Genes, Brain and Behavior
Official publication of the International Behavioural and Neural Genetics Society

Genes, Brain and Behavior (2015) ***14:*** *607–617* doi: 10.1111/gbb.12250

Family-based genome scan for age at onset of late-onset Alzheimer's disease in whole exome sequencing data[1]

M. Saad[†,‡], Z. Brkanac[§] and E. M. Wijsman[*,†,‡]

[†]Department of Biostatistics, [‡]Division of Medical Genetics, Department of Medicine, and [§]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

[1]All authors participated in study design, analysis and drafted the manuscript.

**Corresponding author: E. M. Wijsman, Division of Medical Genetics, School of Medicine, University of Washington, BOX 357720, Seattle, WA 98195-7720, USA. E-mail: wijsman@uw.edu*

Alzheimer's disease (AD) is a common and complex neurodegenerative disease. Age at onset (AAO) of AD is an important component phenotype with a genetic basis, and identification of genes in which variation affects AAO would contribute to identification of factors that affect timing of onset. Increase in AAO through prevention or therapeutic measures would have enormous benefits by delaying AD and its associated morbidities. In this paper, we performed a family-based genome-wide association study for AAO of late-onset AD in whole exome sequence data generated in multigenerational families with multiple AD cases. We conducted single marker and gene-based burden tests for common and rare variants, respectively. We combined association analyses with variance component linkage analysis, and with reference to prior studies, in order to enhance evidence of the identified genes. For variants and genes implicated by the association study, we performed a gene-set enrichment analysis to identify potential novel pathways associated with AAO of AD. We found statistically significant association with AAO for three genes (*WRN*, *NTN4* and *LAMC3*) with common associated variants, and for four genes (*SLC8A3*, *SLC19A3*, *MADD* and *LRRK2*) with multiple rare-associated variants that have a plausible biological function related to AD. The genes we have identified are in pathways that are strong candidates for involvement in the development of AD pathology and may lead to a better understanding of AD pathogenesis.

Keywords: DNA sequencing, family-based association, linear mixed model, pedigree

Received 27 May 2015, revised 8 August 2015, accepted for publication 24 August 2015

Alzheimer's disease (AD: MIM104300) is a common and complex neurodegenerative disease. In the United States, it is reported as the sixth leading cause of death (Murphy *et al.* 2013), with direct costs in caring for subjects estimated as >200 billion dollars annually (Thies *et al.* 2013). Unlike a number of other common diseases associated with aging for which death rates have declined, including heart disease, stroke and some cancers, the death rate attributable to AD has substantially increased since 2000 (Thies *et al.* 2013). These issues lead to a projection of substantial rising costs in the near future, not only in the United States, but also in other countries as well (Banerjee 2014; Thies *et al.* 2013). Even a modest increase in age at onset (AAO) of AD through prevention or therapeutic measures would have enormous benefits as it would delay the disease and associated morbidities. To date, there have been no successful pharmacological or other therapies that achieve this goal (Thies *et al.* 2013).

Alzheimer's disease risk has a genetic basis, with evidence for familial aggregation first noted more than 60 years ago (Sjögren *et al.* 1952). Evidence for a genetic basis was later strengthened with results from twin (Gatz & Pedersen 2013; Gatz *et al.* 1997; Pedersen *et al.* 2004) and family (Mayeux *et al.* 1991; Van Duijn *et al.* 1991) studies. Identification of four genes with variation that contributes to AD risk provided definitive confirmation (Corder *et al.* 1993; Goate *et al.* 1991; Levy-Lahad *et al.* 1995; Sherrington *et al.* 1995). Rare mutations in *APP*, *PSEN1* and *PSEN2* are typically characterized by highly penetrant early-onset AD (EOAD, age < 65 years; Bertram *et al.* 2008; Bird *et al.* 1996), while common variation in *APOE* is associated with altered AD risk in more typical and common late-onset AD (LOAD, age ≥ 65 years). Risk increases as a function of number of *APOE* ε4 alleles, and decreases as a function of number of *APOE* ε2 alleles, relative to the baseline ε3 allele (Corder *et al.* 1993, 1994). A recent study employing targeted high-throughput sequencing also implicates rare variation in *APP*, *PSEN1* and *PSEN2* in risk of LOAD (Cruchaga *et al.* 2012) and rare variants in *TREM2* and *PLD3* genes have been implicated in LOAD as well (Benitez *et al.* 2013; Cruchaga *et al.* 2014; Guerreiro *et al.* 2013; Jonsson *et al.* 2013). Finally, large genome-wide studies (GWAS) have recently implicated multiple additional risk loci (Harold *et al.* 2009; Lambert *et al.* 2009, 2013; Naj *et al.* 2011; Seshadri *et al.* 2010). Currently, efforts are under way to obtain additional evidence for involvement of the genes proposed in GWAS studies (Holton *et al.* 2013; Lord *et al.* 2014).

In AD, age is an important factor. While AD is very rare in younger individuals, especially below age 60 years, mutations in the known early-onset genes are believed to account for ~50% of early-onset cases (Finckh *et al.* 2005; Ikeuchi *et al.* 2008; Lleo *et al.* 2002; Tandon & Fraser 2002).

Incidence of AD increases with age in all populations surveyed (Fratiglioni *et al.* 2000; Hebert *et al.* 2003; Lobo *et al.* 2000; Rocca *et al.* 1991; Sosa-Ortiz *et al.* 2012), with annual incidence in the United States increasing from ~1% at ages 65–70 years, to 6–8% by age 85 years and up (Mayeux 2003). As a result of this high-annual incidence rate, the prevalence of AD among individuals of age 85 years and above is ~32% (Thies *et al.* 2013). In the study of AD, age may be used either as a covariate, or as a phenotype of direct interest. As a phenotype that is directly relevant to AD, AAO is correlated among family-members, with wide variability among families (Bird *et al.* 1996). Transmission models of AAO in family-based samples support a genetic basis (Daw *et al.* 2000). Also consistent with a genetic basis, AAO differs among *APOE* genotypes, with genotype-specific risk inversely proportional to genotype-specific AAO. Differences among *APOE* genotype-specific onset-distributions are consistent across studies, whether measured on a genetic background of an EOAD mutation in one of the presenilin genes (Pastor *et al.* 2003; Wijsman *et al.* 2005) or in more typical LOAD (Farrer *et al.* 1997). These observations all suggest that AAO may be a useful phenotype for study of the genetic basis of AD.

To date AAO as the phenotype of interest has been used in only a few genome scans of LOAD. Family-based linkage analysis is the primary approach that has been used to identify regions of interest (Choi *et al.* 2011; Dickson *et al.* 2008; Holmans *et al.* 2005; Lee *et al.* 2008; Zhao *et al.* 2013b), with more limited recent use of GWAS in samples of unrelated, affected subjects (Kamboh *et al.* 2012; Naj *et al.* 2014). Together with regional analyses (Wijsman *et al.* 2004), these genome scans implicate regions containing AAO loci. In particular, regions on chromosome 6 and 19p replicated across independent sets of pedigrees. No inclusion of sequence data in AAO studies has yet been reported. There are two main issues with use of AAO. First, study subjects typically are identified through retrospective sampling designs. Analysis of AAO as a continuous variable from a GWAS study can give biased or misleading results when analyses are conducted as if a prospective sampling design had been used (Lin & Zeng 2009). This contrasts with logistic regression typically used for analysis of case–control studies, which provides equivalent estimates of relative risk for both prospective and retrospective studies (Prentice & Pyke 1979). Second, with AAO it is necessary to address age-censoring. As a result, analysis has either been carried out using only affected subjects (Holmans *et al.* 2005; Kamboh *et al.* 2012; Lee *et al.* 2008; Naj *et al.* 2014), or with a model that accounts for age-censoring (Choi *et al.* 2011; Dickson *et al.* 2008; Zhao *et al.* 2013b). Restricting analysis of AAO to AD cases leads to different interpretation of results than does analysis that includes unaffected subjects and incorporates age censoring. Analysis of AAO only in AD cases provides information about the genetic basis only of AAO modification, given predisposition to AD. Inclusion of unaffected subjects and incorporation of age-censoring allow a broader interpretation. In this context, reduced AAO in cases compared to controls as inferred from age-censored data may lead to the inference of increased age-specific risk that is a function of genotype in a region of interest.

In our study here, we performed a family-based GWAS for AAO of LOAD in whole exome sequencing (WES) data in families with multiple AD cases. Our WES data consisted of only AD cases, so our analysis is a case-only analysis of AAO. While GWAS single nucleotide polymorphism (SNP) chips provide a relatively high genomic coverage of the common (minor allele frequency, MAF > 0.05) genetic variation (>80%, Li *et al.* 2008), a portion of common variation and all rare genetic variation is poorly covered. The sequence data allow access to both rare and common variants. Therefore, the use of WES family data allows the study of both common and rare variants, which are likely to be enriched in families (Wijsman 2012). In order to identify new loci (both rare and common variants) and genes that may be a modifying factor of AAO, we conducted a classical single-marker association analysis for common variants, and a gene-based burden association analysis for multiple rare variants in genes. We combined association analyses with variance component linkage analysis, in addition to referring to previous analyses of AAO in other family-based samples, in order to enhance evidence of the identified genes. We also performed a gene-set enrichment analysis on the list of identified genes and explored the biological function of these genes to help understanding of AAO/AD pathogenesis.

Materials and methods

Subjects and phenotyping

Our data consisted of 77 subjects diagnosed with AD. These subjects were selected from pedigrees with multiple cases of late onset AD (>60 years onset age) in multigenerational families from public repositories: 58 subjects from the NIA-LOAD/NCRAD collection (Wijsman *et al.* 2011) and 19 subjects from the NIMH (Blacker *et al.* 1997) collection. Single subjects, but not families, from the NIA-LOAD/NCRAD families have been incorporated into GWAS studies of AAO, and the NIMH families have previously been used for a linkage-analysis genome scan of AAO (Choi *et al.* 2011). Neither set of families has been used for evaluation of sequence-based variants as contributors to AAO, which was the goal of the current study. Pedigrees ranged in size from 10 to 25 subjects, and consisted of 3–4 generations/pedigree. The assumption was that use of families increases ability to detect effects of (potentially) rare alleles. Selection of specific families additionally required availability of DNA from at least two relatively distantly related LOAD cases (e.g. avuncular to second cousin relationships), thus minimizing their relatedness relative to subjects from other available pedigrees, and reducing the sizes of regions of interest identified in the families. DNA samples from up to four cases per family were used when they were available, therefore also including additional, closer relationships. For inclusion in the current analysis, subjects also were required to be of European ancestry. Additional Hispanic pedigrees selected at the same time with the same criteria were only used for principal component analysis in order to ensure a homogenous group of subjects. They were not used for association analysis because of analytical complications of joint analysis of an admixed set of pedigrees with European- and Hispanic-descent families. The AD affection status was defined as meeting NINCDS-ADRDA criteria for definite, probable or possible AD (Mckhann *et al.* 1984; Wijsman *et al.* 2011). In both the NIMH and NIA-LOAD/NCRAD samples, AAO for AD cases was defined as the age at which first symptoms of AD were reported (Choi *et al.* 2011; Wijsman *et al.* 2011). In the subjects used for our analysis, the mean of AAO and its SD are 70.6 and 9.24, respectively. Our study was approved by the University of Washington institutional review board. All samples used were collected with appropriate consent for this study.

Whole exome sequencing data

The gene coding sequences were captured using Nimblegen SeqCap EZ Human Exome Library v2.0 kit (Roche, Basel, Switzerland) following the manufacturer's recommended instruction. The capture kit targets 28 858 genes with total size of the target regions 36.5 Mbp. The sequencing library clusters were generated on Illumina flowcells using cBlot (Illumina Inc., San Diego, CA, USA) and pair-end 101 bp sequencing was performed on the Illumina HIseq2000 sequencing platform at the Department of Genome Sciences, University of Washington. The raw base calling was performed with CASAVA (Illumina Inc.). Sequenced reads were aligned to NCBI human reference genome GRCh37 (hg19) using the Burrows-Wheeler Aligner version 0.5.6 (Li & Durbin 2010). BAM files were generated using SAMtools (Li *et al.* 2009). Polymerase chain reaction duplicates were marked using Picard (http://broadinstitute.github.io/picard/). After base recalibration the sequence reads were realigned around indels and mapped. For single nucleotide variant calling, the Genome Analyzer Toolkit version 4.1 (GATK) was utilized (Mckenna *et al.* 2010). The average read depth for called positions was 70.89.

In all analyses, we considered autosomal di-allelic polymorphic variants (i.e. 102 603 SNPs). Quality control (QC) steps were performed to filter out possible sequencing errors. We used the following exclusion criteria: ABHet (Allele balance for heterozygotes) > 0.75, HRun (largest contiguous Homopolymer Run of variant allele in either direction) > 4.0, QUAL (Phred-scaled quality score) ≤ 50, QD (Variant Confidence/Quality by Depth) < 5 or SB (Strand Bias) ≥ 0.10. We also excluded SNPs with evidence of Hardy-Weinberg (HW) disequilibrium (P value $< 10^{-3}$). HW testing was based on the 32 unrelated individuals in our WES dataset. Quality control procedures led to the exclusion of 15 937 SNPs (i.e. 14 948 due to sequence call quality filters and 989 due to HW disequilibrium). The number of remaining SNPs was 86 666.

Statistical analysis

Principal component analysis

Families in the NIA-LOAD/NCRAD sample were recruited by multiple sites across the United States (Wijsman *et al.* 2011) and families in the NIMH sample were recruited by three different sites (Blacker *et al.* 1997). As a result, subjects may be drawn from different genetic backgrounds even within subjects declared as of European ancestry. For the NIA-LOAD/NCRAD sample, several families are known to be Caribbean Hispanic. As described previously, there was strong evidence for population stratification in the NIA-LOAD/NCRAD European-American sample (Wijsman *et al.* 2011), and an indication of possible stratification in the NIMH sample (Choi *et al.* 2011). To account and correct for population stratification in the full sample used here, we performed a supervised principal component analysis (PCA) using EIGENSTRAT (Price *et al.* 2006) using SNPs from the WES data. For the PCA, we added to our AD cases the 1000 Genomes (1KG) project (Abecasis *et al.* 2010) subjects (release November 2010) of European (EUR), African (AFR) and Asian (ASN) descent. Principal component analysis was performed using common variants (MAF > 0.05) with low pairwise linkage disequilibrium (LD), and with available genotypes in both our data and the 1KG data. We used PLINK (Purcell *et al.* 2007) to select SNPs as follows: in a window of 150 SNPs, we estimated LD for all pairs of SNPs and filtered out one of each pair having an $r^2 > 0.2$. We used overlapping sliding windows with a step-size of five SNPs. The procedure led to a set of 8822 SNPs. Our PCA results identified 16 WES subjects who do not cluster with the remaining subjects of European descent (Fig. S1, Supporting Information). These subjects were all of known Caribbean Hispanic descent and were excluded from subsequent analysis, which led to exclusion of entire pedigrees, with no exclusion of any of the European-descent subjects. Among the remaining subjects, 56 subjects had available AAO information, and out of these 56 subjects, *APOE* genotypes were available for 47 subjects.

Imputation analysis for APOE

As recently shown (Radmanesh *et al.* 2014), missing *APOE* genotypes can now be accurately imputed using the very dense 1000 Genomes data as a reference. To avoid decrease of the sample size due to missing *APOE* genotypes for nine subjects, we performed an imputation analysis to infer the missing *APOE* genotypes using the 1000 Genomes data (August 2010). Briefly, we used SHAPEIT2 (Delaneau *et al.* 2012) to phase the NIA-LOAD/NCRAD GWAS subjects and minimac version 0.1.1 (http://genome.sph.umich.edu/wiki/Minimac) to impute the missing genotypes. Then we extracted the allelic dosages for the two *APOE* SNPs (i.e. rs7412 and rs429358) in the case of missing *APOE* genotypes and we used them along with the known *APOE* genotypes in subsequent association analyses. With these procedures, all 56 subjects were used.

Association analysis

We conducted two family-based genome-wide association analyses for the log-transformed AAO. The first analysis used a single-marker test where one SNP was tested at a time. The second analysis used a gene-based association burden test (Weighted Sum Approach; Madsen & Browning 2009). The aim of the first analysis was to find evidence of association between AAO and SNPs with MAF > 0.05. The aim of the second analysis was to find evidence of association between AAO and multiple rare variants in a gene (SNPs with MAF ≤ 0.05). All association analyses were performed using the statistical package R (http://www.r-project.org/). To account for family relationship among subjects, we used a linear mixed model (LMM) implemented in the R-package 'kinship'. In all models, we used the theoretical kinship matrix obtained from the pedigree structure information.

Single-marker association test for common variants

For the single-marker association test, assuming an additive model, we considered 39 993 SNPs with MAF > 0.05 and with at least 50 non-missing genotypes. Among these SNPs, 30 384, 20 574 and 13 590 have minor allele frequencies greater than 0.1, 0.2 and 0.3, respectively. We used the model: $\log(\mathrm{AAO}) = \beta \times X + \delta Z + \varepsilon$, where X is the vector of genotypes coded additively as 0, 1 or 2 copies of the minor allele, Z is the vector of observed covariates (e.g. *APOE* genotypes when used), β and δ are the marker and covariate fixed effect coefficients, respectively, and $\varepsilon \sim N\left(0, \sigma_g^2\Phi + \sigma_e^2 I\right)$ where Φ is a matrix of twice the coefficient of kinship between pairs of subjects, I is an identity matrix, and σ_g^2 and σ_e^2 are the polygenic and residual variances, respectively. To test for association, we used the Wald test ($H_1 : \beta \neq 0$ vs. $H_0 : \beta = 0$).

Burden association test for multiple rare variants

We performed the weighted sum burden association test (Madsen & Browning 2009) for rare variants (MAF ≤ 0.05). This approach collapses rare variants within genes by giving them weights that are inversely proportional to MAF. Single nucleotide polymorphisms were annotated using ANNOVAR (Wang *et al.* 2010). A gene was tested if it had at least two non-synonymous SNPs with MAF < 0.05 and if the sum of the MAFs was greater than 0.05. The association model we used was: $\log\left(\mathrm{AAO}\right) = \beta \times \sum_{i=1}^{p} w_i \times X_i + \delta Z + \varepsilon$ where $\sqrt{w_i} \sim d\beta(5, 25)$ (Wu *et al.* 2011), where Z and ε are defined above. Again, we used the Wald test to test for association.

Variance component linkage analysis

In order to determine which of our association signals are also supported by evidence for linkage and to incorporate this evidence into the overall evaluation of the signals, we performed a variance component linkage analysis using SOLAR (Almasy & Blangero 1998) for all 26 NIA-LOAD pedigrees, which have GWAS SNP data. The small number of NIMH pedigrees had a different, microsatellite marker scan, which could not be combined with the NIA-LOAD SNP markers for joint analysis. These pedigrees were therefore not used for the linkage analysis since the 19 available subjects is too small a sample to support estimation of the multiple parameters needed for a separate VC linkage analysis. For each chromosome, we first

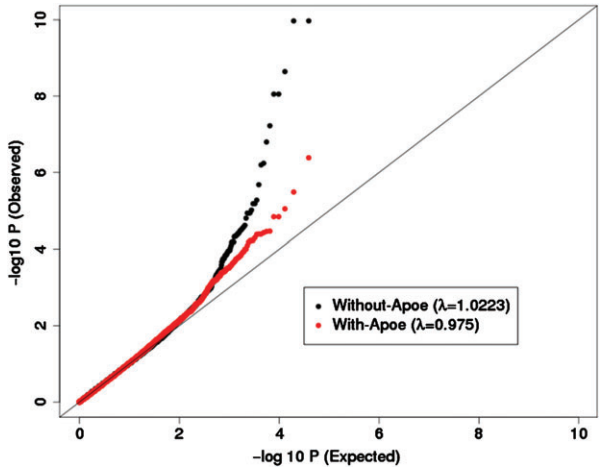

Figure 1: QQ-plot of the single-marker test analysis that adjusts (red) and does not adjust (black) for *APOE*. The plot is based on all 39 993 tested SNPs.

selected a set of equally-spaced SNPs (~0.5 cM) with relatively high MAF (>0.4) and in linkage equilibrium. Then, we estimated the Identity-By-Descent distribution at each marker position in a full multipoint computation using a Markov Chain Monte Carlo (MCMC) method for pedigrees that have more than 15 transmitted meioses and exact computation for the remaining pedigrees. This analysis was performed using the program gl_auto in the MORGAN package (Thompson 2011). Identity-By-Descent estimates at 5 cM intervals were converted to SOLAR input-format and a model with additive variance components only was fitted and compared to a model with polygenic variance, only, in a likelihood ratio test. We ran two versions of this model: the first one adjusted for APOE genotypes (VC wApoe) and the second one did not (VC sansApoe).

Results

Single-marker test

We first ran association analyses without any covariates. The QQ-plot and genomic control coefficient (Devlin & Roeder 1999; $\lambda = 1.022$) showed slight inflation in the statistical test distribution (Fig. 1). This excess of significant results may be driven by the effect of *APOE* on sample ascertainment, and thus on sample structure as suggested in Wijsman *et al.* (2011). To reduce the observed inflation, and to also avoid detecting signals driven by *APOE* (a very well established associated factor with AD; Corder *et al.* 1994; Farrer *et al.* 1997), we ran a second association analysis adjusting for *APOE* as covariate in the LMM. The statistical distribution obtained by this analysis was better controlled, as both the genomic control coefficient ($\lambda = 0.975$) and QQ-plot showed (Fig. 1).

As expected, we observed significant evidence of association for SNPs in the *APOE* region on chromosome 19 in the first analysis without adjustment for *APOE* (from 44 Mbp to 47 Mbp). None of the SNPs tested are the two SNPs that define the three critical *APOE* alleles, as those SNPs fail QC analysis because of low read depth. At a nominal threshold ($\alpha = 0.05$), 10 SNPs were significant in this region. The SNP rs11879355 provided the strongest evidence for association with a P value of 6.2×10^{-3} (Table S1). After adjusting for *APOE*, all 10 significant SNPs were no longer significant at a nominal threshold of $\alpha = 0.05$ (Table S1).

Our main results from single SNP analyses are based on the second analysis, which adjusts for *APOE*, and are shown in the Manhattan plot in Fig. 2. Our strategy was to focus on the most significant non-synonymous SNPs with P values less than 5×10^{-4} (i.e. 15 SNPs). The results of these SNPs are shown in Table 1. All these SNPs, except rs1800378, had negative effect sizes, indicating that they decrease AAO. The MAF of 13 of these SNPs was less than 0.1. The significance of association tests of the 15 SNPs ranged from 4.99×10^{-4} to 4.12×10^{-7}. We identified one SNP on chromosome 19 (rs2291516, MAF = 0.08, P value = 4.12×10^{-7}) with Bonferroni-corrected significance ($0.05/39\,993 = 1.25 \times 10^{-6}$). This SNP is in the gene *RGL3*. The remaining 14 SNPs are located in 11 different genes on nine different chromosomes.

Burden test

The number of tested genes with at least two non-synonymous SNPs with MAF < 0.05 and the sum of the MAFs greater than 0.05 was 1949. The first quartile, median, mean and third quartile of the number of rare variants in a gene were 3, 4, 4.96 and 6, respectively. We ran two versions of burden tests. The first one adjusts for *APOE* (wApoe) and the second one does not (sansApoe). Genomic control coefficients (0.776 for wApoe and 0.996 for sansApoe) and QQ-plots showed better properties of the burden test obtained by sansApoe. This trend was different from what we obtained in the single-marker test. This is due to the fact that the weighted sum of variants in a gene is likely to be less correlated with *APOE* than the alleles of each marker alone. Therefore, our results were based on the sansApoe analysis. Manhattan and QQ-plots can be found in Figs. S2,S3. In the burden test analysis, we focused on the 10 most significant genes (Table 2) from our genome scan, with P values ranging from 3.11×10^{-4} to 5.3×10^{-6}. Again, to ensure that our most significant results were not driven by the effect of *APOE*, we explored the wApoe analysis of the 10 selected genes. The significance of all genes decreased slightly in the wApoe analysis.

Variance component linkage analysis

In the VC sansApoe, we identified five regions with lod-scores ≥1.5 (Chromosome 2, 4, 9, 13 and 19) (Supporting Information). The region on chromosome 19 is the longest one and has the highest maximum lod-score, reaching ~2, out of these five regions (Fig. 3). The relevant gene in this region is most likely *APOE* since the lod-score maximizes at approximately the location of *APOE* on chromosome 19. Indeed, the VC wApoe analysis showed a drastic decrease of lod-scores in this region (i.e. lod-scores drop from ~2 for VC sansApoe to <0.2 for VC wApoe at the position of *APOE* at ~72 cM, Fig. 3). In the same analysis, most, but not all, lod-scores of other regions also decreased. The region on chromosome 9 was the most robust to *APOE* adjustment with its lod-score decreasing modestly from 1.47 to 1. New regions appeared when adjusting for *APOE*: one around 210 cM

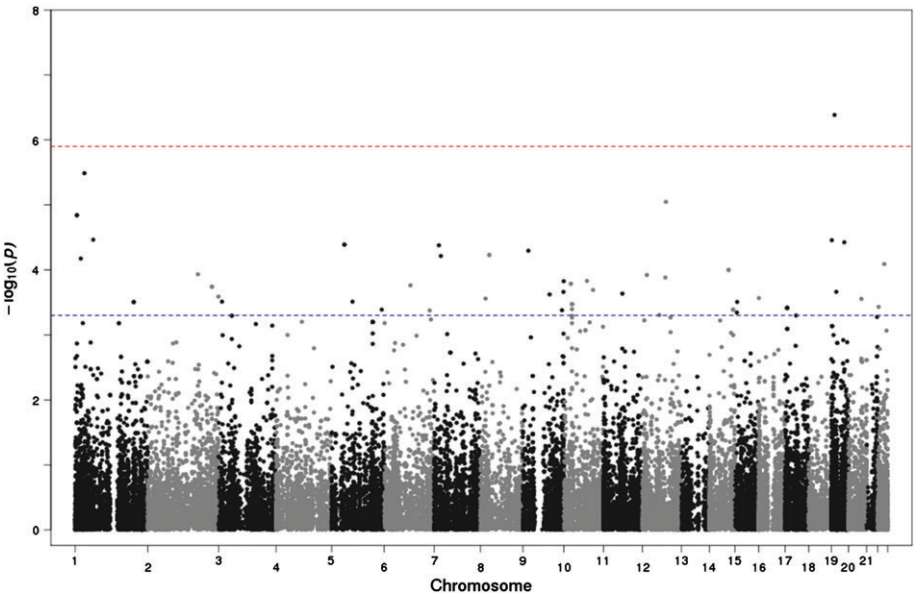

Figure 2: Manhattan plot of the single-marker test analysis that adjusts for *APOE*. The plot is based on all 39 993 tested SNPs. The horizontal blue line is the threshold we used to decide the most significant SNPs. The horizontal red line is the genome-wide significance threshold.

Table 1: Most significant SNPs obtained by the single-marker test

Chr	Position	Gene	rsname	Minor	Major	MAF	*F*	*β*	*P*	*β**	*P**
1	201,180,100	*IGFN1*	—	T	C	0.071	—	−0.147	1.18E−03	−0.144	3.12E−04
1	201,180,340	*IGFN1*	rs139390045	G	A	0.071	0.084	−0.147	1.18E−03	−0.144	3.12E−04
5	43,613,046	*NNT*	rs35201656	G	A	0.054	0.042	−0.116	2.59E−02	−0.157	4.08E−05
5	43,653,243	*NNT*	rs41271083	T	C	0.054	0.042	−0.116	2.59E−02	−0.157	4.08E−05
8	30,921,935	*WRN*	rs2230009	A	G	0.054	0.056	−0.346	1.07E−10	−0.296	5.88E−05
9	133,963,008	*LAMC3*	rs4740412	A	G	0.214	0.266	−0.077	1.31E−02	−0.089	4.17E−04
9	139,118,673	*QSOX2*	rs12380852	C	T	0.107	0.117	−0.132	1.71E−04	−0.113	2.17E−04
10	23,729,362	*OTUD1*	—	T	A	0.063	—	−0.102	8.63E−02	−0.203	1.63E−04
15	28,230,318	*OCA2*	rs1800407	T	C	0.054	0.078	−0.142	4.34E−03	−0.157	3.11E−04
16	615,048	*C16orf11*	rs113068385	A	G	0.055	0.036	−0.053	3.05E−01	−0.151	2.71E−04
17	7,324,788	*SPEM1*	rs33989543	A	G	0.054	0.070	−0.258	6.57E−06	−0.151	2.71E−04
17	7,735,063	*DNAH2*	rs57985356	G	T	0.071	0.085	−0.141	3.84E−03	−0.138	3.82E−04
17	7,735,934	*DNAH2*	rs61745181	A	G	0.071	0.076	−0.141	3.84E−03	−0.138	3.82E−04
19	11,508,177	*RGL3*	rs2291516	A	G	0.080	0.101	−0.155	3.60E−04	−0.168	4.12E−07
19	17,361,116	*USHBP1*	rs1043963	A	G	0.063	0.086	−0.133	8.46E−03	−0.157	2.17E−04

β, effect size not adjusted for *APOE*; *F*, minor allele frequency in the ESP project data of European-American sample; MAF, minor allele frequency in the sample; *P*, *P* value not adjusted for *APOE*.
*Analysis that adjusts for *APOE*. All analyses were based on $N = 56$ subjects.

on chromosome 2 (lod-score adjusted for *APOE* = 1.1) and other around 50 cM on chromosome 4 (lod-score adjusted for *APOE* = 1.4). Moreover, a modest signal on chromosome 19p around 30–40 cM (lod ~0.6) is effectively immune to whether or not there is adjustment for *APOE*.

Analysis of known genes from the literature

We explored our association results for the SNPs and genes reported in a previous AAO GWAS (Naj *et al.* 2014). The authors considered previously identified genes associated with AD (i.e. *CR1*, *BIN1*, *CD2AP*, *EPHA1*, *CLU*, *MS4A4A*, *PICALM*, *ABCA7* and *CD33*). They identified association between AAO and SNPs in *CR1* (rs6701713, *P* value $= 7.2 \times 10^{-4}$), *BIN1* (rs7561528, *P* value $= 4.8 \times 10^{-4}$), and *PICALM* (rs561655, *P* value $= 2.2 \times 10^{-3}$). From the list of SNPs reported in this previous study, only one SNP was found in our WES data (rs3752246, *ABCA7*). This SNP was not significant in both our and their study (*P* value = 0.69 and 0.064, respectively). Note that the remaining SNPs were located in introns. Nonetheless, several SNPs with MAF greater than 0.05, in *CR1* and *ABCA7*, were nominally significant in our single-marker test analysis of AAO. In *CR1*, two SNPs (rs2274567 and rs3811381) had *P* values of 0.034. In *ABCA7*, two SNPs (rs3764645 and rs3752234) were nominally significant with *P* values equal to 0.018 and 0.023, respectively. These results are shown in Table S2. Using the burden test, three genes (i.e. *BIN1*, *EPHA1* and *ABCA7*) out of the nine considered in Naj *et al.* (2014) had at least two

Table 2: Burden test results for the ten most significant genes

Chr	Gene	Pos first SNP	Pos last SNP	V0.05	S_MAF	β	*P*	β*	*P**
2	*SLC19A3*	228,552,234	228,563,911	3	0.054	−0.266	5.85E−06	−0.166	2.77E−03
11	*MADD*	47,296,533	47,315,499	2	0.054	−0.254	2.00E−05	−0.169	2.85E−03
11	*C11orf82*	82,625,814	82,644,904	4	0.080	−0.208	3.08E−05	−0.105	2.71E−02
11	*KDELC2*	108,352,777	108,357,137	2	0.054	−0.241	5.24E−05	−0.219	5.21E−04
12	*LRRK2*	40,619,082	40,758,652	4	0.054	−0.223	2.84E−04	−0.133	2.09E−02
14	*SLC8A3*	70,527,576	70,634,200	3	0.071	−0.196	2.56E−04	−0.049	3.68E−01
14	*SLC24A4*	92,909,807	92,959,940	5	0.063	−0.177	5.30E−06	−0.288	1.00E−04
15	STARD9	42,930,972	43,011,009	8	0.080	−0.109	3.11E−04	−0.053	3.45E−01
19	*GRIN3B*	1,000,785	1,009,585	3	0.063	−0.190	9.00E−05	−0.119	3.47E−02
19	*LENG8*	54,966,557	54,968,038	3	0.063	−0.202	2.35E−04	−0.112	3.52E−02

β, effect size not adjusted for *APOE*; *P*, *P* value not adjusted for *APOE*; S_MAF, sum of the V0.05 minor allele frequencies; V0.05, SNPs with MAF < 0.05.
*Analysis that adjusts for *APOE*. All analyses were based on $N = 56$ subjects.

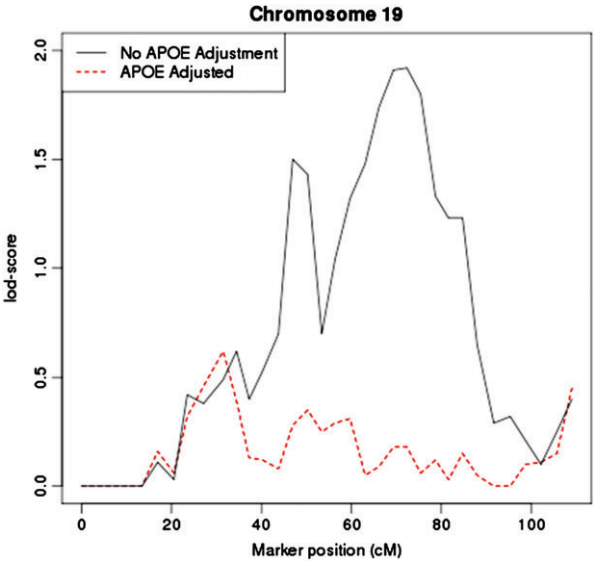

Figure 3: Lod-score plot for chromosome 19. The red dashed line represents the lod-scores of VC analysis adjusting for *APOE*. The black line represents the lod-scores of VC analysis without adjustment for *APOE*.

SNPs with MAF less than 0.05, which means that they were tested. However, none of them was significant (Table S3).

Bioinformatic enrichment analysis

We ran a gene-set enrichment analysis (GSEA) using a user-friendly web-based tool, WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/, Wang *et al.* 2013). We performed GSEA enrichment analysis using both GO and KEGG databases on the list of: (1) the 16 most significant genes obtained by the single-marker test, (2) the 10 most significant genes obtained by the burden test, and (3) the combined list of 26 genes from both tests. We focused on pathways with *P* values less than 0.05 after a Bonferroni correction for multiple testing (Table 3). Only the burden test achieved this threshold, and using the list of 'burden test genes', six GO pathways containing five genes (*SLC8A3*, *SLC24A4*, *SLC19A3*, *GRIN3B* and *LRRK2*) from the burden list showed significant results after Bonferroni correction (adjusted *P* value range = [0.0096 – 0.0288]; Table 3). It is notable that identified genes belong to multiple associated pathways.

Discussion

In this paper, we presented the result of a (WES) family-based association study for AAO of LOAD subjects. Risk of AD and its AAO are related traits, as shown by overlap on both traits of effects of known genetic factors, such as *APOE* (Corder *et al.* 1993; Farrer *et al.* 1997). Nonetheless, the interpretation and implication of results focused on AD risk vs. AAO are different. A focus on AAO among AD cases, as in our study, may bring new insights into factors affecting onset of AD in those that are at risk. This allows a more nuanced and useful measure than simple risk of AD, and is highly pertinent to downstream investigation. Even a modest increase in AAO of AD through prevention or therapeutic measures would have enormous benefits by simply delaying the onset of disease.

In our study design, WES was obtained on 2–3 affected subjects per family that has multiple affected subjects. This family-based design might be more efficient than a population-based design, especially for rare variants that are enriched in pedigrees (Wijsman 2012). In addition, the rare variants that might be implicated in AD or the AAO of AD are likely to be shared by these affected subjects. Despite the modest size of our WES dataset, our results suggest that a good design that uses a carefully selected set of subjects can provide promising results. This is demonstrated by our replication of several signals from (Naj *et al.* 2014), a study based on more than 9000 unrelated subjects, by our identification of several additional candidate genes for AD, and also finding that some of the identified genes are located in regions with evidence of linkage for AAO in this study, as well as in other studies.

Table 3: Gene-set enrichment analysis using WebGestalt for the most significant genes obtained by the burden test

Pathway ID	Pathway name	Database	*C*	*O*	*E*	*R*	raw*P*	adj*P*
GO:0015291	Secondary active transmembrane transporter activity	GO molecular function	192	3 (*SLC24A4*, *SLC8A3*, *SLC19A3*)	0.11	26.53	2.0E−04	9.6E−03
GO:0043025	Neuronal cell body	GO cellular component	291	3 (*SLC8A3*, *GRIN3B*, *LRRK2*)	0.15	19.53	4.0E−04	1.5E−02
GO:0044297	Cell body	GO cellular component	312	3 (*SLC8A3*, *GRIN3B*, *LRRK2*)	0.16	18.22	5.0E−04	1.9E−02
GO:0015297	Antiporter activity	GO molecular function	60	2 (*SLC24A4*, *SLC8A3*)	0.04	56.61	5.0E−04	2.4E−02
GO:0006816	Calcium ion transport	GO biological process	216	3 (*SLC24A*, *SLC8A3*, *GRIN3B*)	0.13	22.58	2.0E−04	2.7E−02
GO:0022804	Active transmembrane transporter activity	GO molecular function	309	3 (*SLC24A4*, *SLC8A3n SLC19A3*)	0.18	16.49	6.0E−04	2.9E−02

adj*P*, Bonferroni-corrected *P* value; *C*, the number of reference genes in the pathway; *E*, the expected number in the pathway; *O*, the number of genes in the gene set and also in the pathway; *R*, ratio of enrichment; raw*P*, *P* value from hypergeometric test.

There are also several aspects of both the design and the analysis that should contribute to robust results. First, as is common in similar studies, we took a number of steps to make sure that our results are statistically robust and not explained by artifacts of confounders (e.g. poor SNP quality and population stratification). Second, by focusing on AAO in affected subjects, only, we avoid the problem posed by the censored age data in unaffected subjects, for which there are not yet analytical methods that give statistically-robust results for variance-components analysis in pedigree samples. Third, our choices of analysis details were chosen to make the results robust to small sample size. By using direct sequence data and methods that allow for the possibility of multiple variants within relevant genes while also capitalizing on the increased information that can be obtained from a continuous trait, we eliminate many of the reasons that very large sample sizes became necessary during the era of GWAS case–control studies.

The use of WES data permits the evaluation of rare genetic variations in the functional parts of the genome (exons), which are not genotyped or well-tagged in classical GWAS SNP chips (Li *et al.* 2008). In addition, WES may provide a direct observation of common variants that are not well-tagged in GWAS SNP chips. Using a single-marker test analysis for common variants and a burden test analysis for multiple rare variants have permitted us to detect novel candidate genes that may play a functional role in modifying AAO of AD. On the basis of literature review, among genes we have identified, three (*WRN*, OMIM 604611; *NTN4*, OMIM 610401 and *LAMC3*, OMIM 604349) with common-associated SNPs and four with multiple rare variants associated with AAO (*SLC8A3*, OMIM 607991; *SLC19A3*, OMIM 606152; *MADD*, OMIM 603584; and *LRRK2*, OMIM 609007) have strong prior evidence for involvement in AD (http://www.genecards.org/).

Two genes, *NTNA4* and *LAMC3*, belong to a family of proteins related to laminins. The gene Netrin 4 (*NTN4*) [rs17288108, P value $= 3.48 \times 10^{-5}$, MAF = 0.125, p.(Y205H)] was originally found to have a role in neuronal axon migration and may play an important role in development (Cirulli & Yebra 2007). Shen *et al.* (2012) recently showed that *NTN4* expression is upregulated during β-amyloid induced injury of neurite outgrowth, an effect that is reversed with addition of acetylcholinestarase inhibitors. This is consistent with the possibility that *NTN4* might play a role in the development of AD pathology (Shen *et al.* 2012). The gene *LAMC3* [rs4740412, P value $= 1.44 \times 10^{-4}$, MAF = 0.214, p.(R1459Q)] is a member of the laminin family of heterotrimeric molecules that function in stabilization of epithelial structures. *LAMC3* is strongly expressed in developing human fetal brain with highest expression in temporooccipital regions. Recessive mutations in *LAMC3* cause a syndrome with cortical malformations and seizures (Barak *et al.* 2011) further underlining its importance in brain development. Laminin interacts with β-amyloid supporting its role in AD (Morgan & Inestrosa 2001). In addition, *LAMC3* is a part of network that includes *PICALM*, a gene previously reported as associated with AD (Carrasquillo *et al.* 2010; Harold *et al.* 2009; Lambert *et al.* 2013).

Among genes identified by the burden test, there was *LRRK2* (leucine-rich repeat serine/threonine-protein kinase 2, P value $= 2.84 \times 10^{-4}$) which is the most common cause of dominant inherited Parkinson's disease (*PARK8*; Lesage & Brice 2009). Common variants in this gene have been also found to increase the susceptibility to Parkinson's disease (MIM168600; Gilks *et al.* 2005; Nalls *et al.* 2011; Simon-Sanchez *et al.* 2009). Several LRRK2 protein functions might lead to its effects in AD. *LRRK2* regulates autophagy through a calcium-dependent activation of the CaMKK/AMPK signaling pathway (Gomez-Suaga *et al.* 2012) and mediates the synaptotoxic effects of Amyloid beta oligomers through tau phosphorylation (Mairet-Coello *et al.* 2013). *LRRK2* might also contribute to Lewy Body pathology in AD (Linnertz *et al.* 2014). The remaining four genes that have functions related to AD can be found in the Supporting Information.

Finally, the VC linkage analysis we performed gives strength to some genes identified in our single-marker and burden association analyses. An interesting gene is *RGL3*, which both gave the strongest single-variant results in the current analysis, for SNP rs2291516, and is in a region with

evidence of linkage on chromosome 19p located upstream of *APOE*. Even though the evidence for linkage in the current sample is moderate, it is interesting because at the same position, a strong signal was identified when adjusting for *APOE* in an earlier evaluation of regions containing AAO loci (Wijsman *et al.* 2004) with later confirmatory evidence provided by two other independent samples (Choi *et al.* 2011; Zhao *et al.* 2013a). The pedigrees used in these studies do not overlap with pedigrees used here. The region with evidence for linkage to AAO also includes SNP rs1043963 in *USHBP1*, which also gave positive results in the single-marker association test performed here. Another gene, *SLC19A3*, which was identified by the burden test, is also located in a region with evidence of linkage on chromosome 2. This gene has a biological function that might be related to AD (Supporting Information).

Current advances in sequencing allow for whole exome and genome sequencing in tens of thousands of samples. However, in the context of sequencing data, replication is not simple. Combining or replicating studies that use next-generation sequence data has new challenges. Read depth and analytical procedures for alignment and variant calling affect results, requiring both re-calling of the sequence data, and use of methods that include modeling read depth to avoid spurious results (Derkach *et al.* 2014). Such procedures do not yet include related subjects. In addition, even such large samples might not have sufficient power to account for multiple testing and large samples introduce additional complications of aggregating large datasets, including variation in phenotypic measurements, such as AAO across studies.

In conclusion, with our statistical approach that uses a family-based association study design, we have identified several candidate genes that have additional functional evidence for association with AD. Our family-based design and focus on coding regions of the genome attenuates the issues of multiple testing that complicate classic case–control designs. The identification of candidate genes is important from the perspective of a follow-up in larger case–control samples or in other family-based samples. Our approach that identifies a focused list of candidate genes, specific types of variants that are responsible for association (single low frequency variant vs. burden of rare variants) and the specific AAO phenotype allows for increase in power and replication study in smaller datasets.

References

Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. & McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073.

Almasy, L. & Blangero, J. (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* **62**, 1198–1211.

Banerjee, S. (2014) The macroeconomics of dementias – will the world economy get Alzheimer's disease? *Arch Med Res* **43**, 705–709.

Barak, T., Kwan, K.Y., Louvi, A. *et al.* (2011) Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat Genet* **43**, 590–594.

Benitez, B.A., Cooper, B., Pastor, P., Jin, S.C., Lorenzo, E., Cervantes, S. & Cruchaga, C. (2013) TREM2 is associated with the risk of Alzheimer's disease in Spanish population. *Neurobiol Aging* **34**, e1715–e1717.

Bertram, L., Lange, C., Mullin, K. *et al.* (2008) Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet* **83**, 623–632.

Bird, T.D., Levy-Lahad, E., Poorkaj, P., Sharma, V., Nemens, E., Lampe, T. & Schellenberg, G.D. (1996) Wide range in age of onset for chromosome 1-related familial Alzheimer's disease. *Ann Neurol* **40**, 932–936.

Blacker, D., Haines, J.L., Rodes, L., Terwedow, H., Go, R.C., Harrell, L.E., Perry, R.T., Bassett, S.S., Chase, G., Meyers, D., Albert, M.S. & Tanzi, R. (1997) ApoE-4 and age at onset of Alzheimer's disease: the NIMH genetics initiative. *Neurology* **48**, 139–147.

Carrasquillo, M.M., Belbin, O., Hunter, T.A., Ma, L., Bisceglio, G.D., Zou, F., Crook, J.E., Pankratz, V.S., Dickson, D.W., Graff-Radford, N.R., Petersen, R.C., Morgan, K. & Younkin, S.G. (2010) Replication of CLU, CR1, and PICALM Associations with Alzheimer disease. *Arch Neurol* **67**, 961–964.

Choi, Y., Marchani, E.E., Bird, T.D., Steinbart, E.J., Blacker, D. & Wijsman, E.M. (2011) Genome scan of age-at-onset in the NIMH Alzheimer disease sample uncovers multiple loci, along with evidence of both genetic and sample heterogeneity. *Am J Med Genet B Neuropsychiatr Genet* **156B**, 785–798.

Cirulli, V. & Yebra, M. (2007) Netrins: beyond the brain. *Nat Rev Mol Cell Biol* **8**, 296–306.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L. & Pericak-Vance, M.A. (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923.

Corder, E.H., Saunders, A.M., Risch, N.J., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Rimmler, J.B., Locke, P.A., Conneally, P.M., Schmader, K.E., Small, G.W., Roses, A.D., Haines, J.L. & Pericak Vance, M.A. (1994) Protective effect of apolipoprotein-E type-2 allele for late-onset Alzheimer-disease. *Nat Genet* **7**, 180–184.

Cruchaga, C., Chakraverty, S., Mayo, K., Vallania, F.L., Mitra, R.D., Faber, K., Williamson, J., Bird, T., Diaz-Arrastia, R., Foroud, T.M., Boeve, B.F., Graff-Radford, N.R., St Jean, P., Lawson, M., Ehm, M.G., Mayeux, R. & Goate, A.M. (2012) Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS ONE* **7**, e31039.

Cruchaga, C., Karch, C.M., Jin, S.C. *et al.* (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**, 550–554.

Daw, E., Bird, T., Nemens, E., Nochlin, D., Schellenberg, G. & Wijsman, E. (2000) ApoE and other loci affect age of Alzheimer's disease onset in families with PS2 mutation. *Am J Hum Genet* **67** (Suppl), 192.

Delaneau, O., Marchini, J. & Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179–181.

Derkach, A., Chiang, T., Gong, J., Addis, L., Dobbins, S., Tomlinson, I., Houlston, R., Pal, D.K. & Strug, L.J. (2014) Association analysis using next generation sequence data from publicly available control groups: the robust variance score statistics. *Bioinformatics* **30**, 2179–2188.

Devlin, B. & Roeder, K. (1999) Genomic control for association studies. *Biometrics* **55**, 997–1004.

Dickson, M.R., Li, J., Wiener, H.W., Perry, R.T., Blacker, D., Bassett, S.S. & Go, R.C.P. (2008) A genomic scan for age at onset of Alzheimer's disease in 437 families from the NIMH genetic initiative. *Am J Med Genet B Neuropsychiatr Genet* **147B**, 784–792.

Farrer, L.A., Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W.A., Mayeux, R., Myers, R.H., PericakVance, M.A., Risch, N. & vanDuijn, C.M. (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease – a meta-analysis. *JAMA* **278**, 1349–1356.

Finckh, U., Kuschel, C., Anagnostouli, M., Patsouris, E., Pantes, G.V., Gatzonis, S., Kapaki, E., Davaki, P., Lamszus, K., Stavrou, D. & Gal, A. (2005) Novel mutations and repeated findings of mutations in familial Alzheimer disease. *Neurogenetics* **6**, 85–89.

Fratiglioni, L., Launer, L.J., Andersen, K., Breteler, M.M.B., Copeland, J.R.M., Dartigues, J.F., Lobo, A., Martinez-Lage, J., Soininen, H., Hofman, A. & Neurologic Diseases in Elderly Research Group (2000) Incidence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. *Neurology* **54**, S10–S15.

Gatz, M. & Pedersen, N.L. (2013) Study of dementia in Swedish twins. *Twin Res Hum Genet* **16**, 313–316.

Gatz, M., Pedersen, N.L., Berg, S., Johansson, B., Johansson, K., Mortimer, J.A., Posner, S.F., Viitanen, M., Winblad, B. & Ahlbom, A. (1997) Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci* **52**, M117–M125.

Gilks, W.P., Abou-Sleiman, P.M., Gandhi, S., Jain, S., Singleton, A., Lees, A.J., Shaw, K., Bhatia, K.P., Bonifati, V., Quinn, N.P., Lynch, J., Healy, D.G., Holton, J.L., Revesz, T. & Wood, N.W. (2005) A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet* **365**, 415–416.

Goate, A., Chartier-Harlin, M.C., Mullan, M., *et al.* (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704–706.

Gomez-Suaga, P., Luzon-Toro, B., Churamani, D., Zhang, L., Bloor-Young, D., Patel, S., Woodman, P.G., Churchill, G.C. & Hilfiker, S. (2012) Leucine-rich repeat kinase 2 regulates autophagy through a calcium-dependent pathway involving NAADP. *Hum Mol Genet* **21**, 511–525.

Guerreiro, R., Wojtas, A., Bras, J. *et al.* (2013) TREM2 variants in Alzheimer's disease. *N Engl J Med* **368**, 117–127.

Harold, D., Abraham, R., Hollingworth, P. *et al.* (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **62**, 1088–1093.

Hebert, L.E., Scherr, P.A., Bienias, J.L., Bennett, D.A. & Evans, D.A. (2003) Alzheimer disease in the US population – prevalence estimates using the 2000 census. *Arch Neurol* **60**, 1119–1122.

Holmans, P., Hamshere, M., Hollingworth, P. *et al.* (2005) Genome screen for loci influencing age at onset and rate of decline in late onset Alzheimer's disease. *Am J Med Genet B Neuropsychiatr Genet* **135B**, 24–32.

Holton, P., Ryten, M., Nalls, M. *et al.* (2013) Initial assessment of the pathogenic mechanisms of the recently identified Alzheimer risk loci. *Ann Hum Genet* **77**, 85–105.

Ikeuchi, T., Kaneko, H., Miyashita, A., Nozaki, H., Kasuga, K., Tsukie, T., Tsuchiya, M., Imamura, T., Ishizu, H., Aoki, K., Ishikawa, A., Onodera, O., Kuwano, R. & Nishizawa, M. (2008) Mutational analysis in early-onset familial dementia in the Japanese population. *Dement Geriatr Cogn Disord* **26**, 43–49.

Jonsson, T., Stefansson, H., Steinberg, S. *et al.* (2013) Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* **368**, 107–116.

Kamboh, M.I., Barmada, M.M., Demirci, F.Y., Minster, R.L., Carrasquillo, M.M., Pankratz, V.S., Younkin, S.G., Saykin, A.J., Initiative, The Alzheimer's Disease Neuroimaging, Sweet, R.A., Feingold, E., Dekosky, S.T. & Lopez, O.L. (2012) Genome-wide association analysis of age-at-onset in Alzheimer's disease. *Mol Psychiatry* **17**, 1340–1346.

Lambert, J.C., Heath, S., Even, G. *et al.* (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094–1099.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D. *et al.* (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452–1458.

Lee, J.H., Barral, S., Cheng, R., Chacon, I., Santana, V., Williamson, J., Lantigua, R., Medrano, M., Jimenez-Velazquez, I.Z., Stern, Y., Tycko, B., Rogaeva, E., Wakutani, Y., Kawarai, T., St George-Hyslop, P. & Mayeux, R. (2008) Age-at-onset linkage analysis in Caribbean Hispanics with familial late-onset Alzheimer's disease. *Neurogenetics* **9**, 51–60.

Lesage, S. & Brice, A. (2009) Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet* **18**, R48–R59.

Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D.M., Oshima, J., Pettingell, W.H., Yu, C.E., Jondro, P.D., Schmidt, S.D., Wang, K., Crowley, A.C., Fu, Y.H., Guenette, S.Y., Galas, D., Nemens, E., Wijsman, E.M., Bird, T.D., Schellenberg, G.D. & Tanzi, R.E. (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973–977.

Li, H. & Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595.

Li, M., Li, C. & Guan, W. (2008) Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Genet* **16**, 635–643.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Lin, D.Y. & Zeng, D. (2009) Proper analysis of secondary phenotype data in case–control association studies. *Genet Epidemiol* **33**, 256–265.

Linnertz, C., Lutz, M.W., Ervin, J.F., Allen, J., Miller, N.R., Welsh-Bohmer, K.A., Roses, A.D. & Chiba-Falek, O. (2014) The genetic contributions of SNCA and LRRK2 genes to Lewy Body pathology in Alzheimer's disease. *Hum Mol Genet* **23**, 4814–4821.

Lleo, A., Blesa, R., Queralt, R., Ezquerra, M., Molinuevo, J.L., Pena-Casanova, J., Rojo, A. & Oliva, R. (2002) Frequency of mutations in the presenilin and amyloid precursor protein genes in early-onset Alzheimer disease in Spain. *Arch Neurol* **59**, 1759–1763.

Lobo, A., Launer, L.J., Fratiglioni, L., Andersen, K., Di Carlo, A., Breteler, M.M.B., Copeland, J.R.M., Dartigues, J.F., Jagger, C., Martinez-Lage, J., Soininen, H., Hofman, A. & Neurologic Dis Elderly Res, G. (2000) Prevalence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. *Neurology* **54**, S4–S9.

Lord, J., Lu, A.J. & Cruchaga, C. (2014) Identification of rare variants in Alzheimer's disease. *Front Genet* **5**, 369.

Madsen, B.E. & Browning, S.R. (2009) A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* **5**, e1000384.

Mairet-Coello, G., Courchet, J., Pieraut, S., Courchet, V., Maximov, A. & Polleux, F. (2013) The CAMKK2-AMPK kinase pathway mediates the synaptotoxic effects of Abeta oligomers through Tau phosphorylation. *Neuron* **78**, 94–108.

Mayeux, R. (2003) Epidemiology of neurodegeneration. *Annu Rev Neurosci* **26**, 81–104.

Mayeux, R., Sano, M., Chen, J., Tatemichi, T. & Stern, Y. (1991) Risk of dementia in 1st-degree relatives of patients with Alzheimer's-disease and related disorders. *Arch Neurol* **48**, 269–273.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. & Stadlan, E.M. (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–944.

Morgan, C. & Inestrosa, N.C. (2001) Interactions of laminin with the amyloid beta peptide. Implications for Alzheimer's disease. *Braz J Med Biol Res* **34**, 597–601.

Murphy, S.L., Xu, J. & Kochanek, K.D. (2013) Deaths: final data for 2010. In Prevention, National Vital Statistics Reports, Vol. 61, No. 4. pp. 1–118.

Naj, A.C., Jun, G., Beecham, G.W. *et al.* (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436–441.

Naj, A.C., Jun, G., Reitz, C. *et al.* (2014) Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. *JAMA Neurol* **71**, 1394–1404.

Nalls, M.A., Plagnol, V., Hernandez, D.G., Sharma, M., Sheerin, U.M., Saad, M., Simon-Sanchez, J., Schulte, C., Lesage, S., Sveinbjornsdottir, S., Stefansson, K., Martinez, M., Hardy, J., Heutink, P., Brice, A., Gasser, T., Singleton, A.B. & Wood, N.W. (2011) Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649.

Pastor, P., Roe, C., Villegas, A., Bedoya, G., Chakraverty, S., Garcia, G., Tirado, V., Norton, J., Rios, S., Martinez, M., Kosik, K., Lopera, F. & Goate, A. (2003) Apolipoprotein E $\varepsilon 4$ modifies Alzheimer's disease onset in an E280A PS1 Kindred. *Ann Neurol* **54**, 163–169.

Pedersen, N.L., Gatz, M., Berg, S. & Johansson, B. (2004) How heritable is Alzheimer's disease late in life? Findings from Swedish twins. *Ann Neurol* **55**, 180–185.

Prentice, R.L. & Pyke, R. (1979) Logistic disease incidence models and case–control studies. *Biometrika* **66**, 403–411.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. & Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. & Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575.

Radmanesh, F., Devan, W.J., Anderson, C.D., Rosand, J. & Falcone, G.J. (2014) Accuracy of imputation to infer unobserved APOE epsilon alleles in genome-wide genotyping data. *Eur J Hum Genet* **22**, 1239–1242.

Rocca, W.A., Hofman, A., Brayne, C. *et al.* (1991) Frequency and distribution of Alzheimer's-disease in Europe – a collaborative study of 1980–1990 prevalence findings. *Ann Neurol* **30**, 381–390.

Seshadri, S., Fitzpatrick, A.L., Ikram, M.A. *et al.* (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832–1840.

Shen, J.N., Wang, D.S. & Wang, R. (2012) The protection of acetylcholinesterase inhibitor on beta-amyloid-induced the injury of neurite outgrowth via regulating axon guidance related genes expression in neuronal cells. *Int J Clin Exp Pathol* **5**, 900–913.

Sherrington, R., Rogaev, E.I., Liang, Y., *et al.* (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754–760.

Simon-Sanchez, J., Schulte, C., Bras, J.M. *et al.* (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet* **41**, 1308–1312.

Sjögren, T., Sjögren, H. & Lindgren, A.G. (1952) *Morbus Alzheimer and Morbus Pick: A Genetic, Clinical and Patho-anatomical Study*. Munksgaard, Copenhagen.

Sosa-Ortiz, A.L., Acosta-Castillo, I. & Prince, M.J. (2012) Epidemiology of dementias and Alzheimer's disease. *Arch Med Res* **43**, 600–608.

Tandon, A. & Fraser, P. (2002) The presenilins. *Genome Biol* **3**, article 3014.1.

Thies, W., Bleiler, L. & Alzheimer's, A. (2013) 2013 Alzheimer's disease facts and figures Alzheimer's Association. *Alzheimers Dement* **9**, 208–245.

Thompson, E. (2011) The structure of genetic linkage data: from LIPED to 1 M SNPs. *Hum Hered* **71**, 86–96.

Van Duijn, C.M., Clayton, D., Chandra, V., Fratiglioni, L., Graves, A.B., Heyman, A., Jorm, A.F., Kokmen, E., Kondo, K., Mortimer, J.A., Rocca, W.A., Shalat, S.L., Soininen, H. & Hofman, A. (1991) Familial aggregation of Alzheimers-disease and related disorders – a collaborative reanalysis of case–control studies. *Int J Epidemiol* **20**, S13–S20.

Wang, K., Li, M. & Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164.

Wang, J., Duncan, D., Shi, Z. & Zhang, B. (2013) WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77–W83.

Wijsman, E.M. (2012) The role of large pedigrees in an era of high-throughput sequencing. *Hum Genet* **131**, 1555–1563.

Wijsman, E.M., Daw, E.W., Yu, C.E., Payami, H., Steinbart, E.J., Nochlin, D., Conlon, E.M., Bird, T.D. & Schellenberg, G.D. (2004) Evidence for a novel late-onset Alzheimer's disease locus on chromosome 19p13.2. *Am J Hum Genet* **75**, 398–409.

Wijsman, E.M., Daw, E.W., Yu, X., Steinbart, E.J., Nochlin, D., Bird, T.D. & Schellenberg, G. (2005) APOE and other loci affect age-at-onset in Alzheimer's disease families with PS2 mutation. *Am J Med Genet B Neuropsychiatr Genet* **132B**, 14–20.

Wijsman, E.M., Pankratz, N.D., Choi, Y., Rothstein, J.H., Faber, K.M., Cheng, R., Lee, J.H., Bird, T.D., Bennett, D.A., Diaz-Arrastia, R., Goate, A.M., Farlow, M.R., Ghetti, B., Sweet, R.A., Foroud, T.M. & Mayeux, R. (2011) Genome wide association of familial late onset Alzheimer's disease replicates *BIN1* and *CLU*, and nominates *CUGBP2* in interaction with *APOE*. *PLoS Genet* **7**, e1001308.

Wu, M.C., Lee, S., Cai, T.X., Li, Y., Boehnke, M. & Lin, X.H. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82–93.

Zhao, W., Marchani, E.E., Cheung, C.Y., Steinbart, E.J., Schellenberg, G.D., Bird, T.D. & Wijsman, E.M. (2013a) Genome scan in familial late-onset Alzheimer's disease: a locus on chromosome 6 contributes to age-at-onset. *Am J Med Genet B Neuropsychiatr Genet* **162B**, 201–212.

Zhao, W., Marchani, E.E., Cheung, C.Y.K., Steinbart, E.J., Schellenberg, G.D., Bird, T.D. & Wijsman, E.M. (2013b) Genome scan in familial late-onset Alzheimer's disease: a locus on chromosome 6 contributes to age at onset. *Am J Med Genet B Neuropsychiatr Genet* **162B**, 201–212.

Acknowledgments

This work was partially supported by grant funding from NIH R01 AG039700 and NIH P50 AG005136. Subjects and samples used here were originally collected with grant funding from NIH U24 AG026395, U24 AG021886, P50 AG008702, P01 AG007232, R37 AG015473, P30 AG028377, P50 AG05128, P50 AG16574, P30 AG010133, P50 AG005681, P01 AG003991, U01 MH046281, U01 MH046290 and U01 MH046373. The funders had no role in study design, analysis or preparation of the manuscript. The authors declare no competing interests.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Figure S1: Principal component analysis results. Our WES subjects and subjects from the three populations of the 1000 genomes project are plotted on the principal component 1 (PC1) vs. principal component 2 (PC2). Our WES subjects are represented by red crosses, and the EUR, AFR and ASN subjects are represented by blue, cyan and pink circles, respectively.

Figure S2: Manhattan plot of the burden test analysis (Weighted Sum Approach) that does not adjust for *APOE*. The plot is based on all 1949 tested genes. The horizontal blue line represents the least significant gene among the 10 most significant genes. The horizontal red line is the genome-wide significance threshold.

Figure S3: QQ-plot of the burden test analysis (Weighted Sum Approach) that adjusts (red) or does not adjust (black) for *APOE*. The plot is based on all 1949 tested genes.

Appendix S1: Genes' biological functions.

Table S1: Single-marker test results of the APOE region.

Table S2: Single-marker test results of SNPs in genes are discussed in Naj *et al.* (2014).

Table S3: Burden test results for genes are discussed in Naj *et al.* (2014).