

Common variants at *MS4A4/MS4A6E*, *CD2AP*, *CD33* and *EPHA1* are associated with late-onset Alzheimer's disease

The Alzheimer Disease Genetics Consortium (ADGC) performed a genome-wide association study of late-onset Alzheimer disease using a three-stage design consisting of a discovery stage (stage 1) and two replication stages (stages 2 and 3). Both joint analysis and meta-analysis approaches were used. We obtained genome-wide significant results at *MS4A4* (**rs4938933**; stages 1 and 2, meta-analysis P (P_M) = 1.7×10^{-9} , joint analysis P (P_J) = 1.7×10^{-9} ; stages 1, 2 and 3, P_M = 8.2×10^{-12}), *CD2AP* (**rs9349407**; stages 1, 2 and 3, P_M = 8.6×10^{-9}), *EPHA1* (**rs11767557**; stages 1, 2 and 3, P_M = 6.0×10^{-10}) and *CD33* (**rs3865444**; stages 1, 2 and 3, P_M = 1.6×10^{-9}). We also replicated previous associations at *CR1* (**rs6701713**; P_M = 4.6×10^{-10} , P_J = 5.2×10^{-11}), *CLU* (**rs1532278**; P_M = 8.3×10^{-8} , P_J = 1.9×10^{-8}), *BIN1* (**rs7561528**; P_M = 4.0×10^{-14} , P_J = 5.2×10^{-14}) and *PICALM* (**rs561655**; P_M = 7.0×10^{-11} , P_J = 1.0×10^{-10}), but not at *EXOC3L2*, to late-onset Alzheimer's disease susceptibility^{1–3}.

Alzheimer's disease is a neurodegenerative disorder affecting more than 13% of individuals aged 65 years and older and 30–50% of individuals aged 80 years and older^{4,5}. Early work identified mutations in *APP*, *PSEN1* and *PSEN2* that cause early-onset autosomal dominant Alzheimer's disease^{6–9} and variants in *APOE* that affect late-onset Alzheimer's disease (LOAD) susceptibility¹⁰. Recent genome-wide association studies (GWAS) identified variants in *CR1*, *CLU*, *PICALM* and *BIN1* as LOAD susceptibility loci^{1–3}. However, because LOAD heritability estimates are high ($h^2 \approx 60$ –80%)¹¹, much of the genetic contribution to this condition remains unknown.

To identify genetic variants associated with risk for Alzheimer's disease, the ADGC assembled a discovery dataset (stage 1, 8,309 individuals with LOAD (cases) and 7,366 cognitively normal elders (CNEs) as controls) using data from eight cohorts and a ninth newly assembled cohort from the 29 National Institute on Aging (NIA)-funded Alzheimer Disease Centers (ADCs) (Supplementary Tables 1,2 and Supplementary Note), with data coordinated by the National Alzheimer Coordinating Center (NACC) and samples coordinated by the National Cell Repository for Alzheimer Disease (NCRAD). For the stage 2 replication, we used four additional datasets and additional samples from the ADCs (3,531 LOAD cases and 3,565 CNEs). The stage 3 replication used the results of association analyses provided by three other consortia, including 6,992 LOAD cases and 24,666 mixed-age controls, reported in a companion manuscript¹². For stages 1 and 2, we used both a meta-analysis approach that integrated results

from the association analyses of individual datasets and a joint analysis approach in which genotype data from each study were pooled. The latter method has improved power over the meta-analysis in the absence of between-study heterogeneity¹³ and has a more direct correction for confounding sampling bias¹⁴. We were limited to meta-analysis for stage 3 analyses.

Because the cohorts were genotyped using different platforms, we used imputation to generate a common set of 2,324,889 SNPs. We applied uniform stringent quality control measures to all datasets to remove low-quality and redundant samples and problematic SNPs (Supplementary Tables 3,4 and Online Methods). We performed an association analysis assuming an additive model on the log odds ratio scale with adjustment for population substructure using logistic regression for case-control data and generalized estimating equations (GEE) with a logistic model for family data. We combined results from individual datasets in the meta-analysis using the inverse variance method, applying a genomic control to each dataset. We performed the joint analysis using GEE and incorporated terms to adjust for population substructure and site-specific effects (Online Methods). For both approaches, we also examined an extended model of covariate adjustment that adjusted for age (age at onset or death in cases and age at exam or death in controls), sex and number of *APOE* $\epsilon 4$ alleles (0, 1 or 2). Genomic inflation factors (λ) for both the discovery meta-analysis and the joint analysis and extended models were less than 1.05, indicating that there was not substantial inflation of the test statistics (Supplementary Table 3 and Supplementary Fig. 1). Association findings from the meta-analysis and joint analysis were comparable.

In stage 1, the strongest signal was from the *APOE* region (rs4420638; P_M = 1.1×10^{-266} , P_J = 1.3×10^{-253} ; Supplementary Table 5). Excluding the *APOE* region, SNPs at nine distinct loci yielded P_M or $P_J \leq 10^{-6}$ (Table 1; all SNPs with $P < 10^{-4}$ are shown in Supplementary Table 5). SNPs from these nine loci were carried forward to stage 2. Five of these loci had not previously been associated with LOAD at a genome-wide significance level of $P \leq 5.0 \times 10^{-8}$ (loci in *MS4A*, *EPHA1*, *CD33*, *ARID5B* and *CD2AP*). Because the companion study¹² identified SNPs at *ABCA7* to be within a new LOAD locus, we included *ABCA7*-region SNPs in our stage 2 analysis and provided our results to researchers from that study. For all loci listed in Table 1, we did not detect evidence for effect heterogeneity (Supplementary Fig. 2). One newly associated locus (in *MS4A*) was significant in the stage 1+2 analysis. Four other loci approached but

*A full list of authors and affiliations appears at the end of the paper.

Table 1 Genome-wide association results for LOAD in the ADGC stage 1 and stage 2 datasets

Chr.:Mb, chromosome:position (build 19); MA, minor allele; MAF, minor allele frequency; # SNPs, the number of SNPs for which $P \leq 1 \times 10^{-5}$ in meta-analysis from the combined analysis in stage 1+2; OR_{MA} , odds ratio in meta-analysis; $OR_{P_{MA}}$, odds ratio in joint analysis; $P_{P_{MA}}$, P value in joint analysis. Genes with previous case-control genome-wide statistically significant associations: *CR1* (ref. 3), *CLU* (refs. 1, 3), *PICALM* (ref. 1), *BN1* (ref. 2). Association signals represent SNPs with the strongest associations within each locus showing $P \leq 10^{-5}$ in the stage 1 dataset or in or near previously reported genes, excluding the *APOE* region (Supplementary Table 5). Genes with previously published case-control association signals at $P \leq 5.0 \times 10^{-8}$. This locus did not meet this level of statistical significance². [†]Locus previously reported in a family-based association study as genome-wide significant¹⁵. [‡]Locus identified in the companion study¹² with genome-wide significant evidence for association.

Our stage 1+2 analyses also confirmed the association of previously reported loci (in *BIN1*, *CRI1*, *CLU* and *PICALM*) with LOAD (**Table 1**). For each locus, supporting data were *P* values that were less than $P = 5.0 \times 10^{-8}$ in one or both types of analysis.

Table 2 Meta-analysis of stage 1+2 with stage 3 (CHARGE/GERAD/EADI1 Consortia²) GWAS results

Gene:SNP	Cases	Controls	Total	OR _M (95% CI)	P _M	OR _J (95% CI)	P _J
CD2AP: rs9349407							
ADGC	11,840	10,931	22,771	1.12 (1.07–1.18)	1.0 × 10 ^{−6}	1.12 (1.07–1.17)	2.1 × 10 ^{−6}
External	6,922	18,896	25,818	1.09 (1.03–1.15)	0.002	—	—
ADGC + External	18,762	29,827	48,589	1.11 (1.07–1.15)	8.6 × 10 ^{−9}	—	—
EPHA1: rs11767557							
ADGC	11,840	10,931	22,771	0.87 (0.83–0.92)	2.4 × 10 ^{−7}	0.87 (0.83–0.91)	4.9 × 10 ^{−8}
External	6,922	24,666	31,588	0.91 (0.87–0.96)	2.9 × 10 ^{−4}	—	—
ADGC + External	18,762	35,597	54,359	0.90 (0.86–0.93)	6.0 × 10 ^{−10}	—	—
ARID5B: rs2588969							
ADGC	11,840	10,931	22,771	0.93 (0.89–0.97)	0.001	0.93 (0.89–0.97)	7.8 × 10 ^{−4}
External	6,922	18,896	25,818	1.06 (1.01–1.11)	0.018	—	—
ADGC + External	18,762	29,827	48,589	0.99 (0.95–1.02)	0.362	—	—
MS4A4A: rs4938933							
ADGC	11,840	10,931	22,771	0.88 (0.85–0.92)	1.7 × 10 ^{−9}	0.88 (0.85–0.92)	1.7 × 10 ^{−9}
External	6,922	18,896	25,818	0.92 (0.88–0.97)	5.4 × 10 ^{−4}	—	—
ADGC + External	18,762	29,827	48,589	0.89 (0.87–0.92)	8.2 × 10 ^{−12}	—	—
CD33: rs3865444							
ADGC	11,840	10,931	22,771	0.89 (0.86–0.93)	1.1 × 10 ^{−7}	0.89 (0.86–0.93)	2.0 × 10 ^{−7}
External	6,922	18,896	25,818	0.92 (0.88–0.97)	0.002	—	—
ADGC + External	18,762	29,827	48,589	0.91 (0.88–0.93)	1.6 × 10 ^{−9}	—	—

Meta-analysis using an external replication case-control sample (stage 3) for SNPs from previously unidentified loci at which associations did not exceed the genome-wide statistical significance threshold ($P = 5.0 \times 10^{-8}$) in the ADGC meta-analysis (stage 1+2). Results for *MS4A* are also included to show association results from the ADGC and accompanying manuscript¹². The external replication dataset is described in the accompanying paper¹² and includes the stage 1 discovery sample and the CHARGE sample² but does not include results from the TGEN, ADNI and MAYO cohorts (Supplementary Tables 1 and 2).

We also examined SNPs with statistically significant GWAS results reported by others (*GAB2* (ref. 16), *PCDH11X*¹⁷, *GOLM1* (ref. 18) and *MTHFD1L*¹⁹; Supplementary Table 7). Stage 1 data were used, except for *PCDH11X*, for which stage 1+2 data were used because Affymetrix platforms do not contain the appropriate SNP. Only SNPs in the *APOE*, *CR1*, *PICALM* and *BIN1* loci had $P < 10^{-6}$. For *MTHFD1L*¹⁹, we obtained **modest independent association evidence** at rs11754661 (previously reported $P = 4.7 \times 10^{-8}$; this study $OR_M = 1.16$, 95% CI 1.04–1.29, $P_M = 0.006$, $OR_J = 1.19$, 95% CI 1.08–1.32, $P_J = 7.5 \times 10^{-4}$). For the remaining sites, we obtained only nominal evidence ($P < 0.05$) or no evidence of association. For the *GAB2* locus¹⁶ at rs10793294 (previously reported $P = 1.60 \times 10^{-7}$), we obtained **nominal statistically significant** results ($P_M = 0.017$, $P_J = 0.029$). The association for rs5984894 in the

PCDH11X locus¹⁷ (previously reported $P = 3.9 \times 10^{-12}$) **did not replicate** ($P_M = 0.89$, $P_J = 0.26$). Likewise, findings at *GOLM1* (ref. 18) for rs10868366 (previously reported $P = 2.40 \times 10^{-4}$) did not replicate ($P_M = 0.71$, $P_J = 0.62$). Another gene consistently implicated in LOAD is *SORL1* (ref. 20), where at rs3781835 (previously reported $P = 0.006$), we obtained modest evidence for association ($OR_M = 0.72$, 95% CI 0.60–0.86, $P_M = 2.9 \times 10^{-4}$, $OR_J = 0.78$, 95% CI 0.59–0.86, $P_J = 3.8 \times 10^{-4}$).

We examined the influence of the *APOE* $\epsilon 4$ allele on the loci in Table 1 stratified by and in interactions with *APOE* $\epsilon 4$ allele carrier status. After adjustment, all loci had similar effect sizes as the unadjusted analyses, with some loci showing a modest reduction in statistical significance. We previously reported evidence for a *PICALM*-*APOE* (ref. 21) interaction using a dataset that largely overlaps with the stage 1

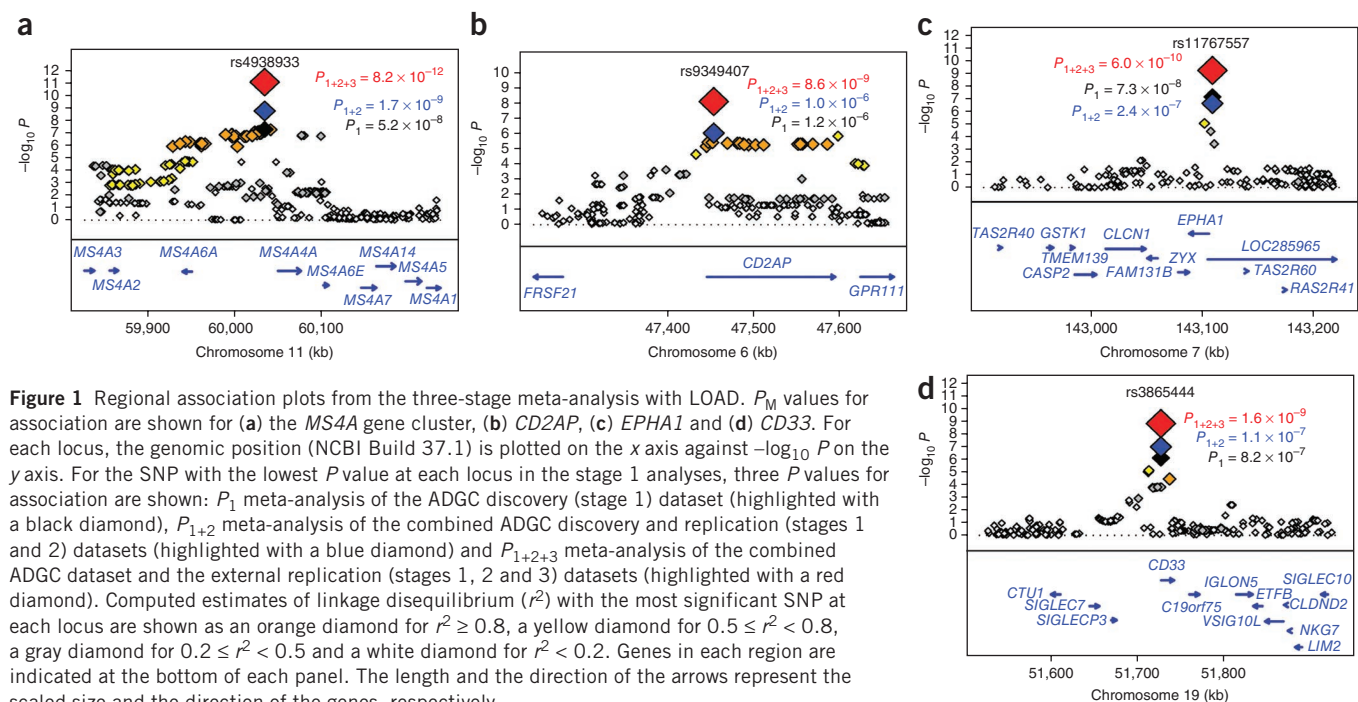


Figure 1 Regional association plots from the three-stage meta-analysis with LOAD. P_M values for association are shown for (a) the *MS4A* gene cluster, (b) *CD2AP*, (c) *EPHA1* and (d) *CD33*. For each locus, the genomic position (NCBI Build 37.1) is plotted on the x axis against $-\log_{10} P$ on the y axis. For the SNP with the lowest P value at each locus in the stage 1 analyses, three P values for association are shown: P_1 meta-analysis of the ADGC discovery (stage 1) dataset (highlighted with a black diamond), P_{1+2} meta-analysis of the combined ADGC discovery and replication (stages 1 and 2) datasets (highlighted with a blue diamond) and P_{1+2+3} meta-analysis of the combined ADGC dataset and the external replication (stages 1, 2 and 3) datasets (highlighted with a red diamond). Computed estimates of linkage disequilibrium (r^2) with the most significant SNP at each locus are shown as an orange diamond for $r^2 \geq 0.8$, a yellow diamond for $0.5 \leq r^2 < 0.8$, a gray diamond for $0.2 \leq r^2 < 0.5$ and a white diamond for $r^2 < 0.2$. Genes in each region are indicated at the bottom of each panel. The length and the direction of the arrows represent the scaled size and the direction of the genes, respectively.

dataset used here. However, using the stage 1+2 data, we did not replicate this finding or see evidence of *APOE* genotype interactions with the loci listed in **Table 1** (data not shown).

Previous work reported an association between *LOAD* and the chromosome 19 SNP rs597668, which is located 7.2 kb proximal to *EXOC3L2* and 296 kb distal of *APOE*². Although we did observe a signal for this SNP (stage 1, $P_M = 1.5 \times 10^{-9}$, $P_I = 7.7 \times 10^{-10}$) and other SNPs in the *EXOC2L3-MARK4* region, the evidence was completely extinguished for all SNPs after adjustment for *APOE* (Online Methods and **Supplementary Table 8**), suggesting that signal in this region is from *APOE*.

Our observation of genome-wide significant associations at *MS4A4A*, *CD2AP*, *EPHA1* and *CD33* extends our understanding of the genetic architecture of *LOAD* and confirms the emerging consensus that common genetic variation plays an important role in the etiology of *LOAD*. With our findings and those in the companion study¹², there are now ten *LOAD* susceptibility loci (in *APOE*, *CR1*, *CLU*, *PICALM*, *BIN1*, *EPHA1*, *MS4A*, *CD33*, *CD2AP* and *ABCA7*). Examining the amount of genetic effect attributable to these candidate genes, the most strongly associated SNPs at each locus other than that in *APOE* had population attributable fractions between 2.72% and 5.97% (**Supplementary Table 9**), with a cumulative population-attributable fraction for non-*APOE* loci estimated to be as much as 35%; however, these estimates may vary widely between studies²², and the actual effect sizes are likely to be much smaller than those estimated here because of the 'winner's curse'. Also, the results do not account for interaction among loci and are not derived from appropriate population-based samples.

A recent review of GWAS²³ noted that risk alleles with small effect sizes ($0.80 < OR < 1.2$) likely exist for complex diseases such as *LOAD* but remain undetected, even with thousands of samples, because of insufficient power²⁴. Our discovery dataset (stage 1, 8,309 cases and 7,366 controls) was well powered to detect associations exceeding the statistical significance threshold of $P < 10^{-6}$ (**Supplementary Table 9**). If there are many loci of more modest effects, some, but not all, will likely be detected in any one study. This likely explains the genome-wide statistical significance for the *ABCA7* locus in the accompanying manuscript¹², which reached only modest statistical significance in our dataset (rs3752246; $P_M = 1.0 \times 10^{-5}$, $P_I = 1.9 \times 10^{-5}$). Finding additional *LOAD* loci will require larger studies with increased depth of genotyping to test for the effects of both common and rare variants.

URLs. The Alzheimer Disease Genetics Consortium (ADGC), <http://alois.med.upenn.edu/adgc/about/overview.html>; ADNI database, <http://www.loni.ucla.edu/ADNI>; ADNI investigators, http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf; *APOE* Genotyping kit from TIB MOLBIOL, http://www.roche-as.es/logs/LightMix%C2%AE_40-0445-16_ApoE-112-158_V080904.pdf; PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink/>; PREST, <http://utstat.toronto.edu/sun/Software/Prest/>; MACH, <http://www.sph.umich.edu/csg/abecasis/mach/>; EIGENSTRAT, <http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm>; The R Project for Statistical Computing, <http://www.r-project.org/>; Package *GWAF* in R, <http://cran.r-project.org/web/packages/GWAF/index.html>; Package *gee* in R, <http://cran.r-project.org/web/packages/gee/index.html>; UCSC Genome Browser, <http://genome.ucsc.edu/>; METAL, <http://www.sph.umich.edu/csg/abecasis/Metal/>; FUGUE, <http://www.sph.umich.edu/csg/abecasis/fugue/>.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The National Institutes of Health, National Institute on Aging (NIH-NIA) supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; NACC, U01 AG016976; NCRAD, U24 AG021886; NIA LOAD, U24 AG026395, U24 AG026390; Boston University, P30 AG013846, R01 HG02213, K24 AG027841, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG009029, R01 AG017173, R01 AG025259; Columbia University, P50 AG008702, R37 AG015473; Duke University, P30 AG028377; Emory University, AG025688; Indiana University, P30 AG10133; Johns Hopkins University, P50 AG005146, R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, MO1RR00096, and UL1 RR029893; Northwestern University, P30 AG013854; Oregon Health & Science University, P30 AG008017, R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, R01 AG30146; University of Alabama at Birmingham, P50 AG016582, UL1RR02777; University of Arizona/TGEN, P30 AG019610, R01 AG031581, R01 NS059873; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573, P50, P50 AG016575, P50 AG016576, P50 AG016577; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501, P01 AG019724; University of Kentucky, P30 AG028383; University of Michigan, P50 AG008671; University of Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, AG019757; University of Washington, P50 AG005136, U01 AG06781, U01 HG004610; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681, P01 AG03991. ADNI Funding for ADNI is through the Northern California Institute for Research and Education by grants from Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc., F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., Alzheimer's Association, Alzheimer's Drug Discovery Foundation, the Dana Foundation, and by the National Institute of Biomedical Imaging and Bioengineering and NIA grants U01 AG024904, RC2 AG036535, K01 AG030514. We thank C. Phelps, M. Morrison-Bogorad and M. Miller from NIA who are *ex-officio* ADGC members. Support was also from the Alzheimer's Association (LAF, IIRG-08-89720; MP-V, IIRG-05-14147) and the Veterans Affairs Administration. P.S.-H. is supported by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health Research.

AUTHOR CONTRIBUTIONS

Sample collection, phenotyping and data management: J.D. Buxbaum, G.P.J., P.K.C., E.B.L., T.D.B., B.F.B., N.R.G.-R., P.L.D., D.E., J.A. Schneider, M.M.C., N.E.-T., S.G.Y., C.C., J.S.K.K., P.N., P.K., J.H., M.J.H., A.J.M., M.M.B., F.Y.D., C.T.B., R.C.G., E.R., P.S.G.-H., S.E.A., R.B., T.B., E.H.B., J.D. Bowen, A.B., J.R.B., N.J.C., C.S.C., S.L.C., H.C.C., D.G.C., J.C., C.W.C., J.L.C., C.D., S.T.D., R.D.-A., M.D., D.W.D., W.G.E., K.M.F., K.B.F., M.R.F., S.F., M.P.F., D.R.G., M. Ganguli, M. Gearing, D.H.G., B. Ghetti, J.R.G., S.G., B. Giordani, J.D.G., J.H.G., R.L.H., L.E.H., E.H., L.S.H., C.M.H., B.T.H., G.A.J., L.-W.J., N.J., J.K., A.K., J.A.K., R.K., E.H.K., N.W.K., J.J.L., A.I.L., A.P.L., O.L.L., W.J.M., D.C. Marson, F.M., D.C. Mash, E.M., W.C.M., S.M.M., A.N.M., A.C.M., M.M., B.L.M., C.A.M., J.W.M., J.E.P., D.P.P., E.P., R.C.P., W.W.P., J.F.Q., M.R., B.R., J.M.R., E.D.R., R.N.R., M.S., L.S.S., W.S., M.L.S., M.A.S., C.D.S., J.A. Sonnen, S.S., R.A.S., R.E.T., J.Q.T., J.C.T., V.M.V., H.V.V., J.P.V., S.W., K.A.W.-B., J.W., R.L.W., L.B.C., B.A.D., D. Beekly, M.I.K., A.J.S., E.M.R., D.A.B., A.M.G., W.A.K., T.M.F., J.L.H., R.M., M.A.P.-V., L.A.F.

Study management and coordination: L.B.C., D. Beekly, D.A.B., J.C.M., T.J.M., A.M.G., D. Blacker, D.W.T., H.H., W.A.K., T.M.F., J.L.H., R.M., M.A.P.-V., L.A.F., G.D.S.

Statistical methods and analysis: A.C.N., G.J., G.W.B., L.-S.W., B.N.V., J.B., P.J.G., R.M.C., R.A.R., M.A.S., K.L.L., E.R.M., J.L.H., M.A.P.-V., L.A.F.

Interpretation of results: A.C.N., G.J., G.W.B., L.-S.W., B.N.V., J.B., P.J.G., R.A.R., M.A.S., K.L.L., E.R.M., M.I.K., A.J.S., E.M.R., D.A.B., J.C.M., T.J.M., A.M.G., D. Blacker, D.W.T., H.H., W.A.K., T.M.F., J.L.H., R.M., M.A.P.-V., L.A.F., G.D.S.

Manuscript writing group: A.C.N., G.J., G.W.B., L.-S.W., B.N.V., J.B., P.J.G., J.L.H., R.M., M.A.P.-V., L.A.F., G.D.S.

Study design: D.A.B., J.C.M., T.J.M., A.M.G., D. Blacker, D.W.T., H.H., W.A.K., T.M.F., J.L.H., R.M., M.A.P.-V., L.A.F., G.D.S.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Harold, D. *et al.* Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1088–1093 (2009).
- Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *J. Am. Med. Assoc.* **303**, 1832–1840 (2010).
- Lambert, J.C. *et al.* Genome-wide association study identifies variants at *CLU* and *CRI* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1094–1099 (2009).
- Hebert, L.E., Scherr, P.A., Bienias, J.L., Bennett, D.A. & Evans, D.A. Alzheimer disease in the US population—prevalence estimates using the 2000 census. *Arch. Neurol.* **60**, 1119–1122 (2003).
- Alzheimer's Association. Alzheimer's disease facts and figures. *Alzheimer's Dement.* **5**, 234–270 (2009).
- Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704–706 (1991).
- Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754–760 (1995).
- Rogaev, E.I. *et al.* Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376**, 775–778 (1995).
- Levy-Lahad, E. *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973–977 (1995).
- Corder, E.H. *et al.* Gene dose of apolipoprotein-E type-4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).
- Gatz, M. *et al.* Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J. Gerontol. A Biol. Sci. Med. Sci.* **52**, M117–M125 (1997).
- Hollingsworth, P. *et al.* Common variants at *ABCA7*, *MS4A6A/MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease. *Nat. Genet.* advance online publication, doi:10.1038/ng.803 (3 April 2011).
- Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
- Ioannidis, J.P., Rosenberg, P.S., Goedert, J.J. & O'Brien, T.R. Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am. J. Epidemiol.* **156**, 204–210 (2002).
- Bertram, L. *et al.* Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to *APOE*. *Am. J. Hum. Genet.* **83**, 623–632 (2008).
- Reiman, E.M. *et al.* *GAB2* alleles modify Alzheimer's risk in *APOE* ϵ 4 carriers. *Neuron* **54**, 713–720 (2007).
- Carrasquillo, M.M. *et al.* Genetic variation in *PCDH11X* is associated with susceptibility to late-onset Alzheimer's disease. *Nat. Genet.* **41**, 192–198 (2009).
- Li, H. *et al.* Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch. Neurol.* **65**, 45–53 (2008).
- Naj, A.C. *et al.* Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. *PLoS Genet.* **6**, e1001130 (2010).
- Rogaeva, E. *et al.* The neuronal sortilin-related receptor *SORL1* is genetically associated with Alzheimer disease. *Nat. Genet.* **39**, 168–177 (2007).
- Jun, G. *et al.* Meta-analysis confirms *CRI*, *CLU*, and *PICALM* as Alzheimer disease risk loci and reveals interactions with *APOE* genotypes. *Arch. Neurol.* **67**, 1473–1484 (2010).
- Rockhill, B., Newman, B. & Weinberg, C. Use and misuse of population attributable fractions. *Am. J. Public Health* **88**, 15–19 (1998).
- Ku, C.S., Loy, E.Y., Pawitan, Y. & Chia, K.S. The pursuit of genome-wide association studies: where are we now? *J. Hum. Genet.* **55**, 195–206 (2010).
- Florez, J.C. Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. *J. Clin. Endocrinol. Metab.* **93**, 4633–4642 (2008).

Adam C Naj^{1,115}, Gyungah Jun^{2-4,115}, Gary W Beecham^{1,5}, Li-San Wang⁶, Badri Narayan Vardarajan³, Jacqueline Buross³, Paul J Gallins¹, Joseph D Buxbaum⁷⁻⁹, Gail P Jarvik^{10,11}, Paul K Crane¹², Eric B Larson¹³, Thomas D Bird¹⁴, Bradley F Boeve¹⁵, Neill R Graff-Radford^{16,17}, Philip L De Jager^{18,19}, Denis Evans²⁰, Julie A Schneider^{21,22}, Minerva M Carrasquillo¹⁶, Nilufer Ertekin-Taner^{16,17}, Steven G Younkin¹⁶, Carlos Cruchaga²³, John S K Kauwe²⁴, Petra Nowotny²³, Patricia Kramer^{25,26}, John Hardy²⁷, Matthew J Huentelman²⁸, Amanda J Myers²⁹, Michael M Barmada³⁰, F Yesim Demirci³⁰, Clinton T Baldwin³, Robert C Green^{3,31,32}, Ekaterina Rogaeva³³, Peter St George-Hyslop^{33,34}, Steven E Arnold³⁵, Robert Barber³⁶, Thomas Beach³⁷, Eileen H Bigio³⁸, James D Bowen³⁹, Adam Boxer⁴⁰, James R Burke⁴¹, Nigel J Cairns⁴², Chris S Carlson⁴³, Regina M Carney⁴⁴, Steven L Carroll⁴⁵, Helena C Chui⁴⁶, David G Clark⁴⁷, Jason Corneveaux²⁸, Carl W Cotman⁴⁸, Jeffrey L Cummings⁴⁹, Charles DeCarli⁵⁰, Steven T DeKosky⁵¹, Ramon Diaz-Arrastia⁵², Malcolm Dick⁴⁸, Dennis W Dickson¹⁶, William G Ellis⁵³, Kelley M Faber⁵⁴, Kenneth B Fallon⁴⁵, Martin R Farlow⁵⁵, Steven Ferris⁵⁶, Matthew P Frosch⁵⁷, Douglas R Galasko⁵⁸, Mary Ganguli⁵⁹, Marla Gearing^{60,61}, Daniel H Geschwind⁶², Bernardino Ghetti⁶³, John R Gilbert^{1,5}, Sid Gilman⁶⁴, Bruno Giordani⁶⁵, Jonathan D Glass⁶⁶, John H Growdon⁶⁷, Ronald L Hamilton⁶⁸, Lindy E Harrell⁴⁷, Elizabeth Head⁶⁹, Lawrence S Honig⁷⁰, Christine M Hulette⁷¹, Bradley T Hyman⁶⁷, Gregory A Jicha⁷², Lee-Way Jin⁵³, Nancy Johnson⁷³, Jason Karlawish⁷⁴, Anna Karydas⁴⁰, Jeffrey A Kaye^{26,75}, Ronald Kim⁷⁶, Edward H Koo⁵⁸, Neil W Kowall^{31,77}, James J Lah⁶⁶, Allan I Levey⁶⁶, Andrew P Lieberman⁷⁸, Oscar L Lopez⁷⁹, Wendy J Mack⁸⁰, Daniel C Marson⁴⁷, Frank Martiniuk⁸¹, Deborah C Mash⁸², Eliezer Masliah^{58,83}, Wayne C McCormick¹², Susan M McCurry⁸⁴, Andrew N McDavid⁴³, Ann C McKee^{31,77}, Marsel Mesulam^{85,86}, Bruce L Miller⁴⁰, Carol A Miller⁸⁷, Joshua W Miller⁵³, Joseph E Parisi^{88,89}, Daniel P Perl⁹⁰, Elaine Peskind⁹¹, Ronald C Petersen¹⁵, Wayne W Poon⁴⁸, Joseph F Quinn²⁶, Ruchita A Rajbhandary¹, Murray Raskind⁹¹, Barry Reisberg^{56,92}, John M Ringman⁴⁹, Erik D Roberson⁴⁷, Roger N Rosenberg⁵², Mary Sano⁸, Lon S Schneider^{46,93}, William Seeley⁴⁰, Michael L Shelanski⁹⁴, Michael A Slifer^{1,5}, Charles D Smith⁷², Joshua A Sonnen⁹⁵, Salvatore Spina⁶³, Robert A Stern³¹, Rudolph E Tanzi⁶⁷, John Q Trojanowski⁶, Juan C Troncoso⁹⁶, Vivianna M Van Deerlin⁶, Harry V Vinters^{49,97}, Jean Paul Vonsattel⁹⁸, Sandra Weintraub^{85,86}, Kathleen A Welsh-Bohmer^{41,99}, Jennifer Williamson⁷⁰, Randall L Woltjer¹⁰⁰, Laura B Cantwell⁶, Beth A Dombroski⁶, Duane Beekly¹⁰¹, Kathryn L Lunetta², Eden R Martin^{1,5}, M Ilyas Kamboh^{30,79}, Andrew J Saykin^{54,102}, Eric M Reiman^{28,103-105}, David A Bennett^{22,106}, John C Morris^{42,107}, Thomas J Montine⁹⁵, Alison M Goate²³, Deborah Blacker^{108,109}, Debby W Tsuang⁹¹, Hakon Hakonarson¹¹⁰, Walter A Kukull¹¹¹, Tatiana M Foroud⁵⁴, Jonathan L Haines^{112,113}, Richard Mayeux^{70,114}, Margaret A Pericak-Vance^{1,5}, Lindsay A Farrer^{2-4,31,32} & Gerard D Schellenberg⁶

¹The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, USA. ²Department of Biostatistics, Boston University, Boston, Massachusetts, USA. ³Department of Medicine (Genetics Program), Boston University, Boston, Massachusetts, USA. ⁴Department of Ophthalmology, Boston University, Boston, Massachusetts, USA. ⁵Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida, USA. ⁶Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. ⁷Department of Neuroscience, Mount Sinai School of Medicine, New York, New York, USA. ⁸Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, USA. ⁹Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA. ¹⁰Department of Genome Sciences, University of Washington, Seattle, Washington, USA. ¹¹Department of Medicine (Medical Genetics), University of Washington, Seattle, Washington, USA. ¹²Department of Medicine, University of Washington, Seattle, Washington, USA. ¹³Group Health Research Institute, Seattle, Washington, USA. ¹⁴Department of Neurology, University of Washington, Seattle, Washington, USA. ¹⁵Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA. ¹⁶Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, USA. ¹⁷Department of Neurology, Mayo Clinic, Jacksonville, Florida, USA. ¹⁸Program in Translational Neuropsychiatric Genomics, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, USA. ¹⁹Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. ²⁰Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, USA. ²¹Department of Pathology (Neuropathology), Rush University Medical Center, Chicago, Illinois, USA. ²²Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA. ²³Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, Missouri, USA. ²⁴Department of Biology, Brigham Young University, Provo, Utah, USA. ²⁵Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, Oregon, USA. ²⁶Department of Neurology, Oregon Health and Science University, Portland, Oregon, USA. ²⁷Institute of Neurology, University College London, Queen Square, London, UK. ²⁸Neurogenetics Division, Translational Genomics Research Institute, Phoenix, Arizona, USA. ²⁹Department of Psychiatry and Behavioral Sciences, University of Miami, Miami, Florida, USA. ³⁰Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ³¹Department of Neurology, Boston University, Boston, Massachusetts, USA. ³²Department of Epidemiology, Boston University, Boston, Massachusetts, USA. ³³Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, Ontario, Canada. ³⁴Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, Massachusetts, UK. ³⁵Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. ³⁶Department of Pharmacology and Neuroscience, University of Texas Southwestern, Fort Worth, Texas, USA. ³⁷Civivi Laboratory for Neuropathology, Banner Sun Health Research Institute, Phoenix, Arizona, USA. ³⁸Department of Pathology, Northwestern University, Chicago, Illinois, USA. ³⁹Swedish Medical Center, Seattle, Washington, USA. ⁴⁰Department of Neurology, University of California San Francisco, San Francisco, California, USA. ⁴¹Department of Medicine, Duke University, Durham, North Carolina, USA. ⁴²Department of Pathology and Immunology, Washington University, St. Louis, Missouri, USA. ⁴³Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ⁴⁴Department of Psychiatry, Vanderbilt University, Nashville, Tennessee, USA. ⁴⁵Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama, USA. ⁴⁶Department of Neurology, University of Southern California, Los Angeles, California, USA. ⁴⁷Department of Neurology, University of Alabama at Birmingham, Birmingham, Alabama, USA. ⁴⁸Institute for Memory Impairments and Neurological Disorders, University of California Irvine, Irvine, California, USA. ⁴⁹Department of Neurology, University of California Los Angeles, Los Angeles, California, USA. ⁵⁰Department of Neurology, University of California Davis, Sacramento, California, USA. ⁵¹University of Virginia School of Medicine, Charlottesville, Virginia, USA. ⁵²Department of Neurology, University of Texas Southwestern, Dallas, Texas, USA. ⁵³Department of Pathology and Laboratory Medicine, University of California Davis, Sacramento, California, USA. ⁵⁴Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, USA. ⁵⁵Department of Neurology, Indiana University, Indianapolis, Indiana, USA. ⁵⁶Department of Psychiatry, New York University, New York, New York, USA. ⁵⁷C.S. Kubik Laboratory for Neuropathology, Massachusetts General Hospital, Charlestown, Massachusetts, USA. ⁵⁸Department of Neurosciences, University of California San Diego, La Jolla, California, USA. ⁵⁹Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ⁶⁰Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, USA. ⁶¹Emory Alzheimer's Disease Center, Emory University, Atlanta, Georgia, USA. ⁶²Neurogenetics Program, University of California Los Angeles, Los Angeles, California, USA. ⁶³Department of Pathology and Laboratory Medicine, Indiana University, Indianapolis, Indiana, USA. ⁶⁴Department of Neurology, University of Michigan, Ann Arbor, Michigan, USA. ⁶⁵Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA. ⁶⁶Department of Neurology, Emory University, Atlanta, Georgia, USA. ⁶⁷Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, USA. ⁶⁸Department of Pathology (Neuropathology), University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ⁶⁹Department of Molecular and Biomedical Pharmacology, University of California Irvine, Irvine, California, USA. ⁷⁰Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, New York, USA. ⁷¹Department of Pathology, Duke University, Durham, North Carolina, USA. ⁷²Department of Neurology, University of Kentucky, Lexington, Kentucky, USA. ⁷³Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, Illinois, USA. ⁷⁴Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. ⁷⁵Department of Biomedical Engineering, Oregon Health and Science University, Portland, Oregon, USA. ⁷⁶Department of Pathology and Laboratory Medicine, University of California Irvine, Irvine, California, USA. ⁷⁷Department of Pathology, Boston University, Boston, Massachusetts, USA. ⁷⁸Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA. ⁷⁹University of Pittsburgh Alzheimer's Disease Research Center, Pittsburgh, Pennsylvania, USA. ⁸⁰Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA. ⁸¹Department of Medicine-Pulmonary, New York University, New York, New York, USA. ⁸²Department of Neurology, University of Miami, Miami, Florida, USA. ⁸³Department of Pathology, University of California San Diego, La Jolla, California, USA. ⁸⁴School of Nursing Northwest Research Group on Aging, University of Washington, Seattle, Washington, USA. ⁸⁵Alzheimer's Disease Center, Northwestern University, Chicago, Illinois, USA. ⁸⁶Cognitive Neurology, Northwestern University, Chicago, Illinois, USA. ⁸⁷Department of Pathology, University of Southern California, Los Angeles, California, USA. ⁸⁸Department of Anatomic Pathology, Mayo Clinic, Rochester, Minnesota, USA. ⁸⁹Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. ⁹⁰Department of Pathology, Mount Sinai School of Medicine, New York, New York, USA. ⁹¹Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA. ⁹²Alzheimer's Disease Center, New York University, New York, New York, USA. ⁹³Department of Psychiatry, University of Southern California, Los Angeles, California, USA. ⁹⁴Department of Pathology, Columbia University, New York, New York, USA. ⁹⁵Department of Pathology, University of Washington, Seattle, Washington, USA. ⁹⁶Department of Pathology, Johns Hopkins University, Baltimore, Maryland, USA. ⁹⁷Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, California, USA. ⁹⁸Taub Institute Department of Pathology, Columbia University, New York, New York, USA. ⁹⁹Department of Psychiatry and Behavioral Sciences, Duke University, Durham, North Carolina, USA. ¹⁰⁰Department of Pathology, Oregon Health and Science University, Portland, Oregon, USA. ¹⁰¹National Alzheimer's Coordinating Center, University of Washington, Seattle, Washington, USA. ¹⁰²Department of Radiology and Imaging Sciences, Indiana University, Indianapolis, Indiana, USA. ¹⁰³Department of Psychiatry, University of Arizona, Phoenix, Arizona, USA. ¹⁰⁴Arizona Alzheimer's Consortium, Phoenix, Arizona, USA. ¹⁰⁵Banner Alzheimer's Institute, Phoenix, Arizona, USA. ¹⁰⁶Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA. ¹⁰⁷Department of Neurology, Washington University, St. Louis, Missouri, USA. ¹⁰⁸Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. ¹⁰⁹Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, USA. ¹¹⁰Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ¹¹¹Department of Epidemiology, University of Washington, Seattle, Washington, USA. ¹¹²Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA. ¹¹³Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, USA. ¹¹⁴Gertrude H. Sergievsky Center, Columbia University, New York, New York, USA. ¹¹⁵These authors contributed equally to this work. Correspondence should be addressed to G.D.S. (gerardsc@mail.med.upenn.edu).

ONLINE METHODS

Subjects. A full description of study cohorts is provided in the **Supplementary Note** and **Supplementary Tables 1** and **2**.

Covariate data. Age of onset data was available from some cohorts (Alzheimer's Disease Center (ADC), Translational Genomics Research Institute series 2 (TGEN2), National Institute on Aging Late-onset AD (NIA-LOAD), Multi-Institutional Research on Alzheimer's Genetic Epidemiology (MIRAGE), Adult Changes in Thought (ACT), Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GenADA), University of Pittsburgh (UP) and the Rush University Religious Orders Study/Memory and Aging Project (ROS/MAP)), whereas for others, only age at ascertainment (Washington University (WU) and ADNI), age at diagnosis (Mayo Clinic (MAYO)), or a combination of both age at ascertainment and age at death was available (a subset of autopsy-confirmed samples in the University of Miami/Vanderbilt University/Mt. Sinai School of Medicine (UM/VU/MSSM) cohort). For subjects with autopsy-confirmed diagnosis and no clinical diagnosis, the age at diagnosis was equated to the age at death. For all studies, the age used for CNEs was the age of last exam or age at death. Case and CNE subjects with age at symptom onset or age at death less than 60 were excluded from the analysis. We restricted our association analyses to individuals of European ancestry because there were insufficient subjects from non-European-ancestry groups to obtain meaningful results.

Genotyping, data cleaning and imputation. Genotypes were from either Illumina or Affymetrix high-density SNP microarrays (**Supplementary Table 3**). Genotype data were cleaned by applying minimum call rates (95% and 98%) and minimum minor allele frequencies (0.02 and 0.01) for cohorts genotyped on Affymetrix and Illumina chips, respectively. SNPs not in Hardy-Weinberg equilibrium ($P < 10^{-6}$) were excluded. Subjects where the gender was mis-specified were identified by analysis of X-chromosome SNPs using PLINK²⁵. For cohorts genotyped on multiple chips (MIRAGE and UM/UV/MSSM), genotype and sample quality thresholds were applied within subsets of individuals genotyped on each chip. For all other cohorts, quality thresholds were applied per cohort. Relationships among individuals in the family-based cohorts (MIRAGE and NIALOAD) were confirmed by pairwise genome-wide estimates of proportion identity-by-descent (IBD) using PREST software²⁶. Any discrepancies identified were reviewed in light of available clinical and pedigree data to determine the most likely relationship consistent with a proportion of IBD, and any remaining scenarios were excluded from analysis. Latent relatedness in the case-control cohorts was identified by proportion IBD using PLINK software^{25,27}. Both of each pair of identical samples by IBD ($\hat{\pi} > 0.99$) were dropped, and one subject was selected from each related pair ($0.4 \geq \hat{\pi} > 0.90$), prioritizing non-missing case or non-missing control status and then higher call rate in selection. Duplicate enrollments among studies (**Supplementary Table 4**) were identified using proportion of IBD in a genotyped dataset including all cohorts where pairs with $\hat{\pi} > 0.95$ were considered duplicate enrollments. Duplicates with discordant case or control status by study were dropped from both studies, and those with concordant status were included in only one cohort and selected according to a predetermined priority list of cohorts which considered genotype data, phenotype data and the type of cohort. Genome-wide imputation was performed per cohort using MACH software²⁸ with HapMap phase 2 (release 22) CEPH Utah pedigree (CEU) reference haplotypes and genotype data passing quality control as inference. Imputation quality was determined as R^2 and only SNPs imputed with $R^2 \geq 0.50$ were included in the analysis.

APOE genotyping. APOE genotypes were determined for the ADC, ACT, NIA-LOAD, UM/VU/MSSM, MAYO and GenADA cohorts using SNPs rs7412 and rs429358; for the MIRAGE cohort using the Roche Diagnostics LightCycler 480 instrument (Roche Diagnostics)²⁹ LightMix Kit ApoE C112R R158 (TIB MOLBIOL); for TGEN2, ADNI, UP and WU cohorts by pyrosequencing³⁰ or restriction fragment length polymorphism analysis^{31,32}; and for ROSMAP by high-throughput sequencing of codons 112 and 158 in APOE by Agencourt Bioscience Corporation.

Meta-analysis. Presence of intra-study population substructure was evaluated separately by cohort in a two-step process that first removed outliers before estimating population substructure within the remaining population. For the first step, either the STRUCTURE software package^{33,34} (UM/VU/MSSM and MIRAGE) or the 'smartpca' script in EIGENSTRAT³⁵ (remaining cohorts) was used to remove outliers and/or confirm self-reported ethnicity after filtering to remove SNPs in pairwise LD. In the second step, we used EIGENSTRAT³⁵, often a second time, to estimate principal component loadings for inclusion in association analysis. For each study, the first two, three or four estimated principal components were identified for inclusion as covariates in association analysis (**Supplementary Table 3**). Outlier detection for the ADC, TGEN2, GenADA, ACT, ADNI, ROS/MAP, OHSU, UP, WU and MAYO cohorts was evaluated by comparison to the HapMap 3 CEPH (CEU) population. EIGENSTRAT analyses of family cohort data (NIA-LOAD and MIRAGE) used a sample of unrelated individuals to fit principal components after outliers with respect to European-American ancestry were removed.

Genotyped and imputed SNP data passing quality control were tested for association with Alzheimer's disease in each dataset using logistic generalized linear model (GLM) for case-control analysis and logistic generalized estimating equations (GEE) for family-based cohorts in R³⁶⁻³⁸. All analyses assumed an additive genetic model, coding genotyped SNPs by the number of minor alleles (0, 1 or 2) and imputed SNPs by the posterior probability of the minor allele (range 0–2). Primary association analyses were adjusted for population substructure (baseline model).

SNP association results for each dataset were meta-analyzed using the inverse variance method implemented in the software package METAL³⁹. The meta-analysis P value was estimated by the summarized test statistic after applying a genomic control within each individual study. Heterogeneity among odds ratios in the meta-analysis was assessed using the Cochran's Q and I^2 statistics^{40,41}.

Regional association plots were prepared for the most strongly associated SNPs in *CRI*, *BIN1*, *CD2AP*, *EPHA1*, *CLU*, *MS4A4A/MS4A6A*, *PICALM*, *ABCA7* and *CD33* using the gene locations from UCSC Genome browser (hg19, GRCh37, Feb 2009 release) and SNP locations from the corresponding dbSNP build 131. Estimates of LD were calculated with the FUGUE software⁴² using HapMap phase 2 (release 24, CEU) genotype data and build 131 SNP positions. Forest plots of study-specific effects and analysis results are presented for the same set of SNPs using the 'rmeta' package in R.

Joint analysis. Testing for population substructure across studies was performed in a combined dataset using the set of SNPs genotyped in all study cohorts. After filtering SNPs with pairwise LD ($r^2 < 0.20$), 31,310 SNPs were evaluated using EIGENSTRAT. The top three principal components from EIGENSTRAT were used as covariates in the joint analysis for association in addition to an adjustment for site-specific effects using dummy variables for each cohort. SNP associations with Alzheimer's disease affection status were examined in a pooled analysis of subjects from all cohorts, excluding SNPs missing from one or more individual dataset or with genotypes available on fewer than 98% of individuals overall. In total, 2,312,972 directly genotyped or imputed SNPs common to all datasets were tested for association in 8,309 cases and 7,366 CNEs, including 3,489 individuals in family datasets using GEE analyses in R. Joint analyses of the baseline model, full model and models evaluating robustness to APOE included as covariates the principal components from inter-study and intra-study population substructure and a dummy covariate for cohort-specific effects. Genomic inflation factors for the discovery joint analysis in the basic and extended models of covariate adjustment were 1.05 and 1.04, respectively (**Supplementary Table 3**), which were similar to those from meta-analysis.

Secondary analysis. Association results in regions yielding at least one SNP with $P < 10^{-6}$ (follow-up SNPs) were further evaluated for robustness to APOE $\epsilon 4$ carrier status in analyses stratified according to presence or absence of APOE $\epsilon 4$ and an interaction analysis including effects for SNP, APOE $\epsilon 4$ and their interaction. In addition, we examined the *EXOC3L2* region in chromosome 19 previously reported as independent of APOE genotype² in a full model including covariates for age at onset or age at last exam, gender and the dosage of APOE $\epsilon 4$ alleles.

Internal and external replication analyses. SNPs attaining a $P \leq 1 \times 10^{-6}$ for association with LOAD in the discovery cohort were evaluated in five independent datasets (ADC3, OHSU, MAYO, ROS/MAP and UP) consisting of 3,531 cases and 3,565 CNEs using the same analytical approaches as described above. Replication was performed using both meta-analysis and joint analysis. The datasets included in discovery and replication analyses are summarized in **Supplementary Tables 1 and 2**. Following internal replication, an external replication cohort was sought to evaluate the most strongly associated SNP in each of four newly identified genes (*CD2AP* (rs9349407), *EPHA1* (rs11767557), *ARID5B* (rs2588969) and *CD33* (rs3865444)) for which results did not meet genome-wide significance ($P_M > 5 \times 10^{-8}$, $P_J > 5 \times 10^{-8}$) in the combined discovery and replication datasets (stage 1+2). We obtained summarized results from five independent external datasets generously provided by the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium¹, the European Alzheimer's Disease Initiative (EADI) Consortium³ and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium², and these were used for combined stage 1, 2 and 3 meta analysis. After removing subjects recognized as part of the ADGC cohorts¹², the sample included 6,922 Alzheimer's disease cases and 24,666 controls. These datasets were analyzed using meta-analysis as described above for the stage 1 and 2 datasets. Results from stages 1, 2 and 3 were likewise assessed by meta-analysis as described above.

25. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
26. McPeck, M.S. & Sun, L. Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am. J. Hum. Genet.* **66**, 1076–1094 (2000).

27. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. GRR: graphical representation of relationship errors. *Bioinformatics* **17**, 742–743 (2001).
28. Li, Y. & Abecasis, G.R. Rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.* **S79**, 2290 (2006).
29. Wittwer, C.T. *et al.* The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* **22**, 176–181 (1997).
30. Ahmadian, A. *et al.* Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280**, 103–110 (2000).
31. Hixson, J.E. & Vernier, D.T. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J. Lipid Res.* **31**, 545–548 (1990).
32. Lai, E., Riley, J., Purvis, I. & Roses, A. A 4-Mb high-density single nucleotide polymorphism-based map around human *APOE*. *Genomics* **54**, 31–38 (1998).
33. Pritchard, J.K., Stephens, M., Rosenberg, N.A. & Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170–181 (2000).
34. Pritchard, J.K., Stephens, M. & Donnelly, P.J. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
35. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
36. Carey, V.J. Ported to R by Lumley, T. (versions 3.13.4.4) & Ripley, B. (version 4.13). GEE: Generalized Estimation Equation Solver. R package version 4.13-15. <<http://CRAN.R-project.org/package=gee>> (2010).
37. Chen, M.H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580–581 (2010).
38. R Development Core Team. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria, 2009).
39. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
40. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
41. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
42. Abecasis, G.R. & Wigginton, J.E. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. *Am. J. Hum. Genet.* **77**, 754–767 (2005).