

ARTICLE

Open Access

GWAS on family history of Alzheimer's disease

Riccardo E. Marioni^{1,2}, Sarah E. Harris^{1,2}, Qian Zhang³, Allan F. McRae³, Saskia P. Hagenaars^{2,4}, W. David Hill^{2,5}, Gail Davies², Craig W. Ritchie⁶, Catharine R. Gale^{2,5,7}, John M. Starr^{2,8}, Alison M. Goate⁹, David J. Porteous^{1,2}, Jian Yang^{3,10}, Kathryn L. Evans¹, Ian J. Deary^{2,5}, Naomi R. Wray^{3,10} and Peter M. Visscher^{2,3,10}

Abstract

Alzheimer's disease (AD) is a public health priority for the 21st century. Risk reduction currently revolves around lifestyle changes with much research trying to elucidate the biological underpinnings. We show that self-report of parental history of Alzheimer's dementia for case ascertainment in a genome-wide association study of 314,278 participants from UK Biobank (27,696 maternal cases, 14,338 paternal cases) is a valid proxy for an AD genetic study. After meta-analysing with published consortium data ($n = 74,046$ with 25,580 cases across the discovery and replication analyses), three new AD-associated loci ($P < 5 \times 10^{-8}$) are identified. These contain genes relevant for AD and neurodegeneration: *ADAM10*, *BCKDK/KAT8* and *ACE*. Novel gene-based loci include drug targets such as *VKORC1* (warfarin dose). We report evidence that the association of SNPs in the *TOMM40* gene with AD is potentially mediated by both gene expression and DNA methylation in the prefrontal cortex. However, it is likely that multiple variants are affecting the trait and gene methylation/expression. Our discovered loci may help to elucidate the biological mechanisms underlying AD and, as they contain genes that are drug targets for other diseases and disorders, warrant further exploration for potential precision medicine applications.

Introduction

The genetic epidemiology of late-onset Alzheimer's disease (LOAD) has advanced over the last decade¹, with >20 independent loci associated with the disease in addition to *APOE*². Presently, the largest meta-analytic genome-wide association study (GWAS) for LOAD employed a two-stage study design. First, 17,008 cases were compared to 37,154 controls. A total of 11,632 single-nucleotide polymorphisms (SNPs) with $P < 1 \times 10^{-3}$ from this meta-analysis were included in the second stage that compared 8,572 cases to 11,312 controls. A meta-analysis of the SNPs included in stages 1 and 2 was also performed³.

One difficulty in traditional studies of AD is case ascertainment⁴—either directly for prevalent cases or indirectly through prospective cohort studies for incident cases. A recent GWAS study on a subset of the UK Biobank cohort used information from family history (parent or first-degree relative with AD or dementia) as a proxy-phenotype for the participants⁵. When meta-analysed with the GWAS summary data highlighted above³, four new loci were identified.

The UK Biobank proxy-phenotype AD question, which is used here, does not incorporate biomarker data that are required for a clinical diagnosis. However, it is easy to administer at scale and we show that it has a near-unit genetic correlation with the AD results from the LOAD meta-analysis³, where many of the samples also lacked a confirmed diagnosis by biomarker levels or autopsy.

In the present study, we related proxy-phenotype information on dementia (i.e., reporting a parent with Alzheimer's dementia or dementia) to genetic data from

Correspondence: Riccardo E. Marioni (riccardo.marioni@ed.ac.uk)

¹Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

²Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK

Full list of author information is available at the end of the article



314,278 individuals from the UK Biobank cohort to identify new AD-associated loci. This sample includes individuals from the previous proxy-phenotype AD study by Liu et al.⁵ GWA studies were conducted separately for maternal and paternal AD due to a 1.7-fold difference in disease prevalence—9.6% and 5.5%, respectively. The summary statistics from these models were meta-analysed with those from the largest publicly available case–control study³. Sensitivity analyses showed that an overlap of controls in the maternal and paternal GWAS did not bias the results. Genetic correlation analysis showed the self-reported measure of parental AD to be an accurate proxy for clinical diagnosis, validating the global meta-analysis. In addition, we tested for causal evidence of our SNP–AD associations being mediated through gene expression and DNA methylation in the prefrontal cortex.

Subjects and methods

UK Biobank cohort

UK Biobank data⁶ (<http://www.ukbiobank.ac.uk>) were collected on over 500,000 individuals aged between 37 and 73 years from across Great Britain (England, Wales and Scotland) at the study baseline (2006–2010), including health, cognitive and genetic data.

The Research Ethics Committee (REC) granted ethical approval for the study—reference 11/NW/0382—and the current analysis was conducted under data application 10,279.

Genotyping

Genotyping details for the UK Biobank cohort have been reported previously^{7,8}. Briefly, two custom genotyping arrays were utilised with 49,950 participants typed using the UK BiLEVE Axiom Array and 438,427 participants typed using the UK Biobank Axiom Array^{7,8}. The released genotyped data contained 805,426 markers on 488,377 individuals. Imputed genotypes were supplied with the UK Biobank data with the Haplotype Reference Consortium (HRC) used as the imputation reference panel⁷.

Downstream quality control steps conducted for the current analysis included removing (1) those with non-British ancestry based on both self-report and a principal components analysis, (2) outliers based on heterozygosity and missingness, (3) individuals with sex chromosome configurations that were neither XX nor XY, (4) individuals whose reported sex did not match inferred sex from their genetic data and (5) individuals with >10 putative third-degree relatives from the kinship table. This left a sample of 408,095 individuals. To remove the possibility of double contributions from sibs, whose parents will have the same AD status, we first considered a list of all participants with a relative ($N=131,790$). A genetic relationship matrix was built for these individuals using

GCTA-GRM⁹ and a relationship threshold of 0.025 was applied to exclude related individuals. After removing one person from each pair of related individuals, the sample size was 332,050. Quality control thresholds applied to the GWAS included: minor allele frequency >0.01, imputation quality score >0.3 and restriction to HRC-imputed SNPs, leaving a total of 7,795,605 SNPs for the GWAS.

Phenotypes

Family history of Alzheimer's disease was ascertained via self-report. Participants were asked “Has/did your father ever suffer from Alzheimer's disease/dementia?” and “Has/did your mother ever suffer from Alzheimer's disease/dementia?” Self-report data from the initial assessment visit (2006–2010), the first repeat assessment visit (2012–2013) and the imaging visit (2014+) were aggregated with exclusions made for participants whose parents were: aged under 60 years; dead before reaching age 60 years; without age information. After merging with the genetic data, this left 27,696 cases of maternal AD with 260,980 controls, and 14,338 cases of paternal AD with 245,941 controls. There were 314,278 instances where AD information was available on at least one parent. Given the expected difference in disease prevalence due to sex differences in longevity—AD prevalence was 1.7-fold higher in mothers compared to fathers—GWA studies were performed separately for maternal and paternal AD.

Genome-wide association study

The GWA studies were conducted using BGENIE⁷. The outcome variable was the residuals from a linear regression model of maternal or paternal AD on age of parent at death or at time of the offspring's self-report, assessment centre, genotype batch, array and 40 genetic principal components. The predictor variable was the autosomal SNP and an additive model was considered.

The GWAS linear regression coefficients were converted to odds ratios using observed sample prevalences of 0.096 and 0.055 for maternal and paternal AD, respectively¹⁰. Subsequently, the log-odds were multiplied by two so that the effect sizes are reported on the same scale as a traditional case–control design⁵. Briefly, the conversion to odds ratios uses the following equation, derived in Lloyd-Jones et al.¹⁰, where k =disease prevalence, p =population allele frequency and β =the estimated SNP regression coefficient on the binary disease scale from the GWAS: $OR = \frac{(k+\beta(1-p)) \times (1-k+\beta p)}{(k-\beta p) \times (1-k-\beta(1-p))}$. SEs for the log-odds were then calculated based on the adjusted OR and the P -value from the initial GWAS (Supplementary Note 1). The ORs and SEs were then carried forward to a SE-weighted meta-analysis in METAL¹¹, first to create a UK Biobank parental meta-analysis, and then with the stage 2 summary output from the International

Genomics of Alzheimer's Project (IGAP) study³ and the stage 1 output for the SNPs that did not contribute to stage 2. Linkage disequilibrium score (LDSC) regression was used to estimate the genetic correlation between the maternal and paternal AD GWAS results and to test for residual confounding in the meta-analysis by examining the LDSC intercept^{12,13}.

The number of independent loci from the meta-analysis was determined by using the default settings in FUMA¹⁴. Independent lead SNPs had $P < 5 \times 10^{-8}$ and were independent at $r^2 < 0.6$. Within this pool of independent SNPs, lead SNPs were defined as those in LD at $r^2 < 0.1$. Loci were defined by combining lead SNPs within a 250 kb window and all SNPs in LD of at least $r^2 = 0.6$ with one of the independent SNPs. A gene-based analysis was carried out on all SNP output using the MAGMA software¹⁵ with default settings (SNP-wise (mean) model for each gene), and assuming a constant sample size for all genes. A Bonferroni-adjusted P -value of $0.05/18,251 = 2.7 \times 10^{-6}$ was used to identify significant genes. The 1000 genomes phase 3 data¹⁶ were used to map LD in both the independent locus and MAGMA analyses.

Summary data-based Mendelian randomisation

To test for pleiotropic associations between SNPs and AD and gene expression/DNA methylation in the brain, summary data-based Mendelian randomisation (SMR) was performed¹⁷. GWAS summary output from the meta-analysis of UK Biobank and IGAP (sample size specified as $314,278 + 74,046 = 388,324$) were included along with expression Quantitative Trait Loci (QTL) summary output from the Common Mind Consortium, which contains data on >600 dorsolateral prefrontal cortex samples, and DNA methylation QTL summary output on 258 prefrontal cortex samples (age >13)¹⁸. The reference genotypes were based on the Health and Retirement Study, imputed to the 1000 Genomes phase 1 reference panel. SNP exclusions included: imputation score <0.3, Hardy-Weinberg P -value < 1×10^{-6} and a minor allele frequency <0.01. Related individuals, based on a genomic-relationship matrix cutoff of 0.05, were removed. Two sets of eQTL summary data were considered (1) after adjustment for diagnosis, institution, sex, age of death, post-mortem interval, RNA integrity number (RIN), RIN^2 , and clustered library batch (2) with additional adjustments for 20 surrogate variables to account for additional possible confounders. Five ancestry vectors were included as covariates in the eQTL analyses. Further details are available at: <https://www.synapse.org/#!Synapse:syn4622659>. Default parameters for the SMR analysis were used and *cis* eQTLs/methQTLs were considered for analysis. Bonferroni-corrected P -value thresholds were applied ($P < 0.05/2011 = 2.5 \times 10^{-5}$ for eQTL data set 1, $P < 0.05/4380 = 1.1 \times 10^{-5}$ for eQTL data set 2 and $P < 0.05/54,624 = 9.2 \times 10^{-7}$ for the methQTL data set). The SMR P -value highlights candidate transcripts or methylation sites

through which a *cis* SNP may be acting on the outcome, AD. The heterogeneity in dependent instruments (HEIDI) P -value indicates evidence for a single causal SNP (effect on AD is mediated through the transcript/methylation site if $P > 0.05$) or different SNPs affecting AD and the transcript/methylation site if $P < 0.05$.

Results

UK Biobank GWAS

There were 27,696 cases of maternal AD (260,980 controls, prevalence of 9.6%) and 14,338 cases of paternal AD (245,941 controls, prevalence of 5.5%) in the UK Biobank. The genetic correlation between maternal and paternal AD was not significantly different from unity ($r_g = 0.71$, SE 0.43), although the SE is large. Both traits had a high genetic correlation with the case-control summary output from the IGAP study: r_g with maternal and paternal AD was 0.91 (SE 0.24) and 0.66 (0.39), respectively, both not significantly different from unity but with large SEs.

Prior to meta-analysing the UK Biobank parental summary statistics with the IGAP output, we investigated the influence of overlapping proxy controls in UK Biobank. The P -values from a single GWAS of parental AD status (0, 1 or 2 parents with AD) correlated 0.99 with those from a meta-analysis of separate maternal AD and paternal AD; the regression of $-\log_{10} P$ -values on each other gave an intercept of 0 and a slope of 1. A meta-analysis of the summary statistics from the maternal and paternal results is therefore essentially equivalent to the analysis of parental AD status. The linear regression effect sizes from the GWAS were converted to odds ratios prior to the meta-analysis¹⁰.

UK Biobank paternal and maternal meta-analysis

Six genome-wide significant loci were identified in the UK Biobank paternal and maternal GWAS meta-analysis (Supplementary Table 1, GWAS summary statistics available at www.ccace.ed.ac.uk/node/335). All were located in established AD loci. The top lead SNPs were rs679515 ($P = 5.3 \times 10^{-9}$, chr1, *CRI* locus), rs6733839 ($P = 1.1 \times 10^{-27}$, chr2, *BINI* locus), rs7384878 ($P = 1.3 \times 10^{-10}$, chr7, *PILRA* locus), rs3851179 ($P = 1.8 \times 10^{-12}$, chr11, *PICALM* locus), rs3845261 ($P = 4.0 \times 10^{-8}$, chr17, *ZNF232* locus) and rs10119 ($P = 1.1 \times 10^{-307}$, chr19, *APOE/TOMM40* locus). The chromosome 17 locus is ~100 kb from the *SCIMP* gene-variant identified by Liu et al.⁵

UK Biobank parental and IGAP meta-analysis

The meta-analysis of the maternal and paternal AD history in UK Biobank with the IGAP data identified 69 lead SNPs and 218 independent significant SNPs with $P < 5 \times 10^{-8}$ from 21 genomic risk loci. The majority ($n = 42$)

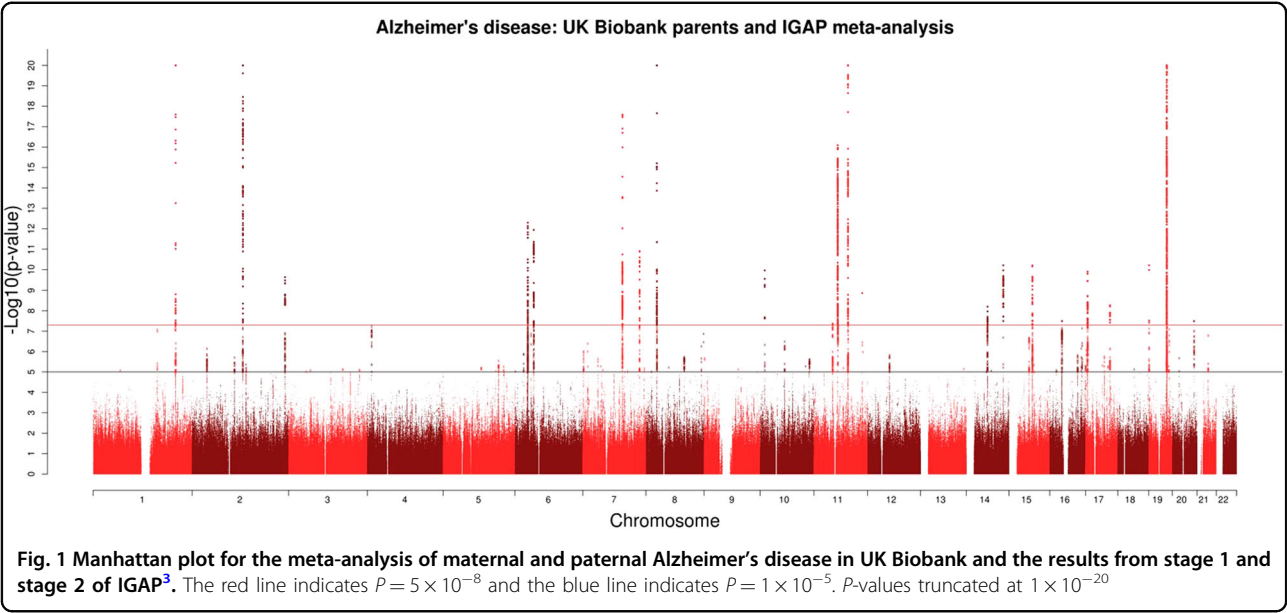


Table 1 Novel SNPs ($P < 5 \times 10^{-8}$) from the meta-analysis of UK Biobank parental history of Alzheimer's disease with results from IGAP

Locus	Chr	Lead SNP	A1	A2	BP	Freq	Locus start	Locus end	OR 95% CI	P	Direction ^a
1	15	rs593742	A	G	59045774	0.69	58873555	59120077	1.06 (1.04, 1.07)	6.20E-11	++?+
2	16	rs889555	T	C	31122571	0.29	30916129	31155458	0.95 (0.94, 0.97)	3.20E-08	--?--
3	17	rs6504163	T	C	61545779	0.63	61545779	61578207	1.05 (1.03, 1.07)	5.40E-09	+++?

Chr chromosome, BP base pair, Freq frequency of allele A1, OR odds ratio and 95% confidence interval
^aUK Biobank maternal AD, UK Biobank paternal AD, IGAP stage 1, IGAP stage 2

of the lead SNPs were located in the gene-dense *APOE/TOMM40* locus on chromosome 19 (Fig. 1 and Supplementary Tables 2 and 3; GWAS summary statistics are available for the 7,795,605 meta-analysed SNPs online: www.ccace.ed.ac.uk/node/335). The LDSC regression intercept term from the meta-analysis summary output was 1.021 (SE 0.0099), and around 85% of the genomic inflation ($\lambda = 1.09$) was due to the polygenic signal, rather than sources of confounding including population stratification and cryptic relatedness.

Novel genome-wide significant loci

Of the 21 significant risk loci, three were novel (Table 1 and Supplementary Figures 1–3). The top lead SNPs were: rs593742 (*ADAM10*, locus chr15:58873555–59120077); rs889555 210 (*BCKDK/KAT8*, locus chr16:30916129–31155458); and rs6504163 (*ACE*, locus 211 chr17:61545779–61578207).

Replication of IGAP SNPs

Fifteen of the 21 previously reported SNPs³ associated with AD were genome-wide significant ($P < 5 \times 10^{-8}$) in

the current meta-analysis, with four other SNPs (rs2718058, rs10838725, rs17125944 and rs10498633) having $P < 1 \times 10^{-5}$ (Supplementary Table 4). The *MEF2C* variant, rs190982, had a meta-analysis P -value of 7.2×10^{-3} and rs8093731 (a *DSG2* variant), which was genome-wide significant in stage 1 but not stage 2 of IGAP, had a meta-analysis P -value of 0.07. There was complete sign-concordance between UK Biobank and IGAP for all 21 SNPs (Supplementary Table 4). The odds ratios between the maternal and paternal analysis for the top 21 IGAP SNPs were correlated $r = 0.91$. Both also correlated highly with the effect sizes reported in the IGAP analysis ($r = 0.87$ and 0.84 , respectively).

Gene-based analysis

A total of 92 genes were significant at a Bonferroni threshold of $P < 2.7 \times 10^{-6}$ (Supplementary Table 5). Gene ontology analysis, using a Fisher exact test to compare the number of significant genes in each gene set with the expected number, showed significant enrichment for 46 gene sets, including those linked to the regulation of amyloid-beta, neurofibrillary tangle assembly, immune

response, and cholesterol and lipid transport (Supplementary Table 6).

Summary data-based Mendelian randomisation

Pleiotropic associations between AD and gene expression in the brain were tested using SMR¹⁷. GWAS summary data for AD were taken from the UK Biobank and IGAP meta-analysis. eQTL summary data came from the Common Mind Consortium ($n > 600$ dorsolateral prefrontal cortex samples: data set 1 adjusted for age at death, sex and institution; data set 2 made additional adjustments for 20 surrogate variables). MethQTL data came from 258 dorsolateral prefrontal cortex samples (participants aged 13 years and older—adjustments were made for the first 5 genetic MDS components and first 11 methylation PCs)¹⁸. We found evidence of brain expression and DNA methylation associated with AD in the *TOMM40* gene (part of the *APOE/TOMM40* cluster on chromosome 19) in both the second eQTL model and also in the methQTL model (Supplementary Tables 7–9). However, the HEIDI *P*-values were < 0.05 for both analyses, indicating that the associations were unlikely to be driven by a single causal variant affecting both expression/methylation and AD.

Different top QTL SNPs in *TOMM40* were identified in the two analyses: rs760136 and rs7259620 although they were in perfect LD in European samples^{19,20} ($R^2 = 1.00$). The SNPs have differing LD patterns with the *APOE* allele defining SNPs, rs7412 ($R^2 = 0.001$, $D' = 0.03$) and rs429358 ($R^2 = 0.15$, $D' = 1.0$). A significant SMR association with HEIDI *P*-value > 0.05 , indicating pleiotropy, was observed for *KAT8* in the eQTL analysis (Supplementary Table 8 and Supplementary Figure 4) and for *EXOC3L2* (chr19, *APOE/TOMM40* locus) and *STAG3* (chr7, *PILRA* locus) in the methQTL analysis (Supplementary Table 9 and Supplementary Figures 5 and 6).

Discussion

Using recently established proxy-phenotype methods for case ascertainment, we show this approach to be valid in GWAS of AD. Meta-analysing the proxy-phenotype summary statistics from UK Biobank with those from a case-control study (IGAP), we identified three new genome-wide significant loci that contain several relevant genes.

ACE determines levels of angiotensin II, which has trophic actions within the brain and contributes to the regulation of cerebral blood flow²¹. Previous meta-analyses of candidate gene studies identified variants within *ACE* to be associated with AD, though not at genome-wide significance^{22,23}. *ACE* variants have also been linked to atrophy of the hippocampus and amygdala²⁴, and CSF-*ACE* protein levels correlate with CSF tau and phosphorylated tau^{25,26}.

ADAM10 is involved in the cleavage of amyloid-beta precursor protein²⁷, which is involved in the deposition of amyloid beta, a major neurological hallmark of AD. *ADAM10* has been proposed as potential therapeutic agent in AD therapy^{27,28}. Rare variants in *ADAM10* have also been linked to LOAD²⁹.

The *BCKDK/KAT8* locus contains the *VKORC1* gene, which was a genome-wide significant gene-based finding ($P = 3.4 \times 10^{-8}$). The *VKORC1* variant, rs9923231, whose T allele was suggestively associated with an increased risk of AD ($P = 6.9 \times 10^{-8}$), is strongly associated with the need for a reduced dose of warfarin anticoagulation^{30,31}. *KAT8* is regulated by *KANSL1*, which has been linked to AD as a genome-wide significant finding in *APOE* e4-negative individuals³².

Our cutoff for relatedness of 0.025 in UK Biobank was very stringent, in particular since disease status was on parents of genotyped individuals, so that correlations in proxy-phenotypes from relatives (e.g., cousins) are likely to be very small. When we used a more relaxed threshold of 0.4 (excluding first-degree relatives only), we increased our sample size to 385,869 and found four additional loci and top lead SNPs: rs4575098, $P = 2.7 \times 10^{-8}$, *ADAMTS4* (chr1:161097241–161156033); a gene desert on chr4 (rs6448453, $P = 1.2 \times 10^{-9}$, chr4:11026028–11040406) that is ~400 kb from the *CLNK* gene; in *CCDC6* on chr10 (rs1171812, $P = 3.5 \times 10^{-8}$, chr10:61631416–61710540); and in a poorly annotated region on chr16 (rs12444183, $P = 5.3 \times 10^{-9}$, chr16:81773003–81773816), proximal to the *PLCG2* gene. The corresponding *P*-values for these four lead SNPs in the GWAS meta-analysis with the more stringent relatedness cutoff were $P = 8.4 \times 10^{-8}$, 5.9×10^{-8} , 5.7×10^{-7} and 7.4×10^{-8} , respectively.

The gene-set enrichment analysis implicated pathways involved in AD neuropathology (amyloid and neurofibrillary tangles) in addition to immune response, lipid and cholesterol metabolism, as previously reported in a gene-set analysis on the IGAP data³³. A separate gene-set analysis on the meta-analysed paternal and maternal summary statistics from UK Biobank showed a highly overlapping set of enriched pathways (Supplementary Table 10).

An integrative analysis of eQTL and methQTL with the GWAS summary data identified one previously identified AD gene, *TOMM40*, as having its gene expression and methylation levels associated with AD. The most parsimonious explanation of these results is the existence of multiple causal variants, some affecting AD and others affecting expression or methylation. A previous SMR analysis of AD and LDL cholesterol identified evidence of 16 pleiotropic SNPs, 12 of which were located in the *APOE* region³⁴. Given the involvement of immunity in the aetiology of AD, as highlighted by the pathway analysis, it may be relevant to consider immune cells, such as monocytes, for the SMR analysis.

The main strength of the study is the proxy-phenotype approach, which resulted in over 42,000 proxy cases for the GWAS analysis. However, the question used to determine parental AD status may have resulted in some responders being unable to discriminate Alzheimer's disease and dementia from other dementia sub-types, which have different presentations and genetic architectures^{35,36}. This method of proxy-case ascertainment may have influenced the loci uncovered. Parental dementia status is partly dependent on longevity, with age being the biggest risk factor for AD. We partially controlled for this by excluding participants whose parents were younger than or died prior to reaching the age of 60 years when AD incidence is extremely low. The misclassification of case status via incorrect informant reporting will have reduced the power to detect true effects. This, along with a possible winner's curse effect for the IGAP study, might explain the reduction in the meta-analytic odds ratios compared to those previously reported³.

Conclusion

We identified three new AD-associated loci that have known and putative biological processes associated with Alzheimer's disease. These findings help to elucidate the biological mechanisms underlying AD and, given that some loci (*VKORC1*, *ACE*) are existing drug targets for other diseases and disorders, warrant further exploration for potential precision medicine and clinical trial applications.

Disclaimer

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Acknowledgements

R.E.M. is supported by Alzheimer's Research UK major project grant number ARUK-PG2017B-10. The research was conducted, using the UK Biobank Resource, in The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), part of the cross-council Lifelong Health and Wellbeing Initiative (MR/K026992/1); funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) are gratefully acknowledged. CCACE supports I.J.D., with some additional support from Dementias Platform UK (MR/L015382/1). This report represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The University of Queensland authors acknowledge funding from the Australian National Health and Medical Research Council (grants 1113400, 1078901 and 1078037) and the Charles & Sylvia Viertel Fellowship (J.Y.). W.D.H. is supported by a grant from Age UK (Disconnected Mind Project). A.G. is supported by the National Institutes of Health (AG049508, AG052411 and RF1AG054014) and the JPB Foundation. A.G. is on the scientific advisory board for Denali Therapeutics and has served as a consultant for AbbVie and Cognition Therapeutics.

Author details

¹Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. ²Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh,

Edinburgh EH8 9JZ, UK. ³Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia. ⁴Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK. ⁵Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK. ⁶Centre for Dementia Prevention, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH8 9YL, UK. ⁷MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton SO16 6YD, UK. ⁸Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK. ⁹Departments of Neuroscience, Neurology and Genetics and Genomic Sciences, Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY 10029-5674, USA. ¹⁰Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia

Authors' contributions

Conception and design of the work, drafting the manuscript: R.E.M. Interpretation: R.E.M., P.M.V., A.F.M. and Q.Z. Analysis: R.E.M., W.D.H. and Q.Z. Critical revision of the manuscript: all authors.

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41398-018-0150-6>.

Received: 19 March 2018 Revised: 4 April 2018 Accepted: 4 April 2018

Published online: 18 May 2018

References

- Huang, Y. & Mucke, L. Alzheimer mechanisms and therapeutic strategies. *Cell* **148**, 1204–1222 (2012).
- Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet. Med.* **18**, 421–430 (2016).
- Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
- Sibbett, R. A., Russ, T. C., Deary, I. J. & Starr, J. M. Dementia ascertainment using existing data in UK longitudinal and cohort studies: a systematic review of methodology. *BMC Psychiatry* **17**, 239 (2017).
- Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using family history of 363 disease. *Nat. Genet.* **49**, 325–331 (2017).
- Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. (2017) Preprint at <https://www.biorxiv.org/content/early/2017/07/20/166298>.
- Wain, L. V. et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med.* **3**, 769–781 (2015).
- Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
- Lloyd-Jones, L. R., Robinson, M. R., Yang, J. & Visscher, P. M. Transformation of summary statistics from linear mixed model association on all-or-none traits to odds ratio. *Genetics* **208**, 1397–1408 (2018).
- Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

14. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
15. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
16. 1000 Genomes Project Consortium et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
17. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
18. Jaffe, A. E. et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2016).
19. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
20. Machiela, M. J. & Chanock, S. J. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics* **34**, 887–889 (2017).
21. Starr, J. M. & Whalley, L. J. *ACE Inhibitors: Central Actions* (Raven Press, New York, 1994).
22. Elkins, J. S., Douglas, V. C. & Johnston, S. C. Alzheimer disease risk and genetic variation in ACE: a meta-analysis. *Neurology* **62**, 363–368 (2004).
23. Lehmann, D. J. et al. Large meta-analysis establishes the ACE insertion-deletion polymorphism as a marker of Alzheimer's disease. *Am. J. Epidemiol.* **162**, 305–317 (2005).
24. Sleegers, K. et al. ACE gene is associated with Alzheimer's disease and atrophy of hippocampus and amygdala. *Neurobiol. Aging* **26**, 1153–1159 (2005).
25. Jochimsen, H. M. et al. The association of angiotensin-converting enzyme with biomarkers for Alzheimer's disease. *Alzheimers Res. Ther.* **6**, 27 (2014).
26. Kauwe, J. S. K. et al. Genome-wide association study of CSF levels of 59 alzheimer's disease candidate proteins: significant associations with proteins involved in amyloid processing and inflammation. *PLoS Genet.* **10**, e1004758 (2014).
27. Suh, J. et al. ADAM10 missense mutations potentiate β -amyloid accumulation by impairing prodomain chaperone function. *Neuron* **80**, 385–401 (2013).
28. Yuan, X.-Z., Sun, S., Tan, C.-C., Yu, J.-T. & Tan, L. The role of ADAM10 in Alzheimer's disease. *J. Alzheimers Dis.* **58**, 303–322 (2017).
29. Kim, M. et al. Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate α -secretase activity. *Hum. Mol. Genet.* **18**, 3987–3996 (2009).
30. Takeuchi, F. et al. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet.* **5**, e1000433 (2009).
31. Wadelius, M. et al. Common VKORC1 and GGCC polymorphisms associated with warfarin dose. *Pharm. J.* **5**, 262–270 (2005).
32. Jun, G. et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol. Psychiatry* **21**, 108 (2016).
33. International Genomics of Alzheimer's Disease Consortium (IGAP). Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement.* **11**, 658–671 (2015).
34. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* **9**, 224 (2018).
35. Ferrari, R. et al. Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *J. Neurol. Neurosurg. Psychiatry* **88**, 152–164 (2017).
36. Ikram, M. A. et al. Genetics of vascular dementia—review from the ICVD working group. *BMC Med.* **15**, 48 (2017).