Online Methods

Sample ascertainment, diagnostic criteria and genotyping.
Stage 1 comprised a meta-analysis of four AD GWAS datasets (6688 cases, 13685 controls), including: GERAD1[3], EADI1[4], TGEN1[13] and ADNI[14]. All AD cases were diagnosed according NINCDS-ADRDA[36], DSM-IV or CERAD[37] criteria for either probable or definite AD. AD cases were predominantly female (62.4%). The mean age at disease onset and ascertainment in AD cases were 71.6 and 77.3 years, respectively. Stage 1 included a total of 7915 aged (≥60 years), screened controls (59.9% female; mean age at collection, 74.5 years) and 5770 population based, unscreened controls from the GERAD1 study (50.8% female, mean age at collection 48.6 years). The complete genome-wide meta-analysis results of stage 1 are available to researchers upon application.

Stage 2 included 4896 cases and 4903 controls, which comprised individual genotyping of the GERAD2 sample and *in silico* replication in the deCODE and AD-IG GWAS datasets. All cases were diagnosed according NINCDS-ADRDA[36], DSM-IV or CERAD[37] criteria for possible, probable or definite AD. AD cases were predominantly female (63.4%). The mean age at disease onset and ascertainment in AD cases were 72.3 and 76.8 years, respectively. The stage 2-control group (55.1% female, mean age at ascertainment 70.0 years) were predominantly aged (≥60 years) and screened for dementia (77.2%). Stage 3 comprised 8286 cases and 21258 controls, which included new genotyping in the EADI2[4] and Mayo2 samples, and *in silico* replication in the CHARGE sample[11]. All individuals included in these analyses have provided consent to take part in genetic association studies. We have obtained ethical approval to use these samples to search for susceptibility genes for Alzheimer's disease (MREC 04/09/030; Amendment 2 and 4; approved 27 July 2007). Full descriptions of all samples, genotyping methods and quality control can be found in the Supplementary Note. Clinical characteristics of all samples can be found in Supplementary Table 1.

Stage 1: combined analysis of 4 AD GWA Studies
An inverse variance weighted fixed effects meta-analysis was used to test for association with AD in the GERAD1, ADNI and TGEN datasets. The *P*-values from this meta-analysis were then combined with the publicly available *P*-values from the EADI1 study using Fisher's combined probability test. The combined analysis tested 496795 autosomal SNPs. These SNPs passed QC in each of the GERAD1, ADNI and EADI1 GWA studies. 457509 of these SNPs were successfully imputed in the TGEN sample (which unlike the other studies employed the Affymetrix 500K array). In the combined analysis, 67 SNPs were associated with AD at $P \leq 1 \times 10^{-5}$ (see Supplementary Table 2). Full summary statistics were obtained from the EADI consortium for these 67 SNPs and the analysis was repeated using inverse variance-weighted meta-analysis. 61 SNPs remained significantly associated with AD at $P \leq 1 \times 10^{-5}$. In selecting SNPs for replication in Stage 2 we chose to exclude variants at the *APOE* locus (26 SNPs), as this is a known susceptibility gene for AD and also variants at the *CLU* and *PICALM* loci (1 SNP and 8 SNPs respectively), as these data were already reported in the GERAD2 sample[3]. We restricted genotyping of *CRI* and *BIN1* SNPs to the most significant

markers at each locus, as they had not previously been tested in the Stage 2 sample (6 SNPs excluded). We excluded 8 SNPs that were in high LD with the other SNPs selected for genotyping (see Supplementary Table 2).

Stage 2, 3 and Final Meta-analyses
Both new genotyping data (GERAD2) and *in silico* replication (deCODE and Reimenschneider GWAS datasets) were included in Stage 2. Data from these three samples were combined using an inverse-variance weighted fixed effects meta-analysis (total sample of up to 4896 cases and 4903 controls). We employed a Bonferroni adjusted threshold for significance, taking account of 12 tests, of $P=4.2\text{x}10^{-3}$. For the three novel SNPs that showed evidence for association in Stage 2 ($P<4.2\text{x}10^{-3}$), summary data (including odds ratios and variances) from the EADI2, Mayo2 and CHARGE and EADI2 studies were combined in an inverse-variance weighted fixed effects meta-analysis (total sample of up to 8286 cases and 21258 controls), employing a Bonferroni adjusted threshold for significance, taking account of 3 tests, of P=0.0167.

Finally, we combined full summary data (including odds ratios and variances) from all datasets (GERAD1&2, EAD1&2, ADNI, TGEN, deCODE, AD-IG, Mayo2, CHARGE) using an inverse-variance weighted fixed effects meta-analysis (total sample of up to 19870 cases and 39846 controls). For meta-analyses at all stages Cochran's Q-test was performed and I^2 calculated to assess heterogeneity. The total sample size tested for each SNP is shown in Supplementary Table 7. Summary statistics for all datasets are shown in Supplementary Table 8.

Secondary analyses
We tested the genome-wide significant SNPs for relationships with age at onset (AAO). To this end, age at onset (in years) was used as the dependent variable in a linear regression analysis and an additive model was assumed. Covariates were included in the logistic regression analysis to allow for country of origin. We found no evidence for association between these loci and age at onset of AD (rs3764650: P=0.17; rs670139: P=0.38; rs610932: P=0.95). We also tested the SNPs for association with AD, whilst adjusting for the presence of at least one *APOE* ε4 allele, within a logistic regression framework. Inclusion of presence/absence of *APOE* ε4 as a covariate had little effect on the results (see Supplementary Table 5B). Finally, we tested for genetic interactions between all pairwise combinations of genome-wide significant SNPs in *APOE*, *CLU*, *PICALM*, *BIN1*, *CR1*, *ABCA7* and *MS4A*. Logistic regression analyses were performed including terms for covariates, main effects of each SNP, and a SNPxSNP interaction term. *e.g.* $\ln(P/(1-P)) = \alpha + \beta X_A + \gamma X_B + \delta Y_1 + \varepsilon Y_2 + \zeta Y_3 + \eta Y_4 + \theta Y_5 + \iota Y_6 + \kappa X_A X_B$, where P is the probability of having the disease, X_A and X_B correspond to variables representing the number of minor alleles at SNPs A and B respectively, Y_1 to Y_6 correspond to the covariates included in the main analysis (country of origin and principal components), α is a constant and β, γ, δ, ε, ζ, η, θ, ι and κ are regression coefficients. We tested whether the regression coefficient that represents the interaction term, κ, equals zero or not. *P*-values for this 1 degree of freedom test are presented in Supplementary Table 5A.

Expression quantitative trait loci (eQTL) analysis

Expression profiles were analysed within two eQTL datasets. The first, published by Stranger and colleagues[17] consists of gene expression profiles generated using RNA extracted from lymphoblastoid cell lines generated from 60 unrelated CEU HapMap individuals. Expression analysis was performed using Illumina's commercial whole genome expression array, Sentrix Human-6 Expression BeadChip. The second eQTL dataset[18] analysed 143 neurologically normal subjects of European ancestry. Frozen tissue samples were obtained from four brain regions (cerebellum, pons, frontal and temporal cortices) for each subject. Genotyping was performed using Infinium HumanHap550 BeadChips (Illumina). Expression analysis was performed using Illumina HumanRef-8 Expression BeadChips. Genotype data were used as presented in the original publications[17,18,38]. The expression data were normalised and log transformed as described in the original papers[17,18,38]. eQTLs were tested by linear regression of normalised expression level on SNP genotypes (coded as the number of minor alleles at each SNP: 0, 1 or 2).