nature genetics

Common variants at *ABCA7*, *MS4A6A/MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease

We sought to identify new susceptibility loci for Alzheimer's disease through a staged association study (GERAD+) and by testing suggestive loci reported by the Alzheimer's Disease Genetic Consortium (ADGC) in a companion paper. We undertook a combined analysis of four genome-wide association datasets (stage 1) and identified ten newly associated variants with $P \leq 1 \times 10^{-5}$. We tested these variants for association in an independent sample (stage 2). Three SNPs at two loci replicated and showed evidence for association in a further sample (stage 3). Meta-analyses of all data provided compelling evidence that *ABCA7* (rs3764650, meta $P = 4.5 \times 10^{-17}$; including ADGC data, meta $P = 5.0 \times 10^{-21}$) and the *MS4A* gene cluster (rs610932, meta $P = 1.8 \times 10^{-14}$; including ADGC data, meta $P = 1.2 \times 10^{-16}$) are new Alzheimer's disease susceptibility loci. We also found independent evidence for association for three loci reported by the ADGC, which, when combined, showed genome-wide significance: *CD2AP* (GERAD+, $P = 8.0 \times 10^{-4}$; including ADGC data, meta $P = 8.6 \times 10^{-9}$), *CD33* (GERAD+, $P = 2.2 \times 10^{-4}$; including ADGC data, meta $P = 1.6 \times 10^{-9}$) and *EPHA1* (GERAD+, $P = 3.4 \times 10^{-4}$; including ADGC data, meta $P = 6.0 \times 10^{-10}$).

Alzheimer's disease is the most common form of dementia, with both environmental and genetic factors contributing to risk. Alzheimer's disease is genetically complex and shows heritability of up to 79% (ref. 1). Rare variants in three genes (*APP*, *PSEN1* and *PSEN2*)[1] cause disease in a minority of cases, but until recently, *APOE* (encoding apolipoprotein E) was the only gene known to increase disease risk for the common form of Alzheimer's disease with late onset[2]. In 2009, we published a genome-wide association study (GWAS) of Alzheimer's disease in a sample designated GERAD1 (Genetic and Environmental Risk in Alzheimer's Disease Consortium 1), in which we identified two new genome-wide significant susceptibility loci: clusterin (*CLU*: $P = 8.5 \times 10^{-10}$) and the phosphatidylinositol-binding clathrin assembly protein gene (*PICALM*: $P = 1.3 \times 10^{-9}$). We also observed more variants with $P < 1 \times 10^{-5}$ than were expected by chance ($P = 7.5 \times 10^{-6}$)[3]. These included variants in *CR1* (the complement receptor 1 gene), *BIN1* (the bridging integrator 1 gene) and the *MS4A* (membrane-spanning 4A gene) cluster. A second independent Alzheimer's disease GWAS[4] using the EADI1 (European Alzheimer's Disease Initiative 1) sample showed genome-wide significant evidence for association with *CLU* ($P = 7.5 \times 10^{-9}$) and *CR1* ($P = 3.7 \times 10^{-9}$) and support for association with *PICALM* ($P = 3 \times 10^{-3}$). A combined analysis of the GERAD1 and EADI1 data yielded highly significant support for all three loci (*CLU* meta $P = 6.7 \times 10^{-16}$; *PICALM* meta $P = 6.3 \times 10^{-9}$; and *CR1* meta $P = 3.2 \times 10^{-12}$). The associations in *CLU*, *PICALM* and *CRI* have since been replicated in several independent datasets[5–8], have shown trends in another dataset[9] and have shown relationships with the neurodegenerative processes underlying disease[10]. In addition, members of this consortium have since reported genome-wide significant association for *BIN1* ($P = 1.6 \times 10^{-11}$) and support for *EPHA1* (encoding ephrin receptor A1) ($P = 1.7 \times 10^{-6}$)[11].

This study sought to identify new common susceptibility variants for Alzheimer's disease by first undertaking a three-stage association study based upon predominantly European samples (GERAD+; **Fig. 1**) and then by testing these samples for loci showing suggestive evidence for association in the ADGC GWAS[12].

The first stage of this study comprised a meta-analysis of four Alzheimer's disease GWAS datasets (6,688 affected individuals (cases) and 13,685 controls) including: the GERAD1 (ref. 3), EADI1 (ref. 4), Translational Genomics Research Institute (TGEN1)[13] and the Alzheimer's Disease Neuroimaging Initiative (ADNI)[14] datasets. SNPs which remained significant at $P \leq 1 \times 10^{-5}$ were then tested for replication in the second stage of this study, comprising 4,896 cases and 4,903 controls, including genotyping of the GERAD2 sample and *in silico* replication in the deCODE and German Alzheimer's Disease Integrated Genome Research Network (AD-IG) GWAS datasets. In stage 3, previously unidentified SNPs showing significant evidence of replication in stage 2 were then tested for association in a sample comprising 8,286 cases and 21,258 controls, which included new genotyping in the EADI2 (ref. 4) and Mayo2 samples and *in silico* replication in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) sample[11]. Sample descriptions and characteristics can be found in the **Supplementary Note** and **Supplementary Table 1**.

In stage 1, we identified 61 SNPs associated with Alzheimer's disease at $P \leq 1 \times 10^{-5}$ following meta-analysis of 496,763 SNPs in the GERAD1, TGEN1, ADNI and EADI1 samples (**Supplementary Table 2** and **Supplementary Note**). Ten SNPs at newly associated loci and two SNPs at previously identified susceptibility loci that surpassed the $P \leq 1 \times 10^{-5}$ threshold were selected for further analysis (see below). One SNP, rs610932 (stage 1 $P = 1.8 \times 10^{-8}$) at the *MS4A* (encoding membrane spanning 4A) gene cluster, surpassed the threshold for genome-wide significance ($P < 5.0 \times 10^{-8}$)[15]. We also observed strong evidence for association at *ABCA7* (encoding ATP-binding cassette, sub-family A, member 7; rs3764650; stage 1 $P = 2.6 \times 10^{-7}$).

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

Received 9 September 2010; accepted 10 March 2011; published online 3 April 2011; doi:10.1038/ng.803

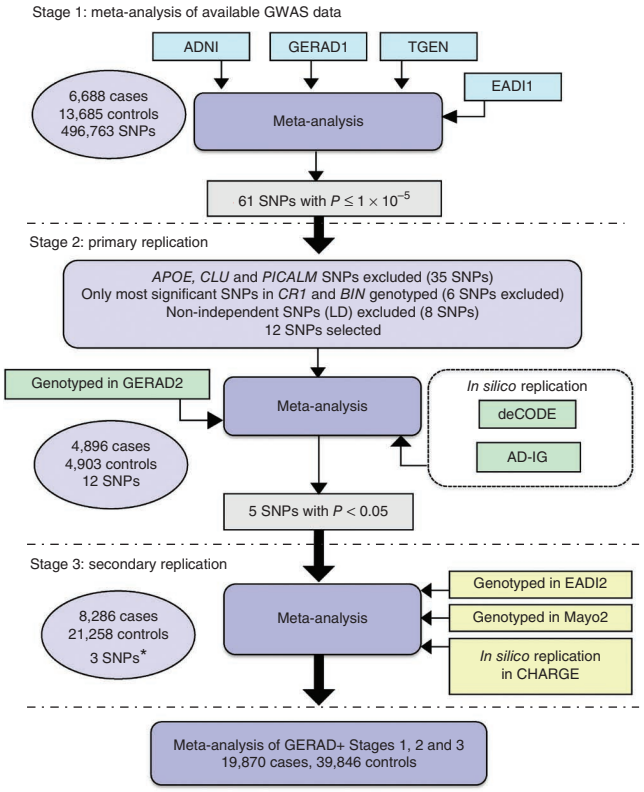

Figure 1 GERAD+ study design. *Data for rs744373 and rs3818361 in the CHARGE consortium have been presented elsewhere[15], as has data for rs381861 in the EADI2 samples[4]; as such these SNPs were not included in stage 3.

When selecting SNPs for testing in stage 2, we excluded known susceptibility loci that had previously been tested in GERAD2, and we limited our analysis of *BIN1* and *CR1*, which had not been tested in GERAD2, to the most significant SNPs at each locus (**Supplementary Table 2**). Following pruning for linkage disequilibrium (LD), 12 SNPs were taken forward for replication in stage 2 (10 SNPs, excluding *BIN1* and *CR1*).

Five of the 12 SNPs tested in stage 2 showed significant evidence for replication using a Bonferroni-adjusted threshold for significance of $P = 4.2 \times 10^{-3}$ (**Table 1** and **Supplementary Table 3**). In addition to SNPs at *BIN1* and *CR1*, one SNP within *ABCA7* (rs3764650, stage 2 $P = 1.9 \times 10^{-5}$) and two SNPS at the *MS4A* gene cluster (rs610932, stage 2 $P = 1.6 \times 10^{-3}$; and rs670139, stage 2 $P = 1.1 \times 10^{-3}$) showed evidence of replication in stage 2. We tested the three SNPs implicating new risk loci for association in the stage 3 sample and showed further evidence of replication (rs3764650, stage 3 $P = 2.9 \times 10^{-7}$; rs610932, stage 3 $P = 2.1 \times 10^{-5}$; and rs670139, stage 3 $P = 3.2 \times 10^{-3}$; **Table 1** and **Supplementary Table 3**).

We conducted an inverse variance weighted meta-analysis of data from stages 1, 2 and 3 (**Table 1** and **Supplementary Table 3**). This provided strong evidence for association with rs3764650 at *ABCA7* (meta $P = 4.5 \times 10^{-17}$) and two SNPs at the *MS4A* gene cluster: rs610932 (meta $P = 1.8 \times 10^{-14}$) and rs670139 (meta $P = 1.4 \times 10^{-9}$). When combining GERAD+ and ADGC results (after removing overlapping samples), *ABCA7* had $P = 5.0 \times 10^{-21}$ (odds ratio (OR) = 1.22). The two SNPs at the *MS4A* gene cluster, rs610932 and rs670139, showed $P = 1.2 \times 10^{-16}$ (OR = 0.91) and $P = 1.1 \times 10^{-10}$ (OR = 1.08), respectively, in the combined analysis of the GERAD+ and ADGC results. It is noteworthy that the most significant ADGC SNP at the *MS4A* locus is in LD with our top SNP (rs4938933 with rs610932, $r^2 = 0.62$, $D' = 0.86$), and thus both datasets may be detecting the same underlying signal.

This study also provides additional independent support for association with *CR1* (stage 2 $P = 1.4 \times 10^{-3}$) and *BIN1* (stage 2 $P = 3.8 \times 10^{-5}$; see **Table 1** for meta-analysis.) We did not observe interaction between *APOE* and the new variants identified in this study, and indeed, we did not find evidence of epistasis between any of the genome-wide significant variants identified to date (*ABCA7*, *MS4A*, *BIN1*, *CR1*, *PICALM*, *CLU* or *APOE*) (**Supplementary Table 4a**). Likewise, adjusting for the presence of at least one *APOE* ε4 allele had little effect on the results of the analysis of the three newly associated variants (**Supplementary Table 4b**). We also found no evidence for association between these loci and age at onset of Alzheimer's disease (rs3764650, $P = 0.17$; rs670139, $P = 0.38$; rs610932, $P = 0.95$; rs744373, $P = 0.87$; and rs3818361, $P = 0.58$).

This study therefore identifies two new Alzheimer's disease susceptibility loci, which replicate over a number of independent case-control samples. The first of these is the *ABCA7* (encoding ATP-binding cassette, sub-family A, member 7) locus (**Fig. 2a**). The associated marker is rs3764650, which is located in intron 13. This SNP was the only variant in the gene that passed our stage 1 criterion, which is not unexpected given the low levels of LD between this SNP and others included in the GWAS. However, in a preliminary attempt to identify an associated functional variant at the *ABCA7* locus, we genotyped the GERAD2 sample for rs3752246, a non-synonymous SNP in exon 32 of the gene, which showed the highest LD with rs3764650 out of all HapMap *ABCA7*-coding variants based on r^2 values ($r^2 = 0.36$, $D' = 0.89$). This variant (which was not genotyped in stage 1) was also associated with Alzheimer's disease (GERAD2 $P = 1 \times 10^{-3}$, OR = 1.17). rs3752246 encodes a glycine to alanine substitution at position 1527 of the protein, which is predicted to be a benign change[16] and is unlikely to be the relevant functional variant. We used data from two published expression quantitative trait loci (eQTL) datasets (derived from lymphoblastoid cell lines[17] and brain[18]) to determine if rs3764650 is associated with the expression of *ABCA7*. However,

Table 1 Results of the GERAD+ study

SNP	Closest gene	Chr.	MAF	Stage 1[a]			Stage 2[b]			Stage 3[c]			Meta-analysis of GERAD+ and stage 1, 2 and 3[d]			Meta-analysis of GERAD+ and ADGC		
				P	OR	95% CI	*P*	OR	95% CI	*P*	OR	95% CI	*P*	OR	95% CI	*P*	OR	95% CI
rs3764650	*ABCA7*	19	0.10	2.6×10^{-7}	1.22	1.13–1.32	1.9×10^{-5}	1.28	1.14–1.44	2.9×10^{-7}	1.22	1.13–1.32	4.5×10^{-17}	1.23	1.18–1.30	5.0×10^{-21}	1.23	1.17–1.28
rs610932	*MS4A6A*	11	0.42	1.8×10^{-8}	0.88	0.85–0.92	1.6×10^{-3}	0.90	0.84–0.96	2.1×10^{-5}	0.91	0.87–0.95	1.8×10^{-14}	0.90	0.87–0.92	1.2×10^{-16}	0.91	0.88–0.93
rs670139	*MS4A4E*	11	0.41	1.0×10^{-5}	1.11	1.06–1.16	1.1×10^{-3}	1.11	1.04–1.19	3.2×10^{-3}	1.06	1.02–1.11	1.4×10^{-9}	1.09	1.06–1.12	1.1×10^{-10}	1.08	1.06–1.11
rs3818361	*CR1*	1	0.19	3.2×10^{-12}	1.21	1.14–1.27	1.4×10^{-3}	1.14	1.05–1.23	NA	NA	NA	3.7×10^{-14}	1.18	1.13–1.24	NA	NA	NA
rs744373	*BIN1*	2	0.29	1.5×10^{-10}	1.17	1.11–1.22	3.8×10^{-5}	1.17	1.08–1.25	NA	NA	NA	2.6×10^{-14}	1.17	1.12–1.21	NA	NA	NA

Chr., chromosome; MAF, minor allele frequency.
[a]GERAD1, EADI1, ADNI and TGEN1, less than 6,688 cases and 13,685 controls. [b]GERAD2, deCODE and AD-IG, 4,896 Alzheimer's disease cases and 4,903 controls. [c]EADI2, CHARGE and Mayo2, less than 8,286 Alzheimer's disease cases and 21,258 controls. [d]GERAD1 and GERAD2, EADI1 and EADI2, ADNI, TGEN1, deCODE, AD-IG, CHARGE and Mayo2, less than 19,870 Alzheimer's disease cases and 39,846 controls.

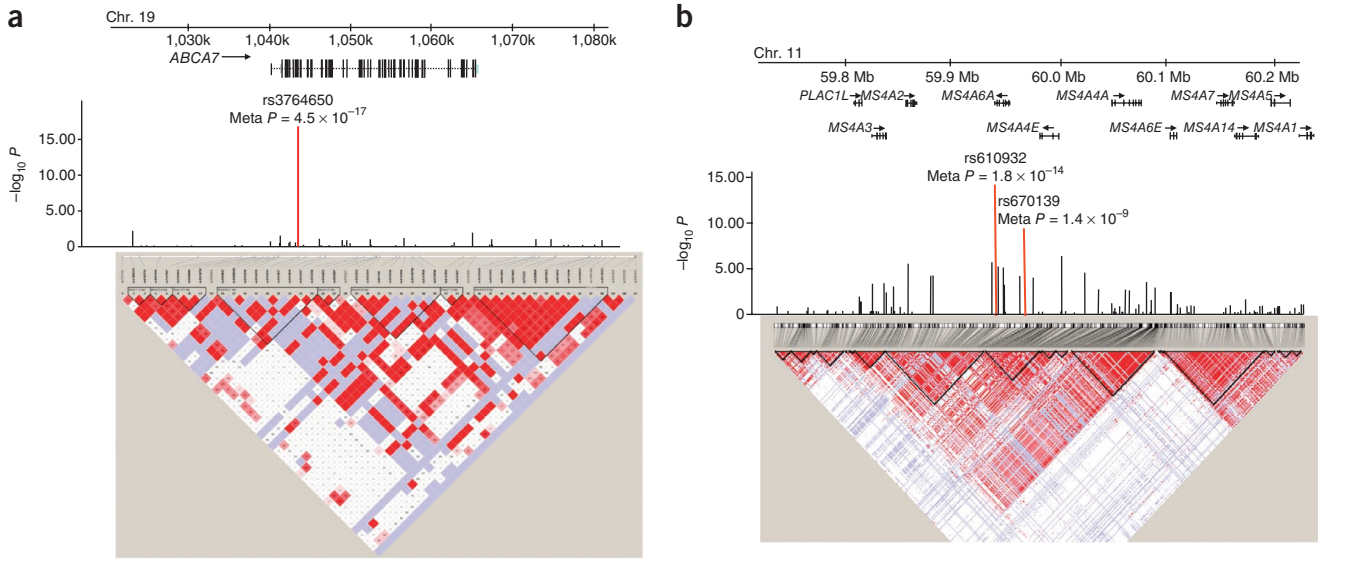

Figure 2 Schematic of the associated variants reported in reference to (**a**) *ABCA7* and (**b**) chromosomal region chr11: 59.81Mb–60.1Mb harboring members of the *MS4A* gene cluster. Chromosome positions are shown at the top of the schematics (UCSC Feb 2009). Gene schematic: horizontal arrows indicate directions of transcription, black boxes indicate gene exons and the untranslated region. The $-\log_{10} P$ of the SNPs analyzed in stage 1 are shown in the chart and graph. The GERAD+ stage 1, 2 and 3 meta-analysis *P* values for rs3764650 (*ABCA7*), rs610932 (*MS4A6A*) and rs670139 (*MS4A4E*) are indicated by the red lines. The *D′* LD block structure of *ABCA7* plus the surrounding region and chr11: 59.81Mb–60.1Mb according to the CEPH HapMap data are provided at the bottom of each schematic with lines indicating where each SNP genotyped on the Illumina 610-quad chip is represented.

we observed no association (**Supplementary Table 5**). Further work will be required to identify the causal variant(s) at this locus.

Second, we implicate the *MS4A* (encoding membrane-spanning 4A) gene cluster (**Fig. 2b**). The association spans an LD block of 293 kb (chr11: 59,814,287–60,107,105) and includes 6 of 16 known genes comprising the membrane-spanning 4-domains subfamily A (encoded by *MS4A*). These are *MS4A2*, *MS4A3*, *MS4A4A*, *MS4A4E*, *MS4A6A* and *MS4A6E*. The associated SNPs are found in the 3′ untranslated region of *MS4A6A* (rs610932) and the intergenic region between *MS4A4E* and *MS4A6A* (rs670139). rs610932 showed nominally significant association with expression levels of *MS4A6A* in cerebellum and temporal cortex ($0.01 < P < 0.05$; **Supplementary Table 5**) but not in the frontal cortex, pons or lymphoblastoid cell lines. The non-synonymous SNP that was most strongly associated with the genome-wide significant variants was rs2304933. This SNP was analyzed in stage 1 but showed weaker evidence for association ($P = 0.006$) than the genome-wide significant variant at this locus in the same sample. **Figure 3** shows forest plots depicting association in the different datasets for SNPs at the *ABCA7* (rs3764650) and *MS4A* (rs610932 and rs670139) loci.

We also sought to follow up four additional loci showing suggestive evidence for association with Alzheimer's disease ($1 \times 10^{-6} \geq P > 5 \times 10^{-8}$) from the ADGC GWAS[12]. These loci included those in *CD33*, *EPHA1*, *CD2AP* and *ARID5B*. It should be noted that evidence for suggestive association with *EPHA1* and *CD33* has been reported previously. Members of this collaboration were the first to report *EPHA1* as showing suggestive evidence of association with Alzheimer's disease (rs11771145, $P = 1.7 \times 10^{-6}$; LD with ADGC rs11767557, $r^2 = 0.28$, $D' = 0.75$)[11], which included the GERAD1 and EADI1 samples reported on here. Similarly, researchers from another study were the first to show suggestive evidence for *CD33* (rs3826656, $P = 4.0 \times 10^{-6}$; LD with ADGC rs3865444, $r^2 = 0.13$, $D' = 1.0$)[19].

We combined data from the GERAD+ dataset, comprising the GERAD1, EADI1, deCODE and AD-IG GWAS datasets (up to 6,992 cases and 13,472 controls), using inverse variance meta-analysis. We included the TGEN1, ADNI and Mayo1 datasets in the ADGC discovery set and were thus excluded from these particular analyses. We observed support for association with *CD2AP* (rs9349407, $P = 8.0 \times 10^{-4}$, OR = 1.11), *CD33* (rs3865444, $P = 2.2 \times 10^{-4}$, OR = 0.89) and *EPHA1* (rs11767557, $P = 3.4 \times 10^{-4}$, OR = 0.90).

When these data were combined with the ADGC data, we observed genome-wide evidence for association with Alzheimer's disease (rs9349407, GERAD+ and ADGC meta $P = 8.6 \times 10^{-9}$, OR = 1.11; rs3865444, GERAD+ and ADGC meta $P = 1.6 \times 10^{-9}$, OR = 0.91; rs11767557, GERAD+ and ADGC meta $P = 6.0 \times 10^{-10}$, OR = 0.90). We observed nominally significant evidence of association with *ARID5B* (rs2588969, $P = 3.3 \times 10^{-2}$, OR = 1.06); however, the direction of effect was opposite to that reported by ADGC[12] and was not significant overall (GERAD+ and ADGC meta $P = 3.6 \times 10^{-1}$, OR = 0.99). See **Table 2** for results of the GERAD+ and ADGC combined analyses and **Supplementary Table 6** for results of additional SNPs at these loci.

Taken together, these results show compelling evidence for an additional five Alzheimer's disease susceptibility loci. *ABCA7* encodes an ATP-binding cassette (ABC) transporter. The ABC transporter superfamily has roles in transporting a wide range of substrates across cell membranes[20]. *ABCA7* is highly expressed in brain, particularly in the hippocampal CA1 neurons[21] and microglia[22]. ABCA7 is involved in the efflux of lipids from cells to lipoprotein particles. Notably, the main lipoproteins in brain are APOE followed by CLU. We observed no evidence for epistatic interactions between the three genetic loci (**Supplementary Table 4a**), however, this is not a prerequisite for biological interaction between these molecules. In addition, ABCA7 has been shown to regulate APP processing and inhibit β-amyloid secretion in cultured cells overexpressing *APP* (ref. 23). *ABCA7* also modulates phagocytosis of apoptotic cells by macrophages mediated through the C1q complement receptor protein on the apoptotic cell surface[23]. *ABCA7* is an ortholog of *Caenorhabditis elegans ced-7*, the product of which is known to clear apoptotic cells, and the high levels of expression of *ABCA7* in microglia are consistent with such a role.

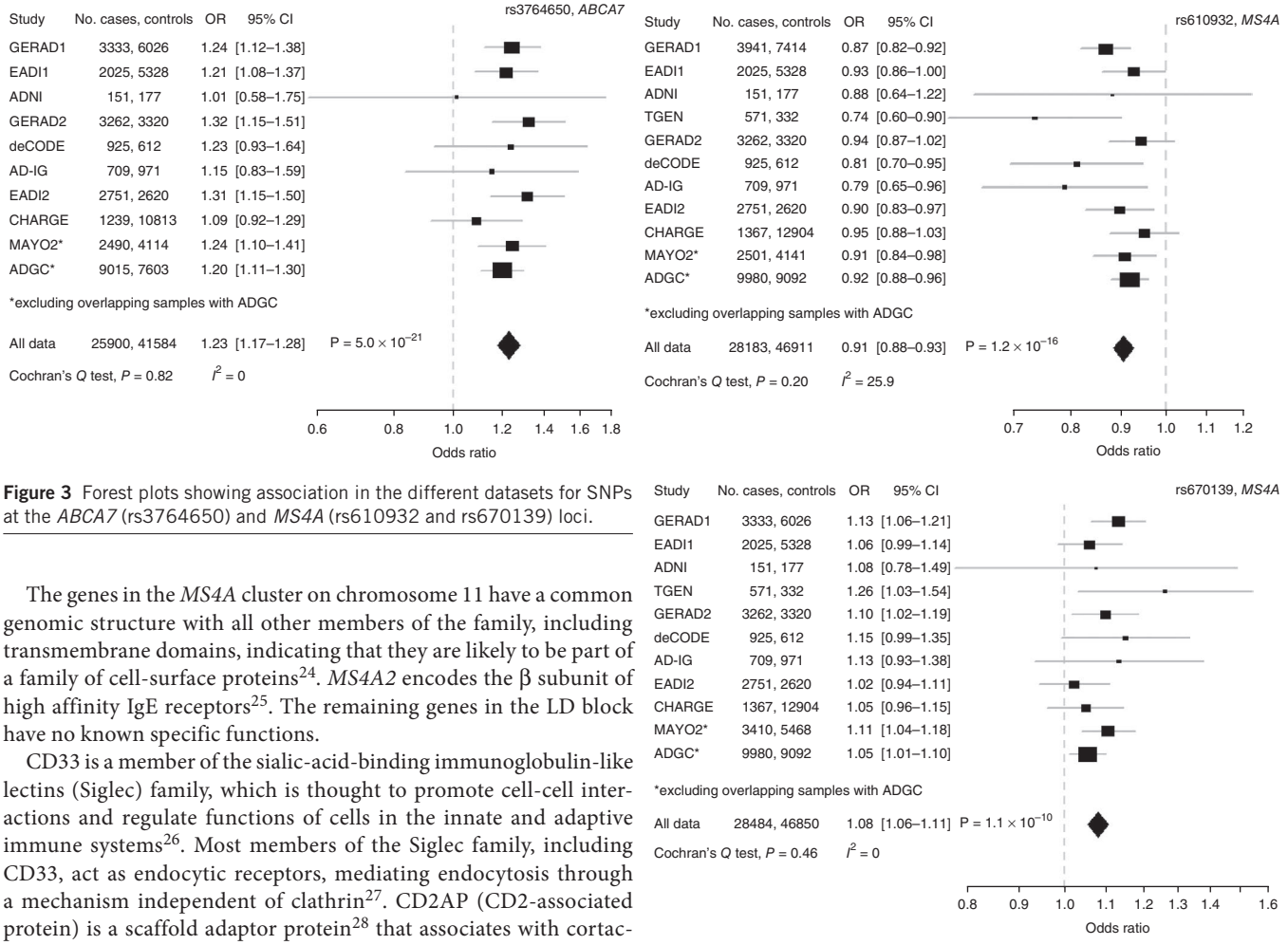

Figure 3 Forest plots showing association in the different datasets for SNPs at the *ABCA7* (rs3764650) and *MS4A* (rs610932 and rs670139) loci.

The genes in the *MS4A* cluster on chromosome 11 have a common genomic structure with all other members of the family, including transmembrane domains, indicating that they are likely to be part of a family of cell-surface proteins[24]. *MS4A2* encodes the β subunit of high affinity IgE receptors[25]. The remaining genes in the LD block have no known specific functions.

CD33 is a member of the sialic-acid-binding immunoglobulin-like lectins (Siglec) family, which is thought to promote cell-cell interactions and regulate functions of cells in the innate and adaptive immune systems[26]. Most members of the Siglec family, including CD33, act as endocytic receptors, mediating endocytosis through a mechanism independent of clathrin[27]. CD2AP (CD2-associated protein) is a scaffold adaptor protein[28] that associates with cortactin, a protein also involved in the regulation of receptor-mediated endocytosis[29]. It is striking that these two new susceptibility genes for Alzheimer's disease, and the recently established susceptibility genes *PICALM* and *BIN1*, are all implicated in cell-cell communication and transduction of molecules across the membrane. EPHA1 is a member of the ephrin receptor subfamily. Ephrins and Eph receptors are membrane-bound proteins which play roles in cell and axon guidance[30] and in synaptic development and plasticity[31]. However, EPHA1 is expressed mainly in epithelial tissues[32] where it regulates cell morphology and motility[33]. Additional roles in apoptosis[34] and inflammation[35] have also been proposed.

Our study has identified variants at *ABCA7* and the *MS4A* gene cluster associated with susceptibility to Alzheimer's disease with replication over a number of independent case-control samples. We also provide independent support for three loci showing suggestive evidence in a companion paper[12]: loci in *CD33*, *CD2AP* and *EPHA1*, which, when the data were combined, showed genome-wide levels of significance. Finally, we provide further replication evidence for *BIN1* and *CR1* loci as Alzheimer's disease susceptibility loci. What is striking about our findings is the emerging consistency in putative function of the genes identified. Five of the recently identified Alzheimer's disease susceptibility loci in *CLU*, *CR1*, *ABCA7*, *CD33* and *EPHA1* have putative functions in the immune system; *PICALM*, *BIN1*, *CD33* and *CD2AP* are involved in processes at the cell membrane, including endocytosis, and *APOE*, *CLU* and *ABCA7* are involved in lipid processing. It is conceivable that these processes would play strong roles in neurodegeneration and Aβ clearance from the brain. These findings therefore provide new impetus for focused studies aimed at understanding the pathogenesis of Alzheimer's disease.

Table 2 Results of the combined analysis of the GERAD+ consortia with the ADGC GWAS ×

				LD with the top ADGC SNP at each loci		GERAD+ consortia[a]					GERAD+ and ADGC meta-analysis		
SNP	Closest gene	Chr.	MAF	r^2	D'	Cases	Controls	P	OR	95% CI	P	OR	95% CI
rs9349407[b]	*CD2AP*	6	0.29	N/A	N/A	6,283	7,165	8.0×10^{-4}	1.11	1.04–1.18	8.6×10^{-9}	1.11	1.07–1.15
rs9296559[c]	*CD2AP*	6	0.29	0.71	0.95	6,283	7,165	1.5×10^{-3}	1.10	1.04–1.17	NA	NA	NA
rs11767557	*EPHA1*	7	0.21	N/A	N/A	6,283	12,935	3.4×10^{-4}	0.90	0.85–0.95	6.0×10^{-10}	0.90	0.86–0.93
rs2588969[b]	*ARID5B*	10	0.40	N/A	N/A	6,283	7,165	3.3×10^{-2}	1.06[c]	1.01–1.13	3.6×10^{-1}	0.99	0.95–1.02
rs4948288	*ARID5B*	10	0.26	0.55	0.78	6,992	13,472	3.6×10^{-3}	1.07[c]	1.03–1.15	NA	NA	NA
rs3865444[d]	*CD33*	19	0.31	N/A	N/A	6,283	7,165	2.2×10^{-4}	0.89	0.84–0.95	1.6×10^{-9}	0.91	0.88–0.93

Chr., chromosome; MAF, minor allele frequency; LD, linkage disequilibrium.
[a]GERAD1, EADI1, deCODE and AD-IG. [b]Results generated from imputed data. The results from the top genotyped SNP are also shown. See **Supplementary Table 6** for full details. [c]Opposite direction of effect to that reported by Naj *et al.*[12]. [d]Data imputed in the deCODE dataset.

URLs. ADNI database, http://www.loni.ucla.edu/ADNI; ADNI investigators, http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Authorship_List.pdf.

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

For complete acknowledgments, please see the **Supplementary Note**. We thank the individuals and families who took part in this research and those who funded the groups who contributed to this study: Wellcome Trust; Medical Research Council (UK); Alzheimer's Research UK; the Welsh Assembly Government; Mercer's Institute for Research on Ageing; Alzheimer's Society; Ulster Garden Villages; Northern Ireland R&D Office; Royal College of Physicians; Dunhill Medical Trust; Bristol Research into Alzheimer's and Care of the Eldery; US National Institutes of Health, the Barnes Jewish Foundation; Charles and Joanne Knight Alzheimer's Research Initiative; University College London (UCL) Hospital/UCL Biomedical Centre; Lundbeck; German Federal Ministry of Education and Research Competence Network Dementia and Competence Network Degenerative Dementia; Alfried Krupp von Bohlen und Halbach-Stiftung; Intramural Research Program of the National Institute of Ageing Department of Health and Human Services; University of Antwerp, Fund for Scientific Research-Flanders; Foundation for Alzheimer Research; Methusalem Excellence grant; Federal Science Policy Office Interuniversity Attraction Poles program; Mayo AD Research Center; National Institute of Neurological Disorders and Stroke; Robert and Clarice Smith Postdoctoral Fellowship and AD Research Program; Palumbo Professorship in AD Research; Carl Edward and Susan Bass Bolch Gift; Institut Pasteur de Lille; Centre National de Génotypage; Fondation pour la Recherche Médicale Caisse; Nationale Maladie des Travailleurs Salariés, Direction Générale de la Sant; Institut de la Longévité; Agence Française de Sécurité Sanitaire des Produits de Santé; Aquitaine and Bourgogne Regional Councils; Fondation de France; French Ministry of Research/Institute national de la santé et de la recherché medicale; Eisai; Health Research Council of the Academy of Finland; Nordic Centre of Excellence in Neurodegeneration; Italian Ministry of Research; Carimonte Foundation; Italian Ministry of Health; Fondazione Monzino; Ministerio de Educación y Ciencia the Ministerio de Sanidad y Consumo; Fundación Ramón Areces; National Institute of Biomedical Imaging Bioengineering; Abbott; AstraZeneca, Bayer Schering Pharma; Bristol-Myers Squibb; Elan; Genentech; GE; GlaxoSmithKline; Innogenetics; Johnson and Johnson; Eli Lilly; Medpace; Merck; Novartis; Pfizer; Hoffman-La Roche; Schering-Plough; Synarc; Wyeth; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; US Food and Drug Administration; Foundation for the National Institutes of Health Northern California Institute for Research and Education; Dana Foundation; German National Genome Research Network; German Ministry for Education and Research; National Institute on Deafness and Other Communication Disorders; Hjartavernd; Althingi; National Heart Lung and Blood Institute; National Institute of Diabetes and Digestive and Kidney Diseases; Robert Dawson Evans Endowment; Netherlands Organisation of Scientific Research; Netherlands Genomics Initiative; Erasmus Medical Center; Netherlands organization for scientific research; Netherlands Organization for the Health Research and Development; the Research Institute for Diseases in the Elderly; Ministry of Education, Culture and Science; Ministry for Health, Welfare and Sports; European Commission and the Municipality of Rotterdam. G.W. was partly funded by the NIHR Biomedical Research Centre Programme, Oxford. The Leeds sample collection was funded by the Health Foundation.

AUTHOR CONTRIBUTIONS

J. Williams directed this study, assisted by M.J.O. and M.O.D. and was also helped by P.H., R.S., A.G., R.A., L.J. and D. Harold. J. Williams, P.H. and D.H. took primary responsibility for drafting the manuscript, assisted by R.S., A.G., R.A., M.O.D. and M.J.O. All authors contributed to the sample collection, sample preparation, genotyping and/or conduct of the GWAS upon which this study is based. J. Williams, R.A., P.H., R.S., A.G., C.W., J. Chapman, K.D., N.J., A.S., C. Thomas, S. Lovestone, J.P., P. Proitsi, M.K.L., C. Brayne, D.C.R., M.G., B.L., A.L., K. Morgan, K.S.B., P.A.P., D. Craig, B.M., S.T., C.H., D.M., A.D.S., S. Love, P.G.K., J.H., S. Mead, N.C.F., M. Rossor, J. Collinge, W.M., F.J., B.S., E.R., R.H., H.K., H.v.d.B., I.H., J.K., J. Wiltfang, M. Dichgans, L.F., H.H., M. Hüll, J.G., A.M.G., D.R., I.G., J.S.K.K., C.C., P.N., J.C.M., K. Mayo, K. Sleegers, K.B., S.E., P.P.D., C.V.B., G.L., N.J.B., H.G., A.M., M.T., T.W.M., M.M.N., S. Moebus, K.-H.J., N.K. and H.-E.W. contributed to clinical sample collection, ascertainment, diagnosis and preparation of samples from the independent GERAD2 sample genotyped as part of this study. R.S., D. Harold, A.G., D.R. and I.G. were responsible for procedures related to genotyping the GERAD2 sample. V.C., B.G.-B., M. Hiltunen, O.C., D.Z., M. Delepine, M.J.B., F. Pasquier, I.M., A.F.-G., E.P., O.H., E. Coto, V.A., P. Bosco, G.S., M. Mancuso, F. Panza, B.N., S. Sorbi, P. Bossù, P. Piccardi, B.A., G.A., D.S., E.S., D.G., A.B., D. Hannequin, F.L., H. Soininen, J.-C.L. and P.A. were responsible for sample collection, sample preparation, genotyping and analysis of the EADI2 Sample. S.S., A.L.D., O.L. and L.J.L., as well as M.A.I., C.M.v.D. and M.M.B.B. contributed clinical and genotypic data to the CHARGE GWAS. J.-C.L. and P.A. contributed clinical and genotypic data. M.M.C. played a leading role, along with H.B., D.W., G.W., N.M.H., E.R.L.C.V., S.B.S., J.O.A., M.B., Z.K.W., D.W.D., N.R.G.R., R.C.P., K. Morgan and S.G.Y. in sample collection, sample preparation, genotyping and analysis of the Mayo2 sample. M. Riemenschneider, T.M.F., P.F., C.R., M.K., S. Schreiber, M. Mayhaus, S.N. and S.W. were responsible for sample collection, conduct and analysis of the AD-IG GWAS. S. Steinberg, T.J., H. Stefansson, K. Stefansson, J.S., S.B. and P.V.J were responsible for sample collection, conduct and analysis of the deCODE GWAS. D. Harold and P.H. completed statistical quality control and produced association statistics under the supervision of J. Williams and P.A.H. All authors discussed the results and approved the manuscript.

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/.

1. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
2. Saunders, A.M. *et al.* Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**, 1467–1472 (1993).
3. 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).
4. Lambert, J.C. *et al.* Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1094–1099 (2009).
5. Corneveaux, J.J. *et al.* Association of *CR1*, *CLU* and *PICALM* with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *Hum. Mol. Genet.* **19**, 3295–3301 (2010).
6. Zhang, Q. *et al.* Complement receptor 1 polymorphisms and risk of late onset Alzheimer's disease. *Brain Res.* **1348**, 216–221 (2010).
7. Carrasquillo, M.M. *et al.* Replication of *CLU*, *CR1*, and *PICALM* associations with alzheimer disease. *Arch. Neurol.* **67**, 961–964 (2010).
8. Jun, G. *et al.* Meta-analysis confirms *CR1*, *CLU*, and *PICALM* as Alzheimer disease risk loci and reveals interactions with *APOE* genotypes. *Arch. Neurol.* **67**, 1473–1484 (2010).
9. Kamboh, M.I. *et al.* Association of *CLU* and *PICALM* variants with Alzheimer's disease. *Neurobiol. Aging* published online, doi:10.1016/j.neurobiolaging.2010.04.015 (4 June 2010).
10. Biffi, A. *et al.* Genetic variation and neuroimaging measures in Alzheimer disease. *Arch. Neurol.* **67**, 677–685 (2010).
11. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *J. Am. Med. Assoc.* **303**, 1832–1840 (2010).
12. Naj, A.C. *et al.* Common variants in *MS4A4/MS4A6E*, *CD2AP*, *CD33* and *EPHA1* are associated with late-onset Alzheimer's disease. *Nat. Genet.* advance online publication, doi:10.1038/ng.801 (3 April 2011).
13. Reiman, E.M. *et al.* GAB2 alleles modify Alzheimer's risk in *APOE* ε4 carriers. *Neuron* **54**, 713–720 (2007).
14. Petersen, R.C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201–209 (2010).
15. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
16. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
17. Stranger, B.E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1**, e78 (2005).
18. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
19. 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).
20. Kim, W.S., Weickert, C.S. & Garner, B. Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. *J. Neurochem.* **104**, 1145–1166 (2008).
21. Kim, W.S. *et al. Abca7* null mice retain normal macrophage phosphatidylcholine and cholesterol efflux activity despite alterations in adipose mass and serum cholesterol levels. *J. Biol. Chem.* **280**, 3989–3995 (2005).
22. Kim, W.S., Guillemin, G.J., Glaros, E.N., Lim, C.K. & Garner, B. Quantitation of ATP-binding cassette subfamily-A transporter gene expression in primary human brain cells. *Neuroreport* **17**, 891–896 (2006).

23. Jehle, A.W. *et al.* ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. *J. Cell Biol.* **174**, 547–556 (2006).
24. Liang, Y., Buckley, T.R., Tu, L., Langdon, S.D. & Tedder, T.F. Structural organization of the human *MS4A* gene cluster on chromosome 11q12. *Immunogenetics* **53**, 357–368 (2001).
25. Kinet, J.P. *et al.* Isolation and characterization of cDNAs coding for the beta subunit of the high-affinity receptor for immunoglobulin E. *Proc. Natl. Acad. Sci. USA* **85**, 6483–6487 (1988).
26. Crocker, P.R., Paulson, J.C. & Varki, A. Siglecs and their roles in the immune system. *Nat. Rev. Immunol.* **7**, 255–266 (2007).
27. Tateno, H. *et al.* Distinct endocytic mechanisms of CD22 (Siglec-2) and Siglec-F reflect roles in cell signaling and innate immunity. *Mol. Cell. Biol.* **27**, 5699–5710 (2007).
28. Dustin, M.L. *et al.* A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts. *Cell* **94**, 667–677 (1998).
29. Lynch, D.K. *et al.* A Cortactin-CD2-associated protein (CD2AP) complex provides a novel link between epidermal growth factor receptor endocytosis and the actin cytoskeleton. *J. Biol. Chem.* **278**, 21805–21813 (2003).
30. Martinez, A. & Soriano, E. Functions of ephrin/Eph interactions in the development of the nervous system: emphasis on the hippocampal system. *Brain Res. Brain Res. Rev.* **49**, 211–226 (2005).
31. Lai, K.O. & Ip, N.Y. Synapse development and plasticity: roles of ephrin/Eph receptor signaling. *Curr. Opin. Neurobiol.* **19**, 275–283 (2009).
32. Coulthard, M.G. *et al.* Characterization of the Epha1 receptor tyrosine kinase: expression in epithelial tissues. *Growth Factors* **18**, 303–317 (2001).
33. Yamazaki, T. *et al.* EphA1 interacts with integrin-linked kinase and regulates cell morphology and motility. *J. Cell Sci.* **122**, 243–255 (2009).
34. Duffy, S.L. *et al.* Generation and characterization of EphA1 receptor tyrosine kinase reporter knockout mice. *Genesis* **46**, 553–561 (2008).
35. Ivanov, A.I. & Romanovsky, A.A. Putative dual role of ephrin-Eph receptor interactions in inflammation. *IUBMB Life* **58**, 389–394 (2006).

Paul Hollingworth[1,110], Denise Harold[1,110], Rebecca Sims[1,110], Amy Gerrish[1,110], Jean-Charles Lambert[2–4,110], Minerva M Carrasquillo[5,110], Richard Abraham[1], Marian L Hamshere[1], Jaspreet Singh Pahwa[1], Valentina Moskvina[1], Kimberley Dowzell[1], Nicola Jones[1], Alexandra Stretton[1], Charlene Thomas[1], Alex Richards[1], Dobril Ivanov[1], Caroline Widdowson[1], Jade Chapman[1], Simon Lovestone[6,7], John Powell[7], Petroula Proitsi[7], Michelle K Lupton[7], Carol Brayne[8], David C Rubinsztein[9], Michael Gill[10], Brian Lawlor[10], Aoibhinn Lynch[10], Kristelle S Brown[11], Peter A Passmore[12], David Craig[12], Bernadette McGuinness[12], Stephen Todd[12], Clive Holmes[13], David Mann[14], A David Smith[15], Helen Beaumont[15], Donald Warden[15], Gordon Wilcock[16], Seth Love[17], Patrick G Kehoe[17], Nigel M Hooper[18], Emma R L C Vardy[14,18,19], John Hardy[20,21], Simon Mead[22], Nick C Fox[22], Martin Rossor[22], John Collinge[22], Wolfgang Maier[23,24], Frank Jessen[23], Eckart Rüther[24–26], Britta Schürmann[23,26], Reiner Heun[23,27], Heike Kölsch[23], Hendrik van den Bussche[28], Isabella Heuser[29], Johannes Kornhuber[30], Jens Wiltfang[31], Martin Dichgans[32,33], Lutz Frölich[34], Harald Hampel[35], John Gallacher[36], Michael Hüll[37], Dan Rujescu[37], Ina Giegling[36], Alison M Goate[38–40], John S K Kauwe[41], Carlos Cruchaga[38], Petra Nowotny[38], John C Morris[39], Kevin Mayo[38], Kristel Sleegers[42,43], Karolien Bettens[42,43], Sebastiaan Engelborghs[42,44], Peter P De Deyn[42,44], Christine Van Broeckhoven[42,43], Gill Livingston[45], Nicholas J Bass[45], Hugh Gurling[45], Andrew McQuillin[45], Rhian Gwilliam[46], Panagiotis Deloukas[46], Ammar Al-Chalabi[47], Christopher E Shaw[47], Magda Tsolaki[48], Andrew B Singleton[49], Rita Guerreiro[49], Thomas W Mühleisen[50,51], Markus M Nöthen[25,50,51], Susanne Moebus[52], Karl-Heinz Jöckel[52], Norman Klopp[53], H-Erich Wichmann[53–55], V Shane Pankratz[56], Sigrid B Sando[57,58], Jan O Aasly[57,58], Maria Barcikowska[59], Zbigniew K Wszolek[60], Dennis W Dickson[5], Neill R Graff-Radford[5,60], Ronald C Petersen[61,62], the Alzheimer's Disease Neuroimaging Initiative[63], Cornelia M van Duijn[64,65], Monique M B Breteler[64,65], M Arfan Ikram[64,65], Anita L DeStefano[66,67], Annette L Fitzpatrick[68], Oscar Lopez[69,70], Lenore J Launer[71], Sudha Seshadri[67,72], CHARGE consortium, Claudine Berr[73], Dominique Campion[74], Jacques Epelbaum[75], Jean-François Dartigues[76], Christophe Tzourio[77], Annick Alpérovitch[77], Mark Lathrop[78,79], EADI1 consortium, Thomas M Feulner[80], Patricia Friedrich[80], Caterina Riehle[80], Michael Krawczak[81–83], Stefan Schreiber[82,83], Manuel Mayhaus[80], S Nicolhaus[83], Stefan Wagenpfeil[84], Stacy Steinberg[85], Hreinn Stefansson[85], Kari Stefansson[86], Jon Snædal[87], Sigurbjörn Björnsson[87], Palmi V Jonsson[87], Vincent Chouraki[2–4], Benjamin Genier-Boley[2–4], Mikko Hiltunen[88], Hilkka Soininen[88], Onofre Combarros[89,90], Diana Zelenika[91], Marc Delepine[91], Maria J Bullido[90,92], Florence Pasquier[4,93], Ignacio Mateo[89,90], Ana Frank-Garcia[90,94], Elisa Porcellini[95], Olivier Hanon[96], Eliecer Coto[97], Victoria Alvarez[97], Paolo Bosco[98], Gabriele Siciliano[99], Michelangelo Mancuso[99], Francesco Panza[100], Vincenzo Solfrizzi[100], Benedetta Nacmias[101], Sandro Sorbi[101], Paola Bossù[102], Paola Piccardi[103], Beatrice Arosio[104], Giorgio Annoni[105], Davide Seripa[106], Alberto Pilotto[106], Elio Scarpini[107], Daniela Galimberti[107], Alexis Brice[108], Didier Hannequin[109], Federico Licastro[95], Lesley Jones[1], Peter A Holmans[1], Thorlakur Jonsson[85], Matthias Riemenschneider[80], Kevin Morgan[11], Steven G Younkin[5], Michael J Owen[1], Michael O'Donovan[1], Philippe Amouyel[2–4,92] & Julie Williams[1]

[1]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Neurosciences and Mental Health Research Institute, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK. [2]INSERM U744, F-59019 Lille, France. [3]Institut Pasteur de Lille, F-59019, Lille, France. [4]Université de Lille Nord de France, F-59000 Lille, France. [5]Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, Florida, USA. [6]National Institute for Health Research Biomedical Research Centre for Mental Health at the South London and Maudsley National Health Service Foundation Trust and Institute of Psychiatry, Kings College, London, UK. [7]Department of Neuroscience, Institute of Psychiatry, Kings College, London, UK. [8]Institute of Public Health, University of Cambridge, Cambridge, UK. [9]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. [10]Mercer's Institute for Research on Aging, St. James Hospital and Trinity College, Dublin, Ireland. [11]Institute of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham, UK. [12]Ageing Group, Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK. [13]Division of Clinical Neurosciences, School of Medicine, University of Southampton, Southampton, UK. [14]Neurodegeneration and Mental Health Research Group, School of Community Based Medicine, University of Manchester, Hope Hospital, Stott Lane, Salford, Manchester, UK. [15]Oxford Project to Investigate Memory and Ageing (OPTIMA), University of Oxford, John Radcliffe Hospital, Oxford, UK. [16]Nuffield Department of Clinical Medicine, Medical Sciences Division. University of Oxford, Headington, Oxford, UK. [17]Dementia Research Group, University of Bristol Institute of Clinical Neurosciences, Frenchay Hospital, Bristol, UK. [18]Institute of Molecular and Cellular Biology, Faculty of Biological Sciences, LIGHT Laboratories, University of Leeds, Leeds, UK. [19]Cerebral Function Unit, Salford Royal National Health Service (NHS) Trust, Stott Lane, Salford, UK. [20]Department of Molecular Neuroscience, Institute of Neurology, London, UK. [21]Reta Lilla Weston Laboratories, Institute of Neurology, London, UK. [22]Department of Neurodegenerative Disease, University College London, Institute of Neurology, London, UK. [23]Department of Psychiatry, University of Bonn, Bonn, Germany. [24]German Centre for Neurodegenerative Diseases, Bonn, Bonn, Germany. [25]Institute for Molecular Psychiatry, University of Bonn, Bonn, Germany. [26]Department of Psychiatry, University of Göttingen, Göttingen, Germany. [27]Department of Psychiatry, Royal Derby Hospital, Derby, UK. [28]Institute of Primary Medical Care, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [29]Department of Psychiatry, Charité Berlin, Berlin, Germany. [30]Department of Psychiatry and Psychotherapy, University of Erlangen, Nuremberg, Germany. [31]Landschaftsverband Rheinland-Hospital Essen, Department of Psychiatry and Psychotherapy, University Duisburg-Essen, Essen, Germany. [32]Department of Neurology, Klinikum der Universität München, Munich, Germany. [33]Institute for Stroke and Dementia Research, Klinikum der Universität München, Munich, Germany. [34]Department of Geriatric Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. [35]Department of Psychiatry, Psychsomatic Medicine and Psychotherapy, Johann Wolfgang Goethe-University, Frankfurt, Germany. [36]Department of Primary Care and Public Health, School of Medicine, Cardiff University, Cardiff, UK. [37]Centre for Geriatric Medicine and Section of Gerontopsychiatry and Neuropsychology, University of Freiburg, Freiburg, Germany. [38]Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri, USA. [39]Department of Neurology, Washington University School of Medicine, St Louis, Missouri, USA. [40]Department of Genetics, Washington University School of Medicine, St Louis, Missouri, USA. [41]Department of Biology, Brigham Young University, Provo, Utah, USA. [42]Institute Born-Bunge, University of Antwerp, Antwerpen, Belgium. [43]Neurodegenerative Brain Diseases group, Department of Molecular Genetics, Vlaams Interuniversitair Instituut voor Biotechnologie, Antwerpen, Belgium. [44]Memory Clinic and Department of Neurology, ZiekenhuisNetwerk Antwerpen, Middelheim, Antwerpen, Belgium. [45]Department of Mental Health Sciences, University College London, London UK. [46]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [47]MRC Centre for Neurodegeneration Research, Department of Clinical Neuroscience, King's College London, Institute of Psychiatry, London, UK. [48]Third Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece. [49]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA. [50]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [51]Institute of Human Genetics, University of Bonn, Bonn, Germany. [52]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, University Duisburg-Essen, Essen, Germany. [53]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [54]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [55]Klinikum Grosshadern, Munich, Germany. [56]Division of Biomedical Statistics and Informatics, Mayo Clinic and Mayo Foundation, Rochester, Minnesota, USA. [57]Department of Neurology, St. Olav's Hospital, Trondheim, Norway. [58]Department of Neuroscience, Norwegian University of Science and Technology, NTNU, 7491 Trondheim Norway. [59]Department of Neurodegenerative Disorders, Medical Research Centre, Polish Academy of Sciences, Warsaw, Poland. [60]Department of Neurology, Mayo Clinic College of Medicine, Jacksonville, Florida, USA. [61]Department of Neurology, Mayo Clinic College of Medicine, Rochester, Minnesota, USA. [62]Mayo Alzheimer Disease Research Center, Mayo Clinic College of Medicine, Rochester, Minnesota, USA. [63]Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (see URLs). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. See URLs for a complete listing of the ADNI investigators. [64]Department of Epidemiology, Erasmus Medical Center, University Medical Center, Rotterdam, The Netherlands. [65]Netherlands Consortium for Healthy Aging, Leiden, The Netherlands. [66]Departments of Neurology and Biostatistics, Boston University School of Medicine, Boston, Massachusetts, USA. [67]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [68]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [69]Department of Neurology, The Alzheimer's Disease Research Center, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. [70]Department of Psychiatry, The Alzheimer's Disease Research Center, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. [71]Neuroepidemiology Section, Laboratory of Epidemiology, Demography and Biometry (LJL), National Institute on Aging, Washington DC, USA. [72]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA. [73]INSERM U888, Hôpital La Colombière, Montpellier, France. [74]INSERM U614, Faculté de Médecine-Pharmacie de Rouen, Rouen, France. [75]Unité Mixte de Recherche 894, Inserm Faculté de Médecine, Université Paris Descartes, Paris, France. [76]INSERM U897, Victor Segalen University, Bordeaux, France. [77]INSERM U708, Paris, France. [78]Centre National de Genotypage, Institut Genomique, Commissariat à l'énergie Atomique, Evry, France. [79]The Foundation Jean Dausset-Centre d'Etude du Polymorphisme Humain, Paris, France. [80]Department of Psychiatry and Psychotherapy, Universitätsklinikum des Saarlandes, Universität des Saarlandes Saarbruecken Germany. [81]Institute of Medical Informatics and Statistics, Christian-Albrechts-University, Kiel, Germany. [82]Biobank Popgen, Institute of Experimental Medicine, Section of Epidemiology, Christian-Albrechts-University, Kiel, Germany. [83]Institute for Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany. [84]Institute of Medical Statistics and Epidemiology, Klinikum Rechts der Isar, Technische Universität, München, Germany. [85]deCODE Genetics, Reykjavik, Iceland. [86]deCODE Genetics and University of Iceland, Faculty of Medicine, Reykjavik, Iceland. [87]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [88]Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [89]Neurology Service, 'Marqués de Valdecilla' University Hospital (University of Cantabria), Santander, Spain. [90]Centro Investigacion Biomedica en Red Enfermedades Neurodegenerativas 'Marqués de Valdecilla' University Hospital (University of Cantabria), Santander, Spain. [91]Centre National de Genotypage, Institut Genomique, Commissariat à l'énergie Atomique, Evry, France. [92]Centro de Biologia Molecular Severo Ochoa, Universidad Autonoma, Campus de Cantoblanco, Madrid, Spain. [93]Centre Hospitalier Régional Universitaire de Lille, Lille, France. [94]Servicio de Neurologia, Hospital Universitario La Paz (UAM), Madrid, Spain. [95]Department of Experimental Pathology, School of Medicine, University of Bologna, Bologna, Italy. [96]University Paris Descartes, Department of Geriatrics, Broca Hospital, Paris, France. [97]Genetic Molecular Unit, Hospital Universitario Central de Asturias, Oviedo, Spain. [98]Istituto Di Ricovero e Cura a Carattere Scientifico Oasi Maria SS, Troina, Italy. [99]Department of Neuroscience, Neurological Clinic, University of Pisa, Italy. [100]Department of Geriatrics, Center for Aging Brain, Memory Unit, University of Bari, Policlinico, Bari, Italy. [101]Department of Neurological and Psychiatric Sciences, University of Florence, Florence, Italy. [102]Department of Clinical and Behavioral Neurology, IRCCS Santa Lucia Foundation, Roma, Italy. [103]Lab of Molecular Genetics, Section of Clinical Pharmacology, Department of Neuroscience, University of Cagliari, Cagliari, Italy. [104]Department of Internal Medicine, Università degli Studi di Milano, Fondazione IRCCS, Ospedale Maggiore, Mangiagalli e Regina Elena, Milan, Italy. [105]Department of Clinical Medicine and Prevention, University of Milano-Bicocca, Monza, Italy. [106]Geriatric Unit and Gerontology-Geriatric Research Laboratory, Department of Medical Science, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy. [107]Department of Neurological Sciences, Università degli Studi di Milano, Fondazione IRCCS, Ospedale Maggiore, Mangiagalli e Regina Elena, Milan, Italy. [108]INSERM, UMR_S679, Hopital de la salpétirère, Paris, France. [109]INSERM U614, Faculté de Médecine-Pharmacie de Rouen, Rouen, France. [110]These authors contributed equally to this work. Correspondence should be addressed to M.J.O. (owenmj@cardiff.ac.uk) or J.W. (williamsj@cardiff.ac.uk).

ONLINE METHODS

Sample ascertainment, diagnostic criteria and genotyping. Stage 1 comprised a meta-analysis of four Alzheimer's disease GWAS datasets (6,688 cases and 13,685 controls) including: GERAD1 (ref. 3), EADI1 (ref. 4), TGEN1 (ref. 13) and ADNI[14]. All Alzheimer's disease cases were diagnosed according to National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA)[36], The Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) or the Consortium to Establish a Registry for Alzheimer's Disease (CERAD)[37] criteria for either probable or definite Alzheimer's disease. Alzheimer's disease cases were predominantly female (62.4%). The mean age at disease onset and mean age at ascertainment of cases with Alzheimer's disease were 71.6 years and 77.3 years, respectively. Stage 1 included a total of 7,915 aged (≥60 years) screened controls (59.9% female; mean age at collection, 74.5 years) and 5,770 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 4,896 cases and 4,903 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 to NINCDS-ADRDA[36], DSM-IV or CERAD[37] criteria for possible, probable or definite Alzheimer's disease. Alzheimer's disease cases were predominantly female (63.4%). The mean age at disease onset and the mean age at ascertainment in cases with Alzheimer's disease were 72.3 years 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 8,286 cases and 21,258 controls, which included new genotyping in the EADI2 (ref. 4) and Mayo2 samples and *in silico* replication in the CHARGE sample[11]. All individuals included in these analyses provided written consent or assent to take part in the genetic association studies. We 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 measures can be found in the **Supplementary Note**. Clinical characteristics of all samples can be found in **Supplementary Table 1**.

Stage 1: combined analysis of four Alzheimer's disease GWAS. An inverse variance-weighted fixed effects meta-analysis was used to test for association with Alzheimer's disease 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 496,795 autosomal SNPs. These SNPs passed quality control in each of the GERAD1, ADNI and EADI1 GWAS. We successfully imputed 457,509 of these SNPs in the TGEN sample (which, unlike the other studies, employed the Affymetrix 500K array). In the combined analysis, 67 SNPs were associated with Alzheimer's disease at $P \leq 1 \times 10^{-5}$ (**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. Sixty-one SNPs remained significantly associated with Alzheimer's disease 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 Alzheimer's disease, 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 (**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 4,896 cases and 4,903 controls). We employed a Bonferroni-adjusted threshold for significance, accounting for 12 tests, of $P = 4.2 \times 10^{-3}$. For the three new SNPs that showed evidence for association in stage 2 ($P < 4.2 \times 10^{-3}$), summary data (including ORs 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 8,286 cases and 21,258 controls), employing a Bonferroni adjusted threshold for significance, accounting for three tests, of $P = 0.0167$.

Finally, we combined full summary data (including ORs and variances) from all datasets (GERAD1 and GERAD2, EAD1 and EAD2, ADNI, TGEN, deCODE, AD-IG, Mayo2 and CHARGE) using an inverse-variance–weighted fixed effects meta-analysis (total sample of up to 19,870 cases and 39,846 controls). For meta-analyses at all stages, a Cochran's *Q* test was performed and I^2 was 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. 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 Alzheimer's diesase (rs3764650, $P = 0.17$; rs670139, $P = 0.38$; and rs610932, $P = 0.95$). We also tested the SNPs for association with Alzheimer's disease while adjusting for the presence of at least one *APOE* ε4 allele within a logistic regression framework. Inclusion of presence or absence of *APOE* ε4 as a covariate had little effect on the results (**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 SNP × SNP interaction term

$$\ln(P/(1-P)) = \alpha + \beta X_A + \gamma X_B + \delta Y_1 + \varepsilon Y_2 + \zeta Y_3 + \eta Y_4 + \theta Y_5 + \iota Y_6 + \kappa X_A X_B$$

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 analyzed 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 European CEU HapMap individuals. Expression analysis was performed using Illumina's commercial whole genome expression array, the Sentrix Human-6 Expression BeadChip. The second eQTL dataset[18] analyzed 143 neurologically normal subjects of European ancestry. Frozen tissue samples were obtained from four brain regions (cerebellum, pons, and 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 normalized and log transformed as described in the original papers[17,18,38]. eQTLs were tested by linear regression of normalized expression level on SNP genotypes (coded as the number of minor alleles at each SNP: 0, 1 or 2).

36. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–944 (1984).
37. Mirra, S.S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479–486 (1991).
38. Myers, A.J. *et al.* A survey of genetic human cortical gene expression. *Nat. Genet.* **39**, 1494–1499 (2007).