ARTICLE

Genome-wide Association Study Implicates a Chromosome 12 Risk Locus for Late-Onset Alzheimer Disease

Gary W. Beecham,[1] Eden R. Martin,[1] Yi-Ju Li,[2] Michael A. Slifer,[1] John R. Gilbert,[1] Jonathan L. Haines,[3] and Margaret A. Pericak-Vance[1,*]

Only *Apolipoprotein E* polymorphisms have been consistently associated with the risk of late-onset Alzheimer disease (LOAD), but they represent only a minority of the underlying genetic effect. To identify additional LOAD risk loci, we performed a genome-wide association study (GWAS) on 492 LOAD cases and 498 cognitive controls using Illumina's HumanHap550 beadchip. An additional 238 cases and 220 controls were used as a validation data set for single-nucleotide polymorphisms (SNPs) that met genome-wide significance. To validate additional associated SNPs ($p < 0.0001$) and nominally associated candidate genes, we imputed SNPs from our GWAS using a previously published LOAD GWAS[1] and the IMPUTE program. Association testing was performed with the Cochran-Armitage trend test and logistic regression, and genome-wide significance was determined with the False Discovery Rate-Beta Uniform Mixture method. Extensive quality-control methods were performed at both the sample and the SNP level. The GWAS confirmed the known APOE association and identified association with a 12q13 locus at genome-wide significance; the 12q13 locus was confirmed in our validation data set. Four additional highly associated signals (1q42, 4q28, 6q14, 19q13) were replicated with the use of the imputed data set, and six candidate genes had SNPs with nominal association in both the GWAS and the joint imputated data set. These results help to further define the genetic architecture of LOAD.

Introduction

Alzheimer Disease (AD [MIM 104300]) is the leading cause of dementia in the elderly and has a complex etiology, with strong genetic and environmental determinants. *Apolipoprotein E* (*APOE* [MIM 107741]) is the single most significant genetic risk factor identified for late-onset AD (LOAD) and was identified as a risk gene primarily through genetic mapping.[2–5] Though APOE has been universally confirmed as a risk gene for LOAD, the risk polymorphism is neither necessary nor sufficient to cause AD, given that as much as 50% of the genetic-risk effect remains unexplained.[6] Efforts to identify additional AD loci have primarily taken the form of genome-wide linkage scans in multiplex families (two or more individuals with AD) and candidate-gene association studies. Though linkage scans were instrumental in detecting the effect of the *APOE* gene, they suffer from low resolution (signals often cover over 30 million base pairs) and have low power to detect smaller signals.[7] Candidate-gene studies use increased resolution, but their ability to replicate positive results has been both difficult and inconsistent.[8]

With the advent of genome-wide association studies (GWAS), we can now interrogate the entire genome with increased resolution and power. GWAS have already been completed for a variety of complex genetic diseases, with varying degrees of success.[1,9–15] Two published GWAS have examined LOAD, and both studies[13,14] convincingly confirmed the association of *APOE* to LOAD (p value = 1.0×10^{-39} and 2.3×10^{-44}, respectively), but neither has shown genome-wide significance at any SNP unlinked to *APOE*. This suggests that the remaining LOAD risk loci must be of small effect.

To identify the loci underlying the remaining genotypic effect, we present here a GWAS of LOAD, with 492 cases and 498 controls, using the Illumina HumanHap 550 beadchip. SNPs significant at the genome-wide level were genotyped in an independent validation data set. SNPs with strong association (p values < 0.0001) and nominally associated SNPs (p values < 0.05) in and near candidate genes were examined in a previous GWAS of AD (by Reiman et al.[1]) using an imputation procedure.[16]

Subjects and Methods

Ascertainment and Genotyping

Our analysis uses a clinic-based case-control design. The sample set is derived from the Collaborative Alzheimer Project (CAP, the Miami Institute for Human Genomics at the University of Miami Medical Center and the Center for Human Genetics Research at Vanderbilt University Medical Center). The CAP data set utilized for this report is independent from previously published data sets.

After complete description of the study to the subjects, written informed consent was obtained from all participants, in agreement with protocols approved by the institutional review board at each contributing center. For inclusion, each LOAD affected individual met the NINCDS-ADRDA criteria for probable or definite AD and had an age at onset (AAO) greater than 60 years of

[1]Miami Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL 33101, USA; [2]Center for Human Genetics and Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA; [3]Vanderbilt Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN 37232, USA
*Correspondence: mpericak@med.miami.edu
DOI 10.1016/j.ajhg.2008.12.008. ©2009 by The American Society of Human Genetics. All rights reserved.

age.[17] Subjects' AAO for LOAD was determined from specific probe questions within the clinical history provided by a reliable family informant or from documented significant impairment in the medical record. Cognitive controls were spouses, friends, and other biologically unrelated individuals who were frequency matched by age and gender to the cases and were from within the same clinical catchment areas. All cognitive controls were examined, and none showed signs of dementia in clinical history or upon interview. Additionally, each cognitive control had a documented Mini-Mental State Exam (MMSE) score ≥ 27 or a Modified Mini-Mental State (3MS) Exam score ≥ 87;[18] of the controls, 78% had documented 3MS Exams and 22% had documented MMSEs. The preliminary GWAS cohort contained a total of 1086 individuals of European descent. There were 529 LOAD cases, average age 71.7 years at onset (+/− 7.2 years), and 557 cognitive controls, average age 74.4 years at exam (+/− 5.9 years). Each group was 63.5% female.

From this preliminary GWAS cohort, we genotyped 1049 individuals (518 cases and 531 controls; Table 1). After genotyping and before the statistical analysis, samples had to pass a stringent set of quality control tests, so that the integrity of the genetic data was ensured. The final GWAS data set analyzed contains a total of 988 individuals of European descent. There are 492 LOAD cases, average age 72.9 years at onset (+/− 6.6 years), and 496 cognitive controls, average age 74.3 years at exam (+/− 6.5 years). Cases are 61% female, and controls are 63% female.

The validation data set consisted of 238 LOAD cases and 220 controls—independent of the preliminary cohort—that were subjected to the same inclusion criteria as those in the GWAS data set. The cases averaged 67.7 years AAO (+/− 8.6 years), and the controls averaged 70.5 years age at exam (+/− 6.5).

Table 1. GWAS Sample Information

	All	Cases	Controls
Total	988	492	496
Male:Female Ratio	372:616 (1:1.66)	180:312 (1:1.73)	192:304 (1:1.58)
AAO[a] or AAE[b]		72.9 (+/− 5.5)	74.2 (+/− 5.6)
APOE −/−[c] carriers	547 (55.7%)	169 (34.5%)	378 (76.6%)
APOE −/4[d] carriers	339 (34.5%)	234 (47.9%)	105 (21.3%)
APOE 4/4[e] carriers	96 (9.8%)	86 (17.6%)	10 (2.0%)
Efficiency[f]	99.8%	99.8%	99.8%

A description of the GWAS analysis cohort.
[a] Age at onset (cases).
[b] Age at exam (controls).
[c] Samples with no APOE e4 alleles.
[d] Samples with only one APOE e4 allele.
[e] Samples with two APOE e4 alleles.
[f] Percentage of successfully genotyped SNPs among those attempted.

Genotyping

We extracted DNA for individuals ascertained by the CAP by using Puregene chemistry (QIAGEN, Germantown, MD, USA). We performed genotyping using the Illumina Beadstation and the Illumina HumanHap 550 beadchip, following the recommended conditions, with the exception that we required the more conservative gencall score of 0.25. Genotyping efficiency was greater than 99%, and quality assurance was achieved by the inclusion of two CEPH controls that were genotyped multiple times. The lab was blinded to affection status and quality-control samples. The ABI 7900 Taqman system was used for generating *APOE* genotypes corresponding to allele combinations at SNP +3937/rs429358 and SNP +4075/rs7412.

Sample-Quality Control

After genotyping, samples were subjected to a battery of quality-control tests. One measure of the overall quality of a sample's data is sample efficiency; the proportion of valid genotype calls to attempted calls within a sample. Samples with efficiency less than 0.98 were dropped from the analysis. Many of these samples were previously genotyped on the Illumina Goldengate and/or ABI Taqman platforms for SNPs that were in the GWAS (80% of samples were previously typed at 100 or more SNPs; average = 346, median = 428). This duplication validates that the sample was correctly acquisitioned and that the Infinium II assay was accurate. Samples with less than 90% genotype-concordance rates on 100 or more previously typed SNPs were dropped from the analysis. Reported gender and genetic gender were examined with the use of X-linked SNPs; inconsistent samples were dropped from the analysis. Relatedness between samples was tested via the program Graphical Representation of Relatedness (GRR),[19] and related samples were dropped from the analysis.

A set of 3500 independent SNPs (not in strong linkage disequilibrium [LD], $r^2 < 0.16$) spread evenly across the autosomal chromosomes were analyzed in STRUCTURE[20] for evidence of population substructure (burn in: 1000, iterations: 20,000). In addition to this first run, we ran 250 SNPs with twice the number of iterations. We also used the program EigenStrat to look for population substructure. EigenStrat is a principle-components-analysis program that utilizes eigenvalues to investigate substructure and to potentially correct for it.[21] A set of 20,000 SNPs across the genome was used.

SNP-Quality Control

SNPs were subjected to several tests for quality before being analyzed. Genotypes were first recalled on the basis of our own data, per Illumina's recommendations. Recalling corrects missed calls due to ill-defined HapMap clusters and eliminates SNPs for which the platform is inconsistent. Only samples with efficiency greater than 0.98 were used for redefining the genotype clusters. SNP efficiency is calculated as the percentage of samples that have genotype calls for a given SNP. All SNPs with less than 90% efficiency were dropped from the analysis. SNPs with MAF < 0.005 were dropped, because even under highly optimistic conditions (high risk ratio, direct ascertainment of the disease locus), these SNPs have 50% power at best. To reduce error, we subjected SNPs with MAF < 0.10 to a more stringent efficiency cutoff of 99%. SNPs could have significant Hardy-Weinberg disequilibrium statistics for legitimate biological reasons and could have even been used for disease inference.[22,23] Laboratory-process errors typically lead to very extreme disequilibrium, so SNPs were only dropped when the HWD statistic was significant at the $p < 10^{-6}$ level. HWD statistics were calculated with the Fisher's exact test in the PLINK package.[24]

Association Analysis

Association analysis was performed with the use of the Cochran-Armitage trend test for association.[25] This method tests for a linear

trend in the number of alleles at a single locus. That is, two copies of an allele have more of an effect than one copy, which in turn has more of an effect than no copies. The effect is in the same direction for each genotype. This test is equivalent to the score statistic from a logistic-regression model with no covariates. In addition to the standard trend test, we performed logistic regression, with *APOE* status, age at onset (cases) or exam (controls), and gender as covariates. All analyses were performed via PLINK.[24] *APOE* status was designated as the number of e4 alleles. A genome-wide multiple-testing correction was applied with a false-discovery rate, with the use of the beta-uniform distribution.[26] SNPs with FDR q values less than 0.20 were declared significant. Initial haplotyping was performed with the Haploview software[27] using the confidence-interval-based block definitions,[28] and follow-up was performed with the Haplo.Stats software.[29,30]

Imputation Analysis

The software IMPUTE[16] was used for imputing genotype data. Both our data and the data from the previous GWAS[1] were imputed, independently, to a HapMap reference of over 2.5 million SNPs. Individual genotypes with probability less than 0.90 were not included, and SNPs missing > 10% of genotypes within either data set were dropped from the joint analysis. Joint analysis was performed with PLINK.[24] Association testing was performed in PLINK, with logistic regression, with an indicator variable of study of origin included as a covariate.

Results

Genotypes were initially generated on 518 LOAD cases and 531 cognitive controls for 555,000 SNPs. Stringent quality-control criteria were required for all samples and markers. Of the initial 1049 samples, 988 met the quality-control criteria (492 cases, 498 controls; average genotyping efficiency > 99.8%). There were 31 samples (3%) dropped because their efficiencies were less than 98%, and 17 samples were dropped because their concordance rates were less than 90%. Nine samples were dropped because the genotypic gender disagreed with the clinical information (five males that tested female, four females that tested male), and three samples were dropped because of their relatedness to other samples. One additional sample was dropped for clinical reasons. Of the 555,000 SNPs, only 23,000 (4%) were dropped from the analysis (average minor-allele frequency of the remaining SNPs = 0.246). Samples were tested for population substructure, and none was found. In STRUCTURE, there were no samples that consistently clustered in the same groups and there was no observation of bimodality or outliers in the plots. In Eigenstrat, the top PCA components accounted for only a small percentage of variation (< 3%) and there was no bimodality or outliers in the plots of the top principal components.

There were 38 SNPs with uncorrected p values < 0.00005 for association to LOAD using the Cochran-Armitage trend test, six of which were in or near the *APOE* gene (Table 2; complete results in Figure 1), including the top three (not shown). The LOAD association at *APOE* represents a positive control. The remaining 32 SNPs span the genome,

Table 2. Single-Nucleotide Polymorphisms with a p Value Less Than 5E−5

SNP	Chr.	BP[a]	p Value	Gene	Role
rs1415985	1	49,703,336	1.23E−05		
rs11205641	1	49,957,662	8.41E−06		
rs4926831	1	50,062,688	1.23E−05		
rs9659092	1	50,216,176	4.54E−06		
rs11583200	1	50,332,407	1.83E−05		
rs11683103	2	34,766,354	8.58E−06		
rs2119067	2	165,835,529	4.38E−05		
rs10184275	2	165,836,174	2.20E−05		
rs2681411	3	123,268,321	4.21E−05	*CD86*	Intron
rs12639920	4	42,107,444	4.85E−05	*ATP8A1*	Downstream
rs3807031	6	30,141,863	1.16E−05	*PPP1R11*	Promoter
rs929156	6	30,247,678	1.69E−05	*TRIM15*	Intron
rs11754661	6	151,248,771	2.01E−05	*MTHFD1L*	Intron
rs9455973	6	168,325,855	4.47E−05		
rs6942930	7	1,518,946	1.61E−05		
rs2039461	9	20,135,988	3.48E−05		
rs7893928	10	44,398,949	2.31E−05		
rs11610206*	12	45,925,793	1.43E−06	*FAM113B*	Downstream
rs2387100	13	27,324,759	3.82E−05		
rs9544105	13	75,456,154	5.41E−06		
rs659628	13	76,361,237	4.46E−05	*KCTD12*	Promoter
rs12146962	14	32,450,849	7.25E−06		
rs4555132	15	95,740,242	3.08E−05		
rs1480090	15	96,533,184	3.52E−05		
rs1383139	15	96,535,200	3.48E−05		
rs1402627	18	4,123,739	4.42E−05		
rs4459653	19	49,291,455	8.00E−06	*ZNF224*	Intron
rs4802207	19	49,292,217	9.23E−06	*ZNF224*	Intron
rs3746319	19	49,304,071	2.96E−05	*ZNF224*	Coding exon
rs2061332	19	49,305,501	3.93E−05	*ZNF224*	Downstream
rs6059244	20	29,474,144	4.76E−05		
rs2180566	20	29,482,515	3.80E−05	*DEFB123*	Promoter

SNPs in the GWAS with p values $< 5 \times 10^{-5}$, based on 492 cases and 496 controls. p values are calculated with the Cochran-Armitage trend test and are uncorrected for multiple testing. *APOE*-linked SNPs have been removed. Asterisk indicates the SNP that met genome-wide significance.
[a] BP indicates position in base pairs.

representing 19 distinct signals across 16 chromosomes. There was little change in this list when logistic regression with covariates was applied instead of the trend test (sex, age at onset or at exam, and *APOEe4*-carrier status as covariates). The majority of these signals (12 of 19) lie in regions that have previously shown genetic linkage to LOAD through other studies.[8]

The most significant non-*APOE* SNP was rs11610206 on 12q13 (45.92 Mb). This SNP met genome-wide significance criteria with the use of the False Discovery Rate-Beta Uniform Mixture (FDR-BUM)[26] multiple-testing-correction criteria. The uncorrected p value was 1.93×10^{-6} (FDR = 0.17). Because this SNP met our significance criteria, we genotyped the marker in an independent data set. The marker was significant in our independent replication data set of 238 cases and 220 controls (p = 0.0496). The association was in the same direction, and the joint analysis had a p value of 3.452×10^{-7}, nearly an order of magnitude more significant than in the initial data set. There is some mild LD structure in this region, but a haplotype analysis

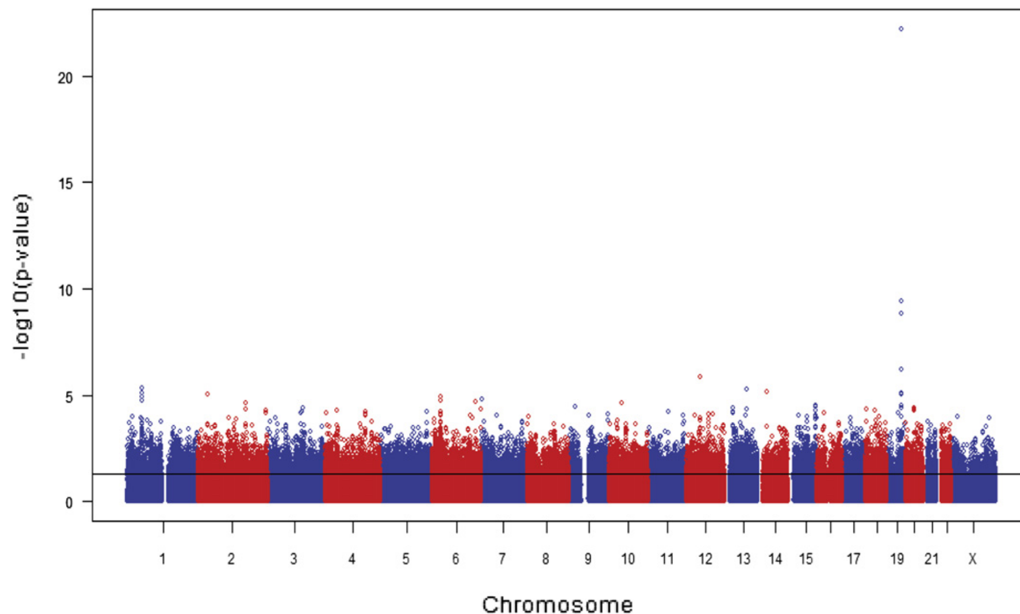

Figure 1. Plot of GWAS Results
This plot shows the results of our GWAS. The results are reported as –log10(p value) by genomic position. The horizontal line indicates the 0.05 p value cutoff.

of this and surrounding SNPs does not reveal any stronger association than that of the rs11610206 SNP alone. There are a number of genetic linkage results on 12q13.[31–34] In particular, the broad linkage signals observed on 12q were narrowed considerably in the Liang et al. study[34] using an ordered-subset analysis (44 Mb–48 Mb). This association lies directly under the Liang et al. linkage signal and represents a confirmation of that signal in an independent data set; no individuals from the families in the Liang et al. study were used in our case-control cohort. Other than three of the *APOE*-linked SNPs, there were no additional loci that met the FDR threshold.

To validate additional associated SNPs, we used an imputation approach. Both our GWAS and the previously published GWAS[1] were imputed to a HapMap reference with the use of IMPUTE,[16] and the common SNPs were the basis for comparison. We first compared the strongly associated results from each study ($p < 0.0001$), and we then examined nominally associated markers within known candidate genes.

Among the top signals in the GWAS, there were four that showed association in both studies (Table 3). Two of these signals, 1q42 and 19q13, are within genes. The 1q42 signal (rs12044355) has the following p values: $p_B = 0.026$; $p_R = 0.000044$; $p_J = 0.0000020$ (in which p_B is the p value in our data set, p_R is the p value in the Reiman[1] data set, and p_J is the p value in the joint analysis). It is within an intron of the *disrupted in schizophrenia 1* (*DISC1*) gene.

The 19q13 signal is in and near exon 6 of *zinc finger protein 224* (*ZNF224* [MIM 194555]). Two of the associated markers (rs4508518 and rs3746319) are within the exon. The first, rs4508518 ($p_B = 0.000039$, $p_R = 0.0082$, $p_J = 0.0000092$), is a coding but synonymous polymorphism, whereas the second (rs3746319; $p_B = 0.000036$, $p_R = 0.01$, $p_J = 0.000011$) leads to a missense mutation. The *ZNF224* signal is 800 kb proximal to *APOE* but is not in LD to *APOE* (Table 4). Additionally, logistic regression of our data showed that the association of the *ZNF224* signal was not greatly diminished when *APOEe4*-carrier status was included as a covariate. The rs20612332 SNP has a p value equal to 0.000030 without *APOEe4*-carrier status as a covariate and a p value equal to 0.000038 with carrier status as a covariate. This confirms that the signal is independent of *APOE*.

The two other signals replicated in both data sets are not in known genes. The gene nearest the chromosome 6 signal is *branched chain keto acid dehydrogenase E1, beta polypeptide* (*BCKDHB* [MIM 248611]) but is over 800 kb proximal to the SNP. The chromosome 4 signal is 200 kb proximal to *protocadherin 18* (*PCDH18* [MIM 608287]), a protocadherin precursor that is thought to play a role in cell-cell connections in the brain.

In addition to these top hits, nine candidate genes from the over 500 genes in the AlzGene candidate-gene list[1] have SNPs with nominal association in both GWASs (Table 5). These genes (*ADAM12*, *CSF1*, *GBP2*, *KCNMA1*, *NOS2A*, *SORCS2*, *SORCS3*, *SORL1*, *WWC1* [MIM 602714, 120420, 600412, 600150, 163730, 606284, 606285, 602005, 611675, respectively]) had p values ranging from 0.003 to 0.05 in the individual GWAS and from 0.0001 to 0.01 in the joint analysis. Of the 21 nominally associated SNPs, 19 were intronic, and the remaining 2 are downstream from the gene.

Discussion

We have shown genome-wide association of the SNP rs11610206 with LOAD and have validated this signal in

Table 3. Top Association Signals that Were Replicated in Both GWAS

SNP	Chr.	BP[a]	Type	P_B[b]	P_R[c]	P_J[d]	Gene	Role
rs12044355	1	229,910,970	R	3.90E-05	0.008216	9.20E-06	*DISC1*	Intron
rs1425967	4	138,508,340	R	3.90E-05	0.01052	1.25E-05		
rs4416533	4	138,546,322	I	3.61E-05	0.0101	1.13E-05		
rs13213247	6	81,572,755	R	4.73E-05	0.01587	2.40E-05		
rs4508518	19	49,303,260	I	0.02627	4.37E-05	1.95E-06	*ZNF224*	Coding exon
rs3746319	19	49,304,071	B	6.05E-05	0.02326	3.01E-06	*ZNF224*	Coding exon
rs2061332	19	49,305,501	B	6.19E-05	0.03786	4.91E-06	*ZNF224*	Downstream
rs2061333	19	49,306,048	I	0.01745	2.51E-05	1.51E-06	*ZNF224*	Downstream

SNPs in either GWAS with a p value < 0.0001 that was replicated in the other GWAS. p values calculated are uncorrected for multiple testing. "Type" refers to how the marker was genotyped; Type B markers were genotyped in the Beecham GWAS and imputed in the Reiman samples, Type R markers were genotyped in the Reiman GWAS and imputed in the Beecham samples, and Type I markers were imputed in both GWAS.
[a] BP indicates position in base pairs.
[b] P_B indicates the p values from this study.
[c] P_R indicates the p values from the study by Reiman et al.[1]
[d] P_J indicates the p values from the joint analysis.

an independent case-control data set. This provides strong evidence for a risk locus on 12q13. The SNP is not in a known gene but is less than 10 kb from the hypothetical gene *FAM113B*. Additionally, there are a number of nearby candidate genes, such as the *vitamin D (1,25- dihydroxyvitamin D3) receptor* (*VDR* [MIM 601769]) and *adhesion molecule with Ig-like domain 2* (*AMIGO2*). *VDR* is the most appealing of the candidate genes. There has been association with *VDR* reported,[35] and *VDR* has been associated with memory performance.[36] There is no known connection between our top SNP and *VDR*, but the region between the two is largely uncharacterized; it is possible that the top SNP could be in a long-range regulatory element that influences *VDR*.

It is of note that the rs11610206 SNP was not imputed in the Reiman[1] data with enough confidence to allow inclusion in the imputation analysis. This demonstrates one of the weaknesses of imputation. If there is not strong LD between a genotyped SNP and an untyped SNP of interest, the untyped SNP will not be imputed with high confidence. In this case, there is not extended LD around rs11610206, so the nearest SNPs in the Reiman GWAS were not sufficiently informative for imputation. This same phenomenon was seen at the *APOE* locus. The two data sets did not share any SNP near *APOE*, and nearby HapMap SNPs were not imputed with confidence. In the end, the signal at *APOE*—highly significant in each individual GWAS—is missed entirely in the joint imputation analysis unless quality control standards are lowered. Indeed, nearly 20% of the top SNPs from our GWAS failed to be imputed in the Reiman data.

Four of the top hits among the GWAS were validated in the imputation analysis. The 1q42 and 19q13 signals are of particular interest. The 1q42 signal resides in the *DISC1* gene, a gene that has been associated with schizophrenia and has links to bipolar disorder, depression, and cognitive function.[37–41] The 19q13 signal lies in the exon of the *ZNF224* gene, and several of the SNPs were coding SNPs, including one missense mutation. Although this is not the first report of a non-*APOE* signal on 19q13,[42,43] it is the first time the *ZNF224* gene has been implicated specifically.

There were eight candidate genes from the AlzGene list with SNPs associated in both GWASs. Principal among these genes is *sortlin-related receptor* (*SORL1*), a gene that has received much attention in LOAD genetics. *SORL1* (alternatively *LR11* or *SorLA*) has been associated with LOAD in a variety of populations.[44–47] Replication has been inconsistent,[45,48,49] and it is thought that there could be extensive locus and allelic heterogeneity involved.[44,50]

Table 4. LD between *ZNF224* SNPs and *APOE*-Linked SNPs

Gene	SNP	Position	rs4802207	rs3746319	rs2061332	rs2075650	rs8106922	rs405509	rs439401
ZNF224	rs4459653	19: 49,305,501	0.94	0.92	0.92	0	0.01	0.01	0
	rs4802207	19: 49,306,048		0.95	0.95	0	0.01	0.01	0
	rs3746319	19: 49,304,071			1.00	0	0.01	0.01	0
	rs2061332	19: 49,303,260				0	0.01	0.01	0
APOE	rs2075650	19: 50,087,459					0.16	0.24	0.08
	rs8106922	19: 50,093,506						0.60	0.08
	rs405509	19: 50,100,675							0.17
	rs439401	19: 50,106,291							

LD between the *ZNF224* SNPs on 19q13 (rs4459653, rs4802207, rs3746319, rs2061332) and SNPs most linked to *APOE* on 19q13 (rs2075650, rs8106922, rs405509, rs439401). Disequilibrium is reported as r^2. Position is reported in base pairs. This shows that there is a single *ZNF224* signal that is independent from the *APOE* signal.

Table 5. **Candidate Genes with SNPs Significant in Both GWAS**

Gene	SNP	Chr.	BP[a]	Type	P_B[b]	P_R[c]	P_J[d]	OR[e]
ADAM12	rs11244841	10	127,824,556	B	0.04379	0.04551	0.003386	1.2180
CSF1	rs7537752	1	110,186,484	R	0.03273	0.0378	0.002087	0.8082
GBP2	rs10922573	1	89,300,401	B	0.01481	0.00854	0.000833	1.2190
GBP2	rs12725861	1	89,278,117	I	0.008497	0.01599	0.000945	1.2170
GBP2	rs6428503	1	89,296,090	B	0.01165	0.00854	0.000665	1.2230
KCNMA1	rs16934131	10	78,407,601	I	0.04519	0.03569	0.003338	1.2210
NOS2A	rs11653716	17	23,108,659	I	0.003526	0.007249	0.00014	0.4845
SORCS2	rs3846421	4	7,403,428	B	0.003131	0.0206	0.000117	0.7805
SORCS3	rs10786828	10	106,599,890	B	0.04546	0.03495	0.004627	1.1800
SORCS3	rs7894737	10	106,603,320	I	0.04694	0.04736	0.004345	1.1840
SORL1	rs11218342	11	120,939,638	I	0.04825	0.04859	0.008507	0.5509
SORL1	rs11218343	11	120,940,797	I	0.04825	0.04859	0.008507	0.5509
SORL1	rs1784919	11	120,944,875	I	0.04825	0.04813	0.008433	0.5505
SORL1	rs1792124	11	120,946,730	I	0.04825	0.04813	0.008433	0.5505
SORL1	rs2298814	11	120,930,092	I	0.04825	0.04906	0.008583	0.5513
SORL1	rs3781835	11	120,953,464	B	0.04825	0.03458	0.006237	0.5353
SORL1	rs3781838	11	120,958,727	I	0.03072	0.03314	0.004064	0.5157
SORL1	rs6589885	11	120,931,252	I	0.04825	0.04906	0.008583	0.5513
SORL1	rs720099	11	120,939,003	I	0.04825	0.04859	0.008507	0.5509
SORL1	rs7946599	11	120,928,850	I	0.04825	0.04906	0.008583	0.5513
WWC1	rs12514426	5	167,826,286	I	0.03592	0.004984	0.000928	0.5430

SNPs in candidate genes associated with LOAD in both GWAS and the joint analysis. p values are uncorrected for multiple testing. "Type" refers to how the marker was genotyped; Type B markers were genotyped in this GWAS and imputed in the Reiman samples, Type R markers were genotyped in the Reiman GWAS and imputed in the samples from this GWAS, and Type I markers were imputed in both GWAS.
[a] BP indicates position in base pairs.
[b] P_B indicates the p values from this study.
[c] P_R indicates the p values from the Reiman et al. study.[1]
[d] P_J indicates the p values from the joint analysis.
[e] OR indicates odds-ratio estimates.

There are also multiple studies that show that *SORL1* expression is decreased in Alzheimer disease and in the cognitively impaired brain.[51–53] Although there are no consensus *SORL1* mechanisms that confer LOAD risk, it is known that the SORL1 protein interacts with both APOE protein and amyloid beta (A4) precursor protein (APP [MIM 104760]).[44,54] The findings of association in our GWAS, as well as in the joint analysis with the Reiman GWAS, further confirm *SORL1* as a risk gene for LOAD.

Also among the nominally associated genes are *guanylate binding protein 2, interferon-inducible* (*GBP2*), which is upregulated in the hippocampus in AD and has previously shown nominal significance to AD,[55] and the gene *WW and C2 domain containing 1* (*WWC1*), which has shown association with AD in a Spanish population.[56] *WWC1* has also been associated with memory performance based on a verbal-memory task.[57]

It is of note that multiple testing is an issue with the imputation analysis. There are many tests, because the imputation provides a dense map; this suggests a more stringent threshold. However, the tests are highly correlated as a result of LD, and there is a priori evidence for the candidate-gene SNPs, suggesting a more relaxed threshold. Rather than arbitrarily quantifying a statistical prior or establishing a highly arbitrary significance threshold, we report uncorrected p values and look for concordance between the two GWAS.

We have shown a genome-wide significant association between the 12q13 SNP rs11610206 and late-onset Alzheimer disease. This signal was replicated in an independent case-control cohort. The region around this SNP is largely uncharacterized, and further delineation of possible candidates near this SNP is needed. We have also identified four regions (1q42, 4q28, 6q14, 19q13) with strong association to AD that were replicated in the imputation analysis, confirmed the association of *SORL1* to LOAD, and validated a number of candidate genes with nominal association in both GWAS. Detailed functional examination of these signals and genes could lead to a better understanding of the complex pathophysiology of Alzheimer disease.

Acknowledgments

We thank the AD patients and their families for participating in our study. We also thank the clinical and research staff of the Miami Institute for Human Genomics and the Vanderbilt Center for Human Genetics Research; particularly Paul Gallins, Carol Haynes, Chris Potocky, Karen Soldano, Heidi Munger, Simon Gregory, and Susan Slifer. This work was supported by the National Institute on Aging (grants AG20135 and AG19757) and the National Institute of Neurological Disorders and Stroke (grant NS31153), both of the National Institutes of Health; the Alzheimer's Association; and the Louis D. Scientific Award of the Institut de France. A subset of the participants was ascertained while M.A.P.-V. was a faculty member at Duke University.

Received: September 4, 2008
Revised: November 4, 2008
Accepted: December 12, 2008
Published online: December 31, 2008

Web Resources

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

References

1. Reiman, E.M., Webster, J.A., Myers, A.J., Hardy, J., Dunckley, T., Zismann, V.L., Joshipura, K.D., Pearson, J.V., Hu-Lince, D., Huentelman, M.J., et al. (2007). GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. Neuron *5*, 713–720.
2. Pericak-Vance, M.A., Bebout, J.L., Gaskell, P.C., Yamaoka, L.H., Hung, W.Y., Alberts, M.J., Walker, A.P., Bartlett, R.J., Haynes, C.S., Welsh, K.A., et al. (1991). Linkage studies in familial Alzheimer's disease: Evidence for chromosome 19 linkage. Am. J. Hum. Genet. *48*, 1034–1050.
3. Saunders, A.M., Strittmatter, W.J., Schmechel, D., George-Hyslop, P.H., Pericak-Vance, M.A., Joo, S.H., Rosi, B.L., Gusella, J.F., Crapper-MacLachlan, D.R., and Alberts, M.J. (1993). Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. Neurology *43*, 1467–1472.
4. Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S., and Roses, A.D. (1993). Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc. Natl. Acad. Sci. USA *90*, 1977–1981.
5. Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., and 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 *512*, 921–923.
6. Bergen, A.L. (1994). Heredity in dementia of the Alzheimer type. Clin. Genet. *46*, 144–149.
7. Lange, C., Blacker, D., and Laird, N.M. (2004). Family-based association tests for survival and times-to-onset analysis. Stat. Med. *30*, 179–189.
8. Bertram, L., McQueen, M.B., Mullin, K., Blacker, D., and Tanzi, R.E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat. Genet. *1*, 17–23.
9. Maraganore, D.M., De Andrade, M., Lesnick, T.G., Strain, K.J., Farrer, M.J., Rocca, W.A., Pant, P.V., Frazer, K.A., Cox, D.R., and Ballinger, D.G. (2005). High-resolution whole-genome association study of Parkinson disease. Am. J. Hum. Genet. *77*, 685–693.
10. Herbert, A., Liu, C., Karamohamed, S., Liu, J., Manning, A., Fox, C.S., Meigs, J.B., and Cupples, L.A. (2006). BMI modifies associations of IL-6 genotypes with insulin resistance: the Framingham Study. Obesity (Silver Spring) *8*, 1454–1461.
11. Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A., et al. (2006). A genome-wide association study identifies il23r as an inflammatory bowel disease gene. Science *314*, 1461–1463.
12. Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., Vermeire, S., Dewit, O., de Vos, M., Dixon, A., et al. (2007). Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet. *4*, e58.
13. Coon, K.D., Myers, A.J., Craig, D.W., Webster, J.A., Pearson, J.V., Lince, D.H., Zismann, V.L., Beach, T.G., Leung, D., Bryden, L., et al. (2007). 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 *4*, 613–618.
14. Li, H., Wetten, S., Li, L., St Jean, P.L., Upmanyu, R., Surh, L., Hosford, D., Barnes, M.R., Briley, J.D., Borrie, M., et al. (2008). Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch. Neurol. *1*, 45–53.
15. International Multiple Sclerosis Genetics Consortium, Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B., et al. (2007). Risk alleles for multiple sclerosis identified by a genomewide study. N. Engl. J. Med. *9*, 851–862.
16. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies via imputation of genotypes. Nat. Genet. *39*, 906–913.
17. McKhann, G., Drachman, D., and Folstein, M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology *34*, 939–944.
18. Teng, E.L., and Chui, H.C. (1987). The modified Mini-Mental State (3MS) examination. J. Clin. Psychiatry *48*, 314–318.
19. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2001). GRR: graphical representation of relationship errors. Bioinformatics *17*, 742–743.
20. Pritchard, J.K., Stephens, M., Rosenberg, N.A., and Donnelly, P. (2000). Association mapping in structured populations. Am. J. Hum. Genet. *67*, 170–181.
21. Patterson, N., Price, A.L., and Reich, D. (2006). Population Structure and Eigenanalysis. PLoS. Genet. *1*, e190.
22. Wittke-Thompson, J.K., Pluzhnikov, A., and Cox, N.J. (2005). Rational inferences about departures from Hardy-Weinberg equilibrium. Am. J. Hum. Genet. *6*, 967–986.
23. Ryckman, K., and Williams, S.M. (2008). Calculation and use of the Hardy-Weinberg model in association studies. Curr Protoc Hum Genet. (New York: John Wiley & Sons).
24. 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., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *3*, 559–575.
25. Armitage, P. (1955). Tests for linear trends in proportions and frequencies. Biometrics *11*, 375–386.
26. Pounds, S., and Morris, S.W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p values. Bioinformatics *1*, 1236–1242.
27. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263–265.
28. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart,

M., et al. (2002). The structure of haplotype blocks in the human genome. Science *557*, 2225–2229.
29. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M., and Poland, G.A. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am. J. Hum. Genet. *70*, 425–434.
30. Lake, S.L., Lyon, H., Tantisira, K., Silverman, E.K., Weiss, S.T., Laird, N.M., and Schaid, D.J. (2003). Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. Hum. Hered. *55*, 56–65.
31. Myers, A., Wavrant De-Vrieze, F., Holmans, P., Hamshere, M., Crook, R., Compton, D., Marshall, H., Meyer, D., Shears, S., Booth, J., et al. (2002). Full genome screen for Alzheimer disease: stage II analysis. Am. J. Med. Genet. *114*, 235–244.
32. Pericak-Vance, M.A., Bass, M.P., Yamaoka, L.H., Gaskell, P.C., Scott, W.K., Terwedow, H.A., Menold, M.M., Conneally, P.M., Small, G.W., Vance, J.M., et al. (1997). Complete genomic screen in late-onset familial Alzheimer disease. Evidence for a new locus on chromosome 12. JAMA *278*, 1237–1241.
33. Mayeux, R., Lee, J.H., Romas, S.N., Mayo, D., Santana, V., Williamson, J., Ciappa, A., Rondon, H.Z., Estevez, P., Lantigua, R., et al. (2002). Chromosome-12 mapping of late-onset Alzheimer disease among Caribbean Hispanics. Am. J. Hum. Genet. *70*, 237–243.
34. Liang, X., Schnetz-Boutaud, N., Kenealy, S.J., Jiang, L., Bartlett, J., Lynch, B., Gaskell, P.C., Gwirtsman, H., McFarland, L., Bembe, M.L., et al. (2006). Covariate analysis of late-onset Alzheimer disease refines the chromosome 12 locus. Mol Psych *11*, 280–285.
35. Gezen-Ak, D., Dursun, E., Ertan, T., Hanağasi, H., Gürvit, H., Emre, M., Eker, E., Oztürk, M., Engin, F., and Yilmazer, S. (2007). Association between vitamin D receptor gene polymorphism and Alzheimer's disease. Tohoku J. Exp. Med. *212*, 275–282.
36. Przybelski, R.J., and Binkley, N.C. (2007). Is vitamin D important for preserving cognition? A positive correlation of serum 25-hydroxyvitamin D concentration with cognitive function. Arch. Biochem. Biophys. *460*, 202–205.
37. Kilpinen, H., Ylisaukko-Oja, T., Hennah, W., Palo, O.M., Varilo, T., Vanhala, R., Nieminen-von Wendt, T., von Wendt, L., Paunio, T., and Peltonen, L. (2008). Association of Dros. Inf. Serv.C1 with autism and Asperger syndrome. Mol. Psychiatry *13*, 187–196.
38. Perlis, R.H., Purcell, S., Fagerness, J., Kirby, A., Petryshen, T.L., Fan, J., and Sklar, P. (2008). Family-based association study of lithium-related and other candidate genes in bipolar disorder. Arch. Gen. Psychiatry *65*, 53–61.
39. Hodgkinson, C.A., Goldman, D., Jaeger, J., Persaud, S., Kane, J.M., Lipsky, R.H., and Malhotra, A.K. (2004). Disrupted in schizophrenia 1 (Dros. Inf. Serv.C1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. Am. J. Hum. Genet. *75*, 862–872.
40. Cannon, T.D., Hennah, W., van Erp, T.G., Thompson, P.M., Lonnqvist, J., Huttunen, M., Gasperoni, T., Tuulio-Henriksson, A., Pirkola, T., Toga, A.W., et al. (2005). Association of Dros. Inf. Serv.C1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. Arch. Gen. Psychiatry *62*, 1205–1213.
41. Hennah, W., Tuulio-Henriksson, A., Paunio, T., Ekelund, J., Varilo, T., Partonen, T., Cannon, T.D., Lönnqvist, J., and Peltonen, L. (2005). A haplotype within the Dros. Inf. Serv.C1 gene is associated with visual memory functions in families with a high density of schizophrenia. Mol. Psychiatry *10*, 1097–1103.
42. Li, Y., Nowotny, P., Holmans, P., Smemo, S., Kauwe, J.S., Hinrichs, A.L., Tacey, K., Doil, L., van Luchene, R., Garcia, V., et al. (2004). Association of late-onset Alzheimer's disease with genetic variation in multiple members of the GAPD gene family. Proc. Natl. Acad. Sci. USA *101*, 15688–15693.
43. Luedecking, E.K., DeKosky, S.T., Mehdi, H., Ganguli, M., and Kamboh, M.I. (2000). Analysis of genetic polymorphisms in the transforming growth factor-beta1 gene and the risk of Alzheimer's disease. Hum. Genet. *106*, 565–569.
44. R ogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat. Genet. *2*, 168–177.
45. Li, Y., Rowland, C., Catanese, J., Morris, J., Lovestone, S., O'Donovan, M.C., Goate, A., Owen, M., Williams, J., and Grupe, A. (2008). SORL1 variants and risk of late-onset Alzheimer's disease. Neurobiol. Dis. *29*, 293–296.
46. Lee, J.H., Cheng, R., Schupf, N., Manly, J., Lantigua, R., Stern, Y., Rogaeva, E., Wakutani, Y., Farrer, L., St George-Hyslop, P., and Mayeux, R. (2007). The association between genetic variants in SORL1 and Alzheimer disease in an urban, multiethnic, community-based cohort. Arch. Neurol. *64*, 501–506.
47. Bettens, K., Brouwers, N., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., and Sleegers, K. (2008). SORL1 is genetically associated with increased risk for late-onset Alzheimer disease in the Belgian population. Hum. Mutat. *29*, 769–770.
48. Minster, R.L., DeKosky, S.T., and Kamboh, M.I. (2008). No association of SORL1 SNPs with Alzheimer's disease. Neurosci. Lett. *440*, 190–192.
49. Shibata, N., Ohnuma, T., Baba, H., Higashi, S., Nishioka, K., and Arai, H. (2008). Genetic Association between SORL1 Polymorphisms and Alzheimer's Disease in a Japanese Population. Dement. Geriatr. Cogn. Disord. *26*, 161–164.
50. Mayeux, R., and Hyslop, P.S. (2008). Alzheimer's disease: advances in trafficking. Lancet Neurol. *7*, 2–3.
51. Dodson, S.E., Gearing, M., Lippa, C.F., Montine, T.J., Levey, A.I., and Lah, J.J. (2006). LR11/SorLA expression is reduced in sporadic Alzheimer disease but not in familial Alzheimer disease. J. Neuropathol. Exp. Neurol. *65*, 866–872.
52. Zhao, Y., Cui, J.G., and Lukiw, W.J. (2007). Reduction of sortilin-1 in Alzheimer hippocampus and in cytokine-stressed human brain cells. Neuroreport *18*, 1187–1191.
53. Sager, K.L., Wuu, J., Leurgans, S.E., Rees, H.D., Gearing, M., Mufson, E.J., Levey, A.I., and Lah, J.J. (2007). Neuronal LR11/sorLA expression is reduced in mild cognitive impairment. Ann. Neurol. *62*, 640–647.
54. Jacobsen, L., Madsen, P., Moestrup, S.K., Lund, A.H., Tommerup, N., Nykjaer, A., Sottrup-Jensen, L., Gliemann, J., and Petersen, C.M. (1996). Molecular characterization of a novel human hybrid-type receptor that binds the alpha2-macroglobulin receptor-associated protein. J. Biol. Chem. *271*, 31379–31383.
55. Taguchi, K., Yamagata, H.D., Zhong, W., Kamino, K., Akatsu, H., Hata, R., Yamamoto, T., Kosaka, K., Takeda, M., Kondo, I., and Miki, T. (2005). Identification of hippocampus-related candidate genes for Alzheimer's disease. Ann. Neurol. *57*, 585–588.

56. Rodríguez-Rodríguez, E., Infante, J., Llorca, J., Mateo, I., Sánchez-Quintana, C., García-Gorostiaga, I., Sánchez-Juan, P., Berciano, J., and Combarros, O. (2007). Age-dependent association of KIBRA genetic variation and Alzheimer's disease risk. Neurobiol. Aging *30*, 322–324.

57. Papassotiropoulos, A., Stephan, D.A., Huentelman, M.J., Hoerndli, F.J., Craig, D.W., Pearson, J.V., Huynh, K.D., Brunner, F., Corneveaux, J., Osborne, D., et al. (2006). Common Kibra alleles are associated with human memory performance. Science *314*, 475–478.