



Online article and related content
current as of August 12, 2010.

Genome-wide Analysis of Genetic Loci Associated With Alzheimer Disease

Sudha Seshadri; Annette L. Fitzpatrick; M. Arfan Ikram; et al.

JAMA. 2010;303(18):1832-1840 (doi:10.1001/jama.2010.574)

<http://jama.ama-assn.org/cgi/content/full/303/18/1832>

Supplementary material

eSupplement

<http://jama.ama-assn.org/cgi/content/full/303/18/1832/DC1>

Correction

[Contact me if this article is corrected.](#)

Citations

[This article has been cited 3 times.](#)

[Contact me when this article is cited.](#)

Topic collections

Neurology; Alzheimer Disease; Neurogenetics; Genetics; Genetic Disorders; Genetics, Other

[Contact me when new articles are published in these topic areas.](#)

Related Articles published in
the same issue

Reaching the Limits of Genome-wide Significance in Alzheimer Disease: Back to the Environment

Nancy L. Pedersen. *JAMA*. 2010;303(18):1864.

Subscribe

<http://jama.com/subscribe>

Permissions

permissions@ama-assn.org

<http://pubs.ama-assn.org/misc/permissions.dtl>

Email Alerts

<http://jamaarchives.com/alerts>

Reprints/E-prints

reprints@ama-assn.org

Genome-wide Analysis of Genetic Loci Associated With Alzheimer Disease

Sudha Seshadri, MD; Annette L. Fitzpatrick, PhD; M. Arfan Ikram, MD, PhD; Anita L. DeStefano, PhD; Vilmondur Gudnason, MD, PhD; Merce Boada, MD, PhD; Joshua C. Bis, PhD; Albert V. Smith, PhD; Minerva M. Carrasquillo, PhD; Jean Charles Lambert, PhD; Denise Harold, PhD; Elisabeth M. C. Schrijvers, MD; Reposo Ramirez-Lorca, PhD; Stephanie Debette, MD, PhD; W. T. Longstreth Jr, MD; A. Cecile J. W. Janssens, PhD; V. Shane Pankratz, PhD; Jean François Dartigues, PhD; Paul Hollingworth, PhD; Thor Aspelund, PhD; Isabel Hernandez, MD; Alexa Beiser, PhD; Lewis H. Kuller, MD; Peter J. Koudstaal, MD, PhD; Dennis W. Dickson, MD; Christophe Tzourio, MD; Richard Abraham, PhD; Carmen Antunez, MD; Yangchun Du, PhD; Jerome I. Rotter, MD; Yuri S. Aulchenko, PhD; Tamara B. Harris, MD; Ronald C. Petersen, MD; Claudine Berr, MD, PhD; Michael J. Owen, MB, ChB, PhD; Jesus Lopez-Arrieta, MD; Badri N. Vardarajan, MS; James T. Becker, PhD; Fernando Rivadeneira, MD, PhD; Michael A. Nalls, PhD; Neill R. Graff-Radford, MD; Dominique Campion, MD, PhD; Sanford Auerbach, MD; Kenneth Rice, PhD; Albert Hofman, MD, PhD; Palmi V. Jonsson, MD; Helena Schmidt, MD, PhD; Mark Lathrop, PhD; Thomas H. Mosley, PhD; Rhoda Au, PhD; Bruce M. Psaty, MD, PhD; Andre G. Uitterlinden, PhD; Lindsay A. Farrer, PhD; Thomas Lumley, PhD; Agustin Ruiz, MD, PhD; Julie Williams, PhD; Philippe Amouyel, MD, PhD; Steve G. Younkin, PhD; Philip A. Wolf, MD; Lenore J. Launer, PhD; Oscar L. Lopez, MD; Cornelia M. van Duijn, PhD; Monique M. B. Breteler, MD, PhD for the CHARGE, GERAD1, and EADI Consortia

ONE OF EVERY 5 PERSONS AGED 65 years is predicted to develop Alzheimer disease (AD) in their lifetime, and genetic variants may play an important part in the development of the disease.¹

For editorial comment see p 1864.

Context Genome-wide association studies (GWAS) have recently identified *CLU*, *PICALM*, and *CR1* as novel genes for late-onset Alzheimer disease (AD).

Objectives To identify and strengthen additional loci associated with AD and confirm these in an independent sample and to examine the contribution of recently identified genes to AD risk prediction in a 3-stage analysis of new and previously published GWAS on more than 35 000 persons (8371 AD cases).

Design, Setting, and Participants In stage 1, we identified strong genetic associations ($P < 10^{-3}$) in a sample of 3006 AD cases and 14 642 controls by combining new data from the population-based Cohorts for Heart and Aging Research in Genomic Epidemiology consortium (1367 AD cases [973 incident]) with previously reported results from the Translational Genomics Research Institute and the Mayo AD GWAS. We identified 2708 single-nucleotide polymorphisms (SNPs) with $P < 10^{-3}$. In stage 2, we pooled results for these SNPs with the European AD Initiative (2032 cases and 5328 controls) to identify 38 SNPs (10 loci) with $P < 10^{-5}$. In stage 3, we combined data for these 10 loci with data from the Genetic and Environmental Risk in AD consortium (3333 cases and 6995 controls) to identify 4 SNPs with $P < 1.7 \times 10^{-8}$. These 4 SNPs were replicated in an independent Spanish sample (1140 AD cases and 1209 controls). Genome-wide association analyses were completed in 2007-2008 and the meta-analyses and replication in 2009.

Main Outcome Measure Presence of Alzheimer disease.

Results Two loci were identified to have genome-wide significance for the first time: rs744373 near *BIN1* (odds ratio [OR], 1.13; 95% confidence interval [CI], 1.06-1.21 per copy of the minor allele; $P = 1.59 \times 10^{-11}$) and rs597668 near *EXOC3L2/BLOC1S3/MARK4* (OR, 1.18; 95% CI, 1.07-1.29; $P = 6.45 \times 10^{-9}$). Associations of these 2 loci plus the previously identified loci *CLU* and *PICALM* with AD were confirmed in the Spanish sample ($P < .05$). However, although *CLU* and *PICALM* were confirmed to be associated with AD in this independent sample, they did not improve the ability of a model that included age, sex, and *APOE* to predict incident AD (improvement in area under the receiver operating characteristic curve from 0.847 to 0.849 in the Rotterdam Study and 0.702 to 0.705 in the Cardiovascular Health Study).

Conclusions Two genetic loci for AD were found for the first time to reach genome-wide statistical significance. These findings were replicated in an independent population. Two recently reported associations were also confirmed. These loci did not improve AD risk prediction. While not clinically useful, they may implicate biological pathways useful for future research.

JAMA. 2010;303(18):1832-1840

www.jama.com

The apparent substantial heritability of late-onset AD² is inadequately explained by genetic variation within the well-replicated genes (apolipoprotein E [*APOE*; RefSeq NG_007084], presenilin-1 [*PSEN1*; RefSeq NG_007386], presenilin-2 [*PSEN2*; RefSeq NG_007381], and amyloid beta precursor protein [*APP*;

RefSeq NM_000484]).³ Initial genome-wide association studies (GWAS) identified putative new candidate genes

Author Affiliations are available online at <http://www.jama.com>.

Corresponding Author: Monique M. B. Breteler, MD, PhD, Department of Epidemiology Erasmus MC, University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, the Netherlands (m.breteler@erasmusmc.nl).

(GRB2-associated binding protein [GAB2; RefSeq NG_016171], protocadherin 11 x-linked [PCDH11X; RefSeq NG_016251], lecithin retinol acyltransferase [LRAT; RefSeq NG_009110], and transient receptor potential cation channel, subfamily C, member 4-associated protein [TRPC4AP; RefSeq NM_015638])^{4,6} and regions of interest (eg, on chromosomes 14q, 10q, and 12q),⁷⁻¹⁰ but no locus outside of the APOE region consistently reached genome-wide significance.^{4,11,12} These disappointing results are most likely explained by the modest sample size and, hence, limited statistical power of early studies to detect genes with small effects. Recently, 2 large GWAS, the United Kingdom-led Genetic and Environmental Risk in Alzheimer Disease 1 consortium (GERAD1)¹³ and the European Alzheimer Disease Initiative stage 1 (EADI1),¹⁴ reported 3 new genome-wide significant loci for AD: within the *CLU* gene (GenBank AY341244) encoding clusterin (also called apolipoprotein J), near the *PICALM* gene (GenBank BC073961) encoding phosphatidylinositol-binding clathrin assembly protein, and within the *CRI1* (RefSeq NG_007481) gene encoding complement component (3b/4b) receptor 1.^{13,14}

We performed a 3-stage analysis of GWAS data to identify additional loci associated with late-onset AD. Moreover, we sought to replicate genome-wide significant loci, from both the current analysis and previous reports, in an independent case-control population. Finally, we used 2 large, prospective, population-based studies to assess the improvement in incident AD risk prediction conferred by the recently described loci.

METHODS

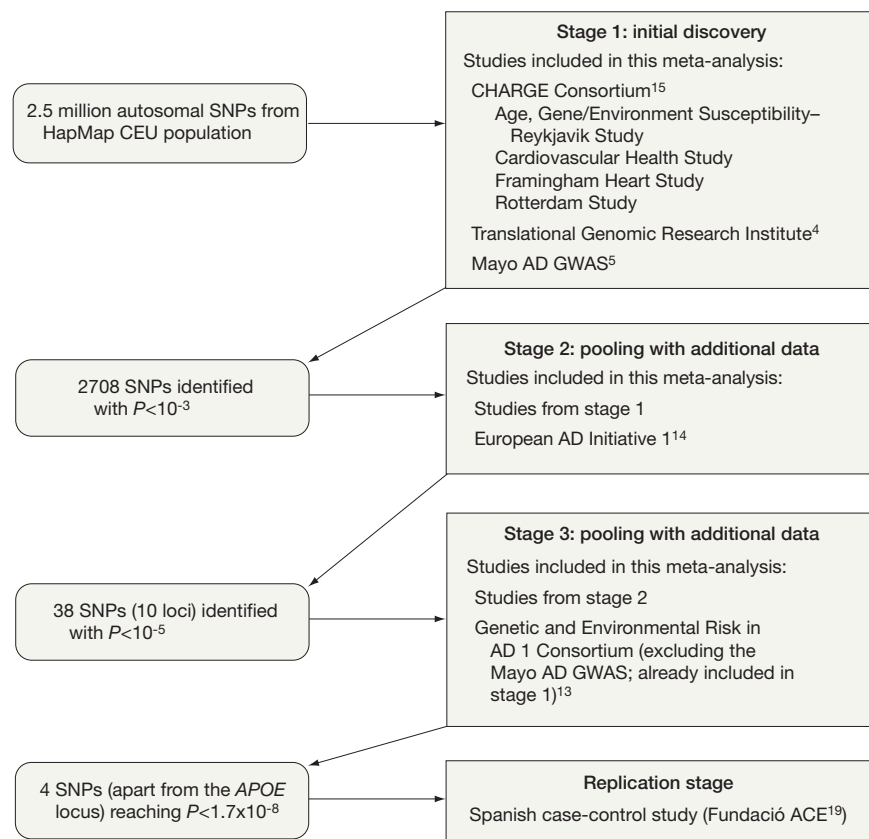
We used a 3-stage sequential analysis to identify novel loci associated with late-onset AD (FIGURE 1). Our initial investigation (stage 1) was a meta-analysis combining new genome-wide association data from white participants in the large, population-based Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)

consortium¹⁵ with GWAS data from the Translational Genomics Research Institute (TGEN) public release database⁴ and the Mayo AD GWAS.⁵ The sample characteristics of the participants contributing to this discovery stage are summarized in TABLE 1. In stage 2, we combined results for our most suggestive findings (single-nucleotide polymorphisms [SNPs] with $P < 10^{-3}$) with corresponding results in the EADI1 consortium.¹⁴ In stage 3, we combined results for the most promising hits in stage 2 (selecting top SNPs from all loci that reached $P < 10^{-5}$) with data from the nonoverlapping studies within the GERAD1 consortium (ex-

cluding the Mayo AD GWAS, the only overlapping study).¹³ All participants (or their authorized proxies) in the contributing studies gave written informed consent including for genetic analyses. Local institutional review boards approved study protocols. Details of study sample selection for the contributing studies are described in section 2 of the eAppendix (section 1 lists abbreviations used in this article) and in eFigure 1, parts A through D (available at <http://www.jama.com>).

In each study, dementia was defined using the *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition Revised or Fourth Edition (DSM-IV)

Figure 1. The 3-Stage Approach and the Various Studies Included in the Different Stages



AD indicates Alzheimer disease; APOE, apolipoprotein E; CHARGE, Cohorts for Heart and Aging Research in Genomic Epidemiology; Fundació ACE, Fundació Alzheimer Centre Educacional; GWAS, genome-wide association study; SNP, single-nucleotide polymorphism.

criteria.¹⁶ Among persons with dementia, all studies used the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) criteria to define AD and included persons with definite (diagnosis of AD pathologically confirmed at autopsy), probable, or possible AD.¹⁷

The individual studies in stage 1 were genotyped on different platforms, as shown in Table 1. The EAD11 used the Illumina Quad 6.0 (Illumina Inc, San Diego, California), and GERAD1 was genotyped on various Illumina chips. In each of the CHARGE cohorts and in the TGEN database, we used the genotype data to impute to the 2.5 million nonmonomorphic autosomal SNPs described in HapMap (Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain collection [CEU population]). Imputations are needed to meta-analyze genome-wide association data across stud-

ies that have used different genotyping platforms because the platforms differ in the SNPs genotyped. Imputation methods and quality control filters in each sample are described in section 3 of the eAppendix.

All analyses were restricted to white persons, racial identity being self-defined by the participants (see section 2 of the eAppendix). We included white Hispanics and adjusted for population structure. Since only 1 of the CHARGE studies, the Cardiovascular Health Study (CHS), had a small number of African American participants (n=574 with genotyping), this racial subgroup was too small for independent analysis. Linkage disequilibrium patterns are very different in individuals of African heritage, which leads to greater uncertainty in imputation, as well as the possibility of false-positive associations if data from 2 racial groups are combined when disease risk differs by race (referred to as population stratification); hence, African American participants in the CHS were excluded from these analyses. Each

study fit an additive genetic model (a 1-degree-of-freedom trend test) relating genotype dosage (0, 1, or 2 copies of the minor allele) to study trait. In the CHARGE cohorts, prevalent cases were compared with controls free of dementia at the DNA draw date. Participants were excluded if they declined consent or genotyping failed. For analysis of prevalent events in the CHARGE cohorts and for the case-control data from the TGEN and Mayo cohorts, we used logistic regression models. For the analysis of incident events in the CHARGE cohorts, participants who were free of dementia entered the analysis at the time of the DNA sample collection and were followed up until development of incident AD; participants were censored at death, at the time of their last follow-up examination or health status update when they were known to be free of clinical dementia, and when they developed dementia from a cause other than AD. We used Cox proportional hazards models to calculate hazard ratios with corresponding 95% confidence intervals

Table 1. Characteristics of Studies in Stage 1 of the Analysis

Characteristics	CHS		FHS		Rotterdam Study		AGES		TGEN		Mayo AD GWAS	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Study design	Cohort		Cohort		Cohort		Cohort		Case-control		Case-control	
Genotype platform	Illumina HumanCNV370 Duo		Affymetrix GeneChip Human Mapping 500K Array Set +50K Gene Focused Panel		Illumina Infinium HumanHap550 Chip Version 3.0		Illumina HumanCNV370 Duo		Affymetrix GeneChip Human Mapping 500K Array Set		Illumina HumanHap300 Version 2 Duo BeadChips	
Prevalence studies ^a												
No. of participants ^b	93	2429	52	2091	171	5700	78	2684	829	536	810	1202
Women, No. (%)	49 (53)	1506 (62)	42 (81)	1192 (57)	128 (75)	3347 (59)	39 (50)	1557 (58)	431 (52)	338 (63)	462 (57)	601 (50)
Age, mean (SD), y	80 (6)	75 (5)	87 (6)	76 (7)	84 (9)	69 (9)	81 (5)	76 (5)	81 (10)	80 (7)	73 (4)	74 (5)
APOE ε4 positive, No. (%) ^c	35 (38)	583 (24)	20 (38)	418 (20)	62 (36)	1549 (28)	38 (49)	725 (27)	481 (58)	107 (20)	535 (66)	337 (28)
Incidence studies ^a												
Cohort at risk, No. ^b	2429		806		5700							
Women, No. (%)	1506 (62)		484 (60)		3347 (59)							
Age at start/age at incident dementia, mean (SD), y	75 (5)/ 82 (5)		82 (6)/ 88 (5)		69 (9)/ 82 (7)							
Incident AD cases, No.	435		76		462							
Follow-up, mean (SD), y	6.8 (3.6)		4.8 (3.0)		9.3 (3.2)							
APOE ε4 positive, No. (%) ^c	632 (26)		153 (19)		1549 (28)							

Abbreviations: AD, Alzheimer disease; AGES, Age, Gene/Environment Susceptibility-Reykjavik Study; APOE, apolipoprotein E; CHS, Cardiovascular Health Study; FHS, Framingham Heart Study; GWAS, genome-wide association study; TGEN, Translational Genomics Research Institute.

^aIn the prevalence studies, cases were persons who had AD at time of DNA draw. Controls were those who were free of any dementia. In the incidence studies, cases were persons from the cohort at risk who developed dementia during follow-up. Persons who developed another type of dementia were censored at date of onset.

^bIncludes only persons with genotyping who also provided consent for these analyses and had high-quality genotyping (met quality control criteria) (eAppendix). In the FHS, only original cohort participants were included in incident analyses.

^cAmong those with APOE genotyping available.

(CIs) after ensuring that assumptions of proportionality of hazards were met. In the CHS, Framingham Heart Study (FHS), and Rotterdam Study, controls contributed one set of person-years to the prevalent analysis and a second, nonoverlapping set of person-years to the incident analyses. Under the Martingale property of Cox models, the 2 analyses are independent, and their independence was confirmed in simulation studies. Primary analyses were adjusted for age and sex and any evidence of population stratification. Details of the screening for latent population substructure in each discovery sample are available in section 4 of the eAppendix. In addition, the CHS also adjusted for study site and the FHS accounted for familial relationships (by using a Cox model with robust variance estimator clustering on pedigree to account for family relationships) and for whether the DNA had been whole-genome amplified.

Our stage 1 meta-analysis combined results from 9 discrete sources: incident AD in the CHS, FHS, and Rotterdam Study; prevalent AD in the Age, Gene/Environment Susceptibility–Reykjavik Study, CHS, FHS, and Rotterdam Study; and the TGEN and Mayo AD GWAS case-control studies. We used inverse-variance weighting (also known as a fixed-effects analysis) for meta-analysis, applying genomic control to each study of stage 1. This approach assigns greater weight to more precise (study-specific) estimators; thus, greater weight is given to studies in which a given SNP was genotyped or more effectively imputed and to studies with larger sample sizes. Details of the meta-analyses are available in section 5 of the eAppendix. We retained only SNP-phenotype associations that were based on results from at least 2 of the 9 discovery samples and for which the minor allele frequency was at least 2%. For stages 2 and 3, we again used inverse-variance meta-analysis but without genomic control adjustment. We decided a priori on a genome-wide significance threshold of $P < 1.7 \times 10^{-8}$, which gives, for a 3-stage sequential analysis, the same control of false-positives as a single

study's use of $P < 5 \times 10^{-8}$.¹⁸ The 3 stages of meta-analyses were completed in May to August 2009.

Significant hits from stage 3 of the discovery phase were replicated in an independent Spanish case-control sample (the Fundació Alzheimer Centre Educacional [ACE]) of 1140 patients with AD (mean age, 78.8 [SD, 7.9] years; 69.9% women) compared with 1209 general population controls (mean age, 49.9 [SD, 9.2] years; 52.8% women).^{19,20} All AD patients fulfilled DSM-IV criteria for dementia and NINCDS-ADRDA criteria for possible and probable AD.^{16,17} Both cases and controls were white. Details of the sample are provided in section 6 of the eAppendix. Genotyping was undertaken using real-time polymerase chain reaction coupled with fluorescence resonance energy transfer. Effect sizes for single markers were calculated by unconditional logistic regression analysis using SPSS software, version 13.0 (SPSS Inc, Chicago, Illinois). Replication was completed in October 2009.

In secondary analyses, we also examined results for previously reported loci.^{5,13,14} For these loci, which included the recently reported loci by the EADI1 and GERAD1 consortia, we restricted our analysis to the previously unpublished CHARGE data. We did not assess the association with *PCDH11X* because we focused only on autosomal SNPs in these analyses. We did examine associations with the top 15 candidate genes listed in the Alzgene database (<http://www.alzgene.org>)²¹ as of August 12, 2009, including the *APOE/TOMM40/APOC1* locus and 12 genes outside that locus. Details of SNPs selected and results for these SNPs are provided in section 7 of the eAppendix and in eTable 1.

We sought to estimate the effect of recently identified loci on 10-year risk prediction in the general population using the data for prospectively ascertained incident AD in the 2 largest community-based cohort studies available to us (Rotterdam Study and CHS). In these analyses, we only included SNPs from the 2 loci that were shown to have

genome-wide significance in previous publications and that we replicated nominally within CHARGE: *PICALM* and *CLU* ($P < .05$). Moreover, the analysis was restricted to incident AD to avoid survival bias and was restricted to population-based samples because case-control studies may overestimate the effects of the genes if cases and controls were not randomly selected from the populations in which AD risk prediction is to be applied.²² The improvement in risk prediction was investigated by comparing 3 sequentially incremental AD risk prediction models that first incorporated age and sex alone, then added data on risk allele status at the *APOE* locus and, finally, data on risk allele status at the *CLU* and *PICALM* loci. We did not assess the utility of novel loci uncovered in this article (using CHARGE as part of the discovery sample) to avoid the risk of overestimating effects by using the same sample for gene discovery and risk prediction.²² Prediction models were constructed using Cox proportional hazards methods using the R package *survcomp*. *APOE* $\epsilon 4$ status was included as a discrete variable (0, 1, or 2 alleles) and the other 2 genetic loci as dosages; all gene effects were examined using additive models. The accuracy of risk prediction for each model was assessed as discriminative accuracy, measured by the area under the receiver operating characteristic curve (AUC). The AUC theoretically ranges from 0.50 (as predictive as tossing a coin) to 1.00 (perfect prediction).

RESULTS

The stage 1 meta-analysis had 8935 dementia-free individuals (mean age, 72 [SD, 7] years), of whom 973 developed incident AD over an average follow-up time of 8 [SD, 3] years, and 2033 prevalent cases of AD who were compared with 14 642 dementia-free controls. Of these, 1367 AD cases (973 incident) were from the CHARGE cohorts. In this discovery analysis based on the CHARGE cohorts, TGEN, and Mayo GWAS, there was no evidence of spurious inflation of *P* values or sig-

nificant population stratification (see eFigure 2 for the quantile-quantile plot comparing the observed and expected *P* value distributions). eFigure 3 illustrates the primary findings from the stage 1 meta-analysis in a Manhattan plot showing genome-wide *P* values for all interrogated SNPs across the 22 autosomal chromosomes. After stage 1, 2708 SNPs had a $P < 10^{-3}$ and were studied in stage 2. In stage 2, pooling these results with data from EADI1 (2032 cases and 5328 controls), 38 SNPs in 10 loci had a $P < 10^{-5}$. Finally, in stage 3, the most significant SNPs from these 10 loci were meta-analyzed with the nonoverlapping studies from GERAD1 (3333 cases and 6995 controls). The findings of the stages 1, 2, and 3 analyses at these 10 loci are presented in TABLE 2. Additional details are provided in eTable 2, which shows chromosomal location, adjacent genes,

sample- and stage-specific estimates of relative risks, 95% CIs, and *P* values for each of the 38 SNPs selected in stage 2 analyses. FIGURE 2 and FIGURE 3 show regional association plots for the 2 SNPs not previously reported to have reached genome-wide significance, rs744373 and rs597668 on chromosomes 2 and 19, respectively. In these figures, we show the linkage disequilibrium (with the index SNP) and stage 1, 2, and 3 association results for the index SNP and stage 1 results for all SNPs within 200 kilobase (kb) on either side of the index SNP at that locus, as well as gene locations and recombination rates in the region. Regional association plots for the other loci listed in Table 2 are presented in eFigure 4, eFigure 5, eFigure 6, eFigure 7, and eFigure 8.

In stage 1, 11 SNPs in the *APOE*/*TOMM40*/*APOC1* region reached our preset threshold for genome-wide sig-

nificance (eTable 2 and eFigure 3). In stage 2, 2 additional loci, rs11136000 in *CLU* and a locus (rs11771145) at chromosome 7 in the 5' upstream promoter/regulatory region of ephrin receptor A1 (*EPHA1*; GenBank AH007960) reached genome-wide significance. However, the latter became nonsignificant after adding GERAD1 data in stage 3, though the effect seen in GERAD1 was in the same direction in that the same allele was associated with an increased (but nonsignificant) risk of AD. In stage 3, genome-wide significant evidence for association with AD was reached at the *APOE* (rs2075650; $P = 1.04 \times 10^{-295}$), *CLU* (rs11136000; $P = 1.62 \times 10^{-16}$), and *PICALM* (rs3851179; $P = 3.16 \times 10^{-12}$) loci, as well as for 2 novel loci on chromosomes 2 (rs744373; $P = 1.59 \times 10^{-11}$) and 19 (rs597668; $P = 6.45 \times 10^{-9}$). Table 2 shows the odds ratios (ORs)

Table 2. Genetic Loci at Which SNPs Were Associated With Alzheimer Disease at $P < 10^{-5}$ in Stage 2 Meta-analysis and That Underwent Further Meta-analysis in Stage 3

Top SNP ^a	Chromosome: Position	Additional SNPs ^b	Nearest Gene ^c	Minor Allele ^d	MAF	Stage 1 Meta-analysis		Stage 2 Meta-analysis		Stage 3 Meta-analysis	
						Meta-OR (95% CI) ^e	Meta- <i>P</i> Value	Meta-OR (95% CI) ^e	Meta- <i>P</i> Value	Meta-OR (95% CI) ^e	Meta- <i>P</i> Value
rs2075650	19:50087459	18	<i>APOE</i> (RefSeq NG_007084)	G	13.7	2.23 (2.04-2.44)	3.18×10^{-68}	2.61 (2.45-2.80)	4.67×10^{-172}	2.53 (2.41-2.66)	1.04×10^{-295}
rs11136000	8:27520436		<i>CLU</i> (GenBank AY341244)	T	39.2	0.89 (0.83-0.94)	4.98×10^{-4}	0.85 (0.81-0.90)	1.49×10^{-9}	0.85 (0.82-0.88)	1.62×10^{-16}
rs3851179	11:85546288		<i>PICALM</i> (GenBank BC073961)	T	37.1	0.86 (0.81-0.92)	1.22×10^{-5}	0.89 (0.84-0.93)	2.81×10^{-6}	0.87 (0.84-0.91)	3.16×10^{-12}
rs744373	2:127611085		<i>BIN1</i> (RefSeq NG_012042)	G	29.1	1.13 (1.06-1.21)	4.93×10^{-4}	1.14 (1.08-1.20)	1.02×10^{-6}	1.15 (1.11-1.20)	1.59×10^{-11}
rs597668	19:50400728	1	<i>EXOC3L2</i> (RefSeq NM_138568)	C	15.4	1.18 (1.07-1.29)	5.91×10^{-4}	1.18 (1.10-1.26)	2.16×10^{-6}	1.17 (1.11-1.23)	6.45×10^{-9}
rs11771145	7:142820884		<i>EPHA1</i> (GenBank AH007960)	A	34.7	0.87 (0.81-0.94)	2.14×10^{-4}	0.86 (0.81-0.90)	1.32×10^{-8}	0.91 (0.87-0.94)	1.70×10^{-6}
rs2043948	14:74142801		<i>LTBP2</i> (RefSeq NM_000428)	T	7.7	1.25 (1.10-1.42)	6.96×10^{-4}	1.27 (1.16-1.39)	4.44×10^{-7}	1.13 (1.06-1.22)	4.46×10^{-4}
rs2825544	21:19662423		<i>PRSS7</i> (RefSeq NG_012207)	C	34.6	1.13 (1.06-1.21)	2.55×10^{-4}	1.14 (1.08-1.20)	4.85×10^{-7}	1.09 (1.05-1.13)	2.10×10^{-5}
rs7527934 ^f	1:14231011	9	<i>PRDM2</i> (RefSeq NM_012231)	G	25.7	0.86 (0.79-0.93)	3.50×10^{-4}	0.87 (0.82-0.92)	5.87×10^{-6}	0.97 (0.91-1.03)	
rs4296166 ^f	14:32022118		<i>AKAP6</i> (RefSeq NM_004274)	A	47.8	1.14 (1.07-1.21)	8.36×10^{-5}	1.12 (1.07-1.18)	4.08×10^{-6}	0.98 (0.89-1.08)	

Abbreviations: CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.

^aAt each locus, the SNP with the lowest (most statistically significant) *P* value was selected for stage 3 meta-analysis.

^bNumber of additional SNPs at the locus with $P < 10^{-5}$.

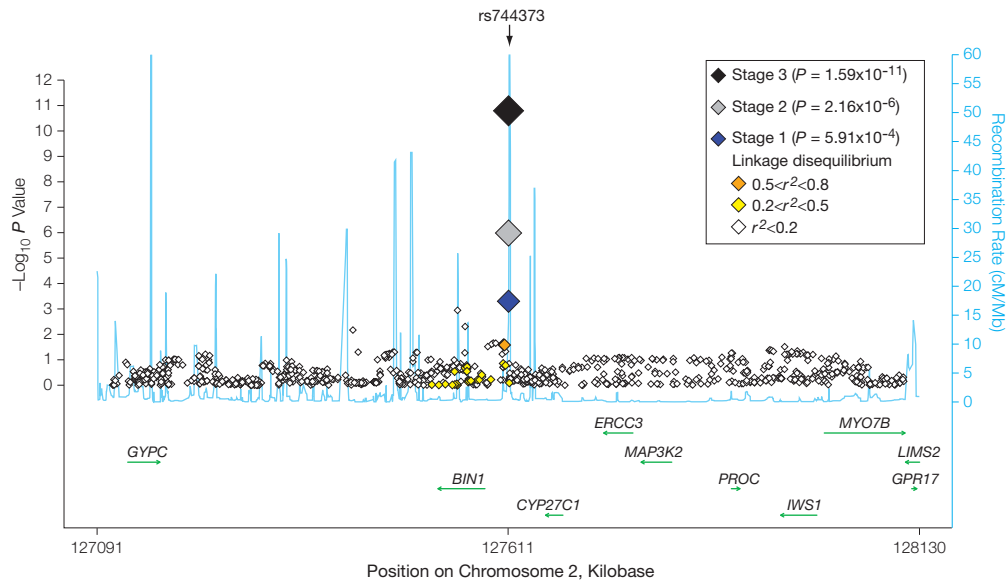
^cColumn shows the Human Gene Organization Gene Nomenclature System symbols for the gene located closest to each SNP. Standardized gene annotations for all SNP results were derived programmatically from the University of California, Santa Cruz Genome Browser RefSeq gene track (hg18).

^dAlleles were coded on the forward strand of the genome.

^eThe minor allele was taken as the coded allele. The ORs represent the relative increase of disease risk per increase of 1 copy of the minor allele.

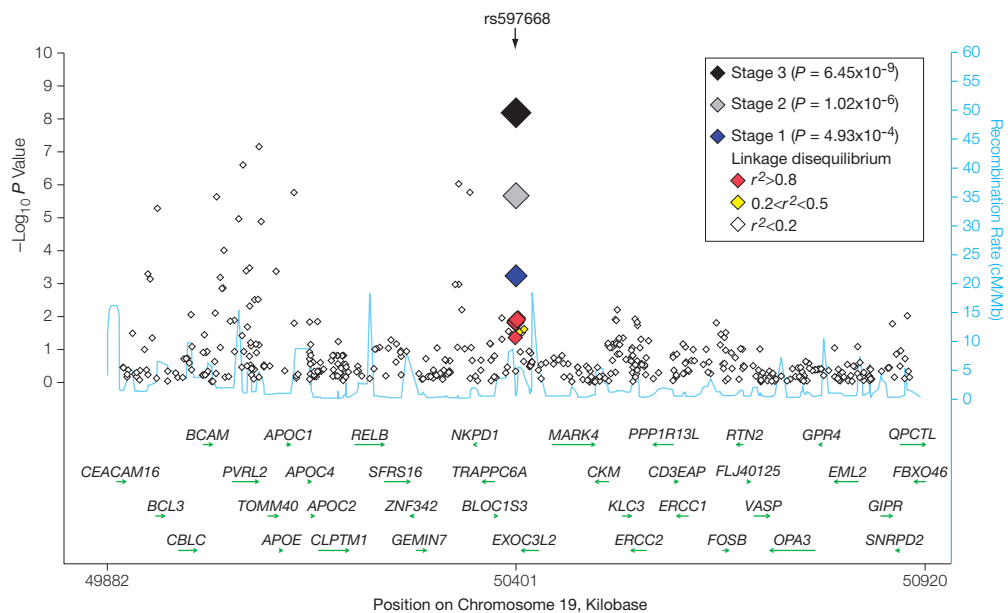
^fThese SNPs did not pass quality control criteria in the Genetic and Environmental Risk in Alzheimer Disease 1 consortium. Stage 3 results are presented from this database for proxy SNPs: rs2206586 for rs7527934 and rs10146289 for rs4296166.

Figure 2. Regional Association Plot for Novel Locus rs744373 Near *BIN1* on Chromosome 2 Significantly Associated ($P < 5 \times 10^{-8}$) With Alzheimer Disease in Stage 3 Analyses



Each data marker represents the statistical significance (P value) of each single-nucleotide polymorphism (SNP) plotted on the $-\log_{10}$ scale against its chromosomal position (NCBI Build 36). The blue diamonds show stage 1 P values for the sentinel (top) SNP at each locus, whereas the gray and black diamonds show the P values for the same SNP following stage 2 and stage 3 meta-analyses, respectively. The fine-scale recombination rate is shown by the blue line, which shows the average frequency with which recombination occurs at that site. Genes located in the region shown are shown as green arrows with HUGO gene nomenclature committee gene symbols. The length of the green arrow represents the size/extent of the gene and the arrowhead the direction in which transcription of mRNA occurs.

Figure 3. Regional Association Plot for Novel Locus rs597668 Near *BLOC1S3* and *MARK4* on Chromosome 19 Significantly Associated ($P < 5 \times 10^{-8}$) With AD in Stage 3 Analyses



Each data marker represents the statistical significance (P value) of each single-nucleotide polymorphism (SNP) plotted on the $-\log_{10}$ scale against its chromosomal position (NCBI Build 36). The blue diamonds show stage 1 P values for the sentinel (top) SNP at each locus, whereas the gray and black diamonds show the P values for the same SNP following stage 2 and stage 3 meta-analyses, respectively. The fine-scale recombination rate is shown by the blue line, which shows the average frequency with which recombination occurs at that site. Genes located in the region shown are shown as green arrows with HUGO gene nomenclature committee gene symbols. The length of the green arrow represents the size/extent of the gene and the arrowhead the direction in which transcription of mRNA occurs.

Table 3. Replication of Genome-wide Significant Results From an Independent Discovery Sample¹⁹

Gene	SNP	MAF, Cases/Controls	OR (95% CI)	P Value
<i>CLU</i>	rs11136000	0.36/0.39	0.82 (0.77-0.99)	.03
<i>PICALM</i>	rs3851179	0.30/0.34	0.84 (0.74-0.95)	.007
<i>BIN1</i>	rs744373	0.30/0.27	1.17 (1.03-1.33)	.02
<i>EXOC3L2</i>	rs597668	0.13/0.11	1.26 (1.05-1.51)	.01

Abbreviations: CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.

associated with the minor allele for each of these SNPs. Locus rs744373 is within 30 kb of the gene bridging integrator 1 (*BIN1*; RefSeq NG_012042) (Figure 2), while rs597668 is within 60 kb of 6 genes including exocyst complex component 3–like 2 (*EXOC3L2*; RefSeq NM_138568), biogenesis of lysosomal organelles complex 1, subunit 3 (*BLOC1S3*; RefSeq NG_008372), and microtubule-associated protein/microtubule affinity-regulating kinase 4 (*MARK4*; GenBank BC071948) (Figure 3).

Independent Replication

We replicated the 4 associations that reached our preset genome-wide significance threshold (1.7×10^{-8}) in an independent sample of cases and controls (TABLE 3). Effect sizes in the replication cohort were similar to those observed in the discovery sample; each of these associations reached $P < .05$.

Conditional Analyses at Chromosome 19 Locus

Because rs597668 is on chromosome 19, fairly close to the *APOE* locus, we undertook conditional analyses to examine whether its association with AD was independent of *APOE* $\epsilon 4$. We conducted 2 analyses with AD (among persons with directly genotyped *APOE* $\epsilon 4$ status) in the CHARGE, TGEN, and Mayo samples, adjusting (1) for our strongest association in the *APOE/TOMM40/APOC1* locus (rs2075650) and (2) for the actual *APOE* $\epsilon 4$ SNP, rs429358. In each case, we found that the association was attenuated but a marginal signal remained when adjusting for *APOE* $\epsilon 4$ (OR, 1.18; 95% CI, 1.08-1.24; $P = 3.9 \times 10^{-4}$ without adjustment; OR, 1.17; 95% CI, 1.07-1.23; $P = 8.7 \times 10^{-4}$ for analysis 1; and OR, 1.10; 95% CI, 1.00-

1.16; $P = .05$ for analysis 2). We also examined the effect of adjusting for age, sex, and presence of at least 1 *APOE* $\epsilon 4$ allele (using a dominant genetic inheritance model) in the Spanish replication sample. Again, the results were attenuated (OR, 1.24; 95% CI, 1.02-1.51; $P = .03$). These findings are consistent with the moderate to low level of linkage disequilibrium observed between rs597668 and SNPs within the *APOE* and *TOMM40* region ($r^2 < 0.01$ according to HapMap CEU data) (Figure 3).

Replication of Previously Reported Associations in the CHARGE Sample

In our secondary analyses examining replication of published findings in the previously unreported CHARGE data, 6 intronic or 3' untranslated region SNPs in the *APOE/TOMM40/APOC1* region (rs6857, rs2075650, rs4420638, rs157582, rs6859, and rs10119) reached a genome-wide significance threshold of $P < 1.7 \times 10^{-8}$, and we replicated the most statistically significant SNPs within 2 of the 3 recently reported genetic loci associated with AD in prior GWAS: *CLU* (rs11136000; OR, 0.90; 95% CI, 0.82-0.98; $P = .02$) and *PICALM* (rs3851179; OR, 0.90; 95% CI, 0.83-0.99; $P = .02$) (eTable 2 and eAppendix). We did not find a significant association with the most statistically significant *CRI* SNP (rs3818361) in the CHARGE data. However, 13 SNPs within the gene showed nominal significance ($P > .001$ but $P < .05$) (eTable 3). Furthermore, adding CHARGE and TGEN data on rs3818361 to the previously reported EADI1 and GERAD1 data (Mayo AD GWAS data were included in the GERAD1 data for this analysis) showed that results now reached genome-wide significance (OR, 1.15; 95% CI, 1.11-1.20; $P = 1.04 \times 10^{-11}$) (eFigure 9).

Among the 54 SNPs selected from the top 12 candidate genes (outside the *APOE/TOMM40/APOC1* locus) listed in the Alzgene Web site, we found evidence for a nominal association of rs4362 in the angiotensin-converting enzyme (*ACE*; RefSeq NC_000017.10) gene and rs1784933 in the sortilin-related receptor L (DLR class A) repeats-containing (*SORL1*; RefSeq NC_000011.9) gene with AD (relative risks associated with each copy of the minor allele were 0.92; 95% CI, 0.85-0.99; $P = .03$ for *ACE* and 1.33; 95% CI, 1.03-1.72; $P = .03$ for *SORL1*) (eTable 1).

Genetic Risk Prediction

We assessed the extent to which *APOE* $\epsilon 4$, *PICALM*, and *CLU* can improve predictive models for risk of incident AD in the general population (represented by the cohorts of the Rotterdam Study and CHS). The addition of *APOE* $\epsilon 4$ carrier status to a prediction model including age and sex only increased the AUC from 0.826 (95% CI, 0.806-0.846) to 0.847 (95% CI, 0.828-0.865) in the Rotterdam Study and from 0.670 (95% CI, 0.625-0.723) to 0.702 (95% CI, 0.654-0.754) in the CHS. Further inclusion of risk allele status for *CLU* and *PICALM* improved the AUC only minimally to 0.849 (95% CI, 0.831-0.867) in the Rotterdam Study and to 0.705 (95% CI, 0.654-0.751) in the CHS. The corresponding receiver operating characteristic curves are shown in eFigure 10.

COMMENT

We report results of an international 3-stage genome-wide analysis to study genetic variation underlying late-onset, sporadic AD. We studied more than 35 000 persons (8371 AD cases), constituting the largest sample analyzed to date. In the gene discovery phase, we showed genome-wide significance for 2 loci related to AD, one on chromosome 2 and a second locus on chromosome 19, that had not previously been found to achieve genome-wide significance and that appear to be independent of *APOE*. *BIN1* was previously identified as having a possible

association with AD in the recent GWAS from the GERAD1,¹³ but until our analysis, this association was not significant at the genome-wide level. Furthermore, we replicated both these loci as well as the recently identified loci, *CLU* and *PICALM*, in an independent sample. Although genetic variation at the *CLU* and *PICALM* loci did modify the risk of AD in our population-based sample, these polymorphisms added very little to prediction of AD risk.

The locus on chromosome 2q14.3 is adjacent to the *BIN1* gene, which is 1 of 2 amphiphysins and is expressed most abundantly in the brain and muscle.²³ Amphiphysins promote caspase-independent apoptosis and also play a critical role in neuronal membrane organization and clathrin-mediated, synaptic vesicle formation,²⁴ a process disrupted by β -amyloid.²⁵ Knockout mice with decreased expression of the amphiphysins have seizures and major learning deficits.²⁶ Altered expression of *BIN1* has been demonstrated in aging mice, in transgenic mouse models of AD, and in persons with schizophrenia.^{27,28}

The 19q13.3 locus (rs597668), a site distal to and not in linkage disequilibrium with SNPs in the *APOE* locus, had been suspected in an early linkage study to harbor a gene for AD.²⁹ There are 6 genes adjacent to this locus, 2 of which are part of pathways linked to AD pathology. The protein product of *BLOC1S3*, biogenesis of lysosomal organelles complex 1, subunit 3, is expressed in the brain, regulates endosomal to lysosomal routing,³⁰ and has been implicated in schizophrenia.³¹ The second gene, *MARK4*, is inducible, expressed only in the brain, and plays a role in neuronal differentiation.³² *MARK4* is a kinase that phosphorylates tau, is polyubiquitinated in vivo, and is a substrate of the aging-related deubiquitinating enzyme USP9X; hence, it may play a role in the abnormal tau phosphorylation seen in AD.³³ Little is known of the function of the gene closest to rs597668, *EXOC3L2*, also referred to as protein 7 transacti-

vated by hepatitis B virus X antigen (*XTP7*) gene.

When evaluating the added value of the new AD genes in clinical risk prediction, we focused on the 2 recently reported AD genes^{13,14} that were replicated in our population-based studies, *CLU* and *PICALM*, and found that they only slightly improved prediction of incident AD beyond age, sex, and *APOE* $\epsilon 4$ based models. The increase in AUC was 0.002 in the Rotterdam Study and 0.003 in the CHS, which would not be of value in the clinical setting. There are 2 reasons for this. First, the associations of *CLU* and *PICALM* with AD risk were markedly lower than those of age and *APOE*; therefore, a major improvement was not expected. This fits with recent insights on polygenic models that assume there are tens of thousands of risk alleles, each with a small (approximately 5% increase in relative risk) effect throughout the whole genome, rather than a discrete number of alleles with moderate effects. Such models appear to underlie the susceptibility to schizophrenia risk, and a similar model may be applicable to AD.³⁴ Second, the extent to which risk factors improve risk prediction depends on the predictive performance of the initial risk model. Added risk factors need to have stronger effects to improve a risk model with high AUC than to improve a model with lower AUC. Alzheimer disease risk prediction based on age, sex, and *APOE* already has very high discriminative accuracy: the AUC was 0.826 in the Rotterdam Study and 0.670 in the CHS, which implies that further improvements require many new variants or variants with strong effects. Whether such improvements are to be expected will depend in large part on the ability to unravel the underlying genetic architecture and to identify and quantify environmental risk factors, including complex interactions.³⁵ A next step for genetic research in AD will be to further increase the sample sizes of GWAS and evaluate further genetic models.

Strengths of this study include the large sample of clinic- and community-

based cases and controls and the subsample of prospectively ascertained incident AD that permitted the exploration of incident risk prediction algorithms. Alzheimer disease was diagnosed using standard NINCDS-ADRDA criteria. The observed associations are unlikely to be due to population stratification since the discovery and replication samples were restricted to white individuals of European origin and were also investigated for latent population substructure.

The study also has limitations. Despite our large sample size, we had limited power to detect associations with small effect sizes and associations with rare variants. Although all studies used accepted clinical or pathological criteria to define dementia and AD, phenotypic heterogeneity between samples may have limited our ability to detect some associations. Moreover, the controls in the Spanish replication sample were younger than the cases, and their cognitive status had not been formally examined. However, whereas this could reduce our power to observe an association, it would not invalidate the associations we did observe. Furthermore, the frequency distribution of minor and major alleles among the Spanish controls was similar to that noted in the discovery sample and in the HapMap CEU sample.

In conclusion, this meta-analysis of GWAS data from several of the largest AD GWAS studies to date confirms previously known and recently described associations (*CLU* and *PICALM*) and shows genome-wide significance and replication for 2 biologically plausible, novel loci on chromosomes 2 and 19. However, the predictive ability of *CLU* and *PICALM* to identify individuals at risk of AD is not clinically significant. The value of these associations may lie in the insights they could provide for research into the pathophysiological mechanisms of AD.

Author Contributions: Drs Seshadri, Ikram, DeStefano, and Breteler had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Seshadri, Fitzpatrick, Ikram, DeStefano, Gudnason, and Boada contributed equally as first authors. Drs Ruiz, Williams, Amouyel, Younkin, Wolf, Launer, Lopez, van Duijn, and Breteler contributed equally as last authors.

Study concept and design: Seshadri, Bis, Rotter, Hofman, Schmidt, Mosley, Lumley, Amouyel, Wolf, Lopez, Duijn, Breteler.

Acquisition of data: Seshadri, Ikram, Gudnason, Boada, Carrasquillo, Lambert, Schrijvers, Ramirez-Lorca, Janssens, Pankratz, Dartigues, Hernandez, Beiser, Kuller, Koudstaal, Dickson, Tzourio, Antunez, Rotter, Berr, Owen, Lopez-Arrieta, Becker, Rivadeneira, Nalls, Graff-Radford, Campion, Auerbach, Jonsson, Lathrop, Au, Psaty, Uitterlinden, Farrer, Ruiz, Williams, Amouyel, Younkin, Lopez, Breteler.

Analysis and interpretation of data: Seshadri, Fitzpatrick, Ikram, DeStefano, Bis, Smith, Carrasquillo, Harold, DeBette, Longstreth, Janssens, Hollingworth, Aspelund, Beiser, Abraham, Du, Rotter, Aulchenko, Harris, Petersen, Owen, Vardarajan, Rivadeneira, Nalls, Rice, Lathrop, Mosley, Farrer, Lumley, Ruiz, Williams, Amouyel, Younkin, Launer, Lopez, Duijn, Breteler.

Drafting of the manuscript: Seshadri, Fitzpatrick, Ikram, DeStefano, Bis, Ramirez-Lorca, Du, Vardarajan, Lopez, Duijn, Breteler.

Critical revision of the manuscript for important intellectual content: Seshadri, Ikram, Gudnason, Boada, Bis, Smith, Carrasquillo, Lambert, Harold, Schrijvers, Ramirez-Lorca, DeBette, Longstreth, Janssens, Pankratz, Dartigues, Hollingworth, Aspelund, Hernandez, Beiser, Kuller, Koudstaal, Dickson, Tzourio, Abraham, Antunez, Rotter, Aulchenko, Harris, Petersen, Berr, Owen, Lopez-Arrieta, Becker, Rivadeneira, Nalls, Graff-Radford, Campion, Auerbach, Rice, Hofman, Jonsson, Schmidt, Lathrop, Mosley, Au, Psaty, Uitterlinden, Farrer, Lumley, Ruiz, Williams, Amouyel, Younkin, Wolf, Launer, Lopez, Breteler.

Statistical analysis: Ikram, DeStefano, Bis, Smith, Harold, Schrijvers, Janssens, Hollingworth, Aspelund, Beiser, Abraham, Du, Aulchenko, Vardarajan, Rivadeneira, Nalls, Rice, Lathrop, Farrer, Lumley, Ruiz, Duijn.

Obtained funding: Seshadri, Gudnason, Boada, Dartigues, Kuller, Antunez, Rotter, Harris, Petersen, Becker, Jonsson, Schmidt, Uitterlinden, Ruiz, Younkin, Launer, Lopez, Breteler.

Administrative, technical, or material support: Seshadri, Fitzpatrick, Bis, Carrasquillo, Ramirez-Lorca, Pankratz, Hernandez, Kuller, Koudstaal, Antunez, Rotter, Becker, Nalls, Campion, Auerbach, Mosley, Psaty, Uitterlinden, Farrer, Wolf.

Study supervision: Seshadri, Gudnason, Boada, Koudstaal, Tzourio, Antunez, Rotter, Berr, Owen, Hofman, Lathrop, Uitterlinden, Williams, Amouyel, Duijn, Breteler.

Financial Disclosures: None reported.

Funding/Support: The funding/support for this study is described in the eAppendix.

Role of the Sponsor: The funding organizations and sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review or approval of the manuscript. The final version submitted was approved without changes by the National Heart, Lung, and Blood Institute and the National Institute on Aging.

Online-Only Material: The eAppendix, eTables 1 through 3, and eFigures 1 through 10 are available at <http://www.jama.com>.

Additional Contributions: Additional contributions are listed in the eAppendix.

REFERENCES

1. Seshadri S, Wolf PA. Lifetime risk of stroke and dementia. *Lancet Neurol*. 2007;6(12):1106-1114.
2. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168-174.
3. Ertekin-Taner N. Genetics of Alzheimer's disease. *Neurol Clin*. 2007;25(3):611-667.
4. Reiman EM, Webster JA, Myers AJ, et al. GAB2 alleles modify Alzheimer's risk in APOE ε4 carriers. *Neuron*. 2007;54(5):713-720.
5. Carrasquillo MM, Zou F, Pankratz VS, et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet*. 2009;41(2):192-198.
6. Poduslo SE, Huang R, Huang J, Smith S. Genome screen of late-onset Alzheimer's extended pedigrees identifies TRPC4AP by haplotype analysis. *Am J Med Genet B Neuropsychiatr Genet*. 2009;150B(1):50-55.
7. Beecham GW, Martin ER, Li YJ, et al. Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am J Hum Genet*. 2009;84(1):35-43.
8. Bertram L, Lange C, Mullin K, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet*. 2008;83(5):623-632.
9. Myers A, Holmans P, Marshall H, et al. Susceptibility locus for Alzheimer's disease on chromosome 10. *Science*. 2000;290(5500):2304-2305.
10. Li H, Wetten S, Li L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol*. 2008;65(1):45-53.
11. Coon KD, Myers AJ, Craig DW, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry*. 2007;68(4):613-618.
12. Feulner TM, Laws SM, Friedrich P, et al. Examination of the current top candidate genes for AD in a genome-wide association study. *Mol Psychiatry*. 2009.
13. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1088-1093.
14. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1094-1099.
15. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. *Circ Cardiovasc Genet*. 2009;2(1):73-80.
16. Association AP. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. Washington, DC: American Psychiatric Association; 1994.
17. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease. *Neurology*. 1984;34(7):939-944.
18. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide

association studies of nearly all common variants. *Genet Epidemiol*. 2008;32(4):381-385.

19. Antúnez C, Boada M, Lopez-Arrieta J, et al. *GOLPH2* gene markers are not associated with Alzheimer's disease in a sample of the Spanish population. *J Alzheimers Dis*. 2009;18(4):751-754.

20. Ramirez-Lorca R, Boada M, Saez ME, et al. *GAB2* gene does not modify the risk of Alzheimer's disease in Spanish APOE 4 carriers. *J Nutr Health Aging*. 2009;13(3):214-219.

21. Bertram L, McQueen MB, Mullin K, et al. Systematic meta-analyses of Alzheimer disease genetic association studies. *Nat Genet*. 2007;39(1):17-23.

22. Janssens AC, van Duijn CM. Genome-based prediction of common diseases. *Genome Med*. 2009;1(2):20.

23. Wechsler-Reya R, Sakamuro D, Zhang J, et al. Structural analysis of the human *BIN1* gene. *J Biol Chem*. 1997;272(50):31453-31458.

24. Wigge P, Kohler K, Vallis Y, et al. Amphiphysin heterodimers. *Mol Biol Cell*. 1997;8(10):2003-2015.

25. Kelly BL, Ferreira A. Beta-amyloid disrupted synaptic vesicle endocytosis in cultured hippocampal neurons. *Neuroscience*. 2007;147(1):60-70.

26. Di Paolo G, Sankaranarayanan S, Wenk MR, et al. Decreased synaptic vesicle recycling efficiency and cognitive deficits in amphiphysin 1 knockout mice. *Neuron*. 2002;33(5):789-804.

27. English JA, Dicker P, Focking M, et al. 2-D DIGE analysis implicates cytoskeletal abnormalities in psychiatric disease. *Proteomics*. 2009;9(12):3368-3382.

28. Yang S, Liu T, Li S, et al. Comparative proteomic analysis of brains of naturally aging mice. *Neuroscience*. 2008;154(3):1107-1120.

29. Poduslo SE, Yin X. A new locus on chromosome 19 linked with late-onset Alzheimer's disease. *Neuroreport*. 2001;12(17):3759-3761.

30. Starcevic M, Dell'Angelica EC. Identification of snapin and three novel proteins (BLOS1, BLOS2, and BLOS3/reduced pigmentation) as subunits of biogenesis of lysosome-related organelles complex-1 (BLOC-1). *J Biol Chem*. 2004;279(27):28393-28401.

31. Morris DW, Murphy K, Kenny N, et al. Dysbindin (*DTNBP1*) and the biogenesis of lysosome-related organelles complex 1 (BLOC-1). *Biol Psychiatry*. 2008;63(1):24-31.

32. Moroni RF, De Biasi S, Colapietro P, et al. Distinct expression pattern of microtubule-associated protein/microtubule affinity-regulating kinase 4 in differentiated neurons. *Neuroscience*. 2006;143(1):83-94.

33. Trinczek B, Brajenovic M, Ebner A, Drewes G. MARK4 is a novel microtubule-associated proteins/microtubule affinity-regulating kinase that binds to the cellular microtubule network and to centrosomes. *J Biol Chem*. 2004;279(7):5915-5923.

34. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-752.

35. Janssens AC, van Duijn CM. Genome-based prediction of common diseases. *Hum Mol Genet*. 2008;17(R2):R166-R173.