



King's Research Portal

DOI:

[10.1038/ng.440](https://doi.org/10.1038/ng.440)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., ... Williams, J. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics*, 41(10), 1088-1093. [10.1038/ng.440](https://doi.org/10.1038/ng.440)

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Published in final edited form as:

Nat Genet. 2009 October ; 41(10): 1088–1093. doi:10.1038/ng.440.

Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease, and shows evidence for additional susceptibility genes

Denise Harold^{1,*}, Richard Abraham^{1,*}, Paul Hollingworth^{1,*}, Rebecca Sims¹, Amy Gerrish¹, Marian Hamshere¹, Jaspreet Singh Pahwa¹, Valentina Moskvina¹, Kimberley Dowzell¹, Amy Williams¹, Nicola Jones¹, Charlene Thomas¹, Alexandra Stretton¹, Angharad Morgan¹, Simon Lovestone², John Powell³, Petroula Proitsi³, Michelle K Lupton³, Carol Brayne⁴, David C. Rubinsztein⁵, Michael Gill⁶, Brian Lawlor⁶, Aoibhinn Lynch⁶, Kevin Morgan⁷, Kristelle Brown⁷, Peter Passmore⁸, David Craig⁸, Bernadette McGuinness⁸, Stephen Todd⁸, Clive Holmes⁹, David Mann¹⁰, A. David Smith¹¹, Seth Love¹², Patrick G. Kehoe¹², John Hardy¹³, Simon Mead¹⁴, Nick Fox¹⁵, Martin Rossor¹⁵, John Collinge¹⁴, Wolfgang Maier¹⁶, Frank Jessen¹⁶, Britta Schürmann¹⁶, Hendrik van den Bussche¹⁷, Isabella Heuser¹⁸, Johannes Kornhuber¹⁹, Jens Wiltfang²⁰, Martin Dichgans^{21,22}, Lutz Frölich²³, Harald Hampel^{24,25}, Michael Hüll²⁶, Dan Rujescu²⁵, Alison Goate²⁷, John S.K. Kauwe²⁸, Carlos Cruchaga²⁷, Petra Nowotny²⁷, John C. Morris²⁷, Kevin Mayo²⁷, Kristel Slegers^{29,30}, Karolien Bettens^{29,30}, Sebastiaan Engelborghs^{30,31}, Peter De Deyn^{30,31}, Christine van Broeckhoven^{29,30}, Gill Livingston³², Nicholas J. Bass³², Hugh Gurling³², Andrew McQuillin³², Rhian Gwilliam³³, Panagiotis Deloukas³³, Ammar Al-Chalabi³⁴, Christopher E. Shaw³⁴, Magda Tsolaki³⁵, Andrew Singleton³⁶, Rita Guerreiro³⁶, Thomas W. Mühleisen^{37,38}, Markus M. Nöthen^{37,38}, Susanne Moebus³⁹, Karl-Heinz Jöckel³⁹, Norman Klopp⁴⁰, H-Erich Wichmann^{40,41}, Minerva M. Carrasquillo⁴², V. Shane Pankratz⁴³, Steven G. Younkin⁴², Peter Holmans¹, Michael O'Donovan¹, Michael J. Owen^{1,†}, and Julie Williams^{1,†}

¹ MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK. ² NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS 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,

† Corresponding authors.

*These authors contributed equally to this work

Author Contributions

JW (Cardiff), MJO, MOD directed this study, assisted by RA, PH, and PAH. DH and JW (Cardiff) took primary responsibility for drafting the manuscript assisted by RA, PH, RS, AG, MOD and MJO. JW (Cardiff), RA, PH, RS, AG, KD, AW, NJ, CT, AS, ARM, SL (IOP), JP, PP, MKL, CB, DCR, MG, BL, AL, KM (Notts), KSB, PAP, DC, BM, ST, CH, DM, ADS, SL (Bristol) PGK, JH, SM (UCL), NF, MR, JC, WM, FJ, BS, HvB, IH, JK, JW (Essen), MD, LF, AMG, JSK, CC, PN, JCM, KM (WashU), GL, NJB, HG and AM 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. RA and PH were responsible for the coordination, collection, transit and selection of samples for genotyping from the '610 group'. RG (Sanger) and PD were responsible for procedures related to genotyping the 610 group on the Illumina platform. AA, CES, ABS, RG (NIA), TWM, MMN, SM, KJ, NK, HW, MMC, SVP and SGY 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'). KS, KB, SE, PDD, CVB and MT contributed towards sample collection, diagnosis and preparation of case-control material for the stage 2 'replication sample'. Replication genotyping was coordinated and performed by RA, assisted by RS and AG. JSP developed the database for the GWA project in which the data were stored. DH completed statistical quality control and produced association statistics, under the supervision of MH (Cardiff), VM and PAH. All authors discussed the results and approved the manuscript.

Accession numbers

GenBank: *CLU* isoform 1 mRNA, NM_001831.2; *CLU* isoform 2 mRNA, NM_203339.1; *PICALM* isoform 1 mRNA, NM_007166.2; *PICALM* isoform 2 mRNA, NM_001008660.1.

University of Cambridge, Cambridge, UK. 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, NG7 2UH, UK. ⁸ Ageing Group, Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University 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 (OPTIMA), University of Oxford, Level 4, John Radcliffe Hospital, Oxford OX3 9DU, 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, Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK. ¹⁵ Dementia Research Centre, Department of Neurodegenerative Diseases, University College London, Institute of Neurology, London, UK. ¹⁶ Department of Psychiatry, University of Bonn, Sigmund-Freud-Str. 25, D-53105 Bonn, Germany. ¹⁷ Institute of Primary Medical Care, University Medical Center Hamburg-Eppendorf, Germany. ¹⁸ Department of Psychiatry, Charité Berlin, Germany. ¹⁹ Department of Psychiatry and Psychotherapy, University of Erlangen-Nürnberg, Germany. ²⁰ LVR-Hospital Essen, Department of Psychiatry and Psychotherapy, University Duisburg-Essen, Germany. ²¹ Institute for Stroke and Dementia Reserach, Klinikum der Universität München, Marchioninstr. 15, 81377, Munich, Germany. ²² Department of Neurology, Klinikum der Universität München, Marchioninstr. 15, 81377, Munich, Germany. ²³ Department of Geriatric Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Square J5, D-68161 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, 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, Germany. ²⁷ Departments of Psychiatry, Neurology and Genetics, Washington University School of Medicine, St Louis, MO 63110, US. ²⁸ Department of Biology, Brigham Young University, Provo, UT, 84602, 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, ZNA Middelheim, Antwerpen, Belgium. ³² Department of Mental Health Sciences, University College 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, SE5 8AF, UK. ³⁵ 3rd Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece. ³⁶ Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, 20892, USA. ³⁷ Department of Genomics, Life & Brain Center, University of Bonn, Sigmund-Freud-Str. 25, D-53127 Bonn, Germany. ³⁸ Institute of Human Genetics, University of Bonn, Wilhelmstr. 31, D-53111 Bonn, Germany. ³⁹ Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, University Duisburg-Essen, Hufelandstr. 55, D-45147 Essen, Germany. ⁴⁰ Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany. ⁴¹ Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. ⁴² Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, Florida 32224, USA. ⁴³ Division of Biomedical Statistics and Informatics, Mayo Clinic and Mayo Foundation, Rochester, Minnesota 55905, USA.

Abstract

We undertook a two-stage genome-wide association study of Alzheimer's disease involving over 16,000 individuals. In stage 1 (3,941 cases and 7,848 controls), we replicated the established association with the *APOE* locus (most significant SNP: rs2075650, $p=1.8\times 10^{-157}$) and observed genome-wide significant association with SNPs at two novel loci: rs11136000 in the *CLU* or *APOJ* gene ($p=1.4\times 10^{-9}$) and rs3851179, a SNP 5' to the *PICALM* gene ($p=1.9\times 10^{-8}$). Both novel associations were supported in stage 2 (2,023 cases and 2,340 controls), producing compelling evidence for association with AD 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). We also observed more variants associated at $p<1\times 10^{-5}$ than expected by chance ($p=7.5\times 10^{-6}$), including polymorphisms at the *BINI*, *DAB1* and *CR1* loci.

Alzheimer's disease (AD) is the most common form of dementia, is highly heritable (heritability of up to 76%) but genetically complex. Neuropathologically, the disease is characterized by extracellular senile plaques containing β -amyloid (A β) and intracellular neurofibrillary tangles containing hyperphosphorylated τ protein. Four genes have been definitively implicated in its etiology. Mutations of the *amyloid precursor protein* (*APP*) gene and the *presenilin 1* and *2* genes (*PSEN1*, *PSEN2*) cause rare, Mendelian forms of the disease usually with an early-onset. However, in the more common form of AD, only *apolipoprotein E* (*APOE*) has been established unequivocally as a susceptibility gene. Aiming to identify novel AD loci, several genome-wide association studies (GWAS) have been conducted prior to the present study. All have identified strong evidence for association to *APOE*, but less convincing evidence implicating other genes²⁻⁹. This outcome is consistent with the majority of findings from GWAS of other common phenotypes, where susceptibility alleles typically have effect sizes with odds ratios (OR) of 1.5 or less, rather than of the magnitude for *APOE* (OR~3). Detecting such modest effects requires much larger samples than those that have been applied in the GWAS of AD to date¹⁰, which have all included fewer than 1,100 cases. Based upon the hypothesis that risk alleles for AD are likely to confer ORs in the range seen in other common diseases, we undertook a more powerful GWAS than has been undertaken to date.

We established a collaborative consortium from Europe and the USA 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, 14,639 subjects were genotyped on Illumina platforms. 5,715 samples were genotyped for the present study using the Illumina 610-quadchip; genotypes for the remaining subjects were either made available to us from population control datasets or through collaboration and were genotyped on the Illumina HumanHap550 or the 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 AD cases, 7,848 controls of which 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 stratification.

In addition to the known association with the *APOE* locus, GWA analysis identified two novel loci at a genome-wide level of significance (see Fig. 1). Table 1 shows SNPs which were genome-wide significant (GWS) in stage 1; 13 GWS SNPs map within or close to the *APOE* locus on chromosome 19 ($p<3\times 10^{-8}$ – 2×10^{-157}) and the top 5 are shown in Table 1 (see Supplementary Table 2 for the complete list). The other two SNPs represent novel associations. One of the novel SNPs (rs11136000) is located within an intron of clusterin (*CLU*, also known as *APOJ*) on chromosome 8 ($p=1.4\times 10^{-9}$, OR=0.840); the other SNP (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 novel GWS SNPs were genotyped in an independent sample comprising 2,023 AD 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 and 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 Translational Genomics Research Institute (TGEN) study⁹ and the Li *et al.* study⁸, two publicly available AD GWAS datasets, but neither of the novel GWS SNPs had been genotyped or could be imputed. As secondary analyses, we tested each novel finding for interaction with *APOE* status and for association with age at onset. No significant interactions of the novel SNPs with *APOE* status were observed influencing AD risk (rs11136000x*APOE*- ϵ 4 interaction $p=0.674$; rs3851179x*APOE*- ϵ 4 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 novel GWS SNP or that might plausibly have functional effects (see Supplementary Table 4). A synonymous SNP (rs7982) in the *CLU* gene 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 AD in the whole sample (meta- $p=8\times 10^{-10}$; stage 1 genotypes were imputed). This SNP is in exon 5 of the *CLU* gene, which codes for part of the beta chain of the protein and may influence a predicted exon splicing enhancer. We note that Tycko and colleagues¹² previously published a negative association study on the *CLU* gene, analyzing 4 SNPs in an AD case-control sample of African-American, Hispanic and Caucasian/non-Hispanic individuals. Although they identified rs7982 through mutation screening (referred to as VB in their study), the SNP was not tested for association in their sample. The 4 SNPs that were analyzed were rare in Caucasians (minor allele frequency <2%) and there was very limited power to detect association in their Caucasian sub-sample of 53 AD cases and 43 controls. As these 4 SNPs were not genotyped in the present 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 the gene 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 (*i.e.* rs561655: meta- $p=1\times 10^{-7}$ and rs592297: meta- $p=2\times 10^{-7}$). A number of SNPs in LD with rs3851179 and showing moderate evidence of association in the GWAS ($p<1\times 10^{-4}$) were also followed up, most notably rs541458. This SNP is 8kb 5' to the *PICALM* gene and was directly genotyped in the extension sample, the TGEN study and the Li *et al.* study, with $p<0.05$ in each. Following meta-analysis, this is one of our most significant SNPs (meta- $p=8\times 10^{-10}$) and is also supported by the study of Amouyel *et al.* published in this issue ($p=3\times 10^{-3}$). Further genetic analyses will be required to characterize the true nature of the associations observed at these loci, which is beyond the scope of this paper.

We also tested whether the observed 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 (see Methods) we focused on those which showed most evidence for association ($p<1\times 10^{-5}$). Approximately 13 independent signals were observed; less than 4 would be expected by chance ($p=7.5\times 10^{-6}$). Table 2 shows the loci implicated

and provides strong evidence for association with the complement receptor 1 (*CR1*) gene, particularly in the context of the results of the Amouyel *et al.* study published in this issue. Also noteworthy are the bridging integrator 1 (*BIN1*) gene, which produces a protein involved in synaptic vesicle endocytosis¹³, and the disabled homolog 1 (*DABI*) gene, whose product is involved with tyrosine phosphorylation and microtubule function in neurons¹⁴. These data thus provide strong evidence that there are several genes associated with AD which remain to be identified. We have also tested over 100 variants highlighted by previous AD GWA studies 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 AD. It is therefore intriguing that we find compelling evidence for association with *CLU*, a gene that encodes another major brain apolipoprotein¹⁵, suggesting that susceptibility genes are not randomly distributed through functional pathways. The predominant form of clusterin is a secreted heterodimeric glycoprotein of 75-80kDa. The single copy gene spans about 16kb on chromosome 8p21-p12 and encodes an mRNA of approximately 2kb that translates into a 449 amino acid primary polypeptide chain. Clusterin is expressed in all mammalian tissues¹⁶ and there is strong evidence that *CLU* levels are elevated in a number of pathological conditions involving injury or chronic inflammation of the brain¹⁷. In Alzheimer's disease brain, *CLU* expression is reported to be increased in affected cortical areas and is present in amyloid plaques and in the cerebrospinal fluid of AD cases¹⁷⁻²⁰.

Clusterin is a multi-functional 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 barrier²¹⁻²⁴. Interestingly, APOE also appears to act as a molecular chaperone for Aβ and influences when it aggregates and deposits²⁵ as well as influencing Aβ conformation and toxicity^{26,27}. In a similar way, clusterin appears to regulate both the toxicity and conversion of Aβ into insoluble forms²⁸⁻³². Furthermore, APOE and CLU have been shown to cooperate in suppressing Aβ deposition³³ and APOE and CLU may critically modify Aβ clearance at the blood brain barrier, suggesting a role for clusterin in the amyloidogenic pathway³⁴. Levels of APOE protein appear to be inversely proportional to *APOE-ε4* allele dose levels, *i.e.* expression levels are reduced in ε4 homozygotes compared with heterozygotes. Conversely, *CLU* levels are increased in proportion to *APOE-ε4* allele dose suggesting an induction of clusterin in individuals with low APOE levels³⁵. Thus, the strong statistical evidence for the involvement of this gene in AD has additional support in terms of biological plausibility.

The second gene locus to show compelling evidence for association with AD 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 post synaptic structures. It has been shown that like BIN1, *PICALM* is involved in clathrin-mediated endocytosis (CME), an essential step in the intracellular trafficking of proteins and lipids such as nutrients, growth factors and neurotransmitters³⁶⁻³⁸. Of relevance to AD, *PICALM* appears to be involved in directing the trafficking of VAMP2. VAMP2 is a SNARE protein that plays a prominent role in the fusion of synaptic vesicles to the presynaptic membrane in neurotransmitter release, a process that is crucial to neuronal function³⁹. AD brains show a reduced number of synapses and stereological and biochemical analysis has shown that this reduction in synaptic density correlates with cognitive defects better than the accumulation of plaques and tangles⁴⁰. More recent analysis indicates synapses within AD brains may be dysfunctional even before they visibly degenerate⁴¹. 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 AD. Alternatively, PICALM could influence risk of AD through APP processing via endocytic pathways resulting in changes in A β levels. Cell culture experiments have shown that full length APP is retrieved from the cell surface by CME⁴² and inhibition of endocytosis reduces APP internalization and reduces A β production and release^{43,44}. Increased synaptic activity is known to lead to the elevated endocytosis of synaptic vesicle proteins and Cirrito *et al.* have since provided evidence *in vivo* that the increased CME, triggered by increased synaptic activity, drives more APP into endocytotic compartments resulting in an increase of A β production and release⁴⁵. Thus, as for clusterin, the strong statistical evidence for the involvement of PICALM in AD has support in terms of biological plausibility.

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 over-estimates of the true effects⁴⁶. In the extension study, rs11136000 had an OR of 0.905, while rs3851179 had an OR of 0.897. Assuming that these 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, while failing to reach this stringent level of statistical significance, might still reflect true associations with disease, *e.g.* variants at the *CR1* locus, also highlighted by the results of Amouyel and colleagues (this issue). As with other GWAS research into complex disorders, such as Type 2 diabetes where 18 susceptibility loci have now been confirmed through GWAS meta-analysis of 54,000 subjects⁴⁷, further larger GWAS are required to identify new susceptibility variants for AD.

Methods

Sample ascertainment and diagnostic criteria

The study comprised a Stage 1 discovery sample of 4,957 AD cases and 9,682 controls and a Stage 2 follow-up sample of 2,023 AD 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. 200ng of input DNA per sample were used and prepared for genotyping using the Illumina Infinium™ system (Illumina® Inc., San Diego, CA, USA). Manufacturer's protocols were followed throughout. The Illuminus algorithm for cluster analysis was used for genotype calling⁴⁹.

Stage 1: Individual Quality Control

4,113 AD cases and 1,602 controls were genotyped on the Illumina 610-quad chip as part of this study (the 610 group). In addition, 844 AD 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 7 different studies, making 8 separate groups in total: 1) 610; 2) Mayo; 3) 1958 birth cohort (Sanger); 4) 1958 birth cohort (T1DGC); 5) ALS control; 6) Coriell control; 7) Heinz Nixdorf Recall (HNR) study; 8) KORA F4. As we used genotype data from multiple sources, it was important to apply stringent QC filters, as differential genotyping error rates between groups could result in spurious associations when the data are combined^{50,51}. These filters were applied separately to each of these 8 groups to remove poorly performing samples using tools implemented in PLINK v1.0552.

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. 71 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 $P < 1 \times 10^{-5}$ and a minor allele frequency > 0.01 . As a result, 506 individuals were excluded (note that this includes 311 individuals that 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 linkage disequilibrium (chr5:44-51.5 Mb; chr6: 25-33.5 Mb; chr8: 8-12 Mb; chr11: 45-57 Mb)⁵³, we further pruned 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 $P < 1 \times 10^{-5}$ and a minor allele frequency > 0.01). The resulting matrix of IBS distances was used as input for classical multi-dimensional scaling (MDS) in R v2.7.1. When the first two dimensions were extracted and plotted against each other, three clusters were observed 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 (PCA) as implemented in EIGENSTRAT⁵⁴ 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 standard deviations from the mean on one of the top ten axes of variation. As a result, 188 outliers were identified and excluded. Following sample QC 3,941 AD 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 AD by Carrasquillo *et al.*⁵; see Supplementary Table 5). Individuals were genotyped on either the Illumina 610-quad as part of this project, or were previously genotyped on the Illumina HumanHap550 or the Illumina HumanHap300 array, and the genotypes made available to us. Note that SNPs had already been filtered out of some groups prior to inclusion in this study. Moreover, where different versions of the same array were used (*e.g.* 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 4 different categories; 1) 266,714 SNPs common to all 3 arrays and genotyped in all individuals; 2) 202,516 SNPs common to the 610 and 550 arrays, but not present or without genotypes in individuals typed on the 300 array; 3) 7,744 SNPs common

to the 610 and 300 arrays, but not present or without genotypes in individuals typed on the 550 array; 4) 105,614 SNPs with genotypes only in the 610 data.

We assessed the effects of different missing data rate and Hardy-Weinberg filters, aiming to remove poorly performing SNPs without excluding markers that may show genuine association with AD. For each of the 4 SNP categories, markers were excluded if they had a minor allele frequency (MAF) < 0.01 or a Hardy-Weinberg $P = 1 \times 10^{-5}$, 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 employed. 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 if the PCs could assuage any population structure within our sample, we performed logistic regression tests of association with AD, 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 . We found that including the first 4 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 4 PCs as covariates as previously described. Comparisons were only performed between individuals from the same geographical region (*i.e.* British Isles, Germany or USA). For each of the 4 categories of SNPs, a quantile-quantile (Q-Q) plot was produced for each cohort control comparison, and the significance threshold employed 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 AD in this study.

Stage 1 Statistical Analysis

SNPs were tested for association with AD using logistic regression, assuming an additive model. Covariates were included in the logistic regression analysis to allow for geographical region and chip, *i.e.* to distinguish between 1) individuals from the British Isles, 2) individuals from Germany, 3) individuals from the US typed on the 610 or 550 chip, 4) individuals from the US 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 8 groups, as only two included both cases and controls (610 and Mayo groups). The first 4 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 $< 1 \times 10^{-4}$. 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 employed. The overall genomic control inflation factor, 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 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.

Expected number of significant SNPs

We assessed our results to determine if we observed more significant SNPs than would be expected by chance. We first removed SNPs within 500 kb either side of risk SNPs, *i.e.*

rs429358 (the *APOE* $\epsilon 4$ 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 56. Of 16 SNPs significant at a significance level $\alpha=10^{-5}$ (excluding *APOE*, *CLU* and *PICALM* SNPs, see Table 2) we estimated 12.6 “independent” tests. We calculated the mean ($N*\alpha = 397224.7*10^{-5} \approx 4.0$) and variance ($N*\alpha*(1-\alpha) = 3.97$) 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 5 European cohorts (described in Supplementary Table 3). Putative functional SNPs were identified using PupaSuite57. Genotyping was performed at Cardiff using the MassARRAY and iPLEXGOLD systems (Sequenom, San Diego, CA) 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-value ≤ 0.05 in cases or controls. SNPs were tested for association with AD using logistic regression, assuming an additive model. Covariates were included in the logistic regression analysis to allow for each cohort, i.e. 1) Belgium, 2) MRC, 3) ART, 4) Bonn, 5) 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 employed genotype data from the TGEN study and the Li *et al.* study, two publicly available AD GWAS datasets. The TGEN sample is comprised of 861 AD cases and 550 controls genotyped on the Affymetrix 500K chip. The Li *et al.* discovery sample is comprised of 753 AD cases and 736 controls also genotyped on the Affymetrix 500K chip. If a SNP of interest was not genotyped in our GWAS or the TGEN 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 greater than 0.8 were included in analysis (see PLINK website). Individual level data were not available for the Li study and so SNPs could not be imputed. Where individual level data was available, SNPs were tested for association with AD 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 TGEN sample distinguished between samples from the Netherlands Brain Bank and samples from the USA. Meta-analyses incorporating data from the Li *et al.* study are based on Mantel-Haenszel 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 employed as the dependent variable in a linear regression analysis and an additive model was assumed. AAO data was available for 2,856 AD cases. Covariates were included in the logistic regression analysis to allow for geographical region and chip, i.e. to distinguish between 1) cases from the British Isles, 2) cases from Germany, 3) cases from the US typed on the 610 chip, 4) cases from the US typed on the 300 chip.

In addition, we stratified our sample based on presence/absence of at least 1 *APOE* $\epsilon 4$ allele. We had *APOE* genotype data for 6045 individuals; our $\epsilon 4$ -positive sample consisted of 2,203 AD cases and 632 controls; our $\epsilon 4$ -negative sample consisted of 1,446 cases and 1,764 controls. We performed genome-wide tests for association with AD in each sub-

sample, but no SNP achieved genome-wide significance (see Supplementary Tables 6 and 7).

URLs

VIB Genetic Service Facility, <http://www.vibgeneticservicefacility.be>

TGEN, <http://www.tgen.org>

Haploview, <http://www.broad.mit.edu/mpg/haploview>

PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink>

R, <http://www.r-project.org>

The GWAS data will be made available to *bona fide* researchers within 6 months (contact the corresponding authors for access).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the patients and families who took part in this research. Cardiff University was supported by the Wellcome Trust, Medical Research Council (MRC), Alzheimer's Research Trust (ART) and the Welsh Assembly Government. ART supported sample collections at the Institute of Psychiatry, the South West Dementia Bank, Universities of Cambridge, Nottingham, Manchester and Belfast. The Belfast group acknowledges support from the Alzheimer's Society, Ulster Garden Villages, N.Ireland R&D Office and the Royal College of Physicians/Dunhill Medical Trust. The MRC and Mercer's Institute for Research on Ageing supported the Trinity College group. The South West Dementia Brain Bank acknowledges support from Bristol Research into Alzheimer's and Care of the Elderly. The Charles Wolfson Charitable Trust supported the OPTIMA group. Ammar Al-Chalabi and Christopher Shaw thank the MND and MRC for support. DCR is a Wellcome Trust Senior Clinical Research Fellow. Washington University was funded by NIH grants, Barnes Jewish Foundation and the Charles and Joanne Knight Alzheimer's Research Initiative. The Mayo GWAS was supported by NIH grants, the Robert and Clarice Smith and Abigail Van Buren AD Research Program and the Palumbo Professorship in AD Research. Patient recruitment for the MRC Prion Unit/UCL Department of Neurodegenerative Disease collection was supported by the UCLH/UCL Biomedical Centre. LASER-AD was funded by Lundbeck SA. The Bonn group was supported by the German Federal Ministry of Education and Research (BMBF), Competence Network Dementia and Competence Network Degenerative Dementia, and by the Alfried Krupp von Bohlen und Halbach-Stiftung. The KORA F4 studies were financed by Helmholtz Zentrum München; German Research Center for Environmental Health; BMBF; German National Genome Research Network and the Munich Center of Health Sciences. The Heinz Nixdorf Recall cohort was funded by the Heinz Nixdorf Foundation (Dr. jur. G.Schmidt, Chairman) and BMBF. Coriell Cell Repositories is supported by NINDS and the Intramural Research Program of the National Institute on Aging. We acknowledge use of DNA from the 1958 Birth Cohort collection, funded by the MRC and the Wellcome Trust which was genotyped by the Wellcome Trust Case Control Consortium and the Type-1 Diabetes Genetics Consortium, sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development and Juvenile Diabetes Research Foundation International. The Antwerp site was supported by the VIB Genetic Service Facility, the Biobank of the Institute Born-Bunge, the Special Research Fund of the University of Antwerp; the Fund for Scientific Research-Flanders, the Foundation for Alzheimer Research, the Interuniversity Attraction Poles program P6/43 of the Belgian Federal Science Policy Office. K.S. is a postdoctoral fellow and K.B. a PhD fellow (Fund for Scientific Research-Flanders). We thank Robert Brown, John Landers, Donald Warden, Donald Lehmann, Nigel Leigh, James Uphill, Jon Beck, Tracy Campbell, Steffi Klier, Gary Adamson, Julia Wyatt, Marc Lucia Perez, T. Meitinger, P. Lichtner, G. Eckstein, Neill Graff-Radford, Ronald Petersen, Dennis Dickson, G. Fischer, Horst Bickel, Michael Hüll, Holger Jahn, Hanna Kaduszkiewicz, Christian Luckhaus, Steffi Riedel-Heller, Stefanie Wolf, Siegfried Weyerer, the Helmholtz Zentrum München genotyping staff, Eric Reiman, TGEN and the NIMH AD Genetics Initiative. We thank Advanced Research Computing @Cardiff (ARCCA) who facilitated data analysis.

References

1. Avramopoulos D. Genetics of Alzheimer's disease: recent advances. *Genome Med.* 2009; 1:34. [PubMed: 19341505]
2. Abraham R, et al. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics.* 2008; 1:44. [PubMed: 18823527]
3. Beecham GW, et al. Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am J Hum Genet.* 2009; 84:35–43. [PubMed: 19118814]
4. Bertram L, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet.* 2008; 83:623–32. [PubMed: 18976728]
5. Carrasquillo MM, et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet.* 2009; 41:192–8. [PubMed: 19136949]
6. Coon KD, 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.* 2007; 68:613–8. [PubMed: 17474819]
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.* 2007; 16:865–73. [PubMed: 17317784]
8. Li H, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol.* 2008; 65:45–53. [PubMed: 17998437]
9. Reiman EM, et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron.* 2007; 54:713–20. [PubMed: 17553421]
10. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447:661–78. [PubMed: 17554300]
11. Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999; 55:997–1004. [PubMed: 11315092]
12. Tycko B, et al. Polymorphisms in the human apolipoprotein-J/clusterin gene: ethnic variation and distribution in Alzheimer's disease. *Hum Genet.* 1996; 98:430–6. [PubMed: 8792817]
13. Cousin MA, Robinson PJ. The dephosphins: dephosphorylation by calcineurin triggers synaptic vesicle endocytosis. *Trends Neurosci.* 2001; 24:659–65. [PubMed: 11672811]
14. 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.* 1999; 24:481–9. [PubMed: 10571241]
15. Jenne DE, Tschopp J. Clusterin: the intriguing guises of a widely expressed glycoprotein. *Trends Biochem Sci.* 1992; 17:154–9. [PubMed: 1585460]
16. Jones SE, Jomary C. Clusterin. *Int J Biochem Cell Biol.* 2002; 34:427–31. [PubMed: 11906815]
17. Calero M, et al. Apolipoprotein J (clusterin) and Alzheimer's disease. *Microsc Res Tech.* 2000; 50:305–15. [PubMed: 10936885]
18. Giannakopoulos P, et al. Possible neuroprotective role of clusterin in Alzheimer's disease: a quantitative immunocytochemical study. *Acta Neuropathol.* 1998; 95:387–94. [PubMed: 9560017]
19. Liang WS, et al. Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. *Physiol Genomics.* 2008; 33:240–56. [PubMed: 18270320]
20. McGeer PL, Kawamata T, Walker DG. Distribution of clusterin in Alzheimer brain tissue. *Brain Res.* 1992; 579:337–41. [PubMed: 1378350]
21. 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.* 1993; 293(Pt 1):27–30. [PubMed: 8328966]
22. Golabek A, Marques MA, Lalowski M, Wisniewski T. Amyloid beta binding proteins in vitro and in normal human cerebrospinal fluid. *Neurosci Lett.* 1995; 191:79–82. [PubMed: 7659297]
23. Zlokovic BV, et al. Brain uptake of circulating apolipoproteins J and E complexed to Alzheimer's amyloid beta. *Biochem Biophys Res Commun.* 1994; 205:1431–7. [PubMed: 7802679]

24. Zlokovic BV, 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 U S A*. 1996; 93:4229–34. [PubMed: 8633046]
25. Bales KR, et al. Apolipoprotein E is essential for amyloid deposition in the APP(V717F) transgenic mouse model of Alzheimer's disease. *Proc Natl Acad Sci U S A*. 1999; 96:15233–8. [PubMed: 10611368]
26. Castano EM, et al. Fibrillogenesis in Alzheimer's disease of amyloid beta peptides and apolipoprotein E. *Biochem J*. 1995; 306(Pt 2):599–604. [PubMed: 7534068]
27. Ma J, Yee A, Brewer HB Jr, Das S, Potter H. Amyloid-associated proteins alpha 1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer beta-protein into filaments. *Nature*. 1994; 372:92–4. [PubMed: 7969426]
28. Boggs LN, et al. Clusterin (Apo J) protects against in vitro amyloid-beta (1-40) neurotoxicity. *J Neurochem*. 1996; 67:1324–7. [PubMed: 8752142]
29. DeMattos RB, 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 U S A*. 2002; 99:10843–8. [PubMed: 12145324]
30. Lambert MP, et al. Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. *Proc Natl Acad Sci U S A*. 1998; 95:6448–53. [PubMed: 9600986]
31. Matsubara E, Soto C, Governale S, Frangione B, Ghiso J. Apolipoprotein J and Alzheimer's amyloid beta solubility. *Biochem J*. 1996; 316(Pt 2):671–9. [PubMed: 8687416]
32. 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*. 1995; 136:22–31. [PubMed: 7589331]
33. DeMattos RB, et al. ApoE and clusterin cooperatively suppress Abeta levels and deposition: evidence that ApoE regulates extracellular Abeta metabolism in vivo. *Neuron*. 2004; 41:193–202. [PubMed: 14741101]
34. Bell RD, 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*. 2007; 27:909–18. [PubMed: 17077814]
35. Bertrand P, Poirier J, Oda T, Finch CE, Pasinetti GM. Association of apolipoprotein E genotype with brain levels of apolipoprotein E and apolipoprotein J (clusterin) in Alzheimer disease. *Brain Res Mol Brain Res*. 1995; 33:174–8. [PubMed: 8774959]
36. Dreyling MH, 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 U S A*. 1996; 93:4804–9. [PubMed: 8643484]
37. Tebar F, Bohlander SK, 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*. 1999; 10:2687–702. [PubMed: 10436022]
38. Yao PJ, Petralia RS, Bushlin I, Wang Y, Furukawa K. Synaptic distribution of the endocytic accessory proteins AP180 and CALM. *J Comp Neurol*. 2005; 481:58–69. [PubMed: 15558718]
39. Harel A, Wu F, Mattson MP, Morris CM, Yao PJ. Evidence for CALM in directing VAMP2 trafficking. *Traffic*. 2008; 9:417–29. [PubMed: 18182011]
40. Masliah E, et al. Altered expression of synaptic proteins occurs early during progression of Alzheimer's disease. *Neurology*. 2001; 56:127–9. [PubMed: 11148253]
41. Fitzjohn SM, 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*. 2001; 21:4691–8. [PubMed: 11425896]
42. Nordstedt C, Caporaso GL, Thyberg J, Gandy SE, Greengard P. Identification of the Alzheimer beta/A4 amyloid precursor protein in clathrin-coated vesicles purified from PC12 cells. *J Biol Chem*. 1993; 268:608–12. [PubMed: 8416966]

43. Carey RM, Balcz BA, Lopez-Coviella I, Slack BE. Inhibition of dynamin-dependent endocytosis increases shedding of the amyloid precursor protein ectodomain and reduces generation of amyloid beta protein. *BMC Cell Biol.* 2005; 6:30. [PubMed: 16095541]
44. Koo EH, Squazzo SL. Evidence that production and release of amyloid beta-protein involves the endocytic pathway. *J Biol Chem.* 1994; 269:17386–9. [PubMed: 8021238]
45. Cirrito JR, et al. Endocytosis is required for synaptic activity-dependent release of amyloid-beta in vivo. *Neuron.* 2008; 58:42–51. [PubMed: 18400162]
46. Goring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet.* 2001; 69:1357–69. [PubMed: 11593451]
47. 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.* 2008; 40:638–45. [PubMed: 18372903]
48. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21:263–5. [PubMed: 15297300]
49. Teo YY, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics.* 2007; 23:2741–6. [PubMed: 17846035]
50. Clayton DG, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet.* 2005; 37:1243–6. [PubMed: 16228001]
51. Moskvina V, Craddock N, Holmans P, Owen MJ, O'Donovan MC. Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered.* 2006; 61:55–64. [PubMed: 16612103]
52. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–75. [PubMed: 17701901]
53. Price AL, et al. Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet.* 2008; 83:132–5. author reply 135–9. [PubMed: 18606306]
54. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–9. [PubMed: 16862161]
55. Schlesselman, JJ. *Case-Control Studies: Design, Conduct, Analysis.* University Press; Oxford: 1982.
56. Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. *Genet Epidemiol.* 2008; 32:567–73. [PubMed: 18425821]
57. Conde L, et al. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res.* 2006; 34:W621–5. [PubMed: 16845085]

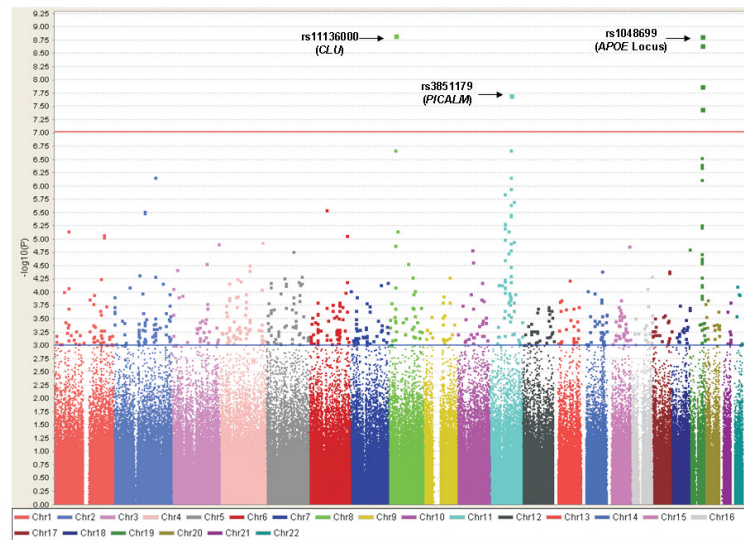


Figure 1. Scatterplot of chromosomal position (x-axis) against $-\log_{10}$ GWAS P-value (y-axis). The y-axis scale has been limited to 9.25 ($p = 5.6 \times 10^{-10}$), although highly significant association was observed with SNPs in the vicinity of the *APOE* locus (e.g. rs2075650 with $p = 1.8 \times 10^{-157}$). The threshold for genome-wide significance ($p = 9.4 \times 10^{-8}$) is indicated by the red horizontal line. 761 SNPs with $p < 1 \times 10^{-3}$ lie above the blue horizontal line and are listed in Supplementary Table 2. The plot was produced using Haploview v4.048.

Table 1

SNPs showing genome-wide significant association with AD in stage 1 of the GWAS. 5 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).

SNP	Chr	Closest RefSeq Gene	Location Relative to Gene	MAF	GWAS: 3941 cases 7848 controls		Extension: 2023 cases 2340 controls		Combined: 5964 cases 10188 controls		Population Attributable Risk (%)
					GWAS OR (95% CI)	P-value (two-tailed)	Extension OR (95% CI)	P-value (one-tailed)	Extension OR (95% CI)	P-value (two-tailed)	
rs2075650 *	19	<i>TOMM40</i>	Intron	0.15	1.8×10^{-157}	2.53 (2.37-2.71)					18.7%
rs157580	19	<i>TOMM40</i>	Intron	0.39	9.6×10^{-54}	0.63 (0.59-0.66)					
rs6859	19	<i>PVRL2</i>	3' UTR	0.43	6.9×10^{-41}	1.46 (1.38-1.54)					
rs8106922	19	<i>TOMM40</i>	Intron	0.40	5.4×10^{-39}	0.68 (0.64-0.72)					
rs405509	19	<i>APOE</i>	5'	0.52	4.9×10^{-37}	0.70 (0.66-0.74)					
rs11136000	8	<i>CLU</i>	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	<i>PICALM</i>	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%

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 ($r^2 = 0.48$).

Table 2

SNPs showing association with AD at $P < 1 \times 10^{-5}$ (excluding SNPs at the *APOE*, *CLU* and *PICALM* loci).

SNP	Chr	MB	Closest RefSeq Gene	Location Relative to Gene	GWAS P-value	OR	95% CI
rs11894266	2	170.3	<i>SSB</i>	5'	6.9×10^{-7}	0.86	0.81-0.91
rs610932	11	59.7	<i>MS4A6A</i>	3'UTR	1.4×10^{-6}	0.87	0.82-0.92
rs10501927	11	99.3	<i>CNTN5</i>	Intronic	2.0×10^{-6}	1.18	1.10-1.26
rs9446432	6	72.4		Intergenic	2.8×10^{-6}	1.28	1.15-1.41
rs7561528	2	127.6	<i>BINI</i>	5'	3.0×10^{-6}	1.16	1.09-1.24
rs744373	2	127.6	<i>BINI</i>	5'	3.2×10^{-6}	1.17	1.09-1.25
rs662196	11	59.7	<i>MS4A6A</i>	Intronic	5.2×10^{-6}	0.88	0.83-0.93
rs583791	11	59.7	<i>MS4A6A</i>	Intronic	5.3×10^{-6}	0.88	0.83-0.93
rs676309	11	59.8	<i>MS4A4E</i>	5'	6.3×10^{-6}	1.14	1.08-1.20
rs1157242	8	37.2		Intergenic	7.0×10^{-6}	1.17	1.10-1.26
rs1539053	1	57.9	<i>DABI</i>	Intronic	7.1×10^{-6}	0.88	0.83-0.93
rs11827375	11	76.0	<i>CHIT1orf30</i>	3'	7.2×10^{-6}	1.23	1.12-1.35
rs1408077	1	205.9	<i>CRI</i>	Intronic	8.3×10^{-6}	1.17	1.09-1.25
rs9384428	6	156.5		Intergenic	8.5×10^{-6}	1.14	1.08-1.21
rs6701713	1	205.9	<i>CRI</i>	Intronic	8.7×10^{-6}	1.17	1.09-1.25
rs3818361	1	205.9	<i>CRI</i>	Intronic	9.2×10^{-6}	1.17	1.09-1.25

Chr = Chromosome; MB = position in megabases; OR = odds ratio for the minor allele; 95% CI = 95% confidence interval; UTR= untranslated region.