Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease

Eleven susceptibility loci for late-onset Alzheimer’s disease (LOAD) were identified by previous studies; however, a large portion of the genetic risk for this disease remains unexplained. We conducted a large, two-stage meta-analysis of genome-wide association studies (GWAS) in individuals of European ancestry. In stage 1, we used genotyped and imputed data (7,055,881 SNPs) to perform meta-analysis on 4 previously published GWAS data sets consisting of 17,008 Alzheimer’s disease cases and 37,154 controls. In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer’s disease cases and 11,312 controls. In addition to the APOE locus (encoding apolipoprotein E), 19 loci reached genome-wide significance ($P < 5 \times 10^{-8}$) in the combined stage 1 and stage 2 analysis, of which 11 are newly associated with Alzheimer’s disease.

Alzheimer’s disease is a devastating neurological disorder primarily affecting the elderly. The disease manifests with progressive deterioration in cognitive functions, leading to loss of autonomy. The APOE gene (encoding apolipoprotein E) is a major genetic risk factor for Alzheimer’s disease. Previous GWAS in individuals of European ancestry identified nine other genomic regions associated with LOAD. Recently, a rare susceptibility variant in $TREM2$ was identified. The search for additional genetic risk factors requires large-scale meta-analysis of GWAS to increase statistical power. Under the banner of I-GAP (International Genomics of Alzheimer’s Project), we conducted a meta-analysis of 4 GWAS samples of European ancestry totaling 17,008 cases and 37,154 controls (stage 1) followed up by genotyping of 11,632 SNPs showing moderate evidence of association ($P < 1 \times 10^{-3}$ in stage 1) in an independent sample that included 8,572 cases and 11,312 controls (stage 2).

In the stage 1 meta-analysis, we used data from four consortia: the Alzheimer’s Disease Genetic Consortium (ADGC), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the European Alzheimer’s Disease Initiative (EADI) and the Genetic and Environmental Risk in Alzheimer’s Disease (GERAD) Consortium (Table 1, Online Methods, Supplementary Table 1 and Supplementary Note). We used European population reference (EUR) haplotype data from the 1000 Genomes Project (2010 interim release based on sequence data freeze from 4 August 2010 and phased haplotypes from December 2010) to impute genotypes for up to 11,863,202 SNPs per data set. We excluded SNPs that did not pass quality control in each study (Supplementary Table 2 and Supplementary Note).

Our meta-analysis included SNPs either genotyped or successfully imputed in at least 40% of the Alzheimer’s disease cases and 40% of the control samples across all data sets (7,055,881 SNPs; Online Methods). In each data set, genotype dosages were analyzed as described in the Supplementary Note (Supplementary Table 2). We performed meta-analysis of the results after applying genomic control correction to each study. The genomic control inflation factor for the meta-analysis was 1.087 for the full set of SNPs and 1.082 after excluding SNPs within the APOE locus (chr. 19: 45,409,039–45,412,650) and within 500 kb of SNPs associated with Alzheimer’s disease at a prespecified level of genome-wide significance ($P < 5 \times 10^{-8}$) in stage 1 (see Supplementary Fig. 1 for quantile-quantile plots).

In addition to the APOE locus, 14 genomic regions had associations that reached the genome-wide significance level (Fig. 1). Nine had been previously identified by GWAS as genetic susceptibility factors, and five ($HLA-DRB5$–$HLA-DRB1$, $PTK2B$, $SORL1$, $SLC24A4$–$RIN3$ and $DSG2$) represent newly associated loci (Table 2). $SORL1$ had previously been identified as an Alzheimer’s disease gene through candidate gene approaches and in a GWAS combining ADGC and Asian samples. Genes attributed to a signal were those closest to the most significantly associated SNP. However, we are aware that these are potentially not the causative genes. Detailed results for each region are given in Supplementary Figures 2–7.

In stage 2, we selected for genotyping all stage 1 SNPs with a $P$ value less than $1 \times 10^{-3}$, excluding SNPs flanking APOE (chr. 19: 45,409,039–45,412,650) ($n = 19,532$; see URLs for database access). From the initial set of SNPs, 14,445 could be genotyped using Illumina iSelect technology. After quality control procedures (Online Methods), we considered 11,632 SNPs for association analysis. The stage 2 sample included 8,572 cases and 11,312 controls of European ancestry originating from Austria, Belgium, Finland, Germany, Greece, Hungary, Italy, Spain, Sweden, the UK and the United States (Table 1 and Supplementary Note). We observed 116 SNPs showing the same risk allele and direction of association in stages 1 and 2 that were significantly associated with Alzheimer’s disease risk in stage 2 after a strict Bonferroni correction for multiple testing ($P < 4.3 \times 10^{-6}$). Of these 116 SNPs, 80 had been associated at genome-wise significance with Alzheimer’s disease risk in stage 1. Additionally, in analyses in stage 2, 2,562 SNPs were associated with Alzheimer’s disease at a nominal level of significance ($P < 0.05$), having the same risk allele and direction of association as in stage 1.

The results from stages 1 and 2 and from the combined stage 1 and stage 2 data sets, which represent a secondary discovery effort,
are shown in Table 2. With the exception of CD33 and DSG2, we nominally replicated all loci that surpassed the genome-wide significance level in stage 1. Inability to replicate DSG2 is not surprising, as evidence of association for this locus was based on data for a single SNP and was not supported by data from surrounding SNPs in linkage disequilibrium (LD, $r^2 > 0.8$; Supplementary Fig. 7b). Moreover, seven new loci reached the genome-wide significance level in the combined analysis (Table 2). More detailed results for the seven newly identified LOAD loci are provided in Supplementary Figures 8–11. There was no significant heterogeneity across studies at any of the loci, except at DSG2 (Table 2 and Supplementary Figs. 12–16). To identify potential causative genes, we also examined all SNPs with association $P < 5 \times 10^{-8}$ that were within 500 kb of the top SNP at each locus to identify cis expression quantitative trait locus (cis-eQTL) associations (Online Methods and Supplementary Table 3).

The results from the combined stage 1 and stage 2 data sets also identified 13 loci with suggestive evidence of association ($P < 1 \times 10^{-6}$) (Supplementary Table 4). Among these, we detected a signal for rs93810490 ($P = 6.3 \times 10^{-7}$), which is located approximately 5.5 kb away from the 3′ end of TREM1 and 24 kb away from the 5′ end of TREM2. TREM2 was recently reported to carry a rare variant (encoding p.Arg47His) associated with three- to fourfold increased risk of developing Alzheimer’s disease.7,9 This region also reached genome-wide significance in a study of cerebral spinal fluid levels of phosphorylated tau, a biomarker for Alzheimer’s disease.11

Beyond the already known, GWAS-defined genes (ABCA7, BIN1, CD33, CLU, CR1, CD2AP, EPHA1, MS4A6A-MS4A4E and PICALM), the most significant new association was in the $HLA-DRB5–DRB1$ region (encoding major histocompatibility complex, class II, DRB5 and DRB1, respectively). This region is associated with immunocompetence and histocompatibility and, interestingly, with risk of both multiple sclerosis and Parkinson disease.12,13 Owing to the complex genetic organization of the human leukocyte antigen (HLA) region on chromosome 6, we were unable to define which gene(s) are responsible for this signal (Supplementary Fig. 6a).

The second strongest signal was within the SORL1 gene (encoding sortilin-related receptor, L(DLR class) 1). Our data clearly demonstrated that this gene was associated at genome-wide significance in European samples. SORL1 is noteworthy, as it is associated with increased risk of both autosomal dominant and sporadic forms of Alzheimer’s disease4,15 and represents the first LOAD gene that directly connects aberrant trafficking and metabolism of the amyloid precursor protein (APP) to LOAD.14

The third locus, PTK2B (encoding protein tyrosine kinase 2β), is only approximately 130 kb away from CLU, but we believe the two signals are independent because (i) the two most strongly associated SNPs within each of these two genes are not in LD ($D’ = 0.06$ and $r^2 = 0.003$ as computed using 1000 Genomes Project data); (ii) a recombination peak exists between the two loci (Fig. 2); and (iii) conditional analysis in the stage 2 data confirmed the independence of the PTK2B association (Supplementary Fig. 17 and Supplementary Table 5). The protein encoded by PTK2B may be an intermediate between neuropeptide-activated receptors or neurotransmitters that increase calcium flux and the downstream
Table 2  Summary of stage 1, stage 2 and overall meta-analyses for SNPs reaching genome-wide significance after stages 1 and 2

<table>
<thead>
<tr>
<th>SNPa</th>
<th>Chr.</th>
<th>Positionb</th>
<th>Closest genec</th>
<th>Major/minor alleles</th>
<th>MAF2</th>
<th>Stage 1</th>
<th>Meta</th>
<th>Stage 2</th>
<th>Meta</th>
<th>Overall</th>
<th>P (%)</th>
<th>P valuesd</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6656401</td>
<td>1</td>
<td>207692049</td>
<td>CR1</td>
<td>G/A</td>
<td>0.197</td>
<td>1.17</td>
<td>7.7×10−15</td>
<td>1.21</td>
<td>7.9×10−11</td>
<td>1.18</td>
<td>5.7×10−24</td>
<td>0.78×10−1</td>
</tr>
<tr>
<td>rs6733839</td>
<td>2</td>
<td>127892810</td>
<td>BIN1</td>
<td>C/T</td>
<td>0.409</td>
<td>1.21</td>
<td>1.7×10−13</td>
<td>1.24</td>
<td>3.4×10−10</td>
<td>1.22</td>
<td>6.9×10−14</td>
<td>2.8×10−4</td>
</tr>
<tr>
<td>rs10948633</td>
<td>6</td>
<td>47487762</td>
<td>CD2AP</td>
<td>A/G</td>
<td>0.266</td>
<td>1.10</td>
<td>3.1×10−10</td>
<td>1.09</td>
<td>4.1×10−10</td>
<td>1.10</td>
<td>5.2×10−11</td>
<td>0.9×10−1</td>
</tr>
<tr>
<td>rs11771145</td>
<td>7</td>
<td>143110762</td>
<td>EPHA1</td>
<td>G/A</td>
<td>0.338</td>
<td>0.90</td>
<td>8.8×10−10</td>
<td>0.90</td>
<td>2.8×10−5</td>
<td>0.90</td>
<td>1.1×10−3</td>
<td>14.2×10−1</td>
</tr>
<tr>
<td>rs9331896</td>
<td>8</td>
<td>27476786</td>
<td>CLU</td>
<td>T/C</td>
<td>0.379</td>
<td>0.86</td>
<td>9.6×10−11</td>
<td>0.90</td>
<td>4.5×10−10</td>
<td>0.86</td>
<td>2.8×10−29</td>
<td>0.49×10−1</td>
</tr>
<tr>
<td>rs983392</td>
<td>11</td>
<td>59923508</td>
<td>MS4A6A</td>
<td>A/G</td>
<td>0.403</td>
<td>0.90</td>
<td>2.8×10−11</td>
<td>0.90</td>
<td>4.5×10−10</td>
<td>0.90</td>
<td>6.1×10−16</td>
<td>1.4×10−1</td>
</tr>
<tr>
<td>rs10792821</td>
<td>11</td>
<td>85867875</td>
<td>PICALM</td>
<td>G/A</td>
<td>0.358</td>
<td>0.88</td>
<td>6.5×10−16</td>
<td>0.85</td>
<td>1.1×10−11</td>
<td>0.87</td>
<td>9.3×10−26</td>
<td>0.94×10−1</td>
</tr>
<tr>
<td>rs4147929</td>
<td>15</td>
<td>1063443</td>
<td>ABCA7</td>
<td>G/A</td>
<td>0.190</td>
<td>1.14</td>
<td>1.7×10−9</td>
<td>1.17</td>
<td>9.9×10−8</td>
<td>1.15</td>
<td>0.94×10−1</td>
<td>0.94×10−1</td>
</tr>
<tr>
<td>rs38654444a</td>
<td>19</td>
<td>51727962</td>
<td>CD33</td>
<td>C/A</td>
<td>0.307</td>
<td>0.91</td>
<td>5.1×10−8</td>
<td>0.99</td>
<td>6.9×10−1</td>
<td>0.94</td>
<td>3.0×10−6</td>
<td>0.69×10−1</td>
</tr>
<tr>
<td>New loci reaching genome-wide significance in the discovery analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9271192</td>
<td>6</td>
<td>32578530</td>
<td>HLA-DRB5− HLA-DRB1</td>
<td>A/C</td>
<td>0.276</td>
<td>1.11</td>
<td>1.6×10−8</td>
<td>1.12</td>
<td>4.2×10−5</td>
<td>1.11</td>
<td>2.9×10−12</td>
<td>0.54×10−1</td>
</tr>
<tr>
<td>rs28834970</td>
<td>8</td>
<td>27195121</td>
<td>PTK2B</td>
<td>A/C</td>
<td>0.366</td>
<td>1.10</td>
<td>3.3×10−9</td>
<td>1.11</td>
<td>4.3×10−6</td>
<td>1.10</td>
<td>7.4×10−14</td>
<td>1.0×10−1</td>
</tr>
<tr>
<td>rs11218343</td>
<td>11</td>
<td>121435587</td>
<td>SORL1</td>
<td>T/C</td>
<td>0.039</td>
<td>0.76</td>
<td>5.0×10−11</td>
<td>0.78</td>
<td>4.0×10−6</td>
<td>0.77</td>
<td>9.7×10−15</td>
<td>0.83×10−1</td>
</tr>
<tr>
<td>rs10498633</td>
<td>14</td>
<td>92926952</td>
<td>SLC24A4− RIN3</td>
<td>G/T</td>
<td>0.217</td>
<td>0.90</td>
<td>1.5×10−7</td>
<td>0.93</td>
<td>7.8×10−3</td>
<td>0.91</td>
<td>5.5×10−9</td>
<td>0.63×10−1</td>
</tr>
<tr>
<td>rs80937314</td>
<td>19</td>
<td>29088958</td>
<td>DSG2</td>
<td>C/T</td>
<td>0.017</td>
<td>0.54</td>
<td>4.6×10−9</td>
<td>1.01</td>
<td>9.0×10−1</td>
<td>0.73</td>
<td>1.0×10−4</td>
<td>38.3×10−2</td>
</tr>
<tr>
<td>New loci reaching genome-wide significance in the combined discovery and replication analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs355496692</td>
<td>23</td>
<td>234068476</td>
<td>INPP5D</td>
<td>C/T</td>
<td>0.488</td>
<td>1.07</td>
<td>9.6×10−5</td>
<td>1.10</td>
<td>5.7×10−5</td>
<td>1.08</td>
<td>3.2×10−8</td>
<td>0.8×10−1</td>
</tr>
<tr>
<td>rs190982</td>
<td>5</td>
<td>88223420</td>
<td>MEFC</td>
<td>A/G</td>
<td>0.408</td>
<td>0.92</td>
<td>2.5×10−6</td>
<td>0.93</td>
<td>3.4×10−3</td>
<td>0.93</td>
<td>3.2×10−8</td>
<td>0.64×10−1</td>
</tr>
<tr>
<td>rs2718058</td>
<td>7</td>
<td>38741534</td>
<td>NME8</td>
<td>A/G</td>
<td>0.373</td>
<td>0.93</td>
<td>1.3×10−5</td>
<td>0.91</td>
<td>6.3×10−5</td>
<td>0.91</td>
<td>4.8×10−9</td>
<td>0.92×10−1</td>
</tr>
<tr>
<td>rs1476679</td>
<td>7</td>
<td>100004446</td>
<td>ZCWPW1</td>
<td>T/C</td>
<td>0.287</td>
<td>0.92</td>
<td>7.4×10−6</td>
<td>0.89</td>
<td>9.7×10−6</td>
<td>0.91</td>
<td>5.6×10−10</td>
<td>0.7×10−1</td>
</tr>
<tr>
<td>rs10838725</td>
<td>11</td>
<td>47557871</td>
<td>CELF1</td>
<td>T/C</td>
<td>0.316</td>
<td>1.08</td>
<td>6.7×10−6</td>
<td>1.09</td>
<td>4.1×10−4</td>
<td>1.08</td>
<td>1.1×10−8</td>
<td>0.76×10−1</td>
</tr>
<tr>
<td>rs17125944</td>
<td>14</td>
<td>53400629</td>
<td>FERMT2</td>
<td>T/C</td>
<td>0.092</td>
<td>1.13</td>
<td>1.0×10−5</td>
<td>1.17</td>
<td>1.6×10−4</td>
<td>1.14</td>
<td>7.9×10−9</td>
<td>1.0×10−1</td>
</tr>
<tr>
<td>rs727458</td>
<td>20</td>
<td>55018260</td>
<td>CASS4</td>
<td>T/C</td>
<td>0.083</td>
<td>0.87</td>
<td>1.6×10−6</td>
<td>0.89</td>
<td>4.1×10−3</td>
<td>0.88</td>
<td>2.5×10−8</td>
<td>0.99×10−1</td>
</tr>
</tbody>
</table>

Chr., chromosome.
SNPs showing the best level of association after meta-analysis of studies 1 and 2. *Build 37, assembly hg19. †Genes located ± 100 kb of the top SNP. *Average in the discovery sample.
*Calculated with respect to the minor allele. ‡Cochran’s Q test. †Not replicated in stage 2.
signals regulating neuronal activity such as mitogen-activated protein kinase (MAPK) signaling16. PTK2B is involved in the induction of long-term potentiation in the hippocampal CA1 (cornu ammonis 1) region, a central process in the formation of memory17. We cannot, however, exclude the possibility that there are multiple signals in the PTK2B−CLU region that are functionally connected to a single gene. For instance, two SNPs associated with genome-wide significance in the PTK2B−CLU region are eQTLs for the gene DPYSL2 that has been implicated in Alzheimer’s disease18 (Supplementary Table 3).

The fourth locus was SLC24A4 (encoding solute carrier family 24 (sodium/potassium/calcium exchanger), member 4). The SLC24A4 gene encodes a protein involved in iris development and hair and skin color variation in humans in addition to being associated with the risk of developing hypertension19,20. SLC24A4 is also expressed in the brain and may be involved in neural development21. Of note, in the vicinity of the most strongly associated SNP is another gene called RIN3 (encoding Ras and Rab interactor 3), and its gene product directly interacts with the BIN1 gene product22, a protein that may be connected to tau-mediated pathology23.

In addition to these four loci reaching genome-wide significance in stage 1, seven new loci reached genome-wide significance in the combined analysis.

The strongest association at one of these new loci was intronic in the ZCWPW1 gene (encoding zinc finger, CW type with PWWP
domain 1), whose corresponding protein modulates epigenetic regulation. However, the region defined by all the SNPs associated with Alzheimer’s disease in our data is large and contains about ten genes (Supplementary Fig. 9b). Another interesting possible candidate gene in the ZCWPW1 region is NYAP1, as disruption of the corresponding gene in mice affects brain size, neurite elongation and, more generally, neuronal morphogenesis. Our data do not resolve which gene in this region may be causal.

A second locus was within the CELF1 gene (encoding CUGBP, Elav-like family member 1), whose gene product is a member of the protein family that regulates pre-mRNA alternative splicing. As with the ZCWPW1 locus, the region of interest is large and contains about ten genes (Supplementary Fig. 10a). Among these genes is MADD (encoding MAP kinase-activating death domain), the reduced expression of which may affect long-term neuronal viability in Alzheimer’s disease. A discrete signal was observed adjacent to NME8 (encoding NME/NM23 family member 8), which is responsible for primary ciliary dyskinesia type 6 (ref. 28).

The FERM2 gene (encoding fermitin family member 2) is expressed in the brain. Its corresponding protein localizes to cell matrix adhesion structures, activates integrins, is involved in the orchestration of actin assembly and cell shape modulation, and is an important mediator of angiogenesis. An association between the Drosophila melanogaster ortholog of FERM2 (fit1/fit2) and tau-mediated toxicity was recently described.

We identified a fifth signal on chromosome 20 at CASS4 (encoding Cas scaffolding protein family member 4).Little is known about the function of the encoded protein. However, the Drosophila CASS family ortholog (p130CAS) binds to CMS, the Drosophila ortholog of CD2AP (CMS), a known Alzheimer’s disease susceptibility gene (Table 2) that is involved in actin dynamics.

Another locus was identified at INPP5D (encoding inositol polyphosphate-5-phosphatase, 145 kDa) on chromosome 2. INPP5D is expressed at low levels in the brain, but the encoded protein has been shown to interact with CD2AP, whose corresponding gene is one of the Alzheimer’s disease genes previously identified by GWAS, and to modulate, along with GRB2, metabolism of APP.

We identified a seventh signal adjacent to MEF2C (encoding myocyte enhancer factor 2). Mutations at this locus are associated with severe mental retardation, stereotypic movements, epilepsy and cerebral malformation. MEF2C protein limits excessive synapse formation during activity-dependent refinement of synaptic connectivity and thus may facilitate hippocampal-dependent learning and memory.

In summary, our Alzheimer’s disease GWAS meta-analysis has identified 11 new susceptibility loci in addition to the already known ABCA7, APOE, BIN1, CLU, CR1, CD2AP, EP5H1, M4S4A6-M4S4A4E and PICALM genes. However, we were not able to replicate association of CD33 in our stage 2 analysis (P = 0.61). We did not detect any biases in terms of imputation in our discovery data sets or genotyping in our replication data sets (data not shown), suggesting a potential statistical fluctuation across our populations as an explanation for the lack of replication. However, recent data suggest that genetically determined decreased CD33 expression might reduce Alzheimer’s disease risk and interfere with amyloid β peptide clearance, a dys-regulation thought to be central in late-onset forms of Alzheimer’s disease. Further investigations in independent case-control studies will thus be required to confirm or refute the association of CD33 with Alzheimer’s disease.

The newly associated loci reinforce the importance of some previously suspected pathways such as APP (SORL1 and CASS4) and tau (CASS4 and FERMT2) in pathology. Several candidate genes at these loci are involved in pathways already shown to be enriched for association signal in Alzheimer’s disease GWAS, such as immune response and inflammation (HLA-DRB5-DRB1, INPP5D and MEF2C), which is also supported by the described association of Alzheimer’s disease with CR1 (ref. 3) and TREM2 (refs. 8, 9), cell migration (PTK2B) and lipid transport and endocytosis (SORL1). Our results also suggest the existence of new pathways underlying Alzheimer’s disease. These pathways could include hippocampal synaptic function (MEF2C and PTK2B), cytoskeletal function and axonal transport (CELF1, NME8 and CASS4), regulation of gene expression and post-translational modification of proteins, and microglial and myeloid cell function (INPP5D).

Examining the genetic effect attributable to all the associated loci, we demonstrated that the most strongly associated SNPs at each locus other than APOE had population-attributable fractions (PAFs) or preventive fractions between 1.0–8.0% in the stage 2 sample (Supplementary Table 6). Strong efforts in sequencing and post-GWAS analyses will now be required to fully characterize the candidate genes and functional variants responsible for the association of these GWAS-identified loci with Alzheimer’s disease risk and to understand their exact roles in the pathophysiology of Alzheimer’s disease.

### URLs


### METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.
ACKNOWLEDGMENTS

This work was made possible by the generous participation of the control subjects, the patients and their families. Select costs were funded by the French National Foundation on Alzheimer’s Disease and Related Disorders. Data management involved the Centre National de Génotypage and was supported by the Institut Pasteur de Lille, INSERM, FRC (Fondation pour la Recherche sur le Cerveau) and Rotary. This work has been developed and supported by the LABEX (Laboratory of Excellence Program Investment for the Future) DISTALZ grant (Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer’s Disease). The French National Foundation on Alzheimer’s Disease and Related Disorders and the Alzheimer’s Association (Chicago, Illinois) grant provided some funds to each consortium for analyses.

GERAD was supported by the Welcome Trust, the MRC, Alzheimer’s Research UK (ARUK) and the Welsh government. ADGC and CHARGE were supported by the US National Institutes of Health, National Institute on Aging (NIH-NIA), including grants U01 AG032984 and R01 AG033193 (additional US National Institutes of Health grant numbers are listed in the Supplementary Note).

CHARGE was also supported by Erasmus Medical Center and Erasmus University.

Complete acknowledgments are detailed in the Supplementary Note.

AUTHOR CONTRIBUTIONS


COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1INSERM, U744, Lille, France. 2Université Lille 2, Lille, France. 3Institut Pasteur de Lille, Lille, France. 4Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. 5Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. 6Institute of Psychological Medicine and Clinical Neurosciences, Medical Research Council (MRC) Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, UK. 7Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 8Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 9Department of Medicine (Biomedical Genetics), Boston University School of Public Health, Boston, Massachusetts, USA. 10Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts, USA. 11Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. 12Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA. 13John P. Hussman Institute for Human Genetics, University of Miami, Miami, Florida, USA. 14Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida, USA. 15Functional Genomics Center Zurich, Eidenössische Technische Hochschule (ETH)/University of Zurich, Zurich, Switzerland. 16Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA. 17Faculty of Medicine, University of Iceland, Reykjavik, Iceland. 18Icelandic Heart Association, Kopavogur, Iceland. 19Department of Radiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. 20Netherlands Consortium for Healthy Aging, Leiden, The Netherlands. 21Centre National de Génotypage, Institut Génomique, Commissariat à l’Energie Atomique, Evry, France. 22Fondation Jean Dausset, Centre d’Etude du Polymorphisme Humain (CEPH), Paris, France. 23Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA. 24Institute for Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria. 25Memory Clinic of Fundación Alzheimer Centre Educational (ACE). Institut Català de Neurociències Aplicades, Barcelona, Spain. 26Taub Institute on Alzheimer’s Disease and the Aging Brain,
ONLINE METHODS

All case-control studies are described in Table 1, in the Supplementary Note (see full description of the I-GAP data sets) and in Supplementary Tables 1, 7 and 8. Written informed consent was obtained from study participants or, for those with substantial cognitive impairment, from a caregiver, legal guardian or other proxy, and the study protocols for all populations were reviewed and approved by the appropriate institutional review boards.

Imputation and SNP selection for stage 1 analysis. After quality control criteria were finalized for each individual and each sample collection (SNPs with call rates of < 95% were excluded; Supplementary Note), IMPUTE2 (ref. 42) or MaCh/Minimac43 software (Supplementary Table 2) was used to impute the genotypes of all participants with haplotypes derived from samples of European ancestry in the 1000 Genome Project (2010 interim release based on the sequence data freeze from 4 August 2010 and phased haplotypes from December 2010). In each data set, SNPs with R2 or info score quality estimates of less than 0.3, as indicated by MaCh or IMPUTE2, respectively (with these two quality estimates described to be equivalent), were excluded from analyses. Similarly, SNPs with a MAF of <1% were also excluded. After these procedures, a maximum of 8,133,148 SNPs were retained that were present in at least 1 data set.

In each case-control data set, the association of LOAD with genotype dosage was analyzed by a logistic regression model including covariates for age, sex and principal components to adjust for possible population stratification (Supplementary Table 2). For the three CHARGE cohorts with incident Alzheimer’s disease data, Cox proportional hazards models were applied to the remaining 12,446 SNPs to detect potential associations. SNPs exhibiting an Illumina score superior or equal to 0.4 were selected for a devoted Illumina website to develop an iSelect microarray. A total of 16,732 SNPs associated with Alzheimer’s disease risk and exhibiting evidence of non-European ancestry. Duplicated and related individuals were identified (Supplementary Table 10). Briefly, discrepancies in sex were examined using genetic sex as estimated by PLINK on 40 SNPs on chromosome X. We also removed 93 individuals from a single plate for whom an abnormal number of discrepancies in sex were observed, suggesting that sample mixing had occurred. Using a panel of 261 ancestry-informative markers (AIMs), we performed a principal-component analysis (PCA) on HapMap 2 data with the function SMARTPCA from EIGENSOFT 4.2 software47. For each country, individuals were projected onto the first two PCA axes to define their genetic ancestry. Individuals with evidence of non-European ancestry were then identified by applying a Bayesian clustering approach48 to their coordinates on the first two axes. Identity by descent (IBD) was computed for all pairs of individuals using PLINK, and individuals in a pair with IBD greater than 0.98 were considered to be duplicates. If clinical data for duplicated individuals were discordant, both individuals were excluded. Otherwise, the individual with the greater proportion of missing genotype was excluded. Similarly, IBD was computed for all pairs of individuals in data from each country separately, using 6,764 autosomal SNPs with MAF of >1% and selected to minimize LD. Individuals in pairs with IBD greater than 0.2 were considered to be related and were iteratively removed so as to obtain a sample of unrelated individuals within each country data set.

Stage 2 sample quality control. The iSelect microarray contained 33,368 SNPs, of which 11,632 were devoted to stage 2. These supplementary SNPs included various genetic data that allowed us to further refine our quality control processes. On the basis of data for all of these SNPs, we excluded individuals who had more than 3% missing genotypes, showed a discrepancy between reported sex and sex estimated on the basis of genetic data (genetic sex) or showed evidence of non-European ancestry. Duplicated and related individuals were identified (Supplementary Table 10). Briefly, discrepancies in sex were examined using genetic sex as estimated by PLINK on 40 SNPs on chromosome X. We also removed 93 individuals from a single plate for whom an abnormal number of discrepancies in sex were observed, suggesting that sample mixing had occurred. Using a panel of 261 ancestry-informative markers (AIMs), we performed a principal-component analysis (PCA) on HapMap 2 data with the function SMARTPCA from EIGENSOFT 4.2 software47. For each country, individuals were projected onto the first two PCA axes to define their genetic ancestry. Individuals with evidence of non-European ancestry were then identified by applying a Bayesian clustering approach48 to their coordinates on the first two axes. Identity by descent (IBD) was computed for all pairs of individuals using PLINK, and individuals in a pair with IBD greater than 0.98 were considered to be duplicates. If clinical data for duplicated individuals were discordant, both individuals were excluded. Otherwise, the individual with the greater proportion of missing genotype was excluded. Similarly, IBD was computed for all pairs of individuals in data from each country separately, using 6,764 autosomal SNPs with MAF of >1% and selected to minimize LD. Individuals in pairs with IBD greater than 0.2 were considered to be related and were iteratively removed so as to obtain a sample of unrelated individuals within each country data set.

Finally, individuals with missing clinical data and controls less than 25 years of age were excluded from the analysis. After sample quality control (Supplementary Table 10), 19,884 individuals (8,572 cases and 11,312 controls) were available for analysis in stage 2.

Statistical analysis. For the stage 1 meta-analysis, we undertook fixed-effects inverse variance-weighted meta-analysis with the standard errors of the β-coefficient scaled by the square roots of study-specific genomic inflation factors estimated before combining the summary statistics across data sets. Each consortium performed an independent stage 1 meta-analysis after downloading the data files available on the I-GAP website. Two software packages were used for meta-analysis: METAL49 and GWAMA50. Very similar results were generated independently of the software used and as expected, perfect matching was observed between the analyses undertaken by each of the 4 consortia.

For stage 2, association tests were performed for each country for all high-quality genotyped SNPs under an additive model, using logistic regression as implemented in PLINK. Analysis was adjusted for age, sex and principal
components, when necessary. Using SMARTPCA, PCA was performed on individuals from each country separately. Difference in PCA coordinates between cases and controls were tested for the first four principal components, and analysis was further adjusted on principal components if the P value of this test was lower than 0.05. PCA for Bonn stage 2 samples was based on GWAS data. For imputed data sets, association tests were performed using likelihood score tests for missing data as implemented in SNPTEST. Genotyped and imputed German samples were analyzed separately, and results were then combined by fixed-effects meta-analysis using the inverse variance approach as implemented in METAL. Using this approach, a fixed-effects meta-analysis was then performed to combine stage 2 results from the different countries. We also performed the analysis separately for each center in stage 2 and combined the results by fixed-effects meta-analysis. Results were similar to those obtained when analysis was performed by country (data not shown).

We finally generated fixed-effects inverse variance–weighted meta-analyses by combining summary statistics across ADGC, CHARGE, EADI, GERAD and stage 2 data by country. At this point, we performed Cochran’s Q test for heterogeneity and generated I^2 estimates with METAL to evaluate the possible effect of study heterogeneity on the results.

A graphic representation of the association signal in the stage 1 data was generated with LocusZoom software for all the loci of interest reaching a genome-wide significant level after combined stage 1 and stage 2 analyses. PAF was calculated using the Levin equation.

Annotation of I-GAP top SNPs for eQTLs. To gain further biological insights, we explored reported associations between SNPs in the top I-GAP loci and gene expression. We first selected all SNPs that reached genome-wide significance (P value ≤ 5 × 10^{-8}) in the combined stage 1 and stage 2 analysis and were located in a 500-kb window upstream or downstream of the top SNP at each locus (Table 2). We then searched for published data on gene expression associated with each of these SNPs in the eQTL database from the Pritchard laboratory (see URLs). For each reported eQTL gene and each type of eQTL association as defined in this database, we then counted the number of reported eQTL SNPs and selected the one with the lowest P value.