



(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2011/030199

(22) International Filing Date:
28 March 2011 (28.03.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/317,900 26 March 2010 (26.03.2010) US

(71) Applicant (for all designated States except US): **UNIVERSITY OF MIAMI** [US/US]; 1475 N.W. 12th Avenue, Suite 2012, Miami, Florida 33136 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PERICAK-VANCE, Margaret** [US/US]; 9250 Old Cutler Road, Coral Gables, Florida 33156 (US). **HAINES, Jonathan L.** [US/US]; 1609 Knox Dr., Brentwood, Tennessee 37027 (US). **BUXBAUM, Joseph D.** [US/US]; 467 Central Park West, Apt. 2A, New York, New York 10025 (US). **GILBERT, John** [US/US]; 7370 SW 156th St., Palmetto Bay, Florida 33157 (US). **BEECHAM, Gary** [US/US]; 1213 Andora Avenue, Coral Gables, Florida 33146 (US). **MARTIN, Eden** [US/US]; 7740 SW 176th Street, Palmetto Bay, Florida 33157 (US). **NAJ, Adam** [US/US]; 2850 Coconut Avenue, #15, Miami, Florida 33133 (US).

(74) Agent: **DOBBELAERE, Amy, A.**; Novak Druce + Quigg LLP, City Place Tower, 525 Okeechobee Blvd., 15th Floor, West Palm Beach, Florida 33401 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: METHODS AND COMPOSITIONS FOR DIAGNOSING AND PREDICTING RISK OF LATE-ONSET ALZHEIMER'S DISEASE

(57) Abstract: An analysis of genome-wide association in a discovery dataset of 931 cases and 1,104 controls and results from a replication analysis on the strongest associations ($P < 10^{-5}$) using genotype data from four existing studies totaling 1,338 cases and 2,003 controls identified SNPs that are associated with late-onset Alzheimer's Disease (LOAD), providing methods, assays, reagents and kits for predicting a subject's risk of developing LOAD. A typical method of predicting a subject's risk of developing LOAD includes obtaining a biological sample from the subject; analyzing the biological sample for the presence of at least one SNP in the MTHFD1L gene; and correlating the presence of the at least one SNP with an increased risk of developing LOAD. The at least one SNP can be one of the SNPs listed in Table 2 (e.g., one or more of rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422).



METHODS AND COMPOSITIONS FOR DIAGNOSING AND PREDICTING RISK OF LATE-ONSET ALZHEIMER'S DISEASE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of Provisional Application Serial No. 61/317,900 filed March 26, 2010, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The invention relates generally to the fields of genetics and medicine. More particularly, the invention relates to methods of and biomarkers for diagnosing and predicting risk of late-onset Alzheimer's Disease.

BACKGROUND

[0003] Alzheimer Disease (AD) [MIM 104300] is a neurodegenerative disorder characterized by memory and cognitive impairment affecting more than 13% of individuals aged 65 years and older (Alzheimer's Association 2009 Alzheimer's Disease Facts and Figures, Washington, D.C.; Hebert et al., Arch Neurol vol. 60:1119-1122, 2003) and constitutes the most common form of dementia among older adults. While several major genes contributing to risk of Alzheimer Disease have been identified (APP, PS1, PS2), all but one (APOE (Saunders et al., Neurology vol. 43:1467-1472, 1993; Corder et al., Science vol. 261:921-923, 1993; Strittmatter et al., Proc Natl Acad Sci USA vol. 90:1977-1981, 1993) contributed predominantly to early-onset forms of AD that cluster within families; other than APOE, few consistent association signals have been observed for late-onset AD (LOAD). Recent estimates of the heritability of LOAD fall between 60% and 80% (Gatz et al., J Gerontol A Biol Sci Med Sci vol. 52:M117-125, 1997). However, while APOE ϵ 4-alleles elevate AD risk, only 50% of AD cases carry an APOE ϵ 4 allele, suggesting genetic factors elsewhere in the genome contribute to AD risk (Huang et al., Arch Neurol vol. 61:1930-1934, 2004).

[0004] Studies have been conducted that have tested association with LOAD on genome-wide panels of single nucleotide polymorphisms (SNPs). However, it remains unlikely that these studies, incorporating cases and controls from multiple samples with varying case/control inclusion criteria, have identified all loci with modest effect sizes in LOAD.

SUMMARY

[0005] The invention relates to the development of reagents and methods for diagnosing LOAD and for determining a subject's risk of developing LOAD. Studies looking for genetic variants across the genome that affect LOAD have had little success identifying genes other than APOE. Here, an expanded set of AD cases and controls was used to improve the power to detect genetic variants driving LOAD risk. Analyzing 483,399 genetic variants across the genome in a discovery dataset of 931 cases and 1,104 controls, a strong association to the marker rs11754661 on chromosome 6 in the gene *MTHFD1L* was found, in addition to the highly replicated chromosome 19 APOE association. Adjacent variants on chromosome 6 were also genotyped in these same cases and controls, and it was found that these variants were also associated with LOAD. The association was replicated with rs11754661 and additional SNPs in *MTHFD1L* in a combined dataset of cases and controls, some from publicly-available datasets. This finding is important because the gene is known to be involved in biological pathways influencing levels of homocysteine, a significant risk factor for AD. Described herein are methods diagnosing LOAD in an individual and methods of identifying an individual's risk of developing LOAD involving the use of SNPs. Examples of these methods include obtaining a sample from the individual, purifying genomic DNA from the sample; genotyping one or more single nucleotide polymorphisms SNPs associated with LOAD (e.g., one or more SNPs in the *MTHFD1L* gene); and correlating the presence of one or more of the SNPs associated with LOAD with an increased risk of developing LOAD in the individual.

[0006] Accordingly, described herein is a method of predicting a subject's risk of developing LOAD. The method includes the steps of: obtaining a biological sample from the subject; analyzing the biological sample (e.g., blood, plasma, serum, saliva, urine, and tissue) for the presence of at least one SNP in the *MTHFD1L* gene (e.g., an SNP selected from Table 2); and correlating the presence of the at least one SNP with an increased risk of developing LOAD. The at least one SNP can be an SNP such as rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422, e.g., rs11754661 of chromosome 6q25.1. The method can further include analyzing the biological sample for the presence of a mutation in the APOE gene, and correlating the presence of the at least one SNP in the *MTHFD1L* gene and the presence of the mutation in the APOE gene with an increased risk of developing LOAD. In one embodiment, the method can additionally include the step of analyzing homocysteine levels in the biological sample and correlating increased levels of homocysteine and the presence of the at least one SNP in the *MTHFD1L*

gene with an increased risk of developing LOAD. The at least one SNP can be two or more SNPs selected from Table 2.

[0007] Also described herein is a method of diagnosing LOAD in a subject. The method includes the steps of: obtaining a biological sample (e.g., blood, plasma, serum, saliva, urine, and tissue, etc.) from the subject; analyzing the biological sample for the presence of at least one SNP in the MTHFD1L gene; and correlating the presence of the at least one SNP with a diagnosis of LOAD in the subject. The at least one SNP can be, e.g., an SNP selected from Table 2. The at least one SNP can be one or more of the following SNPs: rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422, e.g., rs11754661 of chromosome 6q25.1. The method can further include analyzing the biological sample for the presence of a mutation in the APOE gene, and correlating the presence of the at least one SNP in the MTHFD1L gene and the presence of the mutation in the APOE gene with a diagnosis of LOAD in the subject. In one embodiment, the method can additionally include analyzing homocysteine levels in the biological sample and correlating increased levels of homocysteine and the presence of the at least one SNP in the MTHFD1L gene with a diagnosis of LOAD in the subject. The at least one SNP can be two or more SNPs selected from Table 2. Unless otherwise defined, all technical terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

[0008] Further described herein is a kit for predicting a subject's risk of developing LOAD. A typical kit includes: at least one reagent for analyzing a biological sample from the subject for the presence of at least one SNP in the MTHFD1L gene; at least one control; and instructions for use. The at least one SNP can be an SNP selected from Table 2. The at least one SNP can be one of the following SNPs: rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422, e.g., rs11754661 of chromosome 6q25.1.

[0009] As used herein, "protein" and "polypeptide" are used synonymously to mean any peptide-linked chain of amino acids, regardless of length or post-translational modification, e.g., glycosylation or phosphorylation.

[00010] By the term "gene" is meant a nucleic acid molecule that codes for a particular protein, or in certain cases, a functional or structural RNA molecule.

[00011] As used herein, a "nucleic acid" or a "nucleic acid molecule" means a chain of two or more nucleotides such as RNA (ribonucleic acid) and DNA (deoxyribonucleic acid).

[0010] As used herein, an "allele" may refer to a nucleotide at a SNP position (wherein at least two alternative nucleotides are present in the population at the SNP position, in

accordance with the inherent definition of a SNP) or, for cSNPs, may refer to an amino acid residue that is encoded by the codon which contains the SNP position (where the alternative nucleotides that are present in the population at the SNP position form alternative codons that encode different amino acid residues). An "allele" may also be referred to herein as a "variant". Also, an amino acid residue that is encoded by a codon containing a particular SNP may simply be referred to as being encoded by the SNP.

[0011] The term, "complementary" means that two sequences are complementary when the sequence of one can bind to the sequence of the other in an anti-parallel sense wherein the 3'-end of each sequence binds to the 5'-end of the other sequence and each A, T(U), G, and C of one sequence is then aligned with a T(U), A, C, and G, respectively, of the other sequence.

[0012] The term "complementary sequence" as it refers to a polynucleotide sequence, relates to the base sequence in another nucleic acid molecule by the base-pairing rules. More particularly, the term or like term refers to the hybridization or base pairing between nucleotides or nucleic acids, such as, for instance, between the two strands of a double stranded DNA molecule or between an oligonucleotide primer and a primer binding site on a single stranded nucleic acid to be sequenced or amplified. Complementary nucleotides are, generally, A and T (or A and U), or C and G. Two single stranded RNA or DNA molecules are said to be substantially complementary when the nucleotides of one strand, optimally aligned and compared and with appropriate nucleotide insertions or deletions, pair with at least about 95% of the nucleotides of the other strand, usually at least about 98%, and more preferably from about 99 % to about 100%. Complementary polynucleotide sequences can be identified by a variety of approaches including use of well-known computer algorithms and software, for example the BLAST program.

[0013] The terms "biomarker" and "genetic marker" and "marker" are used interchangeably herein and as used herein refer to a region of a nucleotide sequence (e.g., in a chromosome) that is subject to variability (i.e., the region can be polymorphic for a variety of alleles). For example, a SNP in a nucleotide sequence is a biomarker that is polymorphic for two alleles. Other examples of biomarkers of this invention can include but are not limited to microsatellites, restriction fragment length polymorphisms (RFLPs), repeats (i.e., duplications), insertions, deletions, etc.

[0014] By the phrases "risk genes," "risk genes for diseases," and "disease loci" is meant genetic variants that confer an increased likelihood of developing disease.

[0015] 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. Microarrays can be prepared and used by a number of methods, including those described in U.S. Pat. No. 5,837,832 (Chee et al.), PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (Nat. Biotech. 14:1675-1680, 1996) and Schena, M. et al. (Proc. Natl. Acad. Sci. 93:10614-10619, 1996), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays can be produced by the methods described by Brown et al., U.S. Pat. No. 5,807,522.

[0016] By the terms "genomewide association study" and "GWAS" is meant a study of genetic variation across the entire human genome designed to identify genetic association with observable traits or the presence or absence of a disease.

[0017] As used herein, the term "convergence study" means an evidence-based evaluation of multiple independent experimental methods that substantiate the role of a gene in disease.

[0018] The terms "patient," "subject" and "individual" are used interchangeably herein, and mean a mammalian (e.g., human) subject to be diagnosed and/or treated and/or to obtain a biological sample from.

[0019] Although methods and reagents similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods are described below. All publications, patent applications, and patents mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. The particular embodiments discussed below are illustrative only and not intended to be limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] FIG. 1 is a pair of Quantile-Quantile plots for 483,399 single SNP tests of association. Plots depict expected versus observed $-\log_{10}$ P-values for 483,399 single SNP tests of association (in 931 LOAD cases and 1,104 cognitive controls, with adjustment for principal components as covariates for population substructure). Plot A includes the most-strongly associated SNPs within the APOE locus, whereas plot B excludes the three most-strongly associated SNPs for clarity.

[0021] FIG. 2 is a Manhattan plot of SNP associations in *MTHFD1L*, on chromosome 6 between 151.2Mb and 151.3Mb. Plot of $-\log_{10}$ P-values for single SNP tests of association with LOAD with adjustment for population substructure for the chromosome 6

region from 151.2Mb to 151.3Mb in *MTHFDIL*. Blue circles depict results for SNPs examined as part of genomewide association testing, whereas the green circle depicts the genome wide significant association of rs11754661 and the green triangles show the association results of the additional six SNPs proximal to rs11754661 genotyped on the Taqman platform. The orange line below the x-axis depicts the exons (thick line) and introns (thin line) of the *MTHFDIL*, oriented 5' to 3' from left to right.

[0022] FIG. 3 is a pair of plots of $-\log_{10}$ P-values for 483,399 single SNP tests of association (in 931 LOAD cases and 1,104 cognitive controls, with adjustment for principal components as covariates for population substructure). Plot A includes association results from all SNPs within the *APOE* locus, whereas plot B excludes the three most-strongly associated SNPs for clarity.

[0023] FIG. 4 shows an LD (Plot A: D' , Plot B: r^2) between 130 SNPs genotyped in 931 cases and 1,104 controls in and around the gene *MTHFDIL* (± 50 kilobasepairs). The SNP with the most significant association, rs11754661, is highlighted with a blue arrow in the diagram below.

DETAILED DESCRIPTION

[0024] Described herein are results from an analysis of genome-wide association in a discovery dataset of 931 cases and 1,104 controls and results from a replication analysis on the strongest associations ($P < 10^{-5}$) using genotype data from four existing studies totaling 1,338 cases and 2,003 controls. As described in greater detail in Example 1, SNPs were identified that are associated with LOAD, providing methods, assays, reagents and kits for predicting a subject's risk of developing LOAD. A typical method of predicting a subject's risk of developing LOAD includes obtaining a biological sample from the subject; analyzing the biological sample for the presence of at least one SNP in the *MTHFDIL* gene; and correlating the presence of the at least one SNP with an increased risk of developing LOAD. The at least one SNP can be one of the SNPs listed in Table 2 (e.g., one or more of rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422).

[0025] The below described preferred embodiments illustrate adaptations of these methods. Nonetheless, from the description of these embodiments, other aspects of the invention can be made and/or practiced based on the description provided below.

Biological Methods

[0026] Methods involving conventional molecular biology techniques are described herein. Such techniques are generally known in the art and are described in detail in

methodology treatises such as *Molecular Cloning: A Laboratory Manual*, 3rd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001; and *Current Protocols in Molecular Biology*, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1992 (with periodic updates). Genotyping techniques and SNP analysis methods are generally known in the art and are described in detail in methodology treatises such as *Genetic Analysis of Complex Disease* by Jonathan L. Haines and Margaret A. Pericak-Vance, 2nd ed., 2006, Wiley-Liss Publishing, Hoboken, NJ; and *Single Nucleotide Polymorphisms Methods and Protocols (Methods in Molecular Biology)* by Pui-Yan Kwok, 1st ed., 2002, Humana Press, New York, New York. Genome-wide association studies are reviewed in Pearson and Manolio, *JAMA* vol. 299:1335-1344, 2008. A human haplotype map of over 3.1 million SNPs is described in International HapMap Consortium et al., *Nature* vol. 449:851-861, 2007.

Biomarkers and SNPs

[0027] In the experiments described herein, biomarkers were identified as being associated with risk of LOAD and include SNPs which demonstrated strong association with LOAD risk. In a typical embodiment, a biomarker for predicting LOAD includes at least one SNP on chromosome 6q25.1 and/ or at least one SNP set forth as: rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422, variants, mutants, alleles or complementary sequences thereof. In some embodiments, a biomarker for predicting LOAD includes at least one SNP from Table 2. One example of a method of identifying the risk of developing LOAD or diagnosing LOAD in a patient includes identifying in a patient a biomarker set including at least one SNP on chromosome 6q25.1 and/ or at least one SNP set forth as: rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422. The biomarkers of this invention can be used individually or in combination. Thus, in some embodiments, a method of predicting a patient's risk of developing LOAD includes identifying one of the SNPs described herein in a sample from the patient, while in other embodiments, a method of predicting a patient's risk of developing LOAD includes identifying two or more of the SNPs described herein.

[0028] The coexistence of multiple forms of a genetic sequence gives rise to genetic polymorphisms, including SNPs. SNPs are single base positions in DNA at which different alleles, or alternative nucleotides, exist in a population, and are the most common form of genetic variation in the genome. The SNP position (interchangeably referred to herein as SNP, SNP site, SNP locus, SNP marker, biomarker or marker) is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than

1/100 or 1/1000 members of the populations). An individual may be homozygous or heterozygous for an allele at each SNP position. In some embodiments, an SNP is referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP is an amino acid coding sequence.

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

[0030] 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, 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.

[0031] Causative SNPs do not necessarily occur in coding regions; causative SNPs can occur in, for example, any genetic region that can ultimately affect the expression, structure, and/or activity of the protein encoded by a nucleic acid. Such genetic regions include, for example, those involved in transcription, such as SNPs in transcription factor binding domains, SNPs in promoter regions, in areas involved in transcript processing, such as SNPs at intron-exon boundaries that may cause defective splicing, or SNPs in mRNA processing signal sequences such as polyadenylation signal regions. Some SNPs that are not causative SNPs nevertheless are in close association with, and therefore segregate with, a disease-causing sequence. In this situation, the presence of a SNP correlates with the presence of, or predisposition to, or an increased risk in developing the disease. These SNPs, although not causative, are nonetheless also useful for diagnostics, disease predisposition screening, and other uses.

[0032] An association study (e.g., GWAS) 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 neurodegenerative disease, and comparing the information to that of controls (i.e., individuals (subjects) who do not have the disorder; controls may be also referred to as "healthy" or "normal" individuals (subjects)) 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 desirable.

[0033] 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 neurodegenerative disease (e.g., LOAD). Once a statistically significant association is established between one or more SNPs and a pathological condition (or other phenotype) of interest, then the regions around the SNPs can optionally be thoroughly screened to identify the causative genetic locus or sequences (e.g., the causative SNP/mutation, gene, regulatory region, etc.) that influences the pathological condition or phenotype. Association studies such as GWAS are generally conducted within the general population, in contrast to studies performed on related individuals in affected families (linkage studies).

GWAS and Genomic Convergence

[0034] Described herein are SNPs associated with LOAD that were identified by methods involving GWAS and genomic convergence studies. GWAS use high-throughput genotyping technologies to assay hundreds of thousands of the most common form of genetic variant, the SNP, and relate these variants to diseases or health-related traits. With the advent of GWAS and genome-wide expression studies, the entire genome can now be interrogated with increased resolution, and increased power to detect disease loci for common diseases. A typical GWAS genotypes 300,000 to one million SNPs and has four parts: selection of a large number of individuals with the disease or trait of interest and a suitable comparison group; DNA isolation, genotyping, and data review to ensure high genotyping quality; statistical tests for associations between the SNPs passing quality thresholds and the disease/trait; and replication of identified associations in an independent population sample or examination of functional implications experimentally. GWAS have already been completed for a variety of complex genetic diseases with varying degrees of success (Maragonore et al., *Am. J. Hum. Genet.* 77:685, 2005; Herbert et al., *Obesity* 14:1454, 2006; Duerr et al., *Science* 2006;

Libioulle et al., PLoS Genet. 3: e58, 2007). Two published GWAS have examined LOAD (Coon et al., J. Clin. Psychiatry 68:613, 2007; Li et al., Arch. Neurol. 65:45, 2008).

[0035] Genomic convergence is the combining of results from multiple sample sets and multiple genomic technologies to further facilitate the dissection of LOAD and other complex diseases. This convergence across multiple studies effectively increases sample sizes, decreases false-positives, and adds layers of replication to the initial results. This allows the detection of consistent effects that may otherwise be lost in the flood of statistical-genomic data.

Methods of Predicting Risk of Developing LOAD

[0036] In the methods described herein, the detection of a biomarker (e.g., one or more of the SNPs described herein associated with a risk of developing LOAD) in a subject can be carried out according to methods well known in the art. For example DNA is obtained from any suitable sample from the subject that will contain DNA, preferably genomic DNA, and the DNA is then prepared and analyzed according to well-established protocols for the presence of biomarkers. In some embodiments, analysis of the DNA can be carried out by amplification of the region of interest according to amplification protocols well known in the art (e.g., polymerase chain reaction, ligase chain reaction, strand displacement amplification, transcription-based amplification, self-sustained sequence replication (3SR), Q β replicase protocols, nucleic acid sequence-based amplification (NASBA), repair chain reaction (RCR) and boomerang DNA amplification (BDA)). The amplification product can then be visualized directly in a gel by staining or the product can be detected by hybridization with a detectable probe. When amplification conditions allow for amplification of all allelic types of a biomarker, the types can be distinguished by a variety of well-known methods, such as hybridization with an allele-specific probe, secondary amplification with allele-specific primers, by restriction endonuclease digestion, or by electrophoresis. Also described herein are oligonucleotides for use as primers and/or probes for detecting and/or identifying biomarkers according to the methods described herein.

[0037] A typical method of predicting a subject's risk of developing LOAD includes obtaining a biological sample from the subject; analyzing the biological sample for the presence of at least one SNP in the MTHFD1L gene; and correlating the presence of the at least one SNP with an increased risk of developing LOAD. The at least one SNP can be one of the SNPs listed in Table 2 (e.g., one or more of rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422). In some embodiments, the at least one SNP is the at least one SNP is rs11754661 of chromosome 6q25.1. The method can further

include analyzing the biological sample for the presence of a mutation in the APOE gene, and correlating the presence of the at least one SNP in the MTHFD1L gene and the presence of the mutation in the APOE gene with an increased risk of developing LOAD. Additionally or alternatively, the method can include analyzing homocysteine levels in the biological sample and correlating increased levels of homocysteine and the presence of the at least one SNP in the MTHFD1L gene with an increased risk of developing LOAD. Typically, an increased risk is relative to the risk in an individual who does not have the at least one SNP in the MTHFD1L gene.

[0038] Any suitable biological samples can be used to obtain DNA for genotyping. Typical sources of DNA include amniotic fluid, serum, blood, buccal cells, urine, whole genome amplified DNA, plasma, cells, and tissue (e.g. brain). Any suitable platform or method of genotyping a plurality of samples can be used. A platform is a series of arrays or chips on which high-throughput genotyping is performed. Several genotyping platforms are commercially available, including those from Affymetrix, Illumina, and Perlegen. For example, one can use Illumina's Infinium II assay on the HumanHap550 Beadchip to genotype over half a million SNPs in each case and control participant.

[0039] All genes, gene names, and gene products disclosed herein are intended to correspond to homologs from any species for which the compositions and methods disclosed herein are applicable. Thus, the terms include, but are not limited to, genes and gene products from humans and mice. It is understood that when a gene or gene product from a particular species is disclosed, this disclosure is intended to be exemplary only, and is not to be interpreted as a limitation unless the context in which it appears clearly indicates. Thus, for example, for the genes disclosed herein, which in some embodiments relate to mammalian nucleic acid and amino acid sequences are intended to encompass homologous and/or orthologous genes and gene products from other animals including, but not limited to other mammals, fish, amphibians, reptiles, and birds. In preferred embodiments, the genes or nucleic acid sequences are human.

EXAMPLES

[0040] The present invention is further illustrated by the following specific examples. The examples are provided for illustration only and should not be construed as limiting the scope of the invention in any way.

Example 1 – Novel Chromosome 6 Locus for Late-Onset AD Provides Genetic Evidence for Folate-Pathway Abnormalities

[0041] Genome-wide association studies (GWAS) of LOAD have consistently observed strong evidence of association with polymorphisms in APOE, however, until recently, variants at few other loci with statistically significant associations have replicated across studies. The present study combines data on 483,399 SNPs from a previously-reported GWAS of 492 LOAD cases and 496 controls and from an independent set of 439 LOAD cases and 608 controls to strengthen power to identify novel genetic association signals. Associations exceeding the experiment-wide significance threshold ($\alpha=1.03\times 10^{-7}$) were replicated in an additional 1,338 cases and 2,003 controls. As expected, these analyses unequivocally confirmed APOE's risk effect (rs2075650, $P=1.9\times 10^{-36}$). Additionally, the SNP rs11754661 at 151.2Mb of chromosome 6q25.1 in the gene MTHFD1L (which encodes the methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like protein) was significantly associated with LOAD ($P=4.70\times 10^{-8}$; Bonferroni-corrected $P=0.022$). Subsequent genotyping of SNPs in high linkage disequilibrium ($r^2 > 0.8$) with rs11754661 identified statistically significant associations in multiple SNPs (rs803424, $P=0.016$; rs2073067, $P=0.03$; rs2072064, $P=0.035$), reducing the likelihood of association due to genotyping error. In the replication case-control set, an association of rs11754661 was observed in the same direction as the previous association at $P=0.002$ ($P=1.90\times 10^{-10}$ in combined analysis of discovery and replication sets), with associations of similar statistical significance at several adjacent SNPs (rs17349743, $P=0.005$; rs803422, $P=0.004$). In summary, a novel statistically significant association in MTHFD1L, a gene involved in the tetrahydrofolate synthesis pathway, was observed and replicated. This finding is noteworthy, as MTHFD1L may play a role in the generation of methionine from homocysteine and influence homocysteine-related pathways, and levels of homocysteine are a significant risk factor for LOAD development.

[0042] MTHFD1L is an excellent candidate for predicting LOAD risk on account of its involvement in folate-pathway abnormalities linked with homocysteine. Detection of abnormalities in the folate-pathway linked with homocysteine may be used as an index (e.g., a panel of biomarkers) to predict the probability of developing LOAD. Thus, high homocysteine levels can be used in the methods described herein as a biomarker to predict the increased risk of LOAD. The aging population may find particular use for this biomarker in predicting a risk of developing LOAD. In some embodiments, this biomarker may be used for diagnosing LOAD in a subject.

Results

Dataset Characteristics

[0043] Table 1 depicts the demographic characteristics of the case and control samples examined in initial association analyses. We examined 931 LOAD cases, average age 74.4 years at onset (standard deviation: ± 8.1 years), and 1,104 cognitive controls, average age 73.8 years at exam (± 7.8 years) (Table 1). Cases were 64.5% female, while controls were 61.9% female.

Table 1. Demographic characteristics of participants in the study sample (Mean $\pm$ SD or Number (Percent)).

	All	Cases	Controls
Number of subjects	2,036	931	1,104
Females (%)	1,284 (63.0%)	601 (64.5%)	683 (61.9%)
Age-at-onset [cases] (yr)/			
Age-at-exam [controls] (yr)	--	74.4 $\pm$ 8.1	73.8 $\pm$ 7.8
APOE $\epsilon 4$ carrier status			
--/-- carriers (0 copies)	1223 (60.1%)	399 (42.8%)	824 (74.6%)
$\epsilon 4$ /-- carriers (1 copy)	629 (30.9%)	398 (42.7%)	231 (20.9%)
$\epsilon 4/\epsilon 4$ carriers (2 copies)	144 (7.1%)	127 (13.6%)	17 (1.5%)
carrier status missing	40 (2.0%)	8 (0.9%)	32 (2.9%)

*Percentage of successfully genotyped SNPs among those attempted.

GWAS Results

[0044] 11 SNPs had association p-values (P) $< 10^{-5}$ after adjustment for population substructure (Table 2; $P < 10^{-4}$ in Tables 3 and 4; all association results in FIG. 3).

Table 2:

SNP	Chr	Location	Gene**	Function**	Minor Allele (Freq.)	Discovery GWAS		Replication GWAS		Combined GWAS	
						OR* (95% CI**)	P	OR* (95% CI**)	P	OR* (95% CI**)	P
rs207565 0	19	50087459	<i>TOMM4</i> 0	intron	G (0.2)	2.96 (2.50, 3.50)	1.30×10 ⁻³⁶	5.72 (3.63, 9.02)	6.24×10 ⁻¹⁴	2.94 (2.48, 3.47)	4.87×10 ⁻³⁶
rs405509	19	50100676	<i>APOE</i>		C (0.48)	0.62 (0.55, 0.70)	1.47×10 ⁻¹³	0.65 (0.49, 0.87)	0.00342	0.61 (0.54, 0.70)	8.13×10 ⁻¹⁴
rs810692 2	19	50093506	<i>TOMM4</i> 0	intron	G (0.36)	0.62 (0.54, 0.71)	3.10×10 ⁻¹²	0.79 (0.6, 1.05)	0.108	0.62 (0.54, 0.71)	2.94×10 ⁻¹²
rs157580	19	50087106	<i>TOMM4</i> 0	intron	G (0.35)	0.66 (0.57, 0.75)	1.22×10 ⁻⁹	0.49 (0.36, 0.68)	0.000015 3	0.63 (0.55, 0.71)	7.78×10 ⁻¹³
rs439401	19	50106291	<i>LOC100</i> <i>129500</i>	intron	A (0.34)	0.66 (0.57, 0.75)	1.76×10 ⁻⁹	0.33 (0.22, 0.51)	4.23×10 ⁻⁷	0.63 (0.55, 0.72)	3.80×10 ⁻¹²
rs117546 61	6	15124877 1	<i>MTHFD</i> <i>1L</i>	intron	A (0.07)	2.03 (1.58, 2.62)	4.70×10 ⁻⁸	2.34 (1.37, 3.98)	0.00187	2.10 (1.67, 2.64)	1.90×10 ⁻¹⁰
rs6859	19	50073874	<i>PVRL2</i>	intron	A (0.46)	1.41 (1.24, 1.60)	1.06×10 ⁻⁷	1.7 (1.35, 2.13)	6.13×10 ⁻⁶	1.41 (1.24, 1.60)	9.60×10 ⁻⁸
rs104022 71	19	50021054			C (0.36)	1.39 (1.22, 1.59)	7.26×10 ⁻⁷	1.23 (1.1, 1.38)	0.000277	1.26 (1.16, 1.38)	2.14×10 ⁻⁷
rs650991 6	19	60254214	<i>RDH13</i>	intron	G (0.46)	1.34 (1.18, 1.51)	5.83×10 ⁻⁶	0.87 (0.78, 0.97)	0.0223	1.10 (1.01, 1.20)	0.0334

rs439401	19	50106291	<i>LOC100129500</i>	intron	A (0.34)	0.66 (0.57, 0.75)	1.76×10^{-9}	0.33 (0.22, 0.51)	4.23×10^{-7}	0.63 (0.55, 0.72)	3.80×10^{-12}
rs117546 61	6	15124877 1	<i>MTHFD1L</i>	intron	A (0.07)	2.03 (1.58, 2.62)	4.70×10^{-8}	2.34 (1.37, 3.98)	0.00187	2.10 (1.67, 2.64)	1.90×10^{-10}
rs6859	19	50073874	<i>PVRL2</i>	intron	A (0.46)	1.41 (1.24, 1.60)	1.06×10^{-7}	1.7 (1.35, 2.13)	6.13×10^{-6}	1.41 (1.24, 1.60)	9.60×10^{-8}
rs104022 71	19	50021054			C (0.36)	1.39 (1.22, 1.59)	7.26×10^{-7}	1.23 (1.10, 1.38)	0.000277	1.26 (1.16, 1.38)	2.14×10^{-7}
rs650991 6	19	60254214	<i>RDH13</i>	intron	G (0.46)	1.34 (1.18, 1.52)	5.83×10^{-6}	0.87 (0.78, 0.98)	0.0223	1.10 (1.01, 1.20)	0.0334
rs509512	11	10535013 3	<i>GRI14</i>	intron	C (0.43)	0.75 (0.66, 0.85)	7.37×10^{-6}	1.04 (0.94, 1.16)	0.439	0.94 (0.86, 1.02)	0.133
rs679670	6	13817924 4			G (0.37)	0.74 (0.65, 0.85)	9.83×10^{-6}	1.11 (0.93, 1.34)	0.25	0.87 (0.78, 0.97)	0.016
rs124409 6	12	12348104 4	<i>NCOR2</i>	intron	A (0.1)	1.61 (1.30, 1.99)	0.0000113	1.26 (0.95, 1.67)	0.116	1.59 (1.29, 1.97)	0.0000192
rs799419	6	13818054 8			G (0.37)	0.74 (0.65, 0.85)	0.0000122	1.12 (0.93, 1.35)	0.223	0.87 (0.78, 0.98)	0.0176
rs669397	11	10535159 7	<i>GRI14</i>	intron	G (0.43)	0.75 (0.66, 0.85)	0.0000128	1.04 (0.93, 1.16)	0.48	0.93 (0.86, 1.02)	0.123
rs807429 4	17	61902137	<i>PRKC4</i>	intron	G (0.39)	0.75 (0.66,	0.0000136	1.08 (0.96,	0.193	0.94 (0.86,	0.198

rs811303 2	19	60245950						0.85)		1.21)		1.03)	
							A (0.41)	1.32 (1.17, 1.50)	0.0000158	0.87 (0.78, 0.97)	0.0149	1.07 (0.98, 1.17)	0.11
rs709181 9	10	26028836					C (0.24)	1.38 (1.19, 1.61)	0.0000187	1.06 (0.94, 1.20)	0.36	1.18 (1.07, 1.30)	0.001
rs467604 9	2	10900168 9					A (0.08)	1.62 (1.30, 2.03)	0.0000188	2.40 (1.57, 3.66)	0.000047 4	1.76 (1.44, 2.15)	4.31×10 ⁻⁸
rs929156	6	30247678	TRIM15	missen se (S- >N)			A (0.24)	1.38 (1.19, 1.60)	0.0000257	1.03 (0.90, 1.16)	0.702	1.21 (1.09, 1.33)	0.000238
rs136087 3	13	63489710					A (0.18)	0.70 (0.59, 0.83)	0.0000265	0.98 (0.85, 1.14)	0.831	0.86 (0.77, 0.96)	0.00879
rs170348 06	2	10900233 7					G (0.08)	1.61 (1.29, 2.01)	0.0000266	2.39 (1.56, 3.64)	0.000054 1	1.75 (1.43, 2.15)	5.14×10 ⁻⁸
rs690391 2	6	31099538					A (0.23)	0.72 (0.62, 0.84)	0.0000373	0.95 (0.83, 1.09)	0.483	0.84 (0.76, 0.94)	0.00187
rs597724 8	23	12932916 8	SLC25A 14	intron			A (0.44)	0.74 (0.65, 0.86)	0.0000395	--	--	0.74 (0.64, 0.85)	0.0000325
rs284477 9	6	30313386					A (0.3)	1.33 (1.16, 1.52)	0.0000405	0.99 (0.88, 1.11)	0.847	1.16 (1.06, 1.27)	0.00199
rs998976 1	2	13285587 2					C (0.23)	1.37 (1.18, 1.59)	0.000041	1.02 (0.89, 1.16)	0.809	1.15 (1.04, 1.27)	0.00788

rs102443 38	7	70359758	<i>WBSR</i> <i>17</i>	intron	A (0.21)	1.38 (1.18, 1.60)	0.0000418	0.88 (0.76, 1.01)	0.0598	1.12 (1.01, 1.25)	0.0292
rs206417 9	23	12923687 4			G (0.42)	1.34 (1.16, 1.54)	0.0000445	--	--	1.35 (1.17, 1.55)	0.0000344
rs230134 3	2	40533653			C (0.26)	0.74 (0.64, 0.85)	0.0000452	1.09 (0.94, 1.27)	0.25	0.90 (0.81, 1.01)	0.0623
rs169749 80	16	83530064			G (0.3)	1.32 (1.16, 1.53)	0.0000466	0.88 (0.76, 1.01)	0.0735	1.32 (1.16, 1.51)	0.0000538
rs589104	11	10531298 2	<i>GRIA4</i>	intron	A (0.46)	0.77 (0.68, 0.87)	0.0000466	1.07 (0.95, 1.19)	0.269	0.96 (0.88, 1.05)	0.385
rs271406 8	11	12290475 1	<i>GRAMD</i> <i>1B</i>	intron	A (0.4)	1.31 (1.15, 1.49)	0.0000524	0.99 (0.85, 1.14)	0.855	1.30 (1.14, 1.48)	0.0000631
rs193915 3	11	10525693 6	<i>GRIA4</i>	intron	A (0.49)	0.77 (0.68, 0.87)	0.0000552	1.01 (0.89, 1.14)	0.901	0.77 (0.67, 0.87)	0.0000529
rs256335	19	39007736			A (0.47)	0.77 (0.68, 0.87)	0.0000575	1.30 (1.10, 1.53)	0.00157	0.77 (0.68, 0.88)	0.0000889
rs379826 7	6	46058786	<i>CLIC5</i>	intron	A (0.3)	1.33 (1.16, 1.52)	0.000061	0.95 (0.84, 1.07)	0.414	1.13 (1.03, 1.24)	0.0109
rs284478 0	6	30312505			A (0.3)	1.32 (1.15, 1.51)	0.0000668	0.99 (0.88, 1.12)	0.921	1.15 (1.05, 1.26)	0.00230
rs118424 68	13	55833285			A (0.07)	1.64 (1.28,	0.000069	--	--	1.33 (1.13,	0.000697

							2.09)					1.57)	
rs110619 95	12	1815303	<i>CACNA2D4, LRTM2</i>	intron, 3'- UTR		A (0.18)	1.39 (1.18, 1.64)	0.0000772	--	--	--	1.34 (1.15, 1.57)	0.000245
rs252948 9	7	11094713 2	<i>IMMP2L</i>	intron		G (0.38)	0.77 (0.68, 0.88)	0.0000803	--	--	--	0.90 (0.83, 0.99)	0.0231
rs133150 1	9	92432152	<i>DIRAS2</i>	intron		G (0.28)	1.32 (1.15, 1.52)	0.0000869	--	--	--	1.11 (1.01, 1.22)	0.0356
rs423423 2	3	34051850				A (0.32)	0.76 (0.66, 0.87)	0.0000903	--	--	--	0.76 (0.66, 0.87)	0.0000622
rs139357 6	4	16070766 2				G (0.43)	1.29 (1.14, 1.47)	0.0000911	--	--	--	1.16 (1.06, 1.26)	0.000920

* OR = Odds Ratio *** Freq. = Frequency

** CI = Confidence Interval **** Gene Annotation using SNPper database (Riva and Kohane, Bioinformatics vol. 18:1681-1685, 2002)

Table 4:

SNP	Chr	Location	Gene**	Function****	Minor Allele (Freq.**)	Discovery GWAS		Replication GWAS		Combined GWAS	
						OR* (95% CI**)	P	OR* (95% CI**)	P	OR* (95% CI**)	P
rs2075650	19	50087459	<i>TOMM40</i>	Intron (boundary)	G (0.2)	2.95 (2.50,	1.62×10 ⁻³⁶	6.87 (4.23,	6.03×10 ⁻¹⁵	3.29 (2.82,	2.46×10 ⁻⁵²

rs157582	19	50088059	<i>TOMM40</i>	Intron (boundary)	A (0.28)	2.45 (2.11, 2.84)	1.18×10^{-31}	5.70 (3.72, 8.72)	1.31×10^{-15}	2.72 (2.37, 3.13)	3.55×10^{-45}
rs405509	19	50100676	<i>APOE</i>	Promoter	C (0.48)	0.62 (0.55, 0.70)	1.49×10^{-13}	0.80 (0.54, 1.17)	0.253	0.63 (0.56, 0.71)	8.14×10^{-14}
rs810692	19	50093506	<i>TOMM40</i>	Intron	G (0.36)	0.62 (0.54, 0.71)	3.34×10^{-12}	1.17 (0.81, 1.70)	0.408	0.67 (0.59, 0.76)	3.05×10^{-10}
rs157580	19	50087106	<i>TOMM40</i>	Intron	G (0.35)	0.66 (0.57, 0.75)	1.32×10^{-9}	0.41 (0.28, 0.60)	5.42×10^{-6}	0.63 (0.55, 0.71)	7.78×10^{-13}
rs439401	19	50106291	<i>APOE</i>	Downstream	A (0.34)	0.66 (0.58, 0.76)	2.45×10^{-9}	0.35 (0.23, 0.55)	3.15×10^{-6}	0.63 (0.55, 0.72)	3.80×10^{-12}
rs117546	61	151248771	<i>MTHFD1L</i>	Intron	A (0.07)	2.03 (1.57, 2.62)	4.97×10^{-8}	2.23 (1.24, 4.01)	0.00777	2.10 (1.67, 2.64)	1.90×10^{-10}
rs6859	19	50073874	<i>PVRL2</i>	3' UTR	A (0.46)	1.41 (1.24, 1.60)	1.05×10^{-7}	1.73 (1.35, 2.22)	1.64×10^{-5}	1.55 (1.39, 1.73)	7.33×10^{-15}
rs104022	19	50021054	<i>BCAM</i>	Downstream	C (0.36)	1.39 (1.22, 1.59)	6.91×10^{-7}	1.16 (1.03, 1.31)	0.0178	1.26 (1.16, 1.38)	2.14×10^{-7}
rs225281	10	49862126			C (0.46)	0.73 (0.65, 0.83)	1.55×10^{-6}	0.95 (0.83, 1.08)	0.410	0.84 (0.77, 0.92)	0.000111
rs677356	3	23633034			C (0.13)	1.61 (1.32, 1.97)	2.88×10^{-6}	0.83 (0.70, 0.99)	0.0361	1.09 (0.96, 1.24)	0.174

rs769979 4	4	138565789			A (0.49)	0.74 (0.65, 0.84)	4.34×10^{-6}	0.92 (0.81, 1.05)	0.203	0.83 (0.76, 0.91)	4.37×10^{-5}
rs100057 76	4	111235463	<i>ELOVL6</i>	Intron	G (0.45)	1.37 (1.20, 1.57)	4.48×10^{-6}	0.96 (0.85, 1.09)	0.548	1.12 (1.02, 1.22)	0.0142
rs126451 60	4	160688870			A (0.42)	1.35 (1.19, 1.54)	4.96×10^{-6}	1.03 (0.91, 1.16)	0.685	1.18 (1.08, 1.28)	0.000303
rs120471 55	1	50189517			A (0.42)	1.34 (1.18, 1.53)	5.42×10^{-6}	0.92 (0.82, 1.04)	0.204	1.11 (1.02, 1.21)	0.0218
rs650991 6	1 9	60254214	<i>RDH13</i>	Intron	G (0.46)	1.34 (1.18, 1.52)	6.85×10^{-6}	0.90 (0.79, 1.02)	0.101	1.10 (1.01, 1.20)	0.0334
rs509512 1	1 1	105350133	<i>GRI14</i>	Intron	C (0.43)	0.75 (0.66, 0.85)	7.54×10^{-6}	1.13 (1.01, 1.27)	0.0410	0.94 (0.86, 1.02)	0.133
rs679670	6	138179244			G (0.37)	0.74 (0.65, 0.85)	1.03×10^{-5}	1.22 (0.99, 1.51)	0.0589	0.87 (0.78, 0.97)	0.0160
rs966027 8	1	50272772			C (0.42)	1.33 (1.17, 1.51)	1.06×10^{-5}	--	--	--	--
rs492682 5	1	49831163			A (0.41)	1.33 (1.17, 1.51)	1.06×10^{-5}	0.93 (0.82, 1.04)	0.205	1.10 (1.01, 1.20)	0.0279
rs124409 6	1 2	123481044	<i>NCOR2</i>	Intron (boundary)	A (0.1)	1.61 (1.30, 1.99)	1.20×10^{-5}	1.14 (0.82, 1.59)	0.423	1.49 (1.25, 1.78)	8.00×10^{-6}
rs799419	6	138180548			G (0.37)	0.74 (0.65,	1.28×10^{-5}	1.23 (1.00,	0.0511	0.87 (0.78,	0.0176

rs669397	1 1	105351597	<i>GRIA4</i>	Intron	G (0.43)	0.75 (0.66, 0.86)	1.31×10^{-5}	1.13 (1.00, 1.27)	0.0495	0.93 (0.86, 1.02)	0.123
rs102254 70	7	54155997			A (0.15)	0.68 (0.57, 0.81)	1.37×10^{-5}	0.95 (0.83, 1.09)	0.474	0.81 (0.73, 0.91)	0.000161
rs807429 4	1 7	61902137	<i>PRKCA</i>	Intron	G (0.39)	0.75 (0.66, 0.85)	1.41×10^{-5}	1.16 (1.03, 1.31)	0.0184	0.94 (0.86, 1.03)	0.198
rs116727 2	1	49658574			A (0.31)	1.36 (1.18, 1.56)	1.58×10^{-5}	0.99 (0.87, 1.13)	0.864	1.14 (1.04, 1.26)	0.00491
rs173797 21	1	50049596			A (0.41)	1.32 (1.16, 1.50)	1.67×10^{-5}	0.93 (0.82, 1.04)	0.211	1.10 (1.01, 1.19)	0.0374
rs110252 37	1 1	19726288	<i>NAI2</i>	Intron	A (0.2)	1.42 (1.21, 1.67)	1.71×10^{-5}	1.01 (0.88, 1.17)	0.873	1.17 (1.05, 1.30)	0.00377
rs709181 9	1 0	26028836			C (0.24)	1.39 (1.19, 1.61)	1.71×10^{-5}	1.02 (0.89, 1.17)	0.808	1.18 (1.07, 1.30)	0.00109
rs120838 87	1	118683212			A (0.4)	0.75 (0.65, 0.85)	1.74×10^{-5}	0.94 (0.83, 1.07)	0.359	0.84 (0.77, 0.92)	0.000185
rs811303 2	1 9	60245950	<i>RDHI3</i>	Downstream	A (0.41)	1.32 (1.16, 1.50)	1.80×10^{-5}	0.88 (0.78, 1.00)	0.0467	1.07 (0.98, 1.17)	0.110
rs121358 21	1	118744972			A (0.47)	1.33 (1.17, 1.51)	1.81×10^{-5}	1.11 (0.98, 1.27)	0.107	1.20 (1.09, 1.31)	9.13×10^{-5}

rs111310 99	3	823802			G (0.37)	0.75 (0.65, 0.85)	1.85×10^{-5}	0.94 (0.83, 1.07)	0.368	0.84 (0.77, 0.92)	0.000222
rs189395 3	4	160693858			C (0.47)	1.32 (1.16, 1.50)	1.85×10^{-5}	1.04 (0.92, 1.18)	0.494	1.17 (1.08, 1.28)	0.000308
rs467604 9	2	109001689			A (0.08)	1.62 (1.30, 2.03)	1.98×10^{-5}	1.32 (0.75, 2.32)	0.342	1.76 (1.44, 2.15)	4.31×10^{-8}
rs730387 6	1 2	58421479	<i>SLC16A</i> 7	Intron	G (0.26)	0.72 (0.63, 0.84)	1.98×10^{-5}	0.67 (0.41, 1.11)	0.118	0.72 (0.62, 0.83)	4.87×10^{-7}
rs185629 7	1	49810849			A (0.41)	1.32 (1.16, 1.50)	2.14×10^{-5}	0.93 (0.82, 1.04)	0.205	1.10 (1.01, 1.2)	0.0369
rs438874 4	1	50294191			A (0.32)	1.34 (1.17, 1.53)	2.15×10^{-5}	--	--	--	--
rs111002 38	4	160696507			G (0.47)	1.32 (1.16, 1.50)	2.16×10^{-5}	--	--	--	--
rs448960 6	1	50287140			G (0.32)	1.34 (1.17, 1.53)	2.23×10^{-5}	0.98 (0.86, 1.12)	0.783	1.14 (1.04, 1.25)	0.00531
rs122556 07	1 0	20181408	<i>PLXDC2</i>	Intron	G (0.17)	1.46 (1.23, 1.74)	2.28×10^{-5}	0.86 (0.70, 1.06)	0.170	1.21 (1.06, 1.38)	0.00432
rs492681 4	1	49729723			A (0.41)	1.32 (1.16, 1.50)	2.34×10^{-5}	0.91 (0.81, 1.03)	0.138	1.09 (1.00, 1.19)	0.0465
rs492654 7	1	50319397			G (0.32)	1.34 (1.17, 1.53)	2.43×10^{-5}	0.98 (0.86, 1.12)	0.706	1.13 (1.03, 1.23)	0.00764

rs110389 13	1 1	46516306	<i>FLJ2029</i> 4	Intron	G (0.1)	1.60 (1.28, 1.98)	2.48×10^{-5}	0.77 (0.58, 1.02)	0.0716	1.28 (1.08, 1.50)	0.00390
rs141598 5	1	49703336			A (0.41)	1.31 (1.16, 1.49)	2.68×10^{-5}	--	--	--	--
rs297513 9	1 2	16393084	<i>MGST1</i>	Intron	C (0.26)	0.74 (0.64, 0.85)	2.78×10^{-5}	0.91 (0.57, 1.44)	0.680	0.85 (0.75, 0.96)	0.00723
rs170348 06	2	109002337			G (0.08)	1.61 (1.29, 2.01)	2.79×10^{-5}	1.36 (0.77, 2.42)	0.289	1.75 (1.43, 2.15)	5.14×10^{-8}
rs230554 3	1 9	60251527	<i>RDH13</i>	Synonymo us Coding (L143L)	A (0.25)	1.36 (1.18, 1.56)	2.81×10^{-5}	0.95 (0.82, 1.11)	0.524	1.18 (1.06, 1.30)	0.00205
rs117207 20	3	65794510	<i>MAGI1</i>	Intron	A (0.08)	1.71 (1.33, 2.20)	2.82×10^{-5}	--	--	--	--
rs252949 1	7	110959870	<i>IMMP2L</i>	Intron	A (0.32)	0.74 (0.65, 0.85)	2.92×10^{-5}	0.92 (0.80, 1.05)	0.225	0.85 (0.77, 0.94)	0.00103
rs136087 3	1 3	63489710			A (0.18)	0.70 (0.60, 0.83)	3.02×10^{-5}	1.03 (0.88, 1.20)	0.733	0.86 (0.77, 0.96)	0.00879
rs695788 3	7	147870112			G (0.38)	0.75 (0.66, 0.86)	3.15×10^{-5}	0.85 (0.75, 0.96)	0.0113	0.81 (0.74, 0.88)	2.29×10^{-6}
rs153898 1	1 0	31451361			G (0.5)	1.32 (1.16, 1.51)	3.34×10^{-5}	0.95 (0.84, 1.07)	0.399	1.10 (1.01, 1.20)	0.0325

rs118522 2	1	49731548			A (0.31)	1.33 (1.16, 1.52)	3.36×10 ⁻⁵	--	--	--	
rs172798 7	1	49761808			A (0.31)	1.33 (1.16, 1.52)	3.36×10 ⁻⁵	0.99 (0.87, 1.12)	0.819	1.14 (1.04, 1.25)	0.00657
rs108886 65	1	49911493			A (0.31)	1.33 (1.16, 1.52)	3.36×10 ⁻⁵	0.97 (0.86, 1.11)	0.696	1.13 (1.03, 1.24)	0.00799
rs205087 6	1 0	31093734			T (0.23)	0.72 (0.62, 0.84)	3.36×10 ⁻⁵	1.34 (1.12, 1.60)	0.00148	0.95 (0.85, 1.07)	0.387
rs116726 2	1	49646567			A (0.31)	1.33 (1.16, 1.52)	3.43×10 ⁻⁵	0.98 (0.87, 1.12)	0.807	1.14 (1.04, 1.25)	0.00605
rs108886 79	1	50338853	ELAVL4	Promoter	G (0.31)	1.33 (1.16, 1.53)	3.43×10 ⁻⁵	0.97 (0.85, 1.10)	0.598	1.13 (1.03, 1.24)	0.00930
rs189166 7	1	49867972			A (0.31)	1.33 (1.16, 1.52)	3.50×10 ⁻⁵	0.98 (0.86, 1.12)	0.792	1.14 (1.04, 1.25)	0.00637
rs200088 6	4	160740930			A (0.47)	0.76 (0.67, 0.87)	3.54×10 ⁻⁵	1.01 (0.89, 1.14)	0.890	0.88 (0.81, 0.97)	0.00623
rs111268 7	1	49841291			A (0.31)	1.33 (1.16, 1.52)	3.57×10 ⁻⁵	0.98 (0.86, 1.11)	0.740	1.13 (1.03, 1.24)	0.00746
rs701970 2	9	132897543	GFI1B	Downstream	C (0.18)	1.43 (1.21, 1.69)	3.58×10 ⁻⁵	1.46 (1.05, 2.02)	0.0240	1.42 (1.22, 1.64)	3.45×10 ⁻⁶
rs3957	1	49630498			C (0.31)	1.33 (1.16,	3.74×10 ⁻⁵	--	--	--	--

rs998976 1	2	132855872							1.52)		3.74×10 ⁻⁵	0.98 (0.85, 1.13)	0.784	1.15 (1.04, 1.27)	0.00788
rs126493 28	1	49463210								T (0.4)	3.75×10 ⁻⁵	--	--	--	--
rs134316 1	1	49883437								G (0.31)	3.80×10 ⁻⁵	0.97 (0.85, 1.10)	0.630	1.13 (1.03, 1.24)	0.0112
rs169749 80	1 6	83530064								G (0.31)	3.81×10 ⁻⁵	0.85 (0.72, 0.99)	0.0386	1.10 (1.00, 1.22)	0.0571
rs131519 52	4	138539922								A (0.45)	3.85×10 ⁻⁵	0.98 (0.87, 1.10)	0.689	0.88 (0.81, 0.96)	0.00379
rs171341 7	1 4	19933651				<i>TEPI</i>	Intron (boundary)			A (0.07)	3.87×10 ⁻⁵	1.02 (0.82, 1.26)	0.883	1.24 (1.06, 1.45)	0.00647
rs669329 4	1	49651709								A (0.31)	3.89×10 ⁻⁵	0.98 (0.85, 1.13)	0.784	1.16 (1.06, 1.28)	0.00196
rs482029 7	2 2	36435927				<i>TRIOBP</i>	Intron			G (0.42)	3.92×10 ⁻⁵	1.06 (0.93, 1.21)	0.386	1.18 (1.08, 1.30)	0.000396
rs157796 9	1	49637882								T (0.31)	3.96×10 ⁻⁵	--	--	--	--
rs107363 88	1	50108115								A (0.31)	3.97×10 ⁻⁵	0.96 (0.85, 1.10)	0.589	1.13 (1.03, 1.24)	0.0122

rs111236 8	1	49932177				G (0.31)	1.32 (1.16, 1.51)	3.99×10^{-5}	--	--	--	--
rs753016 9	1	49941178				A (0.31)	1.32 (1.16, 1.51)	3.99×10^{-5}	0.97 (0.86, 1.11)	0.679	1.13 (1.03, 1.24)	0.00921
rs593040 3	2 3	129318208	<i>SLC25A14</i>	Intron		T (0.44)	0.74 (0.65, 0.86)	4.00×10^{-5}	--	--	--	--
rs117948 4	1	49681231				C (0.31)	1.32 (1.16, 1.51)	4.04×10^{-5}	1.00 (0.87, 1.14)	0.960	1.15 (1.04, 1.26)	0.00404
rs116727 0	1	49685647				A (0.31)	1.32 (1.16, 1.51)	4.04×10^{-5}	0.99 (0.87, 1.12)	0.850	1.14 (1.04, 1.25)	0.00569
rs597724 8	2 3	129329168	<i>SLC25A14</i>	Intron		A (0.44)	0.74 (0.65, 0.86)	4.04×10^{-5}	--	--	--	--
rs149446 2	1	49568744				A (0.4)	1.31 (1.15, 1.49)	4.15×10^{-5}	0.94 (0.83, 1.06)	0.308	1.10 (1.01, 1.2)	0.0308
rs230134 3	2	40533653				C (0.26)	0.74 (0.64, 0.85)	4.16×10^{-5}	1.07 (0.91, 1.26)	0.432	0.90 (0.81, 1.01)	0.0623
rs206417 9	2 3	129236874	<i>SUHW3</i>	Promoter		G (0.42)	1.34 (1.16, 1.54)	4.34×10^{-5}	--	--	--	--
rs284621 5	1 1	105376028	<i>KIAA1826</i>	Downstream		G (0.43)	0.76 (0.66, 0.87)	4.34×10^{-5}	1.11 (0.99, 1.25)	0.0808	0.94 (0.86, 1.03)	0.162
rs492681 2	1	49586400				A (0.31)	1.32 (1.16, 1.51)	4.35×10^{-5}	--	--	--	--

rs133821 4	1	49596084					C (0.31)	1.32 (1.16, 1.51)	4.39×10^{-5}	0.98 (0.86, 1.11)	0.732	1.13 (1.03, 1.24)	0.00909
rs102590 67	7	95069392	<i>PKD4</i>	Promoter			A (0.05)	1.86 (1.38, 2.51)	4.39×10^{-5}	1.23 (0.76, 1.98)	0.401	1.62 (1.28, 2.04)	4.85×10^{-5}
rs102443 38	7	70359758	<i>WBSR 17</i>	Intron			A (0.21)	1.37 (1.18, 1.6)	4.52×10^{-5}	0.89 (0.76, 1.03)	0.111	1.12 (1.01, 1.25)	0.0292
rs658836 2	1	50066939					T (0.31)	1.32 (1.16, 1.51)	4.59×10^{-5}	--	--	--	--
rs108886 69	1	50022860					T (0.37)	1.31 (1.15, 1.48)	4.65×10^{-5}	0.98 (0.87, 1.11)	0.759	1.12 (1.03, 1.23)	0.00922
rs278769 3	1	49688435					C (0.31)	1.32 (1.16, 1.51)	4.65×10^{-5}	0.99 (0.87, 1.13)	0.893	1.14 (1.04, 1.25)	0.00597
rs589104	1 1	105312982	<i>GRI44</i>	Intron			A (0.46)	0.77 (0.68, 0.87)	4.72×10^{-5}	1.17 (1.04, 1.32)	0.0105	0.96 (0.88, 1.05)	0.385
rs309221 7	2 0	39813482					G (0.24)	1.36 (1.17, 1.57)	4.73×10^{-5}	0.87 (0.75, 1.01)	0.0673	1.10 (0.99, 1.22)	0.0756
rs492654 5	1	50152063					A (0.31)	1.32 (1.16, 1.51)	4.76×10^{-5}	0.98 (0.86, 1.12)	0.778	1.13 (1.03, 1.24)	0.00828
rs669783 9	1	50163925					A (0.31)	1.32 (1.16, 1.51)	4.76×10^{-5}	0.98 (0.86, 1.12)	0.778	1.13 (1.03, 1.24)	0.00828

rs107889 24	1	50168618				A (0.31)	1.32 (1.16, 1.51)	4.76×10^{-5}	0.98 (0.86, 1.12)	0.778	1.13 (1.03, 1.24)	0.00828
rs116036 69	1 1	134212161				A (0.18)	1.41 (1.19, 1.65)	4.78×10^{-5}	1.09 (0.94, 1.27)	0.251	1.20 (1.07, 1.33)	0.00126
rs593275 2	2 3	129334460	<i>SLC25A14</i>	Intron		A (0.44)	0.75 (0.65, 0.86)	4.81×10^{-5}	--	--	--	--
rs120365 51	1	49555928				A (0.4)	1.30 (1.15, 1.48)	4.81×10^{-5}	0.94 (0.83, 1.06)	0.2873	1.10 (1.01, 1.20)	0.0322
rs779508 3	7	70376454	<i>WBSCR17</i>	Intron		A (0.21)	1.38 (1.18, 1.61)	4.85×10^{-5}	0.87 (0.75, 1.02)	0.0864	1.13 (1.01, 1.25)	0.0329
rs379826 7	6	46058786	<i>CLIC5</i>	Intron		A (0.3)	1.33 (1.16, 1.53)	4.98×10^{-5}	0.95 (0.83, 1.08)	0.445	1.13 (1.03, 1.24)	0.0109
rs602979 1	2 0	39794198				T (0.16)	1.44 (1.21, 1.72)	5.03×10^{-5}	--	--	--	--
rs104249 69	1 9	60258324	<i>RDH13</i>	Intron		C (0.12)	0.64 (0.52, 0.80)	5.15×10^{-5}	1.06 (0.88, 1.27)	0.537	0.83 (0.72, 0.95)	0.00790
rs669384 6	1	50087939				A (0.31)	1.32 (1.15, 1.51)	5.42×10^{-5}	0.98 (0.86, 1.12)	0.795	1.13 (1.03, 1.24)	0.00920
rs754472 8	1	50112236				A (0.31)	1.32 (1.15, 1.51)	5.42×10^{-5}	0.98 (0.86, 1.12)	0.791	1.13 (1.03, 1.24)	0.00854
rs283259 4	2 1	30388188				G (0.31)	1.33 (1.16, 1.51)	5.51×10^{-5}	0.96 (0.85, 1.07)	0.490	1.12 (1.02, 1.24)	0.0180

rs193915 3	1 1	105256936	<i>GRI44</i>	Intron		A (0.49)	1.52) (0.68, 0.87)	5.54×10 ⁻⁵	0.90 (0.78, 1.04)	0.139	1.11 (1.01, 1.21)	0.0360
rs118424 68	1 3	55833285				A (0.07)	1.65 (1.29, 2.10)	5.60×10 ⁻⁵	1.08 (0.86, 1.37)	0.515	1.33 (1.13, 1.57)	0.000697
rs127134 04	2	59860209				C (0.37)	1.31 (1.15, 1.50)	5.66×10 ⁻⁵	0.94 (0.83, 1.06)	0.307	1.09 (1.00, 1.19)	0.0573
rs127433 69	1	50337003				A (0.4)	1.30 (1.15, 1.49)	5.80×10 ⁻⁵	0.92 (0.81, 1.04)	0.169	1.09 (0.99, 1.18)	0.0700
rs256335	1 9	39007736				A (0.47)	0.77 (0.68, 0.87)	5.81×10 ⁻⁵	1.28 (1.06, 1.54)	0.0109	0.92 (0.82, 1.02)	0.102
rs754019 4	1	49563736				C (0.31)	1.32 (1.15, 1.51)	5.93×10 ⁻⁵	--	--	--	--
rs271406 8	1 1	122904751				A (0.4)	1.30 (1.15, 1.48)	5.96×10 ⁻⁵	1.06 (0.89, 1.26)	0.529	1.21 (1.09, 1.34)	0.000280
rs480663 6	1 9	60240370	<i>GP6</i>	Intron		A (0.49)	1.30 (1.14, 1.48)	6.23×10 ⁻⁵	--	--	--	--
rs389985 6	1	50016441				A (0.37)	1.30 (1.14, 1.48)	6.30×10 ⁻⁵	0.97 (0.86, 1.10)	0.677	1.12 (1.03, 1.22)	0.0120
rs669504 1	1	50018096				A (0.37)	1.30 (1.14, 1.48)	6.30×10 ⁻⁵	0.97 (0.86, 1.10)	0.677	1.12 (1.03, 1.22)	0.0120

rs593273 8	2 3	129242314			A (0.41)	1.33 (1.16, 1.53)	6.42×10 ⁻⁵	--	--	--	
rs165443 1	1 9	60241570	GP6	Promoter	A (0.49)	1.30 (1.14, 1.48)	6.54×10 ⁻⁵	1.03 (0.91, 1.16)	0.661	1.12 (1.02, 1.22)	0.0154
rs241472	1	495444440			G (0.31)	1.31 (1.15, 1.5)	6.66×10 ⁻⁵	0.95 (0.84, 1.09)	0.464	1.11 (1.01, 1.22)	0.0245
rs110619 95	1 2	1815303	LRTM2	3' UTR	A (0.18)	1.40 (1.18, 1.64)	6.89×10 ⁻⁵	1.02 (0.61, 1.72)	0.936	1.34 (1.15, 1.57)	0.000245
rs252948 9	7	110947132	IMMP2L	Intron	G (0.38)	0.77 (0.68, 0.88)	7.22×10 ⁻⁵	1.00 (0.89, 1.13)	0.990	0.90 (0.83, 0.99)	0.0231
rs129764 16	1 9	34053326			C (0.25)	0.74 (0.64, 0.86)	7.39×10 ⁻⁵	1.06 (0.92, 1.23)	0.435	0.90 (0.81, 1.00)	0.0504
rs101097 8	2 3	129307141	SLC25A 14	Intron	G (0.44)	0.75 (0.65, 0.87)	7.45×10 ⁻⁵	--	--	--	--
rs644949 3	5	60082253	ELOVL7	Downstream	A (0.28)	1.32 (1.15, 1.51)	7.48×10 ⁻⁵	1.02 (0.81, 1.28)	0.879	1.18 (1.05, 1.31)	0.00417
rs483018 6	2 3	129340004	GPR119	Downstream	G (0.44)	0.75 (0.65, 0.87)	7.61×10 ⁻⁵	--	--	--	--
rs483018 7	2 3	129340424	GPR119	Downstream	A (0.44)	0.75 (0.65, 0.87)	7.85×10 ⁻⁵	--	--	--	--
rs458284 8	1	50084787			A (0.31)	1.31 (1.14, 1.48)	7.97×10 ⁻⁵	0.97 (0.85, 1.09)	0.618	1.12 (1.02, 1.22)	0.0152

rs477954 2	1 5	29639276	<i>OTUD7</i>	Intron	G (0.19)	1.41 (1.19, 1.67)	8.08×10^{-5}	--	1.23)	--
rs757452 3	2	132858237			T (0.23)	1.36 (1.17, 1.58)	8.21×10^{-5}	--	--	--
rs869058	6	90312689	<i>ANKRD6</i>	Intron	A (0.43)	0.77 (0.68, 0.88)	8.22×10^{-5}	0.95 (0.76, 1.18)	0.84 (0.76, 0.94)	0.00169
rs423423 2	3	34051850			A (0.32)	0.76 (0.66, 0.87)	8.23×10^{-5}	0.96 (0.83, 1.10)	0.85 (0.77, 0.94)	0.00112
rs109125 37	1	169433156	<i>FMO2</i>	Intron	G (0.31)	1.32 (1.15, 1.51)	8.33×10^{-5}	1.02 (0.90, 1.16)	1.15 (1.05, 1.25)	0.00325
rs180029 1	2 3	153811479	<i>F8</i>	Synonymous Coding (D1260E)	C (0.21)	0.71 (0.6, 0.84)	8.40×10^{-5}	--	--	--
rs377759 9	6	46050909	<i>CLIC5</i>	Intron	T (0.3)	1.33 (1.16, 1.54)	8.73×10^{-5}	--	--	--
rs139357 6	4	160707662			G (0.43)	1.29 (1.14, 1.47)	8.95×10^{-5}	1.04 (0.92, 1.17)	1.16 (1.06, 1.26)	0.000920
rs593275 4	2 3	129342752	<i>GPR119</i>	Downstream	A (0.44)	0.75 (0.66, 0.87)	9.09×10^{-5}	--	--	--
rs134270 42	2	132858435			C (0.23)	1.35 (1.16, 1.58)	9.24×10^{-5}	0.98 (0.85, 1.13)	1.14 (1.03, 1.26)	0.0125

rs102822 92	7	110879714	<i>IMMP2L</i>	Intron	G (0.38)	0.77 (0.67, 0.88)	9.28×10^{-5}	1.03 (0.91, 1.17)	0.662	0.91 (0.83, 1.00)	0.0465
rs708020 2	1 0	31097100			C (0.24)	0.74 (0.63, 0.86)	9.38×10^{-5}	1.27 (1.06, 1.51)	0.00797	0.94 (0.84, 1.05)	0.280
rs133150 1	9	92432152	<i>DIRAS2</i>	Intron	G (0.28)	1.32 (1.15, 1.52)	9.55×10^{-5}	0.95 (0.83, 1.08)	0.451	1.11 (1.01, 1.22)	0.0356
rs867991	1 0	31429271			C (0.48)	0.77 (0.68, 0.88)	9.59×10^{-5}	1.01 (0.90, 1.14)	0.856	0.90 (0.82, 0.98)	0.0152
rs229258 5	1 0	49856323			G (0.49)	1.29 (1.13, 1.46)	9.60×10^{-5}	0.97 (0.86, 1.11)	0.6919	0.88 (0.81, 0.96)	0.00530
rs483018 8	2 3	129342104	<i>GPR119</i>	Downstream	A (0.44)	0.76 (0.66, 0.87)	9.60×10^{-5}	--	--	--	--
rs199858 7	9	78268523	<i>GCNT1</i>	Intron	C (0.12)	1.51 (1.23, 1.85)	9.71×10^{-5}	0.99 (0.83, 1.18)	0.8887	1.15 (1.01, 1.31)	0.0356
rs954410 5	1 3	75456154			G (0.27)	1.32 (1.15, 1.53)	9.79×10^{-5}	1.01 (0.88, 1.16)	0.8997	1.15 (1.05, 1.27)	0.00373
rs126106 05	1 9	50062678	<i>PVRL2</i>	Intron	A (0.17)	0.71 (0.60, 0.85)	9.83×10^{-5}	0.53 (0.40, 0.69)	4.64×10^{-6}	0.69 (0.60, 0.79)	1.59×10^{-7}

* OR = Odds Ratio

** CI = Confidence Interval

*** Freq. = Frequency

**** Gene Annotation using SNPper database (Riva and Kohane, Bioinformatics vol. 18:1681-1685, 2002)

[0045] Table 2 shows the strongest associations ($P < 10^{-5}$) from a GWAS of late-onset Alzheimer Disease. Single nucleotide polymorphisms (SNPs) demonstrating association with late-onset Alzheimer Disease at $P < 10^{-5}$ in association tests adjusting for covariates from principal components capturing population substructure, evaluated in the Discovery genome-wide association study (GWAS) dataset of 931 independent cases and 1,104 independent cognitively normal controls, in the Replication GWAS dataset of 1,242 independent cases and 1,737 independent controls, and in the Combined GWAS dataset of 2,174 cases and 2,181 controls.

[0046] Table 3 shows SNPs demonstrating association with late-onset Alzheimer Disease at $P < 10^{-4}$ in association tests adjusting for covariates from principal components capturing population substructure, evaluated in the Discovery genome-wide association study (GWAS) dataset of 931 independent cases and 1,104 independent cognitively normal controls, in the Replication GWAS dataset of 1,242 independent cases and 1,737 independent controls, and in the Combined GWAS dataset of 2,174 cases and 2,181 controls.

[0047] Table 4 shows genotyped and imputed SNPs demonstrating association with late-onset Alzheimer Disease at $P < 10^{-4}$ in association tests adjusting for covariates from principal components capturing population substructure, evaluated in the Discovery genome-wide association study (GWAS) dataset of 931 independent cases and 1,104 independent cognitively normal controls, in the Replication GWAS dataset of 1,242 independent cases and 1,737 independent controls, and in the Combined GWAS dataset of 2,174 cases and 2,181 controls.

[0048] Although the SNPs defining the *APOE* $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ alleles, rs429358 and rs7412, were not included on our genotyping platforms, we independently genotyped these SNPs and tested the association of *APOE* $\epsilon 4$ with LOAD risk (OR (95% CI): 4.18 (3.51, 4.97); $P = 5.49 \times 10^{-58}$). SNPs adjacent to the *APOE* haplotype on chromosome 19 otherwise demonstrated the highest associations observed, with the peak association being rs2075650 with $P = 1.90 \times 10^{-36}$, confirming the expected effect of *APOE* on LOAD risk in this sample. The most significant non-*APOE* SNP in our previous GWAS (Beecham et al., Am J Hum Genet vol. 84:35-43, 2009) (Table 5) was rs11610206 on 12q13 (45.92 Mb) with $P = 1.43 \times 10^{-6}$; in this study, this SNP was still strongly associated with LOAD (OR (95% CI): 0.67 (0.54, 0.85); $P = 7.70 \times 10^{-4}$), but not with experiment-wide statistical significance.

Table 5:

SNP	Chr	Location	Gene**	Function**	Minor Allele (Freq.)	Beecham et al. GWAS		Discovery GWAS	
						P	OR* (95% CI**)	P	P
rs11205641	1	49957662			A (0.42)	8.40×10 ⁻⁶	1.21 (1.07, 1.37)		0.00265
rs4926831	1	50062688			G (0.43)	1.20×10 ⁻⁵	1.21 (1.07, 1.37)		0.00287
rs9659092	1	50216176			A (0.43)	4.50×10 ⁻⁶	1.24 (1.09, 1.41)		0.000888
rs11583200	1	50332407			G (0.38)	1.80×10 ⁻⁵	1.27 (1.12, 1.44)		0.000301
rs6428503	1	89356657	GBP2	Intron	G (0.47)	6.70×10 ⁻⁴	1.10 (0.97, 1.25)		0.124
rs11683103	2	34766354			G (0.3)	8.60×10 ⁻⁶	0.78 (0.68, 0.90)		0.000461
rs2119067	2	165835529			G (0.23)	4.40×10 ⁻⁵	0.79 (0.68, 0.92)		0.00229
rs10184275	2	165836174			G (0.18)	2.20×10 ⁻⁵	0.79 (0.67, 0.93)		0.00499
rs3846421	4	7336257	SORCS2	Intron	G (0.34)	1.20×10 ⁻⁴	0.88 (0.77, 1.01)		0.0670
rs12639920	4	42107444	ATP8A1	Downstream	A (0.28)	4.90×10 ⁻⁵	1.28 (1.11, 1.47)		0.000550
rs3807031	6	30141863	PPP1R1I	Promoter	A (0.2)	1.20×10 ⁻⁵	1.30 (1.11, 1.52)		0.00119
rs929156	6	30247678	TRIM15	Intron	A (0.24)	1.70×10 ⁻⁵	1.38 (1.19, 1.60)		0.0000257
rs11754661	6	151248771	MTHFD1L	Intron	A (0.07)	2.00×10 ⁻⁵	2.03 (1.58, 2.62)		4.70×10 ⁻⁸

rs945597 3	6	16832585 5			A (0.1)	4.50×10^{-5}	0.84 (0.68, 1.03)	0.0958
rs203946 1	9	20135988			A (0.38)	3.50×10^{-5}	1.12 (0.99, 1.28)	0.0784
rs789392 8	10	44398949			A (0.13)	2.30×10^{-5}	1.39 (1.16, 1.67)	0.000422
rs107868 28	10	10659989 0	SORCS3	Intron	A (0.47)	4.60×10^{-3}	1.10 (0.97, 1.24)	0.155
rs112448 41	10	12782455 6	ADAM1 2	Intron	G (0.26)	3.40×10^{-3}	1.08 (0.94, 1.24)	0.282
rs378183 5	11	12095346 4	SORL1	Intron	A (0.02)	6.20×10^{-3}	0.70 (0.44, 1.10)	0.118
rs116102 06	12	45925793	FAM113 B	Downstream	G (0.09)	3.50×10^{-7}	0.68 (0.54, 0.85)	0.000770
rs238710 0	13	27324759			G (0.27)	3.80×10^{-5}	1.28 (1.11, 1.47)	0.000871
rs954410 5	13	75456154			G (0.27)	5.40×10^{-6}	1.32 (1.15, 1.52)	0.000105
rs455513 2	15	95740242			G (0.24)	3.10×10^{-5}	1.20 (1.04, 1.39)	0.0139
rs148009 0	15	96533184			A (0.5)	3.50×10^{-5}	0.86 (0.75, 0.97)	0.0160
rs138313 9	15	96535200			G (0.49)	3.50×10^{-5}	1.13 (1.00, 1.28)	0.0520
rs140262 7	18	4123739			A (0.2)	4.40×10^{-5}	0.79 (0.68, 0.93)	0.00430
rs445965 3	19	49291455	ZNF224	Intron	G (0.17)	8.00×10^{-6}	0.76 (0.64, 0.91)	0.00190
rs480220 7	19	49292217	ZNF224	Intron	G (0.18)	9.20×10^{-6}	0.76 (0.64, 0.89)	0.00108
rs374631 9	19	49304071	ZNF224	Coding exon	A (0.17)	3.00×10^{-5}	0.76 (0.65, 0.90)	0.00162

rs206133 2	19	49305501	ZNF224	Downstream	A (0.17)	3.90×10^{-5}	0.76 (0.65, 0.90)	0.00153
rs605924 4	20	29474144			A (0.35)	4.80×10^{-5}	1.17 (1.03, 1.34)	0.0150
rs218056 6	20	29482515	DEFB12 3	Promoter	G (0.35)	3.80×10^{-5}	1.18 (1.03, 1.34)	0.0144

* OR = Odds Ratio

** CI = Confidence Interval

*** Gene Annotation using SNPper database (Riva and Kohane, Bioinformatics vol. 18:1681-1685, 2002)

[0049] Table 5 shows the results from a follow-up of the strongest associations reported in the Beecham et al. (Am J Hum Genet vol. 84: 35-43, 2009) GWAS of late-onset Alzheimer Disease. 32 SNPs demonstrating the strongest association with late-onset Alzheimer Disease at $P < 10^{-5}$ in the Beecham et al. (Am J Hum Genet vol. 84:35-43, 2009) GWAS of late-onset Alzheimer's Disease, tested here for association with adjustment for covariates from principal components capturing population substructure, evaluated in the Discovery genome-wide association study (GWAS) dataset of 931 independent cases and 1,104 independent cognitively normal controls.

[0050] The SNP rs11754661, located at 151.2Mb of chromosome 6q25.1 in the gene *MTHFDIL*, was significantly associated with LOAD ($P = 4.70 \times 10^{-8}$; Bonferroni-corrected $P = 0.022$). To ensure that this association was not spurious due to differences between subsets of genotyped samples, we performed several post-hoc quality control analyses. We examined clustering plots from genotype calling by platform to determine if misclassification could have affected the associations observed for the top 11 SNPs with $P < 10^{-5}$, and observed discrete clustering by genotype for all 11 SNPs. We found no evidence of a difference in genotype frequencies among controls across subsets by genotyping platform (Fisher's exact test $P = 0.71$) or by study center (Fisher's exact test $P = 0.95$, Table 6).

Table 6:

SNP	Beecham et al. Controls						New AD Controls						Udall Controls					
	Genotype counts (freq.)*			HW E			Genotype counts (freq.)*			HW E			Genotype counts (freq.)*			HW E		
	Min or/ Maj or Allele ^e	1/1	1/2	1/3	P**		Min or/ Maj or Allele ^e	1/1	1/2	1/3	P		Min or/ Maj or Allele ^e	1/1	1/2	1/3	P	FET P***
	(1/2)						(1/2)						(1/2)					
rs20756 50	G/A	11 (2.2%)	106 (21.4%)	379 (76.4%)	0.31		G/A	10 (2.3%)	85 (19.5%)	342 (78.3%)	0.11		G/A	3 (1.7%)	50 (29.1%)	119 (69.2%)	0.58	0.15
rs40550 9	C/A	150 (30.2%)	238 (48%)	108 (21.8%)	0.47		C/A	121 (27.7%)	221 (50.6%)	95 (21.7%)	0.77		A/C	42 (24.4%)	87 (50.6%)	43 (25%)	1.00	0.69
rs81069 22	G/A	90 (18.2%)	239 (48.3%)	166 (33.5%)	0.85		G/A	63 (14.4%)	227 (51.9%)	147 (33.6%)	0.11		G/A	29 (16.9%)	73 (42.4%)	70 (40.7%)	0.20	0.16
rs15758 0	G/A	75 (15.1%)	226 (45.6%)	195 (39.3%)	0.5		G/A	66 (15.1%)	211 (48.3%)	160 (36.6%)	0.84		G/A	30 (17.4%)	78 (45.3%)	64 (37.2%)	0.53	0.84
rs43940 1	A/G	75 (15.1%)	222 (44.8%)	199 (40.1%)	0.34		A/G	67 (15.4%)	205 (47%)	164 (37.6%)	0.84		A/G	15 (8.7%)	92 (53.5%)	65 (37.8%)	0.03	0.13
rs11754 661	A/G	1 (0.2%)	51 (10.3%)	443 (89.5%)	1.00		A/G	1 (0.2%)	41 (9.4%)	394 (90.4%)	1.00		A/G	0 (0%)	18 (10.5%)	154 (89.5%)	1.00	0.95
rs6859	A/G	89	256	151	0.32		A/G	78	207	151	0.62		A/G	28	78	66	0.53	0.36

		(17.9 %)	(51.6 %)	(30.4 %)				(17.9 %)	(47.5 %)	(34.6 %)			(16.3 %)	(45.3 %)	(38.4 %)		
rs10402 271	C/A	57 (11.5 %)	214 (43.1 %)	225 (45.4 %)	0.61	C/A	43 (9.8 %)	186 (42.6 %)	208 (47.6 %)	0.91	C/A	15 (8.7 %)	81 (47.1 %)	76 (44.2 %)	0.38	0.72	
rs65099 16	G/A	99 (20%)	234 (47.2 %)	163 (32.9 %)	0.36	G/A	85 (19.5 %)	211 (48.5 %)	139 (32%)	0.77	G/A	32 (18.6 %)	85 (49.4 %)	55 (32%)	1.00	0.99	
rs50951 2	C/A	107 (21.6 %)	241 (48.6 %)	148 (29.8 %)	0.65	C/A	93 (21.3 %)	232 (53.2 %)	111 (25.5 %)	0.18	C/A	34 (19.8 %)	91 (52.9 %)	47 (27.3 %)	0.45	0.58	
rs67967 0	G/A	89 (17.9 %)	225 (45.4 %)	182 (36.7 %)	0.19	G/A	70 (16%)	204 (46.7 %)	163 (37.3 %)	0.69	G/A	25 (14.5 %)	89 (51.7 %)	58 (33.7 %)	0.43	0.63	

* freq. = frequency (percent)

** P-value for deviation from Hardy-Weinberg Equilibrium (HWE)

*** P-value for Fisher's Exact Test (FET)

[0051] Table 6 shows genotype frequency distributions and differences in three subsets of a GWAS dataset for SNPs with strong associations with late-onset Alzheimer Disease. Genotype counts, P-values for Hardy-Weinberg Equilibrium (HWE), and P-values for differences in genotypic distribution from a Fisher's Exact Test (FET) comparing controls in SNPs with strong associations with late-onset Alzheimer Disease in three subsets of a GWAS dataset: cognitively normal controls from the previously-published Beecham et al. (Am J Hum Genet 84:35-43, 2009) study ("Beecham et al. controls"), cognitively normal controls recruited after the Beecham et al. (Am J Hum Genet 84:35-43, 2009) study ("New AD Controls"), and cognitively normal controls consented for multiple genetic studies whose recruitment was funded through the Udall Parkinson's Disease Collaboration ("Udall Controls").

[0052] We also examined differences in dataset characteristics including variation in age, sex, and *APOE* $\epsilon 4$ genotype distributions, and found limited differences between subsets by study center, autopsy- or clinical-confirmation of case or control status, and by genotyping platform (Table 7).

Table 7:

	HHHG		CHGR		MSBB		Autopsy-Confirmed	Clinically-Confirmed
Number of subjects	1,024	663	349	349	349	349	1,687	1,687
Cases (%)	328 (32.0%)	328 (49.5%)	276 (79.1%)	276 (79.1%)	276 (79.1%)	276 (79.1%)	656 (38.9%)	656 (38.9%)
Cases								
Females (%)	210 (64.0%)	200 (61.0%)	191 (69.2%)	191 (69.2%)	191 (69.2%)	191 (69.2%)	410 (62.5%)	410 (62.5%)
Age-at-onset (yr)	72.8 ± 7.1	73.4 ± 6.5	79.2 ± 10.6	79.2 ± 10.6	79.2 ± 10.6	79.2 ± 10.6	73.1 ± 6.9	73.1 ± 6.9
0 copies of APOE ε4 (%)	112 (34.2%)	127 (38.7%)	160 (58.0%)	160 (58.0%)	160 (58.0%)	160 (58.0%)	239 (36.4%)	239 (36.4%)
1 copy of APOE ε4 (%)	147 (44.8%)	155 (47.3%)	96 (34.8%)	96 (34.8%)	96 (34.8%)	96 (34.8%)	302 (46.0%)	302 (46.0%)
2 copies of APOE ε4 (%)	65 (19.8%)	43 (13.1%)	19 (6.9%)	19 (6.9%)	19 (6.9%)	19 (6.9%)	108 (16.5%)	108 (16.5%)
carrier status missing	4 (1.2%)	3 (0.9%)	1 (0.4%)	1 (0.4%)	1 (0.4%)	1 (0.4%)	7 (1.1%)	7 (1.1%)
Controls								
Females (%)	449 (64.5%)	189 (56.4%)	45 (61.6%)	45 (61.6%)	45 (61.6%)	45 (61.6%)	683 (61.9%)	683 (61.9%)
Age-at-exam (yr)	73.2 ± 7.3	73.1 ± 7.0	82.7 ± 9.5	82.7 ± 9.5	82.7 ± 9.5	82.7 ± 9.5	73.1 ± 7.2	73.1 ± 7.2
0 copies of APOE ε4 (%)	515 (74.0%)	247 (73.7%)	62 (85.0%)	62 (85.0%)	62 (85.0%)	62 (85.0%)	762 (73.9%)	762 (73.9%)
1 copy of APOE ε4 (%)	151 (21.7%)	70 (20.9%)	10 (13.7%)	10 (13.7%)	10 (13.7%)	10 (13.7%)	221 (21.4%)	221 (21.4%)
2 copies of APOE ε4 (%)	8 (1.2%)	9 (2.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	17 (1.7%)	17 (1.7%)
carrier status missing	22 (3.2%)	9 (2.7%)	1 (1.4%)	1 (1.4%)	1 (1.4%)	1 (1.4%)	31 (3.0%)	31 (3.0%)
	Illumina 550K	Illumina 610Quad	Illumina IM	Illumina IM	Illumina IM	Illumina IM		
Number of subjects	2,036	172	877	877	877	877		
Cases (%)	492 (49.9%)	0 (0%)	440 (50.2%)	440 (50.2%)	440 (50.2%)	440 (50.2%)		

Cases				
Females (%)	312 (63.4%)	--	289 (65.7%)	
Age-at-onset (yr)	72.9 ± 6.5	--	76.5 ± 9.7	
0 copies of APOE ε4 (%)	169 (34.4%)	--	230 (52.3%)	
1 copy of APOE ε4 (%)	234 (47.6%)	--	164 (37.3%)	
2 copies of APOE ε4 (%)	86 (17.5%)	--	41 (9.3%)	
carrier status missing	3 (0.6%)	--	5 (1.1%)	
Controls				
Females (%)	304 (61.4%)	119 (69.2%)	260 (59.5%)	
Age-at-exam (yr)	74.2 ± 6.5	71.4 ± 7.6	74.4 ± 9.1	
0 copies of APOE ε4 (%)	377 (76.2%)	114 (66.3%)	333 (76.2%)	
1 copy of APOE ε4 (%)	105 (21.2%)	50 (29.1%)	76 (17.4%)	
2 copies of APOE ε4 (%)	10 (2.0%)	3 (1.7%)	4 (0.9%)	
carrier status missing	3 (0.6%)	5 (2.9%)	24 (5.5%)	

[0053] Table 7 shows demographic characteristics of participants, subsetting by study center, autopsy or clinical confirmation of case or control status, and by genotyping platform (Mean $\pm$ SD or Number (Percent)).

[0054] Subsequently, we examined the first hundred principal components generated from EIGENSTRAT to determine if any of the principal components were associated with both differences in genotyping platform subset and disease status at $P < 0.05$ as markers of potential systematic bias. While two principal components other than those used to adjust for population substructure showed association with both genotyping platform subset and LOAD, additional adjustment for these principal components did not change the strength of association between rs11754661 and LOAD. Models further adjusting for age, sex, and *APOE* $\epsilon 4$ carrier status (+/-) (Table 8) only marginally diminished the effect size and statistical significance of the association of rs11754661 with LOAD (adjustment for age and sex, OR (95% CI): 2.03 (1.56, 2.64), $P = 1.42 \times 10^{-7}$; adjustment for age, sex and *APOE* $\epsilon 4$ (+/-), OR (95% CI): 2.01 (1.51, 2.67), $P = 1.64 \times 10^{-6}$).

Table 8:

SNP	Chr	Location	Gene**	Funcio n***	Minor Allele (Freq.)	(1) Unadjusted			(2) Adjusted for Age* and Sex			(3) Adjusted for Age*, Sex, & +/- APOE ε4		
						OR* (95% CI**)	P		OR* (95% CI**)	P		OR* (95% CI**)	P	
rs207565 0	19	50087459	<i>TOMM4</i> 0	intron	G (0.2)	2.96 (2.50, 3.50)	1.30×10 ⁻³⁶		3.08 (2.59, 3.67)	3.72×10 ⁻³⁶		1.47 (1.17, 1.86)	0.00104	
rs405509	19	50100676	<i>APOE</i>		C (0.48)	0.62 (0.55, 0.70)	1.47×10 ⁻¹³		0.61 (0.54, 0.70)	3.55×10 ⁻¹³		0.79 (0.69, 0.92)	0.002	
rs810692 2	19	50093506	<i>TOMM4</i> 0	intron	G (0.36)	0.62 (0.54, 0.71)	3.10×10 ⁻¹²		0.62 (0.54, 0.72)	2.93×10 ⁻¹¹		0.93 (0.80, 1.10)	0.405	
rs157580	19	50087106	<i>TOMM4</i> 0	intron	G (0.35)	0.66 (0.57, 0.75)	1.22×10 ⁻⁹		0.65 (0.56, 0.74)	1.39×10 ⁻⁹		0.94 (0.80, 1.10)	0.457	
rs439401	19	50106291	<i>LOC100</i> <i>129500</i>	intron	A (0.34)	0.66 (0.57, 0.75)	1.76×10 ⁻⁹		0.66 (0.57, 0.76)	5.60×10 ⁻⁹		0.99 (0.84, 1.17)	0.914	
rs117546 61	6	151248771	<i>MTHFD</i> <i>IL</i>	intron	A (0.07)	2.03 (1.58, 2.62)	4.70×10 ⁻⁸		2.03 (1.56, 2.64)	1.42×10 ⁻⁷		2.01 (1.51, 2.67)	1.64×10 ⁻⁶	
rs6859	19	50073874	<i>PVRL2</i>	intron	A (0.46)	1.41 (1.24, 1.60)	1.06×10 ⁻⁷		1.42 (1.24, 1.61)	1.93×10 ⁻⁷		1.11 (0.96, 1.28)	0.163	
rs104022 71	19	50021054			C (0.36)	1.39 (1.22, 1.56)	7.26×10 ⁻⁷		1.45 (1.26, 1.64)	9.82×10 ⁻⁸		1.13 (0.97, 1.28)	0.114	

rs650991 6	19	60254214						1.59)		1.66)		1.31)	
				intron	G (0.46)			1.34 (1.18, 1.52)	5.83×10^{-6}	1.35 (1.18, 1.54)	8.32×10^{-6}	1.33 (1.16, 1.54)	8.11×10^{-5}
rs509512	11	105350133		intron	C (0.43)			0.75 (0.66, 0.85)	7.37×10^{-6}	0.74 (0.64, 0.84)	5.64×10^{-6}	0.77 (0.67, 0.89)	0.000341
rs679670	6	138179244			G (0.37)			0.74 (0.65, 0.85)	9.83×10^{-6}	0.75 (0.65, 0.86)	3.54×10^{-5}	0.73 (0.63, 0.85)	3.57×10^{-5}

* OR = Odds Ratio

** CI = Confidence Interval

*** Gene Annotation using SNPper database (Riva and Kohane, Bioinformatics vol. 18:1681-1685, 2002)

[0055] Table 8 shows changes in effect size and p-value with additional covariate adjustment for age, sex, and presence/absence of the *APOE* ϵ 4 allele for SNP associations demonstrating $P < 10^{-5}$ in preliminary analyses of late-onset Alzheimer Disease. SNPs demonstrating association with late-onset Alzheimer Disease at $P < 10^{-5}$ as identified in Table 2, here showing results from logistic regression modeling with (1) no additional covariate adjustment, (2) additional covariate adjustment for age-at-onset (years, in cases only) and age-at-exam (years, in controls only) and sex, and (3) additional covariate adjustment for age-at-onset (years, in cases only) and age-at-exam (years, in controls only); sex; and presence/absence of the *APOE* ϵ 4 allele. All models include, at minimum, covariate adjustment for principal components capturing population substructure.

[0056] Furthermore, we examined the associations in 4 SNPs in linkage disequilibrium (LD) (FIG. 4) of $D' > 0.8$ with rs11754661, which demonstrated variable patterns of association with LOAD (Figure 2; rs2839947, $P = 0.0479$; rs11757561, $P = 0.000684$; rs2073066, $P = 0.768$; rs13201018, $P = 0.185$). It should be noted that due to the low minor allele frequency (MAF) of rs11754661 (MAF = 0.07), only one of these SNPs, rs11757561 (MAF = 0.20), had an $r^2 > 0.10$ ($r^2 = 0.23$). This SNP had a similar direction of association as rs11754661 (OR (95% CI): 1.31 (1.12, 1.53)).

[0057] Based on the pattern of LD in the vicinity of rs11754661, we examined several haplotypes of *MTHFD1L* which included this SNP to identify potential markers for untyped variants associated with LOAD. Two haplotypes (the first comprising rs2073066-rs11754661-rs13201018, the second comprising rs2839947-rs11757561-rs2073066-rs11754661-rs13201018) both containing the risk-increasing A allele of rs11754661 had highly statistically significant associations similar to the genotypic association of rs11754661 ($P = 4.60 \times 10^{-8}$ and $P = 6.54 \times 10^{-8}$, respectively) (Table 9). Both haplotypes had similar frequencies (MHF) to the A allele of rs11754661 (MHF = 0.0696 and MHF = 0.0629, respectively).

Table 9:

Haplotype e	Strongest Haplotype Association				Second Strongest Haplotype Association			
	Allele	F*	OR**	P	Allele	F*	OR**	P
Haplotype 1 ^a	AAA	0.0696	2.03	4.60×10 ⁻⁸	AGA	0.654	0.851	0.0157
Haplotype 2 ^b	GG	0.2	1.32	0.00054	AA	0.55	0.882	0.05
Extended Haplotype 1 ^c	GGAAA	0.0629	2.09	6.54×10 ⁻⁸	AAAGA	0.529	0.871	0.0325
Extended Haplotype 2 ^d	GGGAAGAG AAGGAAAG (SEQ ID NO:1)	0.0115	3.51	0.000206	GGAAGAAG GAGGAAAG (SEQ ID NO:2)	0.0218	2.3	0.0003

^a Haplotype 1: rs2073066- rs11754661- rs13201018^b Haplotype 2: rs2839947- rs11757561^c Extended Haplotype 1: rs2839947- rs11757561- rs2073066- rs11754661- rs13201018^d Extended Haplotype 2: rs17348429- rs17426727- rs803410- rs6917461- rs803407- rs803403- rs17348890- rs17427389- rs9397027- rs10484779- rs2839947- rs11757561- rs2073066- rs11754661- rs13201018- rs17349743

* Freq. = Frequency

** OR = Odds Ratio

[0058] Table 9 shows associations with late-onset Alzheimer Disease of *MTHFDIL* haplotypes incorporating SNP rs11754661, with adjustment for covariates from principal components capturing population substructure, evaluated in the Discovery GWAS dataset of 931 independent cases and 1,104 independent cognitively normal controls.

[0059] In order to ensure that the association we observed at rs11754661 was not merely due to genotyping error, we genotyped four additional SNPs in *MTHFDIL* proximal to and in high LD ($r^2 > 0.8$) with rs11754661. All SNPs but one (rs7765521, $P=0.055$) demonstrated associations with nominal statistical significance (rs803424, $P=0.016$; rs2073067, $P=0.030$; and rs2072064, $P=0.035$). Figure 1 shows the $-\log_{10}$ -transformed P-values for single SNP tests of association in the *MTHFDIL* and 50kb flanking region (151.2Mb-151.3Mb) surrounding the chromosome 6 association signal at rs11754661, among both SNPs genotyped in the initial GWAS and those genotyped subsequently.

[0060] Association analyses of pooled datasets combining data on 1,242 cases and 1,737 controls confirmed experiment-wide statistically significant associations for SNPs in/near *APOE* (replication from $P=6.24 \times 10^{-14}$ to $P=0.00187$) for all but one SNP (rs8106922; discovery $P=3.10 \times 10^{-12}$, replication $P=0.108$) (Table 2), however the direction of association in the replication was consistent across each of these SNPs. The association of the *MTHFDIL* SNP rs11754661 in the replication was both statistically significant ($P=0.00187$) and showed similar strength and direction (discovery OR (95%CI): 2.03 (1.58, 2.62); replication OR (95%CI): 2.34 (1.37, 3.98)).

Association in Combined Discovery and Replication Datasets

[0061] In combined analyses, associations in and around the *APOE* locus were unequivocally strengthened, with the p-values observed ranging from $P=3.8 \times 10^{-12}$ to $P=4.87 \times 10^{-36}$. Variation at rs11754461 was strongly associated ($P=1.90 \times 10^{-10}$) with an elevated risk of LOAD with OR=2.10 (95% CI: 1.67, 2.64). Several adjacent SNPs also demonstrated nominal associations with similar direction of effect, including rs11757561 ($P=0.000846$) with OR=1.31 (95% CI: 1.12, 1.53) and rs12195069 ($P=0.0432$) with OR=1.25 (95% CI: 1.01, 1.55).

[0062] Two SNPs with only modest statistical significance of association in the discovery GWAS demonstrated highly statistically significant association in analyses combining both discovery and replication datasets (Tables 3 and 4). SNPs rs4676049 and rs17034806, located at 109Mb on chromosome 2q13, had associations of OR=1.62 ($P=1.88 \times 10^{-5}$) and OR=1.61 ($P=2.66 \times 10^{-5}$) respectively in the discovery dataset. However,

combining discovery and replication datasets, the SNP associations gained modest strength in effect size (OR=1.76 for rs4676049 and OR=1.75 for rs17034806), but the associations now exceeded the threshold for experiment-wide statistical significance, with $P=4.31 \times 10^{-7}$ for rs4676049 and $P=5.14 \times 10^{-8}$ for rs17034806.

Methods

Ethics Statement

[0063] After complete description of the study to the subjects, written informed consent was obtained from all participants, in agreement with protocols approved by the institutional review board at each contributing center.

Ascertainment

[0064] Discovery dataset cases and controls were clinically ascertained through the Collaborative Alzheimer's Project (CAP) comprising the University of Miami John P. Hussman Institute for Human Genomics (HIHG) and the Vanderbilt University Center for Human Genetics Research (CHGR), and autopsy-verified cases and controls were collected through the Mount Sinai Brain Bank (MSBB) at the Mount Sinai School of Medicine (see [47]). Additional controls were also identified in the National Cell Repository for Alzheimer's Disease (NCRAD). 266 cases and 643 controls genotyped in the discovery dataset from the NCRAD, HIHG, and CHGR (Edwards et al., *Ann Hum Genet* vol. 74:97-109, 2010) and NCRAD were independent from previously published data sets including those from our group's previously published GWAS (Beecham et al., *Am J Hum Genet* vol. 84:35-43, 2009). All CAP-ascertained cases and controls were recruited and evaluated using standardized criteria and protocols, and case adjudication in the CAP was performed jointly by a Clinical Advisory Board (CAB) composed of both HIHG and CHGR members, with controls evaluated jointly as well.

[0065] All cases and controls from the HIHG, CHGR, and NCRAD met selection criteria described in the Beecham et al. study (Beecham et al., *Am J Hum Genet* vol. 84:35-43, 2009). Briefly, the study was described and written informed consists were obtained from all participants, in accordance with institutional review board protocols at each study center. Each individual classified as a LOAD case met the NINCDS-ADRDA criteria for probable or definite AD and had an age at onset greater than 60 years of age (McKhann et al., *Neurology* vol. 34:939-944, 1984), as determined from specific questions within the clinical history answered by a reliable family informant or from documented significant cognitive impairment in the medical record. Vascular dementia was diagnosed according to contemporary standards (Roman et al., *Neurology* vol. 43:250-260, 1993) by the CAB, and

individuals with confirmed vascular dementia or phenotypic uncertainty were excluded from analyses. Cognitive controls were individuals who showed signs of dementia in clinical history or upon interview, and were drawn from spouses, friends, and other biologically unrelated individuals of cases, were frequency-matched by age and gender to the cases, and were located in the same clinical catchment areas. All cognitive controls were examined, and none showed signs of dementia in clinical history or upon interview. Also, each cognitive control had a documented Mini-Mental State Exam (MMSE) score ≥ 27 or a Modified Mini-Mental State (3MS) Exam score ≥ 87 . Clinical history and interview data for NCRAD controls, including MMSE scores, were made available and collected along with whole blood for DNA extraction for inclusion in our study.

[0066] 306 cases and 81 controls identified in the MSBB were recently deceased patients at the Mount Sinai Medical Center in New York, NY, and had affection status verified through clinical review and brain autopsy. Neither the cases nor controls examined have been used in previously published studies. Covariates including age at death and sex were abstracted from reviews of medical charts performed by members of the MSBB.

[0067] In total, 572 new cases and 724 new controls were genotyped in this study, and after quality controls measures, combined with data on 492 cases and 496 controls from the previous GWAS [18] for analysis. We also had available for replication from the HIHG, a dataset of 246 cases and 69 cognitively normal controls from a previously described dataset (Slifer et al., Am J Med Genet B Neuropsychiatr Genet vol. 141B:208-213, 2006).

Genotyping

[0068] We extracted DNA for individuals ascertained by the HIHG, CHGR, MSBB, and NCRAD from whole blood by using Puregene chemistry (QIAGEN, Germantown, MD, USA). We performed genotyping using the Illumina Beadstation and the Illumina Infinium Human 1M beadchip on 530 cases and 393 controls following the recommended protocol, only with a more stringent GenCall score threshold of 0.25. Genotyping on 248 controls from the PD GWAS dataset (Edwards et al., Ann Hum Genet vol. 74:97-109, 2010) was performed using the Illumina Infinium Human 610-Quad beadchip. Genotyping efficiency was greater than 99%, and quality assurance was achieved by the inclusion of one CEPH control per 96-well plate that was genotyped multiple times. Technicians were blinded to affection status and quality-control samples. We used Taqman Genotyping Assays for SNPs +3937/rs429358 and +4075/rs7412 and performed allelic discrimination/genotype calling on the ABI 7900 Taqman system, the results of which were used to determine *APOE* $\epsilon 2/\epsilon 3/\epsilon 4$ genotypes.

[0069] After excluding samples which failed quality control (described in the next section) with low genotyping call rates, genotype data was available on 870,954 SNPs (after quality control) using the Illumina 1M BeadChip on 440 cases and 437 controls, while genotype data on 490,960 SNPs (after quality control) from the Illumina 610Quad BeadChip was available on 172 controls. Combining these data with the 522,366 SNPs on 492 cases and 496 controls in our previous GWAS (Beecham et al., Am J Hum Genet vol. 84:35-43, 2009), a set of 483,399 SNPs common to all platforms was generated that passed quality control for each subset individually and in a pooled dataset. The Bonferroni-corrected threshold for experiment-wide statistical significance was thus set at Bonferroni-corrected $\alpha=1.03\times 10^{-7}$.

Sample Quality Control

[0070] After genotyping, multiple quality controls were performed including assessment of sample efficiency, which is the proportion of valid genotype calls to attempted calls within a sample. Samples with efficiency less than 0.98 were dropped from the analysis. Reported gender and genetic gender were examined with the use of X-linked SNPs; 32 inconsistent samples were dropped from the analysis. Relatedness between samples was tested via the program Graphical Representation of Relatedness (GRR) (Abecasis et al., Bioinformatics vol. 17:742-743, 2001), and 3 related samples were dropped from the analysis.

[0071] To determine if population substructure exists in the case-control sample, a set of 10,000 SNPs with MAF > 0.25, selected for minimal between-SNP linkage disequilibrium ($r^2 < 0.20$), and spread evenly across the autosomal chromosomes were analyzed using the program STRUCTURE (Pritchard et al., Genetics vol. 155:945-959, 2000; Pritchard et al., Am J Hum Genet vol. 67:170-181, 2000) (burn in: 5,000, iterations: 25,000) assuming different number of assumed subpopulations (K). The $-\log$ likelihood for K was maximized at K = 3, suggesting population substructure. Further analysis was performed in EIGENSTRAT (Price et al., Nat Genet vol. 38:904-909, 2006), where principal components analysis on the sample of 10,000 SNPs was used to generate principal component loadings for samples and remove outliers by using the top ten principal components over 5 iterations with a threshold of six standard deviations. The top three principal component loadings were used as covariates to account for population structure in the association analysis.

[0072] Removing genotyped individuals with low genotype call rates, incorrect reported gender, high relatedness with other samples, and extreme outliers in substructure

analyses, 440 cases and 608 controls remained for inclusion in analysis, and were combined with 492 cases and 496 controls from the previous GWAS.

SNP Quality Control

[0073] Quality control was performed to remove any low quality SNPs. Genotype clusters were redefined using signal intensities of samples with efficiency greater than 0.98, and genotypes were recalled on the basis of these new clusters per the manufacturer's recommendation. Efficiency of individual SNPs was estimated as the proportion of samples with genotype calls for a given SNP, and SNPs with efficiency less than 0.95 were dropped from analysis. Due to concerns of low statistical power to detect association, SNPs with $MAF < 0.005$ were dropped from analysis. Hardy-Weinberg Disequilibrium (HWD) statistics were calculated among controls with the Fisher's exact test in the PLINK software package (Purcell et al., Am J Hum Genet vol. 81:559-575, 2007); SNPs with $P < 10^{-6}$ for HWD were dropped from analysis. In addition, due to concerns with the spurious association originating from the use of different genotyping platforms on samples in the previous and current GWAS studies, distributions of genotype frequencies at each SNP in each study were examined among controls using a Fisher's exact test, and SNPs with highly-differing genotype distributions across genotyping subsets ($P < 0.001$) were dropped prior to analysis. After these quality control measures, 483,399 SNPs remained for association analysis.

Association Analysis

[0074] Association analysis was performed using logistic regression to test association of genotypes with LOAD under an additive model. Logistic regression was used to permit covariate adjustment for loadings taken from the first three principal components identified in EIGENSTRAT to account for population substructure. Here we report results from logistic regression models adjusting only for population substructure with principal components. Further regression modeling was also performed on SNPs with initial associations of $P < 10^{-5}$, extending models to adjust for *APOE* genotype (designated as the number of $\epsilon 4$ alleles), age-at-onset in cases and age-at-exam in controls, and gender as covariates (Table 7). All analyses were performed using the PLINK software package (Purcell et al., Am J Hum Genet vol. 81:559-575, 2007). Quantile-quantile plots of the associations were made (Figure 1), and suggest the absence of systematic bias in the tests of association.

Imputation and Replication Analysis

[0075] To provide independent replication of the associations observed in the discovery dataset, genome-wide genotyping data were combined from four additional

datasets (one unpublished and three publicly-available datasets) and missing genotype data imputed using IMPUTE v1.0 (Marchini et al., Nat Genet vol. 39:906-913, 2007) (Table 10).

Table 10:

SNP	Chr	Location	Discovery GWAS	Replication GWAS			
				ADRC	ADNI	SHARe	TGEN
rs2075650	19	50087459	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs157582	19	50088059	Genotype d	Imputed	Imputed	Imputed	Imputed
rs405509	19	50100676	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs8106922	19	50093506	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs157580	19	50087106	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs439401	19	50106291	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs11754661	6	151248771	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs6859	19	50073874	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs10402271	19	50021054	Genotype d	Genotyp ed	Genotyp ed	Genotyp ed	Genotyp ed
rs2252811	10	49862126	Genotype d	Imputed	Imputed	Imputed	Imputed
rs6773562	3	23633034	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs7699794	4	138565789	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10005776	4	111235463	Genotype d	Imputed	Imputed	Imputed	Imputed
rs12645160	4	160688870	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs12047155	1	50189517	Genotype d	Imputed	Imputed	Imputed	Imputed
rs6509916	19	60254214	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs509512	11	105350133	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs679670	6	138179244	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs9660278	1	50272772	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4926825	1	49831163	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1244096	12	123481044	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs799419	6	138180548	Genotype d	Imputed	Genotyp ed	Imputed	Imputed

rs669397	11	105351597	Genotype d	Imputed	Genotyp ed	Genotyp ed	Genotyp ed
rs10225470	7	54155997	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs8074294	17	61902137	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs1167272	1	49658574	Genotype d	Imputed	Imputed	Imputed	Imputed
rs17379721	1	50049596	Genotype d	Imputed	Imputed	Imputed	Imputed
rs11025237	11	19726288	Genotype d	Genotyp ed	Imputed	Imputed	Imputed
rs7091819	10	26028836	Genotype d	Genotyp ed	Genotyp ed	Genotyp ed	Genotyp ed
rs12083887	1	118683212	Genotype d	Imputed	Imputed	Imputed	Imputed
rs8113032	19	60245950	Genotype d	Genotyp ed	Genotyp ed	Genotyp ed	Genotyp ed
rs12135821	1	118744972	Genotype d	Imputed	Imputed	Imputed	Imputed
rs11131099	3	823802	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs1893953	4	160693858	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4676049	2	109001689	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs7303876	12	58421479	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs1856297	1	49810849	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4388744	1	50294191	Genotype d	Imputed	Imputed	Imputed	Imputed
rs11100238	4	160696507	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4489606	1	50287140	Genotype d	Imputed	Imputed	Imputed	Imputed
rs12255607	10	20181408	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4926814	1	49729723	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4926547	1	50319397	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs11038913	11	46516306	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1415985	1	49703336	Genotype d	Imputed	Imputed	Imputed	Imputed
rs2975139	12	16393084	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs17034806	2	109002337	Genotype d	Genotyp ed	Genotyp ed	Imputed	Imputed

rs2305543	19	60251527	Genotype d	Imputed	Imputed	Imputed	Imputed
rs11720720	3	65794510	Genotype d	Imputed	Imputed	Imputed	Imputed
rs2529491	7	110959870	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1360873	13	63489710	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs6957883	7	147870112	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1538981	10	31451361	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1185222	1	49731548	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1727987	1	49761808	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10888665	1	49911493	Genotype d	Imputed	Imputed	Imputed	Imputed
rs2050876	10	31093734	Genotype d	Genotyp ed	Genotyp ed	Imputed	Genotyp ed
rs1167262	1	49646567	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10888679	1	50338853	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs1891667	1	49867972	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs2000886	4	160740930	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1112687	1	49841291	Genotype d	Imputed	Imputed	Imputed	Imputed
rs7019702	9	132897543	Genotype d	Genotyp ed	Imputed	Imputed	Imputed
rs3957	1	49630498	Genotype d	Imputed	Imputed	Imputed	Imputed
rs9989761	2	132855872	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs12049328	1	49463210	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1343161	1	49883437	Genotype d	Imputed	Imputed	Genotyp ed	Genotyp ed
rs16974980	16	83530064	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs13151952	4	138539922	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1713417	14	19933651	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs6693294	1	49651709	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4820297	22	36435927	Genotype d	Imputed	Genotyp ed	Imputed	Imputed

rs1577969	1	49637882	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10736388	1	50108115	Genotype d	Imputed	Imputed	Genotyp ed	Genotyp ed
rs1112368	1	49932177	Genotype d	Imputed	Imputed	Imputed	Imputed
rs7530169	1	49941178	Genotype d	Imputed	Imputed	Imputed	Imputed
rs5930403	23	129318208	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs1179484	1	49681231	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs1167270	1	49685647	Genotype d	Imputed	Imputed	Imputed	Imputed
rs5977248	23	129329168	Genotype d	Genotyp ed	Genotyp ed	Imputed	Imputed
rs1494462	1	49568744	Genotype d	Imputed	Imputed	Imputed	Imputed
rs2301343	2	40533653	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs2064179	23	129236874	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs2846215	11	105376028	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4926812	1	49586400	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1338214	1	49596084	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs10259067	7	95069392	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10244338	7	70359758	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs6588362	1	50066939	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10888669	1	50022860	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs2787693	1	49688435	Genotype d	Imputed	Imputed	Imputed	Imputed
rs589104	11	105312982	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs3092217	20	39813482	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4926545	1	50152063	Genotype d	Imputed	Imputed	Imputed	Imputed
rs6697839	1	50163925	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10788924	1	50168618	Genotype d	Imputed	Imputed	Imputed	Imputed
rs11603669	11	134212161	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed

rs5932752	23	129334460	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs12036551	1	49555928	Genotype d	Imputed	Imputed	Imputed	Imputed
rs7795083	7	70376454	Genotype d	Genotyp ed	Imputed	Imputed	Imputed
rs3798267	6	46058786	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs6029791	20	39794198	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10424969	19	60258324	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs6693846	1	50087939	Genotype d	Imputed	Imputed	Imputed	Imputed
rs7544728	1	50112236	Genotype d	Imputed	Imputed	Imputed	Imputed
rs2832594	21	30388188	Genotype d	Imputed	Imputed	Genotyp ed	Imputed
rs1939153	11	105256936	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs11842468	13	55833285	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs12713404	2	59860209	Genotype d	Imputed	Imputed	Imputed	Imputed
rs12743369	1	50337003	Genotype d	Imputed	Imputed	Imputed	Imputed
rs256335	19	39007736	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs7540194	1	49563736	Genotype d	Imputed	Imputed	Imputed	Imputed
rs2714068	11	122904751	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs4806636	19	60240370	Genotype d	Imputed	Imputed	Imputed	Imputed
rs3899856	1	50016441	Genotype d	Imputed	Imputed	Imputed	Imputed
rs6695041	1	50018096	Genotype d	Imputed	Imputed	Imputed	Imputed
rs5932738	23	129242314	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1654431	19	60241570	Genotype d	Imputed	Imputed	Imputed	Imputed
rs241472	1	49544440	Genotype d	Genotyp ed	Imputed	Imputed	Genotyp ed
rs11061995	12	1815303	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs2529489	7	110947132	Genotype d	Genotyp ed	Genotyp ed	Genotyp ed	Genotyp ed
rs12976416	19	34053326	Genotype d	Imputed	Imputed	Imputed	Imputed

rs1010978	23	129307141	Genotype d	Imputed	Imputed	Imputed	Imputed
rs6449493	5	60082253	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4830186	23	129340004	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4830187	23	129340424	Genotype d	Genotyp ed	Imputed	Imputed	Imputed
rs4582848	1	50084787	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs4779542	15	29639276	Genotype d	Imputed	Imputed	Imputed	Imputed
rs7574523	2	132858237	Genotype d	Imputed	Imputed	Imputed	Imputed
rs869058	6	90312689	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4234232	3	34051850	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs10912537	1	169433156	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs1800291	23	153811479	Genotype d	Imputed	Imputed	Imputed	Imputed
rs3777599	6	46050909	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1393576	4	160707662	Genotype d	Imputed	Genotyp ed	Genotyp ed	Genotyp ed
rs5932754	23	129342752	Genotype d	Genotyp ed	Imputed	Imputed	Imputed
rs13427042	2	132858435	Genotype d	Imputed	Imputed	Imputed	Imputed
rs10282292	7	110879714	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs7080202	10	31097100	Genotype d	Imputed	Imputed	Imputed	Imputed
rs1331501	9	92432152	Genotype d	Genotyp ed	Genotyp ed	Genotyp ed	Genotyp ed
rs867991	10	31429271	Genotype d	Imputed	Imputed	Imputed	Imputed
rs2292585	10	49856323	Genotype d	Imputed	Imputed	Imputed	Imputed
rs4830188	23	129342104	Genotype d	Genotyp ed	Imputed	Imputed	Imputed
rs1998587	9	78268523	Genotype d	Genotyp ed	Imputed	Genotyp ed	Genotyp ed
rs9544105	13	75456154	Genotype d	Imputed	Genotyp ed	Imputed	Imputed
rs12610605	19	50062678	Genotype d	Genotyp ed	Genotyp ed	Genotyp ed	Imputed

[0076] Table 10 shows genotyping or Imputation of SNPs associated with LOAD at $P < 10^{-4}$. Index indicating whether SNPs demonstrating association with late-onset Alzheimer Disease at $P < 10^{-4}$ in association tests adjusting for population substructure in the Discovery dataset where genotyped or imputed in the Discovery dataset (931 independent cases and 1,104 independent cognitively normal controls) or any of the Replication Datasets, including the from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Frank et al., *Neurobiol Aging* vol. 24:521-536, 2003) (147 cases and 182 controls), the Framingham Study SHARe dataset (SHARe) (Bachman et al., *Neurology* vol. 2:115-119, 1992) (86 cases and 1,200 controls (all unrelated)), the Reiman et al. LOAD GWAS dataset (TGEN) (Reiman et al., *Neuron* vol. 54:713-720, 2007) (859 cases and 552 controls), and an additional set of LOAD cases and controls independent of the Discovery dataset and not used in prior publications (ADRC) (Slifer et al., *Am J Med Genet B Neuropsychiatr Genet* vol. 141B:208-213, 2006) (246 LOAD cases and 69 cognitively normal controls).

[0077] SNPs with differing genotypic distributions between datasets were excluded from imputation using the Fisher's exact test approached described earlier (Ziegler A., *Genet Epidemiol* vol. 33:S45-S50, 2009). Both primary and replication datasets were imputed to a HapMap reference of over 2.5 million SNPs. Individual genotypes with probability less than 0.90 were not included, and SNPs missing > 10% of genotypes within either data set were dropped. In addition to using the combined Hapmap Phase III CEPH Utah pedigree (CEU) and Tuscan (TSI) haplotype reference panels for imputation, for imputation within each study, we used genotype data on controls from other datasets to improve imputation accuracy, and Affymetrix 5.0 genotype data on 105 individuals genotyped in an independent Ashkenazi Jewish genotyping panel (IntraGenDB population genetics database).

[0078] We analyzed existing pooled and imputed datasets of unrelated individuals from several studies: 147 cases and 182 controls from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Frank et al., *Neurobiolo Aging* vol. 24:521-536, 2003), 86 cases and 1,200 controls (all unrelated) from the Framingham Study SHARe dataset (Bachman et al., *Neurology* vol. 42:115-119, 1992), and 859 cases and 552 controls from the Reiman et al. (*Neuron* vol. 54:713-720, 2007) LOAD GWAS dataset, and a set of 246 LOAD cases and 69 cognitively normal controls previously described (Slifer et al., *Am J Med Genet B Neuropsychiatric Genet* vol. 141B:208-213, 2006) and genotyped on the Affymetrix 6.0 genotyping platform on which results have not been previously published.

Supplementary Methods

Haplotype Analyses

[0079] In order to define haplotypes, Linkage Disequilibrium (LD) structure in the vicinity of SNP rs11754661 was examined with the Haploview program (Barrett et al., Bioinformatics 21: 263-265, 2005) estimating both D' and r^2 measures for this region. Haplotypes were constructed using LD blocks assigned by Haploview, including the LD block containing the SNP (consisting of rs2073066, rs11754661, and rs13201018; analyzed as "Haplotype 1") and immediately adjacent to the SNP (consisting of rs2839947 and rs11757561; analyzed as "Haplotype 2") were examined for association. Extended haplotypes were constructed by further incorporating SNPs from LD Blocks immediately adjacent to the Haplotype 1 block in MTHFD1L, which included the Haplotype 2 block and a third, larger block incorporating SNPs rs17348429, rs17426727, rs803410, rs6917461, rs803407, rs803403, rs17348890, rs17427389, rs9397027, and rs10484779 (labeled here as "Haplotype 3"). This set of haplotypes included "Extended Haploype 1" (comprising SNPs from Haplotype 1 and 2 blocks), and "Extended Haplotype 2" (comprising SNPs from Haplotype 1, 2, and 3 blocks).

[0080] Haplotypic association tests were performed in a manner similar to genotypic association analysis, using a logistic regression approach with covariate adjustment for loadings taken from the first three principal components identified in EIGENSTRAT (Price et al., Nat Genet 38: 904-909) to account for population substructure, 2006). All analyses were performed using the "--hap-logistic" function in the PLINK software package (Purcell et al., Am J Hum Genet 81: 559-575, 2007).

Summary

[0081] *MTHFD1L*, which encodes the methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like protein, is involved in tetrahydrofolate (THF) synthesis, catalyzing the reversible synthesis of 10-formyl-THF to formate and THF, an important step in homocysteine conversion to methionine. On-going biological investigations are continuing to elucidate the pathways connecting elevated homocysteine with AD. Mthfd1l protein has been reported to be decreased in the hippocampus in a mouse model of AD using a proteomic approach (Martin et al., PLoS One 3:e2750, 2008). Homocysteic acid, derived from homocysteine and methionine, is elevated in these mice and treatment with antibodies to homocysteic acid reduced amyloid burden and inhibited cognitive decline in these animals [40]. B6-deficient diets lead to further increases in homocysteic acid in these mice. In this genome-wide association study of LOAD, we identified a novel association with experiment-wide statistical significance in a gene with a potential biological role, *MTHFD1L*. We

replicated this association in additional publicly-available genomewide association datasets, and observed statistically significant association with a similar effect size and direction at this SNP. In summary, *MTHFDIL* is an excellent candidate for LOAD on account of its involvement in folate-pathway abnormalities linked with homocysteine, a significant biological risk factor for AD.

Other Embodiments

[0082] Any improvement may be made in part or all of the method steps. All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference. The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended to illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. Any statement herein as to the nature or benefits of the invention or of the preferred embodiments is not intended to be limiting, and the appended claims should not be deemed to be limited by such statements. More generally, no language in the specification should be construed as indicating any non-claimed element as being essential to the practice of the invention. This invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contraindicated by context.

What is claimed is:

1. A method of predicting a subject's risk of developing late-onset Alzheimer's Disease (LOAD), the method comprising the steps of:
 - a) obtaining a biological sample from the subject;
 - b) analyzing the biological sample for the presence of at least one single nucleotide polymorphism (SNP) in the *MTHFD1L* gene; and
 - c) correlating the presence of the at least one SNP with an increased risk of developing LOAD.
2. The method of claim 1, wherein the at least one SNP is an SNP selected from Table 2.
3. The method of claim 2, wherein the at least one SNP is an SNP selected from the group consisting of: rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422.
4. The method of claim 3, wherein the at least one SNP is rs11754661 of chromosome 6q25.1.
5. The method of claim 1, further comprising analyzing the biological sample for the presence of a mutation in the APOE gene, and correlating the presence of the at least one SNP in the *MTHFD1L* gene and the presence of the mutation in the APOE gene with an increased risk of developing LOAD.
6. The method of claim 1, further comprising analyzing homocysteine levels in the biological sample and correlating increased levels of homocysteine and the presence of the at least one SNP in the *MTHFD1L* gene with an increased risk of developing LOAD.
7. The method of claim 1, wherein the at least one SNP comprises two or more SNPs selected from Table 2.

8. The method of claim 1, wherein the biological sample is selected from the group consisting of: blood, plasma, serum, saliva, urine, and tissue.
9. A method of diagnosing LOAD in a subject, the method comprising the steps of:
 - a) obtaining a biological sample from the subject;
 - b) analyzing the biological sample for the presence of at least one SNP in the *MTHFDIL* gene; and
 - c) correlating the presence of the at least one SNP with a diagnosis of LOAD in the subject.
10. The method of claim 9, wherein the at least one SNP is an SNP selected from Table 2.
11. The method of claim 10, wherein the at least one SNP is an SNP selected from the group consisting of: rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422.
12. The method of claim 11, wherein the at least one SNP is rs11754661 of chromosome 6q25.1.
13. The method of claim 9, further comprising analyzing the biological sample for the presence of a mutation in the APOE gene, and correlating the presence of the at least one SNP in the *MTHFDIL* gene and the presence of the mutation in the APOE gene with a diagnosis of LOAD in the subject.
14. The method of claim 9, further comprising analyzing homocysteine levels in the biological sample and correlating increased levels of homocysteine and the presence of the at least one SNP in the *MTHFDIL* gene with a diagnosis of LOAD in the subject.
15. The method of claim 9, wherein the at least one SNP comprises two or more SNPs selected from Table 2.
16. The method of claim 9, wherein the biological sample is selected from the group consisting of: blood, plasma, serum, saliva, urine, and tissue.

17. A kit for predicting a subject's risk of developing LOAD, the kit comprising:
 - (a) at least one reagent for analyzing a biological sample from the subject for the presence of at least one SNP in the *MTHFD1L* gene;
 - (b) at least one control; and
 - (c) instructions for use.
18. The kit of claim 17, wherein the at least one SNP is an SNP selected from Table 2.
19. The method of claim 18, wherein the at least one SNP is an SNP selected from the group consisting of: rs2075650, rs11754661, rs803424, rs2073067, rs2072064, rs17349743, and rs803422.
20. The method of claim 19, wherein the at least one SNP is rs11754661 of chromosome 6q25.1.

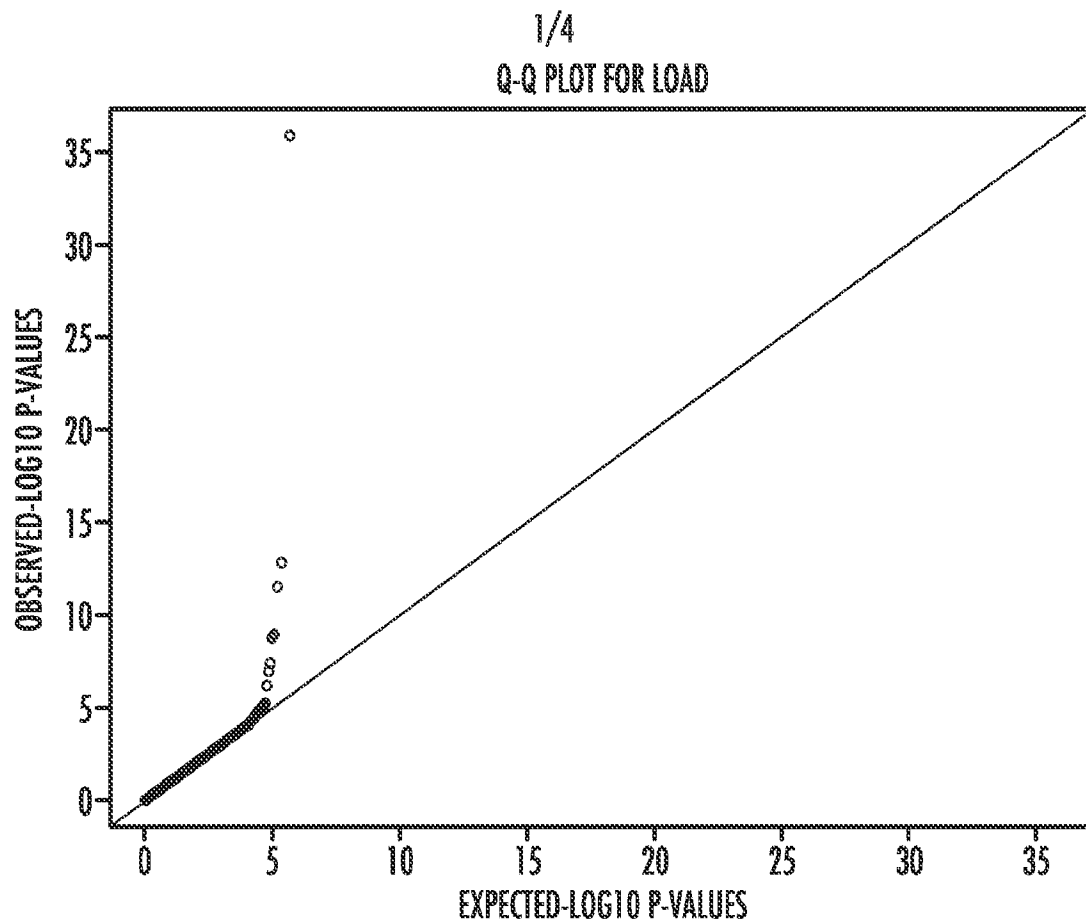


FIG. 1A

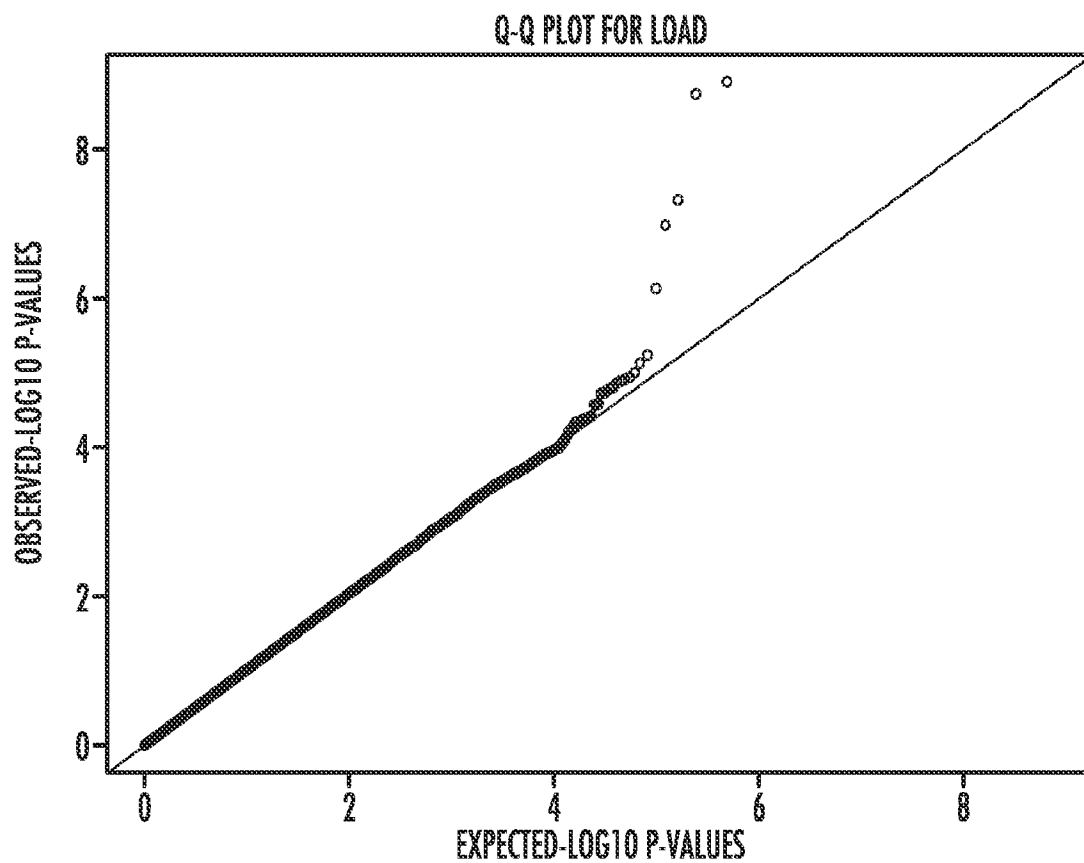


FIG. 1B

2/4

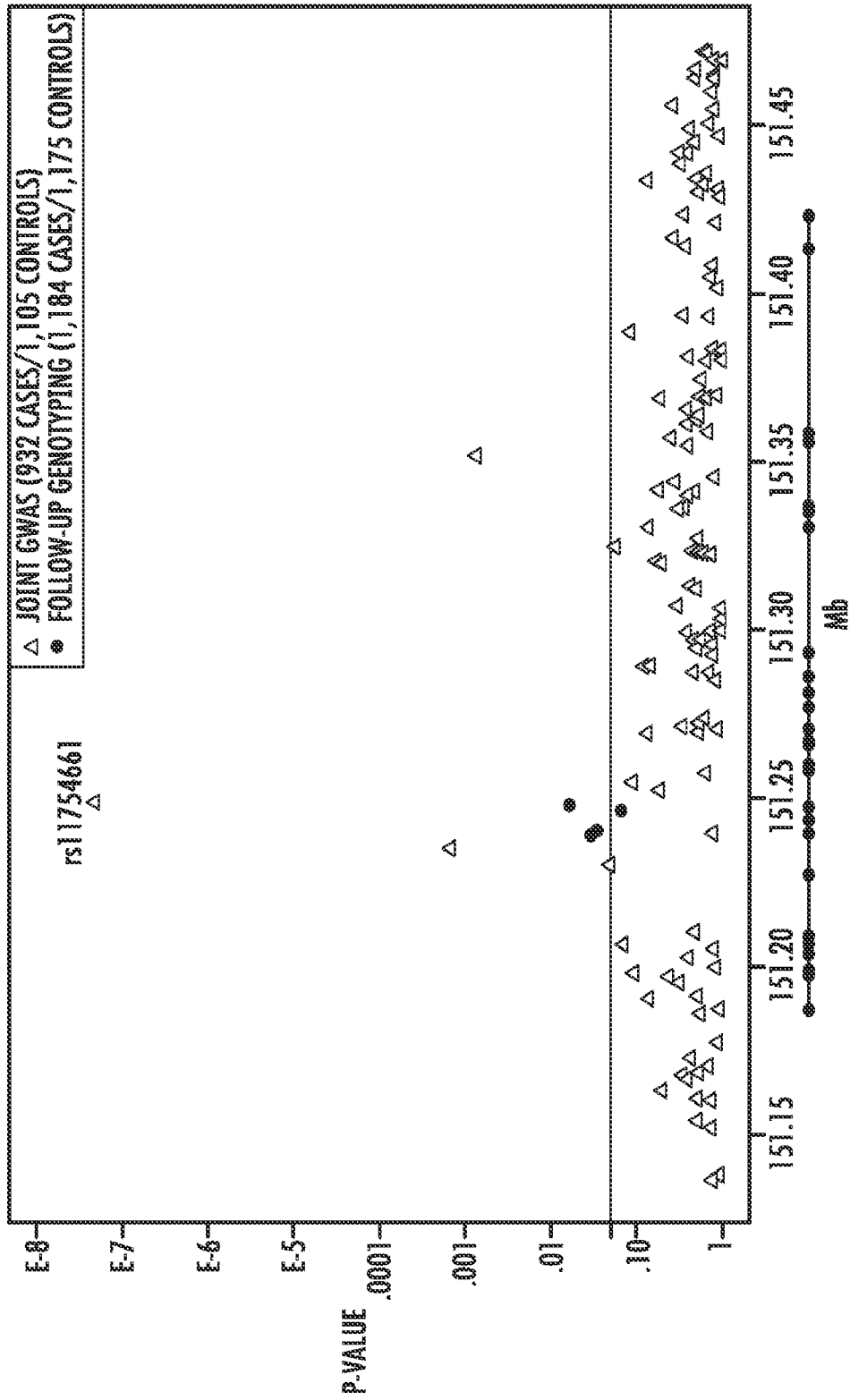
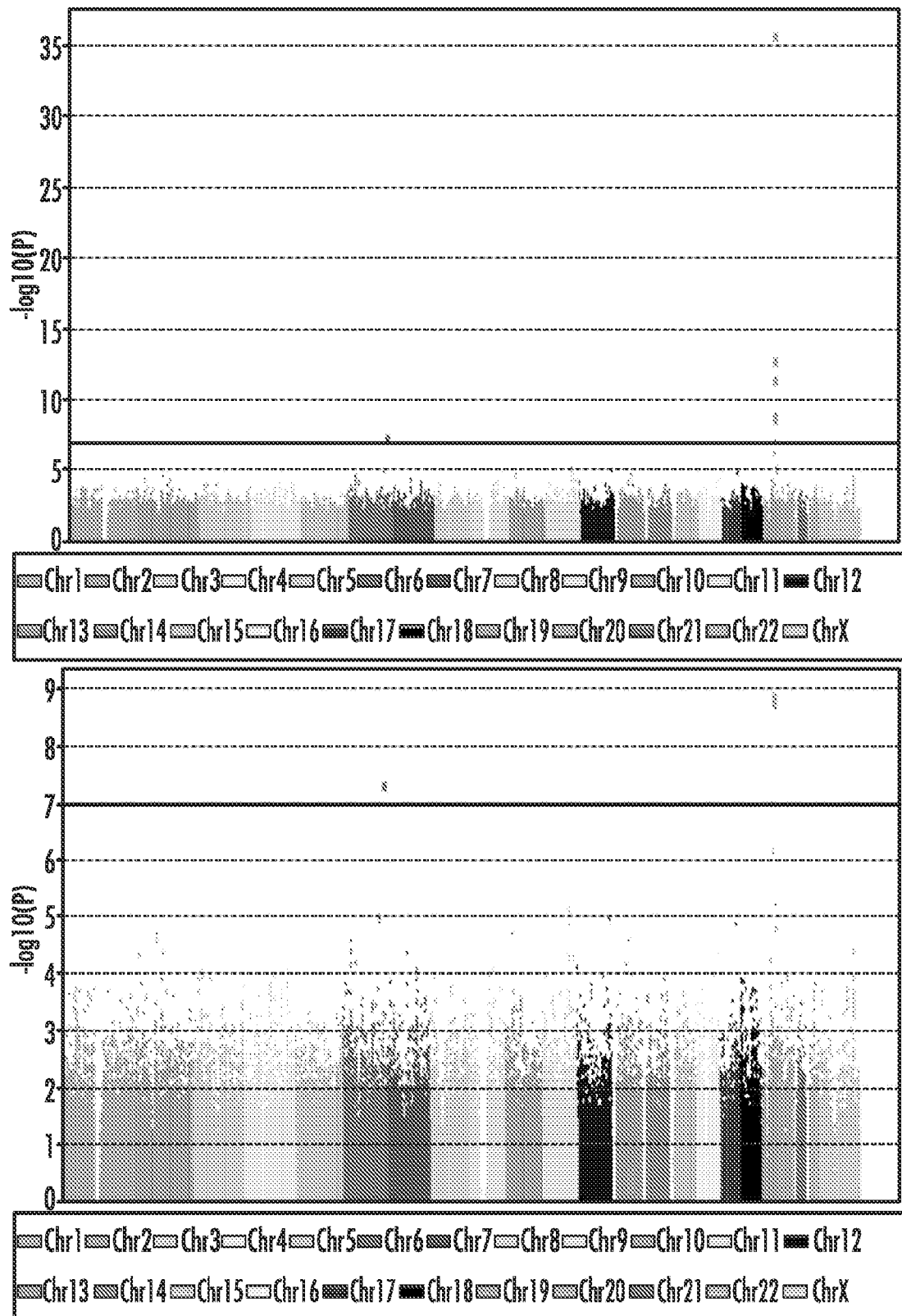


FIG. 2

3/4



4/4

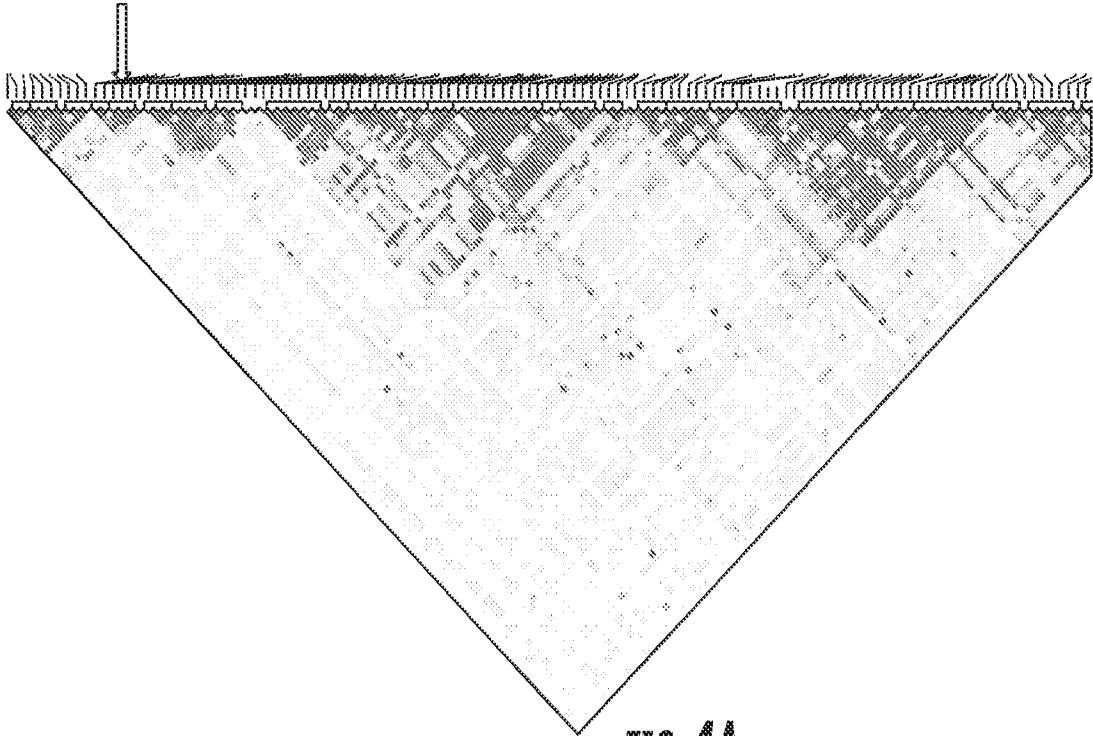


FIG. 4A

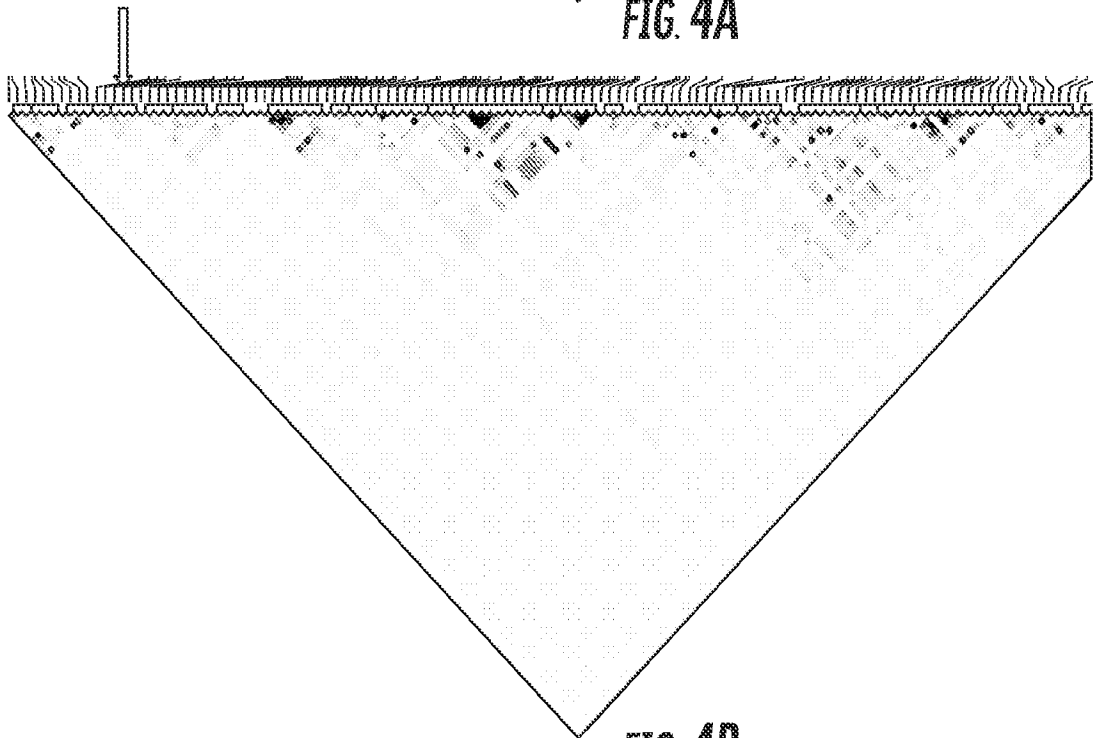


FIG. 4B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 11/30199

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68 (2011.01)

USPC - 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)

USPC: 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)
 PubWEST (PGPB,USPT,USOC,EPAB,JPAB); Google; PubMed: alzheimer's disease, MTHFD1L, SNP, genetic polymorphism, rs11754661, rs2075650, homocysteine

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2008/0286796 A1 (GRUPE et al.) 20 November 2008 (20.11.2008) abstract; para [0012]-[0013], [0032], [0170], [0260]-[0261], [0316], [0427]	1-20
Y	BEECHAM et al. Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. Am. J. Hum. Genet. 2009, 84(1):35-43; abstract; pg 37, para 4; Tables 2, 4	1-20
Y	US 20060171938 A1 (STOCK et al.) 3 August 2006 (03.08.2006) para [0010]	6, 14
A	HAROLD et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat. Genet. 2009, 41(10):1088-1093	1-20
A	US 2008/0213775 A1 (BRODY et al.) 4 September 2008 (04.09.2008)	1-20
X,P	NAJ et al. Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. PLoS Genet. 12 September 2010, 6(9):Article No. e1001130	1-20

☐ Further documents are listed in the continuation of Box C.

* 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 but 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

9 May 2011 (09.05.2011)

Date of mailing of the international search report

20 MAY 2011

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. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
 PCT OSP: 571-272-7774