ARTICLE OPEN ACCESS

Rare genetic variation implicated in non-Hispanic white families with Alzheimer disease

Gary W. Beecham, PhD, Badri Vardarajan, PhD, Elizabeth Blue, PhD, William Bush, PhD, James Jaworski, MPH, Sandra Barral, PhD, Anita DeStefano, PhD, Kara Hamilton-Nelson, MPH, Brian Kunkle, PhD, Eden R. Martin, PhD, Adam Naj, PhD, Farid Rajabli, PhD, Christiane Reitz, MD, PhD, Timothy Thornton, PhD, Cornelia van Duijn, PhD, Allison Goate, PhD, Sudha Seshadri, MD, Lindsay A. Farrer, PhD, Eric Boerwinkle, PhD, Gerard Schellenberg, PhD, Jonathan L. Haines, PhD, Ellen Wijsman, PhD, Richard Mayeux, MD, and Margaret A. Pericak-Vance, PhD, for The Alzheimer's Disease Sequencing Project

Correspondence
Dr. Beecham
gbeecham@med.miami.edu

Neurol Genet 2018;4:e286. doi:10.1212/NXG.0000000000000286

Abstract

Objective

To identify genetic variation influencing late-onset Alzheimer disease (LOAD), we used a large data set of non-Hispanic white (NHW) extended families multiply-affected by LOAD by performing whole genome sequencing (WGS).

Methods

As part of the Alzheimer Disease Sequencing Project, WGS data were generated for 197 NHW participants from 42 families (affected individuals and unaffected, elderly relatives). A two-pronged approach was taken. First, variants were prioritized using heterogeneity logarithm of the odds (HLOD) and family-specific LOD scores as well as annotations based on function, frequency, and segregation with disease. Second, known Alzheimer disease (AD) candidate genes were assessed for rare variation using a family-based association test.

Results

We identified 41 rare, predicted-damaging variants that segregated with disease in the families that contributed to the HLOD or family-specific LOD regions. These included a variant in nitric oxide synthase 1 adaptor protein that segregates with disease in a family with 7 individuals with AD, as well as variants in *RP11-433J8, ABCA1*, and *FISP2*. Rare-variant association identified 2 LOAD candidate genes associated with disease in these families: *FERMT2* (p-values = 0.001) and *SLC24A4* (p-value = 0.009). These genes still showed association while controlling for common index variants, indicating the rare-variant signal is distinct from common variation that initially identified the genes as candidates.

Conclusions

We identified multiple genes with putative damaging rare variants that segregate with disease in multiplex AD families and showed that rare variation may influence AD risk at AD candidate genes. These results identify novel AD candidate genes and show a role for rare variation in LOAD etiology, even at genes previously identified by common variation.

From the John P. Hussman Institute for Human Genomics (G.W.B., J.J., K.H.-N., B.K., E.R.M., F.R., M.A.P.-V.), University of Miami, Miller School of Medicine; Dr. John T. Macdonald Foundation Department of Human Genetics (G.W.B., E.R.M., M.A.P.-V.), University of Miami, Miller School of Medicine, FL; The Taub Institute for Research on Alzheimer's Disease and the Aging Brain (B.V., S.B., C.R., R.M.), Columbia University; The Gertrude H. Segievsky Center (B.V., S.B., C.R., R.M.), Columbia University, New York Presbyterian Hospital; Division of Medical Genetics (E. Blue, E.W.), Department of Medicine, University of Washington, Seattle; Institute for Computational Biology (W.B., J.L.H.), Case Western Reserve University, Cleveland, OH; Department of Neurology (A.D., S.S., L.A.F.), Boston University School of Medicine; Department of Biostatistics (A.D., S.S., L.A.F.), Boston University School of Medicine, MA; School of Medicine (A.N., G.S.), University of Pennsylvania, Philadelphia; Department of Biostatistics (T.T., E.W.), University of Washington, Seattle; Erasmus Medical University (C.D.), Rotterdam, The Netherlands; Icahn School of Medicine at Mount Sinai (A.G.), New York, NY; Department of Medicine (L.A.F.), Boston University School of Medicine, MA; and University of Texas (E. Boerwinkle), Houston.

Funding information and disclosures are provided at the end of the article. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

Co-contributors are listed at links.lww.com/NXG/A132.

The Article Processing Charge was funded by the authors.

Glossary

ADSP = Alzheimer Disease Sequencing Project; **CH** = Caribbean Hispanic; **HLOD** = heterogeneity logarithm of the odds; **LOAD** = late-onset Alzheimer disease; **LOD** = logarithm of the odds; **LRP1** = LDL-receptor-related protein 1; **MAF** = Minor allele frequency; **NHGRI** = National Human Genome Research Institute; **NHW** = non-Hispanic white; **NOS1AP** = nitric oxide synthase 1 adaptor protein; **QC** = quality control; **SNP** = single nucleotide polymorphism; **WES** = whole exome sequencing; **WGS** = whole genome sequencing.

Late-onset Alzheimer disease (LOAD) is a neurodegenerative disease, characterized by progressive dementia, and pathologic changes include neuronal loss, neurofibrillary tangles, and amyloid-beta deposits.[1] LOAD is highly heritable (60%–80%), but most of this heritability remains unexplained, despite the identification of genetic factors that influence LOAD.[2] These factors include the *APOE* gene, as well as other genes identified through genome-wide association studies (GWAS) and a limited number of studies of rare genetic variation.[3–9] While these factors have replicable association with LOAD, few of the underlying causal variants have been definitively identified.

To identify additional genes influencing LOAD and to better understand known LOAD genes, the Alzheimer Disease Sequencing Project (ADSP) was established.[10] A key component of the ADSP is inclusion of whole genome sequencing (WGS) in large, multiply-affected LOAD families of non-Hispanic white (NHW) and Caribbean Hispanic (CH) ancestry. This family-based design enriches the study for risk variation, making it ideal to identify novel risk variants.[11] Family structure facilitates the prioritization of risk variation through linkage and segregation-based approaches. In this study, we report on analyses of the NHW families. Two primary approaches were taken: examination of linkage regions segregating with disease in these families to identify novel genes and a gene-based association analysis to rare variation at known Alzheimer disease (AD) candidate genes. Results indicate that rare variants play a role in LOAD multiplex families, both at novel genes identified through linkage and through rare variation at AD candidate genes.

Methods

The Alzheimer Disease Sequencing Project

Families were assembled as part of the ADSP. The ADSP is a collaboration of the LOAD genetics research community, the National Institutes on Aging, and the National Human Genome Research Institute (NHGRI). The full design is described elsewhere.[11] The ADSP includes contributors from the Alzheimer Disease Genetics Consortium, the neurology working group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), as well as 3 NHGRI sequencing centers at Baylor University, the Broad Institute, and Washington University. Data are available through dbGaP (phs000572).

Family selection and design

Approximately 1,400 multiplex LOAD families were reviewed for inclusion. Families were derived from the National Institute on Aging Late Onset of Alzheimer's Disease family study, the National Cell Repository for Alzheimer's Disease, and families contributed by investigators from Columbia University, University of Miami, University of Washington, University of Pennsylvania, Case Western Reserve University, and Erasmus University. Families analyzed here were of NHW (CH descent families analyzed elsewhere)[12] and were required to have multiple members with LOAD, available genomic DNA, and available *APOE* genotypes. We excluded families known to carry mendelian AD mutations or were pathologically confirmed non-Alzheimer dementia.

Families meeting initial criteria were prioritized and chosen based on the number of affected individuals, number of generations affected, age at onset of clinical symptoms, and presence of APOE ε4 risk alleles (table 1). Details of criteria and selection process are described elsewhere.[10] No ε4/ε4 individuals were included. Cognitively intact participants were selected if available and informative for phasing.

All cases met National Institute of Neurological Diseases-Alzheimer's NINCDS-ADRDA criteria for possible, probable or definite AD.[13] Unaffected individuals were free of clinical AD at the most recent cognitive assessment. In total, 42 NHW families were included. These families included 208 affected individuals and 185 unaffected individuals with array data available, of which 164 affected individuals and 33 unaffected individuals were included in the sequencing experiment.

Standard protocol approvals, registrations, and patient consents

All individuals (or caregivers) provided written informed consent for genomic studies, including broad data sharing, and were assessed with the approval of the relevant institutional review boards.

NGS sequencing

WGS was performed at the NHGRI sequencing centers at the Broad Institute (Boston, MA), Baylor College of Medicine (Houston, TX), and Washington University (St. Louis, MS). Samples were sequenced using Illumina instruments to a minimum average 30× depth. Details of the sequencing experiments are described elsewhere.[14]

NGS calling and quality control

Raw NGS data were aligned to hg19 using BWA.[15] Genotype calling was performed using Atlas (v2).[16] Extensive variant-level quality control (QC) was performed (appendix e-1,

Table 1 Priority variants from consensus linkage regions

Chrm	Position	RS ID	Family	Gene	Alleles	Consequence	CADD	MAF (1kGP)
1	162,167,769		LD0254F	NOS1AP	C/T	Intron variant	1.2	NA
1	162,207,390		LD0254F	NOS1AP	A/T	Intron variant	0.2	NA
1	162,223,640		LD0254F	NOS1AP	A/G	Intron variant	13.6	NA
1	162,479,200		UM0464F	UHMK1	T/G	Intron variant	0.4	0
1	162,564,187		LD1223F	UAP1	A/G	Intron variant	8.7	0.008
1	162,564,187		LD1223F	UAP1	A/G	Intron variant	8.7	0.008
1	162,700,025		UM0464F	DDR2	A/C	Intron variant	0.5	0.009
1	162,735,057		UM0464F	DDR2	G/A	Intron variant	0.7	0.009
1	162,739,064		UM0464F	DDR2	G/A	Intron variant	5.4	0.009
1	162,742,651		LD0254F	DDR2	G/A	Intron variant	0.1	0.001
1	162,751,967		LD0254F	DDR2	T/A	3′ UTR variant	6.6	0.009
1	162,757,273		LD1223F	DDR2	T/C	Upstream gene variant	2.2	0.009
1	162,928,238		UM0464F		G/A	Intergenic variant	0.9	0.002
1	163,032,461		LD0254F		T/A	Intergenic variant	1.9	0.002
1	163,202,875		UM0464F	RGS5	G/A	Intron variant	8.3	0.002
1	163,578,427		UM0464F		C/A	Intergenic variant	4.7	0.003
1	163,749,571		LD0949F		A/G	Intergenic variant	7.2	0.002
1	163,841,625		UM0464F		C/T	Intergenic variant	7.1	0.010
1	164,034,469	rs187504850	UM0464F		A/G	Intergenic variant	19.6	NA
1	164,448,463		UM0464F		G/A	Intergenic variant	3.2	0.003
1	164,622,647		LD0949F	PBX1	G/A	Intron variant	1.2	0.006
1	164,887,661		UM0464F		C/T	Downstream gene variant	8.2	0.005
1	165,253,949		LD0254F	LMX1A	C/T	Intron variant	14.7	0.002
1	165,342,593		UM0464F		G/A	Intergenic variant	5.9	0.010
1	165,532,785		LD1223F	LRRC52	G/A	Synonymous variant	18.3	0.004
14	95,913,507	rs191535004	LD0949F	SYNE3	G/A	Intron variant	0.7	0.008
14	95,923,822		UM0464F	SYNE3	C/T	Intron variant	3.5	0.008
14	96,449,175		UM0464F		C/A	Intergenic variant	0.8	NA
14	96,568,984		UM0464F		C/A	Regulatory region variant	1.5	NA
14	96,923,339		LD0254F	AK7	C/T	Noncoding transcript exon variant	0.1	0.008
14	97,029,358		UM0464F	PAPOLA	C/G	Intron variant	10.4	NA
14	97,228,875		LD0949F	RP11-433J8.2	A/G	Intron variant	12.3	0.003

Abbreviations: CADD = Combined Annotation Dependent Depletion score; Chrm = chromosome; MAF (1kGP) = Minor Allele Frequency among the European samples in the 1,000 Genomes Project data; UTR = untranslated region.

links.lww.com/NXG/A117). Principal components analysis was used to assess population substructure, using the EIGENSTRAT.[17] Array data were compared with WGS data to assess and confirm the pedigree structure for all individuals. Additional details of QC are reported elsewhere.[14]

Linkage analyses

MERLIN v1.1.2 software[18] was used to perform parametric and nonparametric multipoint linkage analyses on the array data available for the entire family. Nonparametric analyses are described in detail elsewhere.[19] For parametric multipoint

analyses, we first pruned markers to minimize linkage disequilibrium ($r^2 < 0.01$) using PLINK v1.07 software.[20] Using this pruned grid of markers, parametric multipoint linkage analyses were performed using a rare disease allele frequency (0.0001) and a dominant model with incomplete penetrance (noncarrier 0.01 and carrier 0.90). Consensus linkage regions (i.e., consistent across multiple families) were defined as peak HLOD ≥ 3.3 per Lander and Kruglyak.[21] Any family with peak family-specific LOD >0.58 in the consensus region was considered a contributor to the consensus signal. Family-specific linkage regions were defined as regions with a parametric family-specific LOD >2.

Annotations

Variants were annotated for location, gene (if applicable), putative function (missense, nonsense, splice site, etc.), combined annotation dependent depletion (CADD) score (a quantitative summary of putative function and conservation),[22] contextual analysis of transcription factor occupancy (CATO) scores[23] for intergenic variation, and allele frequency in the NHW families and in the 1,000 Genomes Project European-ancestry populations.[24] As a QC measure, we used BLAST to interrogate the genome for similar sequence as the high-priority variants, to ensure uniqueness of the relevant sequence.[25]

Variant filtering and prioritization

Variants were filtered based on complete segregation among affected individuals (and absent from unaffected individuals) and rarity (minor allele frequency [MAF] < 0.05 in our data set, <0.01 in 1,000 Genomes Project data). Additional prioritization was applied to variants with high CADD scores, were observed in multiple families, had CATO predictions, had multiple filtered variants in the same gene, or showed nominal association in the ADSP case-control analyses.

Validation genotyping

High-priority genotypes were validated using Sanger sequencing of whole genome sequenced family members to confirm carrier/noncarrier status. Sequencing was performed using standard protocol on genomic DNA (~50 ng). Details of validation typing are in appendix e-1 (links.lww.com/NXG/A117).

Gene-based association tests

Gene-based association tests were performed using the FSKAT v1 software.[26] A cutoff of MAF <0.02 was used among the non-Finnish European ancestry populations in the 1,000 Genomes Project data (1kGP)[24]. Variants were analyzed in 2 sets: (1) damaging rare variants (loss-of-function variants, nonsense, stop-loss, etc and those predicted to be damaging) and (2) damaging variants plus all nonsynonymous variants. Genes with only a single variant were excluded. FSKAT was applied to the remaining genes using 2 models: one adjusted for age, sex, and the top 10 principal components and the other unadjusted.

Candidate gene list

Candidate genes (appendix e-1, links.lww.com/NXG/A117) were selected from replicable population and family AD genetics studies, mostly from GWAS of LOAD or known early-onset AD genes.[4–9]

Data availability

Anonymized data are available by request from qualified investigators through dbGaP (phs000572.v1.p1) and through the National Institute on Aging Genetics of Alzheimer Disease Data Storage Site (www.niagads.org).

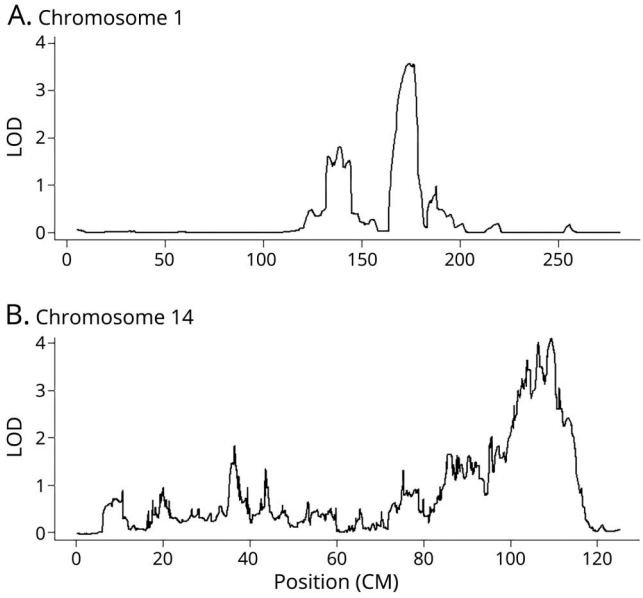
Results

Consensus linkage regions

Linkage scans identified 2 primary "consensus" linkage regions (peak HLOD ≥ 3.3)[21]: a parametric multipoint peak on chromosome 1q23 (peak HLOD = 3.58; 162.2–165.8 Mb; figure 1A) and a nonparametric multipoint peak on chromosome 14q32 (LOD = 4.18; 98.9–99.6 Mb; figure 1B). The 1q23 region was supported (LOD > 0.58) by 8 families (LD0241F, LD0254F, LD0856F, LD1223F, LD1260F, UM0196F, UM0463F, and UM0464F), while the 14q32 region was supported by 4 families (LD0223F, LD0949F, LD1223F, and UP0004F).[19]

In total, there were 86 rare (MAF <0.01 1kGP) variants that segregated with disease in sequenced affected individuals in at least 1 of the 8 families that supported the chromosome 1 peak. Of the 86 variants, 43 were genic (50%) and 43 were intergenic (50%). This initial set of 86 segregating variants was further refined by requiring variants to also be absent from unaffected individuals in the family (if available), have moderate-to-high

Figure Summary of consensus linkage regions on chromosomes 1 and 14



(A) Peak LOD region on chromosome 1. X-axis denotes positions in centimorgans. Y-axis denotes the LOD score at the corresponding position. (B) Peak LOD region on chromosome 14. X-axis denotes positions in centimorgans. Y-axis denotes the LOD score at the corresponding position.

CADD (CADD > 10), CATO predictions, or be seen in multiple families. Of the 86 variants, 24 matched these criteria and were considered "high priority." At the 14q32 locus, we identified 23 rare variants segregating with disease among affected individuals in at least 1 of the 4 families supported the linkage signal. Of this set, there were 7 variants absent from the unaffected individuals, had high CADD predictions, or were seen in multiple families. In total, 31 variants in the consensus linkage regions were prioritized; of these, 29 variants were validated using Sanger sequencing, 1 was a false positive, and 1 could not have a reliable assay developed.

A number of interesting results come out of this set of 29 confirmed variants (table 1). In the 1q23 region, a variant (chr1: 162,223,640 A/G) in nitric oxide synthase 1 adaptor protein (*NOS1AP*) segregates with disease in family UM0464F with 7 individuals with AD (family-specific LOD = 2.62; figure e-1, links.lww.com/NXG/A119); while the variant is intronic, it has a moderate CADD score (13.6) and is completely absent in the 1kGP reference data set. Other variants in *NOS1AP* (chr1: 162,167,769 C/T; chr1:162,207,390 A/T) segregate with disease in family LD0254F (figure e-2, links.lww.com/NXG/A120). In the 14q32 region, an intronic variant in ncRNA RP11-433J8 segregates with disease in family LD1223F (6 AD family members; family-specific LOD = 1.45); the variant was also present in a second family (LD0307F) although it was not present in all AD individuals. The variant is rare in 1kGP (MAF = 0.003) and has a moderate CADD score (CADD = 12.2).

Family-specific linkage regions

In addition to the consensus linkage regions, there were 10 family-specific regions identified using parametric multipoint linkage (table 2). These regions showed family-specific LOD scores >2. Among the 10 regions, there were 647 variants that were rare (MAF <0.01 1kGP) and segregated among the affected individuals in family with the LOD score >2. The 647 variants were further prioritized based on absence in unaffected family members with WGS, high CADD predictions, as well as presence in multiple families, identifying 13 additional variants as high priority (table 3). Twelve of these 13 variants were validated using an orthogonal technology (the last could not have a reliable assay developed).

Among the family-specific regions, a missense variant (rs137854495) in the chromosome 9 *ABCA1* (ATP binding cassette subfamily A member 1) gene segregated with disease in a family with 4 individuals affected with AD (family-specific LOD = 2.04). The variant was absent in the 1kGP reference data set and had a high CADD score (CADD = 34). Two missense variants in *FSIP2* (fibrous sheath interacting protein 2) segregates with disease in a single family with 7 AD family members (family-specific LOD = 2.07). Both variants were rare in 1kGP (MAF < 0.001) and had high CADD scores (CADD = 25.2 and 22.6). This analysis also identified a missense variant with high CADD (CADD = 32) in TTC3, from family UM0146F. This variant was previously identified through whole exome sequencing (WES) in the same family and is described elsewhere.[27]

Candidate genes

FSKAT, a family-based kernel test for association of sets of variants, was used to perform gene-based association in the families[27] (Table 4). A list of 31 candidate genes (identified from GWAS and studies of familial AD) was tested for association with LOAD. Two genes showed association with LOAD in the unadjusted analysis that included nonsynonymous variants: *FERMT2* (p-value = 0.001) and *SLC24A4* (p-value=0.009). The association in *FERMT2* survives a Bonferroni correction for 31 genes tested (p-value = 0.05/31 = 0.0016). Both genes still showed association after adjusting for age, sex, and the top 10

Table 2 Family-specific linkage regions

Family ID	Chrm	cM (start)	cM (end)	BP (start)	BP (end)	Peak LOD
UM0458F	12	0.18	41.96	0.38	20.56	2.95
UM0458F	14	6.03	15.44	21.64	24.64	2.7
UM0146F	21	41.07	49	36.55	40.37	2.63
UM0146F	19	76.43	85.05	48.55	51.38	2.23
UP0005F	1	131.84	144.67	103.72	115.74	2.07
UP0005F	2	188.73	210.97	183.63	213.07	2.07
UP0005F	16	78.58	91.44	58.51	73.81	2.07
UM0463F	5	138.9	146.8	131.38	141.03	2.06
UP0001F	9	106.1	116.2	104.11	112.23	2.04
UP0001F	9	95.12	105.57	92.41	103.64	2.04

Abbreviations: BP (start) and BP (end) = position of the region in megabases; Chrm = Chromosome; cM (start) and cM (end) = position of the region in centimorgans; LOD = logarithm of the odds.

Table 3 Priority variants from family-specific linkage regions

Chrm	Position	Gene	Alleles	Family	Consequence	CADD	MAF (1kGP)
2	186,611,520	FSIP2	C/T	UP0005F	Missense variant	25.2	0.001
2	186,611,521	FSIP2	G/T	UP0005F	Missense variant	22.6	0.001
2	199,347,563	PLCL1	A/G	UP0005F	Intron variant	17.7	0.010
2	208,614,446	CCNYL1	C/G	UP0005F	Intron variant	20.1	0.001
9	100,819,143	NANS	C/T	UP0001F	Missense variant	22.5	0.001
9	107,584,795	ABCA1	G/A	UP0001F	Missense variant	34.0	NA
12	16,342,622	SLC15A5	G/A	UM0458F	Missense variant	24.0	0.008
12	17,149,860		T/A	UM0458F	Downstream gene variant	16.4	0.010
12	18,891,317	CAPZA3	C/T	UM0458F	Missense variant	21.4	0.005
16	61,999,830	CDH8	A/C	UP0005F	Intron variant	15.5	0.007
16	70,546,287	COG4	C/T	UP0005F	Missense variant	23.9	NA
16	73,127,644	HCCAT5	A/G	UP0005F	Noncoding transcript exon variant	16.3	0.002
21	38,534,308	TTC3	C/G	UM0146F	Missense variant	32.0	NA

Abbreviations: CADD = Combined Annotation Dependent Depletion score; Chrm = chromosome; MAF (1kGP) = minor allele frequency among European samples in the 1,000 Genomes Project data.

principal components (*FERMT2* p-value = 0.002; *SLC24A4* p-value = 0.023). The *PICALM* gene also showed nominal association after adjusting for age, sex, and principal components (p-value = 0.032; unadjusted p-value = 0.111). *SLC24A4* also showed nominal association in the damaging variant–only analysis (p-value = 0.026). Because these 3 genes were all initially identified using GWAS, we also performed the association analysis with the genotyped GWAS index single nucleotide polymorphisms (SNPs) as covariates in the FSKAT model (rs17125944 for *FERMT2*, rs10498633 for *SLC24A4*, and rs10792832 for *PICALM*). Each gene still showed evidence of association after including index SNP genotypes as covariates (*FERMT2* p-value = 0.002, *SLC24A4* p-value = 0.015, *PICALM* p-value = 0.021).

Because the gene of interest for a particular associated locus may not be the gene physically closest to the index SNP, as a secondary analysis, we expanded the list of 31 candidate genes to include genes near the GWAS index SNPs (±1,000,000 bp). In this analysis, an additional 586 genes were tested using FSKAT. Near the *FERMT2* locus (within 100 kb downstream), the genes *STYX*, *PSMC6*, and *GNPNAT1* all showed association in the analysis including nonsynonymous variants (p-values = 0.0011, 0.0012, and 0.0016, respectively). *STYX*, in particular, also showed nominal association in a large case-control WES study conducted by the ADSP (p-value = 0.00119).[28] As with *FERMT2*, the p-values did not appreciably change when adjusting for age, sex, and principal components (p-values = 0.0013, 0.0016, and 0.0024, respectively). Additional genes showed association in the nonsynonymous analysis include *MGC45922* (p-value = 0.0030; near *CD33*), *TAP2* (p-value = 0.0043; near *HLA-DRB1/DRB5*; p-value = 0.0047 in the ADSP WES analysis), and *FAM210B* (p-value = 0.0084; near *CASS4*), when adjusting for age, sex, and principal components. In the analysis of damaging variants, the *CPSF2* gene was associated in the adjusted analysis (p-value = 0.000498), which would survive a Bonferroni multiple testing correction for 586 genes. This gene is located near the *SLC24A4* gene and was also nominally associated in the ADSP WES analyses (p-value = 0.034). The *FIS1* gene also showed evidence of association in the unadjusted analysis (p-value = 0.00748, near *ZCWPW1*; unadjusted analysis p-value = 0.0147) and was also nominally associated in the ADSP WES analyses (p-value = 0.034).

Discussion

To identify rare variation influencing LOAD, we performed analyses of WGS data in NHW families multiply affected for LOAD. A two-pronged approach was taken: examination of linkage regions identified through analysis of genome-wide genotyping array data and a gene-based association analysis of rare coding variants, focusing on AD candidate loci identified in GWAS. These results indicate a potential role for rare variants in LOAD etiology. Numerous rare, predicted damaging rare variants were identified that segregate with disease in multiplex LOAD families and were validated with independent technologies. Additionally, rare variation in LOAD candidate genes was associated with AD in these multiplex families. This association persisted even when the common variant index SNPs were included in the models, indicating the rare variant association is likely distinct from the common variants that initially implicated the genes.

Table 4 Gene-based association test results at known AD candidate genes

	Putative damaging + nonsynonomous		Putative damaging	
Gene	***p*-Value (unadj)**	***p*-Value (adj)**	***p*-Value (unadj)**	***p*-Value (adj)**
ABCA7	0.534	0.414	0.782	0.657
AKAP9	0.226	0.167	0.115	0.145
APOE	0.334	0.228	—	—
APP	0.802	0.651	—	—
BIN1	0.422	0.339	—	—
CASS4	0.159	0.164	0.527	0.732
CD2AP	0.939	0.892	0.110	0.216
CD33	0.321	0.250	0.111	0.122
CELF1	0.353	0.223	—	—
CLU	0.608	0.475	0.465	0.631
CR1	0.349	0.182	0.403	0.629
EPHA1	0.481	0.459	0.842	0.896
FERMT2	**0.001**	**0.002**	—	—
GRN	0.310	0.453	—	—
HLA-DRB1	0.262	0.166	—	—
INPP5D	0.140	0.200	—	—
MAPT	0.673	0.668	0.596	0.677
MEF2C	0.266	0.323	—	—
MS4A6A	0.264	0.287	0.751	0.555
NME8	0.228	0.127	—	—
PICALM	0.111	**0.032**	—	—
PLD3	0.169	0.146	0.348	0.305
PSEN1	0.755	0.418	0.154	0.089
PSEN2	0.725	0.444	0.080	0.173
PTK2B	0.653	0.489	0.622	0.447
RIN3	0.419	0.341	0.935	0.958
SLC24A4	**0.009**	**0.023**	0.026	0.076
SORL1	0.642	0.438	0.263	0.172
TREM2	0.678	0.575	0.394	0.358
TREML2	0.381	0.300	0.208	0.143
ZCWPW1	0.560	0.478	—	—

Bold indicates *p*-values < 0.05.

Rare variation in *NOS1AP* was identified. *NOS1AP* lies under one of the HLOD linkage peaks, is expressed in the brain,[29] and is known to interact with the LDL-receptor-related protein 1 (LRP1). LRP1 is an APOE receptor that helps bring APOE into neurons[30,31] and *APP*. In addition, LRP has been associated with AD in the ADSP WES experiment ($p = 0.00018$).[28] *NOS1AP* also interacts with nNOS,[32] encoded by *NOS1*, which has been linked to AD[33] as well as other neurologic diseases.

A missense variant (rs137854495) in ATP binding cassette subfamily A member 1 (*ABCA1*) was found to segregate with disease in one family under a family-specific linkage peak on

chromosome 9. The variant was rare, with a very high CADD score (CADD = 34). *ABCA1* is expressed in brain (though not exclusively; *ABCA7* is expressed in many tissues) and is involved in lipid removal pathways.[29] Variants in *ABCA1* have been associated with HDL deficiency, familial hypercholesterolemia, and *APOA* deficiency.[34–37] The rs137854495 variant, in particular, has been noted in a family with Tangier disease as part of a compound heterozygote.[38] Dyslipidemias and lipid pathways have long been linked to LOAD,[39] starting with the *APOE* gene,[3] and more recently *CLU*, *ABCA7*, etc,[4,22] although exact mechanisms remain unclear. Tangier disease, in particular, has also been proposed as having links to AD, primarily through amyloid-β pathways, although evidence supporting this is mixed.[40–44] The ADSP WES project identified nominal association with 2 additional apolipoproteins (*APOA2*, $p = 0.000636$; *APOA5*, $p = 0.0413$).[28]

Gene-based association tests implicated fermitin family member 2 (*FERMT2*) and surrounding genes *STYX*, *PSMC6*, and *GNPNAT*, all with similar levels of significance ($p = 0.0010–0.0016$). *FERMT2* is involved in cell adhesion, is expressed in brain, and is near an SNP with strong association to AD.[4] *STYX* is likely involved in phosphatase activity and has been associated with diabetes mellitus type 1.[45] *PSMC6* is likely involved in hydrolase activity; *GNPNAT* is involved in sugar metabolism. *SLC24A4* has been associated with AD through a genome-wide meta-analysis,[4] and brain methylation in *SLC24A4* region has been associated with AD risk.[46] Although *FERMT2* and *SLC24A4* were initially identified using common variant approaches, the association observed at these 2 genes was not greatly affected by including the GWAS index SNPs as covariates in the model. If common variants were solely responsible for the association, then we would expect to fail to reject the null hypothesis at the rare variants. This implies the rare variation associated with disease in these families is distinct from the common variant index SNPs initially used to identify the genes.

There are limitations to this study. The sample size was modest relative to GWAS approaches. This of course limits power, particularly for the association-based approaches. However, the use of familial data and linkage and segregation-based approaches mitigates some of these power concerns. Increasing sample sizes and number of multiplex families is an ongoing effort for future studies. Additional limitations include the use of in silico predictions of function (e.g., CADD). While useful as a first pass, these predictions should be seen as a putative,[47] and function will need to be validated by functional genetic approaches.

These results imply a role for rare variation in familial LOAD. The linkage analysis identified 41 high-priority variants, including variants in *NOS1AP* and *ABCA1*, both with plausible roles in AD and AD-related pathways. The analysis of LOAD candidate genes identified several genes with rare variation associated with AD. The tests were still significant while controlling for the common index SNPs, implying a role for rare variation even at GWAS-identified loci. Future directions include a thorough analysis on noncoding variation, particularly the role of enhancers and other regulatory elements in the etiology of AD.

Author contributions

All authors contributed to the work presented in this article. Critical revision: Primary manuscript was prepared by G.W. Beecham, with significant contributions from B. Vardarajan, E. Blue, E. Wijsman, and M.A. Pericak-Vance. All authors participated in the revision and editing of the manuscript. Concept and design: There were significant contributions to concept and design from G.W. Beecham, B. Vardarajan, E. Blue, W. Bush, A. DeStefano, E.R. Martin, A. Naj, C. Reitz, C. van Duijn, A. Goate, S. Seshadri, L.A. Farrer, E. Boerwinkle, G. Schellenberg, J.L. Haines, E. Wijsman, R. Mayeux, and M.A. Pericak-Vance. Analysis and interpretation: Review of family data was performed by M. Pericak-Vance, R. Mayeux, E. Boerwinkle, S. Seshadri, and C. van Duijn. Primary statistical analyses were performed by G.W. Beecham, J. Jaworski, E.R. Martin, and K. Hamilton-Nelson, with additions from B. Vardarajan, W. Bush, and E. Blue. All authors participated in the interpretation and discussion of results. Acquisition of data: Sample data were contributed by C. van Duijn, A. DeStefano, L.A. Farrer, A. Goate, J.L. Haines, M.A. Pericak-Vance, E. Boerwinkle, R. Mayeux, S. Seshadri, and G. Schellenberg. Statistical analyses: Statistical analyses were primarily conducted by G.W. Beecham; additional analyses conducted by J.C.B., A.C.N., E.R. Martin, S.H.C., A. DeStefano, and S. Seshadri (affiliations noted above, all academic). Study supervision and coordination: Primary study supervision and coordination was by R. Mayeux, M.A. Pericak-Vance, and E. Wijsman. Funding: Primary funding was by G. Schellenberg, R. Mayeux, E. Boerwinkle, M.A. Pericak-Vance, J.L. Haines, S. Seshadri, A. Goate, L.A. Farrer, and E. Wijsman. A detailed list of funding is noted in the acknowledgements.

Acknowledgment

The Alzheimer's Disease Sequencing Project (ADSP) comprises 2 Alzheimer disease (AD) genetics consortia and 3 National Human Genome Research Institute–funded Large Scale Sequencing and Analysis Centers (LSAC). The 2 AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by the National Institute on Aging (NIA; U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other NIH institutes and other foreign governmental and nongovernmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Farrer, Haines, Mayeux, Pericak-Vance, and Schellenberg); U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate; and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance, and U01 AG052409 to Drs. Seshadri and Fornage. Data generation and harmonization in the Follow-up Phases

is supported by U54AG052427 to Drs. Schellenberg and Wang. The ADGC cohorts include: Adult Changes in Thought (ACT supported by NIA grant U01AG006781 to Drs. Larson and Crane), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP supported by NIA grant RF1AG054023 to Dr. Mayeux) and the Washington University Sequencing Project (WUSP), the Columbia University Hispanic-Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA supported by NIA grant RF1AG015473 to Dr. Mayeux), the University of Toronto (UT), and Genetic Differences (GD). Analysis of ADGC cohorts was supported by NIA grants R01AG048927 and RF1AG057519 to Dr. Farrer. Efforts of ADGC investigators were also supported by grants from the NIA (R03AG054936) and National Library of Medicine (R01LM012535). The CHARGE cohorts are supported in part by NHLBI infrastructure grant HL105756 (Psaty) and RC2HL102419 (Boerwinkle), and the neurology working group is supported by the NIA R01 grant AG033193. The CHARGE cohorts participating in the ADSP include the following: Austrian Stroke Prevention Study (ASPS), ASPS-Family study, and the Prospective Dementia Registry-Austria (ASPS/PRODEM-Aus), the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). ASPS is funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180 and the Medical University of Graz. The ASPS-Fam is funded by the Austrian Science Fund (FWF) project I904, the EU Joint Programme—Neurodegenerative Disease Research (JPND) in frame of the BRIDGET project (Austria, Ministry of Science) and the Medical University of Graz and the Steiermärkische Krankenanstalten Gesellschaft. PRODEM-Austria is supported by the Austrian Research Promotion agency (FFG) (Project No. 827462) and by the Austrian National Bank (Anniversary Fund, project 15435). ARIC research is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Neurocognitive data in ARIC is collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the NIH (NHLBI, NINDS, NIA, and NIDCD) and with previous brain MRI examinations funded by R01-HL70825 from the NHLBI. CHS research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the NHLBI with additional contribution from the NINDS. Additional support was provided by R01AG023629, R01AG15928, and R01AG20098 from the NIA. FHS research is supported by NHLBI contracts N01-HC-25195 and HHSN2682015000001I. This study was also supported by additional grants from the NIA (R01s AG054076, AG049607, and AG033040) and NINDS (R01 NS017950). The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by a joint grant from the Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the municipality of Rotterdam. Genetic data sets are also supported by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO), Netherlands Consortium for Healthy Aging (NCHA), project 050-060-810. All studies are grateful to their participants, faculty, and staff. The content of these manuscripts is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the U.S. Department of Health and Human Services. The ADES-FR study was funded by grants from the Clinical Research Hospital Program from the French Ministry of Health (GMAJ, PHRC, 2008/067), the CNR-MAJ, the JPND PERADES, the GENMED labex (LABEX GENMED ANR-10-LABX-0013), and the FP7 AgedBrainSysBio. Whole exome sequencing in the 3C-Dijon study was funded by the Fondation Leducq. This work was supported by the France Génomique National infrastructure, funded as part of the Investissements d'Avenir program managed by the Agence Nationale pour la Recherche (ANR-10-INBS-09), the Centre National de Recherche en Génomique Humaine, the National Foundation forAlzheimer's disease and related disorders, the Institut Pasteur de Lille, Inserm, the Lille Métropole Communauté Urbaine council, and the French government's LABEX (laboratory of excellence

program investment for the future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary approach to Alzheimer's disease). The 3C Study supports are listed on the Study Website (three-city-study.com). The FinnAD Study at the University of Tampere was supported by The Academy of Finland: grants 286284 (T.L.), Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association. The three LSACs are the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), and the Washington University Genome Institute (U54HG003079).

Study funding

Supported by the NIH, primarily the National Institute on Aging (NIA), the National Heart, Lung, and Blood Institute, and the National Human Genome Research Institute. Primary support includes the Alzheimer's Disease Genetics Consortium funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), and the Washington University Genome Institute (U54HG003079). Additional funding of contributing sites is noted below in the acknowledgements.

Disclosure

G. Beecham has received research support from the NIH and the Department of Defense. B. Vardarajan reports no disclosures. E. Blue has received research support from the NIH and the Cystic Fibrosis Foundation and has been a grant reviewer for the NIH. W. Bush has served on the editorial board for *BMC Biodata Mining* and *PLoS One* and has received research support from the USDA, the Foundation for Food and Agriculture, NIDDK, and the National Institute on Aging (NIA). J. Jaworski, S. Barral, and A. DeStefano report no disclosures. K. Hamilton-Nelson has received research support from the NIH. B. Kunkle reports no disclosures. E. Martin serves on the editorial board of *Frontiers in Statistical Genetics and Methodology*; and holds a patent for Test for Linkage and Association in General Pedigrees: The Pedigree Disequilibrium Test. A. Naj, F. Rajabli, and C. Reitz report no disclosures. T. Thornton has received research support from the NIH. C. van Duijn reports no disclosures. A. Goate has served on the scientific advisory boards of Denali Therapeutics, Pfizer, and DZNE; has received travel funding/speaker honoraria from the Rainwater Foundation, the Indiana University ADRC advisory board, and Wellcome Trust; serves on the editorial board of *eLife*; holds patents for PSEN mutations in AD, Tau mutations in FTD, and TDP43 mutations in ALS\FTD; has been a consultant for Cognition Therapeutics, AbbVie, and Biogen; has received research support from F Prime, NIA, Rainwater Charitable Foundation, and JPB Foundation; and receives royalty payments from Taconic Industries for tau mutation patent, and from Athena Diagnostics for TDP43 mutation testing. S. Seshadri serves on the editorial boards of the *Journal of Alzheimer's Disease*, *Stroke*, and *Neurology* and has received research support from NIA and NINDS. L. Farrer serves on the editorial boards of the *American Journal of Alzheimer's Disease & Other Dementias*, *Clinical Genetics*, and the *Journal of Clinical Medicine*; holds a patent (pending) for Use of PLXNA4 as a drug target and biomarker for Alzheimer disease; has been a consultant for Novartis Pharmaceuticals, Gerson Lerman, and Guidepoint Global; has received research support from the NIH, the Fidelity Foundation, and the Thome Memorial Foundation; and was a consultant for legal proceedings involving Finnegan & Associates, LLP. E. Boerwinkle has received travel funding and speaker honoraria from the Harvard School of Public Health and the Metabolomics Forum in Cambridge, United Kingdom; serves on the editorial board of *Annals of Epidemiology*; is a scientific officer at Codified Genomics, LLC; and has received research support from the NIH. G. Schellenberg has served on scientific advisory boards for the Alzheimer's Association, the Society of Progressive Supranuclear Palsy, the United Kingdom Parkinson Disease Center, University College London, the Alzheimer's Disease Sequence Project (co-chair), Structural Variant Work Group, the Alzheimer's Disease Sequence Project, Mayo Clinic Rochester Udall Center, University of Miami Udall Center, Discovery Assessment Panel, and the Oxford Parkinson's Disease Centre; has received travel funding/speaker honoraria from Alzheimer's Disease Center, CurePSP, the University of California San Diego, Keystone Symposia, Southern California Alzheimer's Disease Research Conference, University of California Institute for Memory Impairment and Neurological Disorders, NIH, Novartis, McKnight Brain Institute, University of Florida, Keep Memory Alive Event Center, Cleveland Clinic, Accelerated Medicines Program, PSP/Lewy Body Disease Think-Tank, Alzheimer's Disease Center, American Association of Neuropathologists, Fusion Conferences, Center for Public Health Genomics, Columbia University, PSP Genetics Consortium, Tetra Institute, Rockefeller University, Blechman Foundation, Niigata University, Xuan Wu Hospital (Capital Medical University), Genetics of Dementia Summit (United Kingdom), Icahn School of Medicine Mount Sinai, and Biogen; has served on the editorial boards of the *Journal of Neural Transmission, American Journal of Alzheimer's Disease and Other Dementias*, *Alzheimer's Research*, *Neurodegenerative Diseases*, *Current Alzheimer Research*, and *Pathology and Laboratory Medicine International*; is employed by the University of Pennsylvania; has been a consultant for Biogen; and has received research support from the NIH, CurePSP, and CBD Solutions. J. Haines has served on the editorial boards of *Neurogenetics*, *Current Protocols in Human Genetics*, and *Human Molecular Genetics*; receives publishing royalties from John Wiley &

Sons; and has received research support from the NIH. E. Wijsman has served on the editorial boards of *BMC Proceedings* and *Faculty of 1000* and has received research support from the NIH and the Metropolitan Life Foundation. R. Mayeux has received research support from the NIH. R. Pericak-Vance reports no disclosures. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

APPENDIX 1 Co-investigators

Affiliation	Full Name	Contributions
Baylor College of Medicine	Adam English	Baylor College of Medicine site contributed expertise to the study design, sequencing of samples, bioinformatics analyses, quality control, data management, structural variation working group, as well as input into both case-control and family study working groups
	Divya Kalra	
	Donna Muzny	
	Evette Skinner	
	Harsha Doddapeneni	
	Huyen Dinh	
	Jianhong Hu	
	Jireh Santibanez	
	Joy Jayaseelan	
	Kim Worley	
	Michelle Bellair	
	Richard A. Gibbs	
	Sandra Lee	
	Shannon Dugan-Perez	
	Simon White	
	Viktoriya Korchina	
	Waleed Nasser	
	William Salerno	
	Xiuping Liu	
	Yi Han	
	Yiming Zhu	
	Yue Liu	
	Ziad Khan	
Boston University	Adrienne Cupples	The Boston University site contributed expertise to the study design, bioinformatics analyses, quality control, data management, structural variation working group, case-control working group, family study working group, as well as significant sample contributions
	Alexa Beiser	
	Anita DeStefano*	
	Ching Ti Liu	
	Chloe Sarnowski	
	Claudia Satizabal	
	Dan Lancour	
	Devanshi Patel	

Continued

APPENDIX 1 Co-investigators *(continued)*

Affiliation	Full Name	Contributions
	Fangui Jenny Sun	
	Honghuang Lin	
	Jaeyoon Chung	
	John Farrell	
	Josee Dupuis	
	Kathy Lunetta	
	Lindsay Farrer*	
	Sudha Seshadri*	
	Xiaoling Zhang	
	Yiyi Ma	
	Yuning Chen	
Broad Institute	Eric Banks	The Broad Institute site contributed expertise to the study design, sequencing of samples, bioinformatics analyses, quality control, and data management working group
	Namrata Gupta	
	Seung Hoan Choi	
	Stacey Gabriel	
Case Western Reserve University	Jonathan Haines*	The Case Western Reserve University site contributed expertise to the study design, bioinformatics analyses, quality control, data management, structural variation working group, case-control working group, family study working group, annotation working group, as well as sample contributions
	Mariusz Butkiewicz	
	Sandra Smieszek	
	Will Bush*	
	Yeunjoo Song	
Columbia University	Badri Vardarajan*	The Columbia University site contributed expertise to the study design, bioinformatics analyses, quality control, data management, structural variation working group, case-control working group, family study working group, annotation working group, as well as sample contributions
	Christiane Reitz*	
	Dolly Reyes	
	Giuseppe Tosto	
	Phillip L De Jager	
	Richard Mayeux*	
	Sandra Barral*	
Erasmus Medical University/ Rotterdam	Ashley Vanderspek	The Erasmus Medical University site contributed expertise to the study design, the family study working group, as well as sample contributions
	Cornelia van Duijn*	
	M Afran Ikram	
	Najaf Amin	
	Shahzad Amad	
	Sven van der Lee	

Continued

APPENDIX 1 Co-investigators *(continued)*

Affiliation	Full Name	Contributions
Indiana University	Kelley Faber	The Indiana University site contributed expertise to the study design, data and sample collection and management, participated in the family and case-control working groups
	Tatiana Foroud	
Medical University Graz, Austria	Helena Schmidt	Medical University Grace contributed data and samples to the study
	Reinhold Schmidt	
Mount Sinai School of Medicine	Alan Renton	The Mount Sinai site contributed samples as well as expertise to the design of the study, the family and case-control working groups, and the protective variant working group
	Alison Goate*	
	Edoardo Marcora	
	Manav Kapoor	
National Center for Biotechnology Information	Adam Stine	National Center for Biotechnology Information site contributed expertise to data management
	Michael Feolo	
National Institutes of Aging	Lenore J. Launer	National Institute on Aging site contributed expertise to data and study management
Rush University	David A Bennett	Rush University site contributed samples and data to the study
Stanford University	Li Charlie Xia	The Stanford University site contributed expertise to the structural variation working group
University of Miami	Brian Kunkle*	The University of Miami site contributed expertise to the study design, bioinformatics analyses, quality control, data management, structural variation working group, case-control working group, family study working group, as well as significant sample contributions
	Eden Martin*	
	Farid Rajabli*	
	Gary Beecham*	
	James Jaworski*	
	Kara Hamilton-Nelson*	
	Margaret Pericak-Vance*	
	Michael Schmidt	
University of Mississippi	Thomas H. Mosley	The University of Mississippi site contributed samples and data to the study
University of Pennsylvania	Amanda Kuzma	The University of Pennsylvania site contributed expertise to the study design, bioinformatics analyses, quality control, data management, structural variation working group, case-control working group, and family study working groups
	Han-Jen Lin	
	Liming Qu	
	Li-San Wang	
	Micah Childress	
	Otto Valladares	
	Prabhakaran Gangadharan	
	Rebecca Cweibel	
	Yi Zhao	
	Yi-Fan Chou	

Continued

APPENDIX 1 Co-investigators *(continued)*

Affiliation	Full Name	Contributions
	Adam Naj*	
	Elisabeth Mlynarski	
	Gerard Schellenberg*	
	John Malamon	
	Laura Cantwell	
	Nancy Zhang	
	Weixin Wang	
	Yuk Yee Leung	
University of Texas Houston	Eric Boerwinkle*	The University of Texas Houston site contributed expertise to the study design, sequencing of samples, bioinformatics analyses, quality control, data management, structural variation working group, as well as both case-control and family study working groups
	Jan Bressler	
	Jennifer E. Below	
	Myriam Fornage	
	Xiaoming Liu	
	Xueqiu Jian	
University of Washington	Alejandro Q Nato Jr.	The University of Washington site contributed expertise to the bioinformatics analyses, quality control, data management, structural variation working group, case-control working group, and family study working group
	Andrea R Horimoto	
	Bowen Wang	
	Bruce Psaty	
	Daniela Witten	
	Debby Tsuang	
	Elizabeth Blue*	
	Ellen Wijsman*	
	Harkirat Sohi	
	Hiep Nguyen	
	Joshua C. Bis	
	Kenneth Rice	
	Lisa Brown	
	Michael Dorschner	
	Mohamad Saad	
	Pat Navas	
	Rafael Nafikov	
	Timothy Thornton*	
	Tyler Day	

Continued

APPENDIX 1 Co-investigators *(continued)*

Affiliation	Full Name	Contributions
Washington University St. Louis	Carlos Cruchaga	The Washington University St. Louis site contributed expertise to the study design, sequencing of samples, bioinformatics analyses, quality control, data management, structural variation working group, as well as both case-control and family study working groups
	Daniel C. Koboldt	
	David E. Larson	
	Elizabeth Appelbaum	
	Jason Waligorski	
	Lucinda Antonacci-Fulton	
	Richard K. Wilson	
	Robert S. Fulton	

Asterisks (*) indicate coinvestigators whose contributions were sufficient for authorship.

Received February 6, 2018. Accepted in final form October 3, 2018.

References

1. Lane CA, Hardy J, Schott JM. Alzheimer's disease. Eur J Neurol 2018;25:59–70.
2. Gatz M, Pedersen NL, Berg S, et al. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. J Gerontol A Biol Sci Med Sci 1997;52: M117–M125.
3. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261: 921–923.
4. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45: 1452–1458.
5. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 2011;43:436–441.
6. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. N Engl J Med 2013;368:117–127.
7. Reitz C, Mayeux R. Alzheimer's Disease Genetics C. TREM2 and neurodegenerative disease. N Engl J Med 2013;369:1564–1565.
8. Le Guennec K, Nicolas G, Quenez O, et al. ABCA7 rare variants and Alzheimer disease risk. Neurology 2016;86:2134–2137.
9. Vardarajan BN, Zhang Y, Lee JH, et al. Coding mutations in SORL1 and Alzheimer disease. Ann Neurol 2015;77:215–227.
10. Beecham GW, Bis JC, Martin ER, et al. The Alzheimer's disease sequencing project: study design and sample selection. Neurol Genet 2017;3:e194.
11. Preston MD, Dudbridge F. Utilising family-based designs for detecting rare variant disease associations. Ann Hum Genet 2014;78:129–140.
12. Vardarajan BN, Barral S, Jaworski J, et al. Whole genome sequence analysis of caribbean hispanic families with late-onset Alzheimer's disease. Ann Clin Transl Neurol 2018;5:406–417.
13. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadian EM. 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 1984;34:939–944.
14. Naj AC, Lin H, Vardarajan BN, et al. Quality control and integration of genotypes from two calling pipelines for whole genome sequence data in the Alzheimer's disease sequencing project. Genomics Epub 2018 May 29.
15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–1760.
16. Challis D, Yu J, Evani US, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics 2012;13:8.
17. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909.
18. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97–101.
19. Kunkle BW, Jaworski J, Barral S, et al. Genome-wide linkage analyses of non-Hispanic white families identify novel loci for familial late-onset Alzheimer's disease. Alzheimers Dement 2016;12:2–10.
20. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 2007;81:559–575.
21. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–247.
22. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–315.
23. Maurano MT, Haugen E, Sandstrom R, et al. Large-scale identification of sequence variants influencing human transcription factor occupancy. Nat Genet 2015;47:1393–1400.
24. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526:68–74.
25. Butkiewicz M, Blue E, Leung F, et al. Functional annotation of genomic variants in studies of Late-onset Alzheimer's disease. Bioinformatics 2018;34:2724–2731.
26. Yan Q, Tiwari HK, Yi N, et al. A sequence kernel association test for dichotomous traits in family samples under a generalized linear mixed model. Hum Hered 2015;79: 60–68.
27. Kholi MA, Cukier HN, Hamilton-Nelson KL, et al. Segregation of a rare TTC3 variant in an extended family with late-onset Alzheimer disease. Neurol Genet 2016;2:e41.
28. Bis JC, Jian X, Kunkle BW, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. Mol Psychiatry Epub 2018 Aug 14.
29. Melé M, Ferreira PG, Reverter F, et al. Human genomics: the human transcriptome across tissues and individuals. Science 2015;348:660–665.
30. Beisiegel U, Weber W, Ihrke G, Herz J, Stanley KK. The LDL-receptor-related protein, LRP, is an apolipoprotein E-binding protein. Nature 1989;341: 162–164.
31. Vázquez-Higuera JL, Mateo I, Sánchez-Juan P, et al. Genetic interaction between tau and the apolipoprotein E receptor LRP1 Increases Alzheimer's disease risk. Dement Geriatr Cogn Disord 2009;28:116–120.
32. Jaffrey SR, Snowman AM, Eliasson MJ, Cohen NA, Snyder SH. CAPON: a protein associated with neuronal nitri oxide synthase that regulates its interactions with PSD95. Neuron 1998;20:115–124.
33. Galimberti D, Scarpini E, Venturelli E, et al. Association of a NOS1 promoter repeat with Alzheimer's disease. Neurobiol Aging 2008;29:1359–1365.
34. Villarreal-Molina MT, Aguilar-Salinas CA, Rodríguez-Cruz M, et al. The ATP-binding cassette transporter A1 R230C variant affects HDL cholesterol levels and BMI in the Mexican population: association with obesity and obesity-related comorbidities. Diabetes 2007;56:1881–1887.
35. Rhyne J, Mantaring MM, Gardner DF, Miller M. Multiple splice defects in ABCA1 cause low HDL-C in a family with hypoalphalipoproteinemia and premature coronary disease. BMC Med Genet 2009;10:1.
36. Cenarro A, Artieda M, Castillo S, et al. A common variant in the ABCA1 gene is associated with a lower risk for premature coronary heart disease in familial hypercholesterolaemia. J Med Genet 2003;40:163–168.
37. Brooks-Wilson A, Marcil M, Clee SM, et al. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. Nat Genet 1999;22:336–345.
38. Bodzioch M, Orsó E, Klucken J, et al. The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. Nat Genet 1999;22:347–351.
39. Reitz C. Dyslipidemia and the risk of Alzheimer's disease. Curr Atheroscler Rep 2013; 15:307.
40. Kim WS, Hill AF, Fitzgerald ML, et al. Wild type and Tangier disease ABCA1 mutants modulate cellular amyloid-β production independent of cholesterol efflux activity. J Alzheimers Dis 2011;27:441–452.

41. Pervaiz MA, Gau G, Jaffe AS, Saenger AK, Baudhuin L, Ellison J. A non-classical presentation of tangier disease with three ABCA1 mutations. JIMD Rep 2012;4:109–111.
42. Fan J1, Donkin J, Wellington C. Greasing the wheels of Abeta clearance in Alzheimer's disease: the role of lipids and apolipoprotein E. Biofactors 2009;35:239–248.
43. Fitz NF, Cronican AA, Saleem M, et al. Abca1 deficiency affects Alzheimer's disease-like phenotype in human ApoE4 but not in ApoE3-targeted replacement mice. J Neurosci 2012;32:13125–13136.
44. Shahim P1, Bochem AE, Mattsson N, et al. Plasma amyloid-β in patients with Tangier disease. J Alzheimers Dis 2013;35:307–312.
45. Palomer X1, Mauricio D, Rodríguez-Espinosa J, et al. Evaluation of two nonisotopic immunoassays for determination of glutamic acid decarboxylase and tyrosine phosphatase autoantibodies in serum. Clin Chem 2004;50:1378–1382.
46. Yu L, Chibnik LB, Srivastava GP, et al. Association of Brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. JAMA Neurol 2015;72:15–24.
47. Mather CA, Mooney SD, Salipante SJ. CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. Genet Med 2016;18:1269–1275.