BMC Neurology

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Genome-wide association reveals genetic effects on human Abeta42 and tau protein levels in cerebrospinal fluids: a case control study

BMC Neurology 2010, **10**:90 doi:10.1186/1471-2377-10-90

Mi-Ryung Han (hanmr@mail.med.upenn.edu)
Gerard D Schellenberg (gerardsc@mail.med.upenn.edu)
Li-San Wang (lswang@mail.med.upenn.edu)
Alzheimer's Disease Neuroimaging Initiative (adnidpc@bu.edu)

ISSN 1471-2377

Article type Research article

Submission date 14 April 2010

Acceptance date 8 October 2010

Publication date 8 October 2010

Article URL http://www.biomedcentral.com/1471-2377/10/90

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

http://www.biomedcentral.com/info/authors/

© 2010 Han *et al.* , licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Genome-wide association reveals genetic effects on human $A\beta_{42}$ and τ protein levels in cerebrospinal fluids: a case control study

Mi-Ryung Han[1,4], Gerard D. Schellenberg[2,3], Li-San Wang[2,3,4§], the Alzheimer's Disease Neuroimaging Initiative

[1]Genomics and Computational Biology Program, University of Pennsylvania, Philadelphia, PA 19104, USA

[2]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[3]Institute on Aging, University of Pennsylvania, Philadelphia, PA 19104, USA

[4]Penn Center for Bioinformatics, University of Pennsylvania, Philadelphia, PA 19104, USA

[§]Corresponding author

Email addresses:

M-RH: hanmr@mail.med.upenn.edu

GDS: gerardsc@mail.med.upenn.edu

L-SW: lswang@mail.med.upenn.edu

Abstract

Background

Alzheimer's disease (AD) is common and highly heritable with many genes and gene variants associated with AD in one or more studies, including APOE ε2/ε3/ε4. However, the genetic

backgrounds for normal cognition, mild cognitive impairment (MCI) and AD in terms of changes in cerebrospinal fluid (CSF) levels of $A\beta_{1\text{-}42}$, T-tau, and P-tau_{181P}, have not been clearly delineated. We carried out a genome-wide association study (GWAS) in order to better define the genetic backgrounds to these three states in relation to CSF levels.

Methods

Subjects were participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI). The GWAS dataset consisted of 818 participants (mainly Caucasian) genotyped using the Illumina Human Genome 610 Quad BeadChips. This sample included 410 subjects (119 Normal, 115 MCI and 176 AD) with measurements of CSF $A\beta_{1\text{-}42}$, T-tau, and P-tau_{181P} Levels. We used PLINK to find genetic associations with the three CSF biomarker levels. Association of each of the 498,205 SNPs was tested using additive, dominant, and general association models while considering APOE genotype and age. Finally, an effort was made to better identify relevant biochemical pathways for associated genes using the ALIGATOR software.

Results

We found that there were some associations with APOE genotype although CSF levels were about the same for each subject group; CSF $A\beta_{1\text{-}42}$ levels decreased with APOE gene dose for each subject group. T-tau levels tended to be higher among AD cases than among normal subjects. From adjusted result using APOE genotype and age as covariates, no SNP was associated with CSF levels among AD subjects. *CYP19A1* 'aromatase' (rs2899472), *NCAM2*, and multiple SNPs located on chromosome 10 near the *ARL5B* gene demonstrated the strongest associations with $A\beta_{1\text{-}42}$ in normal subjects. Two genes found to be near the top SNPs, *CYP19A1* (rs2899472, $p = 1.90\times10^{-7}$) and *NCAM2* (rs1022442, $p = 2.75\times10^{-7}$) have been reported as genetic factors related to the progression of AD from previous studies. In AD subjects, APOE ε2/ε3 and ε2/ε4 genotypes were associated with elevated T-tau levels and ε4/ε4 genotype was

associated with elevated T-tau and P-tau$_{181P}$ levels. Pathway analysis detected several biological pathways implicated in Normal with CSF β-amyloid peptide (Aβ_{1-42}).

Conclusions

Our genome-wide association analysis identified several SNPs as important factors for CSF biomarker. We also provide new evidence for additional candidate genetic risk factors from pathway analysis that can be tested in further studies.

Background

Alzheimer's disease (AD) is the most common cause of dementia and the most prevalent neurodegenerative disorder. An estimated 10 percent of Americans over the age of 65 and half of those over age 85 have AD. More than 4.5 million Americans currently suffer from the disease. In autosomal dominant early-onset Alzheimer's disease (EOAD, age of onset < 60 years), three susceptible genes (*APP*, *PSEN1*, and *PSEN2*) have been identified [1, 2]. Late-onset AD (LOAD) has ~80% heritability, and is strongly associated with apolipoprotein E (*APOE*) [3]. APOE has three major alleles (ε2/ε3/ε4) that have different effects on the risk of LOAD, with ε4 having between 10 and 30 times of risk of developing AD by 75 years of age [4].

In addition, several genetic studies have identified putative susceptible loci and genetic variants, including sortilin-related receptor (*SORL1*) [5, 6], death-associated protein kinase 1 (*DAPK1*) [7], ubiquilin 1 (*UBQLN1*) [8], adenosine triphosphate–binding cassette transporter 1, subfamily A (*ABCA1*) [9], and low-density lipoprotein receptor–related protein 6 (*LRP6*) [10]. Besides these findings, a large meta-analysis from the AlzGene database [11] reported 598 potential AD-susceptibility genes. For the past few years, genome-wide genotyping association studies brought considerable success by reporting new susceptible loci for AD such as Golgi

membrane protein 1 (*GOLM1*) [12-15]. Recently two groups published the two largest LOAD GWAS [16, 17]; Harold *et al*. [16] reported the association of SNPs in clusterin (*CLU*) and phosphatidylinositol binding clathrin assembly protein (*PICALM*), and Lambert *et al*. [17] reported association of clusterin (*CLU*) with LOAD and additionally reported a novel association with complement component (3b/4b) receptor 1 (*CR1*). These new findings have provided valuable insights in the genetics, neuropathologic mechanisms and pathways associated with AD.

The cerebrospinal fluid (CSF) components β-amyloid peptide ($A\beta_{1\text{-}42}$), total tau protein (T-tau) and phosphorylated tau ($P\text{-}tau_{181P}$) are biomarkers for AD and can be used to aid in diagnosis and to predict progression from mild cognitive impairment (MCI) to AD [18, 19]. These biomarkers can potentially be used in future applications to predict the development of MCI in cognitively normal subjects, progression to AD in MCI patients, and to monitor AD progression [20-23]. These biomarkers may also be used to reveal genes that are important in AD pathogenesis. In the present study, we assessed the several putative AD genes associated with CSF biomarkers that were identified from major public GWAS dataset for Alzheimer's disease, the Alzheimer's disease Neuroimaging Initiative (ADNI). This initiative is the most comprehensive effort to identify neuroimaging measures and biomarkers associated with cognitive and functional changes in healthy elderly people and in people who have MCI and AD [24]. The ADNI data is useful for researchers who are searching for genes that contribute to the development of Alzheimer's disease, which currently affects more than 4.5 million people in the United States alone. We also investigated possible associations of CSF biomarkers ($A\beta_{1\text{-}42}$, T-tau, and $P\text{-}tau_{181P}$) with the number of *APOE* ε4 allele, age and *APOE* genotype in order to improve the characterization of the CSF biomarkers and genome-wide SNP genotyping data from the ADNI cohort. This is the first genome-wide association study to use these AD-related

biomarkers to identify genes critical the pathogenesis of AD.

Methods

Sample

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu\ADNI). ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principle Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research -- approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and

200 people with early AD to be followed for 2 years. For up-to-date information see www.adni-info.org.

The GWAS dataset (downloaded from ADNI website in June 2009) consisted of 243 normal, 235 MCI and 340 AD subjects genotyped using the Illumina Human Genome 610 Quad BeadChips. 410 of these subjects (119 Normal, 115 MCI and 176 AD; 247 males and 163 females) have CSF $A\beta_{1\text{-}42}$, T-tau, and P-tau_{181P} levels. Detailed protocols for subject recruitment and biomarker accrual are available at the ADNI website (http://www.adni-info.org/). Demographics, CSF biomarkers, and *APOE* genotype data of the ADNI dataset are summarized in **Table 1, Table 2** and **Additional file 1**.

Genotyping and sample quality control

Quality control for the genotyping data was performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) [25] as follows. 498,205 SNPs were retained after excluding SNPs with Minor Allele Frequency (MAF) < 5%, call rate < 98%, or significant in Hardy-Weinberg Equilibrium test ($p \leq 10^{-6}$). All samples had genotyping call rate > 95% and were retained. We then examined population stratification by visual inspection using the first two dimensions from principal components analysis, using SmartPCA from EIGENSTRAT (http://genepath.med.harvard.edu/~reich/Software.htm) [26, 27]. Self-reported ethnicity and racial identities for ADNI subjects were used to highlight samples in the PCA plot and are summarized in **Additional file 2.** 390 samples were retained after SmartPCA excluded 20 samples as outliers. We computed the top five principal component coordinates using SmartPCA to correct for stratification in association analysis. SmartPCA removed all but one of the Asian samples and retained Black/African Americans (**Additional file 3**); Visual inspection suggests

that the first principal component (PC0), which explains the most variance in the data, separates the Caucasians and non-Caucasians reasonably well (setting threshold PC0 <= 0.01 can exclude all non-Caucasians). Finally, we excluded 52 non-Caucasian samples as outliers; the genomic control variance inflation factor λ was 1.00983, suggesting minimal population admixture in the final sample used for association analysis. We performed association analysis using age and APOE ε4 genotype as covariate and did not incorporate principal components. Quantile-Quantile plots for each of the three test groups with log10-transformed level of three CSF biomarkers (**Additional file 4** and **Additional file 5**) suggested that population stratification having negligible bias on the genetic associations (**Additional file 6**). Finally, for the whole analysis we performed in the following method section, the study sample of 390 individuals with three CSF biomarkers was used after removing 20 outliers. The level of three CSF biomarkers was log10-transformed.

Effect of *APOE* ε4 copy number on CSF biomarker levels

We performed Kruskal-Wallis test of $\log_{10}$-transformed CSF biomarkers ($A\beta_{1\text{-}42}$, T-tau, and P-tau_{181P}) stratified by the number of APOE ε4 alleles and the APOE genotypes across all 390 samples or within each of the three diagnostic groups.

Association testing of SNPs and CSF analysis

We tested SNP association with the three CSF biomarkers ($A\beta_{1\text{-}42}$, T-tau, and P-tau_{181P}) by PLINK using samples and SNP markers passing QC. CSF biomarker levels were $\log_{10}$-transformed so they become normally distributed. The association analysis used a full linear model comprising three genetic effects: additive effects of allele dosage (ADD), dominance deviation from additivity (DOMDEV) (negative means the allele is recessive), and 2-df joint test

of both additive and dominance (GENO_2DF). In addition, we used age (of being recruited by the study) and APOE ε4 genotype (number of APOE ε4 allele; 0, 1, 2) as covariates after removing outliers. To ensure the significance is not due to population stratification, we also incorporated the top five PCA principal components in the linear regression to further control for the population structure, but found the addition has very little effect on the statistical significance.

Gene ontology and E-SNP analysis

We carried out gene ontology analysis of SNP association results using ALIGATOR (Association LIst Go AnnoTatOR) [28], to find gene-ontology terms enriched with significant SNPs. We used p-value cutoff < 10^{-3} for SNPs, 5000 replicate gene lists and 1000 permutations as parameters to run ALIGATOR. We examined the top associated SNPs and examined nearby SNPs in linkage disequilibrium (LD) that are associated with gene expression from published eQTL studies [29-33].

Results and Discussion

Association of SNPs with CSF biomarkers

We summarize the top SNPs (p-value < 10^{-6}) with and without using covariates (APOE genotype and age) in **Table 3.** Without using covariates, we found some genes near the top SNPs, including *CYP19A1* (rs2899472, p = 1.86×10^{-7}) and *TOMM40* (rs2075650, p = 3.03×10^{-7}) from $A\beta_{1\text{-}42}$ in normal subject. Several genetic studies have identified those genes as putative susceptible loci and genetic variants associated with Alzheimer's disease [34, 35]. However, close examination of nearby SNPs showed rs2899472 and rs2075650 were not supported by

nearby SNPs in LD (nearby SNPs in LD are all non-significant). Because APOE genotypes are strongly associated with AD and TOMM40 is physically close to APOE, we focused on SNPs from adjusted results that consider APOE genotype and age. Here, we found 10 SNPs significantly associated with CSF biomarker level of $A\beta_{1\text{-}42}$, 3 SNPs from T-tau, and 2 SNPs from P-tau_{181P} at 10^{-6} significance level in normal subjects (**Figure 1**). In addition, we found 1 SNP significantly associated with CSF biomarker level of $A\beta_{1\text{-}42}$, 3 SNPs from T-tau, and 2 SNPs from P-tau_{181P} at 10^{-6} significance level in MCI subjects (**Figure 2**). No SNPs were found at 10^{-6} significance level in AD subjects (**Figure 3**). For normal subjects, we found genes near the top SNPs, included *CYP19A1* (rs2899472, $p = 1.90\times10^{-7}$), *NCAM2* (rs1022442, $p = 2.75\times10^{-7}$) for $A\beta_{1\text{-}42}$ association and *UPP2* (rs2074955, $p = 2.07\times10^{-7}$) for P-tau_{181p} association. Again, close examination of SNPs in LD with rs2899472 (CYP19A1) did not support rs2899472. The SNP rs1022442 was in close to genome-wide significance, supported by nearby SNPs (**Figure 4**), and *NCAM2* (neural cell adhesion molecule 2) gene was reported as a genetic factor related to the progression of AD in the Japanese population [36]. The $A\beta_{1\text{-}42}$ level grouped by the SNP rs1022442 genotype over all three cohorts (normal, MCI and AD) supports our finding (**Figure 5**). Boxplots of $A\beta_{1\text{-}42}$ levels in normal subjects stratified by rs1022442 genotype showed significant differences between AA, AB and BB. Previous study indicated an increased risk associated with rs2899472 in AD patients, which was amplified in APOE ε4 carriers in their study [35]. For MCI subjects, we found several genes near the top SNPs, included *FLJ21511* (rs2768975; $p = 1.96\times10^{-7}$, rs6850199; $p = 3.18\times10^{-7}$) by T-tau association and *CHN2* (rs121724, $p = 1.45\times10^{-7}$), *MTUS1* (rs7842088, $p = 2.12\times10^{-7}$) by P-tau_{181p} association.

Association of *APOE* with CSF biomarkers

Previous studies suggesting that CSF $A\beta_{1\text{-}42}$ and T-tau levels are correlated with the number of

APOE ε4 alleles [37]. We analyzed the distribution of levels stratified by diagnosis and the number of *APOE* ε4 in the ADNI cohort after QC **(Figure 6)** and reached the same conclusion. In the AD group, $A\beta_{1-42}$ level was inversely correlated with *APOE* ε4 allele dose. The *APOE* e4 was not associated with T-tau or P-tau_{181P} levels. Analysis of *APOE* genotypes showed that ε4/ε4 is associated with CSF biomarker level of $A\beta_{1-42}$, T-tau and P-tau_{181P} **(Figure 7)**.

Pathway analysis of CSF biomarkers

We ran ALIGATOR to identify top gene ontology terms associated with genes containing SNPs with higher statistical significance, and summarized the results in **Additional file 7.** We found cerebral cortex development, methionine metabolic process, actinin binding, and pallium development to be among the most significant gene ontology terms associated with CSF biomarker level of $A\beta_{1-42}$ in normal subjects. Elevation in Aβ in the cerebral cortex has been implicated in the pathophysiology of AD but its mechanism of action is unknown [38]. It has been known that mammals have a fully developed cortex, but the structure it evolved from pallium which is present in all vertebrates as well as the most primitive ones [39]. The medial pallium forms the precursor of the hippocampus. Since hippocampal disruption is one of the earliest signs for AD, pallium development might be involved in the pathophysiology AD.

E-SNP analysis of CSF biomarkers

We collected eSNPs (SNPs known to be associated with the expression level of some genes) in several published expression quantitative linkage (eQTL) studies, and examined regions around significant SNPs ($p<10^{-4}$) (**Additional file 8**). We found significant SNPs from Dixon *et al.* [29], Stranger *et al*. [31], and Gibbs *et al*. [33], but we could not find any top SNPs that are associated with gene expression from other papers [30, 32]. Dixon *et al.* [29] used lymphoblastoid cell lines

(LCLs) derived from children both with and without asthma.

NCAM2 (Neural Cell Adhesion Molecule 2)

We obtained strong evidence that NCAM2 genotypes are associated with $A\beta_{1\text{-}42}$ levels. *NCAM2* (Ensemble: ENSG00000154654, OMIM: 602040;) is a 541Kb gene at 21q21.1 with no known alternative spicing forms. Novartis SymAtlas human tissue survey shows the gene (Affymetrix probeset ID 205669_at) is ubiquitously expressed and is highly expressed in cardiac myocytes, blood cells, and appendix. Among neuronal tissues *NCAM2* has higher expression levels in prefrontal cortex, superior cervical ganglion, and hypothalamus. The transcript (Ensembl: ENST00000400546) consists of 18 exons and encodes a 93k-Da, 835-residue plasma membrane protein (NP_004531.2). The *NCAM2* protein architecture includes 5 IgC2 (Immunoglobulin C-2 type) domains followed by two FN3 (fibronectin type 3) domains and a transmembrane domain. The gene is conserved in chimpanzee, dog, cow, mouse, rat, chicken, zebrafish, and fruit fly; the eight-domain protein architecture is also conserved in all these organisms except for cow which has only four IgC2 domains. Little is known about *NCAM2* except that the protein interacts with prion protein [40] and estrogen receptor 1 (ESR1) [41], is involved in neuron adhesion and fasciculation of neurons, and may be involved in AD [36], prion disease, and Down syndrome [42, 43].

TOMM40 (translocase of outer mitochondrial membrane 40 homolog (yeast)) and CYP19A1 (Cytochrome P450, family 19, subfamily A, polypeptide 1)

It is interesting to note that we found no significant association in *APOE* (rs769451, chr19: 50102751, p = 0.6682 for association with $A\beta_{1\text{-}42}$ level in normal subjects), but a SNP with strong association in the nearby *TOMM40* (intronic SNP rs2075650, chr19:50087459, p =

3.03×10^{-7} for association with $A\beta_{1-42}$ level in normal subjects), when age and the number of APOE e4 alleles were not included in the regression. The *TOMM40* gene is related to how easily molecules can get into and out of the surface of the mitochondria, the energy center of cells. This gene is a transporter of proteins across the mitochondrial membrane, and Sortillin-related receptor, which functions to partition amyloid precursor protein away from β-secretase and γ-secretase [44]. This is consistent with observations that levels are reduced in the brains of patients with Alzheimer's disease and MCI [44-46]. The *TOMM40* gene has been reported in numerous studies in the study of AD genetics; for example, Yu *et al*. [47] reported possibility that loci in the *TOMM40* gene may have a less effect on the risk for LOAD in Caucasians [47], and recently Roses *et al*. [48] found evidence supporting a poly-T polymorphism (rs10524523, chr19:50094889) in *TOMM40* affecting the AD age of onset in two independent clinical cohorts. The potential association of *TOMM40* and $A\beta_{1-42}$ may be how the gene affects the risk and onset age of AD and should be further investigated.

The *CYP19A1* gene is localized on chromosome 15q21.2 and spans 123 kb. This gene encodes a member of the cytochrome P450 superfamily of enzymes. Cytochrome P450 aromatase is an enzyme that catalyses the conversion of androgens, such as testosterone, to oestrogens, which act as sex steroid hormones but also function during growth and differentiation. There are high levels of expression in both the gonads and the brain [35]. Huang *et al*. [35] indicated an increased risk associated with SNP rs2899472 in the total number of AD patients, which was amplified in APOE ε4 carriers.

Conclusions

Our analysis of the ADNI genome-wide association study identified several putative loci that are in genetic association with $A\beta_{1-42}$, T-tau and P-tau_{181P} levels in cerebrospinal fluids. In particular an intronic SNP rs1022442 of gene *NCAM2* is close to genome-wide significance in association with $A\beta_{1-42}$ in normal subjects. Although the gene is poorly characterized in the literature, prior studies have implicated roles of *NCAM2* in prion disease, Down syndrome, and AD. Our findings suggest *NCAM2* could be part of the pathway on the pathogenesis of senile plaques in human brains with AD.

With only 119 normal subjects and 410 overall, the GWAS dataset is clearly underpowered. The most significant associations were identified using normal subjects since the variances of the CSF biomarker levels are much smaller in MCI and AD subjects due to dementia. Nonetheless, increasing the number of CSF biomarker measurements is challenging especially for normal subjects. An alternative that will substantially increase the sample size is to examine protein levels in blood instead of CSF, given levels of these proteins in blood are informative about AD pathology or prognosis [49].

Our analysis clearly demonstrates that quantitative trait linkage of biomarkers via genome-wide screening can reveal additional insights into the mechanism that connects known genetic factors to the disease [50]. Moving along this path requires further efforts by the research community towards larger sample size and accrual of additional biomarkers in AD and other neurodegenerative disorders.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

M-RH analyzed and interpreted the data, and wrote the manuscript. L-SW conceived the study, analyzed and interpreted the data, and drafted the manuscript. GDS assisted in data analysis and interpretation, and made critical revisions to the manuscript. All authors read and approved the final manuscript.

Acknowledgements

M-RH, GDS, and L-SW are supported by the Alzheimer's Disease Genetic Consortium (U01 AG032984-01 and RC2-AG036528-01). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., and Wyeth, as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH

grants P30 AG010129, K01 AG030514, and the Dana Foundation.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu\ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. Complete listing of ADNI investigators is available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf.

References

1. Raux G, Guyant-Marechal L, Martin C, Bou J, Penet C, Brice A, Hannequin D, Frebourg T, Campion D: **Molecular diagnosis of autosomal dominant early onset Alzheimer's disease: an update**. *J Med Genet* 2005, **42**(10):793-795.
2. Kowalska A, Pruchnik-Wolinska D, Florczak J, Modestowicz R, Szczech J, Kozubski W, Rossa G, Wender M: **Genetic study of familial cases of Alzheimer's disease**. *Acta Biochim Pol* 2004, **51**(1):245-252.
3. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL: **Role of genes and environments for explaining Alzheimer disease**. *Arch Gen Psychiatry* 2006, **63**(2):168-174.
4. Sando SB, Melquist S, Cannon A, Hutton ML, Sletvold O, Saltvedt I, White LR, Lydersen S, Aasly JO: **APOE epsilon 4 lowers age at onset and is a high risk factor for Alzheimer's disease; a case control study from central Norway**. *BMC Neurol* 2008, **8**:9.
5. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H *et al*: **The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease**. *Nat Genet* 2007, **39**(2):168-177.
6. Lee JH, Cheng R, Schupf N, Manly J, Lantigua R, Stern Y, Rogaeva E, Wakutani Y, Farrer L, St George-Hyslop P *et al*: **The association between genetic variants in SORL1 and Alzheimer disease in an urban, multiethnic, community-based cohort**. *Arch Neurol* 2007, **64**(4):501-506.
7. Li Y, Grupe A, Rowland C, Nowotny P, Kauwe JS, Smemo S, Hinrichs A, Tacey K, Toombs TA, Kwok S *et al*: **DAPK1 variants are associated with Alzheimer's disease and allele-specific expression**. *Hum Mol Genet* 2006, **15**(17):2560-2568.
8. Bertram L, Hiltunen M, Parkinson M, Ingelsson M, Lange C, Ramasamy K, Mullin K, Menon R, Sampson AJ, Hsiao MY *et al*: **Family-based association between Alzheimer's disease and variants in UBQLN1**. *N Engl J Med* 2005, **352**(9):884-894.
9. Sundar PD, Feingold E, Minster RL, DeKosky ST, Kamboh MI: **Gender-specific association of ATP-binding cassette transporter 1 (ABCA1) polymorphisms with the risk of late-onset Alzheimer's disease**. *Neurobiol Aging* 2007, **28**(6):856-862.
10. De Ferrari GV, Papassotiropoulos A, Biechele T, Wavrant De-Vrieze F, Avila ME, Major MB, Myers A, Saez K, Henriquez JP, Zhao A *et al*: **Common genetic variation within the low-density lipoprotein receptor-related protein 6 and late-onset Alzheimer's disease**.

Proc Natl Acad Sci U S A 2007, **104**(22):9434-9439.
11. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE: **Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database**. *Nat Genet* 2007, **39**(1):17-23.
12. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M *et al*: **Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease**. *Arch Neurol* 2008, **65**(1):45-53.
13. Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A, Jehu L, Segurado R, Stone D, Schadt E *et al*: **Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants**. *Hum Mol Genet* 2007, **16**(8):865-873.
14. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L *et al*: **A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease**. *J Clin Psychiatry* 2007, **68**(4):613-618.
15. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ *et al*: **GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers**. *Neuron* 2007, **54**(5):713-720.
16. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A *et al*: **Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease**. *Nature Genetics* 2009, **41**(10):1088-U1061.
17. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B *et al*: **Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease**. *Nature Genetics* 2009, **41**(10):1094-U1068.
18. Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, Herukka SK, van der Flier WM, Blankenstein MA, Ewers M *et al*: **CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment**. *JAMA* 2009, **302**(4):385-393.
19. Engelborghs S, Sleegers K, Cras P, Brouwers N, Serneels S, De Leenheir E, Martin JJ, Vanmechelen E, Van Broeckhoven C, De Deyn PP: **No association of CSF biomarkers with APOEepsilon4, plaque and tangle burden in definite Alzheimer's disease**. *Brain* 2007, **130**(Pt 9):2320-2326.
20. Hansson O, Zetterberg H, Buchlave P, Londos E, Blennow K, Minthon L: **Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. (vol 5, pg 228, 2006)**. *Lancet Neurol* 2006, **5**(4):293-293.
21. Andreasen N, Blennow K: **CSF biomarkers for mild cognitive impairment and early Alzheimer's disease**. *Clin Neurol Neurosur* 2005, **107**(3):165-173.
22. Zetterberg H, Wahlund LO, Blennow K: **Cerebrospinal fluid markers for prediction of Alzheimer's disease**. *Neurosci Lett* 2003, **352**(1):67-69.
23. Wahlund LO, Blennow K: **Cerebrospinal fluid biomarkers for disease stage and intensity in cognitively impaired patients**. *Neurosci Lett* 2003, **339**(2):99-102.
24. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, Trojanowski JQ, Toga AW, Beckett L: **The Alzheimer's disease neuroimaging initiative**. *Neuroimaging Clin N Am* 2005, **15**(4):869-877, xi-xii.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al*: **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *Am J Hum Genet* 2007, **81**(3):559-575.

26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies**. *Nat Genet* 2006, **38**(8):904-909.
27. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis**. *PLoS Genet* 2006, **2**(12):e190.
28. Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, Owen MJ, O'Donovan MC, Craddock N: **Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder**. *Am J Hum Genet* 2009, **85**(1):13-24.
29. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M *et al*: **A genome-wide association study of global gene expression**. *Nat Genet* 2007, **39**(10):1202-1207.
30. Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, Welsh-Bohmer KA, Hulette CM, Denny TN, Goldstein DB: **Tissue-specific genetic control of splicing: implications for the study of complex traits**. *PLoS Biol* 2008, **6**(12):e1.
31. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D *et al*: **Population genomics of human gene expression**. *Nat Genet* 2007, **39**(10):1217-1224.
32. Verlaan DJ, Ge B, Grundberg E, Hoberman R, Lam KC, Koka V, Dias J, Gurd S, Martin NW, Mallmin H *et al*: **Targeted screening of cis-regulatory variation in human haplotypes**. *Genome Res* 2009, **19**(1):118-127.
33. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J *et al*: **Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain**. *Plos Genetics* 2010, **6**(5):-.
34. Sager KL, Wuu J, Leurgans SE, Rees HD, Gearing M, Mufson EJ, Levey AI, Lah JJ: **Neuronal LR11/sorLA expression is reduced in mild cognitive impairment**. *Ann Neurol* 2007, **62**(6):640-647.
35. Huang R, Poduslo SE: **CYP19 haplotypes increase risk for Alzheimer's disease**. *J Med Genet* 2006, **43**(8):e42.
36. Kimura R, Kamino K, Yamamoto M, Nuripa A, Kida T, Kazui H, Hashimoto R, Tanaka T, Kudo T, Yamagata H *et al*: **The DYRK1A gene, encoded in chromosome 21 Down syndrome critical region, bridges between beta-amyloid production and tau phosphorylation in Alzheimer disease**. *Hum Mol Genet* 2007, **16**(1):15-23.
37. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P *et al*: **Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects**. *Ann Neurol* 2009, **65**(4):403-413.
38. Yankner BA, Duffy LK, Kirschner DA: **Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides**. *Science* 1990, **250**(4978):279-282.
39. Aboitiz F, Morales D, Montiel J: **The evolutionary origin of the mammalian isocortex: towards an integrated developmental and functional approach**. *Behav Brain Sci* 2003, **26**(5):535-552; discussion 552-585.
40. Schmitt-Ulms G, Hansen K, Liu J, Cowdrey C, Yang J, DeArmond SJ, Cohen FE, Prusiner SB, Baldwin MA: **Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues**. *Nat Biotechnol* 2004, **22**(6):724-731.
41. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR *et al*: **Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1**. *Cell* 2005, **122**(1):33-43.
42. Gratacos M, Costas J, de Cid R, Bayes M, Gonzalez JR, Baca-Garcia E, de Diego Y, Fernandez-Aranda F, Fernandez-Piqueras J, Guitart M *et al*: **Identification of new putative**

susceptibility genes for several psychiatric disorders by association analysis of regulatory and non-synonymous SNPs of 306 genes involved in neurotransmission and neurodevelopment. *Am J Med Genet B Neuropsychiatr Genet* 2009, **150B**(6):808-816.
43. Paoloni-Giacobino A, Chen H, Antonarakis SE: **Cloning of a novel human neural cell adhesion molecule gene (NCAM2) that maps to chromosome region 21q21 and is potentially involved in Down syndrome**. *Genomics* 1997, **43**(1):43-51.
44. Querfurth HW, LaFerla FM: **Alzheimer's disease**. *N Engl J Med*, **362**(4):329-344.
45. Sager KL, Wuu J, Leurgans SE, Rees HD, Gearing M, Mufson EJ, Levey AI, Lah JJ: **Neuronal LR11/SorLA expression is reduced in mild cognitive impairment**. *Annals of Neurology* 2007, **62**(6):640-647.
46. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H *et al*: **The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease**. *Nature Genetics* 2007, **39**(2):168-177.
47. Yu CE, Seltman H, Peskind ER, Galloway N, Zhou PX, Rosenthal E, Wijsman EM, Tsuang DW, Devlin B, Schellenberg GD: **Comprehensive analysis of APOE and selected proximate markers for late-onset Alzheimer's disease: Patterns of linkage disequilibrium and disease/marker association**. *Genomics* 2007, **89**(6):655-665.
48. Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, Huentelman MJ, Welsh-Bohmer KA, Reiman EM: **A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease**. *Pharmacogenomics J* 2010, **10**(5):375-384.
49. Ray S, Britschgi M, Herbert C, Takeda-Uchimura Y, Boxer A, Blennow K, Friedman LF, Galasko DR, Jutel M, Karydas A *et al*: **Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins**. *Nat Med* 2007, **13**(11):1359-1362.
50. Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E *et al*: **Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease**. *PLoS One* 2009, **4**(8):e6501.

Figures

Figure 1. Manhattan plots of the quantitative trait (CSF biomarkers: T-tau, P-tau$_{181P}$ and Aβ_{1-42}) genome wide association analysis in normal subjects. Colors on x-axis indicate an autosomal chromosome (from chromosome 1 to chromosome 22). The y-axis indicates p-values (–$\log_{10}$(observed p-values)). Red arrows indicate rs2899472 on chromosome 15 and blue arrow indicates rs1022442 on chromosome 21.

Figure 2. Manhattan plots of the quantitative trait (CSF biomarkers: T-tau, P-tau$_{181P}$ and Aβ_{1-42}) genome wide association analysis in MCI subjects. Colors on x-axis indicate an autosomal

chromosome (from chromosome 1 to chromosome 22). The y-axis indicates p-values ($-\log_{10}$(observed p-values)).

Figure 3. Manhattan plots of the quantitative trait (CSF biomarkers: T-tau, P-tau$_{181P}$ and Aβ_{1-42}) genome wide association analysis in AD subjects. Colors on x-axis indicate an autosomal chromosome (from chromosome 1 to chromosome 22). The y-axis indicates p-values ($-\log_{10}$(observed p-values)).

Figure 4. Upper image shows genes in linkage disequilibrium with the SNP rs1022442 and lower plot shows the distribution of SNP rs1022442 and nearby SNPs. In lower plot, the x-axis indicates loci on chromosome 21 and y-axis indicates p-values ($-\log_{10}$(observed GENO_2DF p-values)). Green arrow indicates rs1022442 ($p = 2.75\times10^{-7}$) on loci 21277717 and pink arrow indicates rs2826629 ($p = 8.40\times10^{-5}$) on loci 21265887.

Figure 5. Boxplots of Aβ_{1-42} levels in normal, MCI and AD subjects stratified by rs1022442 genotype. The x-axis indicates AA, AB and BB respectively.

Figure 6. Boxplots of the *APOE* ε4 copy number with CSF biomarkers in normal, MCI and AD subjects. The x-axis indicates number of *APOE* ε4 alleles and y-axis indicates CSF biomarkers. P-values are produced by the Kruskal-Wallis test.

Figure 7. Boxplots of the *APOE* genotype with CSF biomarkers in normal, MCI and AD subjects. The x-axis indicates five different *APOE* genotypes and y-axis indicates CSF biomarkers. P-values are produced by the Kruskal-Wallis test.

Tables

Table 1. Demographic, clinical and biomarker data for each subject group after removing 20 outliers (n = 390: Normal, MCI and AD)

Normal (n = 109)	**Mean**
Age (year)	71.7
CSF $A\beta_{1-42}$ levels (pg/ml)	205.8
CSF P-tau_{181P} levels (pg/ml)	25.4
CSF T-tau levels (pg/ml)	71.8
Last MMSE	29
	N
Male	59
Female	50
MCI (n = 109)	**Mean**
Age (year)	70
CSF $A\beta_{1-42}$ levels (pg/ml)	170.4
CSF P-tau_{181P} levels (pg/ml)	33.7
CSF T-tau level (pg/ml)	99.5
Last MMSE	26.5
	N
Male	72
Female	37
AD (n = 172)	**Mean**
Age (year)	71
CSF $A\beta_{1-42}$ levels (pg/ml)	144.7
CSF P-tau_{181P} levels (pg/ml)	40.3
CSF T-tau levels (pg/ml)	115.3
Last MMSE	21.1
	N
Male	105
Female	67

Abbreviations: CSF, cereb al spinal fluid; MMSE, Mini Mental State Examination

Table 2. *APOE* genotype data for each subject group (Normal, MCI and AD)

***APOE* genotype data (after removing outliers: n = 390)**

Additional files

Additional file 1

Title: Demographic, clinical and biomarker data for each subject group before removing 20 outliers (n= 410: Normal, MCI and AD)

Description: Word DOC table

Additional file 2

Title: Ethnic and Racial summary for ADNI data

Description: Word DOC table

Additional file 3

Title: Population stratification assessments using PCA. Colors indicate different ethnic (A) and racial (B) groups. The x-axis indicates the first principal component (PC0) and y-axis indicates the second principal component (PC1).

Description: Word DOC figure

Additional file 4

Title: Log10(CSF levels) for each subject group after removing 20 outliers (n = 390: Normal, MCI and AD)

Description: Word DOC table

Additional file 5

Title: Log10(CSF levels) for each subject group before removing 20 outliers (n = 410: Normal, MCI and AD)

Description: Word DOC table

Additional file 6

Title: Quantile-Quantile plots of three CSF biomarkers in case/control (left: GENO_2DF test, middle: DOMDEV test, right: ADD test). The x-axis indicates quantile values (–log10(quantile values)) and y-axis indicates p-values (–log10(observed p-values)).

Description: Word DOC figure

Additional file 7

Title: Enriched GO Categories (Top 10 GO categories with p-value < 10^{-2})

Description: Word DOC table

Additional file 8

Title: Significant SNPs (p-value < 10^{-4}) known to be associated with gene expression in published eQTL studies.

Description: Word DOC table

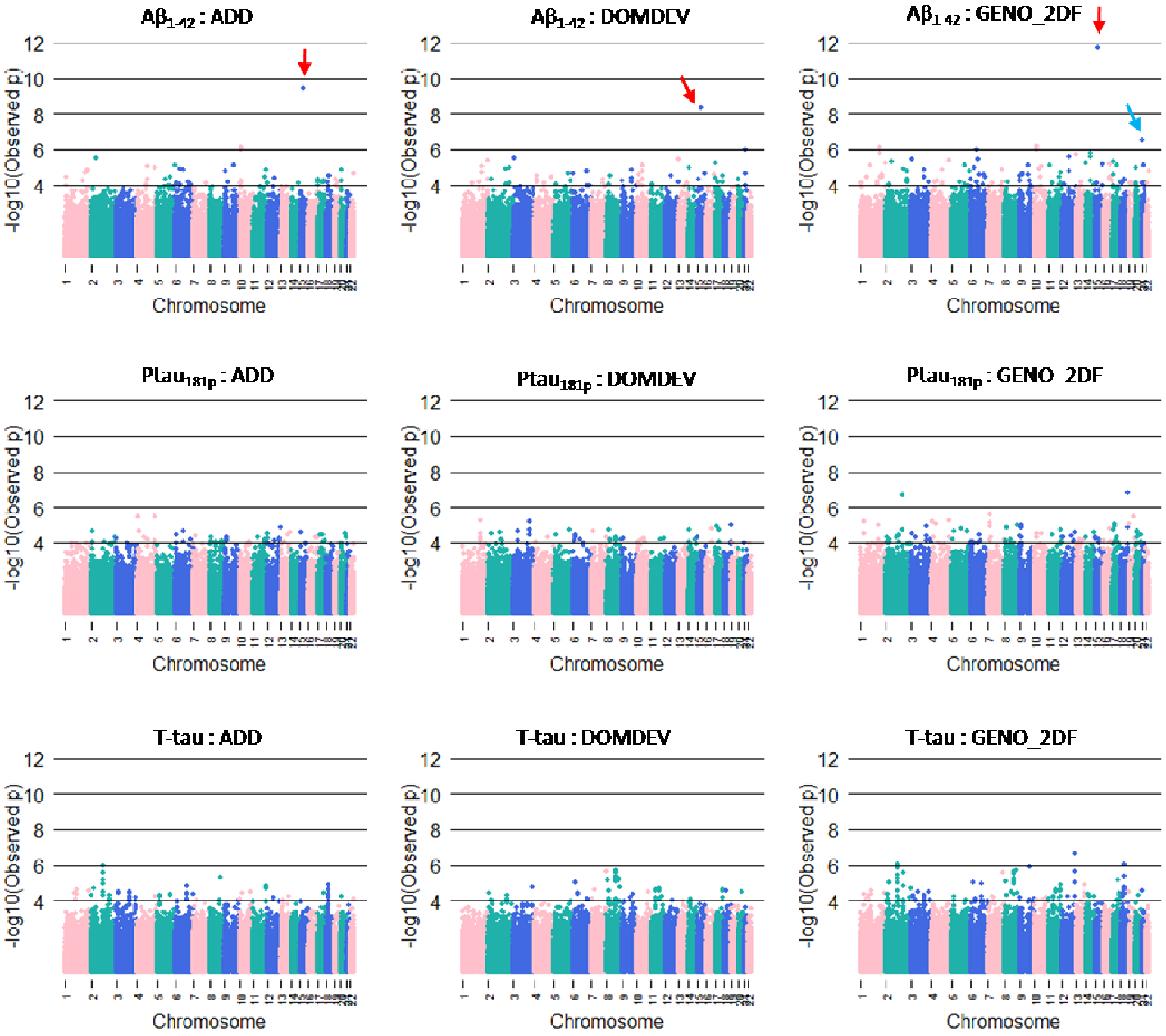
Aβ1-42 : ADD
Aβ1-42 : DOMDEV
Aβ1-42 : GENO_2DF
Ptau181p : ADD
Ptau181p : DOMDEV
Ptau181p : GENO_2DF
T-tau : ADD
T-tau : DOMDEV
T-tau : GENO_2DF
-log10(Observed p)
Chromosome

Figure 1

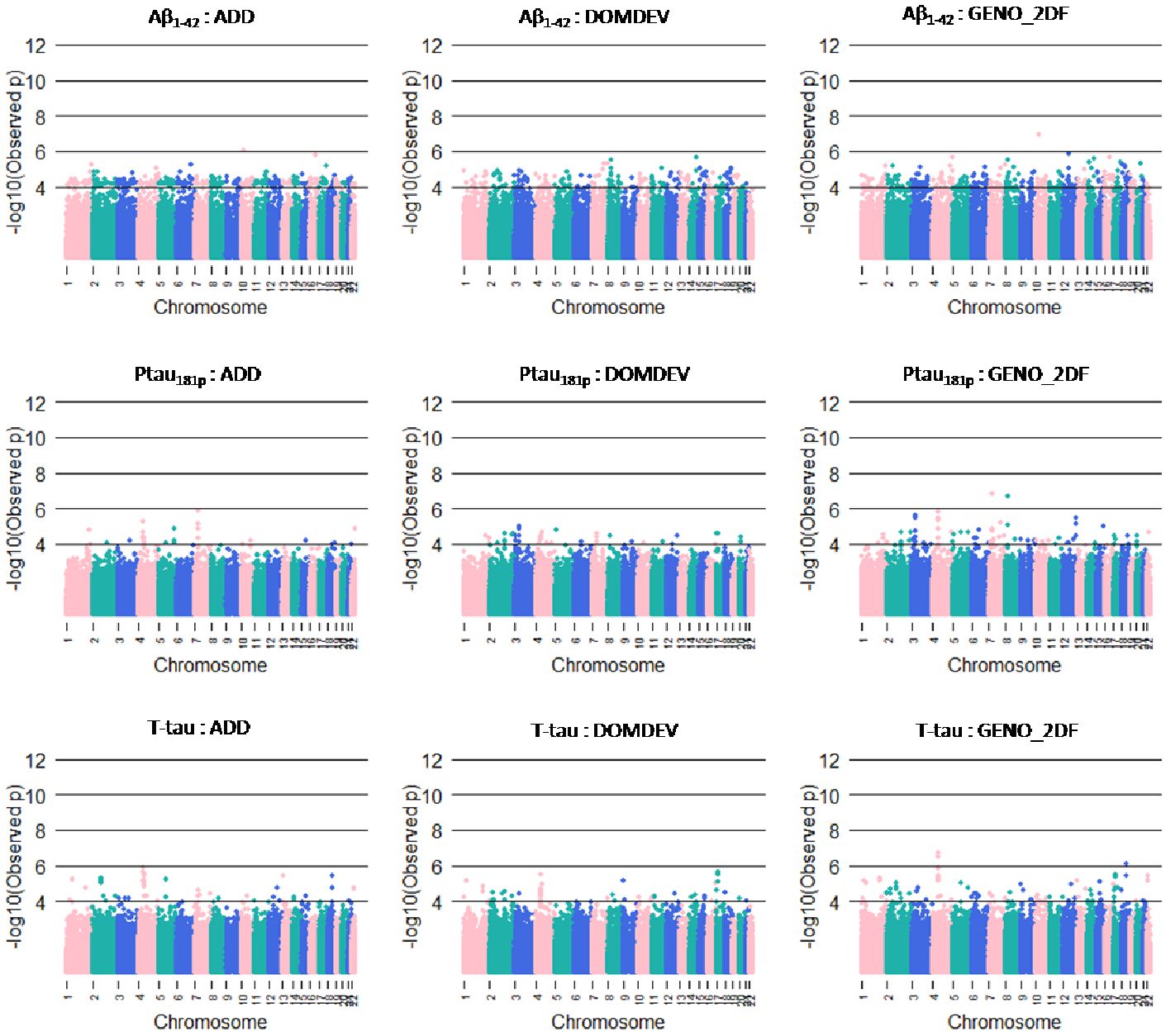
Aβ1-42 : ADD
Aβ1-42 : DOMDEV
Aβ1-42 : GENO_2DF
Ptau181p : ADD
Ptau181p : DOMDEV
Ptau181p : GENO_2DF
T-tau : ADD
T-tau : DOMDEV
T-tau : GENO_2DF
-log10(Observed p)
12
10
8
6
4
Chromosome

Figure 2

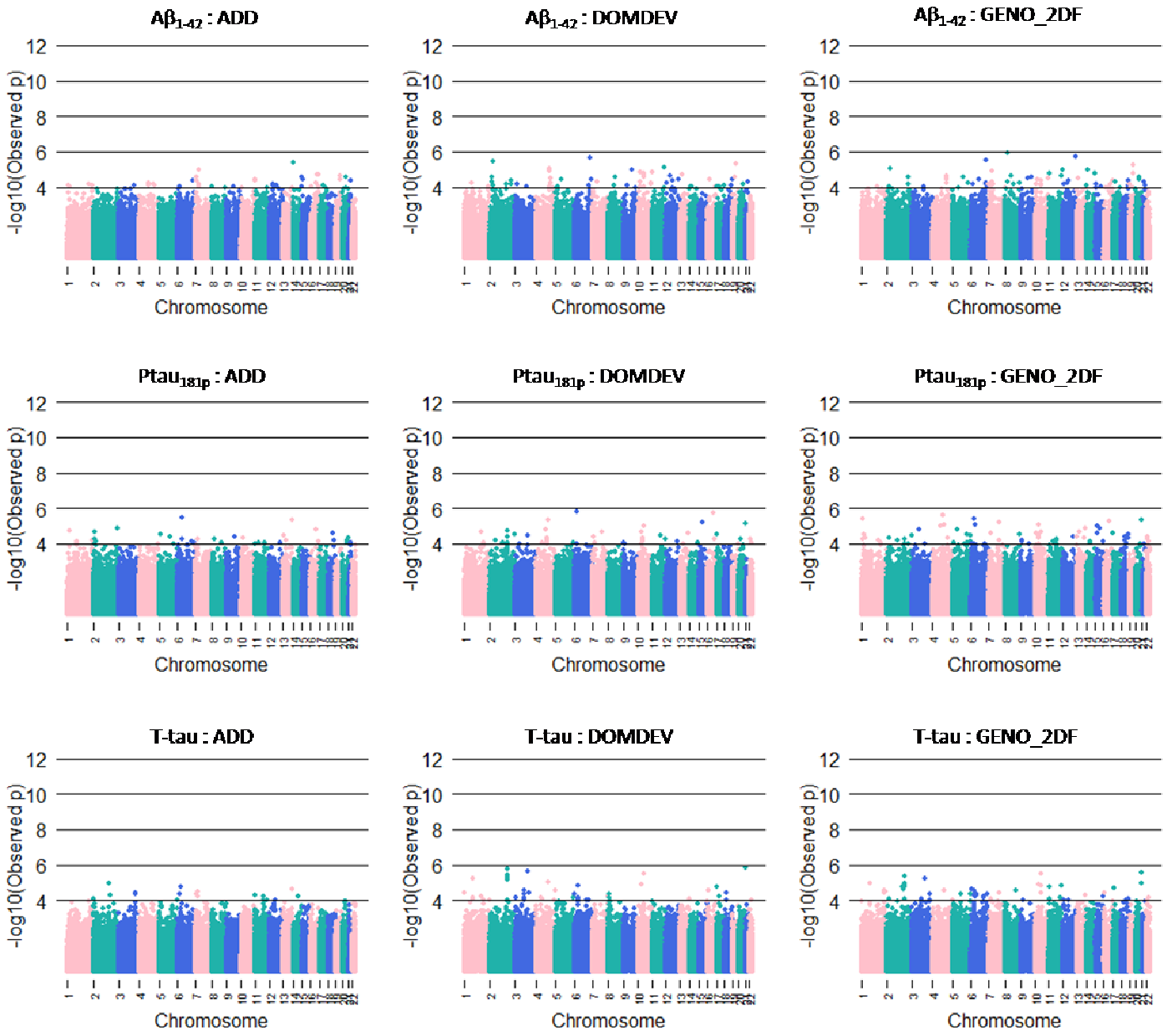
Aβ1-42 : ADD
Aβ1-42 : DOMDEV
Aβ1-42 : GENO_2DF
Ptau181p : ADD
Ptau181p : DOMDEV
Ptau181p : GENO_2DF
T-tau : ADD
T-tau : DOMDEV
T-tau : GENO_2DF
-log10(Observed p)
12
10
8
6
4
Chromosome

Figure 3

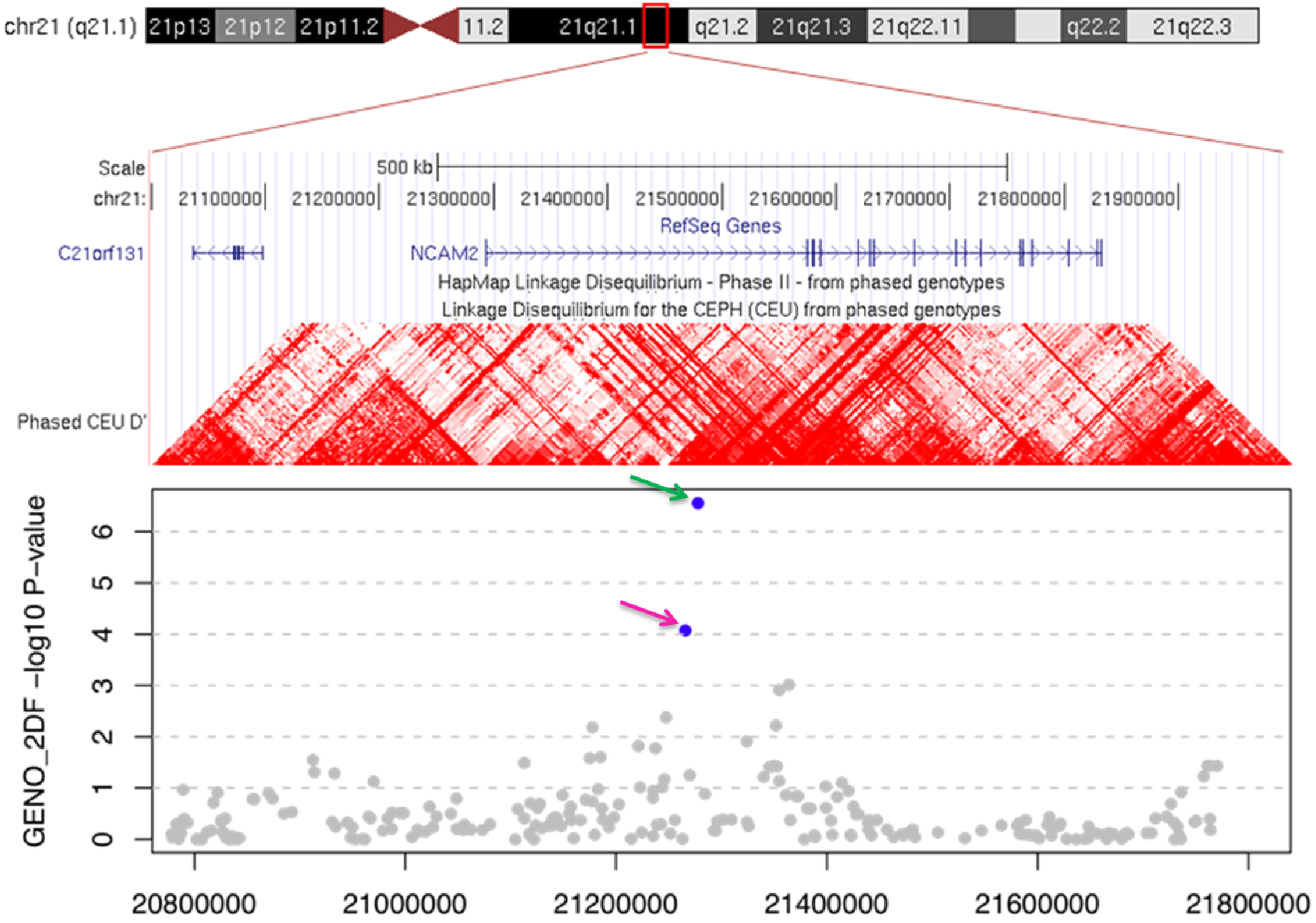
chr21 (q21.1)
21p13
21p12
21p11.2
11.2
21q21.1
q21.2
21q21.3
21q22.11
q22.2
21q22.3
Scale
500 kb
chr21:
21100000
21200000
21300000
21400000
21500000
21600000
21700000
21800000
21900000
RefSeq Genes
C21orf131
NCAM2
HapMap Linkage Disequilibrium - Phase II - from phased genotypes
Linkage Disequilibrium for the CEPH (CEU) from phased genotypes
Phased CEU D'
GENO_2DF −log10 P−value
0
1
2
3
4
5
6
20800000
21000000
21200000
21400000
21600000
21800000

Figure 4

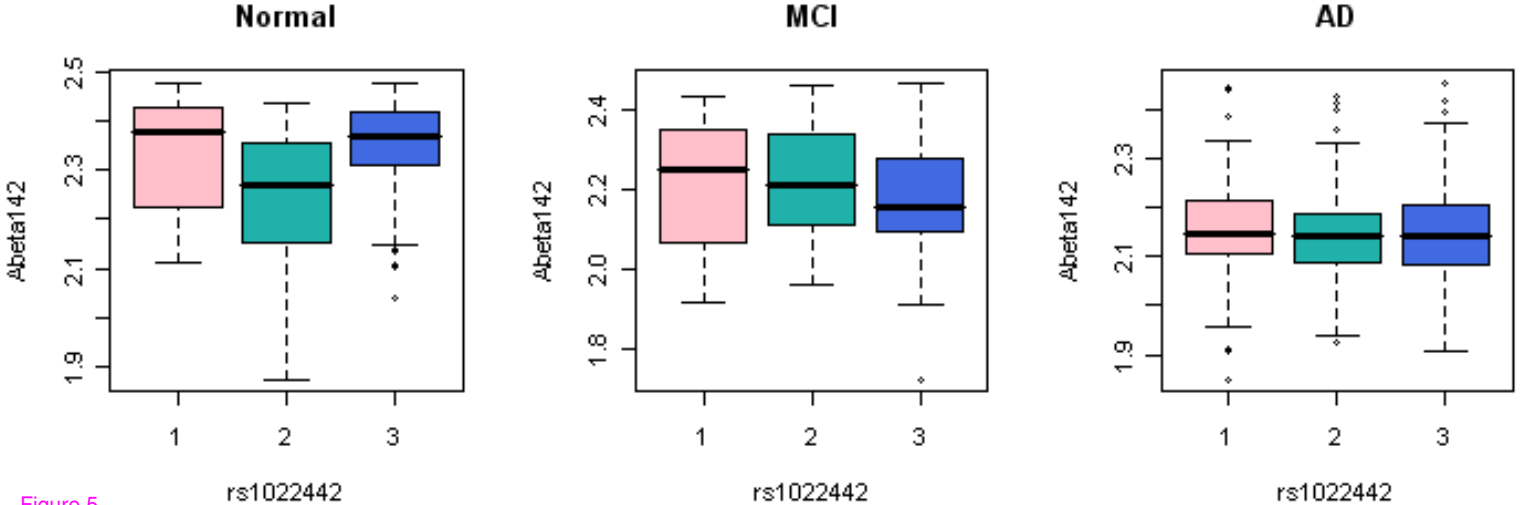

Figure 5

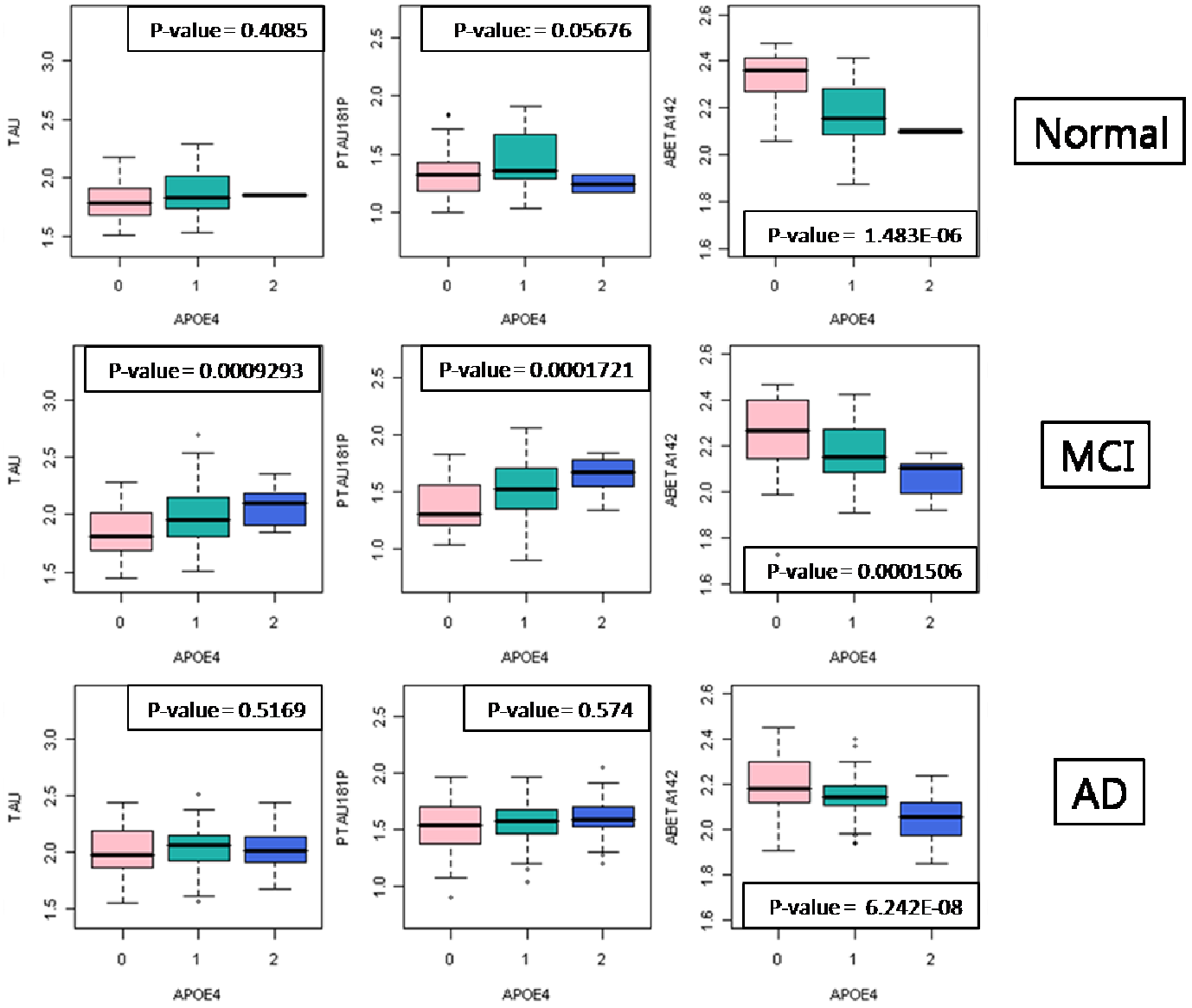
P-value = 0.4085
P-value: = 0.05676
P-value = 1.483E-06
Normal
P-value = 0.0009293
P-value = 0.0001721
P-value = 0.0001506
MCI
P-value = 0.5169
P-value = 0.574
P-value = 6.242E-08
AD
TAU
PTAU181P
ABETA142
APOE4

Figure 6

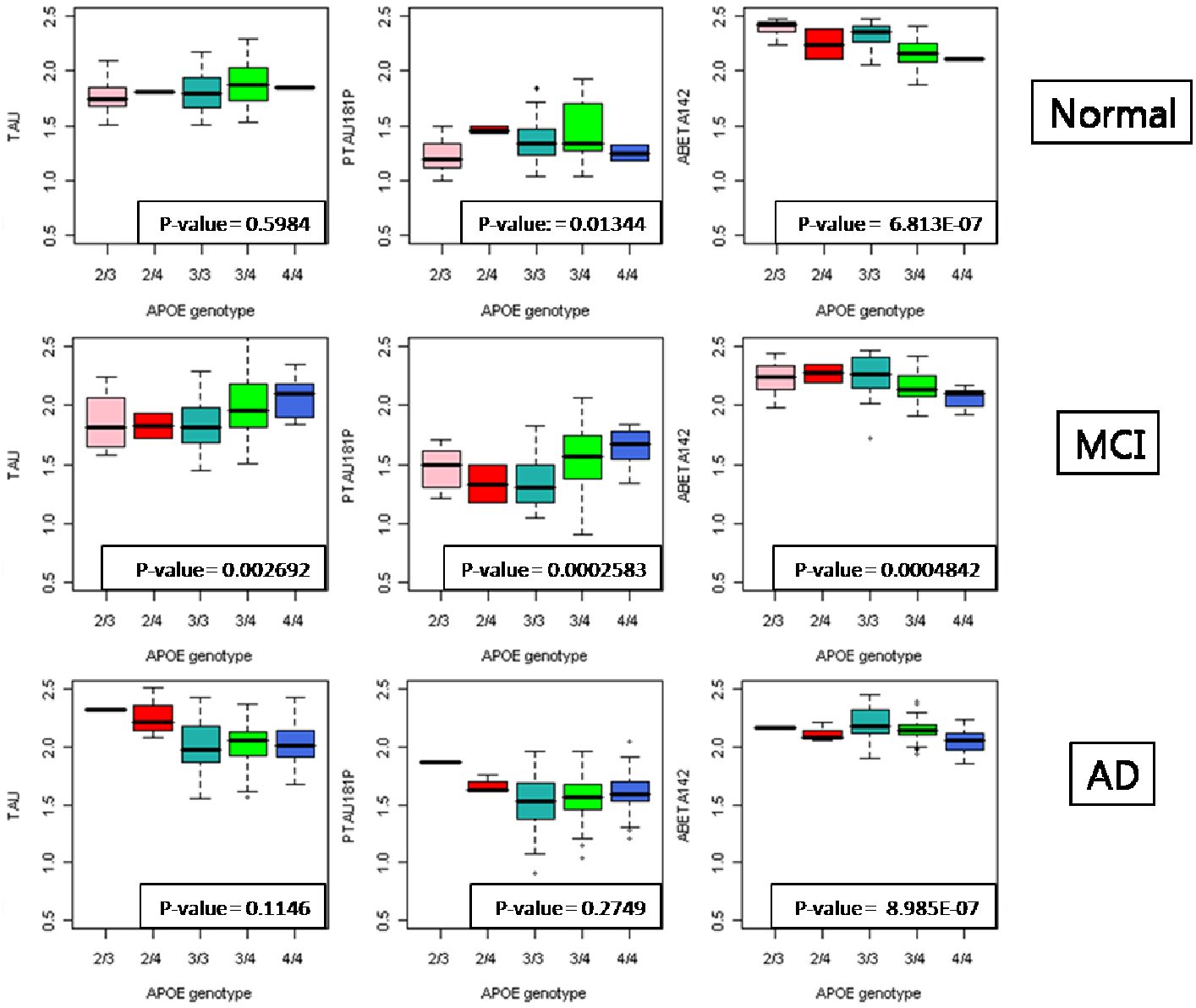
TAU
PTAU181P
ABETA142
APOE genotype
2/3
2/4
3/3
3/4
4/4
P-value = 0.5984
P-value: = 0.01344
P-value = 6.813E-07
P-value = 0.002692
P-value = 0.0002583
P-value = 0.0004842
P-value = 0.1146
P-value = 0.2749
P-value = 8.985E-07
Normal
MCI
AD

Figure 7

Additional files provided with this submission:

Additional file 1: Additional file 1.doc, 41K
http://www.biomedcentral.com/imedia/9455661524587208/supp1.doc
Additional file 2: Additional file 2.doc, 31K
http://www.biomedcentral.com/imedia/4228847424587209/supp2.doc
Additional file 3: Additional file 3.doc, 119K
http://www.biomedcentral.com/imedia/9953273804587209/supp3.doc
Additional file 4: Additional file 4.doc, 35K
http://www.biomedcentral.com/imedia/6680114324587210/supp4.doc
Additional file 5: Additional file 5.doc, 35K
http://www.biomedcentral.com/imedia/6095417245872109/supp5.doc
Additional file 6: Additional file 6.doc, 153K
http://www.biomedcentral.com/imedia/1708276469458721/supp6.doc
Additional file 7: Additional file 7.doc, 147K
http://www.biomedcentral.com/imedia/1455530820458721/supp7.doc
Additional file 8: Additional file 8.doc, 226K
http://www.biomedcentral.com/imedia/2085758225458721/supp8.doc