nature
genetics

Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease

Denise Harold[1,45]*, Richard Abraham[1,45], Paul Hollingworth[1,45], Rebecca Sims[1], Amy Gerrish[1], Marian L Hamshere[1], Jaspreet Singh Pahwa[1], Valentina Moskvina[1], Kimberley Dowzell[1], Amy Williams[1], Nicola Jones[1], Charlene Thomas[1], Alexandra Stretton[1], Angharad R Morgan[1], Simon Lovestone[2], John Powell[3], Petroula Proitsi[3], Michelle K Lupton[3], Carol Brayne[4], David C Rubinsztein[5], Michael Gill[6], Brian Lawlor[6], Aoibhinn Lynch[6], Kevin Morgan[7], Kristelle S Brown[7], Peter A Passmore[8], David Craig[8], Bernadette McGuinness[8], Stephen Todd[8], Clive Holmes[9], David Mann[10], A David Smith[11], Seth Love[12], Patrick G Kehoe[12], John Hardy[13], Simon Mead[14], Nick Fox[15], Martin Rossor[15], John Collinge[14], Wolfgang Maier[16], Frank Jessen[16], Britta Schürmann[16], Hendrik van den Bussche[17], Isabella Heuser[18], Johannes Kornhuber[19], Jens Wiltfang[20], Martin Dichgans[21,22], Lutz Frölich[23], Harald Hampel[24,25], Michael Hüll[26], Dan Rujescu[25], Alison M Goate[27], John S K Kauwe[28], Carlos Cruchaga[27], Petra Nowotny[27], John C Morris[27], Kevin Mayo[27], Kristel Sleegers[29,30], Karolien Bettens[29,30], Sebastiaan Engelborghs[30,31], Peter P De Deyn[30,31], Christine Van Broeckhoven[29,30], Gill Livingston[32], Nicholas J Bass[32], Hugh Gurling[32], Andrew McQuillin[32], Rhian Gwilliam[33], Panagiotis Deloukas[33], Ammar Al-Chalabi[34], Christopher E Shaw[34], Magda Tsolaki[35], Andrew B Singleton[36], Rita Guerreiro[36], Thomas W Mühleisen[37,38], Markus M Nöthen[37,38], Susanne Moebus[39], Karl-Heinz Jöckel[39], Norman Klopp[40], H-Erich Wichmann[40–42], Minerva M Carrasquillo[43], V Shane Pankratz[44], Steven G Younkin[43], Peter A Holmans[1], Michael O'Donovan[1], Michael J Owen[1] & Julie Williams[1]

We undertook a two-stage genome-wide association study (GWAS) of Alzheimer's disease (AD) involving over 16,000 individuals, the most powerful AD GWAS to date. In stage 1 (3,941 cases and 7,848 controls), we replicated the established association with the apolipoprotein E (*APOE*) locus (most significant SNP, rs2075650, $P = 1.8 \times 10^{-157}$) and observed genome-wide significant association with SNPs at two loci not previously associated with the disease: at the *CLU* (also known as *APOJ*) gene (rs11136000, $P = 1.4 \times 10^{-9}$) and 5′ to the *PICALM* gene (rs3851179, $P = 1.9 \times 10^{-8}$). These associations were replicated in stage 2 (2,023 cases and 2,340 controls), producing compelling evidence for association with Alzheimer's disease in the combined dataset (rs11136000, $P = 8.5 \times 10^{-10}$, odds ratio = 0.86; rs3851179, $P = 1.3 \times 10^{-9}$, odds ratio = 0.86).

Alzheimer's disease is the most common form of dementia and is highly heritable (with heritability of up to 76%) but genetically complex[1]. Neuropathologically, the disease is characterized by extracellular senile plaques containing β-amyloid (Aβ) and intracellular neurofibrillary tangles containing hyperphosphorylated tau protein[1]. Thus far, four genes have been definitively implicated in the etiology of Alzheimer's disease. Mutations of the genes encoding amyloid precursor protein (*APP*) and presenilin 1 and 2 (*PSEN1*, *PSEN2*) cause rare, mendelian forms of the disease, usually with an early onset. However, in the more common form of the disease, only *APOE* has been established unequivocally as a susceptibility gene[1]. Aiming to identify new Alzheimer's disease loci, several genome-wide association studies (GWAS) have been previously conducted. All have identified strong evidence for Alzheimer's disease risk association to *APOE* but have found less convincing evidence implicating other genes[2–9]. This outcome is consistent with the majority of findings from GWAS of other common disease phenotypes, where susceptibility alleles typically have effect sizes with odds ratios (OR) of 1.5 or less, in contrast to that for *APOE* and Alzheimer's disease (OR ~ 3). Detecting such modest effects requires much larger samples than those used in the GWAS of Alzheimer's disease to date[10], which have all included fewer than 1,100 cases. Based upon the hypothesis that risk alleles for Alzheimer's disease are likely to confer ORs in the range seen with other common diseases, we undertook a more powerful GWAS than has been carried out to date.

We established a collaborative consortium from Europe and the United States from which we were able to draw upon a combined sample of up to 19,000 subjects (before quality control) and conducted a two-stage study. In stage 1, we genotyped 14,639 subjects on Illumina platforms: 5,715 samples were genotyped using the Illumina 610-quad
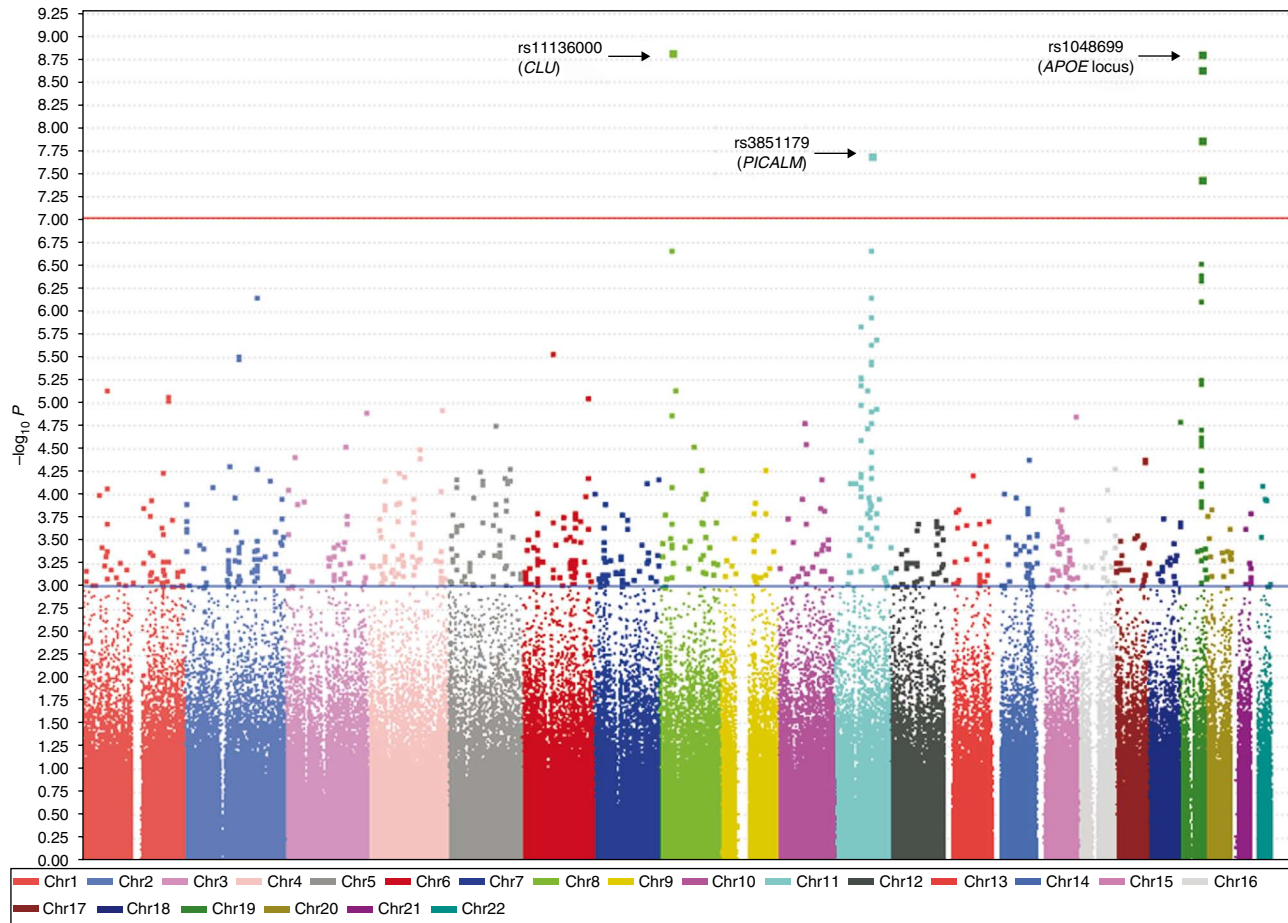
Figure 1 Scatterplot of chromosomal position (*x* axis) against $-\log_{10}$ GWAS *P* value (*y* axis). The *y*-axis scale has been limited to 9.25 ($P = 5.6 \times 10^{-10}$), although highly significant association was observed with SNPs in the vicinity of the *APOE* locus (for example, rs2075650 with $P = 1.8 \times 10^{-157}$). The threshold for genome-wide significance ($P \leq 9.4 \times 10^{-8}$) is indicated by the red horizontal line. 761 SNPs with $P \leq 1 \times 10^{-3}$ lie above the blue horizontal line and are listed in **Supplementary Table 2**. The plot was produced using Haploview version 4.0 (ref 49.).

chip; genotypes for the remaining subjects were made available to us either from population control datasets or through collaboration and were genotyped on the Illumina HumanHap550 or HumanHap300 BeadChips. Prior to association analysis, all samples and genotypes underwent stringent quality control, which resulted in the elimination of 53,383 autosomal SNPs and 2,850 subjects. Thus, in stage 1, we tested 529,205 autosomal SNPs for association in up to 11,789 subjects (3,941 Alzheimer's disease cases and 7,848 controls, of whom 2,078 were elderly screened controls, see **Supplementary Table 1**). The genomic control inflation factor (λ)[11] was 1.037, suggesting little evidence for residual population stratification.

In addition to the known association of Alzheimer's disease with the *APOE* locus, GWAS analysis identified two loci associated at a genome-wide level of significance (see **Fig. 1**). **Table 1** shows SNPs which were genome-wide significant (GWS) in stage 1. Thirteen GWS SNPs map within or close to the *APOE* locus on chromosome 19 ($P = 3 \times 10^{-8}$–2×10^{-157}) and the top five are shown in **Table 1** (see **Supplementary Table 2** for the complete list). The other two SNPs represent newly identified associations. One (rs11136000) is located within an intron of *CLU* (which encodes clusterin) on chromosome 8 ($P = 1.4 \times 10^{-9}$, OR = 0.840); the other (rs3851179) is 88.5 kb 5′ to *PICALM* on chromosome 11 ($P = 1.9 \times 10^{-8}$, OR = 0.849). Note that there was no significant difference in allele frequencies between elderly, screened controls and population controls for these two SNPs. In stage 2, the two newly identified GWS SNPs were genotyped in an

independent sample comprising 2,023 Alzheimer's disease cases and 2,340 age-matched, cognitively screened controls (**Supplementary Table 3**). Both were independently associated in this sample (rs11136000, one-tailed $P = 0.017$, OR = 0.905; rs3851179, one-tailed $P = 0.014$, OR = 0.897). Meta-analysis of the stage 1 and 2 datasets also produced highly significant evidence of association (rs11136000, $P = 8.5 \times 10^{-10}$, OR = 0.861; rs3851179, $P = 1.3 \times 10^{-9}$, OR = 0.859, two-tailed, **Table 1**) for *CLU* and *PICALM* loci, respectively. We sought further evidence from the Reiman *et al.* study[9] and the Li *et al.* study[8], two publicly available Alzheimer's disease GWAS datasets, but neither of the newly identified GWS SNPs had been previously genotyped or could be imputed. We also note that rs11136000 in *CLU* showed the most evidence for association with AD in the GWAS of Amouyel and colleagues published in this issue[12]. As secondary analyses, we tested each one for interaction with *APOE* status and for association with age at onset. No significant interactions of the two newly identified SNPs with *APOE* status were observed as influencing Alzheimer's disease risk (rs11136000 by *APOE*-ε4 interaction $P = 0.674$; rs3851179 by *APOE*-ε4 interaction $P = 0.735$). Although we observed significant effects of the GWS SNPs on age at onset, these were limited to SNPs at the *APOE* locus (data not shown).

In a preliminary attempt to attribute the source of the association to a functional variant, we used publicly available data to identify additional SNPs at each locus that were correlated through linkage disequilibrium (LD) with either of the newly associated SNPs or that

Table 1 SNPs showing genome-wide significant association with Alzheimer's disease in stage 1 of the GWAS

SNP	Chr	Closest RefSeq gene	Location relative to gene	MAF	GWAS: 3,941 cases, 7,848 controls		Extension: 2,023 cases, 2,340 controls		Combined: 5,964 cases, 10,188 controls		Population attributable risk (%)
					P value (two-tailed)	GWAS OR (95% CI)	P value (one-tailed)	Extension OR (95% CI)	P value (two-tailed)	Combined OR (95% CI)	
rs2075650*	19	*TOMM40*	Intron	0.15	1.8×10^{-157}	2.53 (2.37–2.71)					18.7%
rs157580	19	*TOMM40*	Intron	0.39	9.6×10^{-54}	0.63 (0.59–0.66)					
rs6859	19	*PVRL2*	3′ UTR	0.43	6.9×10^{-41}	1.46 (1.38–1.54)					
rs8106922	19	*TOMM40*	Intron	0.40	5.4×10^{-39}	0.68 (0.64–0.72)					
rs405509	19	*APOE*	5′	0.52	4.9×10^{-37}	0.70 (0.66–0.74)					
rs11136000	**8**	***CLU***	**Intron**	**0.40**	**1.4×10^{-9}**	**0.84 (0.79–0.89)**	**0.017**	**0.91 (0.83–0.99)**	**8.5×10^{-10}**	**0.86 (0.82–0.90)**	**8.9%**
rs3851179	**11**	***PICALM***	**5′**	**0.37**	**1.9×10^{-8}**	**0.85 (0.80–0.90)**	**0.014**	**0.90 (0.82–0.99)**	**1.3×10^{-9}**	**0.86 (0.82–0.90)**	**9.3%**

Five of the 13 genome-wide significant SNPs at the *APOE* locus are shown (see **Supplementary Table 2** for the complete list). *P* values in the extension sample and the combined sample are also shown for the two SNPs unlinked to the *APOE* locus (highlighted in bold). Chr, chromosome; MB, position in megabases; MAF, minor allele frequency in controls; OR, odds ratio for the minor allele; 95% CI, 95% confidence interval; UTR, untranslated region. *rs2075650 is in linkage disequilibrium with rs429358, the *APOE* ε4 SNP (r^2=0.48).

might plausibly have functional effects (see **Supplementary Table 4**). A synonymous SNP (rs7982) in *CLU* was in strong LD ($r^2 = 0.95$ in our extension sample) with the GWS SNP and showed a similar level of evidence for association with Alzheimer's disease in our combined stage 1 and 2 sample (meta-analysis $P = 8 \times 10^{-10}$; stage 1 genotypes were imputed). This SNP is in exon 5 of *CLU*, which encodes part of the beta chain of the clusterin protein and may influence a predicted exon splicing enhancer. We note that Tycko and colleagues[13] previously published a negative association study on the *CLU* gene and Alzheimer's disease, analyzing four SNPs in an Alzheimer's disease case-control sample of African-American, Hispanic and European descent/non-Hispanic individuals. Although they identified rs7982 through mutation screening (referred to as VB in their study), this SNP was not tested for association in their sample. The four SNPs that were analyzed were rare in individuals of European descent (minor allele frequency <2%) and there was very limited power to detect association in their European descent subsample of 53 Alzheimer's disease cases and 43 controls. As these four SNPs were not genotyped in our GWAS, there is no overlap between the two studies.

Several potentially functional SNPs were identified at the *PICALM* locus. Of these, two showed good evidence for association: rs561655, which is within a putative transcription factor binding site, and rs592297, which is a synonymous SNP in exon 5 of *PICALM* that may influence a predicted exon splicing enhancer. However, neither of these SNPs showed the strength of evidence for association observed for rs3851179, the GWS SNP at the *PICALM* locus (rs561655, meta-analysis $P = 1 \times 10^{-7}$; rs592297, meta-analysis $P = 2 \times 10^{-7}$). A number of SNPs in LD with rs3851179 and showing moderate evidence of association in the GWAS ($P < 1 \times 10^{-4}$) were also examined in additional samples, most notably rs541458. This SNP is 8 kb 5′ to the *PICALM* gene and was directly genotyped in the extension sample, the Reiman *et al.* study and the Li *et al.* study, with $P < 0.05$ in each. Following meta-analysis, rs541458 is one of the most significant SNPs (meta-analysis $P = 8 \times 10^{-10}$), a finding that is also supported by the study of Amouyel and colleagues published in this issue ($P = 3 \times 10^{-3}$)[12]. Further genetic analyses will be required to characterize the true nature of the associations observed at these loci.

We also tested whether the number of significant associations observed in the GWAS exceeded what would be expected by chance. Having removed SNPs within the *APOE*, *CLU* and *PICALM* loci from the analysis (see Online Methods), we focused on those that showed the most evidence for association ($P < 1 \times 10^{-5}$; see **Table 2**). Approximately 13 independent signals were observed whereas fewer than 4 would be expected by chance ($P = 7.5 \times 10^{-6}$). This shows suggestive evidence for association with the gene encoding complement receptor 1 (*CR1*; rs1408077, $P = 8.3 \times 10^{-6}$). We note that Amouyel and colleagues present an independent GWAS of late-onset Alzheimer's disease and report a genome-wide significant association of SNPs in *CR1* to Alzheimer's disease[12]. Also noteworthy in our study are the genes encoding bridging integrator 1 (*BIN1*), which produces a protein involved in synaptic vesicle endoctyosis[14], and the disabled homolog 1 (*DAB1*), whose product is involved with tyrosine phosphorylation and microtubule function in neurons[15]. We have also tested over 100 variants highlighted by previous GWAS of Alzheimer's disease for association in our sample (see **Supplementary Note** for full discussion). These are summarized in **Supplementary Table 5**.

Until now, the *APOE*-ε4 allele was the only consistently replicated genetic risk factor for Alzheimer's disease. It is therefore intriguing that we find compelling evidence for association with *CLU*, a gene that encodes another major brain apolipoprotein[16], suggesting that susceptibility genes are not randomly distributed through functional pathways. The predominant form of clusterin is a secreted heterodimeric glycoprotein of 75–80 kDa. The single copy gene spans about 16 kb on chromosome 8p21-p12 and encodes an mRNA of approximately 2 kb that translates into a 449-amino-acid primary polypeptide chain. Clusterin is expressed in all mammalian tissues[17], and there is strong evidence that *CLU* expression is elevated in a number of pathological conditions involving injury or chronic inflammation of the brain[18]. In individuals with Alzheimer's disease, *CLU* expression is reported to be increased in affected cortical areas of the brain and to be present in amyloid plaques and in the cerebrospinal fluid[18–21].

Clusterin is a multifunctional molecule. It interacts with the soluble form of Aβ in animal models of disease and binds soluble Aβ in a specific and reversible manner, forming complexes that have been shown to cross the blood-brain barrier[22–25]. Notably, APOE also appears to act as a molecular chaperone for Aβ and influences when Aβ aggregates and deposits[26]; APOE also influences Aβ conformation and toxicity[27,28]. In a similar way, clusterin appears to regulate both the toxicity of Aβ and its conversion into insoluble forms[29–33]. Furthermore, APOE and clusterin have been shown to cooperate in suppressing Aβ deposition[34], and they may critically modify

Table 2 SNPs showing association with Alzheimer's disease at $P \leq 1 \times 10^{-5}$

SNP	Chr	MB	Closest RefSeq gene	Location relative to gene	GWAS P value	OR	95% CI
rs11894266	2	170.3	*SSB*	5′	6.9×10^{-7}	0.86	0.81–0.91
rs610932	11	59.7	*MS4A6A*	3′ UTR	1.4×10^{-6}	0.87	0.82–0.92
rs10501927	11	99.3	*CNTN5*	Intronic	2.0×10^{-6}	1.18	1.10–1.26
rs9446432	6	72.4	*C6orf155*	Intergenic	2.8×10^{-6}	1.28	1.15–1.41
rs7561528	2	127.6	*BIN1*	5′	3.0×10^{-6}	1.16	1.09–1.24
rs744373	2	127.6	*BIN1*	5′	3.2×10^{-6}	1.17	1.09–1.25
rs662196	11	59.7	*MS4A6A*	Intronic	5.2×10^{-6}	0.88	0.83–0.93
rs583791	11	59.7	*MS4A6A*	Intronic	5.3×10^{-6}	0.88	0.83–0.93
rs676309	11	59.8	*MS4A4E*	5′	6.3×10^{-6}	1.14	1.08–1.20
rs1157242	8	37.2	*KCNU1*	Intergenic	7.0×10^{-6}	1.17	1.10–1.26
rs1539053	1	57.9	*DAB1*	Intronic	7.1×10^{-6}	0.88	0.83–0.93
rs11827375	11	76.0	*C11orf30*	3′	7.2×10^{-6}	1.23	1.12–1.35
rs1408077	1	205.9	*CR1*	Intronic	8.3×10^{-6}	1.17	1.09–1.25
rs9384428	6	156.5	*ARID1B*	Intergenic	8.5×10^{-6}	1.14	1.08–1.21
rs6701713	1	205.9	*CR1*	Intronic	8.7×10^{-6}	1.17	1.09–1.25
rs3818361	1	205.9	*CR1*	Intronic	9.2×10^{-6}	1.17	1.09–1.25

SNPs at the *APOE*, *CLU* and *PICALM* loci are excluded. Chr, chromosome; MB, position in megabases; OR, odds ratio for the minor allele; 95% CI, 95% confidence interval; UTR, untranslated region.

Aβ clearance at the blood-brain barrier, which could suggest a role for clusterin in the amyloidogenic pathway[35]. Levels of APOE protein appear to be inversely proportional to *APOE*-ε4 allele dose levels, with protein levels reduced in ε4 homozygotes compared with heterozygotes. Conversely, clusterin levels are increased in proportion to *APOE*-ε4 allele dose levels, suggesting an induction of clusterin in individuals with low APOE levels[36].

The second gene locus to show compelling evidence for association with Alzheimer's disease is *PICALM* (phosphatidylinositol-binding clathrin assembly protein, also known as *CALM*, clathrin assembly lymphoid-myeloid leukemia gene). *PICALM* is ubiquitously expressed in all tissue types with prominent expression in neurons, where it is nonselectively distributed at the pre- and postsynaptic structures. It has been shown that like BIN1 (which shows suggestive evidence for association with AD in our sample), the PICALM protein is involved in clathrin-mediated endocytosis (CME), an essential step in the intracellular trafficking of proteins and lipids such as nutrients, growth factors and neurotransmitters[37–39]. Of relevance to Alzheimer's disease, *PICALM* appears to be involved in directing the trafficking of VAMP2. VAMP2 is a soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) protein that has a prominent role in the fusion of synaptic vesicles to the presynaptic membrane in neurotransmitter release, a process that is crucial to neuronal function[40]. Brains affected with Alzheimer's disease show a reduced number of synapses, and stereological and biochemical analysis has shown that this reduction in synaptic density correlates better with cognitive decline than with the accumulation of plaques[41]. More recent analysis indicates that synapses within the brains of individuals with Alzheimer's disease may be dysfunctional even before they visibly degenerate[42]. Therefore, we can hypothesize that genetically directed changes in *PICALM* function result in perturbations at the synapse, possibly through synaptic vesicle cycling, thereby increasing risk for Alzheimer's disease. Alternatively, *PICALM* could influence Alzheimer's disease risk through APP processing via endocytic pathways, resulting in changes in Aβ levels. Cell culture experiments have shown that full-length APP is retrieved from the cell surface by CME[43] and that inhibition of endocytosis reduces APP internalization and reduces Aβ production and release[44,45]. Increased synaptic activity is known to lead to the elevated endocytosis of synaptic vesicle proteins, and recent work has provided evidence *in vivo* that increased CME triggered by increased synaptic activity drives more APP into endocytotic compartments, resulting in an increase of Aβ production and release[46].

The power of the present study to detect loci of the effect sizes observed in the GWAS for rs11136000 and rs3851179 at a genome-wide significance level is 0.74 and 0.57, respectively. However, it is widely acknowledged that effect sizes of significant loci obtained from genome-wide studies are overestimates of the true effects[47]. In the extension study, rs11136000 had an OR of 0.905 and rs3851179 had an OR of 0.897. Assuming that these odds ratios better reflect the true effect sizes associated with these loci, the GWAS has power to detect genome-wide significant association of 0.026 and 0.041, respectively. Thus, it is likely that there are other genes of similar effect sizes that did not reach genome-wide significance. Indeed, we observed many SNPs which, although failing to reach this level of statistical significance, might still reflect true associations with Alzheimer's disease: for example, variants at the *CR1* locus, also highlighted by the results of Amouyel and colleagues[12]. As with other GWAS for complex disorders (such as type 2 diabetes, for which 18 susceptibility loci have now been confirmed through GWAS meta-analysis of 54,000 subjects[48]), additional and larger GWAS may be required to identify remaining susceptibility variants for Alzheimer's disease.

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.

AUTHOR CONTRIBUTIONS

J. Williams, M.J.O. and M.O. directed this study, assisted by R.A., P.H. and P.A.H. D.H. and J. Williams took primary responsibility for drafting the manuscript assisted by R.A., P.H., R.S., A.G., M.O. and M.J.O. J. Williams, R.A., P.H., R.S., A.G., K.D., A.W., N.J., C.T., A.S., A.R.M., S. Lovestone, J.P., P.P., M.K.L., C.B., D.C.R., M.G., B.L., A.L., K. Morgan, K.S.B., P.A.P., D.C., B.M., S.T., C.H., D.M., A.D.S., S. Love, P.G.K., J.H., S. Mead, N.F., M.R., J.C., W.M., F.J., B.S., H.v.d.B., I.H., J.K., J. Wiltfang, M.D., L.F., A.M.G., J.S.K.K., C.C., P.N., J.C.M., K. Mayo, G.L., N.J.B., H.G. and A.M. contributed towards clinical sample collection, ascertainment, diagnosis and preparation of samples from the '610 group' from the stage 1 'discovery sample' and in some cases also provided stage 2 'follow up' samples. R.A. and P.H. were responsible for the coordination, collection, transit and selection of samples for genotyping from the '610 group'. R. Gwilliam and P.D. were responsible for procedures related to genotyping the 610 group on the Illumina platform. A.A.-C., C.E.S., A.B.S., R. Guerreiro, T.W.M., M.M.N., S.M., K.-H.J., N.K., H.-E.W., M.M.C., V.S.P., S.G.Y., H.H., D.R. and M.H. were involved in clinical sample collection, ascertainment, diagnosis, preparation of samples and genotyping of 'collaborative samples' included in Stage 1 (i.e. samples other than the '610 group'). K.S., K.B., S.E., P.P.D.D., C.v.B. and M.T. contributed towards

sample collection, diagnosis and preparation of case-control material for the stage 2 'replication sample'. Replication genotyping was coordinated and performed by R.A., assisted by R.S. and A.G. and J.S.P. developed the database for the GWA project in which the data were stored, under the supervision of J. Williams. D.H. completed statistical quality control and produced association statistics, under the supervision of J. Williams, M.L.H., V.M. and P.A.H. All authors discussed the results and approved the manuscript.

1. Avramopoulos, D. Genetics of Alzheimer's disease: recent advances. *Genome Med.* **1**, 34 (2009).
2. Abraham, R. *et al.* A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med. Genomics* **1**, 44 (2008).
3. Beecham, G.W. *et al.* Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am. J. Hum. Genet.* **84**, 35–43 (2009).
4. 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).
5. Carrasquillo, M.M. *et al.* Genetic variation in *PCDH11X* is associated with susceptibility to late-onset Alzheimer's disease. *Nat. Genet.* **41**, 192–198 (2009).
6. Coon, K.D. *et al.* A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J. Clin. Psychiatry* **68**, 613–618 (2007).
7. Grupe, A. *et al.* Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum. Mol. Genet.* **16**, 865–873 (2007).
8. Li, H. *et al.* Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch. Neurol.* **65**, 45–53 (2008).
9. Reiman, E.M. *et al.* GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* **54**, 713–720 (2007).
10. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
11. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
12. Amouyel, P. *et al.* Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat. Genet.* advance online publication, doi:10.1038/ng.439 (6 September 2009).
13. Tycko, B. *et al.* Polymorphisms in the human apolipoprotein-J/clusterin gene: ethnic variation and distribution in Alzheimer's disease. *Hum. Genet.* **98**, 430–436 (1996).
14. Cousin, M.A. & Robinson, P.J. The dephosphins: dephosphorylation by calcineurin triggers synaptic vesicle endocytosis. *Trends Neurosci.* **24**, 659–665 (2001).
15. Hiesberger, T. *et al.* Direct binding of Reelin to VLDL receptor and ApoE receptor 2 induces tyrosine phosphorylation of disabled-1 and modulates tau phosphorylation. *Neuron* **24**, 481–489 (1999).
16. Jenne, D.E. & Tschopp, J. Clusterin: the intriguing guises of a widely expressed glycoprotein. *Trends Biochem. Sci.* **17**, 154–159 (1992).
17. Jones, S.E. & Jomary, C. Clusterin. *Int. J. Biochem. Cell Biol.* **34**, 427–431 (2002).
18. Calero, M. *et al.* Apolipoprotein J (clusterin) and Alzheimer's disease. *Microsc. Res. Tech.* **50**, 305–315 (2000).
19. Giannakopoulos, P. *et al.* Possible neuroprotective role of clusterin in Alzheimer's disease: a quantitative immunocytochemical study. *Acta Neuropathol.* **95**, 387–394 (1998).
20. Liang, W.S. *et al.* Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. *Physiol. Genomics* **33**, 240–256 (2008).
21. McGeer, P.L., Kawamata, T. & Walker, D.G. Distribution of clusterin in Alzheimer brain tissue. *Brain Res.* **579**, 337–341 (1992).
22. Ghiso, J. *et al.* The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is complexed to SP-40,40 (apolipoprotein J), an inhibitor of the complement membrane-attack complex. *Biochem. J.* **293**, 27–30 (1993).
23. Golabek, A., Marques, M.A., Lalowski, M. & Wisniewski, T. Amyloid beta binding proteins in vitro and in normal human cerebrospinal fluid. *Neurosci. Lett.* **191**, 79–82 (1995).
24. Zlokovic, B.V. *et al.* Brain uptake of circulating apolipoproteins J and E complexed to Alzheimer's amyloid beta. *Biochem. Biophys. Res. Commun.* **205**, 1431–1437 (1994).
25. Zlokovic, B.V. *et al.* Glycoprotein 330/megalin: probable role in receptor-mediated transport of apolipoprotein J alone and in a complex with Alzheimer disease amyloid beta at the blood-brain and blood-cerebrospinal fluid barriers. *Proc. Natl. Acad. Sci. USA* **93**, 4229–4234 (1996).

26. Bales, K.R. *et al.* Apolipoprotein E is essential for amyloid deposition in the APP(V717F) transgenic mouse model of Alzheimer's disease. *Proc. Natl. Acad. Sci. USA* **96**, 15233–15238 (1999).

27. Castano, E.M. *et al.* Fibrillogenesis in Alzheimer's disease of amyloid beta peptides and apolipoprotein E. *Biochem. J.* **306**, 599–604 (1995).

28. Ma, J., Yee, A., Brewer, H.B. Jr., Das, S. & Potter, H. Amyloid-associated proteins alpha 1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer beta-protein into filaments. *Nature* **372**, 92–94 (1994).

29. Boggs, L.N. *et al.* Clusterin (Apo J) protects against in vitro amyloid-beta (1–40) neurotoxicity. *J. Neurochem.* **67**, 1324–1327 (1996).

30. DeMattos, R.B. *et al.* Clusterin promotes amyloid plaque formation and is critical for neuritic toxicity in a mouse model of Alzheimer's disease. *Proc. Natl. Acad. Sci. USA* **99**, 10843–10848 (2002).

31. Lambert, M.P. *et al.* Diffusible, nonfibrillar ligands derived from Abeta1–42 are potent central nervous system neurotoxins. *Proc. Natl. Acad. Sci. USA* **95**, 6448–6453 (1998).

32. Matsubara, E., Soto, C., Governale, S., Frangione, B. & Ghiso, J. Apolipoprotein J and Alzheimer's amyloid beta solubility. *Biochem. J.* **316**, 671–679 (1996).

33. Oda, T. *et al.* Clusterin (apoJ) alters the aggregation of amyloid beta-peptide (A beta 1–42) and forms slowly sedimenting A beta complexes that cause oxidative stress. *Exp. Neurol.* **136**, 22–31 (1995).

34. DeMattos, R.B. *et al.* ApoE and clusterin cooperatively suppress Abeta levels and deposition: evidence that ApoE regulates extracellular Abeta metabolism in vivo. *Neuron* **41**, 193–202 (2004).

35. Bell, R.D. *et al.* Transport pathways for clearance of human Alzheimer's amyloid beta-peptide and apolipoproteins E and J in the mouse central nervous system. *J. Cereb. Blood Flow Metab.* **27**, 909–918 (2007).

36. Bertrand, P., Poirier, J., Oda, T., Finch, C.E. & Pasinetti, G.M. Association of apolipoprotein E genotype with brain levels of apolipoprotein E and apolipoprotein J (clusterin) in Alzheimer disease. *Brain Res. Mol. Brain Res.* **33**, 174–178 (1995).

37. Dreyling, M.H. *et al.* The t(10;11)(p13;q14) in the U937 cell line results in the fusion of the AF10 gene and CALM, encoding a new member of the AP-3 clathrin assembly protein family. *Proc. Natl. Acad. Sci. USA* **93**, 4804–4809 (1996).

38. Tebar, F., Bohlander, S.K. & Sorkin, A. Clathrin assembly lymphoid myeloid leukemia (CALM) protein: localization in endocytic-coated pits, interactions with clathrin, and the impact of overexpression on clathrin-mediated traffic. *Mol. Biol. Cell* **10**, 2687–2702 (1999).

39. Yao, P.J., Petralia, R.S., Bushlin, I., Wang, Y. & Furukawa, K. Synaptic distribution of the endocytic accessory proteins AP180 and CALM. *J. Comp. Neurol.* **481**, 58–69 (2005).

40. Harel, A., Wu, F., Mattson, M.P., Morris, C.M. & Yao, P.J. Evidence for CALM in directing VAMP2 trafficking. *Traffic* **9**, 417–429 (2008).

41. Masliah, E. *et al.* Altered expression of synaptic proteins occurs early during progression of Alzheimer's disease. *Neurology* **56**, 127–129 (2001).

42. Fitzjohn, S.M. *et al.* Age-related impairment of synaptic transmission but normal long-term potentiation in transgenic mice that overexpress the human APP695SWE mutant form of amyloid precursor protein. *J. Neurosci.* **21**, 4691–4698 (2001).

43. Nordstedt, C., Caporaso, G.L., Thyberg, J., Gandy, S.E. & Greengard, P. Identification of the Alzheimer beta/A4 amyloid precursor protein in clathrin-coated vesicles purified from PC12 cells. *J. Biol. Chem.* **268**, 608–612 (1993).

44. Carey, R.M., Balcz, B.A., Lopez-Coviella, I. & Slack, B.E. Inhibition of dynamin-dependent endocytosis increases shedding of the amyloid precursor protein ectodomain and reduces generation of amyloid beta protein. *BMC Cell Biol.* **6**, 30 (2005).

45. Koo, E.H. & Squazzo, S.L. Evidence that production and release of amyloid beta-protein involves the endocytic pathway. *J. Biol. Chem.* **269**, 17386–17389 (1994).

46. Cirrito, J.R. *et al.* Endocytosis is required for synaptic activity-dependent release of amyloid-beta in vivo. *Neuron* **58**, 42–51 (2008).

47. Göring, H.H., Terwilliger, J.D. & Blangero, J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am. J. Hum. Genet.* **69**, 1357–1369 (2001).

48. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).

49. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).

[1]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK. [2]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. [3]Department of Neuroscience, Institute of Psychiatry, Kings College, London, UK. [4]Institute of Public Health and [5]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. [6]Mercer's Institute for Research on Aging, St. James Hospital and Trinity College, Dublin, Ireland. [7]Institute of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham, UK. [8]Ageing Group, Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK. [9]Division of Clinical Neurosciences, School of Medicine, University of Southampton, Southampton, UK. [10]Clinical Neuroscience Research Group, Greater Manchester Neurosciences Centre, University of Manchester, Salford, UK. [11]Oxford Project to Investigate Memory and Ageing, University of Oxford, John Radcliffe Hospital, Oxford, UK. [12]Dementia Research Group, University of Bristol Institute of Clinical Neurosciences, Frenchay Hospital, Bristol, UK. [13]Department of Molecular Neuroscience and Reta Lilla Weston Laboratories, Institute of Neurology, London, UK. [14]MRC Prion Unit and [15]Dementia Research Centre, Department of Neurodegenerative Diseases, UCL Institute of Neurology, London, UK. [16]Department of Psychiatry, University of Bonn, Bonn, Germany. [17]Institute of Primary Medical Care, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [18]Department of Psychiatry, Charité Berlin, Berlin, Germany. [19]Department of Psychiatry and Psychotherapy, University of Erlangen-Nürnberg, Erlangen, Germany. [20]Landschaftsverband Rheinland-Hospital Essen, Department of Psychiatry and Psychotherapy, University Duisburg-Essen, Essen, Germany. [21]Institute for Stroke and Dementia Research and [22]Department of Neurology, Klinikum der Universität München, Munich, Germany. [23]Department of Geriatric Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. [24]Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Laboratory of Neuroimaging & Biomarker Research, Trinity College, University of Dublin, Dublin, Ireland. [25]Alzheimer Memorial Center and Geriatric Psychiatry Branch, Department of Psychiatry, Ludwig-Maximilian University, Munich, Germany. [26]Centre for Geriatric Medicine and Section of Gerontopsychiatry and Neuropsychology, Medical School, University of Freiburg, Freiburg, Germany. [27]Departments of Psychiatry, Neurology and Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [28]Department of Biology, Brigham Young University, Provo, Utah, USA. [29]Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerpen, Belgium. [30]Institute Born-Bunge and University of Antwerp, Antwerpen, Belgium. [31]Memory Clinic and Department of Neurology, Ziekenhuis Netwerk Antwerpen Middelheim, Antwerpen, Belgium. [32]Department of Mental Health Sciences, University College London, London, UK. [33]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [34]MRC Centre for Neurodegeneration Research, Department of Clinical Neuroscience, King's College London, Institute of Psychiatry, London, UK. [35]Third Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece. [36]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA. [37]Department of Genomics, Life & Brain Center and [38]Institute of Human Genetics, University of Bonn, Bonn, Germany. [39]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, University Duisburg-Essen, Essen, Germany. [40]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [41]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [42]Klinikum Grosshadern, Munich, Germany. [43]Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, Florida, USA. [44]Division of Biomedical Statistics and Informatics, Mayo Clinic and Mayo Foundation, Rochester, Minnesota, USA. [45]These authors contributed equally to this work. Correspondence should be addressed to J.W. (williamsj@cardiff.ac.uk) or M.J.O. (owenmj@cardiff.ac.uk).

ONLINE METHODS

Sample ascertainment and diagnostic criteria. The study comprised a stage 1 discovery sample of 4,957 Alzheimer's disease cases and 9,682 controls and a stage 2 follow-up sample of 2,023 Alzheimer's disease cases and 2,340 controls. See the **Supplementary Note** for a complete description of the study subjects.

Stage 1 genotyping. Genotyping was performed at the Sanger Institute, UK. Two hundred nanograms of input DNA per sample were used and prepared for genotyping using the Illumina Infinium system (Illumina Inc.). Manufacturer's protocols were followed throughout. The Illuminus algorithm for cluster analysis was used for genotype calling[50].

Stage 1 individual quality control. 4,113 Alzheimer's disease cases and 1,602 controls were genotyped on the Illumina 610-quad chip as part of this study (the 610 group). In addition, 844 Alzheimer's disease cases and 8,080 controls previously genotyped using either the Illumina HumanHap550 or Illumina HumanHap300 were included in the analysis. These genotypes were generated as part of seven different studies, making eight separate groups in total: (i) 610; (ii) Mayo; (iii) 1958 birth cohort (Sanger); (iv) 1958 birth cohort (T1DGC); (v) ALS control; (vi) Coriell control; (vii) Heinz Nixdorf Recall (HNR) study; (viii) KORA F4. As we used genotype data from multiple sources, it was important to apply stringent quality control (QC) filters, as differential genotyping error rates between groups could result in spurious associations when the data were combined[51,52]. These filters were applied separately to each of these eight groups to remove poorly performing samples using tools implemented in PLINK version 1.05 (ref. 53).

We removed 1,469 individuals with missing genotype rates > 0.01. We also applied a filter based on mean autosomal heterozygosity, excluding 578 individuals with values above or below empirically determined thresholds. Seventy-one individuals with inconsistencies between reported gender and genotype-determined gender and 22 individuals with ambiguous genotype-determined gender were removed. All individuals passing these QC filters were examined for potential genetic relatedness by calculating identity-by-descent (IBD) estimates for all possible pairs of individuals in PLINK, and removing one of each pair with an IBD estimate ≥0.125 (the level expected for first cousins). IBD estimates were calculated using SNPs that were common to the Illumina 610, 550 and 300 chips with a genotype missing data rate ≤0.01, Hardy-Weinberg $P \geq 1 \times 10^{-5}$ and a minor allele frequency ≥0.01. As a result, 506 individuals were excluded (note that this includes 311 individuals who were included in both the Coriell and ALS control group).

We also sought to detect non-European ancestry. To this end, genotype data from SNPs typed in all cohorts was merged with genotypes at the same SNPs from 210 unrelated European (CEU), Asian (CHB and JPT) and Yoruban (YRI) samples from the HapMap project. Subsequent to removing SNPs in extensive regions of LD (Chr 5:44–51.5 Mb; Chr 6: 25–33.5 Mb; Chr 8: 8–12 Mb; Chr 11: 45–57 Mb)[54], we further excluded SNPs if any pair within a 50-SNP window had $r^2 > 0.2$. Genome-wide average identity-by-state (IBS) distance was calculated in PLINK between each pair of individuals in the resulting dataset, based on 57,966 SNPs (all with a genotype missing data rate ≤0.01, Hardy-Weinberg $P \geq 1 \times 10^{-5}$ and a minor allele frequency ≥0.01). The resulting matrix of IBS distances was used as input for classical multidimensional scaling (MDS) in R version 2.7.1. When the first two dimensions were extracted and plotted against each other, three clusters were observed as corresponding to the European, Asian and Yoruban samples. Sixteen samples appeared to be ethnic outliers from the European cluster and were excluded from further analysis.

We assessed population structure within the data using principal components analysis as implemented in EIGENSTRAT[55] to infer continuous axes of genetic variation. Eigenvectors were calculated based on the previously described LD-pruned subset of 57,966 SNPs common to all arrays. The EIGENSTRAT program also identifies genetic outliers, which are defined as individuals whose ancestry is at least 6 s.d. from the mean on one of the top ten axes of variation. As a result of this analysis, 188 outliers were identified and excluded. Following sample QC, 3,941 Alzheimer's disease cases and 7,848 controls were included in the analysis.

Stage 1 SNP quality control. Due to unresolved genotype-calling issues with a proportion of SNPs on the sex chromosomes, only autosomal SNPs were included in this analysis (with the exception of rs5984894, an X-chromosome SNP that has previously been associated with Alzheimer's disease[5]; **Supplementary Table 5**). Individuals either were genotyped on the Illumina 610-quad chip as part of this project, or had been previously genotyped on the Illumina HumanHap550 or the Illumina HumanHap300 array, with the genotypes made available to us. Note that SNPs had already been filtered out of some groups before inclusion in this study. Moreover, where different versions of the same array were used (for example, HumanHap550v1 used to genotype the 1958 birth cohort (Sanger) cohort compared with the HumanHap550v3 array used to genotype the 1958 birth cohort (T1DGC)), only SNPs common to both versions were considered as present on that array. As such, SNPs included in our analysis fell into four different categories: (i) 266,714 SNPs common to all three arrays and genotyped in all individuals; (ii) 202,516 SNPs common to the 610 and 550 arrays, but not present or without genotypes in individuals typed on the 300 array; (iii) 7,744 SNPs common to the 610 and 300 arrays, but not present or without genotypes in individuals typed on the 550 array; (iv) 105,614 SNPs with genotypes only in the 610 data.

We assessed the effects of different missing data rates and Hardy-Weinberg filters, aiming to remove poorly performing SNPs without excluding markers that may show genuine association with Alzheimer's disease. For each of the four SNP categories, markers were excluded if they had a minor allele frequency (MAF) <0.01 or a Hardy-Weinberg $P \leq 1 \times 10^{-5}$ in either cases or controls. SNPs with a MAF ≥0.05 were excluded if they had a genotype missing rate of >0.03 in either cases or controls; for SNPs with a MAF between 0.01 and 0.05, a more stringent genotype missing rate threshold of 0.01 was used. As a result of this basic SNP QC, 43,542 SNPs were excluded.

Ten principal components (PCs) were extracted using EIGENSTRAT, as previously described. To determine whether the PCs could assuage any population structure within our sample, we performed logistic regression tests of association with Alzheimer's disease, sequentially including between 0 and 10 of the top PCs as covariates. The impact of including the PCs was evaluated by calculating the genomic control inflation factor, λ[11]. We found that including the first four PCs as covariates had the maximum impact on λ.

To minimize inter-chip and inter-cohort differences that could result in an inflation of type I error rate, minor allele frequencies were compared between controls in the different groups using logistic regression analysis, incorporating the top four PCs as covariates as previously described. Comparisons were performed only between individuals from the same geographical region (that is, British Isles, Germany or USA). For each of the four categories of SNPs, a quantile-quantile (Q-Q) plot was produced for each cohort control comparison, and the significance threshold used to exclude SNPs was based on where the observed χ^2 statistics departed from the null expectation. A further 9,828 SNPs were excluded as a result of these comparisons. Thus, a total of 529,218 autosomal SNPs were analyzed for association with Alzheimer's disease in this study.

Stage 1 statistical analysis. SNPs were tested for association with Alzheimer's disease using logistic regression, assuming an additive model. Covariates were included in the logistic regression analysis to allow for geographical region and chip, that is, to distinguish between (i) individuals from the British Isles, (ii) individuals from Germany, (iii) individuals from the United States typed on the 610 or 550 chip and (iv) individuals from the United States typed on the 300 chip. It was not possible to include a covariate for each chip, as only controls were genotyped on the 550 chip. Similarly, it was not possible to include a covariate for each of the eight groups, as only two included both cases and controls (610 and Mayo groups). The first four PCs extracted from EIGENSTRAT were also included as covariates, as previously described. Following analysis, 130 cluster plots were visually inspected for SNPs with a P value ≤ 1×10^{-4}. Thirteen SNPs showing poorly formed clusters were excluded. Thus, our analysis was based on 529,205 SNPs, and a conservative genome-wide significance threshold of $0.05/529205 = 9.4 \times 10^{-8}$ was used. The overall λ was calculated to be 1.037. Population attributable risk (the expected reduction in disease load following removal of a risk factor) was calculated for GWS SNPs according to the following formula:

PAR = $F_{con}(OR − 1)/[F_{con}(OR − 1) + 1]$, where F_{con} is the frequency of the risk allele in controls and OR is the odds ratio associated with the risk allele[56].

Expected number of significant SNPs. We assessed our results to determine whether we observed more significant SNPs than would be expected by chance. We first removed SNPs within 500 kb to either side of risk SNPs of rs429358 (the *APOE* ε4 SNP), rs11136000 (*CLU*) and rs3851179 (*PICALM*). We thus excluded 170 '*APOE*' SNPs, 290 '*CLU*' SNPs and 257 '*PICALM*' SNPs. Of the 528,448 remaining SNPs, we estimated 397,224.7 'independent' tests using the algorithm we described in ref. 57. Of 16 SNPs significant at a level of $\alpha = 10^{-5}$ (excluding *APOE*, *CLU* and *PICALM* SNPs, see **Table 2**), we estimated 12.6 'independent' tests. We calculated the mean ($N \times \alpha = 397224.7 \times 10^{-5} \approx 4.0$) and variance ($N \times \alpha \times (1 − \alpha) = 3.97$) of the expected number of significant tests at $\alpha = 10^{-5}$ level using the binomial distribution. Thus the probability of observing 12.6 significant tests is $P = 7.5 \times 10^{-6}$.

Stage 2 genotyping and statistical analysis. We genotyped SNPs in cases and controls from five European cohorts (described in **Supplementary Table 3**). Putative functional SNPs were identified using PupaSuite[58]. Genotyping was performed at Cardiff using the MassARRAY and iPlexGOLD systems (Sequenom) according to manufacturer's recommendations. All genotyped SNPs had genotype call frequency rates >90% in the follow-up sample, and no SNPs had HWE $P \leq 0.05$ in cases or controls. SNPs were tested for association with Alzheimer's disease using logistic regression, assuming an additive model. Covariates were included in the logistic regression analysis to allow for each cohort: (i) Belgian, (ii) MRC, (iii) ART, (iv) Bonn, (v) Greek.

Meta-analysis. We included genotype data from stages 1 and 2 in a meta-analysis for SNPs at the *CLU* and *PICALM* loci. In addition, we used genotype data from the Reiman *et al.* study and the Li *et al.* study, two publicly available Alzheimer's disease GWAS datasets. The Reiman *et al.* sample comprises 861 Alzheimer's disease cases and 550 controls genotyped on the Affymetrix 500K chip. The Li *et al.* discovery sample comprises 753 Alzheimer's disease cases and 736 controls also genotyped on the Affymetrix 500K chip. If a SNP of interest was not genotyped in our GWAS or the Reiman *et al.* dataset, an attempt was made to impute genotypes in PLINK, using the 60 HapMap CEU founders as a reference panel. Only imputed SNPs with an information content metric value >0.8 were included in analysis. Individual level data were not available for the Li *et al.* study, and so SNPs could not be imputed. Where individual level data was available, SNPs were tested for association with Alzheimer's disease using logistic regression, assuming an additive model. Covariates were included in the logistic regression analysis to allow for geographical region and chip as in stage 1 and for cohort as in stage 2. Covariates included for the Reiman *et al.* sample distinguished between samples from the Netherlands Brain Bank and samples from the United States. Meta-analyses incorporating data from the Li *et al.* study are based on Mantel-Haenzsel χ^2 tests. Results of the meta-analysis are shown in **Table 1** and **Supplementary Table 4**.

Secondary analyses. We also tested the GWS SNPs for relationships with age at onset (AAO). To this end, age at onset (in years) was used as the dependent variable in a linear regression analysis and an additive model was assumed. AAO data was available for 2,856 Alzheimer's disease cases. Covariates were included in the logistic regression analysis to allow for geographical region and chip, that is, to distinguish between (i) cases from the British Isles, (ii) cases from Germany, (iii) cases from the United States typed on the 610 chip and (iv) cases from the United States typed on the 300 chip.

In addition, we stratified our sample based on presence or absence of at least one *APOE* ε4 allele. We had *APOE* genotype data for 6,045 individuals; our ε4-positive sample consisted of 2,203 Alzheimer's disease cases and 632 controls; our ε4-negative sample consisted of 1,446 cases and 1,764 controls. We performed genome-wide tests for association with Alzheimer's disease in each subsample, but no SNP achieved genome-wide significance (see **Supplementary Tables 6** and **7**).

The GWAS genotyping from this study will be made available to researchers within 6 months. Please contact the corresponding authors for access.

URLs. VIB Genetic Service Facility, http://www.vibgeneticservicefacility.be; Haploview, http://www.broad.mit.edu/mpg/haploview; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink; R, http://www.r-project.org.

50. Teo, Y.Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741–2746 (2007).
51. Clayton, D.G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **37**, 1243–1246 (2005).
52. Moskvina, V., Craddock, N., Holmans, P., Owen, M.J. & O'Donovan, M.C. Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum. Hered.* **61**, 55–64 (2006).
53. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
54. Price, A.L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–139, author reply 135–139 (2008).
55. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
56. Schlesselman, J.J. *Case-Control Studies: Design, Conduct, Analysis* (Oxford University Press, Oxford, UK, 1982).
57. Moskvina, V. & Schmidt, K.M. On multiple-testing correction in genome-wide association studies. *Genet. Epidemiol.* **32**, 567–573 (2008).
58. Conde, L. *et al.* PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res.* **34**, W621–W625 (2006).