends to various authentication systems, commonly referred to as biometric systems. Biometric systems convert physical characteristics of humans (or other organisms) into digital data for precise quantification. 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 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. Patent No. 6,119,096).

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*, John Wiley & Sons, N.Y. (2002), 14.1-14.7.

VARIANT PROTEINS, ANTIBODIES, VECTORS & HOST CELLS, & USES THEREOF

Variant Proteins Encoded by SNP-Containing Nucleic Acid Molecules

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:434-866 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.

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.

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.

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.

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.

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, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

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.

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.

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.

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.

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 re-amplified 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.

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.

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., alleleic 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 prior to the present invention.

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, the present invention specifically contemplates 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.

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.

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.

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).

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.

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)).

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.

Known protein modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, 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 phosphotidylinositol, 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, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

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); Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.*, *Meth. Enzymol. 182*: 626-646 (1990); and Rattan *et al.*, *Ann. N.Y. Acad. Sci. 663*:48-62 (1992).

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 prior to the present invention.

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, N.Y. (2002)).

Uses of Variant Proteins

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, Cold Spring Harbor Laboratory Press, Sambrook and Russell, 2000, and Methods in Enzymology: Guide to Molecular Cloning Techniques, Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987).

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: 434-866. 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.

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.

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.

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).

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.

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.

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.

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.

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).

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.

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.

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.

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, $F(ab')_2$, 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).

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.

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.

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 end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the variant protein is derived.

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, *Bio/technology*, 1992, Sept 10(9), 973-80).

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/^{125}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., ^{35}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.

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.

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.

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.

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. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; 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.

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.

Antibodies Directed to Variant Proteins

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).

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, $F(ab)'_2$, 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. Patent 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. Patent Nos. 5,693,762; 5,585,089; and 5,565,332), a single-chain Fv (U.S. Patent 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.

Many methods are known in the art for generating and/or identifying antibodies to a given target antigen (Harlow, Antibodies, Cold Spring Harbor Press, (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.

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. Patent No. 5,637,677).

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.

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, β-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}I, ^{131}I, ^{35}S or ^{3}H.

Antibodies, particularly the use of antibodies as therapeutic agents, are reviewed in: Morgan, "Antibody therapy for Alzheimer's disease", *Expert Rev Vaccines*. 2003 Feb;2(1):53-9; Ross et al., "Anticancer antibodies", *Am J Clin Pathol*. 2003 Apr;119(4):472-85; Goldenberg, "Advancing role of radiolabeled antibodies in the therapy of cancer", *Cancer Immunol Immunother*. 2003 May;52(5):281-96. Epub 2003 Mar 11; Ross et al., "Antibody-based therapeutics in oncology", *Expert Rev Anticancer Ther*. 2003 Feb;3(1):107-21; Cao et al., "Bispecific antibody conjugates in therapeutics", *Adv Drug Deliv Rev*. 2003 Feb 10;55(2):171-97; von Mehren et al., "Monoclonal antibody therapy for cancer", *Annu Rev Med*. 2003;54:343-69. Epub 2001 Dec 03; Hudson et al., "Engineered antibodies", *Nat Med*. 2003 Jan;9(1):129-34; Brekke et al., "Therapeutic antibodies for human diseases at the dawn of the twenty-first century", *Nat Rev Drug Discov*. 2003 Jan;2(1):52-62 (Erratum in: *Nat Rev Drug Discov*. 2003 Mar;2(3):240); Houdebine, "Antibody manufacture in transgenic animals and comparisons with other systems", *Curr Opin Biotechnol*. 2002 Dec;13(6):625-9; Andreakos et al., "Monoclonal antibodies in immune and inflammatory diseases", *Curr Opin Biotechnol*. 2002 Dec;13(6):615-20; Kellermann et al., "Antibody discovery: the use of transgenic mice to generate human monoclonal antibodies for therapeutics", *Curr Opin Biotechnol*. 2002 Dec;13(6):593-7; Pini et al., "Phage display and colony filter screening for high-throughput selection of antibody libraries", *Comb Chem High Throughput Screen*. 2002 Nov;5(7):503-10; Batra et al., "Pharmacokinetics and biodistribution of genetically engineered antibodies", *Curr Opin Biotechnol*. 2002 Dec;13(6):603-8; and Tangri et al., "Rationally engineered proteins or antibodies with absent or reduced immunogenicity", *Curr Med Chem*. 2002 Dec;9(24):2191-9.

Uses of Antibodies

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.

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.

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.

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.

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.

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.

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).

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.

Vectors and Host Cells

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.

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.

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).

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.

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 λ, 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.

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.

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, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

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, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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.

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.

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.

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* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) 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)).

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 (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). 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)).

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, CA).

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., Sf 9 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)).

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 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman et al., *EMBO J.* 6:187-195 (1987)).

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).

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.

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, 2000, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Host cells can contain more than one vector. Thus, different SNP-containing nucleotide sequences can be introduced 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.

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.

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.

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.

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.

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.

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.

Uses of Vectors and Host Cells, and Transgenic Animals

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.

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.

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.

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.

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.

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. Patent Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Patent No. 4,873,191 by Wagner et al., and in Hogan, B., *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 animal 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.

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.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in, for example, Wilmut, I. 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_o 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.

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).

COMPUTER-RELATED EMBODIMENTS

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:1-433, SEQ ID NOS:434-866, SEQ ID NOS:6752-7071, SEQ ID NOS:867-6751, and SEQ ID NOS:7072-54,769; 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.

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:1-433, protein sequences are provided as SEQ ID NOS:434-866, genomic sequences are provided as SEQ ID NOS:6752-7071, transcript-based context sequences are provided as SEQ ID NOS:867-6751, and genomic-based context sequences are provided as SEQ ID NOS:7072-54,769).

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.

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.

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 at, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag et at, *Comp. Chem.* 17:203-207 (1993)) search algorithms.

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.

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.

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 necessary 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.

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.

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).

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.

One exemplary embodiment of a computer-based system comprising SNP information of the present invention is provided in Figure 1. Figure 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.

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.

EXAMPLES:

The following examples are offered to illustrate but not to limit the claimed invention.

EXAMPLE 1 STATISTICAL ANALYSIS OF SNP ASSOCIATION WITH ALZHEIMER'S DISEASE

A case-control genetic study to determine the association of SNPs in the human genome with late onset Alzheimer's Disease (LOAD) was carried out using genomic DNA extracted from 3 independently collected case-control sample sets, totaling 2285 samples

(1089 cases and 1196 controls). The majority of SNPs analyzed in these samples were located on chromosome 9, chromosome 10, and chromosome 12. All patients (cases) were diagnosed with Alzheimer's disease according to NINCDS-ADRDA or related criteria (McKhann et al. 1984, *Neurology* 34:939-44). Controls underwent MMSE testing. All individuals who were included into the study had signed a written informed consent form. The study protocol was IRB approved.

DNA was extracted from blood samples, using conventional DNA extraction methods like the QIA-amp kit from Qiagen. Genotypes were obtained on a PRISM 7900HT sequence detection PCR system (Applied Biosystems) by allele-specific PCR, similar to the method described by Germer et al (Germer S., Holland M.J., Higuchi R. 2000, *Genome Res.* 10: 258-266). Primers for the allele-specific PCR reactions are described in Table 5.

Summary statistics for demographic and environmental traits, APOE allelic and genotypic frequencies, and allele frequencies for the tested SNPs were obtained, and compared between cases and controls. No multiple testing corrections were made.

Significant association was observed between APOE and AD status in all 3 sample sets. For example in sample set 1, the odds ratio for the APOE ε34 genotype vs. ε33 genotype was 3.1 (95% CI 2.22-4.34) while the odds ratio for ε44 vs. ε33 genotype was 9.74 (95% CI 3.74 – 35.33). P-value for genotypic association test was < 0.0001. Allele frequency for APOE ε4 was 31.2% in cases and 13.4% in controls. The odds ratio for APOE ε4 genotypes also compared closely with what has been reported in the literature.

Several tests of association were calculated for both non-stratified and stratified settings: 1) asymptotic chi-square test of allelic association, 2) asymptotic chi-square test of genotypic association, taking three different modes of inheritance into account (dominant, recessive and additive), 3) Cochran-Mantel-Haenszel test (Categorical Data Analysis by Alan Agresti published by Wiley InterScience, 1990) for stratified analyses. Allelic and genotypic p-values were calculated for the combined samples after stratification for sample-set, gender, age of disease onset, and ApoE4 genotype. 4) exact test of Hardy-Weinberg equilibrium (HWE) for cases and controls.

As LOAD is a complex disease that is influenced by APOE genotype, age, and gender, the above analyses were adjusted for the effects of these factors, as indicated in Tables 6-7. P-values, adjusted for these covariates, were considered significant at the level of <0.05.

Effect sizes were estimated through allelic odds ratios and odds ratios for dominant and recessive models, including 95% confidence intervals. Homogeneity of Cochran-Mantel-Haenszel odds ratios was tested across different strata using the Breslow-Day test. The reported allele1 may be under-represented in cases (with a lower allele frequency in cases than in controls, indicating that the minor allele is associated with decreased risk and the major allele is a risk factor for disease) or over-represented in cases (indicating that the minor allele is a risk factor in the development of disease).

A SNP was considered to be a significant genetic marker if it exhibited a p-value < 0.05 in the allelic association test or in any of the 3 genotypic tests (dominant, recessive, additive). SNPs with significant HWE violations in both cases and controls ($p<1x10^{-4}$ in both tests) were not considered for further analysis, since significant deviation from HWE in both cases and controls for individual markers can be indicative of genotyping errors. The association of a marker with Alzheimer's disease was considered replicated if the marker exhibits an allelic or genotypic association test p-value <0.05 in one of the sample sets and the same test and strata are significant ($p<0.05$) in either one or two other independent sample sets, or the Cochran-Mantel-Haenszel p-value was significant ($p<0.05$) for the combined data of the other two sample sets for the same test and strata. SNPs that fall in this category are listed in Table 6.

SNP-strata combinations that are not listed in Table 6, but were significant in one sample set ($p<0.05$) with the allelic or genotypic association test and the Cochran-Mantel-Haenzsel test was significant ($p<0.01$) with the same test and same strata when all available sample set data were analyzed together, are listed in Table 7.

An example of a replicated marker, where the minor allele is associated with increased risk for Alzheimer's disease is hCV8227677 (Table 6). hCV8227677 shows significant association with all individuals (strata = "ALL") of sample set 1 and the "ALL" strata of the jointly analyzed sample sets 2 & 3. In addition, the female ("male=0"), the APOE4 present ("apoe4=1"), and both age of onset substrata ("age_ge75=0", "age_ge75=1") are significantly associated for this marker in 2 independent, non overlapping sample sets. The "ALL" strata for sample set 1 shows significant Cochran-Mantel-Haenszel p-values (corrected for APOE4 status, gender, and age of disease onset) in the allelic ($p=0.00601$), the additive genotypic ($p=0.00587$), and the recessive genotypic tests ($p=0.00001$). The dominant genotypic test is not significant ($p=0.9775$). The "ALL" strata of the combined sample sets 2 and 3 confirm the sample set 1 results with significant Cochran-Mantel-Haenszel test p-values (corrected for sample set, APOE4 status, gender, and age of disease onset) in the

allelic (p=0.00016), additive genotypic (p=0.00017), and the recessive genotypic test (p=0.00324). The allelic and recessive odds ratios for hCV8227677 show similar effects in these sample sets, indicating the C-allele as risk factor, and thereby further strengthening the association of this marker with Alzheimer's disease (sample set 1: OR allelic=1.37 (95%CI=1.09-1.71), OR recessive=2.35 (95%CI=1.61-3.43); sample sets 2 and 3 combined: OR allelic=1.41 (95%CI=1.18-1.69), OR recessive=1.59 (95%CI=1.17-2.16)). The odds ratios are based on the minor allele as observed in the control samples (i.e. C-allele for hCV8227677).

The association of reduced risk with Alzheimer's disease for the G-allele of marker hCV286937 has been replicated in the "ALL" strata and the "male=1" substratum analysis (Table 6). In the male substratum analysis, the significant Cochran-Mantel-Haenszel test p-values ($p<0.05$) for sample set 1 are replicated in the combined sample sets 2 and 3 for the allelic, additive genotypic, and dominant genotypic tests ($p<0.05$). The G-allele confers reduced disease risk, based on similar allelic and dominant odds ratios (OR allelic =0.4/0.46; OR dominant=0.34/0.44; sample set 1/sample sets 2 and 3) in the male substratum of sample set 1 and the combined sample sets 2 and 3.

EXAMPLE 2 STATISTICAL ANALYSIS OF GAPD GENE SNPS ASSOCIATED WITH ALZHEIMER'S DISEASE

The case-control study described in Example 1 was the source of data determining that specific SNPs in GAPD genes are associated with Alzheimer's Disease (LOAD). The results are presented in Table 8. Allele frequencies for the tested SNPs are indicated. P-values and odds ratios (OR) were calculated as described in Example 1. The SNP in the GAPDH gene is located in the 5'untranslated region (UTR5) –8 relative to the first AUG near a Kozak consensus sequence. The SNP in the GAPDS gene is located in a predicted transcription factor binding site (TFBS). The SNP in the p-GAPD gene creates a missense mutation. This polymorphism also maps to an intron of a protein phosphatase 2C homolog (hCG 40446) which is transcribed in the opposite direction to p-GAPD. The data were stratified for analysis. The data for the GAPDH gene were stratified for all samples versus ApoE4- samples. The data for the GAPDS gene were stratified for all samples versus patients whose age of onset of disease was less than 75 years of age (LT75 samples). The data for the p-GAPD gene was not stratified for analysis but the significant association occurs in individuals with an age of onset of disease greater than or equal to 75 years of age (GE75 samplse). The sample sets were obtained from three separate locations and studies.

As the three GAPD genes may function in similar pathways, tests were performed to determine whether there was an association of LOAD with a genotype formed by these genes. A model-free, Williams-corrected G-test was performed separately for each pair-wise combination of genes, showing significant values for GAPDH-GAPDS ($p<0.01$) and GAPDS-p-GAPD ($p<0.005$) but not GAPDH-p-GAPD. Additionally, two-locus genotypic disequillibrium was calculated separately for cases and controls in each of the above comparisons. Significant disequillibrium was not observed at the $p<0.01$ level. However, and interesting pattern was observed when the GAPDH and p-GAPD SNPs were held constant with the genotypes CG and GG, respectively, and the genotype effect size was measured for the GAPDS SNP. All three sample sets, either individually or combined, exhibit a consistent and strong trend in risk estimate across the three GAPDS genotypes, conditional on the CG-GG two-locus genotype at the other two genes. The GAPDS CC genotype is consistently associated with an increased risk for disease and the GAPDS GG genotype is consistently associated with a decreased risk for disease. These results are shown in Table 9. Each individual sample set p-value was calculated with a two-tailed test of proportions. These results show that determining a person's genotype at these specific loci are predictive of the risk for developing the disease. These results also show that those members of the population with a high-risk genotype may benefit from treatment begun earlier in the course of the disease. These results also show that the population may be stratified into groups that are at low risk, neutral risk or high risk of developing LOAD. Identifying individuals that fall into each group may identify those individuals who will respond better to therapeutic treatment.

Table 8. Significant Association of GAPD genes with LOAD

Strata	Gene Symbol/ID	SNP ID/ Type	Samples	Case allele freq.	Control allele freq.	P-value (allelic)	OR (allelic)	OR 95% CI (allelic)	
All	GAPDH Hcg2005673	hCV2655167 UTR5	Set # 1	29.4	23.1	0.0041	1.39	1.11:	1.74
			Set # 2	27.5	22.9	0.0531	1.28	1.00:	1.64
			Set # 3	28.1	30.9	0.2284	0.87	0.70:	1.09
			All	28.5	25.6	0.0471	1.14	1.00:	1.30
ApoE4-			Set # 1	31.9	21.4	0.0003	1.73	1.28:	2.33
			Set # 2	28.4	22.0	0.0598	1.41	0.98:	2.02
			Set # 3	28.0	30.6	0.4156	o.88	0.65:	1.20
	Interaction w/ApoE: p<0.05		All	29.7	24.7	0.0084	1.28	1.06:	1.53
All	GAPDS hCG23577	hCV337151 TFBS	Set # 3	34.8	41.9	0.00393	0.74	0.60:	0.91
			Set # 2	36.9	42.4	0.0314	0.79	0.64:	0.98
			Set # 1	39.0	40.3	0.66854	0.95	0.75:	1.20
			All	36.6	41.5	0.00114	0.81	0.72:	0.92
LT75			Set # 3	27.8	44.4	0.00001	0.48	0.35:	0.67
			Set # 2	33.9	41.3	0.05433	0.73	0.53:	1.01
			Set # 1	36.1	39.8	0.37672	0.85	0.60:	1.21
	Interaction with ApoE: p<0.01		All	32.6	42.0	0.00003	0.66	0.55:	0.80

109

GE75	p-GAPD hCG40445	hCV2950452 Missence	Set # 1	31.4	38.2	0.03882	0.74	0.56:	0.99
			Set # 2	34.9	33.1	0.67978	1.08	0.74:	1.59
			Set # 3	28.8	34.9	0.04709	0.76	0.57:	1.00
			All	30.8	35.2	0.01828	0.81	0.68:	0.97

Table 9 Multi-locus Genotype Analysis of GAPD Genes

Sample	Multi-locus Genotype*	Case	Control	OR	OR 95% CI	Test of Proportions P-value
All	CG-CC-GG	78	54	1.40	1.11 : 1.76	0.047**
	CG-CG-GG	93	90	1	N/C	
	CG-GG-GG	17	38	0.43	0.31 : 0.6	
Set # 3	CG-CC-GG	33	25	1.41	1.01 : 1.98	0.025
	CG-CG-GG	44	47	1	N/C	
	CG-GG-GG	7	16	0.47	0.28 : 0.77	
Set # 1	CG-CC-GG	17	16	1.24	0.77 : 1.99	0.047
	CG-CG-GG	18	21	1	N/C	
	CG-GG-GG	2	10	.23	0.1 : 0.54	
Set # 2	CG-CC-GG	28	13	1.53	0.99 : 2.36	0.045

110

	CG-CG-GG	31	22	1	N/C	
	CG-GG-GG	8	12	0.47	0.28 : 0.81	

*GAPDH-GAPDS-pGAPD

**"All" sample p-value was derived with Fisher's combined p-value from a two tailed test for sample set # 1 and one tailed tests for the other two sample sets. The p-value was corrected for 27 tests with the Dunn-Sidak correction (uncorrected: p=0.0018).

EXAMPLE 3 A METHOD FOR IDENTIFYING COMPOUNDS THAT PREVENT OR LIMIT GAPD NUCLEAR TRANSLOCATION AND/OR GAPD INDUCED APOPTOSIS

Primary neuronal cell cultures (*e.g.* cerebellar granular cells (Chen *et al. Journal of Neuroscience*, 19:9654-9662 (1999), retinal Müller cells (Kusner *et al., IOVS* 45:1553-1561 (2004)) cerebrocortical cells), or established neuronal cell lines (*e.g.* PC-12 cells, human dopaminergic SH-SY5Y cells, human neuroblastoma PAUJ cells, human neuroblastoma mNB41A3 cells) are either stably transfected or transiently transfected with constructs containing wild-type or variant GAPDH, GAPDS or p-GAPD. The vector constructs may contain inducible or constitutive promoters, or the gene insert may contain its own internal promoter. Suitable vectors, promoters, methods of transfecting primary cell cultures or established cell lines and growing primary cell cultures and established cell lines are well known to those of skill in the art (*Current Protocols in Molecular Biology,* Ausubel *et al.* ed., John Wiley and Sons, Inc. (2004)). The GAPDH gene (hCG2005673) used in these experiments comprises a polynucleotide sequence selected from the group consisting of the genomic sequence of SEQ ID NO:6795, the transcript sequences of SEQ ID NOS:125-127, and nucleic acid sequences that encode a polypeptide comprising an amino acid sequence of SEQ ID NOS:558-560. The GAPDS gene (hCG23577) used in these experiments comprises a polynucleotide sequence selected from the group consisting of the genomic sequence of SEQ ID NO:6828, the transcript sequence of SEQ ID NO:212, and nucleic acid sequences that encode a polypeptide comprising an amino acid sequence of SEQ ID NO:645. The p-GAPD gene (hCG40445) used in these experiments comprises a polynucleotide sequence selected from the group consisting of the genomic sequence of SEQ ID NO:6889, the transcript sequence of SEQ ID NO:413, and nucleic acid sequences that encode a polypeptide comprising an amino acid sequence of SEQ ID NO:846.

Once cells are transfected with the gene/construct of interest, the cells may be treated with the compound of interest and then induced to undergo apoptosis. The compounds include but are not limited to monoamine oxydase inhibitors (e.g., R-(-)-deprenyl, Selegiline, CGP 3446), cholinesterase inhibitors (e.g. tacrine), propyl endopeptidase inhibitors (e.g. ONO-1603) and other apoptosis inhibitors such as Rasagiline, TCH 346 (Novartis). The affects of these compounds in preventing or limiting apoptosis and/or nuclear translocation of GAPDH, GAPDS, and p-GAPD on the cells containing the different forms of GAPDH, GAPDS, and p-GAPD are then determined.

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 5, page 1 of 4

hCV	Allele 1	Sequence A	Sequence B	Sequence C
hCV10272	A	ACCCAGTGCTTGTATCAGAA (SEQ ID NO: 54770)	TACCCAGTGCTTGTATCAGAT (SEQ ID NO: 54771)	CAGACTCTGGGTCACAGTGA (SEQ ID NO: 54772)
hCV10546	G	ACAGAGTAATCTAGAATGCAAAGG (SEQ ID NO: 54773)	ACAGAGTAATCTAGAATGCAAAGC (SEQ ID NO: 54774)	TGCAGTACCCATCGTGTATTT (SEQ ID NO: 54775)
hCV11192	A	AACAACATTGCTTCCCAA (SEQ ID NO: 54776)	AACAACATTGCTTCCCAG (SEQ ID NO: 54777)	TGCCCTTTTTCTTCAGAATC (SEQ ID NO: 54778)
hCV11193	G	CAGAAACTCCGAGGAGACA (SEQ ID NO: 54779)	AGAAACTCCGAGGAGACG (SEQ ID NO: 54780)	CCCTGGGTGCAGACATT (SEQ ID NO: 54781)
hCV11200	T	AAGGCACTGAGAGATTCAGTAAC (SEQ ID NO: 54782)	AAGGCACTGAGAGATTCAGTAAT (SEQ ID NO: 54783)	AAACAGGAGCTGAGAGAGAATACTA (SEQ ID NO: 54784)
hCV11214	C	ATTGATGGCACAACTCTGA (SEQ ID NO: 54785)	TGATGGCACAACTCTGC (SEQ ID NO: 54786)	TCAGAGGGCTTCCTCTTCT (SEQ ID NO: 54787)
hCV11214	A	CCATTCCCACTCTAGACTTGA (SEQ ID NO: 54788)	CATTCCCACTCTAGACTTGG (SEQ ID NO: 54789)	GACACTTCCATCAAAGCAGTATTA (SEQ ID NO: 54790)
hCV11278	T	CCAAATGCCACTGAACG (SEQ ID NO: 54791)	ACCAAATGCCACTGAACA (SEQ ID NO: 54792)	TTTGCTAAACAATTCCTCACTACT (SEQ ID NO: 54793)
hCV11396	G	TTTTAGTAGGCCTATAACTTAAGGG (SEQ ID NO: 54794)	AATTTTAGTAGGCCTATAACTTAAGGT (SEQ ID NO: 54795)	CAGCAGGGACAAATCTCTAATC (SEQ ID NO: 54796)
hCV11566	G	CTGTTGCTTCTTTTTGTCTTG (SEQ ID NO: 54797)	TCTGTTGCTTCTTTTTGTCTTC (SEQ ID NO: 54798)	CAAATGATGACCTCTCAGTCTATT (SEQ ID NO: 54799)
hCV11568	G	ACAGTTTCCTCAATCTTTTCC (SEQ ID NO: 54800)	TAACAGTTTCCTCAATCTTTTCA (SEQ ID NO: 54801)	CATGTTGCCAAAATATGATTATAA (SEQ ID NO: 54802)
hCV11574	A	GCTGGACACGAACCAT (SEQ ID NO: 54803)	GCTGGACACGAACCAC (SEQ ID NO: 54804)	GCACCCCCTGGAACAG (SEQ ID NO: 54805)
hCV11595	C	AGAAGAAGCCCTGTACTCAAC (SEQ ID NO: 54806)	AGAAGAAGCCCTGTACTCAAG (SEQ ID NO: 54807)	TTGTGGAATGCATTTCTAATTATAT (SEQ ID NO: 54808)
hCV11597	A	GAAGTTCTTACCACACTGACTACA (SEQ ID NO: 54809)	GAAGTTCTTACCACACTGACTACC (SEQ ID NO: 54810)	AATAACTGTGGGAAAATACTTAACAC (SEQ ID NO: 54811)
hCV11597	A	GAAGTTCTTACCACACTGACTACA (SEQ ID NO: 54812)	GAAGTTCTTACCACACTGACTACC (SEQ ID NO: 54813)	GAGAAACCCTGTGACCATAATA (SEQ ID NO: 54814)
hCV11597	T	CAACATTGCAAGATGCC (SEQ ID NO: 54815)	GCAACATTGCAAGATGCT (SEQ ID NO: 54816)	TTTTGACAAACAAAGTCACTTAGAC (SEQ ID NO: 54817)
hCV11720	T	TGGGAAATTCAAGGCG (SEQ ID NO: 54818)	CTGGGAAATTCAAGGCA (SEQ ID NO: 54819)	TTTAAGTCCTGGGTAAACTAAATAGA (SEQ ID NO: 54820)
hCV11720	T	GGCATGGCAGGACTACG (SEQ ID NO: 54821)	GGCATGGCAGGACTACA (SEQ ID NO: 54822)	GGACTCCAAAGGAAGGTCAA (SEQ ID NO: 54823)
hCV11840	T	ACAATATGCCTAAGATCCCG (SEQ ID NO: 54824)	AACAATATGCCTAAGATCCCA (SEQ ID NO: 54825)	CAGATGAAGAAACTGAGTCATAGAG (SEQ ID NO: 54826)
hCV11841	A	ATAAGTCTTTTATCACCTTTAGGCTA (SEQ ID NO: 54827)	AGTCTTTTATCACCTTTAGGCTG (SEQ ID NO: 54828)	AATCAATTGGCAAATAAGAATGTA (SEQ ID NO: 54829)
hCV11842	T	GGGATTCCAAGCTGACG (SEQ ID NO: 54830)	AAGGGATTCCAAGCTGACT (SEQ ID NO: 54831)	GGGAAGGCCAGGTTCTAC (SEQ ID NO: 54832)
hCV11855	A	GATGTCCCATCTATTAGATGAGT (SEQ ID NO: 54833)	ATGTCCCATCTATTAGATGAGC (SEQ ID NO: 54834)	CGCAAACCTTTCTGAAGATATTA (SEQ ID NO: 54835)
hCV11861	C	TAAATTAGCACAAGGGAACTTC (SEQ ID NO: 54836)	TAAATTAGCACAAGGGAACTTT (SEQ ID NO: 54837)	AATGGCATGCACAGATCTTA (SEQ ID NO: 54838)
hCV11912	G	GGAGCCGGCAAGCA (SEQ ID NO: 54839)	GAGCCGGCAAGCG (SEQ ID NO: 54840)	GGCCTGGTCTGGTTTCAG (SEQ ID NO: 54841)
hCV12029	T	GCACCGTCCTTCG (SEQ ID NO: 54842)	CGCACCGTCCTTCA (SEQ ID NO: 54843)	GCAAGTGTGGAGTAGCTTTCTG (SEQ ID NO: 54844)
hCV12123	C	TCATCAGGTAACTGATTTCCTC (SEQ ID NO: 54845)	TCATCAGGTAACTGATTTCCTT (SEQ ID NO: 54846)	TGGCACTGCTGTGTGTCT (SEQ ID NO: 54847)
hCV12126	C	AACAGTTGTTCTTGTGAATCATC (SEQ ID NO: 54848)	ACAGTTGTTCTTGTGAATCATG (SEQ ID NO: 54849)	CCATTGCCAGAAAATGACT (SEQ ID NO: 54850)
hCV12126	T	CTTTCCGAACAATCTGGG (SEQ ID NO: 54851)	TACTTTCCGAACAATCTGGA (SEQ ID NO: 54852)	AGCTCTGGGAAACAAATGTC (SEQ ID NO: 54853)
hCV12126	A	CAGGTCCATGACCAACAA (SEQ ID NO: 54854)	CAGGTCCATGACCAACAC (SEQ ID NO: 54855)	GGACATCATCCCTACATCTACTAGT (SEQ ID NO: 54856)
hCV12296	C	AGGTGAGTGTCAGGTGC (SEQ ID NO: 54857)	CAGGTGAGTGTCAGGTGT (SEQ ID NO: 54858)	GACCTTGAGTTTCTGTTCACATAC (SEQ ID NO: 54859)
hCV12296	A	CAAAGGAATTTCAGGAGGAA (SEQ ID NO: 54860)	CAAAGGAATTTCAGGAGGAG (SEQ ID NO: 54861)	GCCATAGCCAGCAATCAC (SEQ ID NO: 54862)
hCV12297	A	GGGGATACAGTGCCTGA (SEQ ID NO: 54863)	GGGGATACAGTGCCTGC (SEQ ID NO: 54864)	ACCTCCTGAGGACAAGTCAC (SEQ ID NO: 54865)
hCV12448	C	CGGGAGTACTGAGGGAGAC (SEQ ID NO: 54866)	CGGGAGTACTGAGGGAGAG (SEQ ID NO: 54867)	CAGGGTGAGGATTTCATCAG (SEQ ID NO: 54868)
hCV13056	A	GAGATAAAGGCAAGGAGTCT (SEQ ID NO: 54869)	AGATAAAGGCAAGGAGTCA (SEQ ID NO: 54870)	TCCTGAATGCTGCTCTTCT (SEQ ID NO: 54871)
hCV13224	A	GCCTCTGGGATGAAAAAGA (SEQ ID NO: 54872)	CCTCTGGGATGAAAAAGC (SEQ ID NO: 54873)	GCAGGAGCCTGGGTTCT (SEQ ID NO: 54874)
hCV13458	T	GAGAGTCTCCTCTCCTCTAAGG (SEQ ID NO: 54875)	GGAGAGTCTCCTCTCCTCTAAGT (SEQ ID NO: 54876)	TGCATCCCAAGATTTGTTG (SEQ ID NO: 54877)
hCV13458	C	TAAAGATTTACCTATTTGGTGGAG (SEQ ID NO: 54878)	AAAGATTTACCTATTTGGTGGAA (SEQ ID NO: 54879)	CAGTCAGGGAGAAGAGAAGATC (SEQ ID NO: 54880)
hCV13458	A	GGTTTTCAGTGACATCCA (SEQ ID NO: 54881)	GTTTTCAGTGACATCCG (SEQ ID NO: 54882)	TGTTCTTTTTCTCTAAAGTATCTTT (SEQ ID NO: 54883)
hCV13485	G	CCTTTTAAATGGTAGGGAGAGTAT (SEQ ID NO: 54884)	CCTTTTAAATGGTAGGGAGAGTAC (SEQ ID NO: 54885)	CAAAGCTTGGGAATGTTTTC (SEQ ID NO: 54886)
hCV14068	C	CTACTGTACCTTTCCAACTTATCC (SEQ ID NO: 54887)	GCTACTGTACCTTTCCAACTTATCT (SEQ ID NO: 54888)	TTAACTTGTTTTTGCTGTCTTACAG (SEQ ID NO: 54889)
hCV14132	A	TAGGAGGGTGAAAAGTGGA (SEQ ID NO: 54890)	GGAGGGTGAAAAGTGGG (SEQ ID NO: 54891)	GGCAGCGTGCTCAGAC (SEQ ID NO: 54892)
hCV14199	G	CTAGGCAGTCTGCCTCAAC (SEQ ID NO: 54893)	CTAGGCAGTCTGCCTCAAG (SEQ ID NO: 54894)	AATTAAAGAATTTGTGATCAATGTACT (SEQ ID NO: 54895)
hCV14899	C	ACTTCCTGAGGCTGTAGATG (SEQ ID NO: 54896)	CACTTCCTGAGGCTGTAGATA (SEQ ID NO: 54897)	GCCAACTTACCATTTGATTTTAG (SEQ ID NO: 54898)
hCV15074	G	CAGCTACATTGCTTCTCTTACTTA (SEQ ID NO: 54899)	AGCTACATTGCTTCTCTTACTTG (SEQ ID NO: 54900)	TGTCTTACCCAACAAAAAGTTAGT (SEQ ID NO: 54901)
hCV15585	G	AAAGCCACTTTGAAATCCTC (SEQ ID NO: 54902)	AAGCCACTTTGAAATCCTG (SEQ ID NO: 54903)	AGTTCCTGCTTTGCTTTACAG (SEQ ID NO: 54904)
hCV15585	C	CCATGGTGATTGCCTC (SEQ ID NO: 54905)	CCATGGTGATTGCCTT (SEQ ID NO: 54906)	ATTCCTCGAGCTGTGAGATT (SEQ ID NO: 54907)
hCV15806	C	CTATCAAACCGTATGCTCTTAAG (SEQ ID NO: 54908)	CTATCAAACCGTATGCTCTTAAC (SEQ ID NO: 54909)	AATCGTAAATGGGAGATAGATACTC (SEQ ID NO: 54910)
hCV15811	T	TGGCAGCGTGTGTGG (SEQ ID NO: 54911)	TTTGGCAGCGTGTGTGT (SEQ ID NO: 54912)	TCTAGTCCCCCTGACTCTGTT (SEQ ID NO: 54913)
hCV15870	T	CGCAATTCCATTCCTAGC (SEQ ID NO: 54914)	CCGCAATTCCATTCCTAGT (SEQ ID NO: 54915)	GCCAGGGCAGCAATCT (SEQ ID NO: 54916)
hCV15873	T	TTGCAGGTCCTTATCCAA (SEQ ID NO: 54917)	CTTGCAGGTCCTTATCCAT (SEQ ID NO: 54918)	GAGGGACAAATTCCTTCTTG (SEQ ID NO: 54919)
hCV15887	A	AAGCAGTCCAGGATGGTA (SEQ ID NO: 54920)	AGCAGTCCAGGATGGTG (SEQ ID NO: 54921)	TCTACGTGGGATGAACAGAAG (SEQ ID NO: 54922)
hCV15887	T	CCAATACTTCCTCTTTAGCTTG (SEQ ID NO: 54923)	ACCAATACTTCCTCTTTAGCTTT (SEQ ID NO: 54924)	TCAGGTGGTGGACATCATAC (SEQ ID NO: 54925)
hCV15887	T	GGGCACTTAACAATGGAG (SEQ ID NO: 54926)	GGGCACTTAACAATGGAA (SEQ ID NO: 54927)	GGGTGGTACATTCTCAAGTAAAA (SEQ ID NO: 54928)
hCV15919	A	ATGGCTCACTTTTTATTCCAT (SEQ ID NO: 54929)	ATGGCTCACTTTTTATTCCAC (SEQ ID NO: 54930)	CTGCAGACGCTGAGAACTATAG (SEQ ID NO: 54931)
hCV15961	C	GCTTACTTGTTGGTCTGTGAC (SEQ ID NO: 54932)	GCTTACTTGTTGGTCTGTGAT (SEQ ID NO: 54933)	CCTGAACCTGGTTTCAAATATA (SEQ ID NO: 54934)

TABLE 5, page 2 of 4

hCV	Allele 1	Sequence A	Sequence B	Sequence C
hCV15965	G	TACACTACACTTTCTGTTTCAACTTA (SEQ ID NO: 54935)	ACTACACTTTCTGTTTCAACTTG (SEQ ID NO: 54936)	CTTGAGGTTCATGAGAATGTAATC (SEQ ID NO: 54937)
hCV16111	C	TCCCTTCTGGGTTGTTTATC (SEQ ID NO: 54938)	TCCCTTCTGGGTTGTTTATT (SEQ ID NO: 54939)	TCCTCCAAACAGAACAGGTT (SEQ ID NO: 54940)
hCV16113	A	GGGTTTTGGTCTGAGCA (SEQ ID NO: 54941)	GGGTTTTGGTCTGAGCG (SEQ ID NO: 54942)	TCAACGTCCAAATCTGACTTTA (SEQ ID NO: 54943)
hCV16190	A	AGCGACTCCTGAGTGACTT (SEQ ID NO: 54944)	AGCGACTCCTGAGTGACTC (SEQ ID NO: 54945)	ACATAGCCTGGGAGTAATGAA (SEQ ID NO: 54946)
hCV16221	T	TTTATTCTTCATCTGGCATTC (SEQ ID NO: 54947)	ATTTTATTCTTCATCTGGCATTA (SEQ ID NO: 54948)	TGAACACAGGGCTTTATACTAGATA (SEQ ID NO: 54949)
hCV16248	T	ATCTTGAAAGGTTACGTGATG (SEQ ID NO: 54950)	CAATCTTGAAAGGTTACGTGATA (SEQ ID NO: 54951)	AACCTTAGCAACACTAATTTGTTCT (SEQ ID NO: 54952)
hCV16248	G	CCCTGGGCTTTATTTCC (SEQ ID NO: 54953)	CCCTGGGCTTTATTTCG (SEQ ID NO: 54954)	TGCTGAGTCCCAAAGACTATTT (SEQ ID NO: 54955)
hCV16289	C	AATTGTGGAAATGCTGTCG (SEQ ID NO: 54956)	AAATTGTGGAAATGCTGTCA (SEQ ID NO: 54957)	CTTTGAGGTGCTCAATGTCA (SEQ ID NO: 54958)
hCV16513	A	GACCCTACAGAGCAGCAGA (SEQ ID NO: 54959)	ACCCTACAGAGCAGCAGG (SEQ ID NO: 54960)	CATTGCTACTATTCCTTGATGTG (SEQ ID NO: 54961)
hCV16651	C	AACTCAGACGAAATTGACCC (SEQ ID NO: 54962)	GAACTCAGACGAAATTGACCT (SEQ ID NO: 54963)	AATAGGTACTCCATGAAAATATGTTG (SEQ ID NO: 54964)
hCV16652	T	CCACATTCCCTTGTTTAGTC (SEQ ID NO: 54965)	CCACATTCCCTTGTTTAGTT (SEQ ID NO: 54966)	CCTTACTCTGGCTTTCAATCAC (SEQ ID NO: 54967)
hCV16875	G	CATCAGAGCTTTTTCCTTTG (SEQ ID NO: 54968)	CATCAGAGCTTTTTCCTTTC (SEQ ID NO: 54969)	CCACTTCCCCTCTTCTTTC (SEQ ID NO: 54970)
hCV17806	A	CGTAAGGTTTTTCTTCTGTTACCT (SEQ ID NO: 54971)	GTAAGGTTTTTCTTCTGTTACCC (SEQ ID NO: 54972)	TGTTTCCCTTCCTCTAGAGATATACT (SEQ ID NO: 54973)
hCV17917	G	AAGAAAACCTATTACCAAGTATTTACTATAT (SEQ ID NO: 54974)	AGAAAACCTATTACCAAGTATTTACTATAC (SEQ ID NO: 54975)	CCTGAGGTTGTTTCACAATTAAC (SEQ ID NO: 54976)
hCV17928	C	GGTCATAATCTGGTCATCAG (SEQ ID NO: 54977)	GGTCATAATCTGGTCATCAA (SEQ ID NO: 54978)	GAAGCTAGAATAAACGATCAGAACTAT (SEQ ID NO: 54979)
hCV17928	T	GGAGATTCCCAGAATG (SEQ ID NO: 54980)	GGGAGATTCCCAGAATA (SEQ ID NO: 54981)	GCTCCATAGCATCTTGTAC (SEQ ID NO: 54982)
hCV17928	G	ATGCTAGACAGTTTAATTATCTGGT (SEQ ID NO: 54983)	GCTAGACAGTTTAATTATCTGGC (SEQ ID NO: 54984)	CATACACAGGCAGATGATTTACA (SEQ ID NO: 54985)
hCV18011	G	CCTGAGATGCCTCTTTGG (SEQ ID NO: 54986)	GTCCTGAGATGCCTCTTTGT (SEQ ID NO: 54987)	TCACAGAGCTCTCTGAAACATC (SEQ ID NO: 54988)
hCV18222	C	GATTTTAAAGCCAGGAACATT (SEQ ID NO: 54989)	ATTTTAAAGCCAGGAACATG (SEQ ID NO: 54990)	GCCCATTTTGTTTCTCTACATT (SEQ ID NO: 54991)
hCV18222	A	CGGAGCTCCTTAAGAATTACAT (SEQ ID NO: 54992)	CGGAGCTCCTTAAGAATTACAA (SEQ ID NO: 54993)	TCCAGATGCAGGCATGTAC (SEQ ID NO: 54994)
hCV18249	C	TGGCTTCTTTGATTTCAGGT (SEQ ID NO: 54995)	GGCTTCTTTGATTTCAGGG (SEQ ID NO: 54996)	CAATCACCAGCATTCCTCTT (SEQ ID NO: 54997)
hCV18393	C	ACATACACAACCCGCATTA (SEQ ID NO: 54998)	CATACACAACCCGCATTC (SEQ ID NO: 54999)	GATGTCATTCTTTGGAGTGTTACTA (SEQ ID NO: 55000)
hCV18393	C	AAATTCTGTGGAGAATCTTCAG (SEQ ID NO: 55001)	AAATTCTGTGGAGAATCTTCAA (SEQ ID NO: 55002)	CCTCGATGATTCACAATACAA (SEQ ID NO: 55003)
hCV18393	G	AGGGTTTCTCCTCTGTATGAC (SEQ ID NO: 55004)	AGGGTTTCTCCTCTGTATGAG (SEQ ID NO: 55005)	GAATGGCCAGTTAAAAGAATCT (SEQ ID NO: 55006)
hCV18418	A	CCCTTCCTGAATTTGTCTAAA (SEQ ID NO: 55007)	CCCTTCCTGAATTTGTCTAAG (SEQ ID NO: 55008)	GGCTTGCCCTTCTTTAAAAC (SEQ ID NO: 55009)
hCV18452	C	CTTCAGCGGCTCACG (SEQ ID NO: 55010)	CCTTCAGCGGCTCACA (SEQ ID NO: 55011)	TCCACATCCTCTTGTGTCTATCT (SEQ ID NO: 55012)
hCV18479	G	CATGACTAATGACTCTTCCACAT (SEQ ID NO: 55013)	CATGACTAATGACTCTTCCACAC (SEQ ID NO: 55014)	TCTTTTCCAGCAGATCAATG (SEQ ID NO: 55015)
hCV18534	G	GCTTAGACGCTGCTGGATAT (SEQ ID NO: 55016)	GCTTAGACGCTGCTGGATAC (SEQ ID NO: 55017)	CTACCTTAGTGCATCAAACATTAAT (SEQ ID NO: 55018)
hCV18739	A	ACCATTATAGAAAGACTCACTTTAAGA (SEQ ID NO: 55019)	CCATTATAGAAAGACTCACTTTAAGG (SEQ ID NO: 55020)	TCTTGCATTCAATCAATTTGTAT (SEQ ID NO: 55021)
hCV19112	C	CCAGCTCATTGTAATCCAGAC (SEQ ID NO: 55022)	CCAGCTCATTGTAATCCAGAT (SEQ ID NO: 55023)	CGGATGCCTCCCACAGT (SEQ ID NO: 55024)
hCV19112	C	CCAGCTCATTGTAATCCAGAC (SEQ ID NO: 55025)	CCAGCTCATTGTAATCCAGAT (SEQ ID NO: 55026)	CGGTGCCTTTGGTGAAG (SEQ ID NO: 55027)
hCV19112	T	AGTGGGCTGTGAAACTACAG (SEQ ID NO: 55028)	AGTGGGCTGTGAAACTACAA (SEQ ID NO: 55029)	AAGTGTGGTGGCTGATACTG (SEQ ID NO: 55030)
hCV19130	A	GCCATCAGCCGAACA (SEQ ID NO: 55031)	GCCATCAGCCGAACC (SEQ ID NO: 55032)	CCATCTGGGCCTGACTTATA (SEQ ID NO: 55033)
hCV19206	A	CTGCTCTTGGTGGACA (SEQ ID NO: 55034)	TGCTCTTGGTGGACG (SEQ ID NO: 55035)	GCTATATAAGCTGCTTCTCTCTT (SEQ ID NO: 55036)
hCV19461	G	CAGCCAGATTTCCTCTGTT (SEQ ID NO: 55037)	CAGCCAGATTTCCTCTGTC (SEQ ID NO: 55038)	TCGGGATGCACTGTTCTT (SEQ ID NO: 55039)
hCV19917	A	CCCACAGGTGGAACCA (SEQ ID NO: 55040)	CCACAGGTGGAACCG (SEQ ID NO: 55041)	CAGCGCTGGACTCAAAA (SEQ ID NO: 55042)
hCV20274	A	GAGCTGCCTGCCAATAGT (SEQ ID NO: 55043)	GCTGCCTGCCAATAGC (SEQ ID NO: 55044)	GGGCCATCGTCTTGTAGA (SEQ ID NO: 55045)
hCV20282	T	CAAGATGCATACAGTGCTG (SEQ ID NO: 55046)	CCAAGATGCATACAGTGCTA (SEQ ID NO: 55047)	CCAAGCTAACAGTTCCATACAAAC (SEQ ID NO: 55048)
hCV20283	C	TGGATGATTACTGATATGTGTGTC (SEQ ID NO: 55049)	TGGATGATTACTGATATGTGTGTT (SEQ ID NO: 55050)	GAAGGATTGCCTTCAATAAAGA (SEQ ID NO: 55051)
hCV21160	A	GGCCACGTGGTTAGT (SEQ ID NO: 55052)	GCCACGTGGTTAGC (SEQ ID NO: 55053)	GCTAGGCTGCACATTTAT (SEQ ID NO: 55054)
hCV21164	C	CCAAGCAAACCTAATGACAC (SEQ ID NO: 55055)	CCAAGCAAACCTAATGACAT (SEQ ID NO: 55056)	GGCTCACCTTTTTCTTAAATATCT (SEQ ID NO: 55057)
hCV21319	A	TCTCTAAAGTCCATCTATTTTCACT (SEQ ID NO: 55058)	CTCTAAAGTCCATCTATTTTCACC (SEQ ID NO: 55059)	GAAAGGAAGCCAGGAGTAAA (SEQ ID NO: 55060)
hCV21441	C	CCACTTCAGTCCTGAAGAGC (SEQ ID NO: 55061)	CCACTTCAGTCCTGAAGAGG (SEQ ID NO: 55062)	TCGTAGTGCTGGGAGTTTCT (SEQ ID NO: 55063)
hCV21441	C	GACTTGTGTTCTCATCCAG (SEQ ID NO: 55064)	GACTTGTGTTCTCATCCAC (SEQ ID NO: 55065)	AAGAAGCAAGCTGAGAAA (SEQ ID NO: 55066)
hCV21532	C	AGTGGGTGCAAAGTTCC (SEQ ID NO: 55067)	AGAGTGGGTGCAAAGTTCT (SEQ ID NO: 55068)	CAAGGATGAAGTAGAATTTGTTTT (SEQ ID NO: 55069)
hCV21707	C	CCAAGAAAAAGTGCACAGAC (SEQ ID NO: 55070)	CCAAGAAAAAGTGCACAGAG (SEQ ID NO: 55071)	TCAGGCAAAGAAAGGTAACTAGT (SEQ ID NO: 55072)
hCV22647	T	CTTAGATTCCATCTCTACAAAGAAC (SEQ ID NO: 55073)	CTTAGATTCCATCTCTACAAAGAAT (SEQ ID NO: 55074)	GCCAGGGACCAAACTGA (SEQ ID NO: 55075)
hCV23027	C	ATAAACACCTTTTATCAGGAATTG (SEQ ID NO: 55076)	ATAAACACCTTTTATCAGGAATTC (SEQ ID NO: 55077)	CGATTCCACGGGTTAGATC (SEQ ID NO: 55078)
hCV23027	T	GTATCATCAGCCTCAAAAGAAG (SEQ ID NO: 55079)	TATCATCAGCCTCAAAAGAAA (SEQ ID NO: 55080)	GGGCACATTTTCCACATAG (SEQ ID NO: 55081)
hCV25393	T	GGACGGGGTATCACTCTC (SEQ ID NO: 55082)	GGACGGGGTATCACTCTT (SEQ ID NO: 55083)	GCTGGTGCCCACTACTTG (SEQ ID NO: 55084)
hCV25596	T	CCCCAGATTCCCAAAC (SEQ ID NO: 55085)	CCCCAGATTCCCAAAA (SEQ ID NO: 55086)	CCCGCCCATCAGAGA (SEQ ID NO: 55087)
hCV25602	A	TGTAGCTCTTTGTGATGTATAGAGA (SEQ ID NO: 55088)	CTGTAGCTCTTTGTGATGTATAGAGT (SEQ ID NO: 55089)	TCACTGGCCCGATTTTAC (SEQ ID NO: 55090)
hCV25603	C	AATATCCAGAGGCATTTATCG (SEQ ID NO: 55091)	CAATATCCAGAGGCATTTATCA (SEQ ID NO: 55092)	TGCAGCACTTTGATACTATCTACA (SEQ ID NO: 55093)
hCV25603	T	ATGGTCCTTTGAAAGAGCTAG (SEQ ID NO: 55094)	ATGGTCCTTTGAAAGAGCTAA (SEQ ID NO: 55095)	CATTATCCCCAGAGGAGTTGT (SEQ ID NO: 55096)
hCV25606	T	GGCCTATGAGAGATGATTCC (SEQ ID NO: 55097)	GAGGCCTATGAGAGATGATTCT (SEQ ID NO: 55098)	TCTGAATTGGCTCAATGATG (SEQ ID NO: 55099)

WO 2006/022639 PCT/US2004/024001

TABLE 5, page 3 of 4

hCV	Allele 1	Sequence A	Sequence B	Sequence C
hCV25625	A	AGCTGAGAAGGTGAGCACTA (SEQ ID NO: 55100)	CTGAGAAGGTGAGCACTG (SEQ ID NO: 55101)	CATACCTGATGTTCCAAAAACTAC (SEQ ID NO: 55102)
hCV25636	G	CCTCGGCTTTCTCAAAGT (SEQ ID NO: 55103)	CTCGGCTTTCTCAAAGC (SEQ ID NO: 55104)	GCACGCCAGCAAGTTG (SEQ ID NO: 55105)
hCV25637	C	CTCACACCTTACTTTTCCAG (SEQ ID NO: 55106)	CTCACACCTTACTTTTCCAA (SEQ ID NO: 55107)	CCTGCCGACCCTCTCTT (SEQ ID NO: 55108)
hCV25744	A	CACAAAGGTGACTTCCA (SEQ ID NO: 55109)	CACAAAGGTGACTTCCG (SEQ ID NO: 55110)	TGCCCCTGTTTTTGACA (SEQ ID NO: 55111)
hCV25752	A	AGGCCTTGGGCAGAA (SEQ ID NO: 55112)	AAGGCCTTGGGCAGAT (SEQ ID NO: 55113)	GTCACTGCCACCTCTTTGA (SEQ ID NO: 55114)
hCV25766	T	CAGTGGATGCCTTCACAC (SEQ ID NO: 55115)	CAGTGGATGCCTTCACAT (SEQ ID NO: 55116)	GAGTGCAGCTTCCAAGAAAC (SEQ ID NO: 55117)
hCV25923	G	AACCCAGATACCAAGAGGAC (SEQ ID NO: 55118)	AACCCAGATACCAAGAGGAA (SEQ ID NO: 55119)	GCTGTGTGAGCACACACTTCT (SEQ ID NO: 55120)
hCV25938	T	GGAAAAGAAGAGGCAACATG (SEQ ID NO: 55121)	GGAAAAGAAGAGGCAACATT (SEQ ID NO: 55122)	AACTCGCCAGCATCACA (SEQ ID NO: 55123)
hCV25970	T	GCCATGGTTTTGGAAGAGT (SEQ ID NO: 55124)	GCCATGGTTTTGGAAGAGA (SEQ ID NO: 55125)	TTCCTTTAACTTTCATGATCACTAA (SEQ ID NO: 55126)
hCV25992	G	CAATAATTTTTTCCAGGTTGTC (SEQ ID NO: 55127)	CAATAATTTTTTCCAGGTTGTG (SEQ ID NO: 55128)	CACACTATGATTGTCAGAAACATG (SEQ ID NO: 55129)
hCV26551	C	CATATGAATGGTAGAGATGGG (SEQ ID NO: 55130)	CATATGAATGGTAGAGATGGC (SEQ ID NO: 55131)	AGATGCCCTAGACTCAACTCA (SEQ ID NO: 55132)
hCV26551	T	TCGAAGATTAATTGTAGACATACATAG (SEQ ID NO: 55133)	TCGAAGATTAATTGTAGACATACATAT (SEQ ID NO: 55134)	TGGTGGAATCCTGGCTATTA (SEQ ID NO: 55135)
hCV26551	G	CGAGCCACATCGCTC (SEQ ID NO: 55136)	CGAGCCACATCGCTG (SEQ ID NO: 55137)	CCGCAAGGCTCGTAGAC (SEQ ID NO: 55138)
hCV26827	T	CATTTACCTTCCCAGATGTTC (SEQ ID NO: 55139)	CATTTACCTTCCCAGATGTTT (SEQ ID NO: 55140)	TTTTCTTCAGCTTGAAAGATCTAA (SEQ ID NO: 55141)
hCV26858	A	GTGAAGCCTGCCACACT (SEQ ID NO: 55142)	TGAAGCCTGCCACACC (SEQ ID NO: 55143)	GCCAGTGGCAATGGTAAC (SEQ ID NO: 55144)
hCV27341	G	ATAAAACTGGGCTGCATATC (SEQ ID NO: 55145)	GATAAAACTGGGCTGCATATA (SEQ ID NO: 55146)	AAAGATGCACACATTAAGGTTATC (SEQ ID NO: 55147)
hCV27576	G	GGCAGCTAGGCCGTCT (SEQ ID NO: 55148)	GCAGCTAGGCCGTCC (SEQ ID NO: 55149)	GACCCCCACAGGAAGAAG (SEQ ID NO: 55150)
hCV27576	G	GGCAGCTAGGCCGTCT (SEQ ID NO: 55151)	GGCAGCTAGGCCGTCC (SEQ ID NO: 55152)	ACCCCCACAGGAAGAAG (SEQ ID NO: 55153)
hCV27604	C	GAAGGTAGGGAGAGGAATGAC (SEQ ID NO: 55154)	GAAGGTAGGGAGAGGAATGAG (SEQ ID NO: 55155)	CACTCTGTCTGGCAGAATAATTATA (SEQ ID NO: 55156)
hCV28693	G	AAACAATGTTTCCAGTAAACTAGTAG (SEQ ID NO: 55157)	AAACAATGTTTCCAGTAAACTAGTAC (SEQ ID NO: 55158)	GATCACCCCTGAAAGACTATTT (SEQ ID NO: 55159)
hCV28756	G	CCGTCTGCACTGAATCTG (SEQ ID NO: 55160)	ACCGTCTGCACTGAATCTC (SEQ ID NO: 55161)	CGAACTGGCCTAGAGTCAA (SEQ ID NO: 55162)
hCV29457	T	CTTTCCAGTGGCTATGGAC (SEQ ID NO: 55163)	TTTCCAGTGGCTATGGAA (SEQ ID NO: 55164)	TGCTGGTGGCACTGAAT (SEQ ID NO: 55165)
hCV29504	A	ACCCTTTTGGCTCCCT (SEQ ID NO: 55166)	CCTTTTGGCTCCCC (SEQ ID NO: 55167)	TGCTGGTGCTGAGTATATCATG (SEQ ID NO: 55168)
hCV29522	T	CAGATCTTAATTTTGTCACGATG (SEQ ID NO: 55169)	ACAGATCTTAATTTTGTCACGATT (SEQ ID NO: 55170)	CCTTCCAAGCTGATGATTCT (SEQ ID NO: 55171)
hCV29812	T	ATAAACTCTGGATTTTGCTAATGT (SEQ ID NO: 55172)	AAACTCTGGATTTTGCTAATGA (SEQ ID NO: 55173)	CAGTTCCCCCAACAGTAACA (SEQ ID NO: 55174)
hCV29812	T	CAGCATGAATGCCTATTTATC (SEQ ID NO: 55175)	CAGCATGAATGCCTATTTATT (SEQ ID NO: 55176)	CCCAAAATGCTGGGATTATA (SEQ ID NO: 55177)
hCV29932	T	GTGGTAGGGGAGGAAGTG (SEQ ID NO: 55178)	GGTGGTAGGGGAGGAAGTA (SEQ ID NO: 55179)	ACAGCTTACTGTCTTTATCATTATCAC (SEQ ID NO: 55180)
hCV30273	T	CGGGGGATACAAGGG (SEQ ID NO: 55181)	CCGGGGGATACAAGGA (SEQ ID NO: 55182)	CGCCTCGCTGGATAGAC (SEQ ID NO: 55183)
hCV30394	A	CGTTTGCAAGCTGGAA (SEQ ID NO: 55184)	CGTTTGCAAGCTGGAG (SEQ ID NO: 55185)	CCTCAAATGCTCATTTCTTCT (SEQ ID NO: 55186)
hCV30461	A	TTCTACAATGTCTGAAGAAGTGAA (SEQ ID NO: 55187)	TTCTACAATGTCTGAAGAAGTGAC (SEQ ID NO: 55188)	CTGAACTCCTACCTCTTTTTCTTAG (SEQ ID NO: 55189)
hCV30523	T	TGCCATGATGCCTACG (SEQ ID NO: 55190)	TTGCCATGATGCCTACA (SEQ ID NO: 55191)	ACTGCACTAGCATCAGATGTCT (SEQ ID NO: 55192)
hCV30887	A	AGAAATGACTTGTAAGATCATTCA (SEQ ID NO: 55193)	GAGAAATGACTTGTAAGATCATTCT (SEQ ID NO: 55194)	GCTTTGTGGAAAACATTCTGTA (SEQ ID NO: 55195)
hCV30913	C	AATCCAGGGAAACCTCTAGTT (SEQ ID NO: 55196)	ATCCAGGGAAACCTCTAGTG (SEQ ID NO: 55197)	GGATGGGAGCAAAGATGA (SEQ ID NO: 55198)
hCV31329	A	ACAGGTACAGCAGTGCTTTT (SEQ ID NO: 55199)	ACAGGTACAGCAGTGCTTTC (SEQ ID NO: 55200)	ATAAGGTCCTGATCAGAATCATC (SEQ ID NO: 55201)
hCV31378	C	TCCTCATTATTGGCAGGTG (SEQ ID NO: 55202)	TCCTCATTATTGGCAGGTC (SEQ ID NO: 55203)	GGCATCTGCAGTTTACAATTATT (SEQ ID NO: 55204)
hCV31595	C	AATAATTCCGAATCTAGTTTGAC (SEQ ID NO: 55205)	AATAATTCCGAATCTAGTTTGAT (SEQ ID NO: 55206)	ATCTCAACCTTCTGTCTTGATCT (SEQ ID NO: 55207)
hCV31595	G	GGATAGTTCCATCTGCCT (SEQ ID NO: 55208)	GATAGTTCCATCTGCCC (SEQ ID NO: 55209)	GGTGTTGCTGTTAAGAGAAA (SEQ ID NO: 55210)
hCV31595	T	CAGAATGGCCTAAAGTCTGA (SEQ ID NO: 55211)	CAGAATGGCCTAAAGTCTGT (SEQ ID NO: 55212)	CAGGTGTTTGGGAATTTAAAG (SEQ ID NO: 55213)
hCV31785	G	CGGGAGTACCAGAAAGGGT (SEQ ID NO: 55214)	GGGAGTACCAGAAAGGGC (SEQ ID NO: 55215)	GCATTAGCACTGCACATTACATT (SEQ ID NO: 55216)
hCV31785	T	GAAGACATGGTTTCTCTGTTTC (SEQ ID NO: 55217)	GAAGACATGGTTTCTCTGTTTT (SEQ ID NO: 55218)	ACATTTGGCGGAAGTACTCT (SEQ ID NO: 55219)
hCV31884	C	GACGATTCTGGAATGGTTAC (SEQ ID NO: 55220)	GACGATTCTGGAATGGTTAT (SEQ ID NO: 55221)	CCTAGTCCCTAGACTCCTCTGTT (SEQ ID NO: 55222)
hCV32158	T	TGGTTCTAGGGAGGTAAC (SEQ ID NO: 55223)	ATGGTTCTAGGGAGGTAAT (SEQ ID NO: 55224)	TCCAAAGAGCAGTGTTCT (SEQ ID NO: 55225)
hCV32348	G	TGCAGTCCCCCATCCT (SEQ ID NO: 55226)	GCAGTCCCCCATCCC (SEQ ID NO: 55227)	TGGAATATGGAATACTCCTTTTATCTA (SEQ ID NO: 55228)
hCV32689	C	GGCGACTGGGTGACAG (SEQ ID NO: 55229)	GGCGACTGGGTGACAA (SEQ ID NO: 55230)	GGCTGCCAGGAACAAGT (SEQ ID NO: 55231)
hCV33715	G	CCCCGGCAAGGTTG (SEQ ID NO: 55232)	CCCCGGCAAGGTTC (SEQ ID NO: 55233)	TGCTGGGCTTCCATGTA (SEQ ID NO: 55234)
hCV36839	C	ATTGTTGAGTGTTGGCAATAC (SEQ ID NO: 55235)	ATTGTTGAGTGTTGGCAATAT (SEQ ID NO: 55236)	CAAACACAGCAATCAAGTGTATG (SEQ ID NO: 55237)
hCV36938	C	CAGGAAGCTTCGTGATTTG (SEQ ID NO: 55238)	CCAGGAAGCTTCGTGATTTA (SEQ ID NO: 55239)	TTGAGGACAATAATTTTCTTTACAC (SEQ ID NO: 55240)
hCV47267	C	CACCAAACGGGGTTACTAG (SEQ ID NO: 55241)	CACCAAACGGGGTTACTAC (SEQ ID NO: 55242)	CAGAGGGTTGATTTTCTTTCTAT (SEQ ID NO: 55243)
hCV52509	T	CGGCCTCGGTCTCG (SEQ ID NO: 55244)	CGGCCTCGGTCTCA (SEQ ID NO: 55245)	TGCATCTCGCTCAACAGAC (SEQ ID NO: 55246)
hCV5478	T	CGGCTTTCTGGTGGG (SEQ ID NO: 55247)	ACGGCTTTCTGGTGGA (SEQ ID NO: 55248)	GGCTCCGAGGACGAGA (SEQ ID NO: 55249)
hCV58970	G	CGGCTGATGTTGTTAAATAT (SEQ ID NO: 55250)	CGGCTGATGTTGTTAAATAC (SEQ ID NO: 55251)	GCACACTAGTTGACACCATACT (SEQ ID NO: 55252)
hCV74327	A	ATACCTACCAAGAACATAATCCTT (SEQ ID NO: 55253)	ATACCTACCAAGAACATAATCCTC (SEQ ID NO: 55254)	ACCAACAGCCTCTGAACAA (SEQ ID NO: 55255)
hCV75477	A	TTTTCGCTCAAGTTGTTGTAA (SEQ ID NO: 55256)	ATTTTCGCTCAAGTTGTTGTAC (SEQ ID NO: 55257)	AGGTGCCCTTGATAGTCTGTAT (SEQ ID NO: 55258)
hCV75823	G	GTGAAGGTCCGCTTCTTC (SEQ ID NO: 55259)	TGAAGGTCCGCTTCTTG (SEQ ID NO: 55260)	CCCACCAGAGCTGAAGATC (SEQ ID NO: 55261)
hCV75844	T	CCCCAGGCTGCACAC (SEQ ID NO: 55262)	CCCCAGGCTGCACAT (SEQ ID NO: 55263)	CATCACAACCAACCCTACTG (SEQ ID NO: 55264)

WO 2006/022639 PCT/US2004/024001

TABLE 5, page 4 of 4

hCV	Allele 1	Sequence A	Sequence B	Sequence C
hCV76112	T	CGCCCTCGCACAG (SEQ ID NO: 55265)	CGCCCTCGCACAA (SEQ ID NO: 55266)	AGGATCCCAAGGGAAATACT (SEQ ID NO: 55267)
hCV79952	G	CAGATTCTCATAACTAGCACCAT (SEQ ID NO: 55268)	CAGATTCTCATAACTAGCACCAC (SEQ ID NO: 55269)	TGAAAAGCTGGATGACATGA (SEQ ID NO: 55270)
hCV81132	A	CCTATACTCTAGTCCCAGAGACAA (SEQ ID NO: 55271)	CCTATACTCTAGTCCCAGAGACAC (SEQ ID NO: 55272)	TTCCACTTGGCATGAGTATAGA (SEQ ID NO: 55273)
hCV81610	C	GCAAACCTGGGACTCAG (SEQ ID NO: 55274)	GCAAACCTGGGACTCAA (SEQ ID NO: 55275)	GCTCCATACTTGCTACCTCTTACTA (SEQ ID NO: 55276)
hCV82276	C	AGTGGTCAATAAATTGATTCTAGAC (SEQ ID NO: 55277)	AGTGGTCAATAAATTGATTCTAGAT (SEQ ID NO: 55278)	GTTTTCCCAATTCAAACTTGA (SEQ ID NO: 55279)
hCV84882	C	TCAATCTCCGGAACCTG (SEQ ID NO: 55280)	GTTCAATCTCCGGAACCTA (SEQ ID NO: 55281)	CTGCTGGGTTAGAGAAGACTTAC (SEQ ID NO: 55282)
hCV85597	T	AAAACACCCTCGCTTAGC (SEQ ID NO: 55283)	CAAAACACCCTCGCTTAGT (SEQ ID NO: 55284)	AACTGCAGTGGAGAGTATATAAGGT (SEQ ID NO: 55285)
hCV87151	A	CACAATGTCTAGAGGTGGGTA (SEQ ID NO: 55286)	ACAATGTCTAGAGGTGGGTG (SEQ ID NO: 55287)	TGGACGTAGAAGAACAGTAAGTACA (SEQ ID NO: 55288)
hCV87251	G	GTTCATGTCGTCGAAGTCTT (SEQ ID NO: 55289)	GTTCATGTCGTCGAAGTCTC (SEQ ID NO: 55290)	GGAGCAAATGGCTCAATGTA (SEQ ID NO: 55291)
hCV87806	C	CACATTCCCTCTCTCCG (SEQ ID NO: 55292)	CCACATTCCCTCTCTCCA (SEQ ID NO: 55293)	TTCAGAGGCACACAACTATGA (SEQ ID NO: 55294)
hCV87826	T	CACCTCAGTTCAACAGTTTATATTTA (SEQ ID NO: 55295)	ACCTCAGTTCAACAGTTTATATTTT (SEQ ID NO: 55296)	AAAGAAAGAAAGTGTTAGGATGTCT (SEQ ID NO: 55297)
hCV88562	G	GGAAGGTAGCCCCTAAGAC (SEQ ID NO: 55298)	GGAAGGTAGCCCCTAAGAG (SEQ ID NO: 55299)	CGTGTTCGCTGTTCAGTG (SEQ ID NO: 55300)
hCV88852	A	CACACGTTGTCTCAAAATGAGTA (SEQ ID NO: 55301)	CACGTTGTCTCAAAATGAGTG (SEQ ID NO: 55302)	GATATGGGCAGGTTTCACTG (SEQ ID NO: 55303)
hCV89212	G	CTTCTCATTGCTTTTCTCCTAT (SEQ ID NO: 55304)	CTTCTCATTGCTTTTCTCCTAC (SEQ ID NO: 55305)	GCTGACCGAGAATATAATCATCT (SEQ ID NO: 55306)
hCV89212	G	TTCTCATTGCTTTTCTCCTAT (SEQ ID NO: 55307)	TTCTCATTGCTTTTCTCCTAC (SEQ ID NO: 55308)	GCTGACCGAGAATATAATCATCT (SEQ ID NO: 55309)
hCV89845	C	GGCTTTTAAAACTTCAAATGC (SEQ ID NO: 55310)	AAGGCTTTTAAAACTTCAAATGT (SEQ ID NO: 55311)	GCCACTGAGATTATCAAACTAACT (SEQ ID NO: 55312)
hCV95795	C	ATGACGCTGACCACAGG (SEQ ID NO: 55313)	CATGACGCTGACCACAGA (SEQ ID NO: 55314)	GCCGCTTACCCAAAGTAATT (SEQ ID NO: 55315)
hCV96054	G	CGCAGCTCCCACCA (SEQ ID NO: 55316)	CGCAGCTCCCACCG (SEQ ID NO: 55317)	CCTGCCCACGTTCTCTAT (SEQ ID NO: 55318)
hCV96321	G	AGACCAAGTGGAGACAGACA (SEQ ID NO: 55319)	GACCAAGTGGAGACAGACG (SEQ ID NO: 55320)	CCTTTTCTCCCTTCTTTTCTAAC (SEQ ID NO: 55321)
hCV97656	T	TGGTCAGGTGATAGATCTGC (SEQ ID NO: 55322)	TGTGGTCAGGTGATAGATCTGT (SEQ ID NO: 55323)	CCTCATTGGACAATGAGAGAC (SEQ ID NO: 55324)
hDV68530	C	GGCAGGGCTCTGATACAG (SEQ ID NO: 55325)	GGCAGGGCTCTGATACAA (SEQ ID NO: 55326)	GCGGTTCTTCTGGTGATTA (SEQ ID NO: 55327)
hDV68530	G	CAACGAGTGATTCTTTCACAC (SEQ ID NO: 55328)	CAACGAGTGATTCTTTCACAG (SEQ ID NO: 55329)	AATTCCCTTCACTCTTTCTTTTC (SEQ ID NO: 55330)
hDV68530	C	ACAGGCTGTGTCTAGTTCTCAC (SEQ ID NO: 55331)	ACAGGCTGTGTCTAGTTCTCAT (SEQ ID NO: 55332)	CTCAGAGGCCGGAAGTC (SEQ ID NO: 55333)
hDV68530	T	CGGGGAGGGAGACC (SEQ ID NO: 55334)	CCGGGGAGGGAGACT (SEQ ID NO: 55335)	TAAAAGCCATTTTCAGACACTAA (SEQ ID NO: 55336)
hDV68530	A	CGTGTGACATTTCTGAAATCAA (SEQ ID NO: 55337)	CGTGTGACATTTCTGAAATCAG (SEQ ID NO: 55338)	ACCATTCTCTCCTTATCACTTTATT (SEQ ID NO: 55339)
hDV68531	G	TTAAGAAACAGCCTTTCAACA (SEQ ID NO: 55340)	TAAGAAACAGCCTTTCAACG (SEQ ID NO: 55341)	GCAGCATTCCTGACAGAACTA (SEQ ID NO: 55342)

WO 2006/022639 PCT/US2004/024001

TABLE 6, page 1 of 6

Marker	Sample set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV11200217	1	age_ge75=1	NONE	0.01787	0.01951	0.03656	0.0911	1.39	1.06: 1.82	1.5	1.02: 2.20	1.58	0.93: 2.69	T	41.4	33.7	227	218
hCV11200217	3	age_ge75=1	NONE	0.11846	0.12574	0.03444	0.90777	1.23	0.95: 1.61	1.5	1.03: 2.18	1.03	0.62: 1.70	T	41.0	36.0	221	246
hCV11214795	1	apoe4=1	NONE	0.04774	0.04598	0.07839	0.13579	1.75	1.00: 3.05	1.71	0.94: 3.10			A	15.9	9.8	239	87
hCV11214795	2	apoe4=1	NONE	0.03857	0.03731	0.01631	0.74697	1.82	1.03: 3.23	2.17	1.14: 4.11	0.72	0.10: 5.23	A	16.7	9.9	132	96
hCV11396215	2	male=0	NONE	0.06459	0.04482	0.08886	0.12316	0.71	0.49: 1.02	0.68	0.44: 1.06	0.32	0.07: 1.46	G	21.7	28.1	122	244
hCV11396215	1+3	male=0	source	0.03301	0.03168	0.0858	0.05674	0.81	0.67: 0.98	0.81	0.64: 1.03	0.62	0.37: 1.02	G	23.4	27.3	559	533
hCV11566355	2	apoe4=1	male,age_ge75	0.05462	0.04901	0.01789	0.77701	0.61	0.37: 1.01	0.48	0.26: 0.88	1.23	0.27: 5.70	G	18.1	24.5	119	92
hCV11566355	1+3	apoe4=1	source,male,age_ge75	0.03231	0.03561	0.1035	0.04351	0.74	0.56: 0.98	0.75	0.53: 1.06	0.52	0.27: 1.01	G	21.4	27.1	459	181
hCV11566355	3	male=1	NONE	0.01572	0.01935	0.05005	0.05163	0.53	0.32: 0.89	0.54	0.29: 1.00	0.24	0.05: 1.13	G	17.3	28.2	78	101
hCV11566355	1+2	male=1	source	0.02664	0.02918	0.15309	0.00951	0.73	0.55: 0.96	0.78	0.55: 1.10	0.36	0.17: 0.80	G	21.2	27.1	257	282
hCV11595547	1	apoe4=0	male,age_ge75	0.04368	0.04634	0.04799	0.21518	0.75	0.56: 0.99	0.65	0.43: 1.00	0.73	0.44: 1.20	C	40.1	48.1	162	270
hCV11595547	2+3	apoe4=0	source,male,age_ge75	0.0115	0.01362	0.00886	0.17476	0.73	0.57: 0.93	0.63	0.44: 0.89	0.74	0.48: 1.14	C	39.4	45.8	199	559
hCV11597236	2	ALL	NONE	0.01998	0.01986	0.0562	0.04172	1.4	1.05: 1.86	1.39	0.99: 1.95	2.29	1.01: 5.19	T	23.9	18.3	241	355
hCV11597236	3	ALL	NONE	0.03178	0.03659	0.02195	0.63163	1.33	1.02: 1.73	1.43	1.05: 1.93	1.2	0.58: 2.48	T	19.5	15.4	384	399
hCV11597236	3	male=0	NONE	0.01124	0.01331	0.00886	0.43995	1.49	1.09: 2.03	1.61	1.13: 2.30	1.41	0.59: 3.41	T	20.0	14.4	282	295
hCV11597236	1+2	male=0	source	0.16377	0.16648	0.04545	0.36623	1.2	0.93: 1.55	1.36	1.00: 1.85	0.72	0.34: 1.52	T	21.0	17.9	336	435
hCV11840248	1	age_ge75=1	apoe4,male	0.05662	0.04824	0.01857	0.43982	1.32	0.99: 1.75	1.75	1.10: 2.79	1.21	0.74: 1.98	T	49.5	44.3	212	203
hCV11840248	3	age_ge75=1	apoe4,male	0.10798	0.09389	0.0497	0.48773	1.25	0.95: 1.63	1.55	0.99: 2.41	1.18	0.74: 1.89	T	49.8	45.9	221	246
hCV11841396	3	apoe4=1	NONE	0.02574	0.02598	0.07844	0.02886	0.62	0.41: 0.95	0.64	0.39: 1.05	0.26	0.07: 0.95	A	15.2	22.4	234	96
hCV11841396	1+2	apoe4=1	source	0.03978	0.04694	0.04915	0.31877	0.71	0.51: 0.99	0.67	0.46: 1.00	0.65	0.27: 1.54	A	16.4	21.6	375	160
hCV11861096	1	ALL	apoe4,male,age_ge75	0.00123	0.00174	0.00332	0.05935	1.76	1.24: 2.49	1.78	1.21: 2.63	2.64	0.88: 7.95	C	14.8	10.0	386	350
hCV11861096	2+3	ALL	source,apoe4,male,age_ge75	0.04405	0.04369	0.04073	0.56687	1.33	1.01: 1.75	1.37	1.01: 1.85	1.44	0.45: 4.56	C	12.8	11.3	529	746
hCV11861096	1	male=0	apoe4,age_ge75	0.02222	0.02324	0.06387	0.02973	1.66	1.07: 2.58	1.57	0.97: 2.55	13.1	0.97:176.56	C	14.6	10.5	250	215
hCV11861096	2+3	male=0	source,apoe4,age_ge75	0.01138	0.01039	0.00995	0.45452	1.55	1.11: 2.17	1.63	1.13: 2.36	1.8	0.41: 7.83	C	13.7	11.1	362	520
hCV1191260	1	ALL	apoe4,male,age_ge75	0.03485	0.03382	0.09486	0.07243	1.26	1.02: 1.55	1.34	0.95: 1.88	1.39	0.97: 2.00	G	51.3	44.9	418	375
hCV1191260	2	ALL	apoe4,male,age_ge75	0.04533	0.04221	0.02756	0.29859	1.33	1.01: 1.76	1.65	1.05: 2.59	1.29	0.80: 2.08	G	49.0	45.7	191	396
hCV12029086	1	ALL	NONE	0.00107	0.00134	0.00472	0.01774	0.69	0.56: 0.86	0.67	0.50: 0.88	0.54	0.32: 0.90	T	24.2	31.5	418	376
hCV12029086	2	ALL	NONE	0.2155	0.20584	0.73943	0.01952	0.86	0.68: 1.09	0.95	0.70: 1.29	0.48	0.26: 0.90	T	28.0	31.1	275	408
hCV1212623	1	age_ge75=1	NONE	0.09859	0.10056	0.31965	0.0307	1.3	0.95: 1.78	1.21	0.83: 1.77	2.6	1.06: 6.35	C	25.6	20.9	227	218
hCV1212623	2	age_ge75=1	NONE	0.05989	0.06079	0.20055	0.03059	1.48	0.98: 2.23	1.4	0.84: 2.35	2.8	1.06: 7.37	C	28.9	21.6	76	248
hCV1212623	2	ALL	NONE	0.00509	0.00513	0.01534	0.03412	1.46	1.12: 1.90	1.5	1.08: 2.07	2.08	1.04: 4.15	C	27.9	21.0	247	369
hCV1212623	1+3	ALL	source	0.33223	0.33643	0.79856	0.04993	1.09	0.92: 1.28	1.03	0.84: 1.26	1.55	1.00: 2.40	C	23.8	22.4	802	775
hCV1212623	1	male=1	apoe4,age_ge75	0.5698	0.57217	0.78969	0.02666	1.13	0.74: 1.74	0.93	0.56: 1.55	4.77	1.01: 22.58	C	21.6	19.3	153	140
hCV1212623	2+3	male=1	source,apoe4,age_ge75	0.00543	0.00739	0.05348	0.00474	1.66	1.16: 2.38	1.55	0.99: 2.43	3.54	1.45: 8.66	C	29.5	21.1	168	228
hCV1212684	1	age_ge75=0	NONE	0.04437	0.04212	0.29962	0.01367	1.39	1.01: 1.91	1.28	0.80: 2.02	2.23	1.17: 4.26	T	44.0	36.2	176	145
hCV1212684	2+3	age_ge75=0	source	0.01819	0.02146	0.05864	0.05682	1.36	1.05: 1.74	1.42	0.99: 2.05	1.56	0.99: 2.48	T	44.5	37.2	245	258
hCV1229667	1	apoe4=1	male,age_ge75	0.01063	0.01004	0.00587	0.5827	0.43	0.23: 0.83	0.39	0.20: 0.77			C	5.0	10.5	239	86
hCV1229667	2+3	apoe4=1	source,male,age_ge75	0.04767	0.05058	0.019	0.26075	0.58	0.34: 0.99	0.5	0.29: 0.89			C	5.4	8.6	335	180
hCV1229667	1	male=1	apoe4,age_ge75	0.03556	0.03547	0.02286	0.5827	0.41	0.18: 0.94	0.37	0.16: 0.87			C	3.9	7.6	153	139
hCV1229667	3	male=1	apoe4,age_ge75	0.04405	0.03729	0.03729		0.39	0.15: 1.03	0.37	0.14: 1.00			C	4.5	8.6	78	99
hCV1229682	1	apoe4=1	male,age_ge75	0.0061	0.00465	0.00465		0.39	0.19: 0.77	0.36	0.17: 0.74			A	4.6	10.5	227	81
hCV1229682	3	apoe4=1	male,age_ge75	0.04361	0.05127	0.14203	0.01055	0.47	0.22: 0.98	0.55	0.25: 1.22	0		A	4.1	7.3	234	96
hCV1229777	1	male=1	apoe4,age_ge75	0.05847	0.0635	0.03184	0.6469	1.47	0.99: 2.17	1.74	1.05: 2.87	1.24	0.51: 3.01	A	29.4	25.2	153	139
hCV1229777	3	male=1	apoe4,age_ge75	0.04272	0.04947	0.02225	0.5904	1.62	1.00: 2.62	2.14	1.11: 4.11	1.27	0.49: 3.28	A	33.3	26.7	78	101
hCV1244849	1	age_ge75=1	apoe4,male	0.0471	0.0474	0.02372	0.97131	1.5	1.01: 2.22	1.68	1.07: 2.64	0.98	0.27: 3.59	C	17.4	13.3	210	206
hCV1244849	3	age_ge75=1	apoe4,male	0.01156	0.0113	0.0056	0.73422	1.6	1.11: 2.29	1.79	1.19: 2.72	1.24	0.37: 4.14	C	20.0	14.4	222	246
hCV1322419	3	male=1	apoe4,age_ge75	0.01627	0.01732	0.05447	0.04811	1.68	1.10: 2.57	2.13	0.99: 4.57	1.95	1.00: 3.80	A	59.4	47.1	90	102
hCV1322419	1+2	male=1	source,apoe4,age_ge75	0.0106	0.01174	0.0206	0.07163	1.41	1.08: 1.83	1.64	1.08: 2.51	1.5	0.97: 2.32	A	51.3	45.7	269	268
hCV1413258	1	apoe4=1	male,age_ge75	0.01691	0.01813	0.00986	0.49333	0.61	0.41: 0.92	0.51	0.30: 0.85	0.7	0.25: 1.95	A	21.5	30.2	226	81
hCV1413258	3	apoe4=1	male,age_ge75	0.21242	0.20787	0.04843	0.31005	0.78	0.52: 1.16	0.61	0.37: 1.00	1.87	0.53: 6.57	A	22.0	26.3	218	95
hCV1419932	1	age_ge75=0	apoe4,male	0.0112	0.01334	0.0122	0.4965	3.23	1.33: 7.84	3.37	1.32: 8.56			G	6.5	3.7	176	147

WO 2006/022639 PCT/US2004/024001

TABLE 6, page 2 of 6

Marker	Sample set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV1419932	2+3	age_ge75=0	source,apoe4,male	0.03275	0.03365	0.02488	0.93432	1.95	1.04: 3.65	2.07	1.07: 3.98	0.79	0.01: 61.32	G	7.4	4.1	244	257
hCV1419932	1	ALL	apoe4,male,age_ge75	0.00558	0.00572	0.00677	0.31365	2.02	1.22: 3.32	2.04	1.21: 3.42			G	7.0	4.3	387	352
hCV1419932	2+3	ALL	source,apoe4,male,age_ge75	0.04602	0.05048	0.04077	0.81821	1.49	1.01: 2.20	1.54	1.02: 2.31	1.24	0.21: 7.38	G	6.6	5.0	527	743
hCV1419932	3	apoe4=1	male,age_ge75	0.04844	0.04974	0.05469	0.45939	2.54	0.97: 6.67	2.55	0.95: 6.84			G	6.8	2.6	219	95
hCV1419932	1+2	apoe4=1	source,male,age_ge75	0.03084	0.04417	0.02238	0.88519	2.22	1.09: 4.54	2.47	1.14: 5.38	1.2	0.14: 10.47	G	6.8	3.8	333	158
hCV1419932	2	male=0	apoe4,age_ge75	0.02702	0.02448	0.01952	0.77816	2.44	1.05: 5.67	2.7	1.11: 6.58	0		G	8.8	4.9	80	224
hCV1419932	1+3	male=0	source,apoe4,age_ge75	0.01693	0.01753	0.01728	0.55359	1.7	1.10: 2.62	1.73	1.10: 2.71	3.17	0.12: 84.72	G	6.4	4.3	531	511
hCV1507426	2	age_ge75=1	NONE	0.03209	0.02476	0.04709	0.11127	0.58	0.35: 0.96	0.57	0.32: 1.00			G	13.6	21.5	77	249
hCV1507426	1+3	age_ge75=1	source	0.1043	0.09487	0.04002	0.74122	0.83	0.66: 1.04	0.76	0.58: 0.99	1.12	0.56: 2.26	G	19.6	22.8	448	464
hCV1558518	1	age_ge75=0	apoe4,male	0.00195	0.00316	0.00146	0.11326	0.56	0.38: 0.81	0.41	0.23: 0.72	0.58	0.30: 1.13	G	36.4	47.6	176	147
hCV1558518	2+3	age_ge75=0	source,apoe4,male	0.03783	0.03144	0.80296	0.0005	0.74	0.55: 0.98	0.94	0.60: 1.48	0.36	0.20: 0.64	G	39.6	48.3	245	258
hCV15870743	1	apoe4=0	NONE	0.00234	0.00297	0.00316	0.09552	0.62	0.46: 0.84	0.55	0.37: 0.82	0.56	0.28: 1.12	T	24.7	34.6	162	272
hCV15870743	2+3	apoe4=0	source	0.06426	0.06633	0.04312	0.53516	0.8	0.63: 1.01	0.73	0.55: 0.99	0.84	0.49: 1.45	T	25.6	30.4	248	606
hCV15887512	1	age_ge75=0	apoe4,male	0.5706	0.57672	0.84215	0.03684	1.15	0.72: 1.85	0.94	0.55: 1.63	7.59	1.10: 52.43	A	19.7	17.8	178	146
hCV15887512	3	age_ge75=0	apoe4,male	0.50643	0.51427	0.9785	0.02628	1.19	0.71: 1.98	0.99	0.56: 1.77	12.5	1.00:155.39	A	15.1	13.4	162	153
hCV15919456	3	age_ge75=0	apoe4,male	0.04461	0.04202	0.12163	0.07863	1.45	1.01: 2.08	1.57	0.87: 2.82	1.71	0.94: 3.13	A	52.2	47.4	162	153
hCV15919456	1+2	age_ge75=0	source,apoe4,male	0.01767	0.01772	0.02401	0.10461	1.36	1.05: 1.75	1.59	1.06: 2.40	1.42	0.93: 2.17	A	52.8	46.8	353	285
hCV15919456	2	apoe4=1	male,age_ge75	0.00095	0.00111	0.00224	0.0216	2.1	1.35: 3.26	3.04	1.46: 6.35	2.35	1.14: 4.83	A	57.7	45.9	117	86
hCV15919456	1+3	apoe4=1	source,male,age_ge75	0.1118	0.11352	0.03504	0.63862	1.22	0.95: 1.56	1.52	1.03: 2.23	1.1	0.74: 1.64	A	50.7	45.6	459	181
hCV15965240	1	ALL	apoe4,male,age_ge75	0.2343	0.23343	0.02872	0.76594	1.14	0.92: 1.43	1.49	1.04: 2.13	0.94	0.65: 1.37	G	49.5	46.9	390	352
hCV15965240	3	ALL	apoe4,male,age_ge75	0.14962	0.14311	0.02503	0.98192	1.18	0.94: 1.47	1.5	1.05: 2.14	1	0.68: 1.48	G	48.6	44.9	359	395
hCV16111152	1	apoe4=0	male,age_ge75	0.04186	0.03962	0.06465	0.15809	1.44	1.01: 2.06	1.47	0.97: 2.21	2.4	0.67: 8.58	C	18.8	13.9	178	287
hCV16111152	2	apoe4=0	male,age_ge75	0.02527	0.02544	0.01328	0.79966	1.79	1.07: 3.00	2.18	1.17: 4.05	1.28	0.22: 7.37	C	23.7	17.6	59	273
hCV16289132	1	age_ge75=1	NONE	0.05208	0.05788	0.34336	0.01353	1.33	1.00: 1.76	1.21	0.82: 1.78	2.08	1.15: 3.76	C	38.2	31.8	212	206
hCV16289132	2+3	age_ge75=1	source	0.00882	0.0086	0.07092	0.00634	1.36	1.08: 1.71	1.32	0.98: 1.80	2	1.22: 3.27	C	35.8	31.9	286	494
hCV16289132	1	ALL	NONE	0.01265	0.01696	0.06148	0.0317	1.31	1.06: 1.63	1.32	0.99: 1.77	1.58	1.04: 2.40	C	38.3	32.1	389	352
hCV16289132	2+3	ALL	source	0.17209	0.1784	0.66038	0.03093	1.12	0.95: 1.33	1.05	0.84: 1.31	1.47	1.04: 2.08	C	33.8	32.4	547	756
hCV16289132	3	male=0	apoe4,age_ge75	0.0914	0.08258	0.27686	0.03958	1.27	0.96: 1.67	1.22	0.85: 1.75	2.2	1.06: 4.56	C	31.7	27.6	282	295
hCV16289132	1+2	male=0	source,apoe4,age_ge75	0.02878	0.03565	0.13057	0.0395	1.31	1.03: 1.68	1.29	0.92: 1.80	1.64	1.03: 2.63	C	39.3	34.9	331	446
hCV1665140	3	age_ge75=1	apoe4,male	0.58527	0.59007	0.04038	0.18962	1.08	0.82: 1.41	1.55	1.02: 2.37	0.74	0.47: 1.17	C	47.7	45.1	222	246
hCV1665140	1+2	age_ge75=1	source,apoe4,male	0.02654	0.02936	0.01354	0.28865	1.29	1.03: 1.61	1.54	1.09: 2.19	1.23	0.84: 1.79	C	50.7	44.4	301	486
hCV1665140	3	ALL	NONE	0.37159	0.36916	0.01051	0.22525	1.1	0.90: 1.34	1.53	1.10: 2.12	0.81	0.57: 1.14	C	48.9	46.6	360	396
hCV1665140	1+2	ALL	source	0.05726	0.05881	0.0455	0.28048	1.16	1.00: 1.35	1.28	1.01: 1.64	1.15	0.89: 1.47	C	50.9	47.4	627	777
hCV1665140	1	apoe4=1	male,age_ge75	0.01172	0.01353	0.0123	0.11826	1.57	1.10: 2.24	1.98	1.16: 3.39	1.62	0.88: 2.97	C	51.7	40.7	240	86
hCV1665140	2+3	apoe4=1	source,male,age_ge75	0.61265	0.61109	0.03077	0.18548	1.07	0.82: 1.38	1.59	1.04: 2.41	0.76	0.51: 1.15	C	51.3	47.9	343	191
hCV1665140	1	male=0	apoe4,age_ge75	0.27213	0.27118	0.04834	0.81447	1.16	0.89: 1.53	1.54	1.00: 2.37	0.95	0.60: 1.50	C	49.1	46.4	264	234
hCV1665140	2+3	male=0	source,apoe4,age_ge75	0.21357	0.21441	0.00624	0.48379	1.14	0.93: 1.41	1.61	1.14: 2.28	0.88	0.62: 1.25	C	52.0	48.4	374	548
hCV1665253	1	ALL	apoe4,male,age_ge75	0.03464	0.03372	0.04058	0.25936	0.76	0.59: 0.98	0.72	0.53: 0.99	0.67	0.33: 1.36	T	20.4	25.3	417	375
hCV1665253	2	ALL	apoe4,male,age_ge75	0.0258	0.02641	0.02276	0.34636	0.7	0.52: 0.96	0.65	0.44: 0.94	0.69	0.31: 1.51	T	20.8	26.8	250	399
hCV1665253	1	apoe4=0	male,age_ge75	0.01393	0.01425	0.02172	0.13829	0.66	0.48: 0.92	0.63	0.43: 0.93	0.52	0.21: 1.29	T	18.6	26.4	177	288
hCV1665253	2	apoe4=0	male,age_ge75	0.01119	0.01232	0.0232	0.09029	0.57	0.37: 0.88	0.55	0.33: 0.92	0.39	0.12: 1.28	T	18.3	27.8	93	302
hCV1665253	1	male=1	apoe4,age_ge75	0.03446	0.03229	0.01484	0.78366	0.64	0.43: 0.97	0.54	0.33: 0.89	0.84	0.26: 2.74	T	19.8	27.1	154	140
hCV1665253	2	male=1	apoe4,age_ge75	0.02516	0.02703	0.0477	0.12263	0.61	0.40: 0.94	0.59	0.35: 0.99	0.4	0.13: 1.27	T	20.7	31.5	138	146
hCV1791780	2	apoe4=0	NONE	0.51927	0.49736	0.90966	0.00895	1.14	0.76: 1.71	0.97	0.60: 1.58	4.8	1.32: 17.41	G	22.5	20.3	91	276
hCV1791780	3	apoe4=0	NONE	0.03172	0.02663	0.07792	0.04618	1.41	1.03: 1.92	1.42	0.96: 2.11	2.31	0.99: 5.36	G	29.3	22.8	150	303
hCV1792842	2	apoe4=0	NONE	0.49889	0.47693	0.93946	0.00915	1.15	0.77: 1.72	0.98	0.60: 1.60	4.78	1.32: 17.35	C	22.5	20.2	91	275
hCV1792842	3	apoe4=0	NONE	0.03172	0.02663	0.07792	0.04618	1.41	1.03: 1.92	1.42	0.96: 2.11	2.31	0.99: 5.36	C	29.3	22.8	150	303
hCV1792848	3	male=1	NONE	0.16787	0.15475	0.53822	0.02859	1.38	0.87: 2.16	1.2	0.67: 2.18	3.57	1.07: 11.84	T	34.0	27.2	78	101
hCV1792848	1+2	male=1	source	0.09947	0.08782	0.43623	0.01044	1.26	0.96: 1.66	1.15	0.81: 1.65	2.46	1.21: 5.01	T	32.8	27.5	229	260
hCV1792856	1	apoe4=1	NONE	0.03186	0.02924	0.04099	0.21475	0.62	0.40: 0.96	0.58	0.34: 0.98	0.44	0.11: 1.67	G	16.4	24.1	225	81
hCV1792856	3	apoe4=1	NONE	0.03103	0.03278	0.07844	0.05868	0.63	0.41: 0.96	0.64	0.39: 1.05	0.33	0.10: 1.10	G	15.4	22.4	234	96

WO 2006/022639 PCT/US2004/024001

TABLE 6, page 3 of 6

Marker	Sample set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV1824909	1	ALL	apoe4,male,age_ge75	0.19091	0.19667	0.03041	0.51764	0.86	0.68: 1.08	0.7	0.51: 0.97	1.17	0.72: 1.91	C	31.4	34.4	390	350
hCV1824909	2+3	ALL	source,apoe4,male,age_ge75	0.00181	0.00211	0.01442	0.00784	0.74	0.61: 0.89	0.73	0.57: 0.94	0.58	0.39: 0.86	C	31.7	37.5	526	742
hCV1824909	1	apoe4=1	male,age_ge75	0.03211	0.02906	0.02321	0.3263	0.66	0.45: 0.96	0.54	0.31: 0.92	0.66	0.29: 1.51	C	29.5	39.4	227	80
hCV1824909	2+3	apoe4=1	source,male,age_ge75	0.00796	0.00796	0.00707	0.15844	0.68	0.51: 0.90	0.58	0.39: 0.86	0.66	0.37: 1.17	C	32.2	41.6	325	172
hCV1824909	1	male=0	apoe4,age_ge75	0.26421	0.25986	0.041	0.35324	0.84	0.62: 1.14	0.65	0.44: 0.98	1.37	0.70: 2.70	C	30.4	33.6	250	216
hCV1824909	2+3	male=0	source,apoe4,age_ge75	0.00252	0.00339	0.00817	0.03473	0.7	0.56: 0.88	0.66	0.48: 0.90	0.6	0.37: 0.96	C	31.0	37.4	360	517
hCV1841875	1	apoe4=1	male,age_ge75	0.03102	0.03013	0.0322	0.34162	0.63	0.41: 0.96	0.58	0.35: 0.96	0.54	0.15: 1.96	A	15.8	23.6	240	87
hCV1841875	2+3	apoe4=1	source,male,age_ge75	0.00187	0.00149	0.00174	0.20065	0.57	0.40: 0.81	0.52	0.35: 0.79	0.4	0.10: 1.51	A	14.4	21.6	329	176
hCV1873996	1	apoe4=0	male,age_ge75	0.69193	0.69995	0.72005	0.02805	1.08	0.73: 1.60	0.92	0.59: 1.44	4.01	1.11: 14.53	A	14.3	12.8	178	288
hCV1873996	2+3	apoe4=0	source,male,age_ge75	0.44511	0.45785	0.97161	0.01535	1.15	0.81: 1.64	1.01	0.68: 1.50	3.87	1.23: 12.13	A	12.7	11.3	204	598
hCV1911256	1	ALL	apoe4,male,age_ge75	0.00875	0.0093	0.02001	0.06841	0.72	0.56: 0.92	0.7	0.51: 0.95	0.58	0.31: 1.06	T	22.5	26.8	414	373
hCV1911256	2	ALL	apoe4,male,age_ge75	0.05082	0.05921	0.04337	0.45252	0.71	0.51: 1.00	0.65	0.43: 0.99	0.73	0.33: 1.63	T	22.2	26.4	185	388
hCV1920609	1	age_ge75=1	NONE	0.07824	0.07682	0.03378	0.45855	1.27	0.97: 1.65	1.59	1.03: 2.44	1.18	0.76: 1.83	A	51.8	45.9	227	217
hCV1920609	2+3	age_ge75=1	source	0.06587	0.05746	0.03108	0.37966	1.22	0.99: 1.51	1.47	1.03: 2.09	1.17	0.82: 1.68	A	50.9	46.9	293	503
hCV1920609	1	ALL	NONE	0.03777	0.03595	0.10875	0.07056	1.23	1.01: 1.50	1.31	0.94: 1.81	1.35	0.97: 1.86	A	53.2	48.0	419	376
hCV1920609	2+3	ALL	source	0.05288	0.04738	0.05624	0.1968	1.17	1.00: 1.36	1.28	0.99: 1.65	1.19	0.91: 1.54	A	50.6	47.1	560	772
hCV199172	1	apoe4=1	male,age_ge75	0.02832	0.03477	0.02744	0.4443	1.86	1.06: 3.25	2.03	1.08: 3.81	1.83	0.38: 8.81	A	17.8	11.1	227	81
hCV199172	2	apoe4=1	male,age_ge75	0.06086	0.04997	0.07921	0.15775	1.88	0.98: 3.62	1.9	0.94: 3.86			A	18.2	11.1	110	81
hCV2144148	1	age_ge75=0	NONE	0.01113	0.01236	0.03874	0.02532	0.59	0.39: 0.89	0.61	0.38: 0.98	0.2	0.04: 0.95	C	12.5	19.5	192	159
hCV2144148	3	age_ge75=0	NONE	0.04399	0.04136	0.01381	0.62083	0.63	0.40: 0.99	0.53	0.32: 0.88	1.46	0.32: 6.66	C	13.4	19.7	138	150
hCV2170733	1	male=0	apoe4,age_ge75	0.02417	0.02827	0.02807	0.45693	2.03	1.08: 3.78	2.06	1.07: 3.97	3.36	0.18: 64.45	C	7.6	4.0	264	235
hCV2170733	2	male=0	apoe4,age_ge75	0.00074	0.00061	0.00208	0.00053	3.91	1.71: 8.94	3.91	1.60: 9.57			C	10.1	5.0	89	238
hCV2539346	3	age_ge75=0	apoe4,male	0.00029	0.0003	0.0006	0.02122	0.47	0.31: 0.70	0.38	0.22: 0.67	0.32	0.13: 0.82	T	27.7	45.0	137	150
hCV2539346	1+2	age_ge75=0	source,apoe4,male	0.01327	0.01091	0.09506	0.00842	0.7	0.52: 0.93	0.7	0.46: 1.06	0.48	0.27: 0.85	T	35.2	40.9	281	253
hCV25596081	1	male=1	apoe4,age_ge75	0.04988	0.04739	0.04739		2.92	0.91: 9.39	2.97	0.92: 9.61			T	3.6	1.4	154	140
hCV25596081	2+3	male=1	source,apoe4,age_ge75	0.03199	0.02901	0.02901		2.56	1.09: 6.05	2.65	1.10: 6.35			T	5.3	2.1	171	237
hCV25602413	1	age_ge75=0	apoe4,male	0.05507	0.05588	0.23901	0.0159	1.46	0.99: 2.17	1.34	0.82: 2.18	3.97	1.21: 13.06	A	27.5	22.5	191	158
hCV25602413	2+3	age_ge75=0	source,apoe4,male	0.12785	0.12028	0.354	0.03456	1.29	0.93: 1.80	1.21	0.81: 1.82	2.66	1.06: 6.67	A	24.7	22.2	251	273
hCV25602413	1	ALL	apoe4,male,age_ge75	0.01717	0.01753	0.10483	0.00747	1.35	1.05: 1.72	1.29	0.95: 1.75	2.42	1.24: 4.70	A	27.0	22.7	418	374
hCV25602413	2+3	ALL	source,apoe4,male,age_ge75	0.08936	0.0867	0.31005	0.01959	1.18	0.97: 1.44	1.13	0.89: 1.43	1.92	1.12: 3.29	A	24.5	23.1	634	796
hCV25603905	1	apoe4=0	male,age_ge75	0.00229	0.00475	0.02762	0.00955	1.56	1.17: 2.07	1.58	1.05: 2.38	1.87	1.15: 3.05	C	45.0	34.3	161	270
hCV25603905	2+3	apoe4=0	source,male,age_ge75	0.33631	0.33	0.81503	0.04772	1.12	0.89: 1.42	0.96	0.67: 1.38	1.5	1.00: 2.25	C	47.0	42.8	200	572
hCV25606645	1	apoe4=1	NONE	0.01606	0.0227	0.04067	0.09661	0.57	0.36: 0.91	0.57	0.33: 0.98	0.4	0.13: 1.22	T	14.1	22.2	227	81
hCV25606645	3	apoe4=1	NONE	0.04485	0.05674	0.12293	0.0742	0.63	0.40: 0.99	0.66	0.39: 1.12	0.35	0.10: 1.16	T	12.8	18.9	219	95
hCV25625639	1	apoe4=0	NONE	0.03831	0.03792	0.01501	0.63551	0.73	0.54: 0.98	0.62	0.42: 0.91	0.85	0.43: 1.67	A	26.7	33.4	163	271
hCV25625639	2	apoe4=0	NONE	0.03854	0.03664	0.17389	0.01421	0.61	0.38: 0.98	0.68	0.40: 1.18			A	19.7	28.6	66	271
hCV25625639	1	male=1	apoe4,age_ge75	0.00037	0.00035	0.00012	0.19228	0.48	0.32: 0.72	0.36	0.21: 0.61	0.53	0.20: 1.37	A	22.5	35.9	140	135
hCV25625639	2+3	male=1	source,apoe4,age_ge75	0.03542	0.03212	0.0506	0.16505	0.68	0.48: 0.98	0.64	0.41: 1.00	0.45	0.15: 1.30	A	23.4	29.6	167	226
hCV25636732	2	apoe4=1	NONE	0.01865	0.02175	0.14219	0.0126	0.61	0.40: 0.92	0.65	0.36: 1.16	0.37	0.16: 0.82	G	31.8	43.3	118	82
hCV25636732	3	apoe4=1	NONE	0.31529	0.30405	0.80071	0.01618	0.84	0.59: 1.19	1.07	0.65: 1.75	0.45	0.23: 0.87	G	36.3	40.5	219	95
hCV25636732	2	male=0	apoe4,age_ge75	0.01011	0.01105	0.07719	0.01048	0.55	0.35: 0.87	0.58	0.32: 1.06	0.19	0.05: 0.73	G	26.3	39.8	80	230
hCV25636732	3	male=0	apoe4,age_ge75	0.62704	0.62629	0.4583	0.02779	0.94	0.72: 1.22	1.15	0.80: 1.65	0.52	0.29: 0.94	G	33.5	33.7	281	294
hCV25970515	1	age_ge75=1	apoe4,male	0.02682	0.03136	0.05852	0.10623	1.49	1.04: 2.13	1.5	0.98: 2.28	2.13	0.80: 5.64	T	22.4	16.2	210	204
hCV25970515	2+3	age_ge75=1	source,apoe4,male	0.02472	0.02861	0.01337	0.79582	1.42	1.05: 1.93	1.56	1.10: 2.22	1.14	0.45: 2.88	T	18.9	14.3	283	492
hCV2655167	1	ALL	NONE	0.00414	0.00458	0.0209	0.01498	1.39	1.11: 1.74	1.39	1.05: 1.84	1.99	1.13: 3.51	G	29.4	23.1	418	377
hCV2655167	2	ALL	NONE	0.0531	0.05345	0.01774	0.98329	1.28	1.00: 1.64	1.45	1.07: 1.98	1.01	0.53: 1.90	G	27.5	22.9	275	407
hCV2655167	1	apoe4=0	NONE	0.00032	0.00047	0.01604	0.00011	1.73	1.28: 2.33	1.59	1.09: 2.32	3.95	1.88: 8.28	G	31.9	21.4	177	288
hCV2655167	2	apoe4=0	NONE	0.1053	0.11197	0.03437	0.77083	1.4	0.93: 2.10	1.73	1.04: 2.89	0.85	0.28: 2.57	G	28.4	22.1	74	301
hCV2682758	1	age_ge75=1	apoe4,male	0.16039	0.16029	0.03641	0.77879	0.81	0.60: 1.09	0.65	0.44: 0.97	1.09	0.60: 1.98	T	32.3	37.2	212	207
hCV2682758	3	age_ge75=1	apoe4,male	0.04328	0.03996	0.03198	0.32844	0.75	0.56: 0.99	0.66	0.45: 0.97	0.75	0.42: 1.35	T	32.5	38.0	220	245
hCV2685860	1	apoe4=0	male,age_ge75	0.01901	0.022	0.04667	0.05753	1.74	1.09: 2.79	1.67	1.01: 2.79	5.53	0.68: 44.70	A	11.2	6.8	178	288

TABLE 6, page 4 of 6

Marker	Sample set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV2685860	3	apoe4=0	male,age_ge75	0.03561	0.02852	0.02852		1.69	1.03: 2.78	1.78	1.06: 3.00			A	10.3	6.3	150	303
hCV2734178	1	age_ge75=0	apoe4,male	0.04048	0.03635	0.22904	0.02856	1.43	1.01: 2.03	1.43	0.80: 2.55	1.79	1.03: 3.12	G	52.9	46.2	188	158
hCV2734178	2+3	age_ge75=0	source,apoe4,male	0.1012	0.10581	0.54141	0.03622	1.24	0.96: 1.59	1.14	0.76: 1.70	1.55	1.02: 2.37	G	49.4	45.0	331	290
hCV2734178	1	ALL	NONE	0.11122	0.11686	0.58041	0.04652	1.17	0.96: 1.43	1.1	0.79: 1.52	1.38	1.00: 1.89	G	53.6	49.6	416	376
hCV2734178	2+3	ALL	source	0.17761	0.17805	0.78398	0.04741	1.11	0.96: 1.28	1.03	0.82: 1.30	1.28	1.00: 1.64	G	48.8	46.5	659	806
hCV2757616	1	ALL	apoe4,male,age_ge75	0.01489	0.01664	0.02481	0.14616	1.39	1.06: 1.81	1.43	1.04: 1.95	1.7	0.81: 3.58	G	22.4	16.5	418	375
hCV2757616	2	ALL	apoe4,male,age_ge75	0.05283	0.05457	0.02257	0.8909	1.43	0.99: 2.06	1.65	1.07: 2.53	0.93	0.32: 2.75	G	20.7	15.8	181	380
hCV2757616	1	apoe4=0	male,age_ge75	0.00174	0.00254	0.00279	0.13011	1.72	1.22: 2.43	1.85	1.23: 2.77	1.93	0.80: 4.70	G	23.0	14.9	178	288
hCV2757616	2+3	apoe4=0	source,male,age_ge75	0.00296	0.00284	0.00423	0.12593	1.57	1.16: 2.12	1.66	1.17: 2.35	2.15	0.80: 5.79	G	20.7	15.0	205	590
hCV286937	1	ALL	apoe4,male,age_ge75	0.02842	0.03091	0.02459	0.47352	0.71	0.53: 0.96	0.67	0.47: 0.95	0.71	0.29: 1.77	G	13.9	20.1	389	350
hCV286937	2+3	ALL	source,apoe4,male,age_ge75	0.02821	0.02784	0.05702	0.08088	0.75	0.59: 0.97	0.76	0.57: 1.01	0.41	0.15: 1.14	G	12.7	16.7	526	739
hCV286937	1	male=1	apoe4,age_ge75	0.00079	0.0009	0.00042	0.44728	0.4	0.23: 0.68	0.34	0.18: 0.62	0.54	0.10: 2.83	G	10.0	20.1	140	134
hCV286937	2+3	male=1	source,apoe4,age_ge75	0.00062	0.00063	0.00154	0.03681	0.46	0.29: 0.72	0.44	0.27: 0.74	0.06	0.00: 1.15	G	11.8	20.6	165	223
hCV2875671	2	ALL	NONE	0.27614	0.27371	0.64956	0.0463	0.85	0.64: 1.13	0.93	0.66: 1.29	0.4	0.16: 1.01	G	20.0	22.6	245	358
hCV2875671	1+3	ALL	source	0.0111	0.0126	0.05572	0.01331	0.8	0.67: 0.95	0.82	0.67: 1.00	0.56	0.35: 0.89	G	19.1	22.7	803	776
hCV2875671	3	apoe4=0	NONE	0.02749	0.03649	0.03705	0.26549	0.66	0.46: 0.96	0.63	0.41: 0.97	0.59	0.23: 1.50	G	16.3	22.8	141	301
hCV2875671	1+2	apoe4=0	source	0.03863	0.04484	0.08435	0.11331	0.75	0.57: 0.99	0.75	0.55: 1.04	0.55	0.26: 1.17	G	18.2	23.1	244	558
hCV2875671	3	male=0	NONE	0.1101	0.11158	0.31754	0.03876	0.8	0.61: 1.05	0.84	0.61: 1.18	0.44	0.20: 0.98	G	19.9	23.7	294	297
hCV2875671	1+2	male=0	source	0.07553	0.07808	0.22225	0.04312	0.8	0.63: 1.02	0.84	0.63: 1.11	0.49	0.25: 0.98	G	19.9	23.7	377	466
hCV2950452	1	age_ge75=1	NONE	0.03882	0.04667	0.0063	0.922	0.74	0.56: 0.99	0.58	0.39: 0.86	0.97	0.56: 1.69	A	31.4	38.2	212	207
hCV2950452	3	age_ge75=1	NONE	0.04709	0.05163	0.03405	0.42308	0.76	0.57: 1.00	0.67	0.47: 0.97	0.79	0.44: 1.41	A	28.8	34.9	222	245
hCV2950452	1	apoe4=0	male,age_ge75	0.08524	0.09445	0.04098	0.6852	0.77	0.57: 1.04	0.66	0.45: 0.98	0.89	0.49: 1.60	A	30.7	36.0	163	272
hCV2950452	3	apoe4=0	male,age_ge75	0.04211	0.03847	0.04505	0.2332	0.73	0.54: 0.99	0.67	0.45: 0.99	0.67	0.34: 1.30	A	30.0	36.3	150	302
hCV299325	1	ALL	NONE	0.02899	0.02787	0.03925	0.17922	1.62	1.05: 2.52	1.61	1.02: 2.53			T	6.9	4.4	419	377
hCV299325	2+3	ALL	source	0.05412	0.05997	0.03719	0.83573	1.37	0.99: 1.90	1.44	1.02: 2.03	0.86	0.21: 3.54	T	6.7	5.1	566	789
hCV299325	1	male=0	apoe4,age_ge75	0.04895	0.04503	0.04729	0.60505	1.77	1.00: 3.16	1.83	1.00: 3.33			T	7.0	4.9	264	235
hCV299325	2+3	male=0	source,apoe4,age_ge75	0.04607	0.0468	0.0392	0.94225	1.62	1.00: 2.62	1.68	1.02: 2.76	1.16	0.04: 34.66	T	6.5	4.4	372	542
hCV3039499	1	male=0	apoe4,age_ge75	0.15215	0.14695	0.0273	0.88903	1.23	0.93: 1.64	1.6	1.05: 2.45	0.96	0.56: 1.66	A	42.8	39.0	251	219
hCV3039499	2	male=0	apoe4,age_ge75	0.26232	0.23079	0.02013	0.31435	1.27	0.83: 1.94	2.23	1.14: 4.40	0.62	0.24: 1.61	A	42.6	37.8	81	229
hCV3046185	1	ALL	apoe4,male,age_ge75	0.01403	0.01596	0.04611	0.04934	0.76	0.61: 0.95	0.73	0.53: 1.00	0.68	0.46: 1.00	A	38.8	44.7	418	375
hCV3046185	2+3	ALL	source,apoe4,male,age_ge75	0.1948	0.18907	0.75707	0.03164	0.89	0.74: 1.06	0.96	0.75: 1.24	0.67	0.46: 0.96	A	36.1	38.7	550	794
hCV3046185	1	apoe4=1	male,age_ge75	0.07265	0.07619	0.44312	0.02109	0.72	0.51: 1.03	0.82	0.48: 1.38	0.49	0.27: 0.90	A	39.2	47.7	240	87
hCV3046185	2+3	apoe4=1	source,male,age_ge75	0.04999	0.04976	0.17959	0.04783	0.76	0.59: 1.00	0.77	0.52: 1.13	0.59	0.35: 0.99	A	35.2	41.9	341	191
hCV3088744	1	age_ge75=0	apoe4,male	0.01268	0.01078	0.05961	0.02135	1.56	1.10: 2.21	1.62	0.97: 2.70	2.19	1.10: 4.36	A	48.4	36.6	191	157
hCV3088744	3	age_ge75=0	apoe4,male	0.11248	0.10833	0.55893	0.03335	1.37	0.93: 2.02	1.2	0.66: 2.18	2.07	1.04: 4.10	A	46.4	42.6	137	149
hCV3091316	1	apoe4=1	male,age_ge75	0.49187	0.50675	0.8485	0.00697	0.83	0.49: 1.41	1.06	0.57: 1.96	0.13	0.02: 0.72	C	11.6	13.2	238	87
hCV3091316	2	apoe4=1	male,age_ge75	0.65341	0.63806	0.42881	0.03151	1.16	0.59: 2.27	1.34	0.64: 2.81	0		C	11.4	11.1	110	81
hCV3137872	1	male=0	apoe4,age_ge75	0.00529	0.00659	0.01752	0.03618	1.47	1.12: 1.93	1.68	1.10: 2.56	1.64	1.02: 2.63	C	47.9	40.8	263	234
hCV3137872	2	male=0	apoe4,age_ge75	0.41721	0.43866	0.59366	0.04199	1.18	0.80: 1.74	0.86	0.49: 1.51	1.97	1.02: 3.82	C	45.1	41.3	92	254
hCV3159528	3	age_ge75=0	apoe4,male	0.02004	0.02111	0.15371	0.00961	1.62	1.08: 2.43	1.49	0.86: 2.59	3.47	1.41: 8.57	C	40.2	32.0	138	150
hCV3159528	1+2	age_ge75=0	source,apoe4,male	0.29309	0.2992	0.91247	0.03286	1.16	0.88: 1.51	0.98	0.67: 1.43	1.83	1.07: 3.14	C	40.2	37.4	306	282
hCV3159528	1	age_ge75=1	apoe4,male	0.03855	0.04276	0.66946	0.00132	1.34	1.02: 1.77	1.09	0.73: 1.64	2.32	1.37: 3.91	C	44.9	37.7	225	216
hCV3159528	2+3	age_ge75=1	source,apoe4,male	0.0103	0.00897	0.00364	0.31462	1.35	1.07: 1.70	1.64	1.17: 2.28	1.26	0.80: 2.00	C	39.8	35.6	293	513
hCV3159528	1	ALL	apoe4,male,age_ge75	0.02921	0.03162	0.79558	0.00022	1.28	1.03: 1.59	1.04	0.76: 1.43	2.21	1.45: 3.38	C	42.7	37.8	416	374
hCV3159528	2+3	ALL	source,apoe4,male,age_ge75	0.00097	0.00092	0.00371	0.01452	1.35	1.13: 1.62	1.46	1.13: 1.88	1.58	1.10: 2.27	C	40.0	35.1	546	787
hCV3159528	1	apoe4=1	male,age_ge75	0.27797	0.28061	0.80507	0.02059	1.22	0.85: 1.76	0.94	0.55: 1.59	2.55	1.14: 5.71	C	42.0	37.8	238	86
hCV3159528	2+3	apoe4=1	source,male,age_ge75	0.00215	0.00259	0.00535	0.03973	1.56	1.17: 2.07	1.72	1.18: 2.53	1.92	1.05: 3.52	C	39.3	31.3	338	187
hCV3159528	1	male=0	apoe4,age_ge75	0.14074	0.14883	0.77361	0.02131	1.23	0.93: 1.62	1.06	0.71: 1.59	1.86	1.10: 3.13	C	43.7	39.5	263	234
hCV3159528	2+3	male=0	source,apoe4,age_ge75	0.00904	0.01072	0.05293	0.0189	1.35	1.08: 1.68	1.35	1.00: 1.84	1.71	1.10: 2.65	C	39.5	34.6	372	544
hCV3159529	3	age_ge75=0	apoe4,male	0.02704	0.02999	0.10386	0.03262	1.66	1.07: 2.58	1.58	0.91: 2.73	4.21	1.21: 14.57	G	32.2	23.3	138	150
hCV3159529	1+2	age_ge75=0	source,apoe4,male	0.82632	0.8278	0.3508	0.04419	1.03	0.77: 1.40	0.83	0.56: 1.23	2.11	1.04: 4.25	G	31.4	31.0	285	255

TABLE 6, page 5 of 6

Marker	Sample set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV3159529	1	age_ge75=1	NONE	0.14153	0.14165	0.60157	0.01966	1.24	0.93: 1.67	1.11	0.75: 1.63	2.19	1.12: 4.27	G	34.5	29.8	210	205
hCV3159529	2+3	age_ge75=1	source	0.00124	0.00089	0.00204	0.04062	1.46	1.16: 1.83	1.6	1.19: 2.16	1.8	1.03: 3.17	G	33.6	26.8	298	487
hCV3159529	1	ALL	apoe4,male,age_ge75	0.1259	0.12997	0.88282	0.00043	1.2	0.95: 1.53	0.98	0.71: 1.34	2.62	1.51: 4.54	G	33.1	30.1	388	352
hCV3159529	2+3	ALL	source,apoe4,male,age_ge75	0.00039	0.00031	0.00103	0.0173	1.43	1.17: 1.74	1.53	1.19: 1.97	1.85	1.13: 3.02	G	32.7	26.8	529	745
hCV3159529	1	apoe4=1	male,age_ge75	0.95704	0.95693	0.17523	0.03108	0.99	0.67: 1.47	0.69	0.41: 1.17	3.64	1.05: 12.54	G	31.3	31.9	225	80
hCV3159529	2+3	apoe4=1	source,male,age_ge75	0.00006	0.00009	0.00029	0.00915	1.91	1.39: 2.62	2.07	1.39: 3.06	3.09	1.34: 7.17	G	33.4	22.0	328	173
hCV3159529	3	male=0	apoe4,age_ge75	0.00196	0.0019	0.00878	0.01336	1.58	1.19: 2.11	1.63	1.13: 2.34	2.72	1.25: 5.90	G	32.3	24.0	282	294
hCV3159529	1+2	male=0	source,apoe4,age_ge75	0.21647	0.22307	0.75888	0.03446	1.17	0.91: 1.50	1.05	0.76: 1.47	1.85	1.07: 3.21	G	34.0	30.5	329	442
hCV3178541	1	apoe4=0	male,age_ge75	0.04232	0.04917	0.01333	0.79712	1.35	1.01: 1.80	1.62	1.10: 2.37	1.09	0.58: 2.03	T	33.7	27.5	178	287
hCV3178541	2+3	apoe4=0	source,male,age_ge75	0.0071	0.00946	0.02062	0.05695	1.4	1.10: 1.80	1.47	1.06: 2.05	1.61	0.98: 2.64	T	35.0	28.4	207	596
hCV3215842	1	apoe4=1	male,age_ge75	0.01644	0.01796	0.04044	0.07128	1.61	1.09: 2.39	1.68	1.02: 2.76	2.38	0.90: 6.29	T	34.8	25.0	240	86
hCV3215842	2	apoe4=1	male,age_ge75	0.0503	0.03623	0.03729	0.31284	1.61	1.00: 2.58	1.95	1.04: 3.68	1.9	0.57: 6.34	T	38.6	28.4	110	81
hCV3268994	1	male=1	NONE	0.0448	0.04232	0.1405	0.04512	0.69	0.48: 0.99	0.7	0.43: 1.13	0.42	0.17: 1.00	C	27.1	35.1	140	134
hCV3268994	2+3	male=1	source	0.03918	0.03145	0.02366	0.44352	0.74	0.56: 0.99	0.65	0.45: 0.94	0.75	0.36: 1.57	C	25.9	32.1	224	235
hCV337151	1	age_ge75=0	NONE	0.05433	0.04451	0.03856	0.33213	0.73	0.53: 1.01	0.61	0.39: 0.98	0.72	0.36: 1.41	G	33.9	41.3	174	144
hCV337151	2+3	age_ge75=0	source	0.00005	0.00004	0.00093	0.00064	0.59	0.45: 0.76	0.55	0.38: 0.78	0.39	0.22: 0.68	G	31.1	43.4	246	259
hCV337151	1	ALL	NONE	0.0314	0.03172	0.0511	0.13221	0.79	0.64: 0.98	0.74	0.55: 1.00	0.74	0.49: 1.10	G	36.9	42.4	385	348
hCV337151	2+3	ALL	source	0.00799	0.00768	0.0413	0.01801	0.8	0.68: 0.94	0.79	0.62: 0.99	0.68	0.49: 0.94	G	36.1	41.4	545	746
hCV472673	1	age_ge75=1	apoe4,male	0.05433	0.04497	0.01482	0.47051	1.32	0.99: 1.75	1.79	1.12: 2.84	1.2	0.73: 1.96	C	49.5	44.2	212	206
hCV472673	2+3	age_ge75=1	source,apoe4,male	0.05434	0.04665	0.07595	0.1519	1.24	0.99: 1.55	1.39	0.96: 2.00	1.31	0.90: 1.90	C	50.9	46.9	293	514
hCV589703	1	age_ge75=1	apoe4,male	0.21666	0.21795	0.04782	0.04705	1.28	0.86: 1.90	1.55	1.00: 2.43	0.19	0.03: 1.07	G	14.6	11.5	226	217
hCV589703	2	age_ge75=1	apoe4,male	0.03049	0.03539	0.02945	0.47082	1.8	1.07: 3.04	1.94	1.07: 3.52	1.91	0.36: 10.00	G	20.4	13.2	71	268
hCV7547730	2	apoe4=1	male,age_ge75	0.03911	0.03825	0.11864	0.04237	1.6	1.02: 2.51	1.54	0.89: 2.65	4.45	0.93: 21.29	A	29.4	19.1	158	97
hCV7547730	1+3	apoe4=1	source,male,age_ge75	0.05896	0.05761	0.17171	0.04388	1.32	0.99: 1.78	1.28	0.90: 1.81	2.51	0.98: 6.43	A	26.2	20.9	473	182
hCV7611203	1	male=0	NONE	0.01502	0.01464	0.02194	0.12572	1.4	1.07: 1.83	1.51	1.06: 2.15	1.6	0.87: 2.95	T	34.906	27.8	265.0	236
hCV7611203	3	male=0	NONE	0.04993	0.05699	0.11836	0.10555	1.29	1.00: 1.67	1.3	0.94: 1.80	1.6	0.90: 2.82	T	31.25	26.0	280.0	294
hCV811329	1	ALL	apoe4,male,age_ge75	0.01396	0.01272	0.0136	0.28952	0.71	0.54: 0.93	0.67	0.49: 0.92	0.62	0.26: 1.50	A	17.1	21.0	418	374
hCV811329	2+3	ALL	source,apoe4,male,age_ge75	0.03631	0.03217	0.11103	0.02153	0.8	0.65: 0.99	0.82	0.65: 1.05	0.41	0.19: 0.86	A	18.9	22.1	630	782
hCV8161028	1	male=1	apoe4,age_ge75	0.04534	0.04569	0.04492	0.28183	0.69	0.48: 0.99	0.6	0.36: 0.99	0.66	0.31: 1.41	C	31.8	37.5	154	140
hCV8161028	3	male=1	apoe4,age_ge75	0.04863	0.06271	0.0817	0.18643	0.63	0.40: 1.00	0.56	0.30: 1.07	0.55	0.23: 1.32	C	34.0	43.1	78	101
hCV8227677	1	age_ge75=0	apoe4,male	0.01889	0.01684	0.50634	0.00141	1.55	1.07: 2.23	1.24	0.67: 2.29	2.77	1.46: 5.26	C	55.4	45.5	177	146
hCV8227677	2+3	age_ge75=0	source,apoe4,male	0.00638	0.00868	0.0621	0.01249	1.49	1.12: 2.00	1.52	0.97: 2.38	1.87	1.15: 3.04	C	52.7	43.7	244	253
hCV8227677	2	age_ge75=1	apoe4,male	0.00812	0.01146	0.00843	0.12238	1.78	1.16: 2.73	2.73	1.27: 5.90	1.69	0.87: 3.29	C	58.2	46.6	61	238
hCV8227677	1+3	age_ge75=1	source,apoe4,male	0.02759	0.02462	0.52246	0.00208	1.25	1.02: 1.51	1.11	0.81: 1.52	1.7	1.21: 2.38	C	50.1	45.7	433	450
hCV8227677	1	ALL	apoe4,male,age_ge75	0.00601	0.00587	0.95775	0.00001	1.37	1.09: 1.71	0.99	0.68: 1.43	2.35	1.61: 3.43	C	54.6	46.9	389	350
hCV8227677	2+3	ALL	source,apoe4,male,age_ge75	0.00016	0.00017	0.00136	0.00324	1.41	1.18: 1.69	1.59	1.20: 2.12	1.59	1.17: 2.16	C	50.7	44.7	526	737
hCV8227677	1	apoe4=1	male,age_ge75	0.03439	0.02936	0.97154	0.00079	1.48	1.03: 2.14	1.01	0.54: 1.90	3.32	1.60: 6.86	C	55.5	45.7	226	81
hCV8227677	2+3	apoe4=1	source,male,age_ge75	0.0005	0.00057	0.00844	0.00234	1.64	1.24: 2.17	1.75	1.15: 2.66	2.25	1.33: 3.83	C	49.8	38.2	325	170
hCV8227677	1	male=0	apoe4,age_ge75	0.05246	0.05073	0.91673	0.00264	1.32	1.00: 1.76	1.03	0.63: 1.66	2.03	1.27: 3.24	C	55.4	48.2	250	218
hCV8227677	2+3	male=0	source,apoe4,age_ge75	0.00083	0.00105	0.01265	0.00349	1.45	1.17: 1.81	1.55	1.10: 2.18	1.75	1.20: 2.54	C	50.7	44.4	360	515
hCV848829	1	age_ge75=0	apoe4,male	0.04828	0.044	0.12597	0.0418	1.5	0.99: 2.26	1.46	0.89: 2.38	5.75	1.10: 29.98	C	25.8	17.9	190	156
hCV848829	2+3	age_ge75=0	source,apoe4,male	0.12326	0.12435	0.60075	0.0038	1.29	0.93: 1.78	1.11	0.75: 1.65	3.87	1.56: 9.55	C	25.4	22.8	252	268
hCV855979	1	ALL	NONE	0.0169	0.01833	0.01375	0.63623	1.5	1.07: 2.09	1.57	1.10: 2.26	1.36	0.38: 4.85	T	11.8	8.2	418	377
hCV855979	2+3	ALL	source	0.04501	0.05019	0.05276	0.41742	1.3	1.01: 1.69	1.32	1.00: 1.75	1.49	0.57: 3.86	T	10.9	8.5	562	790
hCV8715115	1	ALL	NONE	0.03239	0.03319	0.04741	0.17723	1.71	1.04: 2.82	1.67	1.00: 2.80			A	5.9	3.5	391	355
hCV8715115	2+3	ALL	source	0.00831	0.00916	0.00997	0.40659	1.73	1.15: 2.61	1.74	1.14: 2.66	3.05	0.23: 40.80	A	4.9	2.9	546	754
hCV8715115	2	apoe4=0	NONE	0.07336	0.08026	0.04784	0.56528	2.08	0.92: 4.72	2.31	0.99: 5.39			A	5.4	2.7	92	279
hCV8715115	1+3	apoe4=0	source	0.00562	0.00545	0.00775	0.15524	1.97	1.21: 3.22	1.96	1.18: 3.25			A	5.4	2.8	314	576
hCV8715115	1	male=0	NONE	0.02755	0.03024	0.0436	0.18457	1.99	1.07: 3.72	1.93	1.01: 3.67			A	6.6	3.4	251	220
hCV8715115	2+3	male=0	source	0.00327	0.00361	0.00544	0.14626	2.09	1.27: 3.44	2.05	1.23: 3.43			A	5.7	2.7	371	523
hCV8725171	1	apoe4=1	male,age_ge75	0.01489	0.01922	0.0309	0.11628	1.83	1.12: 2.98	1.86	1.06: 3.27	3.1	0.70: 13.77	G	24.0	14.8	227	81

TABLE 6, page 6 of 6

Marker	Sample set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV8725171	2+3	apoe4=1	source,male,age_ge75	0.04864	0.04729	0.08495	0.13559	1.38	1.00: 1.92	1.39	0.95: 2.03	2.32	0.75: 7.15	G	23.3	17.3	390	185
hCV8725171	1	male=0	apoe4,age_ge75	0.05495	0.0626	0.04876	0.48103	1.41	1.00: 2.00	1.52	1.00: 2.31	1.42	0.57: 3.55	G	24.2	19.2	250	219
hCV8725171	2+3	male=0	source,apoe4,age_ge75	0.00894	0.01119	0.03595	0.02825	1.43	1.09: 1.86	1.4	1.02: 1.92	2.21	1.09: 4.47	G	24.2	17.9	370	533
hCV8782652	1	age_ge75=1	apoe4,male	0.05378	0.04595	0.01817	0.41898	1.32	1.00: 1.75	1.76	1.10: 2.80	1.22	0.75: 1.99	T	50.0	44.6	212	205
hCV8782652	3	age_ge75=1	apoe4,male	0.11023	0.09478	0.0391	0.55991	1.25	0.95: 1.63	1.59	1.02: 2.48	1.15	0.72: 1.84	T	50.0	46.3	222	246
hCV8856240	1	age_ge75=1	apoe4,male	0.02985	0.02881	0.04189	0.19808	1.52	1.04: 2.21	1.56	1.02: 2.38	2.42	0.61: 9.65	G	19.4	13.2	227	216
hCV8856240	2+3	age_ge75=1	source,apoe4,male	0.03269	0.03653	0.07607	0.08502	1.38	1.03: 1.84	1.36	0.97: 1.92	2.19	0.93: 5.14	G	20.0	17.5	295	509
hCV8856240	1	ALL	apoe4,male,age_ge75	0.00229	0.0025	0.00758	0.02729	1.57	1.17: 2.09	1.57	1.13: 2.18	3.02	1.13: 8.10	G	20.5	14.2	418	374
hCV8856240	2+3	ALL	source,apoe4,male,age_ge75	0.03055	0.03323	0.03945	0.25711	1.28	1.02: 1.60	1.32	1.01: 1.72	1.48	0.77: 2.84	G	20.1	18.0	551	784
hCV8856240	1	apoe4=1	NONE	0.02119	0.02356	0.04583	0.095	1.77	1.08: 2.89	1.75	1.01: 3.03	4.87	0.63: 37.78	G	21.5	13.4	240	86
hCV8856240	2+3	apoe4=1	source	0.03451	0.04378	0.0787	0.1165	1.45	1.03: 2.05	1.43	0.96: 2.12	2.22	0.81: 6.11	G	20.2	14.4	351	187
hCV8921255	3	apoe4=0	NONE	0.03761	0.03326	0.02535	0.37977	0.72	0.52: 0.98	0.63	0.42: 0.95	0.72	0.34: 1.51	G	26.6	33.6	141	301
hCV8921255	1+2	apoe4=0	source	0.03824	0.0422	0.05682	0.18146	0.78	0.61: 0.99	0.74	0.54: 1.01	0.72	0.44: 1.18	G	30.2	35.7	230	544
hCV9579537	1	ALL	NONE	0.03721	0.03826	0.01924	0.56977	1.5	1.02: 2.21	1.63	1.08: 2.45	0.6	0.10: 3.60	C	8.8	6.0	417	375
hCV9579537	3	ALL	NONE	0.04507	0.04904	0.08328	0.12067	1.44	1.01: 2.07	1.41	0.96: 2.07	3.32	0.67: 16.57	C	10.3	7.4	359	393
hCV9632133	3	apoe4=1	NONE	0.01457	0.01599	0.00795	0.38552	1.63	1.10: 2.40	1.94	1.18: 3.16	1.48	0.61: 3.57	G	32.9	23.2	219	95
hCV9632133	1+2	apoe4=1	source	0.03481	0.035	0.16782	0.01781	1.38	1.02: 1.87	1.31	0.89: 1.91	2.73	1.17: 6.38	G	32.0	25.3	344	158
hDV68530985	1	age_ge75=0	NONE	0.00049	0.00051	0.00127	0.0285	1.95	1.34: 2.86	2.11	1.34: 3.33	3.3	1.07: 10.19	C	29.4	17.6	175	145
hDV68530985	3	age_ge75=0	NONE	0.05325	0.05421	0.19047	0.02565	1.47	0.99: 2.16	1.37	0.85: 2.19	3.48	1.09: 11.05	C	26.8	20.0	138	150
hDV68530994	3	age_ge75=0	NONE	0.04207	0.04337	0.15387	0.02565	1.5	1.01: 2.21	1.41	0.88: 2.26	3.48	1.09: 11.05	T	26.8	19.7	138	150
hDV68530994	1+2	age_ge75=0	source	0.02527	0.03175	0.06092	0.09803	1.38	1.04: 1.83	1.39	0.98: 1.96	1.77	0.90: 3.48	T	27.0	21.2	285	255
hDV68530995	3	age_ge75=0	NONE	0.0329	0.03442	0.12278	0.02565	1.53	1.03: 2.26	1.45	0.90: 2.33	3.48	1.09: 11.05	A	26.8	19.3	138	150
hDV68530995	1+2	age_ge75=0	source	0.03873	0.04554	0.07181	0.15451	1.35	1.01: 1.78	1.37	0.97: 1.93	1.65	0.83: 3.28	A	26.8	21.4	282	255

TABLE 7, page 1 of 8

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV1027219	1	male=1	apoe4,age_ge75	0.00771	0.00722	0.09036	0.0017	1.68	1.15: 2.48	1.53	0.94: 2.51	5.35	1.78: 16.07	A	34.4	23.9	154	140
hCV1027219	1+2	male=1	source,apoe4,age_ge75	0.0487	0.0419	0.2554	0.00868	1.34	1.00: 1.79	1.25	0.85: 1.83	2.7	1.30: 5.60	A	34.0	27.8	243	268
hCV1054616	1	male=0	apoe4,age_ge75	0.22834	0.22945	0.60006	0.00712	1.18	0.90: 1.55	0.9	0.59: 1.36	1.96	1.20: 3.20	G	47.9	42.5	264	234
hCV1054616	1+2	male=0	source,apoe4,age_ge75	0.05876	0.05832	0.82983	0.00258	1.24	0.99: 1.55	1.04	0.73: 1.47	1.82	1.23: 2.70	G	49.2	44.9	354	483
hCV11192460	1	age_ge75=1	apoe4,male	0.00014	0.00023	0.00254	0.00069	2.29	1.48: 3.53	2.07	1.29: 3.33		. : .	A	17.0	8.5	227	217
hCV11192460	1+2+3	age_ge75=1	source,apoe4,male	0.00116	0.00137	0.00637	0.00666	1.56	1.19: 2.04	1.51	1.12: 2.02	4.58	1.50: 13.96	A	14.3	10.5	520	731
hCV11192460	1	ALL	apoe4,male,age_ge75	0.00234	0.00303	0.00693	0.04556	1.65	1.19: 2.28	1.64	1.14: 2.36	2.74	0.93: 8.05	A	15.1	10.5	418	375
hCV11192460	1+2+3	ALL	source,apoe4,male,age_ge75	0.00142	0.00159	0.00238	0.11618	1.4	1.14: 1.72	1.42	1.13: 1.79	1.79	0.85: 3.76	A	13.6	11.2	964	1163
hCV11192460	3	male=0	apoe4,age_ge75	0.0253	0.02334	0.02551	0.404	1.58	1.06: 2.37	1.65	1.06: 2.56	2.57	0.33: 19.99	A	13.3	10.0	282	295
hCV11192460	1+2+3	male=0	source,apoe4,age_ge75	0.00442	0.00427	0.01246	0.02438	1.46	1.13: 1.89	1.44	1.08: 1.91	3.9	1.21: 12.55	A	13.8	11.2	636	780
hCV11193939	1	age_ge75=0	apoe4,male	0.00036	0.00034	0.00005	0.32857	2.03	1.37: 3.00	2.8	1.69: 4.65	1.66	0.64: 4.29	G	35.6	22.9	191	157
hCV11193939	1+2+3	age_ge75=0	source,apoe4,male	0.00354	0.00324	0.00295	0.15258	1.43	1.12: 1.81	1.58	1.17: 2.14	1.55	0.87: 2.78	G	32.7	26.6	445	432
hCV11214738	3	ALL	NONE	0.04563	0.04582	0.05589	0.31298	1.36	1.01: 1.83	1.38	0.99: 1.93	1.77	0.57: 5.47	C	14.9	11.4	360	395
hCV11214738	1+2+3	ALL	source	0.00675	0.00649	0.01498	0.05616	1.27	1.07: 1.51	1.27	1.05: 1.55	1.8	0.98: 3.32	C	17.1	13.9	932	1100
hCV11278562	2	male=1	NONE	0.02953	0.02897	0.15597	0.00718	1.57	1.04: 2.37	1.43	0.87: 2.34	6.3	1.38: 28.71	T	27.2	19.2	134	130
hCV11278562	1+2+3	male=1	source	0.03375	0.02955	0.13322	0.00995	1.31	1.02: 1.68	1.26	0.93: 1.69	2.82	1.24: 6.41	T	24.7	19.9	363	366
hCV11568644	1	apoe4=0	male,age_ge75	0.02405	0.02419	0.09906	0.03979	0.73	0.56: 0.96	0.71	0.47: 1.07	0.6	0.37: 0.98	G	41.3	48.8	178	286
hCV11568644	1+2+3	apoe4=0	source,male,age_ge75	0.0131	0.01351	0.17293	0.00532	0.8	0.67: 0.95	0.83	0.64: 1.08	0.62	0.45: 0.87	G	40.4	44.8	382	886
hCV11574262	1	age_ge75=0	apoe4,male	0.11784	0.09633	0.01219	0.96185	1.34	0.93: 1.93	2.15	1.17: 3.95	1.02	0.52: 1.98	A	51.7	42.9	174	147
hCV11574262	1+3	age_ge75=0	source,apoe4,male	0.12259	0.10832	0.00578	0.82526	1.23	0.95: 1.60	1.87	1.19: 2.92	0.95	0.61: 1.49	A	51.3	46.6	312	297
hCV11597077	1	age_ge75=0	NONE	0.12292	0.12657	0.60509	0.04495	0.79	0.59: 1.07	0.88	0.55: 1.42	0.61	0.37: 0.99	A	46.4	52.2	192	159
hCV11597077	1+3	age_ge75=0	source	0.02906	0.03132	0.33082	0.00862	0.78	0.63: 0.98	0.84	0.60: 1.19	0.6	0.41: 0.88	A	43.9	49.7	330	309
hCV11720402	1	age_ge75=1	apoe4,male	0.19273	0.17864	0.97646	0.01586	0.83	0.63: 1.10	0.99	0.66: 1.51	0.52	0.30: 0.89	T	41.0	44.9	227	215
hCV11720402	1+2	age_ge75=1	source,apoe4,male	0.07201	0.06491	0.56292	0.00836	0.81	0.65: 1.02	0.9	0.64: 1.27	0.56	0.36: 0.86	T	40.5	44.5	301	482
hCV11720789	1	apoe4=0	male,age_ge75	0.0492	0.06695	0.02861	0.77578	2.08	1.01: 4.27	2.38	1.10: 5.15	0.69	0.06: 8.24	T	5.3	2.6	178	288
hCV11720789	1+2+3	apoe4=0	source,male,age_ge75	0.00759	0.0104	0.0072	0.67683	1.85	1.17: 2.92	1.92	1.19: 3.10	1.47	0.23: 9.42	T	4.5	2.8	421	893
hCV11842860	1	age_ge75=0	apoe4,male	0.04875	0.0543	0.01039	0.84279	0.68	0.47: 0.99	0.51	0.30: 0.86	0.92	0.44: 1.96	T	32.7	39.7	179	146
hCV11842860	1+3	age_ge75=0	source,apoe4,male	0.01079	0.01189	0.00228	0.51318	0.7	0.53: 0.92	0.56	0.38: 0.82	0.83	0.48: 1.44	T	32.6	39.2	316	296
hCV11855743	1	ALL	NONE	0.18037	0.19033	0.86492	0.00328	1.18	0.93: 1.50	1.03	0.77: 1.37	2.59	1.35: 4.97	A	25.1	22.1	391	355
hCV11855743	1+2	ALL	source	0.46747	0.47429	0.5079	0.00142	1.07	0.89: 1.29	0.93	0.74: 1.16	2.15	1.34: 3.45	A	24.2	22.9	594	749
hCV11855743	2	male=0	NONE	0.80606	0.81008	0.16153	0.0381	0.95	0.65: 1.39	0.71	0.44: 1.15	2.31	1.03: 5.18	A	25.0	25.9	100	251
hCV11855743	1+2	male=0	source	0.92992	0.93115	0.16149	0.00857	0.99	0.78: 1.26	0.81	0.61: 1.09	2.17	1.20: 3.92	A	23.2	24.2	351	471
hCV11861096	1	age_ge75=1	apoe4,male	0.00016	0.00026	0.00247	0.00105	2.38	1.51: 3.77	2.16	1.31: 3.57		. : .	C	16.7	8.0	210	205
hCV11861096	1+2+3	age_ge75=1	source,apoe4,male	0.00058	0.00072	0.00292	0.00998	1.63	1.24: 2.15	1.59	1.17: 2.16	4.31	1.40: 13.28	C	14.3	10.1	494	693
hCV12029086	1	male=0	NONE	0.00304	0.0031	0.00605	0.06	0.66	0.50: 0.87	0.61	0.43: 0.87	0.52	0.26: 1.04	T	23.1	31.4	264	237
hCV12029086	1+2+3	male=0	source	0.00169	0.00143	0.00457	0.02612	0.76	0.65: 0.90	0.74	0.59: 0.91	0.62	0.40: 0.95	T	25.1	30.7	648	786
hCV12123244	1	apoe4=0	male,age_ge75	0.04201	0.03511	0.00619	0.42419	0.71	0.51: 0.99	0.58	0.39: 0.86	1.47	0.56: 3.85	C	18.5	24.7	178	288
hCV12123244	1+2+3	apoe4=0	source,male,age_ge75	0.02194	0.01981	0.00961	0.77366	0.78	0.63: 0.97	0.72	0.56: 0.92	0.91	0.48: 1.74	C	17.9	21.5	420	893
hCV12126867	3	male=1	apoe4,age_ge75	0.03958	0.04997	0.37624	0.00743	0.61	0.38: 0.98	0.75	0.40: 1.41	0.28	0.10: 0.77	A	30.1	39.5	78	100
hCV12126867	1+3	male=1	source,apoe4,age_ge75	0.19199	0.20504	0.8121	0.00161	0.82	0.61: 1.11	1.05	0.71: 1.54	0.36	0.18: 0.71	A	28.3	32.3	230	240
hCV1229667	1	age_ge75=1	apoe4,male	0.00571	0.00442	0.00442		0.44	0.24: 0.80	0.42	0.22: 0.78			C	3.8	7.9	225	216
hCV1229667	1+2+3	age_ge75=1	source,apoe4,male	0.00445	0.00355	0.00254	0.45939	0.59	0.40: 0.86	0.55	0.37: 0.82		. : .	C	4.4	7.4	517	718
hCV1229667	1	ALL	apoe4,male,age_ge75	0.00278	0.00227	0.00163	0.5827	0.5	0.32: 0.79	0.47	0.29: 0.76		. : .	C	4.6	7.6	416	374
hCV1229667	1+2+3	ALL	source,apoe4,male,age_ge75	0.00413	0.0037	0.0018	0.37244	0.66	0.50: 0.88	0.63	0.47: 0.85	4.78	0.22:105.98	C	5.1	7.1	958	1143
hCV1244849	1	apoe4=1	male,age_ge75	0.02575	0.02942	0.02315	0.55282	2.1	1.09: 4.07	2.27	1.11: 4.66	2.03	0.21: 19.85	C	13.9	7.5	226	80
hCV1244849	1+2+3	apoe4=1	source,male,age_ge75	0.0092	0.01094	0.01498	0.16496	1.57	1.12: 2.19	1.59	1.10: 2.30	2.4	0.71: 8.06	C	15.9	11.1	555	257
hCV1305685	1	age_ge75=0	apoe4,male	0.03421	0.03178	0.03102	0.30503	1.53	1.03: 2.26	1.72	1.05: 2.84	1.76	0.64: 4.81	A	29.5	25.6	190	158
hCV1305685	1+2+3	age_ge75=0	source,apoe4,male	0.01249	0.01345	0.00894	0.34492	1.36	1.07: 1.73	1.49	1.11: 2.02	1.34	0.74: 2.43	A	26.3	23.1	516	438
hCV1305685	1	ALL	apoe4,male,age_ge75	0.00842	0.0071	0.00807	0.18163	1.39	1.09: 1.77	1.51	1.11: 2.04	1.58	0.82: 3.06	A	28.8	24.1	417	375
hCV1305685	1+2+3	ALL	source,apoe4,male,age_ge75	0.0059	0.00589	0.00963	0.09594	1.24	1.07: 1.45	1.29	1.06: 1.57	1.39	0.94: 2.06	A	27.1	23.8	961	1147
hCV1305685	1	male=0	apoe4,age_ge75	0.00042	0.00036	0.0002	0.15197	1.74	1.28: 2.37	2.09	1.41: 3.10	1.8	0.82: 3.96	A	31.4	23.0	264	235

WO 2006/022639 PCT/US2004/024001

TABLE 7, page 2 of 8

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV1305685	1+2+3	male=0	source,apoe4,age_ge75	0.00244	0.0026	0.00344	0.09943	1.35	1.11: 1.64	1.43	1.13: 1.82	1.5	0.93: 2.43	A	27.3	23.0	635	769
hCV1322419	1	apoe4=1	NONE	0.00521	0.00696	0.00573	0.08913	1.68	1.16: 2.41	2.14	1.24: 3.69	1.7	0.92: 3.15	A	53.5	40.7	227	81
hCV1322419	1+2+3	apoe4=1	source	0.0046	0.00451	0.01088	0.03659	1.36	1.10: 1.67	1.54	1.10: 2.15	1.47	1.02: 2.10	A	52.6	44.9	564	257
hCV1345818	1	male=0	NONE	0.00932	0.00992	0.00615	0.9369	2.14	1.19: 3.83	2.31	1.25: 4.27	0.89	0.06: 14.37	T	7.4	3.6	265	237
hCV1345818	1+2	male=0	source	0.00537	0.00563	0.00281	0.68412	1.86	1.20: 2.88	2.01	1.27: 3.18	0.58	0.05: 7.43	T	7.3	4.2	365	488
hCV1345858	1	male=0	NONE	0.00678	0.00718	0.00434	0.9369	2.19	1.23: 3.93	2.38	1.29: 4.39	0.89	0.06: 14.37	C	7.5	3.6	265	237
hCV1345858	1+2+3	male=0	source	0.00631	0.00575	0.00374	0.67853	1.57	1.14: 2.18	1.65	1.17: 2.31	0.57	0.05: 7.33	C	7.0	4.7	646	772
hCV1345864	1	male=0	NONE	0.0086	0.00914	0.00564	0.94094	2.15	1.20: 3.86	2.33	1.26: 4.31	0.9	0.06: 14.48	A	7.4	3.6	262	236
hCV1345864	1+3	male=0	source	0.0118	0.01069	0.00863	0.941	1.6	1.11: 2.32	1.67	1.14: 2.45	0.9	0.06: 14.48	A	7.0	4.5	556	533
hCV1348542	1	apoe4=1	male,age_ge75	0.01594	0.01481	0.05726	0.02521	1.73	1.10: 2.72	1.66	0.98: 2.82	6.65	0.93: 47.35	G	29.8	19.5	223	77
hCV1348542	1+3	apoe4=1	source,male,age_ge75	0.03994	0.03578	0.20137	0.00702	1.37	1.01: 1.84	1.26	0.88: 1.81	4.35	1.34: 14.15	G	28.0	21.8	441	172
hCV1406876	1	age_ge75=0	apoe4,male	0.04275	0.03859	0.03204	0.62658	1.65	1.02: 2.68	1.82	1.05: 3.14	1.54	0.27: 8.75	C	18.4	13.4	187	157
hCV1406876	1+2	age_ge75=0	source,apoe4,male	0.0165	0.01439	0.00772	0.91843	1.58	1.09: 2.30	1.78	1.16: 2.72	1.07	0.28: 4.15	C	17.6	13.2	295	272
hCV1406876	1	ALL	apoe4,male,age_ge75	0.01511	0.01502	0.02982	0.08584	1.43	1.07: 1.91	1.44	1.04: 2.01	2.29	0.87: 6.03	C	18.5	14.5	413	373
hCV1406876	1+2	ALL	source,apoe4,male,age_ge75	0.00695	0.00764	0.00907	0.19805	1.38	1.09: 1.75	1.43	1.09: 1.87	1.61	0.78: 3.31	C	18.1	15.0	586	741
hCV1413258	3	age_ge75=1	apoe4,male	0.04971	0.05108	0.00926	0.5652	0.72	0.51: 1.00	0.59	0.39: 0.88	1.32	0.53: 3.27	A	18.3	23.5	221	245
hCV1413258	1+3	age_ge75=1	source,apoe4,male	0.00961	0.01017	0.00191	0.96788	0.74	0.58: 0.93	0.64	0.48: 0.85	0.99	0.53: 1.83	A	19.6	24.6	433	450
hCV1413258	1	ALL	apoe4,male,age_ge75	0.00954	0.01016	0.01026	0.21544	0.71	0.55: 0.92	0.66	0.48: 0.91	0.65	0.33: 1.28	A	21.3	26.4	390	351
hCV1413258	1+3	ALL	source,apoe4,male,age_ge75	0.00649	0.00668	0.00213	0.59702	0.78	0.65: 0.93	0.7	0.56: 0.88	0.88	0.54: 1.42	A	21.7	25.6	749	746
hCV1413258	1	male=0	apoe4,age_ge75	0.00914	0.00925	0.00704	0.29804	0.65	0.47: 0.90	0.58	0.39: 0.86	0.63	0.27: 1.49	A	21.4	27.8	250	218
hCV1413258	1+3	male=0	source,apoe4,age_ge75	0.01028	0.01079	0.00212	0.88031	0.75	0.61: 0.94	0.66	0.50: 0.86	0.96	0.55: 1.67	A	22.4	26.7	531	512
hCV1489917	1	male=0	apoe4,age_ge75	0.1201	0.10708	0.64968	0.0099	0.79	0.59: 1.06	0.91	0.60: 1.37	0.41	0.21: 0.80	C	34.6	39.6	250	217
hCV1489917	1+2	male=0	source,apoe4,age_ge75	0.07933	0.06848	0.54373	0.00533	0.8	0.63: 1.02	0.9	0.64: 1.26	0.45	0.26: 0.78	C	34.7	37.8	329	446
hCV1558531	2	apoe4=0	male,age_ge75	0.0171	0.01495	0.04634	0.03842	0.53	0.31: 0.90	0.54	0.29: 1.00	0	. : .	C	16.7	25.9	60	276
hCV1558531	2+3	apoe4=0	source,male,age_ge75	0.00826	0.00769	0.02591	0.02894	0.68	0.51: 0.90	0.68	0.48: 0.95	0.35	0.13: 0.92	C	19.2	25.8	201	576
hCV15806020	1	male=1	NONE	0.00437	0.00634	0.00927	0.08644	0.58	0.40: 0.85	0.54	0.34: 0.86	0.48	0.21: 1.13	C	21.1	31.4	154	140
hCV15806020	1+3	male=1	source	0.00963	0.01028	0.02187	0.06868	0.69	0.52: 0.91	0.66	0.46: 0.94	0.52	0.26: 1.06	C	23.0	30.4	244	242
hCV15811970	1	ALL	apoe4,male,age_ge75	0.0131	0.01701	0.27857	0.00026	0.72	0.56: 0.94	0.84	0.61: 1.15	0.33	0.18: 0.63	T	22.2	28.1	390	351
hCV15811970	1+3	ALL	source,apoe4,male,age_ge75	0.30678	0.31908	0.90909	0.00982	0.91	0.76: 1.09	1.01	0.81: 1.27	0.57	0.37: 0.88	T	24.6	26.6	749	747
hCV15811970	1	male=1	apoe4,age_ge75	0.04694	0.04471	0.32168	0.00313	0.66	0.43: 1.00	0.77	0.45: 1.29	0.17	0.05: 0.62	T	23.2	28.6	140	133
hCV15811970	1+3	male=1	source,apoe4,age_ge75	0.05059	0.04744	0.30429	0.00321	0.72	0.52: 1.00	0.81	0.54: 1.21	0.24	0.09: 0.67	T	22.0	26.1	218	234
hCV15870743	1	age_ge75=0	NONE	0.00011	0.00016	0.00043	0.01391	0.51	0.36: 0.72	0.45	0.29: 0.70	0.38	0.17: 0.84	T	22.6	36.4	177	147
hCV15870743	1+2+3	age_ge75=0	source	0.00779	0.00777	0.00323	0.39168	0.75	0.61: 0.93	0.67	0.51: 0.87	0.81	0.50: 1.31	T	26.2	32.1	429	422
hCV15870743	1	ALL	NONE	0.0001	0.00014	0.00089	0.00398	0.64	0.51: 0.80	0.61	0.46: 0.82	0.47	0.27: 0.79	T	23.9	33.0	389	353
hCV15870743	1+2+3	ALL	source	0.00109	0.00113	0.00382	0.02003	0.8	0.69: 0.91	0.77	0.65: 0.92	0.68	0.49: 0.94	T	26.0	30.7	950	1140
hCV15870743	1	male=0	NONE	0.00037	0.00037	0.00075	0.0324	0.59	0.44: 0.79	0.53	0.37: 0.77	0.47	0.23: 0.95	T	22.9	33.3	249	219
hCV15870743	1+2+3	male=0	source	0.00118	0.00101	0.00135	0.0763	0.76	0.64: 0.90	0.7	0.57: 0.87	0.69	0.45: 1.04	T	25.6	31.3	630	762
hCV15873426	2	male=0	NONE	0.02509	0.02052	0.02333	0.22872	0.65	0.45: 0.95	0.58	0.37: 0.93	0.55	0.20: 1.48	T	24.0	32.6	100	250
hCV15873426	1+2+3	male=0	source	0.00808	0.00846	0.00953	0.14683	0.8	0.67: 0.94	0.75	0.61: 0.93	0.75	0.52: 1.10	T	27.4	32.0	632	762
hCV15887512	3	ALL	apoe4,male,age_ge75	0.05793	0.0572	0.1702	0.01776	1.33	0.99: 1.80	1.27	0.90: 1.78	3.73	1.14: 12.20	A	16.4	13.3	383	399
hCV15887512	1+3	ALL	source,apoe4,male,age_ge75	0.18896	0.19089	0.60743	0.00847	1.15	0.94: 1.41	1.06	0.84: 1.35	2.37	1.23: 4.57	A	18.5	16.5	772	751
hCV15887521	1	male=0	NONE	0.0004	0.00036	0.00072	0.03366	0.6	0.45: 0.80	0.53	0.37: 0.77	0.48	0.24: 0.96	T	24.1	34.6	249	218
hCV15887521	1+2+3	male=0	source	0.00257	0.00217	0.00598	0.03541	0.78	0.66: 0.92	0.74	0.60: 0.92	0.66	0.44: 0.97	T	27.6	33.1	631	762
hCV15887528	1	male=0	NONE	0.00035	0.00039	0.00061	0.03946	0.6	0.45: 0.79	0.53	0.36: 0.76	0.51	0.27: 0.98	T	24.4	35.1	248	218
hCV15887528	1+2+3	male=0	source	0.00199	0.00178	0.00495	0.03193	0.77	0.65: 0.91	0.73	0.59: 0.91	0.66	0.44: 0.97	T	27.7	33.4	630	760
hCV15919456	1	ALL	NONE	0.17589	0.17995	0.04199	0.8693	1.15	0.94: 1.40	1.39	1.01: 1.92	1.03	0.75: 1.41	A	51.7	48.3	419	376
hCV15919456	1+2+3	ALL	source	0.07805	0.0799	0.00682	0.86956	1.12	0.99: 1.26	1.32	1.08: 1.61	1.02	0.83: 1.24	A	51.5	49.0	979	1150
hCV15961334	3	apoe4=1	male,age_ge75	0.79687	0.79928	0.7969	0.02106	0.92	0.47: 1.79	1.1	0.53: 2.29	0	. : .	C	6.4	7.4	219	95
hCV15961334	1+2+3	apoe4=1	source,male,age_ge75	0.13691	0.13446	0.27225	0.00867	0.74	0.50: 1.11	0.79	0.52: 1.21	0	. : .	C	5.8	8.1	582	278
hCV15965240	3	age_ge75=0	NONE	0.08514	0.08009	0.04786	0.40137	1.33	0.96: 1.85	1.72	1.00: 2.96	1.27	0.73: 2.20	G	52.2	45.0	137	150
hCV15965240	1+3	age_ge75=0	source	0.07365	0.06884	0.00577	0.88222	1.23	0.98: 1.54	1.68	1.16: 2.43	1.03	0.71: 1.50	G	51.3	46.1	315	296

TABLE 7, page 3 of 8

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV15965240	3	apoe4=0	male,age_ge75	0.08829	0.08267	0.01196	0.82819	1.28	0.96: 1.71	1.86	1.14: 3.04	1.06	0.65: 1.73	G	51.4	45.2	141	300
hCV15965240	1+3	apoe4=0	source,male,age_ge75	0.08699	0.08622	0.00311	0.81837	1.19	0.98: 1.45	1.66	1.18: 2.32	0.96	0.69: 1.35	G	51.2	46.5	304	571
hCV15965240	1	male=1	apoe4,age_ge75	0.05983	0.05878	0.00504	0.72088	1.41	0.99: 2.02	2.46	1.30: 4.66	1.11	0.62: 1.99	G	55.0	49.3	140	135
hCV15965240	1+3	male=1	source,apoe4,age_ge75	0.04857	0.04177	0.00271	0.77478	1.33	1.00: 1.75	2.08	1.28: 3.39	1.07	0.67: 1.74	G	52.8	47.2	218	236
hCV16113167	1	age_ge75=0	NONE	0.00005	0.00007	0.00008	0.03408	0.49	0.34: 0.69	0.41	0.26: 0.64	0.43	0.19: 0.96	A	21.1	35.4	178	147
hCV16113167	1+2+3	age_ge75=0	source	0.00932	0.00938	0.00265	0.59322	0.76	0.61: 0.93	0.66	0.51: 0.87	0.87	0.53: 1.44	A	24.8	30.4	432	423
hCV16113167	1	ALL	NONE	0.00002	0.00003	0.00009	0.00734	0.61	0.48: 0.76	0.56	0.42: 0.75	0.48	0.28: 0.83	A	22.4	32.3	390	353
hCV16113167	1+2+3	ALL	source	0.00181	0.00184	0.00257	0.0807	0.8	0.70: 0.92	0.77	0.64: 0.91	0.74	0.53: 1.04	A	24.9	29.3	952	1141
hCV16113167	1	apoe4=0	male,age_ge75	0.0041	0.00504	0.00219	0.2839	0.63	0.46: 0.86	0.54	0.36: 0.80	0.68	0.34: 1.38	A	23.3	33.9	163	271
hCV16113167	1+2+3	apoe4=0	source,male,age_ge75	0.00671	0.00714	0.00331	0.35032	0.76	0.62: 0.93	0.68	0.53: 0.88	0.8	0.49: 1.28	A	24.3	30.5	371	870
hCV16113167	1	male=0	NONE	0.0001	0.00011	0.00009	0.06743	0.56	0.42: 0.75	0.48	0.33: 0.69	0.52	0.25: 1.06	A	21.2	32.4	250	219
hCV16113167	1+2+3	male=0	source	0.00133	0.00116	0.00056	0.24053	0.76	0.64: 0.90	0.68	0.55: 0.85	0.77	0.50: 1.19	A	24.2	29.8	631	763
hCV16190971	2	age_ge75=0	NONE	0.12776	0.13991	0.74861	0.00581	1.34	0.92: 1.94	1.08	0.67: 1.74	3.78	1.39: 10.30	A	31.5	25.6	157	121
hCV16190971	1+2	age_ge75=0	source	0.36878	0.39251	0.52601	0.00281	1.12	0.87: 1.45	0.9	0.65: 1.24	2.48	1.35: 4.58	A	28.6	26.2	334	267
hCV16190971	2	male=1	NONE	0.35683	0.34827	0.90958	0.01804	1.21	0.81: 1.81	0.97	0.57: 1.64	3.2	1.17: 8.75	A	32.5	28.5	97	130
hCV16190971	1+2	male=1	source	0.63111	0.63763	0.34646	0.00583	1.07	0.81: 1.41	0.84	0.59: 1.20	2.48	1.28: 4.81	A	28.5	27.5	237	264
hCV16221181	1	male=1	apoe4,age_ge75	0.00357	0.00295	0.00295		4.66	1.40: 15.53	4.94	1.46: 16.74			T	6.2	1.1	154	140
hCV16221181	1+3	male=1	source,apoe4,age_ge75	0.00775	0.00673	0.00673		2.99	1.25: 7.13	3.1	1.28: 7.48			T	5.2	1.5	231	239
hCV16248263	1	age_ge75=0	apoe4,male	0.53669	0.55579	0.64371	0.04105	1.14	0.75: 1.73	0.88	0.52: 1.50	3.25	1.14: 9.31	T	31.3	27.9	174	140
hCV16248263	1+2+3	age_ge75=0	source,apoe4,male	0.94421	0.94513	0.15644	0.00776	1.01	0.78: 1.30	0.79	0.57: 1.09	2.35	1.27: 4.34	T	29.5	28.1	419	398
hCV16248299	1	male=0	NONE	0.00065	0.00056	0.00129	0.03257	0.61	0.46: 0.81	0.55	0.38: 0.79	0.47	0.23: 0.95	G	23.8	33.9	250	220
hCV16248299	1+2+3	male=0	source	0.00412	0.00336	0.00783	0.05157	0.78	0.66: 0.93	0.75	0.60: 0.93	0.67	0.45: 1.00	G	27.5	32.8	632	766
hCV1651379	1	age_ge75=0	apoe4,male	0.00743	0.00807	0.02484	0.03724	1.63	1.14: 2.33	1.79	1.07: 2.97	2.01	1.02: 3.97	A	41.1	34.8	191	158
hCV1651379	1+2	age_ge75=0	source,apoe4,male	0.00979	0.01081	0.02201	0.06606	1.42	1.09: 1.85	1.56	1.06: 2.29	1.6	0.97: 2.63	A	42.0	37.1	308	284
hCV1687563	1	age_ge75=1	NONE	0.01754	0.01837	0.05609	0.02204	1.6	1.08: 2.37	1.52	0.99: 2.35	7.89	0.98: 63.62	G	16.3	10.8	227	217
hCV1687563	1+3	age_ge75=1	source	0.0208	0.02144	0.08719	0.00608	1.38	1.05: 1.81	1.3	0.96: 1.76	4.96	1.41: 17.45	G	15.1	11.4	449	463
hCV1687563	3	male=0	apoe4,age_ge75	0.03693	0.04021	0.05951	0.18711	1.49	1.03: 2.16	1.49	0.99: 2.27	2.33	0.63: 8.67	G	15.6	11.8	294	297
hCV1687563	1+3	male=0	source,apoe4,age_ge75	0.00955	0.01005	0.03352	0.01842	1.42	1.09: 1.84	1.38	1.02: 1.85	3.48	1.18: 10.21	G	16.1	12.7	558	531
hCV1780695	2	ALL	NONE	0.01239	0.0149	0.02959	0.06815	1.35	1.07: 1.72	1.52	1.04: 2.23	1.43	0.97: 2.11	A	51.7	44.1	207	401
hCV1780695	1+2+3	ALL	source	0.02512	0.02539	0.00851	0.3342	1.15	1.02: 1.30	1.3	1.07: 1.57	1.11	0.90: 1.36	A	49.5	45.8	983	1171
hCV1780695	2	male=1	NONE	0.05985	0.05955	0.03369	0.37516	1.41	0.99: 2.01	1.86	1.04: 3.30	1.32	0.72: 2.42	A	51.0	42.5	105	146
hCV1780695	1+2+3	male=1	source	0.04753	0.04366	0.00653	0.65965	1.24	1.00: 1.52	1.59	1.14: 2.23	1.08	0.76: 1.56	A	49.9	43.9	337	387
hCV1792848	1	ALL	NONE	0.07385	0.07132	0.41139	0.00842	1.23	0.98: 1.53	1.13	0.84: 1.51	2.04	1.19: 3.50	T	32.7	28.4	385	345
hCV1792848	1+2+3	ALL	source	0.02152	0.02023	0.15347	0.00539	1.17	1.02: 1.33	1.13	0.95: 1.35	1.56	1.14: 2.14	T	31.4	28.2	1011	1101
hCV1792856	1	ALL	apoe4,male,age_ge75	0.00014	0.00012	0.00058	0.0064	0.56	0.41: 0.75	0.55	0.39: 0.77	0.21	0.06: 0.67	G	14.2	20.4	388	353
hCV1792856	1+3	ALL	source,apoe4,male,age_ge75	0.00969	0.00846	0.01674	0.08644	0.77	0.63: 0.94	0.75	0.60: 0.95	0.58	0.30: 1.13	G	16.0	19.8	748	749
hCV1792856	1	male=1	apoe4,age_ge75	0.01453	0.01016	0.01189	0.33282	0.53	0.32: 0.88	0.48	0.27: 0.85	0.21	0.01: 4.02	G	13.9	20.0	140	135
hCV1792856	1+3	male=1	source,apoe4,age_ge75	0.01439	0.01018	0.00866	0.52945	0.62	0.43: 0.91	0.56	0.37: 0.86	0.57	0.11: 2.94	G	15.1	21.2	218	236
hCV1801156	1	male=1	apoe4,age_ge75	0.01949	0.02211	0.01836	0.40328	0.58	0.36: 0.92	0.53	0.31: 0.90	0.53	0.12: 2.32	G	13.3	21.4	154	140
hCV1801156	1+2+3	male=1	source,apoe4,age_ge75	0.00885	0.00897	0.01261	0.16858	0.66	0.48: 0.90	0.64	0.45: 0.91	0.42	0.13: 1.43	G	13.6	18.7	327	378
hCV1822206	1	age_ge75=1	NONE	0.03293	0.0338	0.04072	0.26676	0.69	0.49: 0.97	0.66	0.45: 0.98	0.56	0.20: 1.57	C	15.9	21.4	227	217
hCV1822206	1+2	age_ge75=1	source	0.00792	0.00828	0.0108	0.17382	0.68	0.51: 0.90	0.65	0.47: 0.91	0.53	0.21: 1.33	C	14.9	19.2	299	474
hCV1822261	1	male=0	NONE	0.47084	0.4661	0.94999	0.04101	1.11	0.83: 1.49	0.99	0.69: 1.41	2.35	1.01: 5.43	A	25.4	23.4	264	237
hCV1822261	1+3	male=0	source	0.20102	0.20717	0.80587	0.00824	1.14	0.93: 1.38	1.03	0.81: 1.31	1.95	1.18: 3.21	A	26.0	23.7	557	534
hCV1839324	1	age_ge75=0	NONE	0.12292	0.12657	0.60509	0.04495	0.79	0.59: 1.07	0.88	0.55: 1.42	0.61	0.37: 0.99	C	46.4	52.2	192	159
hCV1839324	1+3	age_ge75=0	source	0.03297	0.03532	0.36161	0.00912	0.79	0.63: 0.98	0.85	0.60: 1.20	0.61	0.42: 0.88	C	44.1	49.7	329	309
hCV1839328	1	age_ge75=0	NONE	0.10632	0.10923	0.58473	0.03578	0.78	0.58: 1.05	0.88	0.54: 1.41	0.59	0.36: 0.97	C	46.1	52.2	191	159
hCV1839328	1+3	age_ge75=0	source	0.02886	0.03128	0.36281	0.00699	0.78	0.63: 0.97	0.85	0.60: 1.20	0.6	0.41: 0.87	C	43.9	49.7	327	307
hCV1839328	1	male=1	apoe4,age_ge75	0.06715	0.07205	0.736	0.01024	0.73	0.52: 1.03	0.91	0.52: 1.60	0.49	0.28: 0.85	C	48.0	56.1	153	140
hCV1839328	1+3	male=1	source,apoe4,age_ge75	0.05044	0.04969	0.55136	0.00881	0.76	0.58: 1.00	0.87	0.56: 1.36	0.55	0.35: 0.87	C	47.0	52.7	230	239
hCV1839329	1	age_ge75=0	NONE	0.12267	0.12686	0.62968	0.04133	0.79	0.59: 1.07	0.89	0.55: 1.43	0.6	0.37: 0.98	G	46.4	52.2	192	158

WO 2006/022639 PCT/US2004/024001

TABLE 7, page 4 of 8

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV1839329	1+3	age_ge75=0	source	0.03317	0.03604	0.38839	0.00796	0.79	0.63: 0.98	0.86	0.61: 1.21	0.6	0.41: 0.88	G	43.9	49.5	330	308
hCV1845232	1	age_ge75=0	apoe4,male	0.04177	0.03716	0.09561	0.08272	0.66	0.45: 0.98	0.64	0.38: 1.09	0.46	0.20: 1.08	C	31.5	37.7	178	146
hCV1845232	1+3	age_ge75=0	source,apoe4,male	0.00742	0.00846	0.04887	0.013	0.69	0.52: 0.90	0.69	0.48: 1.00	0.47	0.26: 0.85	C	30.7	36.0	340	299
hCV1847915	2	male=0	apoe4,age_ge75	0.20025	0.20735	0.94906	0.0213	0.77	0.51: 1.15	0.98	0.55: 1.76	0.34	0.14: 0.85	G	35.0	42.6	90	249
hCV1847915	1+2	male=0	source,apoe4,age_ge75	0.01322	0.01453	0.16246	0.00474	0.74	0.59: 0.94	0.79	0.56: 1.10	0.51	0.32: 0.82	G	36.6	41.6	340	468
hCV1853469	1	apoe4=0	male,age_ge75	0.04652	0.03665	0.00951	0.94122	1.35	1.00: 1.83	1.7	1.13: 2.54	1.03	0.49: 2.14	G	35.5	29.2	162	271
hCV1853469	1+3	apoe4=0	source,male,age_ge75	0.07352	0.06584	0.00909	0.71251	1.21	0.98: 1.50	1.46	1.10: 1.95	0.91	0.56: 1.49	G	35.3	31.1	303	570
hCV1911230	3	age_ge75=0	apoe4,male	0.02321	0.02369	0.06782	0.04611	0.59	0.37: 0.93	0.59	0.34: 1.04	0.29	0.08: 1.03	C	21.7	26.0	138	150
hCV1911230	1+3	age_ge75=0	source,apoe4,male	0.05735	0.05792	0.31068	0.00604	0.75	0.56: 1.01	0.83	0.58: 1.19	0.36	0.16: 0.78	C	23.5	25.4	328	307
hCV1913066	1	ALL	NONE	0.07093	0.07069	0.02804	0.95118	1.23	0.98: 1.55	1.37	1.03: 1.82	1.02	0.57: 1.82	A	26.7	22.8	419	377
hCV1913066	1+2	ALL	source	0.03278	0.03259	0.00942	0.94764	1.2	1.02: 1.43	1.32	1.07: 1.63	1.01	0.66: 1.57	A	26.6	23.1	685	763
hCV1920609	2	apoe4=1	male,age_ge75	0.00601	0.00662	0.06107	0.00989	1.81	1.19: 2.78	1.92	0.97: 3.80	2.64	1.25: 5.55	A	55.9	41.9	118	86
hCV1920609	1+2+3	apoe4=1	source,male,age_ge75	0.02203	0.01852	0.36884	0.00426	1.28	1.04: 1.58	1.18	0.82: 1.69	1.7	1.18: 2.45	A	54.2	47.6	577	266
hCV1946182	2	age_ge75=0	apoe4,male	0.09351	0.10165	0.04227	0.98279	1.55	0.94: 2.56	1.9	1.02: 3.52	1.02	0.27: 3.83	G	25.0	21.6	114	118
hCV1946182	1+2	age_ge75=0	source,apoe4,male	0.02591	0.02402	0.0073	0.9492	1.45	1.05: 2.01	1.73	1.16: 2.58	1.03	0.42: 2.49	G	25.1	21.9	293	265
hCV2027467	1	apoe4=1	male,age_ge75	0.168	0.16484	0.61151	0.01095	0.75	0.49: 1.13	0.87	0.52: 1.47	0.28	0.10: 0.78	A	22.5	28.5	227	79
hCV2027467	1+2	apoe4=1	source,male,age_ge75	0.20555	0.20527	0.79174	0.00499	0.81	0.59: 1.12	0.95	0.64: 1.40	0.33	0.15: 0.75	A	22.0	25.0	345	172
hCV2028275	1	age_ge75=0	NONE	0.05679	0.06807	0.00225	0.6732	0.73	0.53: 1.01	0.49	0.30: 0.77	1.13	0.64: 2.01	T	36.7	44.1	173	145
hCV2028275	1+3	age_ge75=0	source	0.09669	0.10815	0.00078	0.21954	0.82	0.65: 1.04	0.56	0.40: 0.79	1.3	0.85: 1.98	T	38.1	42.9	307	294
hCV2028376	1	apoe4=1	NONE	0.19851	0.21114	0.00927	0.42667	0.79	0.55: 1.13	0.46	0.25: 0.83	1.31	0.68: 2.52	C	42.3	48.1	226	80
hCV2028376	1+3	apoe4=1	source	0.17748	0.17581	0.00832	0.49649	0.84	0.66: 1.08	0.58	0.39: 0.87	1.17	0.74: 1.84	C	43.1	47.4	445	175
hCV2116087	1	male=0	apoe4,age_ge75	0.0142	0.01349	0.0228	0.10662	1.45	1.08: 1.95	1.57	1.06: 2.30	1.82	0.91: 3.66	A	35.2	28.5	264	235
hCV2116087	1+2+3	male=0	source,apoe4,age_ge75	0.01046	0.00942	0.01245	0.12787	1.26	1.06: 1.50	1.35	1.07: 1.71	1.37	0.92: 2.05	A	35.4	30.7	670	787
hCV2116434	1	ALL	NONE	0.01843	0.02038	0.03809	0.0673	0.57	0.36: 0.91	0.6	0.37: 0.98		. : .	C	3.6	6.1	418	376
hCV2116434	1+2+3	ALL	source	0.00873	0.00867	0.01088	0.28375	0.68	0.50: 0.91	0.67	0.50: 0.91	0.32	0.03: 2.98	C	3.5	5.1	1071	1176
hCV2131920	1	apoe4=0	male,age_ge75	0.61813	0.59831	0.4903	0.01335	1.08	0.80: 1.45	0.87	0.59: 1.29	2.48	1.19: 5.14	A	33.2	30.6	164	273
hCV2131920	1+2	apoe4=0	source,male,age_ge75	0.50318	0.48742	0.55275	0.00882	1.09	0.85: 1.39	0.91	0.65: 1.26	2.09	1.20: 3.64	A	32.9	31.2	225	549
hCV2153267	1	male=0	NONE	0.02022	0.02464	0.00632	0.77118	1.43	1.06: 1.92	1.68	1.16: 2.43	1.11	0.56: 2.19	C	28.5	21.9	247	217
hCV2153267	1+3	male=0	source	0.01599	0.01641	0.00544	0.66465	1.27	1.05: 1.55	1.42	1.11: 1.81	1.11	0.69: 1.81	C	28.1	23.5	528	512
hCV2170733	2	male=1	apoe4,age_ge75	0.01526	0.01905	0.00577	0.61808	0.42	0.20: 0.87	0.34	0.15: 0.77	2.38	0.11: 50.32	C	6.6	8.9	137	141
hCV2170733	1+2+3	male=1	source,apoe4,age_ge75	0.009	0.00999	0.00482	0.83353	0.56	0.36: 0.88	0.52	0.32: 0.83	1.31	0.14: 12.18	C	5.9	7.7	380	382
hCV2264708	1	male=1	apoe4,age_ge75	0.00412	0.00369	0.01449	0.01585	1.96	1.24: 3.12	1.97	1.14: 3.40	11.3	1.15:110.85	T	23.6	15.6	140	135
hCV2264708	1+2+3	male=1	source,apoe4,age_ge75	0.00197	0.00202	0.00315	0.08663	1.57	1.18: 2.09	1.67	1.19: 2.35	2.11	0.93: 4.76	T	23.9	19.1	314	377
hCV2302732	1	age_ge75=0	NONE	0.00002	0.00003	0.00002	0.03538	0.48	0.34: 0.68	0.4	0.26: 0.61	0.43	0.19: 0.96	C	20.6	34.9	192	159
hCV2302732	1+2+3	age_ge75=0	source	0.00814	0.00805	0.00166	0.70703	0.76	0.61: 0.93	0.66	0.50: 0.85	0.91	0.55: 1.50	C	24.6	30.2	449	440
hCV2302732	1	ALL	NONE	0.00003	0.00004	0.00009	0.01106	0.62	0.50: 0.78	0.57	0.43: 0.76	0.51	0.30: 0.86	C	22.7	32.0	419	376
hCV2302732	1+2+3	ALL	source	0.00155	0.00158	0.00211	0.08162	0.8	0.70: 0.92	0.76	0.64: 0.91	0.75	0.54: 1.04	C	24.8	29.2	988	1170
hCV2302732	1	apoe4=0	male,age_ge75	0.01585	0.01845	0.01087	0.34103	0.69	0.51: 0.93	0.61	0.42: 0.89	0.72	0.37: 1.41	C	24.7	33.3	178	287
hCV2302732	1+2+3	apoe4=0	source,male,age_ge75	0.01429	0.01529	0.00892	0.36077	0.78	0.64: 0.95	0.72	0.56: 0.92	0.81	0.51: 1.28	C	24.8	30.3	387	888
hCV2302732	1	male=0	NONE	0.00023	0.00026	0.00022	0.08057	0.59	0.44: 0.78	0.51	0.36: 0.73	0.54	0.27: 1.09	C	21.3	31.6	265	236
hCV2302732	1+2+3	male=0	source	0.00105	0.00089	0.0005	0.20254	0.75	0.64: 0.89	0.69	0.55: 0.85	0.76	0.49: 1.16	C	24.1	29.6	650	785
hCV2302737	1	ALL	apoe4,male,age_ge75	0.0041	0.00664	0.02948	0.0178	1.39	1.11: 1.74	1.44	1.04: 1.99	1.59	1.07: 2.36	T	45.2	36.3	388	353
hCV2302737	1+2+3	ALL	source,apoe4,male,age_ge75	0.00682	0.0083	0.02982	0.02874	1.2	1.05: 1.37	1.25	1.02: 1.52	1.29	1.02: 1.64	T	45.0	40.6	1023	1151
hCV2302737	1	apoe4=0	male,age_ge75	0.00305	0.00624	0.03316	0.01187	1.53	1.15: 2.04	1.56	1.04: 2.34	1.83	1.13: 2.96	T	45.4	34.9	162	272
hCV2302737	1+2+3	apoe4=0	source,male,age_ge75	0.00516	0.00657	0.09401	0.00337	1.29	1.08: 1.54	1.26	0.96: 1.64	1.57	1.16: 2.13	T	46.5	40.5	370	872
hCV2302737	1	male=1	NONE	0.0329	0.04204	0.06925	0.12402	1.45	1.03: 2.04	1.58	0.96: 2.60	1.6	0.88: 2.92	T	46.4	37.4	139	135
hCV2302737	1+2+3	male=1	source	0.00642	0.00882	0.01304	0.07277	1.33	1.08: 1.63	1.46	1.08: 1.98	1.39	0.97: 1.99	T	46.1	39.5	375	384
hCV2539346	3	ALL	NONE	0.00396	0.00358	0.00649	0.05658	0.74	0.60: 0.91	0.66	0.49: 0.89	0.67	0.44: 1.01	T	35.1	42.3	358	396
hCV2539346	1+2+3	ALL	source	0.00097	0.00096	0.00253	0.02271	0.81	0.71: 0.92	0.75	0.63: 0.91	0.75	0.59: 0.96	T	36.8	41.9	929	1097
hCV2539346	3	male=0	NONE	0.00475	0.00472	0.02174	0.0193	0.71	0.56: 0.90	0.67	0.48: 0.94	0.57	0.35: 0.92	T	34.1	42.2	280	295
hCV2539346	1+2+3	male=0	source	0.00562	0.0054	0.03152	0.01439	0.8	0.68: 0.94	0.78	0.62: 0.98	0.67	0.49: 0.92	T	35.6	40.4	617	736

TABLE 7, page 5 of 8

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV25602413	2	apoe4=1	male,age_ge75	0.15084	0.16251	0.44633	0.04377	1.4	0.88: 2.22	1.24	0.71: 2.15	4.03	0.99: 16.38	A	24.5	19.1	157	97
hCV25602413	1+2+3	apoe4=1	source,male,age_ge75	0.1272	0.13074	0.56564	0.00659	1.21	0.95: 1.55	1.09	0.81: 1.47	2.93	1.31: 6.54	A	24.2	21.0	631	279
hCV25602413	1	male=0	apoe4,age_ge75	0.0172	0.0196	0.08469	0.01923	1.45	1.07: 1.98	1.4	0.95: 2.07	2.5	1.14: 5.48	A	28.8	23.1	264	234
hCV25602413	1+2+3	male=0	source,apoe4,age_ge75	0.00357	0.00371	0.01875	0.0116	1.33	1.10: 1.61	1.33	1.05: 1.69	1.91	1.16: 3.13	A	27.4	24.0	636	776
hCV25602413	2	male=1	apoe4,age_ge75	0.70841	0.71862	0.46547	0.01034	1.09	0.69: 1.71	0.82	0.48: 1.40	5.39	1.34: 21.72	A	19.9	19.3	138	145
hCV25602413	1+2+3	male=1	source,apoe4,age_ge75	0.30802	0.3005	0.9401	0.00563	1.14	0.88: 1.48	1.01	0.74: 1.38	2.93	1.31: 6.58	A	22.6	21.2	382	387
hCV25603905	1	ALL	apoe4,male,age_ge75	0.00418	0.00633	0.02505	0.02072	1.39	1.11: 1.74	1.45	1.05: 2.01	1.59	1.06: 2.37	C	44.3	35.6	386	350
hCV25603905	1+2+3	ALL	source,apoe4,male,age_ge75	0.02173	0.02508	0.21667	0.00968	1.18	1.02: 1.36	1.14	0.93: 1.40	1.37	1.07: 1.76	C	44.3	40.3	913	1095
hCV25603906	1	age_ge75=0	NONE	0.15083	0.16404	0.88053	0.00546	1.28	0.91: 1.80	1.03	0.67: 1.60	3.06	1.35: 6.97	T	33.0	27.7	179	146
hCV25603906	1+2+3	age_ge75=0	source	0.50679	0.51083	0.54924	0.00958	1.07	0.87: 1.33	0.92	0.70: 1.21	1.94	1.17: 3.22	T	29.4	27.7	432	422
hCV25625639	1	age_ge75=0	NONE	0.00562	0.0053	0.00408	0.18717	0.63	0.45: 0.87	0.52	0.33: 0.82	0.61	0.29: 1.28	A	27.2	37.4	178	147
hCV25625639	1+2+3	age_ge75=0	source	0.00857	0.00814	0.00974	0.14947	0.75	0.61: 0.93	0.7	0.53: 0.92	0.69	0.42: 1.14	A	26.1	31.7	424	404
hCV25637868	3	ALL	NONE	0.04981	0.04782	0.05834	0.3295	1.35	1.00: 1.82	1.37	0.99: 1.91	1.83	0.53: 6.31	C	14.3	11.0	384	399
hCV25637868	1+2+3	ALL	source	0.00969	0.00925	0.01777	0.09004	1.25	1.06: 1.49	1.26	1.04: 1.53	1.74	0.92: 3.29	C	16.0	13.1	1020	1113
hCV25744917	1	apoe4=1	male,age_ge75	0.00547	0.00525	0.11518	0.00249	0.59	0.41: 0.86	0.6	0.32: 1.13	0.41	0.23: 0.74	A	44.6	57.0	223	79
hCV25744917	1+3	apoe4=1	source,male,age_ge75	0.00373	0.0046	0.02357	0.0163	0.69	0.53: 0.89	0.62	0.41: 0.94	0.61	0.41: 0.92	A	44.9	54.3	440	174
hCV25752440	1	age_ge75=0	NONE	0.0001	0.00014	0.00036	0.01329	0.51	0.36: 0.71	0.45	0.29: 0.70	0.38	0.17: 0.84	A	22.5	36.4	178	147
hCV25752440	1+2+3	age_ge75=0	source	0.00966	0.00962	0.00461	0.37493	0.76	0.62: 0.94	0.68	0.52: 0.89	0.81	0.50: 1.30	A	26.3	32.0	432	422
hCV25752440	1	ALL	NONE	0.0001	0.00014	0.00091	0.00399	0.64	0.51: 0.80	0.61	0.46: 0.82	0.47	0.27: 0.79	A	23.8	32.9	390	354
hCV25752440	1+2+3	ALL	source	0.0011	0.00113	0.00464	0.01486	0.8	0.69: 0.91	0.78	0.65: 0.93	0.67	0.48: 0.93	A	26.0	30.7	953	1140
hCV25752440	1	apoe4=0	male,age_ge75	0.00671	0.00801	0.00632	0.1905	0.65	0.47: 0.89	0.58	0.39: 0.86	0.63	0.31: 1.26	A	24.5	34.6	163	272
hCV25752440	1+2+3	apoe4=0	source,male,age_ge75	0.00419	0.00481	0.00417	0.15559	0.75	0.61: 0.91	0.69	0.54: 0.89	0.72	0.45: 1.14	A	25.1	31.7	371	870
hCV25752440	1	male=0	NONE	0.00032	0.00032	0.00065	0.0314	0.59	0.44: 0.79	0.53	0.37: 0.76	0.47	0.23: 0.95	A	22.8	33.3	250.0	219
hCV25752440	1+2+3	male=0	source	0.00106	0.00088	0.00146	0.05786	0.75	0.64: 0.89	0.71	0.57: 0.87	0.67	0.44: 1.02	A	25.475	31.3	632.0	763
hCV25766586	2	apoe4=0	male,age_ge75	0.01479	0.0173	0.06066	0.01975	1.79	1.12: 2.86	1.71	0.98: 2.98	4.08	1.21: 13.68	T	22.4	14.9	85	281
hCV25766586	1+2+3	apoe4=0	source,male,age_ge75	0.46484	0.46308	0.91904	0.00473	1.09	0.86: 1.39	0.99	0.75: 1.30	3.04	1.38: 6.72	T	15.8	14.9	396	854
hCV25766586	1	apoe4=1	male,age_ge75	0.01532	0.01478	0.0094	0.69778	0.56	0.34: 0.90	0.48	0.27: 0.84	0.72	0.13: 3.94	T	13.0	20.0	227	80
hCV25766586	1+2+3	apoe4=1	source,male,age_ge75	0.02275	0.02256	0.00838	0.86253	0.72	0.55: 0.96	0.65	0.47: 0.90	1.08	0.42: 2.80	T	13.7	18.4	606	263
hCV25923332	1	male=1	apoe4,age_ge75	0.00054	0.00065	0.00052	0.15627	2.29	1.42: 3.70	2.65	1.52: 4.62	2.6	0.65: 10.42	G	25.4	12.2	140	135
hCV25923332	1+3	male=1	source,apoe4,age_ge75	0.00629	0.00647	0.00112	0.89169	1.64	1.15: 2.35	2.01	1.32: 3.07	0.93	0.33: 2.60	G	23.7	15.5	213	235
hCV25938519	1	male=1	apoe4,age_ge75	0.00605	0.00768	0.00454	0.32979	1.8	1.17: 2.76	2.13	1.25: 3.63	1.6	0.58: 4.39	T	25.7	18.7	138	134
hCV25938519	1+3	male=1	source,apoe4,age_ge75	0.00719	0.00968	0.01696	0.08349	1.54	1.12: 2.11	1.63	1.09: 2.43	1.76	0.88: 3.49	T	28.7	22.1	216	235
hCV25970515	1	apoe4=0	NONE	0.00099	0.00099	0.00506	0.00755	1.8	1.26: 2.55	1.8	1.19: 2.71	4.39	1.35: 14.23	T	23.6	14.7	161	269
hCV25970515	1+2+3	apoe4=0	source	0.00619	0.00596	0.00658	0.23809	1.37	1.09: 1.73	1.44	1.11: 1.87	1.56	0.75: 3.25	T	20.0	15.1	368	845
hCV25992569	1	ALL	NONE	0.00449	0.00499	0.00612	0.0996	0.73	0.59: 0.91	0.67	0.50: 0.89	0.68	0.43: 1.08	G	29.6	36.5	389	352
hCV25992569	1+3	ALL	source	0.00667	0.00681	0.03289	0.01743	0.81	0.70: 0.94	0.8	0.66: 0.98	0.67	0.48: 0.93	G	30.2	34.8	771	750
hCV2655148	1	age_ge75=0	NONE	0.19447	0.21395	0.93571	0.00465	1.25	0.89: 1.76	0.98	0.63: 1.52	3.13	1.37: 7.11	C	32.1	27.4	176	146
hCV2655148	1+2+3	age_ge75=0	source	0.60906	0.61403	0.41954	0.0091	1.06	0.86: 1.31	0.89	0.68: 1.17	1.95	1.18: 3.23	C	29.0	27.7	430	421
hCV2655158	1	ALL	apoe4,male,age_ge75	0.00413	0.00813	0.00254	0.39118	1.48	1.13: 1.93	1.66	1.19: 2.31	1.28	0.72: 2.27	T	25.7	19.5	382	344
hCV2655158	1+2+3	ALL	source,apoe4,male,age_ge75	0.00967	0.01268	0.00915	0.27564	1.23	1.05: 1.44	1.29	1.07: 1.57	1.22	0.85: 1.75	T	24.9	22.9	1017	1142
hCV2655167	1	male=1	apoe4,age_ge75	0.13624	0.13751	0.65137	0.00576	1.34	0.91: 1.97	1.12	0.69: 1.82	3.54	1.28: 9.80	G	29.2	23.6	154	140
hCV2655167	1+2+3	male=1	source,apoe4,age_ge75	0.13637	0.13812	0.69134	0.00523	1.21	0.94: 1.55	1.07	0.78: 1.47	2.35	1.28: 4.33	G	29.4	25.1	332	385
hCV2682758	3	ALL	apoe4,male,age_ge75	0.03315	0.02986	0.00902	0.64938	0.78	0.62: 0.98	0.66	0.48: 0.90	0.9	0.55: 1.46	T	32.1	36.1	358	395
hCV2682758	1+3	ALL	source,apoe4,male,age_ge75	0.03344	0.03071	0.0039	0.98247	0.84	0.71: 0.99	0.72	0.58: 0.90	1	0.71: 1.40	T	33.2	36.1	749	749
hCV2682758	3	apoe4=0	male,age_ge75	0.13254	0.11721	0.01544	0.50973	0.79	0.58: 1.08	0.61	0.40: 0.91	1.24	0.64: 2.41	T	30.1	34.8	141	300
hCV2682758	1+3	apoe4=0	source,male,age_ge75	0.09873	0.09142	0.00532	0.38834	0.84	0.68: 1.03	0.67	0.50: 0.89	1.21	0.78: 1.89	T	32.0	35.5	305	573
hCV2682758	3	male=0	apoe4,age_ge75	0.01491	0.012	0.00659	0.33666	0.72	0.55: 0.94	0.61	0.42: 0.87	0.76	0.43: 1.35	T	31.2	36.3	293	296
hCV2682758	1+3	male=0	source,apoe4,age_ge75	0.03123	0.02838	0.00932	0.60068	0.81	0.66: 0.98	0.7	0.54: 0.92	0.9	0.59: 1.36	T	32.7	36.1	544	515
hCV2685860	3	age_ge75=0	NONE	0.01755	0.01325	0.01325		2.13	1.13: 4.01	2.26	1.17: 4.37			A	9.9	4.9	162	153
hCV2685860	1+2+3	age_ge75=0	source	0.00902	0.00878	0.01112	0.26834	1.59	1.12: 2.26	1.61	1.11: 2.32	3.22	0.36: 28.82	A	9.1	5.9	518	441
hCV2734178	1	apoe4=0	male,age_ge75	0.01102	0.01134	0.39295	0.0013	1.42	1.08: 1.86	1.22	0.77: 1.93	1.96	1.29: 2.98	G	58.0	48.8	176	288

TABLE 7, page 6 of 8

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV2734178	1+2+3	apoe4=0	source,male,age_ge75	0.00347	0.00326	0.09596	0.00164	1.3	1.09: 1.55	1.28	0.96: 1.71	1.56	1.18: 2.07	G	53.8	47.2	385	891
hCV2757616	1	age_ge75=1	NONE	0.02518	0.02728	0.01005	0.88518	1.47	1.05: 2.05	1.68	1.13: 2.49	1.07	0.43: 2.69	G	22.5	16.5	227	218
hCV2757616	1+2+3	age_ge75=1	source	0.00976	0.01052	0.00366	0.80674	1.32	1.07: 1.63	1.44	1.13: 1.85	1.08	0.58: 2.01	G	20.6	16.4	520	721
hCV2760432	1	male=0	NONE	0.0061	0.01022	0.00855	0.31982	2.05	1.22: 3.44	2.1	1.20: 3.69	2.26	0.43: 11.76	C	9.1	4.6	265	237
hCV2760432	1+2	male=0	source	0.00773	0.01075	0.00536	0.76817	1.67	1.15: 2.43	1.78	1.19: 2.68	1.24	0.32: 4.78	C	9.3	6.4	365	488
hCV286937	1	age_ge75=0	NONE	0.00121	0.00189	0.001	0.24138	0.53	0.36: 0.78	0.46	0.29: 0.73	0.56	0.21: 1.50	G	15.2	25.3	178	146
hCV286937	1+2+3	age_ge75=0	source	0.00067	0.00085	0.00065	0.20458	0.63	0.49: 0.83	0.59	0.44: 0.80	0.6	0.28: 1.32	G	13.5	19.4	422	403
hCV286937	1	apoe4=0	NONE	0.01556	0.01669	0.01962	0.22477	0.64	0.44: 0.92	0.6	0.40: 0.92	0.5	0.16: 1.56	G	14.8	21.5	162	270
hCV286937	1+2+3	apoe4=0	source	0.00908	0.00988	0.01998	0.07722	0.72	0.56: 0.92	0.72	0.55: 0.95	0.47	0.20: 1.10	G	14.0	17.9	368	838
hCV286937	2	apoe4=1	NONE	0.05087	0.04567	0.03687	0.67331	0.58	0.34: 1.01	0.52	0.28: 0.96	0.65	0.09: 4.75	G	13.5	21.1	115	76
hCV286937	1+2+3	apoe4=1	source	0.01009	0.00993	0.00742	0.52227	0.68	0.51: 0.91	0.64	0.46: 0.89	0.71	0.25: 2.01	G	12.5	17.5	561	251
hCV2945715	1	male=0	apoe4,age_ge75	0.09244	0.09308	0.00862	0.6672	1.27	0.96: 1.69	1.69	1.14: 2.51	0.88	0.49: 1.56	T	39.0	33.8	264	235
hCV2945715	1+2	male=0	source,apoe4,age_ge75	0.0779	0.07825	0.00932	0.79126	1.23	0.98: 1.55	1.53	1.11: 2.11	0.94	0.58: 1.51	T	38.2	33.3	356	483
hCV29522	1	ALL	apoe4,male,age_ge75	0.07834	0.08509	0.41078	0.0169	1.24	0.98: 1.57	1.14	0.83: 1.56	1.9	1.12: 3.20	T	34.4	29.5	390	352
hCV29522	1+2+3	ALL	source,apoe4,male,age_ge75	0.11891	0.12106	0.68883	0.00639	1.13	0.97: 1.31	1.04	0.86: 1.27	1.58	1.14: 2.19	T	33.3	30.7	919	1097
hCV29522	1	male=0	apoe4,age_ge75	0.26375	0.27263	0.89146	0.03483	1.19	0.88: 1.60	1.03	0.69: 1.54	2.06	1.05: 4.02	T	34.6	31.2	250	218
hCV29522	1+2+3	male=0	source,apoe4,age_ge75	0.0642	0.06616	0.39817	0.00928	1.19	0.99: 1.42	1.11	0.87: 1.41	1.71	1.14: 2.55	T	33.9	31.2	645	744
hCV2981213	1	age_ge75=0	apoe4,male	0.5145	0.53341	0.68569	0.04727	1.14	0.77: 1.70	0.9	0.54: 1.50	2.72	1.05: 7.03	T	32.8	28.5	174	144
hCV2981213	1+2+3	age_ge75=0	source,apoe4,male	0.83883	0.84134	0.21252	0.00852	1.03	0.80: 1.31	0.82	0.59: 1.12	2.23	1.23: 4.04	T	30.2	28.4	419	402
hCV2981216	1	age_ge75=0	apoe4,male	0.42276	0.44228	0.70732	0.02489	1.18	0.79: 1.75	0.91	0.54: 1.51	3.02	1.17: 7.80	T	33.1	28.8	178	146
hCV2981216	1+2+3	age_ge75=0	source,apoe4,male	0.8273	0.82995	0.19738	0.00594	1.03	0.81: 1.31	0.81	0.60: 1.11	2.27	1.27: 4.07	T	29.9	28.3	430	420
hCV299325	1	age_ge75=0	apoe4,male	0.01196	0.01194	0.0115	0.60505	3.01	1.30: 7.00	3.2	1.32: 7.75		. : .	T	6.0	3.8	191	158
hCV299325	1+2+3	age_ge75=0	source,apoe4,male	0.006	0.00695	0.00292	0.67541	1.96	1.20: 3.18	2.16	1.29: 3.60	0.55	0.05: 6.54	T	6.6	4.1	445	434
hCV3027361	1	age_ge75=0	apoe4,male	0.05995	0.06238	0.02149	0.6572	0.71	0.50: 1.01	0.56	0.34: 0.92	0.85	0.42: 1.72	T	33.2	40.8	191	157
hCV3027361	1+3	age_ge75=0	source,apoe4,male	0.05501	0.05498	0.00863	0.93053	0.77	0.59: 1.01	0.61	0.42: 0.88	0.98	0.58: 1.65	T	35.3	40.2	329	306
hCV3052366	1	ALL	NONE	0.01512	0.0216	0.0533	0.03231	1.91	1.12: 3.26	1.72	0.99: 2.99		. : .	T	5.5	3.0	390	355
hCV3052366	1+2+3	ALL	source	0.00531	0.00649	0.01841	0.01221	1.61	1.15: 2.26	1.52	1.07: 2.16		. : .	T	4.4	2.7	951	1128
hCV3052366	1	male=0	NONE	0.00429	0.00902	0.02136	0.03496	2.67	1.33: 5.36	2.3	1.11: 4.75		. : .	T	6.4	2.5	250	220
hCV3052366	1+2+3	male=0	source	0.00824	0.01067	0.02452	0.03515	1.76	1.16: 2.68	1.64	1.07: 2.54		. : .	T	4.7	2.7	631	754
hCV3132900	1	male=0	apoe4,age_ge75	0.00974	0.00764	0.03196	0.02864	0.69	0.52: 0.91	0.65	0.43: 0.97	0.54	0.31: 0.95	A	35.8	43.0	264	235
hCV3132900	1+2+3	male=0	source,apoe4,age_ge75	0.00397	0.00412	0.02085	0.01601	0.78	0.65: 0.92	0.75	0.59: 0.96	0.67	0.48: 0.94	A	36.1	40.3	635	778
hCV3159528	1	apoe4=0	male,age_ge75	0.05536	0.06031	0.60992	0.00403	1.31	0.99: 1.72	1.11	0.75: 1.65	2.07	1.26: 3.40	C	43.5	37.8	178	288
hCV3159528	1+2+3	apoe4=0	source,male,age_ge75	0.011	0.01067	0.16102	0.00249	1.26	1.05: 1.51	1.2	0.93: 1.56	1.67	1.20: 2.33	C	42.2	36.8	386	888
hCV3159528	1	male=1	apoe4,age_ge75	0.09702	0.0987	0.9565	0.00189	1.35	0.95: 1.93	1.01	0.61: 1.68	3.08	1.48: 6.42	C	40.8	35.0	153	140
hCV3159528	1+2+3	male=1	source,apoe4,age_ge75	0.009	0.00699	0.0768	0.00541	1.36	1.08: 1.72	1.36	0.97: 1.90	1.94	1.21: 3.12	C	41.0	35.8	327	383
hCV3159529	1	male=1	apoe4,age_ge75	0.08372	0.09185	0.91238	0.00036	1.4	0.95: 2.07	1.03	0.62: 1.70	4.95	1.80: 13.63	G	32.4	25.9	139	135
hCV3159529	1+2+3	male=1	source,apoe4,age_ge75	0.02684	0.02262	0.14832	0.00764	1.33	1.03: 1.71	1.28	0.92: 1.78	2.27	1.21: 4.23	G	32.2	27.7	306	361
hCV3159576	1	apoe4=1	NONE	0.02939	0.03304	0.02799	0.19725	0.67	0.47: 0.96	0.5	0.27: 0.93	0.69	0.39: 1.22	T	45.6	55.6	227	81
hCV3159576	1+2+3	apoe4=1	source	0.01565	0.0155	0.00407	0.29291	0.77	0.62: 0.95	0.59	0.41: 0.85	0.83	0.59: 1.17	T	47.1	53.4	563	253
hCV3178540	1	apoe4=0	male,age_ge75	0.04353	0.04326	0.00546	0.42381	1.39	1.01: 1.90	1.72	1.17: 2.54	0.67	0.26: 1.74	G	25.8	20.3	178	288
hCV3178540	1+2+3	apoe4=0	source,male,age_ge75	0.01386	0.0157	0.00148	0.62318	1.3	1.05: 1.60	1.5	1.17: 1.93	0.87	0.49: 1.53	G	24.8	20.5	387	882
hCV3188402	3	age_ge75=1	apoe4,male	0.38669	0.38885	0.97514	0.00227	0.84	0.56: 1.25	0.99	0.64: 1.55	0	. : .	C	12.2	13.4	222	246
hCV3188402	1+3	age_ge75=1	source,apoe4,male	0.2435	0.24068	0.57765	0.00685	0.83	0.61: 1.13	0.91	0.65: 1.27	0.13	0.02: 0.87	C	10.8	12.6	434	453
hCV3234889	1	apoe4=1	male,age_ge75	0.00916	0.00932	0.00473	0.36097	0.61	0.42: 0.89	0.48	0.29: 0.80	0.68	0.29: 1.57	G	25.8	36.2	240	87
hCV3234889	1+2	apoe4=1	source,male,age_ge75	0.04116	0.04389	0.0075	0.9222	0.74	0.55: 0.99	0.59	0.40: 0.87	1.03	0.53: 2.02	G	27.4	32.6	354	170
hCV337151	1	apoe4=1	NONE	0.00821	0.00892	0.10503	0.00343	0.61	0.42: 0.88	0.64	0.38: 1.10	0.37	0.19: 0.73	G	32.6	44.3	224	79
hCV337151	1+2+3	apoe4=1	source	0.00702	0.00664	0.0314	0.02086	0.74	0.60: 0.92	0.71	0.52: 0.97	0.62	0.41: 0.93	G	34.7	42.4	560	250
hCV337151	3	male=0	NONE	0.00753	0.00777	0.03237	0.02494	0.72	0.57: 0.92	0.69	0.49: 0.97	0.58	0.36: 0.94	G	34.2	41.9	282	295
hCV337151	1+2+3	male=0	source	0.00687	0.00661	0.04568	0.0114	0.8	0.69: 0.94	0.79	0.63: 1.00	0.66	0.48: 0.91	G	35.5	40.3	619	733
hCV368390	1	age_ge75=1	apoe4,male	0.00014	0.00023	0.00254	0.00069	2.29	1.48: 3.53	2.07	1.29: 3.33		. : .	C	17.0	8.5	227	217
hCV368390	1+2+3	age_ge75=1	source,apoe4,male	0.00109	0.00129	0.00602	0.00659	1.56	1.20: 2.04	1.51	1.13: 2.03	4.59	1.50: 13.99	C	14.4	10.5	519	731

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV368390	1	ALL	apoe4,male,age_ge75	0.00258	0.00334	0.00767	0.04556	1.64	1.18: 2.27	1.63	1.14: 2.35	2.74	0.93: 8.05	C	15.0	10.5	418	375
hCV368390	1+2+3	ALL	source,apoe4,male,age_ge75	0.00117	0.00134	0.00209	0.10273	1.41	1.14: 1.73	1.43	1.14: 1.80	1.83	0.87: 3.83	C	13.7	11.2	963	1163
hCV368390	3	male=0	apoe4,age_ge75	0.02153	0.02055	0.02374	0.3355	1.6	1.07: 2.40	1.66	1.07: 2.58	2.93	0.39: 22.21	C	13.5	10.0	281	295
hCV368390	1+2+3	male=0	source,apoe4,age_ge75	0.00421	0.00416	0.01282	0.02082	1.46	1.13: 1.89	1.44	1.08: 1.91	4.04	1.25: 13.00	C	13.8	11.2	635	780
hCV369380	1	age_ge75=0	NONE	0.0111	0.01153	0.01464	0.27042	0.39	0.19: 0.83	0.4	0.18: 0.85		. : .	C	3.1	7.5	178	147
hCV369380	1+3	age_ge75=0	source	0.00746	0.00715	0.00908	0.27116	0.48	0.28: 0.83	0.48	0.27: 0.84	0	. : .	C	3.1	6.2	340	300
hCV472673	1	apoe4=0	male,age_ge75	0.0218	0.01554	0.00652	0.28757	1.39	1.05: 1.84	1.93	1.20: 3.12	1.31	0.80: 2.17	C	51.2	42.3	163	273
hCV472673	1+2+3	apoe4=0	source,male,age_ge75	0.03476	0.02951	0.00489	0.53431	1.21	1.01: 1.44	1.53	1.13: 2.05	1.1	0.81: 1.50	C	50.4	45.9	371	873
hCV52509	1	age_ge75=0	apoe4,male	0.00813	0.00741	0.02812	0.03083	1.64	1.14: 2.38	1.82	1.05: 3.15	2.11	1.07: 4.17	T	49.2	39.7	177	146
hCV52509	1+2+3	age_ge75=0	source,apoe4,male	0.0096	0.01003	0.02106	0.05947	1.34	1.07: 1.67	1.5	1.06: 2.11	1.42	0.98: 2.07	T	49.1	43.0	428	414
hCV5478	1	age_ge75=0	NONE	0.01421	0.0175	0.0216	0.27042	0.32	0.12: 0.83	0.33	0.12: 0.89		. : .	T	1.7	5.1	178	147
hCV5478	1+3	age_ge75=0	source	0.00312	0.00457	0.00658	0.14473	0.34	0.16: 0.72	0.36	0.17: 0.77	0	. : .	T	1.6	4.4	316	296
hCV7432717	1	apoe4=1	male,age_ge75	0.00861	0.00885	0.00432	0.36515	0.61	0.42: 0.88	0.48	0.29: 0.80	0.68	0.30: 1.58	A	25.7	36.2	239	87
hCV7432717	1+2	apoe4=1	source,male,age_ge75	0.03092	0.03241	0.00577	0.99754	0.73	0.54: 0.97	0.58	0.39: 0.85	1	0.51: 1.97	A	27.1	32.7	356	173
hCV7582334	1	age_ge75=0	NONE	0.00216	0.00164	0.00464	0.02638	0.49	0.31: 0.78	0.49	0.29: 0.80		. : .	G	9.8	18.0	179	147
hCV7582334	1+2+3	age_ge75=0	source	0.00927	0.00864	0.01553	0.10485	0.7	0.54: 0.92	0.69	0.52: 0.93	0.42	0.14: 1.24	G	11.7	15.9	498	421
hCV7582334	1	apoe4=1	male,age_ge75	0.02109	0.0173	0.01324	0.81083	0.57	0.35: 0.93	0.51	0.29: 0.88	0.78	0.08: 7.38	G	10.8	18.5	227	81
hCV7582334	1+2+3	apoe4=1	source,male,age_ge75	0.00575	0.00476	0.00358	0.66224	0.65	0.48: 0.89	0.61	0.43: 0.85	0.73	0.18: 3.06	G	10.8	15.5	556	258
hCV7584409	1	age_ge75=0	apoe4,male	0.00713	0.00881	0.00933	0.12445	0.59	0.40: 0.87	0.51	0.30: 0.85	0.53	0.24: 1.16	T	30.9	40.1	178	146
hCV7584409	1+3	age_ge75=0	source,apoe4,male	0.00869	0.01066	0.0199	0.07228	0.69	0.52: 0.91	0.64	0.44: 0.93	0.59	0.34: 1.04	T	32.0	38.1	316	295
hCV799520	1	age_ge75=1	apoe4,male	0.09533	0.08651	0.03478	0.20152	1.42	0.93: 2.18	1.64	1.03: 2.62	0.31	0.04: 2.51	G	14.6	10.0	212	204
hCV799520	1+3	age_ge75=1	source,apoe4,male	0.04383	0.04506	0.00512	0.04207	1.36	1.00: 1.85	1.62	1.15: 2.27	0.31	0.09: 1.10	G	13.3	9.7	432	448
hCV811329	1	apoe4=0	male,age_ge75	0.01146	0.01077	0.01067	0.30277	0.64	0.45: 0.91	0.59	0.39: 0.89	0.56	0.18: 1.75	A	14.9	21.1	178	287
hCV811329	1+2+3	apoe4=0	source,male,age_ge75	0.0062	0.00551	0.01772	0.02859	0.73	0.58: 0.92	0.73	0.56: 0.95	0.41	0.18: 0.93	A	16.7	21.4	383	877
hCV811329	2	male=1	apoe4,age_ge75	0.00012	0.00011	0.00034	0.0179	0.4	0.25: 0.65	0.37	0.21: 0.65	0.05	0.00: 0.89	A	13.5	26.6	137	141
hCV811329	1+2+3	male=1	source,apoe4,age_ge75	0.00271	0.00223	0.00543	0.04946	0.66	0.51: 0.87	0.64	0.47: 0.88	0.38	0.14: 1.00	A	17.1	23.6	381	383
hCV8227677	1	apoe4=0	male,age_ge75	0.06598	0.06937	0.92784	0.00203	1.3	0.98: 1.72	0.98	0.62: 1.54	2	1.28: 3.12	C	53.4	47.2	163	269
hCV8227677	1+2+3	apoe4=0	source,male,age_ge75	0.0067	0.00701	0.15067	0.00263	1.28	1.07: 1.54	1.24	0.93: 1.67	1.56	1.16: 2.08	C	52.6	46.8	364	836
hCV8227677	1	male=1	apoe4,age_ge75	0.04809	0.04909	0.83364	0.00052	1.44	1.00: 2.06	0.94	0.53: 1.67	3.07	1.62: 5.83	C	53.2	44.7	139	132
hCV8227677	1+2+3	male=1	source,apoe4,age_ge75	0.00765	0.00693	0.1625	0.00234	1.38	1.09: 1.74	1.31	0.90: 1.92	1.88	1.25: 2.82	C	51.8	45.2	305	354
hCV8725171	3	age_ge75=0	NONE	0.00151	0.00143	0.00133	0.15863	1.85	1.26: 2.71	2.11	1.33: 3.35	2.14	0.73: 6.31	G	28.1	17.4	162	152
hCV8725171	1+2+3	age_ge75=0	source	0.01117	0.01248	0.00868	0.36345	1.34	1.07: 1.67	1.43	1.09: 1.87	1.31	0.73: 2.38	G	23.7	18.9	505	428
hCV8725171	3	ALL	NONE	0.01223	0.01401	0.01332	0.25482	1.37	1.07: 1.76	1.45	1.08: 1.96	1.46	0.76: 2.79	G	24.0	18.7	358	395
hCV8725171	1+2+3	ALL	source	0.00901	0.01016	0.01128	0.19773	1.22	1.05: 1.42	1.26	1.05: 1.51	1.3	0.87: 1.95	G	22.7	19.1	949	1127
hCV8780618	1	apoe4=0	male,age_ge75	0.03226	0.02396	0.01159	0.83587	1.47	1.03: 2.09	1.7	1.12: 2.57	0.87	0.21: 3.52	C	22.2	16.3	162	270
hCV8780618	1+2+3	apoe4=0	source,male,age_ge75	0.00776	0.0054	0.00257	0.93874	1.37	1.09: 1.72	1.51	1.15: 1.97	1.03	0.42: 2.55	C	20.5	16.4	364	848
hCV8782652	1	apoe4=0	NONE	0.00955	0.0064	0.00284	0.1939	1.44	1.09: 1.90	2.04	1.27: 3.27	1.38	0.85: 2.25	T	51.8	42.8	163	270
hCV8782652	1+3	apoe4=0	source	0.03425	0.02744	0.00816	0.3968	1.24	1.02: 1.51	1.57	1.12: 2.19	1.16	0.82: 1.63	T	51.2	46.1	304	570
hCV8782652	3	male=0	apoe4,age_ge75	0.09542	0.08558	0.01928	0.7043	1.24	0.96: 1.60	1.65	1.09: 2.52	1.09	0.70: 1.68	T	49.8	46.3	282	295
hCV8782652	1+3	male=0	source,apoe4,age_ge75	0.06549	0.05509	0.00888	0.69313	1.2	0.99: 1.44	1.53	1.12: 2.09	1.07	0.77: 1.49	T	49.5	45.7	532	510
hCV8856240	3	male=1	apoe4,age_ge75	0.00776	0.00927	0.01762	0.07823	2.13	1.22: 3.72	2.22	1.14: 4.30	5.09	0.83: 31.28	G	26.9	16.5	78	100
hCV8856240	1+2+3	male=1	source,apoe4,age_ge75	0.00759	0.00884	0.01734	0.07912	1.5	1.12: 2.02	1.52	1.08: 2.15	2.31	0.93: 5.71	G	20.9	15.4	332	383
hCV8885200	1	apoe4=0	male,age_ge75	0.00006	0.00007	0.00017	0.02036	0.47	0.33: 0.69	0.45	0.30: 0.69	0.19	0.04: 0.88	A	12.6	23.8	178	288
hCV8885200	1+2+3	apoe4=0	source,male,age_ge75	0.00852	0.00895	0.00371	0.61328	0.74	0.60: 0.93	0.68	0.53: 0.88	0.86	0.47: 1.57	A	17.4	22.7	385	887
hCV8921255	1	age_ge75=0	NONE	0.00006	0.00011	0.0003	0.01025	0.51	0.36: 0.71	0.44	0.28: 0.69	0.4	0.20: 0.82	G	24.9	39.5	179	147
hCV8921255	1+2+3	age_ge75=0	source	0.00215	0.00254	0.00229	0.1109	0.72	0.59: 0.89	0.65	0.50: 0.86	0.7	0.45: 1.09	G	28.0	35.0	425	407
hCV8921255	1	ALL	NONE	0.0001	0.00013	0.00027	0.0179	0.65	0.52: 0.81	0.58	0.44: 0.78	0.56	0.35: 0.91	G	26.7	36.1	391	355
hCV8921255	1+2+3	ALL	source	0.00308	0.00321	0.00322	0.11558	0.82	0.71: 0.93	0.77	0.64: 0.92	0.79	0.59: 1.06	G	29.2	33.6	936	1102
hCV8921255	1	male=0	NONE	0.00017	0.00021	0.00031	0.0299	0.59	0.44: 0.78	0.51	0.35: 0.74	0.51	0.27: 0.94	G	25.5	36.8	251	220
hCV8921255	1+2+3	male=0	source	0.00079	0.00074	0.00067	0.08626	0.75	0.64: 0.89	0.69	0.55: 0.85	0.72	0.50: 1.05	G	28.3	34.3	623	740
hCV8984582	1	male=0	apoe4,age_ge75	0.02187	0.01962	0.03738	0.09869	0.72	0.55: 0.95	0.65	0.43: 0.98	0.63	0.36: 1.08	C	36.9	42.3	263	235

TABLE 7, page 8 of 8

Marker	Sample Set	Strata	Adjust	Allelic p-value	Additive p-value	Dominant p-value	Recessive p-value	OR-allelic	OR-allelic 95%CI	OR-dominant	OR-dominant 95%CI	OR-recessive	OR-recessive 95%CI	Allele 1	Case Allele 1 Freq	Control Allele 1 Freq	Case Samples	Control Samples
hCV8984582	1+2	male=0	source,apoe4,age_ge75	0.01039	0.00934	0.01992	0.06651	0.74	0.59: 0.93	0.67	0.48: 0.94	0.66	0.42: 1.03	C	37.0	42.4	354	486
hCV9605432	3	apoe4=0	male,age_ge75	0.01679	0.02303	0.03386	0.13648	0.64	0.44: 0.92	0.62	0.40: 0.96	0.48	0.18: 1.30	G	16.0	22.7	141	300
hCV9605432	1+2+3	apoe4=0	source,male,age_ge75	0.00828	0.01126	0.01013	0.222	0.74	0.59: 0.93	0.71	0.54: 0.92	0.7	0.40: 1.24	G	17.8	23.4	371	872
hCV9605432	3	male=0	NONE	0.09568	0.09517	0.34933	0.01456	0.79	0.59: 1.04	0.85	0.61: 1.19	0.35	0.15: 0.84	G	19.7	23.7	282	295
hCV9605432	1+2+3	male=0	source	0.03261	0.03387	0.17851	0.00704	0.82	0.68: 0.98	0.86	0.69: 1.07	0.49	0.29: 0.83	G	20.5	24.1	631	766
hCV97656	1	age_ge75=1	apoe4,male	0.00003	0.00006	0.00077	0.00055	2.55	1.63: 3.97	2.29	1.41: 3.73		. : .	T	17.0	7.8	227	217
hCV97656	1+2+3	age_ge75=1	source,apoe4,male	0.00072	0.00093	0.00514	0.00389	1.6	1.22: 2.10	1.53	1.14: 2.07	5.17	1.64: 16.33	T	14.0	10.2	520	728
hCV97656	1	ALL	apoe4,male,age_ge75	0.00129	0.00189	0.0046	0.03516	1.71	1.23: 2.38	1.7	1.18: 2.45	2.88	0.98: 8.42	T	15.0	10.1	417	375
hCV97656	1+2+3	ALL	source,apoe4,male,age_ge75	0.00133	0.00158	0.00268	0.087	1.41	1.14: 1.74	1.43	1.13: 1.80	1.88	0.90: 3.94	T	13.3	10.9	963	1157
hDV68530963	1	age_ge75=0	NONE	0.17086	0.18466	0.96827	0.00487	1.27	0.90: 1.77	1.01	0.65: 1.56	3.11	1.37: 7.07	C	32.9	27.9	178	147
hDV68530963	1+2+3	age_ge75=0	source	0.45874	0.46274	0.57878	0.00744	1.08	0.88: 1.34	0.93	0.70: 1.22	2	1.20: 3.33	C	29.7	27.9	424	407
hDV68530976	3	age_ge75=1	NONE	0.43071	0.43582	0.93346	0.01222	0.87	0.63: 1.22	1.02	0.69: 1.50	0.23	0.06: 0.80	G	17.1	19.1	222	246
hDV68530976	1+2+3	age_ge75=1	source	0.88484	0.88733	0.21965	0.00743	1.02	0.81: 1.27	1.18	0.91: 1.53	0.37	0.17: 0.79	G	16.8	16.4	496	694
hDV68531036	1	age_ge75=0	NONE	0.13615	0.15079	0.82263	0.00586	1.3	0.92: 1.82	1.05	0.68: 1.63	3.04	1.34: 6.92	G	32.9	27.4	178	144
hDV68531036	1+2+3	age_ge75=0	source	0.3983	0.40326	0.68354	0.00818	1.1	0.89: 1.36	0.94	0.72: 1.24	1.98	1.19: 3.31	G	29.8	27.7	423	404

CL001496PCT

What Is Claimed Is:

1. A method for identifying an individual who has an altered risk for developing Alzheimer's disease, comprising detecting a single nucleotide polymorphism (SNP) in any one of the nucleotide sequences of SEQ ID NOS:1-433 and 867-54,769 in said individual's nucleic acids, wherein the presence of the SNP is correlated with an altered risk for Alzheimer's disease in said individual.

2. The method of claim 1 in which the altered risk is an increased risk.

3. The method of claim 2 in which said individual has Alzheimer's disease.

4. The method of claim 1 in which the altered risk is a decreased risk.

5. The method of claim 1, wherein the SNP is selected from the group consisting of the SNPs set forth in Tables 6 and 7.

6. The method of claim 1 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.

7. An isolated nucleic acid molecule comprising at least 8 contiguous nucleotides wherein one of the nucleotides is a single nucleotide polymorphism (SNP) selected from any one of the nucleotide sequences in SEQ ID NOS:1-433 and 867-54,769, or a complement thereof.

8. The isolated nucleic acid molecule of claim 7, wherein the SNP is selected from the group consisting of the SNPs set forth in Tables 3 and 4.

9. An isolated nucleic acid molecule that encodes any one of the amino acid sequences in SEQ ID NOS:434-866.

10. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS:434-866.

11. An antibody that specifically binds to a polypeptide of claim 10, or an antigen-binding fragment thereof.

12. The antibody of claim 11 in which the antibody is a monoclonal antibody.

13. An amplified polynucleotide containing a single nucleotide polymorphism (SNP) selected from any one of the nucleotide sequences of SEQ ID NOS:1-433 and 867-54,769, or a complement thereof, wherein the amplified polynucleotide is between about 16 and about 1,000 nucleotides in length.

14. The amplified polynucleotide of claim 13 in which the nucleotide sequence comprises any one of the nucleotide sequences of SEQ ID NOS:1-433 and 867-54,769.

15. 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:1-433 and 867-54,769.

16. The polynucleotide of claim 15 which is 8-70 nucleotides in length.

17. The polynucleotide of claim 15 which is an allele-specific probe.

18. The polynucleotide of claim 15 which is an allele-specific primer.

19. The polynucleotide of claim 15, wherein the polynucleotide comprises a nucleotide sequence selected from the group consisting of the primer sequences set forth in Table 5 (SEQ ID NOS:54,770-55,342).

20. A kit for detecting a single nucleotide polymorphism (SNP) in a nucleic acid, comprising the polynucleotide of claim15, a buffer, and an enzyme.

21. 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:1-433 and 867-54,769 under stringent hybridization conditions, and detecting the formation of a hybridized duplex.

22. The method of claim 21 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.

23. A method of detecting a variant polypeptide, comprising contacting a reagent with a variant polypeptide encoded by a single nucleotide polymorphism (SNP) in any one of the nucleotide sequences of SEQ ID NOS:1-433 and 867-54,769 in a test sample, and detecting the binding of the reagent to the polypeptide.

24. A method for identifying an agent useful in therapeutically or prophylactically treating Alzheimer's disease, comprising contacting the polypeptide of claim 10 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.

25. A method for treating neurodegenerative disease in a human subject, which method comprises administering to said human subject a therapeutically or prophylactically

effective amount of an agent which inhibits the activity of glyceraldehyde-3-phosphate dehydrogenase.

26. A method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene, which method comprises administering to said human subject a therapeutically or prophylactically effective amount of an agent counteracting the neurodegenerative effects of the disease.

27. The method of Claim 26 in which the agent is neuroprotective.

28. The method of Claim 27 in which the agent is anti-apoptotic.

29. The method of Claim 28 in which the agent inhibits the activity of GAPDH.

30. The method of Claim 29 in which the agent inhibits the activity of GAPDH by forming a binding complex with GAPDH.

31. The method of Claim 30 in which the disease is selected from adrenoleukodystrophy, Alexander Disease, Alzheimer's disease, amyotrophic lateral sclerosis, Canavan Disease, cerebellar degeneration, cerebral ischemias, glaucoma, Krabbe Disease, metachromatic leukodystrophy, multiple sclerosis, neuronal ceroid lipofuscinoses, Parkinson's disease, Pelizaeus-Merzbacher Disease, retinitis pigmentosa, stroke, neurodegenerative disease caused by traumatic injury.

32. The method of Claim 29 in which the mutant GAPDH gene comprises a polynucleotide sequence selected from the group consisting of the genomic sequence of SEQ ID NO:6795, the transcript sequences of SEQ ID NOS:125-127, and nucleic acid sequences that encode a polypeptide comprising an amino acid sequence of SEQ ID NOS:558-560.

33. The method of Claim 25 in which the agent is selected from (*R*)-*N*-methyl-*N*-(1-methyl-2-phenyl-ethyl)-*N*-prop-2-ynylamine, dibenzo[*b*,*f*]oxepin-10-ylmethyl-methyl-prop-2-ynyl-amine and (*R*)-indan-1-yl-prop-2-ynyl-amine.

34. A method for identifying an agent useful in therapeutically or prophylactically treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene, which method comprises contacting GAPDH with a candidate agent under conditions suitable to allow formation of a binding complex between the GAPDH and the candidate agent and detecting the formation of the binding complex, wherein the presence of the complex identifies said agent.

35. A method for treating neurodegenerative disease in a human subject wherein said human subject harbors a mutant glyceraldehydes-3-phosphate dehydrogenase (GAPDH) gene, which method comprises:

(i) determining that said human subject harbors the mutant GAPDH gene; and

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

36. A method of determining the risk estimate of an individual for developing late onset Alzheimer's disease comprising:

(i) determining the genotype of an individual at the GAPDH, GAPDS, and p-GAPD loci; and

(ii) correlating the genotype with genotypes of GAPDH, GAPDS, and p-GAPD of known risk.

37. The method of claim 36, wherein the genotype of GAPDH at the variant site identified by hCV2655167 is CG, the genotype of p-GAPD at the variant site identified by hCV2950452 is GG, and the genotype of GAPDS at the variant site identified by hCV337151 is selected from the group consisting of CC, CG, or GG.

38. A method for determining a compound prevents or limits GAPD induced apoptosis and/or nuclear translocation of GAPD comprising:

(i) transfecting neuronal cells with either a wild type or a mutant form of GAPDH, GAPDS, and p-GAPD;

(ii) culturing the transfected cells;

(iii) exposing the cultured cells to a compound or agent suspected of affecting GAPD induced apoptosis or nuclear translocation of GAPD; and

(iv) determining the affect of the compound or agent on apoptosis or nuclear translocation of GAPD.

39. The method of claim 38, wherein the neuronal cells are derived from primary cell cultures.

40. The method of claim 38, wherein the neuronal cells are derived from established cell lines.

41. The method of claim 38, wherein the compound or agent is selected from the group consisting of: monoamine oxydase inhibitors (R-(-)-deprenyl, Selegiline, or CGP 3446), cholinesterase inhibitors (tetrahydroaminoacridine, tacrine), propyl endopeptidase inhibitors (ONO-1603), Rasagiline, TCH346.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US04/24001

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12Q 1/68
US CL : 435/6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	OMIM, Online Mendelian Inheitance in Man database, National Center for Biotechnology Information, National Library of Medicine, NIH, (Bethesda, MD, USA), Accession 602003, Created 9/17/1997 by Lori M. Kelman.	1-6
A	GRIMM et al. Lrmp and Bcat1 are Candidates for the Type I Diabetes Susceptibility Locus Idd6. June 2003, Volume 36, Number 4, pages 241-246.	1-6

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 August 2005 (22.08.2005)	16 SEP 2005
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer Juliet Switzer Telephone No. 571 272 1600

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US04/24001

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: Claims 1-6 regarding polymorphisms within SEQ ID NO: 1

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US04/24001

BOX III. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group 1, claims 1, 2, 3, 4, 5, 6, a method for identifying an individual who has altered risk for developing Alzheimer's disease via detection of a single nucleotide polymorphism.

Group 2, claims 8-9, 13, 14, 15-20, an isolated nucleic acid comprising a polymorphic variant.

Group 3, claim 10, an isolated polypeptide.

Group 4, claims 11-12, a antibody.

Group 5, claim 21-23, a method for detecting a single nucleotide polymorphism.

Group 6, claim 24, a method for identifying an agent useful in treating Alzheimer's disease using a polypeptide of claim 10.

Group 7, claims 25-33, 35, methods for treating neurodegenerative diseases.

Group 8, claims 34, method of identifying an agent useful in treating neurodegenerative disease via binding GAPDH and the candidate agent.

Group 9, claims 36-37, methods for determining the risk estimate of an individual for developing late onset Alzheimer's disease via genotyping of GADPH, GAPDS, and p-GAPD.

Group 10, a method for determining a compound prevents or limits GAPD induced apoptosis.

further species election:

For groups 1, 2, 5 the species in each group are considered each of the 54,336 separately recited polymorphisms within the claims.

For groups 3, 4, 6, the species in each group are the 433 separately recited polymorphism within the claim.

The first named invention which will be searched is Group 1, claims 16, with respect to the first named species which is instant SEQ ID NO: 1 and the polymorphisms contained therein.

There are 164,311 potential groups for search when each species within each group is considered.
The inventions listed as Groups 1-10 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Groups 1 and 2 are related as at least a first named product and method of using the product for each of the particular species within the group. However, these products are not considered to be a special technical feature which joins the claims of group 1 to those of group 2 because first, the methods of group 1 do not recite or require the use of products commensurate with the recitation of the products of group 2. Further, the products of group 2 as recited in the claims are clearly anticipated at least for claim 7 which requires a molecule comprising

"at least 8 contiguous nucleotides" of the recited sequences. Brennan (US 5474796) teaches an array that has thereupon every possible nucleic acid of ten nucleotides or fewer, thus the nucleic acids taught by Brennan comprise EVERY POSSIBLE nucleic acid of eight nucleotides or fewer, and there is no special technical feature that joins groups 1 and 2 (see Col. 9, example 4). The remaining groups include additional products and methods that are not linked by a unifying inventive concept as they are drawn to unique products and methods and are so separately grouped.

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons:

The technical feature that joins all of the nucleic acid species or protien species are that these are nucleic acids and proteins that are asserted to be associated with Alzheimer's disease. However, at the time the invention was made, isolated nucleic acids and protein products associated with Alzhiemer's disease were known, see for example Scangos et al. (US 5015570). Further, each is drawn to a unique nucleic acid sequence or polypeptide sequence that does not share a common structure with the others.

Continuation of B. FIELDS SEARCHED Item 3:
USPATS, USPGPUBS, Medline, Biosis, CAPlus; key words: alzheimer's JAW1, lympoid-restricted membrane protein, polymorphism, mutation