Section S1: Methods

Genotyping for the *APOE* ε2/ε3/ε4 alleles was performed in the AA_M, AA_G and CAU_M datasets using a Roche diagnostics LightCycler® 480 instrument (Roche Diagnostics, Mannheim, Germany) with the LightMix Kit ApoE C112R R158C from TIB MOLBIOL (http://www.roche-as.es/logs/LightMix%C2%AE_40-0445-16_ApoE-112-158_V080904.pdf). *APOE* genotyping in the NIA-LOAD dataset was performed by Prevention Genetics in which the genotypes were determined based on allelic combinations of SNPs rs7412 and rs429358. The same SNPs were genotyped in the GenADA dataset using the ABI 7200 Taqman system. *APOE* genotypes in the ADNI and FHS cohorts were obtained by pyrosequencing or restriction fragment length polymorphism analysis. Genome-wide SNP genotypes for the MIRAGE and AA_G subjects were obtained using the Illumina Beadarray platform and the company's standard protocols. Most subjects were genotyped using the Illumina 610 Quad chip. Approximately 25% of the MIRAGE subjects were genotyped using the Illumina 370 Duo chip. SNPs with a call rate less than 95% were removed prior to imputation. None of the reported SNPs were found to have a significant difference in allele frequencies between the two Illumina chips used for genotyping ($p>.05$). We also excluded individual samples with SNP call rates below 95%. PLINK[1] was used to identify gender-mismatches and cryptic relatedness within the dataset. All relationships were confirmed using PREST[2, 3]. Population of origin was confirmed using STRUCTURE[4]. Principal component (PC) analysis implemented in the EIGENSTRAT[5] software program was used to evaluate population substructure within each dataset. There was no evidence of p-value inflation due to population stratification ($\lambda=1.003$, Supplementary Figure 1).

Individuals deemed as outliers were excluded from the genetic association analysis. None of the top ten PCs generated by EIGENSTRAT were associated with AD status at $p<.05$ in either AA dataset after applying a correction for multiple-testing. An examination of the PCs in both AA datasets did not show any evidence of substructure. In contrast, we observed clustering within several of the Caucasian replication datasets. However, because the PCs were not associated with AD after a multiple testing correction, they were not included in the model. Genotyping for the Caucasian GWAS replication datasets obtained from dbGaP was performed using either the Illumina 610 (ADNI, NIA-LOAD) or Affymetrix 500 (GenADA, FHS) SNP microarray chips.

Imputation of autosomal SNPs was performed using the Markov Chain Haplotyping (MaCH) software[6] based on the HapMap 2 and 3 reference SNP panels (http://hapmap.ncbi.nlm.nih.gov/). Reference haplotypes from the Centre d'Etude du Polymorphisme Humain subjects (CEU) were used to impute genotypes in the Caucasian datasets and a mixture of all of the reference haplotypes from CEU and Yoruban (YRI) subjects were used to impute the genotypes for the AA datasets. When a SNP was present in both HapMap reference panels, the imputation results with the highest R^2 were utilized.

Imputed SNPs were tested for association with AD in the family-based datasets (AA_M, CAU_M, FHS, NIA-LOAD, and GenerAAtions study due to the presence of some half sibling pairs) using generalized estimating equations (GEE)[7, 8] as implemented in the R statistical software package[9] in order to account for non-independence of family members. Analysis of the case-control datasets (ADNI, GenADA) was performed using logistic regression models as implemented in PLINK[1]. A quantitative estimate between 0

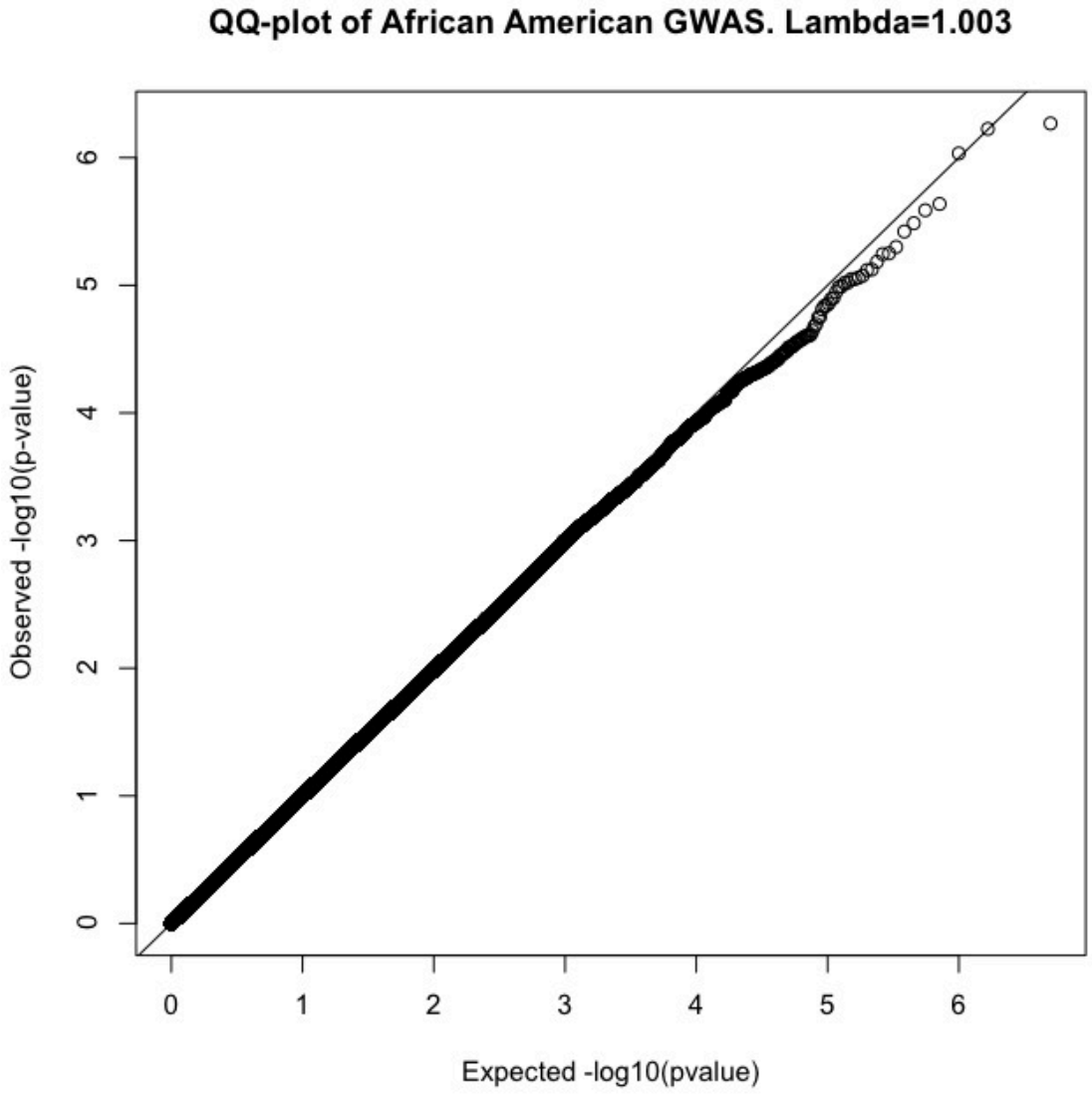
and 2 representing the dose of the minor allele was used in the analysis. All tests of association were adjusted for age (age at exam for cases and controls) and gender. In order to ensure mathematical stability of the GEE estimates only SNPs with minor allele frequencies > 2% and imputation-R^2 values >.5 were analyzed . When the previously reported SNP was within a gene (i.e., *CLU* or *CR1*), we examined SNPs in the gene and its regulatory regions (10 kb upstream and downstream). When the previously reported SNP was outside of the implicated gene (i.e. *PICALM or BIN1*), we examined all SNPs within 100 kb of the reported SNP and all SNPs within the reported candidate gene. The *APOE* region was defined as the chromosome 19 interval between 49.9 Mb and 50.3 Mb according to Map build NCBI36/HG18.

Section S2: Replication of GWAS Results in Caucasian samples.

The minor allele frequencies for most of the SNPs in Table 6 differed substantially between subjects in our AA sample and the HapMap CEU population. Results for these SNPs in Caucasians are presented in Supplementary Table 4 (below). None of these SNPs showed evidence for association in the Caucasian datasets. Next, replication of the top findings from the AA GWAS was sought in the Caucasian datasets by examining all SNPs within 100 kb from any SNP with a p-value $<10^{-5}$. The minimum observed p-value was then adjusted for the effective number of tests in that region as proposed by Li and Ji[10] based on the pattern of LD observed in the CAU_M sample. Examination of SNPs within 100 kb of the top-ranked SNPs in the AA datasets (83-277 SNPs per region) revealed nominally significant results for most of these regions (p_{min} in Supplementary table 4 below), however none survived correction for multiple testing (p_{adj} in

Supplementary Table 4).

Supplementary Figure 1: QQ-Plot of AA GWAS.



Supplementary table 4. Comparison of top-ranked genetic association findings from genome wide survey in African Americans with results from meta analysis of Caucasian datasets.

African American Meta-Analysis								Caucasian Meta-Analysis				
								At SNP		In Region		
CHR	position	gene	SNP	P	Effect Allele	AF	OR (95% CI)	AF_{CEU}	P	SNP_{min}	P_{min}	P_{adj}
1	212,184,713	--	rs340849	7.52E-06	A	.20	.592 (.471-.745)	.40	.486	rs11807590	.055	.912
2	17,291,066	--	rs11889338	8.94E-06	A	.26	1.55 (1.28-1.88)	.01	--	rs12612975	.043	.656
2	27,760,977	SLC4A1AP	rs17006206	2.30E-06	G	.10	2.05 (1.52-2.76)	.00	--	rs13030345	.153	.864
3	28,903,864	--	rs2221154	2.58E-06	T	.19	.566 (.447-.718)	.53	.092	rs1353926	.002	.082
4	2,072,894	POLN	rs1923775	5.61E-06	T	.25	1.60 (1.30-1.95)	.72	.841	rs13141668	.075	.771
7	146,528,336	CNTNAP2	rs10273775	8.94E-06	G	.42	1.52 (1.27-1.84)	.54	.530	rs17170371	.050	.835
8	122,978,868	--	rs956225	8.71E-06	G	.03	.300 (.176-.510)	.13	.204	rs11995962	.037	.740
11	73,710,714	--	rs3888908	9.52E-06	A	.15	1.72 (1.36-2.20)	.17	.711	rs11236101	.086	.912
12	113,864,776	--	rs10850408	9.25E-07	T	.34	.629 (.523-.757)	.30	.359	rs1896331	.0014	.064
13	25,622,328	--	rs17511627	5.01E-06	C	.17	1.75 (1.37-2.22)	.16	.993	rs1886489	.032	.792
13	97,929,295	STK24	rs912330	3.79E-06	T	.14	.536 (.411-.698)	.33	.659	rs11616605	.032	.830
Other SNPs of interest from *APOE*-ε4 adjusted analysis.												
8	144,692,178	ZC3H3	rs3750208	7.28E-06	A	.04	.374 (.243-.574)	.18	.846	rs4403422	.035	.957
12	29,812,934	TMTC1	rs302318	1.97E-06	C	.26	.592 (.477-.735)	.43	.536	rs12817170	.019	.608
13	43,064,019	ENOX1	rs17460623	9.37E-06	C	.10	.493 (.361-.674)	.10	.552	rs1994874	.017	.388

1. Purcell S, Neale B, Todd-Brown K et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-575. Epub 2007 Jul 2025.
2. McPeek MS, Sun L. Statistical Tests for Detection of Misspecified Relationships by Use of Genome-Screen Data. Am J Hum Genet. 2000;66:1076-1094
3. Sun L, Wilder K, McPeek MS. Enhanced pedigree error detection. Hum Hered. 2002;54:99-110
4. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945-959.
5. Price AL, Patterson NJ, Plenge RM et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904-909. Epub 2006 Jul 2023.
6. Li Y, Ding J, Abecasis GR. MACH 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. American Journal of Human Genetics. 2006:2290
7. Liang KY, Zeger SL. Longitudinal data analysis using linear models. Biometrika. 1986;73:13-22
8. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986;42:121-130
9. R Development Core Team. R: A language and environment for statistical computing. 2008
10. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005;95:221-227