doi:10.1093/brain/awaa364 BRAIN 2021: 144; 924–937 | 924

BRAIN
A JOURNAL OF NEUROLOGY

Prediction of Alzheimer's disease using multi-variants from a Chinese genome-wide association study

Longfei Jia,[1] Fangyu Li,[1] Cuibai Wei,[1] Min Zhu,[1] Qiumin Qu,[2] Wei Qin,[1] Yi Tang,[1] Luxi Shen,[1] Yanjiang Wang,[3] Lu Shen,[4] Honglei Li,[5] Dantao Peng,[6] Lan Tan,[7] Benyan Luo,[8] Qihao Guo,[9] Muni Tang,[10] Yifeng Du,[11] Jiewen Zhang,[12] Junjian Zhang,[13] Jihui Lyu,[14] Ying Li,[1] Aihong Zhou,[1] Fen Wang,[1] Changbiao Chu,[1] Haiqing Song,[1] Liyong Wu,[1] Xiumei Zuo,[1] Yue Han,[1] Junhua Liang,[1] Qi Wang,[1] Hongmei Jin,[1] Wei Wang,[1] Yang Lü,[15] Fang Li,[16] Yuying Zhou,[17] Wei Zhang,[18,19] Zhengluan Liao,[20] Qiongqiong Qiu,[1] Yan Li,[1] Chaojun Kong,[1] Yan Li,[1] Haishan Jiao,[1] Jie Lu[21,22] and Jianping Jia[1,23,24,25]

Previous genome-wide association studies have identified dozens of susceptibility loci for sporadic Alzheimer's disease, but few of these loci have been validated in longitudinal cohorts. Establishing predictive models of Alzheimer's disease based on these novel variants is clinically important for verifying whether they have pathological functions and provide a useful tool for screening of disease risk. In the current study, we performed a two-stage genome-wide association study of 3913 patients with Alzheimer's disease and 7593 controls and identified four novel variants (rs3777215, rs6859823, rs234434, and rs2255835; $P_{combined} = 3.07 \times 10^{-19}$, 2.49×10^{-23}, 1.35×10^{-67}, and 4.81×10^{-9}, respectively) as well as nine variants in the apolipoprotein E region with genome-wide significance ($P < 5.0 \times 10^{-8}$). Literature mining suggested that these novel single nucleotide polymorphisms are related to amyloid precursor protein transport and metabolism, antioxidation, and neurogenesis. Based on their possible roles in the development of Alzheimer's disease, we used different combinations of these variants and the apolipoprotein E status and successively built 11 predictive models. The predictive models include relatively few single nucleotide polymorphisms useful for clinical practice, in which the maximum number was 13 and the minimum was only four. These predictive models were all significant and their peak of area under the curve reached 0.73 both in the first and second stages. Finally, these models were validated using a separate longitudinal cohort of 5474 individuals. The results showed that individuals carrying risk variants included in the models had a shorter latency and higher incidence of Alzheimer's disease, suggesting that our models can predict Alzheimer's disease onset in a population with genetic susceptibility. The effectiveness of the models for predicting Alzheimer's disease onset confirmed the contributions of these identified variants to disease pathogenesis. In conclusion, this is the first study to validate genome-wide association study-based predictive models for evaluating the risk of Alzheimer's disease onset in a large Chinese population. The clinical application of these models will be beneficial for individuals harbouring these risk variants, and particularly for young individuals seeking genetic consultation.

1 Innovation Center for Neurological Disorders and Department of Neurology, Xuanwu Hospital, Capital Medical University, National Clinical Research Center for Geriatric Diseases, Beijing, China
2 Department of Neurology, The First Affiliated Hospital of Xi'an Jiaotong University, Shaanxi, China
3 Department of Neurology and Center for Clinical Neuroscience, Daping Hospital, Third Military Medical University, Chongqing, China
4 Department of Neurology, Xiangya Hospital, Central South University, Changsha, China
5 Laboratory of Medical Neurobiology of Zhejiang Province, Zhejiang University School of Medicine, Zhejiang, China
6 Department of Neurology, China-Japan Friendship Hospital, Beijing, China

Received February 14, 2020. Revised July 30, 2020. Accepted August 14, 2020. Advance access publication November 14, 2020

7 Department of Neurology, Qingdao Municipal Hospital, School of Medicine, Qingdao University, Shandong, China
8 Department of Neurology, The First Affiliated Hospital, Zhejiang University, Zhejiang, China
9 Department of Gerontology, Shanghai Jiaotong University Affiliated Sixth People's Hospital, Shanghai, China
10 Department of Geriatrics, Guangzhou Huiai Hospital, Affiliated Hospital of Guangzhou Medical College, Guangzhou, China
11 Department of Neurology, Shandong Provincial Hospital Affiliated to Shandong University, Shandong, China
12 Department of Neurology, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan, China
13 Department of Neurology, Zhongnan Hospital, Wuhan University, Hubei, China
14 Center for Cognitive Disorders, Beijing Geriatric Hospital, Beijing, China
15 Department of Geriatrics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
16 Department of Geriatric, Fuxing Hospital, Capital Medical University, Beijing, China
17 Department of Neurology, Tianjin Huanhu Hospital, Tianjin, China
18 Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
19 Center for Cognitive Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
20 Department of Psychiatry, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou, Zhejiang, China
21 Department of Radiology, Xuanwu Hospital, Capital Medical University, Beijing, China
22 Department of Nuclear Medicine, Xuanwu Hospital, Capital Medical University, Beijing, China
23 Beijing Key Laboratory of Geriatric Cognitive Disorders, Beijing, China
24 Clinical Center for Neurodegenerative Disease and Memory Impairment, Capital Medical University, Beijing, China
25 Center of Alzheimer's Disease, Beijing Institute for Brain Disorders, Beijing, China

Correspondence to: Jianping Jia, MD, PhD
Innovation Center for Neurological Disorders and Department of Neurology, Xuanwu Hospital, Capital Medical University, National Clinical Research Center for Geriatric Diseases, Changchun Street 45, Xicheng District, Beijing, China, 100053
E-mail: jjp@ccmu.edu.cn

Correspondence may also be addressed to: Jie Lu, MD, PhD
Department of Radiology, Department of Nuclear Medicine, Xuanwu Hospital, Capital Medical University, Changchun Street 45, Xicheng District, Beijing, China, 100053
E-mail: imaginglu@hotmail.com

Keywords: Alzheimer's disease; genome-wide association study; Chinese; predictive model; longitudinal cohort

Abbreviations: AUC = area under the curve; eQTL = expression quantitative trait loci; GWAS = genome-wide association study; SNP = single nucleotide polymorphism

Introduction

Alzheimer's disease is the most common type of dementia and is genetically complex with an estimated heritability of 60–80% (Gatz *et al.*, 1997). Previous genome-wide association studies (GWASs) of Alzheimer's disease in Caucasian, African-American, and Asian populations have identified genetic risk variants in *ABCA7*, *BIN1*, *CASS4*, *CD2AP*, *CD33*, *CDK5RAP2*, *CELF1*, *CLU*, *COBL*, *CR1*, *ECHDC3*, *EPHA1*, *EXOC3L2*, *FERMT2*, *HLA-DRB5*, *HLA-DRB1*, *HS3ST1*, *INPP5D*, *KANSL1*, *MEF2C*, *MS4A*, *NME8*, *PICALM*, *PM20D1*, *PTK2B*, *SLC10A2*, *SLC24A4*, *SORL1*, *TREM2*, and *ZCWPW1* (Harold *et al.*, 2009; Lambert *et al.*, 2009, 2013*a*; Seshadri *et al.*, 2010; Hollingworth *et al.*, 2011; Naj *et al.*, 2011; Guerreiro *et al.*, 2013; Miyashita *et al.*, 2013; Reitz *et al.*, 2013; Desikan *et al.*, 2015; Jun *et al.*, 2016; Lacour *et al.*, 2017; Miron *et al.*, 2018; Sanchez-Mut *et al.*, 2018; Kunkle *et al.*, 2019). These variants affect several Alzheimer's disease-related processes, such as lipid metabolism, inflammation, innate immunity, production and clearance of amyloid-β, and endosomal vesicle recycling (Selkoe and Hardy, 2016). However, few of the variants reported in Caucasians have been identified in the Chinese population (Wang *et al.*, 2016). A recent whole genome sequencing study in a Chinese population identified variants in *GCH1* and *KCNJ15*, in addition to the well-known apolipoprotein E (*APOE*) locus; however, the sample size of this study was relatively small (Zhou *et al.*, 2018).

Recently, genetic predictive models have been established for predicting the onset of Alzheimer's disease using a polygenic risk score approach, which was used to reveal polygenetic contributions to Alzheimer's disease risk of common single nucleotide polymorphisms (SNPs) that show a disease association but fail to meet the accepted *P*-value threshold for genome-wide significance (Escott-Price *et al.*, 2015, 2017*a*, *b*, 2019; Chouraki *et al.*, 2016; Stocker *et al.*, 2018; Leonenko *et al.*, 2019). These studies showed variable results. Specifically, Escott-Price *et al.* reported that the area under the curve (AUC) of their predictive models, which included *APOE*, >80 000 SNPs, age, and sex as predictors, was 0.78, whereas in their other study, the AUC of their models including >20 000 SNPs and *APOE* as predictors increased to 0.84 as the included individuals were pathologically but not clinically confirmed (Escott-Price *et al.*, 2015, 2017*a*). However, despite the high predictive accuracy

of these polygenic risk score-based models, it may not be easy to use these models in a clinical setting because an individual may not carry so many risk variants. Thus, simple and effective Alzheimer's disease predictive models are needed for use as tools to screen for the genetic risk of Alzheimer's disease, particularly in young individuals who carry the risk variants.

The current study aimed to investigate novel Alzheimer's disease-related genetic variants in a GWAS, to establish predictive models based on these variants, and to validate the models in a longitudinal cohort. This approach can be applied for early intervention in individuals who are at a risk of developing Alzheimer's disease.

Materials and methods

Subjects

The two-stage GWAS study involved 3913 patients with Alzheimer's disease and 7593 controls from a Chinese population. The cohorts used in the two stages were independent of each other. Patients with Alzheimer's disease were recruited from the outpatient memory clinics at the Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing, China and 46 other participating hospitals across China from 2013 to 2018. All diagnoses of Alzheimer's disease in this study were based on the recommendations of the National Institute on Aging and the Alzheimer's Association workgroup (McKhann *et al.*, 2011) or National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association criteria (McKhann *et al.*, 1984), with an age-at-onset $\geqslant$60 years and no family history of dementia. Controls were recruited from the aforementioned medical centre hospitals. All controls were $\geqslant$60 years of age, cognitively normal (without subjective memory complaints, a Mini-Mental State Examination score of 26–30, and Clinical Dementia Rating Scale score of 0), and free of any general or laboratory evidence of diseases that could impact cognition. Demographic information was collected from each subject using a structured questionnaire.

Furthermore, using associated SNPs from the GWAS data, predictive models of Alzheimer's disease were generated by combining risk variants. To estimate the effectiveness of the predictive models, participants from a longitudinal cohort of the China Cognition and Aging Study (China COAST) (Jia *et al.*, 2014) were selected. China COAST was a longitudinal study established in 2008 as a multicentre cohort study comprising normal, mild cognitive impairment-, and Alzheimer's disease-affected individuals across 30 of 34 provinces in China with yearly follow-up. The inclusion criteria were as follows: (i) the individual was cognitively normal 10 years ago at baseline with indicative blood samples; (ii) the individual developed Alzheimer's disease at the time of sample collection for the present study 10 years later; and (iii) the individual had a detailed clinical data profile including psychometric evaluation every year during follow-up. Finally, 5474 participants were recruited, from among which 2358 developed Alzheimer's disease and 3116 were cognitively normal in 2019 (Supplementary Table 1). The study was approved by the Ethical Committees of Xuanwu Hospital, Capital Medical University. Written informed consent was obtained from either the subjects or their legal guardians according to the Declaration of Helsinki.

GWAS study

First stage

Genomic DNA was extracted from peripheral blood samples using a modified salting-out procedure (Nasiri *et al.*, 2005). In the first stage, we performed genome-wide genotyping of 1679 patients with Alzheimer's disease and 2508 controls using Illumina HumanOmniZhongHua-8 Bead Chips (Illumina). After genotyping, systematic quality control analyses were conducted using PLINK 1.90 software (http://www.cog-genomics.org/plink2) (Purcell *et al.*, 2007; Chang *et al.*, 2015). First, 118 samples (84 patients with Alzheimer's disease and 34 controls) were omitted because of sample duplicates or cryptic relatedness (PI_HAT $>$ 0.1875, which is the identity-by-descent expected between third- and second-degree relatives) (Ellingson and Fardo, 2016), or low individual call rate ($<$0.95). The remaining samples were assessed for population outliers and stratification in principal component analysis using EIGENSTRAT (Patterson *et al.*, 2006). All non-autosomal variants were excluded from statistical analyses, as well as SNPs with a call rate $<$98%, minor allele frequency $<$0.01, and/or significant deviation from Hardy-Weinberg equilibrium in controls ($P < 1.0 \times 10^{-4}$) (Supplementary Table 2). Following quality control processing, the genotypes of 765 144 SNPs in 4069 Chinese individuals (1595 patients with Alzheimer's disease and 2474 controls) were further analysed.

Phasing and imputation were performed by SHAPEIT (Delaneau *et al.*, 2011) and IMPUTE2 (Howie *et al.*, 2009), respectively, and version 3 of the 1000 Genomes Project data was used as the reference set (Genomes Project *et al.*, 2012). Variants with r^2 values $<$ 0.80 or impute information measures $<$ 0.50 from IMPUTE2, missing frequency $>$ 0.02, deviation from Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-4}$), and minor allele frequency $<$ 0.01 were excluded from post-imputation quality control analysis. Logistic regression analysis of GWAS data was conducted before and after imputation to test the differences in allele dosage between cases with Alzheimer's disease and controls under an additive genetic model, adjusted for sex, *APOE* status, age (defined as age-at-onset for cases and age-at-last exam for controls), and population substructure using the first two principal components with PLINK 1.90 software. Manhattan and quantile-quantile plots of the first stage before and after imputation and adjustments for sex and *APOE* status were generated using the R *qqman* package (Version 3.4.2, https://www.r-project.org/). Regional association plots were generated via LocusZoom (http://locuszoom.sph.umich.edu/locuszoom/) (Pruim *et al.*, 2010). Linkage disequilibrium plots of variants in chromosome 19 were generated using Haploview software (https://www.broadinstitute.org/haploview/haploview). Conditional analysis was performed to assess the independence of the novel associations of the genotyped SNPs. In addition, stratified analysis was performed by gender and disease status.

Power calculations with Quanto software were applied to calculate the power of the results from the discovery stage (Gauderman *et al.*, 2006). Alzheimer's disease prevalence was set to 3.21% in accordance with epidemiological studies of Alzheimer's disease in Chinese subjects aged $\geqslant$65 years (Jia

et al., 2014). Parameters included outcome (disease), design (unmatched case-control ratio of 1:1.5), hypothesis (gene only), sample size ($n = 1679$ cases), significance (1.0×10^{-5}, two-sided), mode of inheritance (log-additive), and population risk (0.0321).

Second stage

To replicate the first stage association results, the top 34 variants showing an association with a $P < 1.0 \times 10^{-5}$ after adjusting for age, sex, *APOE* status and the first two principal components were selected and analysed as part of an independent cohort of 7319 Chinese individuals consisting of 2234 cases with Alzheimer's disease and 5085 controls (Table 1 and Supplementary Table 1). These 34 SNPs were genotyped at BioMiao Biological Technology Beijing Co. using the MassArray System (Agena iPLEXassay).

Combined analysis of the first and second stages

To improve statistical power, a meta-analysis was applied to combine the associated results from the first two stages using METAL (Willer *et al.*, 2010) with an inverse variance-based model. Heterogeneity tests between the two groups were performed using the Breslow-Day test (Higgins and Thompson, 2002), and the extent of heterogeneity was assessed using the I^2 and *P*-values of the Q statistics calculated by METAL (Higgins *et al.*, 2003).

Single nucleotide polymorphism annotation

SNPnexus was used for SNP annotation (https://www.snpnexus.org/v4/) (Chelala *et al.*, 2009; Dayem Ullah *et al.*, 2012, 2013, 2018). For co-localization of SNPs with significant associations in both stages of the study, we conducted expression quantitative trait loci (eQTL) analysis using the dataset presented by Ramasamy *et al.* (2014), COLOC analysis (http://coloc.cs.ucl.ac.uk) (Giambartolomei *et al.*, 2014) using the brain-eQTL datasets (Trabzuni *et al.*, 2011; Ramasamy *et al.*, 2013), and summary Mendelian randomization-Heidi analysis (https://cnsgenomics.com/software/smr/) (Zhu *et al.*, 2016) using summary eQTL data from the brain and blood (Westra *et al.*, 2013; Lloyd-Jones *et al.*, 2017; Qi *et al.*, 2018). The expression of novel Alzheimer's disease-associated genes was analysed using data from the National Center for Biotechnology Information Gene Expression Omnibus dataset (http://www.ncbi.nlm.nih.gov/geo). These included the expression of genes in the frontal cortex, hippocampus, and temporal cortex of controls and patients with Alzheimer's disease (Supplementary Table 3). Prism software (version 8.0.0, GraphPad Software, Inc., CA, USA) was used to compare gene expression between cognitively normal and Alzheimer's disease groups (unpaired *t*-test and Welch's *t*-test) and to generate figures. For the Alzheimer's disease-associated genes in this study, STRING analysis was performed to evaluate protein-protein interactions (Szklarczyk *et al.*, 2015). Medium confidence (0.400) was used as the minimum required interaction score and no more than 50 interactors were shown in the first shell. The exported network was analysed using the bioinformatics software platform Cytoscape (Version: 3.7.1, https://cytoscape.org/). In addition, we exported the Gene Ontology information, including molecular function, biological process, and cellular component, as well as the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for the network, from STRING analysis.

APOE genotyping

The *APOE* genotypes for haplotypes derived from rs7412 and rs429358 in samples from both stages of the study were determined using the Sanger sequencing method (Sanger *et al.*, 1977).

Validation of gene associations in Caucasian populations

To determine whether there is underlying heterogeneity in the contributors of genetic risk between Chinese and Caucasian populations, the eligible novel SNPs were examined in the data from 'The International Genomics of Alzheimer's Project summary statistics from stage 1 data' (Lambert *et al.*, 2013*b*). The genetic correlation between the Chinese GWAS and publicly available International Genomics of Alzheimer's Project (IGAP) summary statistics was estimated using linkage disequilibrium score regression implemented in the online software LD Hub (http://ldsc.broadinstitute.org/) (Zheng *et al.*, 2017).

Predictive model study

We performed predictive modelling using the polygenic risk score based on SNP significance in combined analysis and *APOE* status as predictor variables, based on the data of the first stage. The individual polygenic risk scores were generated as sums of the risk variants weighted by effect sizes derived from logistic regression. We also ran the predictive analyses on second-stage data using the same factors. Furthermore, we tested different predictive models with different combinations of SNPs in a population negative for *APOE* ε4. Areas under the receiver operating characteristic curve were calculated by comparing the observed case/control status and polygenic risk score calculated using PRSice2 (Choi and O'Reilly, 2019) profiling in a standard weighted allele-dose manner.

To confirm the capacity of the models to predict Alzheimer's disease, we applied the models to individuals who were recruited in a longitudinal study from 2009 to 2019. To estimate the effectiveness of the GWAS-based predictive models, by measuring the fraction of individuals living without Alzheimer's disease for a certain amount of time from baseline, survival curve analyses were performed using the follow-up data in this longitudinal cohort.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

Results

Demographics of three cohorts

In the GWAS, a total of 11 506 individuals participated in this two-stage study, including 3913 patients with Alzheimer's disease and 7593 controls (Supplementary

Table 1 Association of 13 SNPs with Alzheimer's disease in the first and second stages

Chr.	SNP	Nearest gene	Position	Minor allele	First stage (1679 cases, 2508 controls)				Second stage (2234 cases, 5085 controls)				P_{com}	OR_{com} (95% CI)	P_{het}
					MAF_CA, %	MAF_CL, %	OR	P_{gwas}	MAF_CA, %	MAF_CL, %	OR	P_{re}			
5	rs3777215	*RHOBTB3, GLRX*	95122000	A	12.9	17.1	0.69	9.34×10^{-8}	15.5	20.9	0.69	1.31×10^{-13}	3.07×10^{-19}	0.69 (0.64–0.74)	0.72
5	rs6859823	*CTC-278L1.1*	105554384	T	29.2	35.8	0.74	3.28×10^{-9}	27.1	33.9	0.72	5.32×10^{-15}	2.49×10^{-23}	0.74 (0.69–0.78)	0.79
14	rs234434	*CTD-2506J14.1*	97821020	G	36.6	25.8	1.42	3.70×10^{-11}	37.2	25.4	1.75	1.07×10^{-44}	1.35×10^{-67}	1.71 (1.61–1.82)	0.87
19	rs11668861	*NECTIN2*	45380970	G	33.1	22.7	1.76	1.75×10^{-10}	34.9	22.4	1.86	2.99×10^{-52}	4.13×10^{-63}	1.81 (1.69–1.94)	0.32
19	rs6859	*NECTIN2*	45382034	A	40.9	31.6	1.55	2.92×10^{-8}	43.7	32.6	1.61	9.67×10^{-35}	3.70×10^{-42}	1.58 (1.48–1.69)	0.48
19	rs3852860	*NECTIN2*	45382966	C	33.5	23.6	1.72	1.87×10^{-9}	35.6	24.1	1.75	3.07×10^{-44}	4.20×10^{-54}	1.72 (1.60–1.84)	0.48
19	rs71352238	*TOMM40*	45394336	C	23.3	9.6	3.13	5.64×10^{-15}	21.8	8.7	2.91	2.40×10^{-100}	1.16×10^{-131}	2.90 (2.66–3.17)	0.69
19	rs157580	*TOMM40*	45395266	A	53.9	45.4	1.48	1.10×10^{-6}	54.9	46.6	1.40	1.60×10^{-19}	2.99×10^{-25}	1.40 (1.31–1.49)	0.86
19	rs2075650	*TOMM40*	45395619	G	23.3	9.4	3.27	3.05×10^{-18}	25.2	9.4	3.25	5.30×10^{-132}	2.57×10^{-164}	3.17 (2.91–3.45)	0.36
19	rs157582	*TOMM40*	45396219	T	32.0	18.9	2.33	2.53×10^{-10}	33.7	18.9	2.19	6.57×10^{-76}	2.97×10^{-95}	2.15 (2.00–2.31)	0.62
19	rs439401	*APOE*	45414451	C	50.9	43.3	1.50	2.06×10^{-6}	52.8	42.1	1.54	2.08×10^{-31}	1.17×10^{-35}	1.50 (1.40–1.59)	0.14
19	rs4420638	*APOC1*	45422946	G	27.2	12.2	2.98	2.62×10^{-17}	28.5	11.0	3.20	3.40×10^{-140}	8.32×10^{-171}	3.07 (2.83–3.32)	0.13
21	rs2255835	*CHODL*	19491664	C	33.6	27.7	1.33	8.61×10^{-6}	26.2	23.0	1.19	6.78×10^{-05}	4.81×10^{-9}	1.23 (1.16–1.31)	0.51

Chr. = chromosome; MAF_CA = minor allele frequency of cases; MAF_CL = minor allele frequency of controls; Minor allele = minor allele of the controls; OR = odds ratios calculated according to the minor allele; OR_{com} = the odds ratio of the combined analysis; P_{com} = the *P*-value of the combined stage using meta-analysis; P_{gwas}, the *P*-value of the first stage adjusted for sex, *APOE* status, age (defined as age at onset for cases and age at last exam for controls) and two ancestry principal components; P_{het} = the *P*-value of heterogeneity; Position = base pair position according to hg 19; P_{re} = the *P*-value of the second stage.

Table 1). In the first stage, the mean age at onset of 1679 patients with Alzheimer's disease was 72.07 ± 6.4 years, of which 949 (56.5%) were female, whereas the mean age at examination of 2508 controls was 72.95 ± 19.5 years, of which 1398 (55.7%) were female. In the second stage, the mean age at onset of 2234 patients with Alzheimer's disease was 72.35 ± 7.2 years of which 1253 (56.1%) were female, whereas the mean age at examination of 5085 controls was 69.22 ± 7.0 years, of which 2931 (57.6%) were female. In the longitudinal cohort, 5474 individuals were recruited, comprising 2358 patients with Alzheimer's disease and 3116 controls (Supplementary Table 1). The mean age at onset of 2358 patients with Alzheimer's disease was 71.70 ± 6.8 years, of which 1351 (57.3%) were female, whereas the mean age at examination of 3116 controls was 74.70 ± 7.5 years, of which 1676 (53.8%) were female. The flow chart of the current study is shown in Fig. 1.

Results from genome-wide association studies

In the first stage, 1679 patients with Alzheimer's disease and 2508 controls were genotyped (Supplementary Table 1). Overall, 765 144 autosomal SNPs passed the quality control standards and were included for further analysis. Principal component analysis confirmed that the patients with Alzheimer's disease and controls were well-matched (Supplementary Fig. 1). A quantile-quantile plot indicated that population stratification had negligible effects on the statistical results (λ_{GC} = 1.08; Supplementary Fig. 2). After adjusting for age, sex, and *APOE* status along with the first two principal components, several markers on various chromosomes exhibited genome-wide significance, with the sentinel markers occurring on chromosome 19 (Supplementary Fig. 3). Power calculations indicated that the sample size used in this GWAS provided sufficient statistical power to detect Alzheimer's disease-associated variants. In the second stage, the top 34 SNPs with evidence of associations with Alzheimer's disease (Supplementary Table 4) were selected for genotyping as part of an independent Chinese cohort of 2234 patients with Alzheimer's disease and 5085 controls (Supplementary Table 1). Of these 34 Alzheimer's disease-associated SNPs, 13 surpassed the Bonferroni correction threshold ($P < 1.47 \times 10^{-3}$) with no detectable heterogeneity between stages. This included nine SNPs on chromosome 19 in the *APOE* region (Jun *et al.*, 2012) (*APOE* rs439401, *APOC1* rs4420638, *TOMM40* rs2075650,

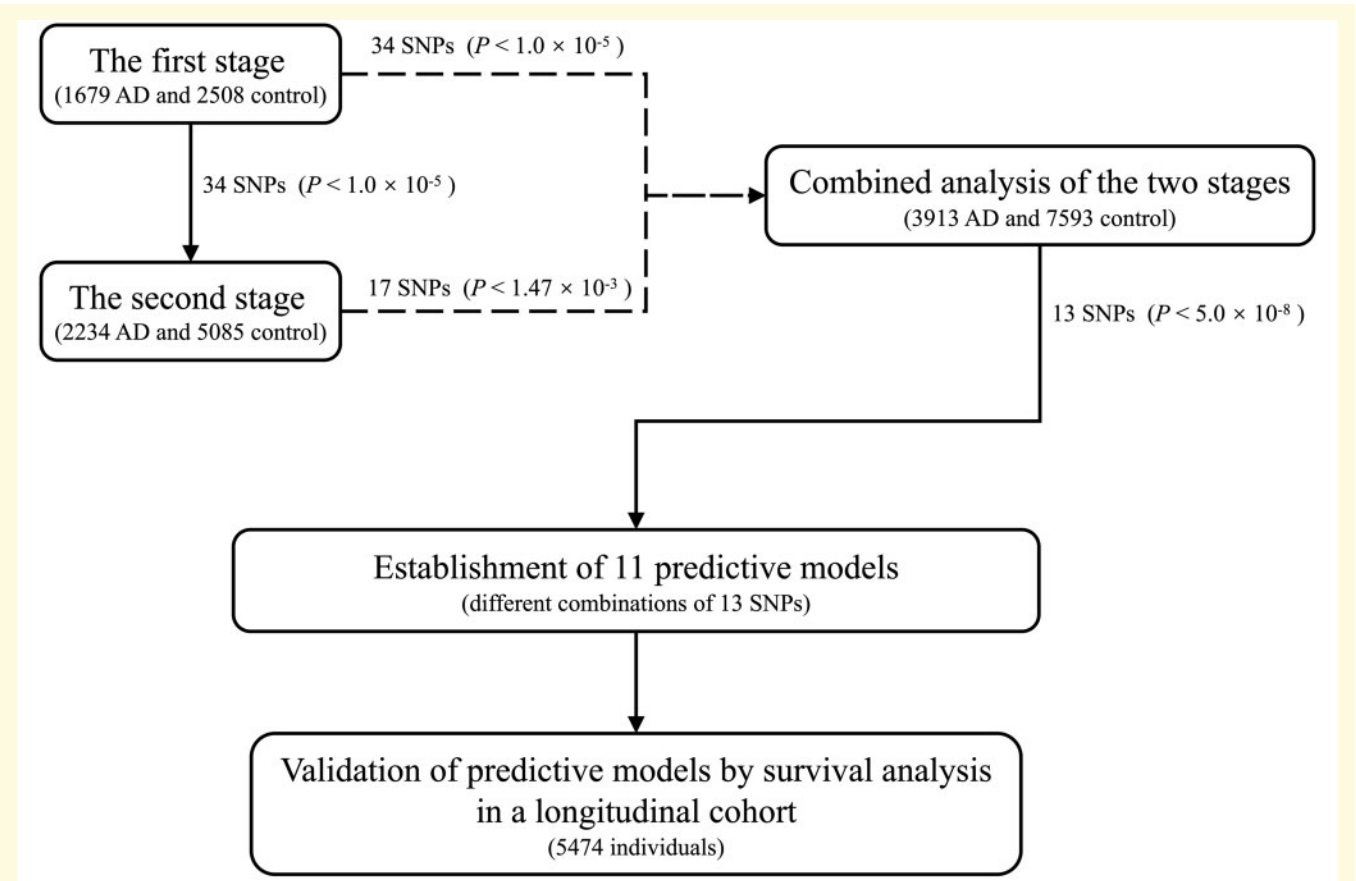

Figure 1 Study flow chart. AD = Alzheimer's disease.

TOMM40 rs71352238, *TOMM40* rs157582, *TOMM40* rs157580, *NECTIN2* rs11668861, *NECTIN2* rs3852860, and *NECTIN2* rs6859; Table 1). The other four SNPs were not located on chromosome 19 (rs3777215, rs6859823, rs234434, and rs2255835; $P = 1.31 \times 10^{-13}$, 5.32×10^{-15}, 1.07×10^{-44}, and 6.78×10^{-05}, respectively). In combined analysis, all 13 of these SNPs showed associations exceeding the genome-wide significance threshold ($P < 5.0 \times 10^{-8}$; Table 1). Furthermore, all four novel variants remained significant after adjusting for the effect of *APOE ε4*. Stratified analyses indicated that the 34 top SNPs in the first stage remained significant after stratification by gender (Supplementary Table 5). Conditional analyses indicated that the four novel SNPs we identified were independent signals (Supplementary Table 6).

Functional annotation of the four novel single nucleotide polymorphisms

Variant rs3777215 was significantly associated with Alzheimer's disease [$P_{combined} = 3.07 \times 10^{-19}$, odds ratio (OR) = 0.69] and was located on chromosome 5 in the intron regions of *RHOBTB3* and *GLRX* (Fig. 2A, Table 1 and Supplementary Table 7). As shown in Fig. 3, *RHOBTB3* was significantly upregulated in the frontal cortex, hippocampus, and temporal cortex of patients with Alzheimer's disease compared to cognitively normal individuals ($P < 0.05$, 0.01 or 0.0001; Fig. 3A). *GLRX* showed lower expression levels in the brain tissues of patients with Alzheimer's disease than in that of cognitively normal individuals ($P < 0.05$, 0.01 or 0.0001; Fig. 3B). Another SNP detected on chromosome 5 was rs6859823 ($P_{combined} = 2.49 \times 10^{-23}$, OR = 0.74; Fig. 2B and Table 1). Variant rs6859823 was intergenic and located between *RNA5SP189* and *CTC-278L1.1* (Supplementary Table 7). *RNA5SP189* and *CTC-278L1.1* were both identified as pseudogenes according to GeneCards and SNPnexus, and neither have been previously reported to be associated with Alzheimer's disease or any other disease. SNP rs234434 ($P_{combined} = 1.35 \times 10^{-67}$, OR = 1.71; Fig. 2C and Table 1) was intergenic and was between two long intergenic non-coding RNAs known as *RP11-359N5.1* and *CTD-2506J14.1*. SNP rs2255835 was located on chromosome 21 in the intron region of *CHODL* and showed genome-wide significance ($P_{combined} = 4.81 \times 10^{-9}$, OR = 1.23; Fig. 2D and Table 1). *CHODL* expression was higher in the hippocampus ($P < 0.05$; Fig. 3C) and lower in the temporal cortex of patients with Alzheimer's disease compared to that in normal individuals ($P < 0.05$; Fig. 3C). Co-localization analyses indicated that the four SNPs were related to various genes being expressed in the blood and different brain regions (Supplementary Tables 8–11). Roadmap epigenomics showed that rs3777215 and rs2255835 were related to transcriptional activation (H3K36me3, H3K4me1, and H3K14ac) in neurons or neuronal progenitor cells, whereas rs234434 was associated with transcriptional repression (H3K27me3) in neurons (Supplementary Table 7). STRING analysis demonstrated that the proteins encoded by the variant genes, except for *CHODL*, interact with APOE (Supplementary Fig. 4). Gene Ontology enrichments of the genes in the STRING network suggested that these genes are involved in several biological processes. For example, *RHOBTB3* was suggested to be involved in the 'establishment of localization' and 'transport' (Supplementary Table 12). Genes enriched in Gene Ontology cellular components and molecular functions, as well as KEGG pathways, are listed in Supplementary Tables 13–15.

Validation in Caucasian genome-wide association studies datasets

Overall, the *APOE ε4* allele frequency in Chinese subjects in the present study was lower than that in Caucasians for both, patients with Alzheimer's disease and controls (Supplementary Table 1). Despite these differences, nine SNPs with significant associations in the *APOE* region in the present study were either reported previously or found in strong linkage disequilibrium with nearby SNPs in Caucasians (Supplementary Fig. 5). One of the four novel SNPs outside chromosome 19 reached genome-wide significance in the IGAP stage 1 data for non-Asian populations (rs6859823; Supplementary Table 16). Moreover, linkage disequilibrium score regression analysis of Chinese GWAS and publicly available IGAP summary statistics revealed a genetic correlation of –0.14 ($P = 0.73$).

Predictive models

We tested 11 predictive models; four models were used to analyse all populations (Fig. 4A–D) and seven models were used to analyse subjects who were negative for *APOE* ε4 (Fig. 4E–K) in this study. The number of SNPs in our models was relatively low, and even for the maximum, the number of SNPs for model A3 was only 13 with an AUC of 0.73 [95% confidence interval (CI): 0.70–0.75] in the first stage. The AUCs of all 11 models were significant ($P < 0.05$; range of AUC: 0.63–0.73), and the specific AUC values for the 11 models are presented in Fig. 4. For model A1, with the four novel SNPs found in this study and *APOE* ε4 status as predictors, training on the first-stage data, prediction accuracy AUC = 0.69 (95% CI: 0.67–0.71) was achieved based on a logistic regression model. However, the prediction accuracy AUC reached 0.73 (95% CI: 0.71–0.74) when using second-stage data (Fig. 4A). Model B1 (with our four novel SNPs found in this study as predictors) and model B4 (with two novel SNPs and two SNPs in the *APOE* region in this study as predictors) covered more individuals who were negative for *APOE* ε4 in our study, and the number of model B2 was maximum in B models with four novel SNPs and three *APOE*-region SNPs. For model B1, the prediction accuracy AUC values were 0.63 (95% CI: 0.61–0.66) and 0.66 (95% CI: 0.63–0.69)

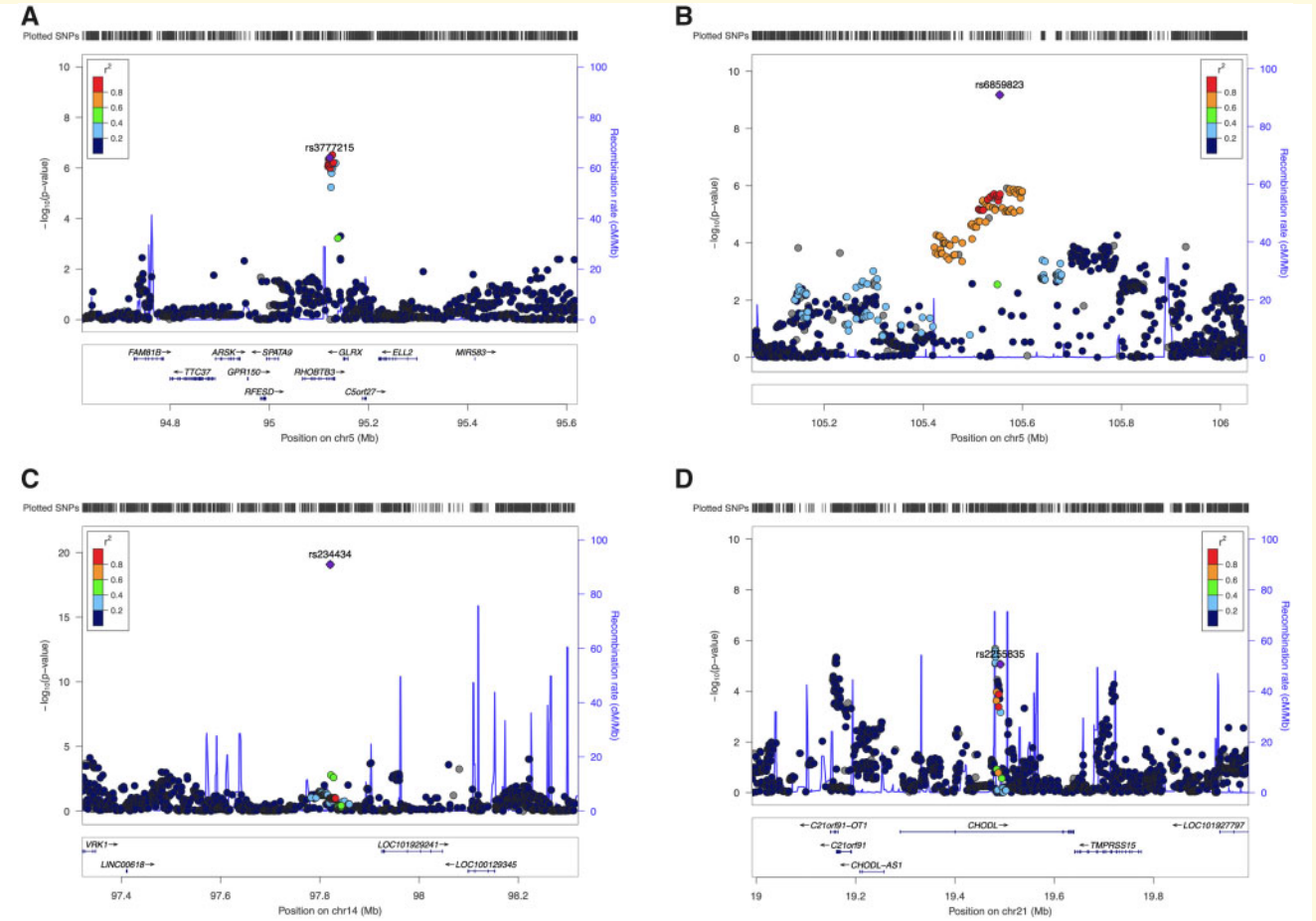

Figure 2 Regional association plots. (**A**–**D**) Association results are shown for the analysed SNPs with recombination rates in the four loci associated with genome-wide significance at chromosome 5 (**A** and **B**), 14 (**C**), and 21 (**D**). The –log10 (*P*-values) (*y*-axis) of SNPs within the ±500 kb region centred on each marker SNP are presented according to the chromosomal positions of the SNPs (*x*-axis; NCBI Build 37). Purple diamonds represent the most significantly associated SNP (marker SNP) in the combined analysis. SNPs are coloured according to their linkage disequilibrium with the marker SNP. Linkage disequilibrium values were based on the 1000 Genome Project Asian data. Blue lines represent the estimated recombination rates based on the 1000 Genome Project samples. Arrows depict genes in the regions of interest annotated from the UCSC Genome Browser.

based on the first and second data, respectively (Fig. 4E); whereas, they were 0.71 (95% CI: 0.68–0.73) and 0.68 (95% CI: 0.65–0.70) for model B4 (Fig. 4H), and 0.72 (95% CI: 0.69–0.74) and 0.68 (95% CI: 0.65–0.70) for model B2, based on the first and second data, respectively (Fig. 4E).

Longitudinal study

In the longitudinal cohort, all 13 SNPs identified in the combined analysis reached significance ($P < 3.85 \times 10^{-3}$; Supplementary Table 17). The AUC values of the 11 predictive models were also significant using the data from the longitudinal cohort ($P < 0.05$; range of AUC: 0.67–0.73; Fig. 4). To confirm the capacity of the models to predict Alzheimer's disease, we performed survival curve analysis. First, the predictive models were applied to individuals (2358 Alzheimer's disease, 3116 controls, 5474 in total) in the longitudinal study. Models A1–A4 were generated based on an *APOE* ε4-positive or *APOE* ε4-negative population, and models B1–B7 were based on an *APOE* ε4-negative population only. Because the same populations were included in models A1–A4, and models B1–B7, we combined them into model A and model B for survival analysis. Individuals were divided into three groups as follows: model A positive (model A+), model B positive (model B+), and model A or B negative (model A– or B–). In total, the number of subjects in model A+ was 213, in model B+ was 1035, and in model A– or B– was 4226. Data from individuals from each group were plotted into the survival curve according to the follow-up data comprising the incidence and onset of Alzheimer's disease from 2009 to 2019. Kaplan-Meier survival curve analysis revealed significant differences between the model A+, model B+, and model A– or B– groups ($P < 0.001$; Fig. 4L), indicating that individuals with model A+ and model B+ had a shorter latency and a higher proportion of

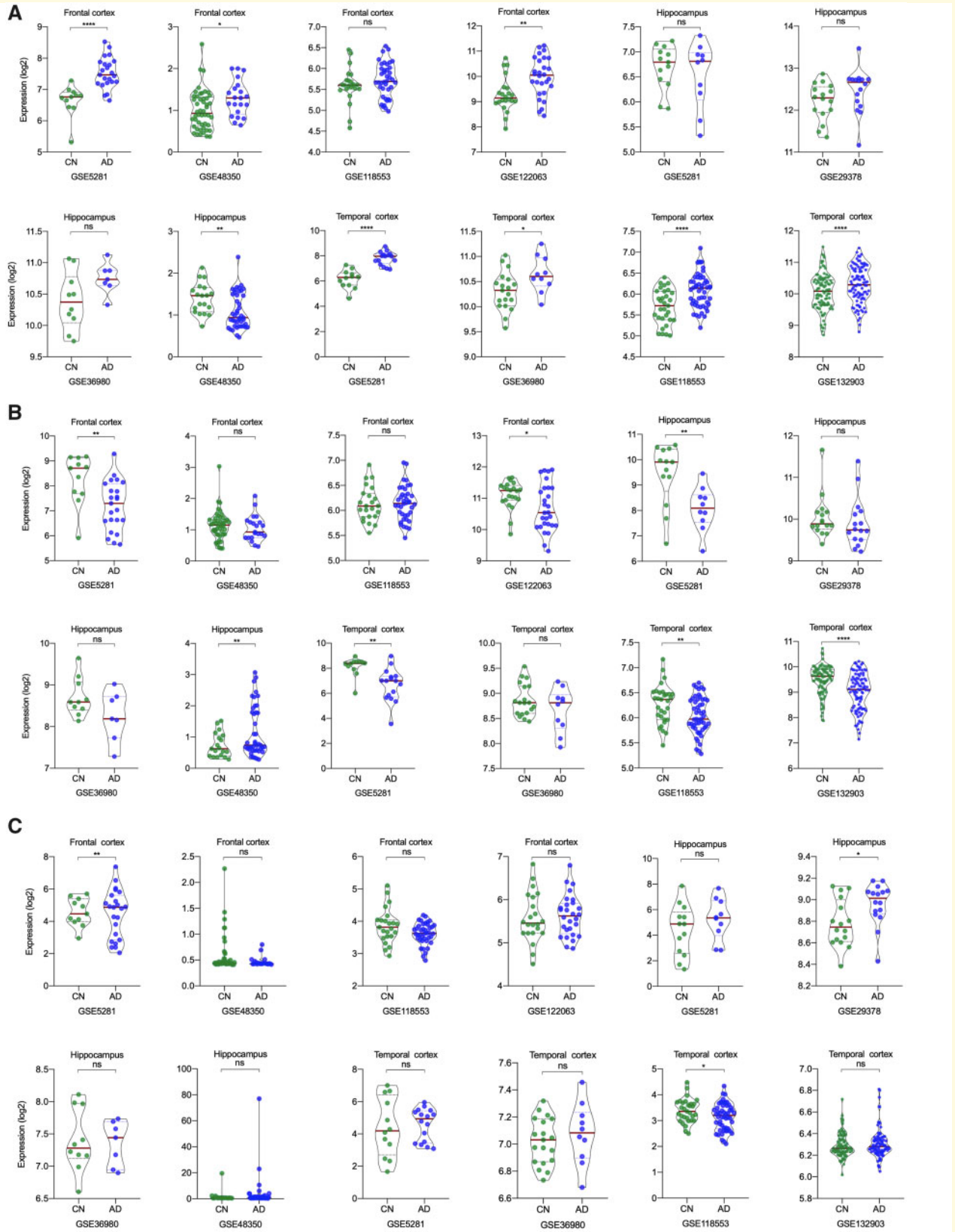

Figure 3 Differential expression of the annotated genes in Gene Expression Omnibus datasets. (**A**–**C**) shows the differential expression of RHOBTB3 (**A**), GLRX (**B**), CHODL (**C**) in frontal cortex, hippocampus, and temporal cortex. The bold red line indicates the median of each group, and the black dotted lines show the quartiles. AD = Alzheimer's disease; CN = cognitively normal; FC = fold change; GSE = Gene Expression Omnibus Series; ns = no significance; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

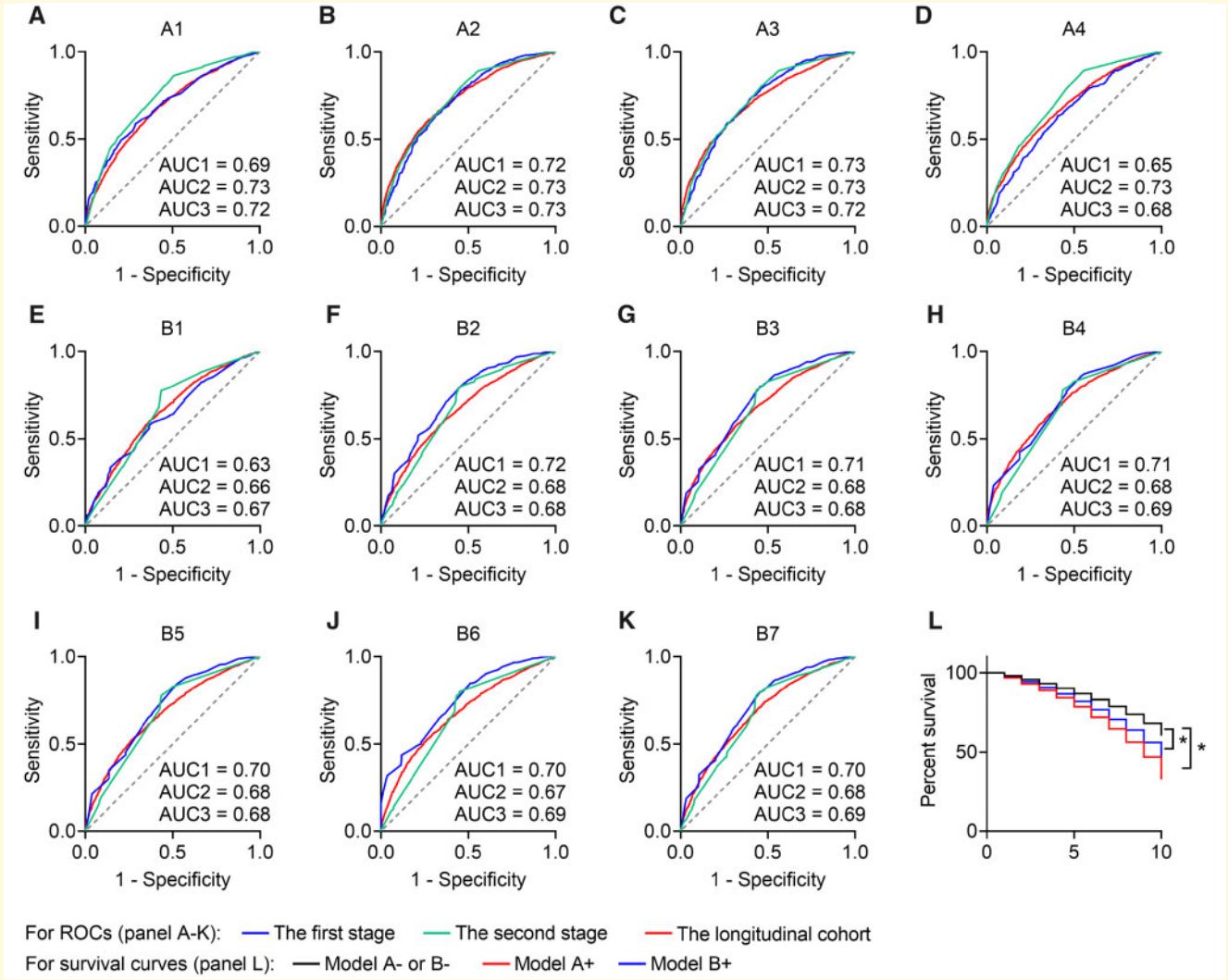

Figure 4 ROC curves for 11 predictive models with different predictors in the three cohorts and survival curves in a longitudinal cohort. The factors included in the 11 models are as follows. (**A**) A1: *APOE* ε4 status, rs3777215, rs6859823, rs234434 and rs2255835; (**B**) A2: *APOE* ε4 status, rs3777215, rs6859823, rs234434, rs2255835, rs11668861, rs71352238 and rs4420638; (**C**) A3: *APOE* ε4 status, rs3777215, rs6859823, rs234434, rs2255835, rs11668861, rs6859, rs3852860, rs71352238, rs157580, rs2075650, rs157582, rs439401 and rs4420638; (**D**) A4: rs3777215, rs6859823, rs234434, rs2255835, rs11668861, rs6859, rs3852860, rs71352238, rs157580, rs2075650, rs157582, rs439401 and rs4420638; (**E**) B1: rs3777215, rs6859823, rs234434 and rs2255835; (**F**) B2: rs3777215, rs6859823, rs234434, rs2255835, rs11668861, rs71352238 and rs4420638; (**G**) B3: rs3777215, rs6859823, rs234434, rs71352238 and rs4420638; (**H**) B4: rs3777215, rs234434, rs71352238 and rs4420638; (**I**) B5: rs6859823, rs234434, rs71352238 and rs4420638; (**J**) B6: rs3777215, rs6859823, rs234434 and rs71352238; (**K**) B7: rs3777215, rs6859823, rs234434 and rs4420638. (**L**) Survival curves of the longitudinal cohort, $^*P < 0.001$. AUC1 indicates AUC of the first stage; AUC2 indicates AUC of the second stage; AUC3 indicates AUC of the longitudinal cohort; ROC = receiver operating characteristic curve.

Alzheimer's disease during the follow-up period. Our data suggest that the 11 predictive models have sufficient capacity for predicting Alzheimer's disease risk.

Discussion

We found four novel variants in addition to nine *APOE*-region variants that were correlated with Alzheimer's disease risk in a Chinese population. Using different combinations of these variants and the *APOE* status, we established 11 predictive models with significant AUC values. Validation of these models in a longitudinal cohort indicated the genetic power of SNPs for Alzheimer's disease prediction. Overall, these findings may improve the understanding of how genetic variants impact the initiation of Alzheimer's disease.

The novel variants identified are functionally involved in the pathogenesis of Alzheimer's disease. One of the annotated genes for the novel SNP rs3777215 was *RHOBTB3*.

Reported as a candidate Alzheimer's disease vulnerability gene through transcription analysis, this gene was found to be overexpressed in the CA1 region following Alzheimer's disease progression (Miller *et al.*, 2013). *RHOBTB3* encodes Rho-related BTB domain-containing protein 3 (RHOBTB3), which is involved in endosome-to-Golgi transport and retrograde transport (Espinosa *et al.*, 2009). Thus, RHOBTB3 may affect APP processing and provide a pathological basis for the development of Alzheimer's disease. The other annotated gene for the novel SNP rs3777215 was *GLRX*, which encodes glutaredoxin-1 (GRX1), an oxidoreductase that contributes greatly to the antioxidant defence system and internal environment homeostasis. Functionally, GRX1 and thioredoxin-1 (TRX1) are antioxidants and their reduced forms can inhibit apoptosis signal regulating kinase (ASK1). It has been reported that amyloid-β can oxidize GRX1 and TRX1, resulting in apoptosis induction via ASK1 (Akterin *et al.*, 2006). Another study found that increasing GRX1 levels in the brain of an Alzheimer's disease mouse model can reverse synaptic dysfunction and cognitive deficits, suggesting GRX1 as a target for Alzheimer's disease intervention (Kommaddi *et al.*, 2019). For the novel SNP rs2255835, the annotated gene was *CHODL*. This gene encodes chondrolectin (CHODL), which is involved in the endocytosis of glycoproteins and exogenous sugar-bearing pathogens (Zelensky and Gready, 2005). CHODL affects cell survival and neuronal outgrowth in animal models (Sleigh *et al.*, 2014). Interestingly, in some early-onset patients with Alzheimer's disease induced by *APP* duplication, the duplicated region also contains *CHODL*, as well as eQTL-associated genes of rs2255835, such as *BTG3*, *C21orf91*, and *TMPRSS15*, which may participate in neurogenesis and/or APP metabolism (McNaughton *et al.*, 2012; Wiseman *et al.*, 2015).

Using different combinations of variants identified in the current study, we established 11 predictive models. Building predictive models based on GWAS data to distinguish asymptomatic population at a high-risk of developing Alzheimer's disease from 'normal' individuals has gained attention (Escott-Price *et al.*, 2015, 2017*a*, *b*, 2019; Chouraki *et al.*, 2016; Stocker *et al.*, 2018), and some recent studies have focused on predicting the conversion from mild cognitive impairment to Alzheimer's disease via models based on genetic factors (Lacour *et al.*, 2017; Chaudhury *et al.*, 2019). Escott-Price and colleagues demonstrated a prediction accuracy of 75–84% for Alzheimer's disease risk with certain predictors (*APOE*, polygenic risk score calculated from more than 20 000 SNPs, sex, and age) (Escott-Price *et al.*, 2015). In addition, Sultan and colleagues reported that the prediction accuracy of their models, with *APOE* SNPs (rs7412 and rs429358), 165 non-*APOE* SNPs, sex, and age as predictors was 82.5% for predicting the conversion from mild cognitive impairment to Alzheimer's disease (Chaudhury *et al.*, 2019). However, most of these models included too many SNPs, preventing their use in clinics. Therefore, we established predictive models to determine the risks of the possibility of developing Alzheimer's disease. In our A models, models A1 and A2 may be easier to use in clinics, as the number of SNPs in model A1 or model A2 was smaller than that in model A3, and the AUC of model A1 or model A2 was similar to that of model A3. However, these models are no longer suitable for populations negative for *APOE* ε4 comprising a more important group. Thus, we constructed models B1–B7. These models incorporate fewer SNPs and show significant AUC values. In the B models, the AUC of model B4 was similar to that of model B2, but the number of SNPs in model B4 was approximately half of that of model B2. Therefore, models B4–B7 are recommended as more of the population can be covered by these models in clinical practice. Overall, the 11 predictive models appear to be useful for identifying the indications of Alzheimer's disease risk in the sectional datasets. To confirm the capacity of the models to predict Alzheimer's disease, we performed survival curve analysis on a longitudinal cohort. The results showed that individuals carrying risk variants included in either model A or model B had a shorter latency and higher incidence of Alzheimer's disease, suggesting that our models can predict Alzheimer's disease onset in a population with genetic susceptibility. The mechanism of this genetic susceptibility requires further analysis.

The current study had some limitations. First, the novel variants we identified are involved in the pathogenesis of Alzheimer's disease based on bioinformatic analysis and literature mining, but we did not conduct functional research on these novel variants. However, validation of the predictive models indicated the contributions of these variants to sporadic Alzheimer's disease development. Second, the prediction accuracies of our models were relatively low compared to those of other predictive models but the models were verified to be effective and accessible for predicting Alzheimer's disease onset based on a 10-year longitudinal cohort and sectional datasets. Third, although the longitudinal study (COAST) we used was prospective, the use of current diagnoses of individuals was compared with the initial condition at baseline and was used to validate our predictive model and to confirm the accuracy and effectiveness of our models to predict Alzheimer's disease. This is rational because the DNA of individuals at baseline may reflect the true genetic conditions. Fourth, the significant SNPs in the discovery stage, which were not replicated in the second stage, did not pass the heterogeneity test. This may be because of differences in the genetic background of the Chinese population (Supplementary Table 18), which is supported by other studies (Chen *et al.*, 2009; Tan *et al.*, 2013; Tao *et al.*, 2017; Wang *et al.*, 2018). Finally, despite the power calculation indicating that our sample size was sufficient to detect associations, larger sample sizes based on Chinese populations are required in future studies, as well as analyses of populations of other ethnicities, to generate more reliable results.

In conclusion, we identified four novel susceptibility variants for Alzheimer's disease, improving the understanding of the genetic predisposition to Alzheimer's disease. Annotated genes of these variants were related to APP

metabolism, antioxidation, and neurogenesis. The contributions of these variants to sporadic Alzheimer's disease development were confirmed to be efficient based on validation of the predictive models in a longitudinal study. This is the first study to validate GWAS-based predictive models for evaluating the risk of Alzheimer's disease onset in a Chinese population. The clinical application of these models is of potential use for individuals harbouring these risk variants but must be validated in a larger population.

Acknowledgements

We thank all participants in this study and all neurologists at relevant academic centres for their help in the recruitment of subjects. In particular, we thank the participating hospitals which were not listed in author list including Tianjin Medical University General Hospital; No. 88 Hospital of PLA; The First Affiliated Hospital of Wenzhou Medical University; The First Bethune Hospital of Jilin University; University Hospital of Hubei University for Nationalities; The Second Hospital of Shandong University; The First Affiliated Hospital of Anhui Medical University; Sir Run Run Shaw Hospital of Zhejiang University; Beijing Chaoyang Hospital, Capital Medical University.

Funding

This study was supported by the Key Project of the National Natural Science Foundation of China (81530036); the National Key Scientific Instrument and Equipment Development Project (31627803); Mission Program of Beijing Municipal Administration of Hospitals (SML20150801); Beijing Scholars Program; Beijing Brain Initiative from Beijing Municipal Science & Technology Commission (Z201100005520016, Z201100005520017, Z161100000216137); Project for Outstanding Doctor with Combined Ability of Western and Chinese Medicine; and Beijing Municipal Commission of Health and Family Planning (PXM2019_026283_000003).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

Akterin S, Cowburn RF, Miranda-Vizuete A, Jimenez A, Bogdanovic N, Winblad B, et al. Involvement of glutaredoxin-1 and thioredoxin-1 in beta-amyloid toxicity and Alzheimer's disease. Cell Death Differ 2006; 13: 1454–65.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci 2015; 4: 7.

Chaudhury S, Brookes KJ, Patel T, Fallows A, Guetta-Baranes T, Turton JC, et al. Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. Transl Psychiatry 2019; 9: 154.

Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics 2009; 25: 655–61.

Chen J, Zheng H, Bei JX, Sun L, Jia WH, Li T, et al. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. Am J Hum Genet 2009; 85: 775–85.

Choi SW, O'Reilly PF. PRSice-2: polygenic Risk Score software for biobank-scale data. Gigascience 2019; 8:1–6. doi: 10.1093/gigascience/giz082.

Chouraki V, Reitz C, Maury F, Bis JC, Bellenguez C, Yu L, et al. Evaluation of a genetic risk score to improve risk prediction for Alzheimer's disease. J Alzheimers Dis 2016; 53: 921–32.

Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res 2012; 40(Web Server issue): W65–70.

Dayem Ullah AZ, Lemoine NR, Chelala C. A practical guide for the functional annotation of genetic variations using SNPnexus. Brief Bioinform 2013; 14: 437–47.

Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res 2018; 46: W109–W13.

Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods 2011; 9: 179–81.

Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, et al. Polygenic overlap between C-reactive protein, plasma lipids, and Alzheimer disease. Circulation 2015; 131: 2061–9.

Ellingson SR, Fardo DW. Automated quality control for genome wide association studies. F1000Res 2016; 5: 1889.

Escott-Price V, Myers A, Huentelman M, Shoai M, Hardy J. Polygenic risk score analysis of Alzheimer's disease in cases without APOE4 or APOE2 alleles. J Prev Alzheimers Dis 2019; 6: 16–9.

Escott-Price V, Myers AJ, Huentelman M, Hardy J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Ann Neurol 2017a; 82: 311–4.

Escott-Price V, Shoai M, Pither R, Williams J, Hardy J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. Neurobiol Aging 2017b; 49: 214 e7–e11.

Escott-Price V, Sims R, Bannister C, Harold D, Vronskaya M, Majounie E, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. Brain 2015; 138: 3673–84.

Espinosa EJ, Calero M, Sridevi K, Pfeffer SR. RhoBTB3: a Rho GTPase-family ATPase required for endosome to Golgi transport. Cell 2009; 137: 938–48.

Gatz M, Pedersen NL, Berg S, Johansson B, Johansson K, Mortimer JA, et al. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. J Gerontol A Biol Sci Med Sci 1997; 52: M117–M25.

Gauderman WJ, Morrison JM, Morrison WGJ. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies [Internet]. 2006. Available from: https://preventivemedicine.usc.edu/download-quanto/

Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491: 56–65.

Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014; 10: e1004383.

Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med 2013; 368: 117–27.

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009; 41: 1088–93.

Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21: 1539–58.

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj 2003; 327: 557–60.

Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 2011; 43: 429–35.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009; 5: e1000529.

Jia J, Wang F, Wei C, Zhou A, Jia X, Li F, et al. The prevalence of dementia in urban and rural areas of China. Alzheimers Dement 2014; 10: 1–9.

Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert JC, Chung J, Naj AC, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. Mol Psychiatry 2016; 21: 108–17.

Jun G, Vardarajan BN, Buros J, Yu CE, Hawk MV, Dombroski BA, et al. Comprehensive search for Alzheimer disease susceptibility loci in the APOE region. Arch Neurol 2012; 69: 1270–9.

Kommaddi RP, Tomar DS, Karunakaran S, Bapat D, Nanguneri S, Ray A, et al. Glutaredoxin1 diminishes amyloid beta-mediated oxidation of F-actin and reverses cognitive deficits in an Alzheimer's disease mouse model. Antioxid Redox Signal 2019; 31: 1321–38.

Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat Genet 2019; 51: 414–30.

Lacour A, Espinosa A, Louwersheimer E, Heilmann S, Hernandez I, Wolfsgruber S, et al. Genome-wide significant risk factors for Alzheimer's disease: role in progression to dementia due to Alzheimer's disease among subjects with mild cognitive impairment. Mol Psychiatry 2017; 22: 153–60.

Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 2009; 41: 1094–9.

Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013a; 45: 1452–8.

Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013b; 45: 1452–8.

Leonenko G, Sims R, Shoai M, Frizzati A, Bossu P, Spalletta G, et al. Polygenic risk and hazard scores for Alzheimer's disease prediction. Ann Clin Transl Neurol 2019; 6: 456–65.

Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, et al. The genetic architecture of gene expression in peripheral blood. Am J Hum Genet 2017; 100: 228–37.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34: 939–44.

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Jr., Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 263–9.

McNaughton D, Knight W, Guerreiro R, Ryan N, Lowe J, Poulter M, et al. Duplication of amyloid precursor protein (APP), but not prion protein (PRNP) gene is a significant cause of early onset dementia in a large UK series. Neurobiol Aging 2012; 33: 426 e13.

Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. Genome Med 2013; 5: 48.

Miron J, Picard C, Nilsson N, Frappier J, Dea D, Theroux L, et al. CDK5RAP2 gene and tau pathophysiology in late-onset sporadic Alzheimer's disease. Alzheimers Dement 2018; 14: 787–96.

Miyashita A, Koike A, Jun G, Wang L-S, Takahashi S, Matsubara E, et al. SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. PLoS One 2013; 8: e58618.

Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 2011; 43: 436–41.

Nasiri H, Forouzandeh M, Rasaee MJ, Rahbarizadeh F. Modified salting-out method: high-yield, high-quality genomic DNA extraction from whole blood using laundry detergent. J Clin Lab Anal 2005; 19: 229–32.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006; 2: e190.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26: 2336–7.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–75.

Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat Commun 2018; 9: 2282.

Ramasamy A, Trabzuni D, Gibbs JR, Dillman A, Hernandez DG, Arepalli S, et al. Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. Nucleic Acids Res 2013; 41: e88.

Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 2014; 17: 1418–28.

Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang L-S, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ4, and the risk of late-onset Alzheimer disease in African Americans. Jama 2013; 309: 1483–92.

Sanchez-Mut JV, Heyn H, Silva BA, Dixsaut L, Garcia-Esparcia P, Vidal E, et al. PM20D1 is a quantitative trait locus associated with Alzheimer's disease. Nat Med 2018; 24: 598–603.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463–7.

Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Mol Med 2016; 8: 595–608.

Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. Jama 2010; 303: 1832–40.

Sleigh JN, Barreiro-Iglesias A, Oliver PL, Biba A, Becker T, Davies KE, et al. Chondrolectin affects cell survival and neuronal outgrowth in in vitro and in vivo models of spinal muscular atrophy. Hum Mol Genet 2014; 23: 855–69.

Stocker H, Mollers T, Perna L, Brenner H. The genetic risk of Alzheimer's disease beyond APOE epsilon4: systematic review of Alzheimer's genetic risk scores. Transl Psychiatry 2018; 8: 166.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43(Database issue): D447–52.

Tan L, Yu JT, Zhang W, Wu ZC, Zhang Q, Liu QY, et al. Association of GWAS-linked loci with late-onset Alzheimer's disease

in a northern Han Chinese population. Alzheimers Dement 2013; 9: 546–53.

Tao QQ, Liu ZJ, Sun YM, Li HL, Yang P, Liu DS, et al. Decreased gene expression of CD2AP in Chinese patients with sporadic Alzheimer's disease. Neurobiol Aging 2017; 56: 212 e5–e10.

Trabzuni D, Ryten M, Walker R, Smith C, Imran S, Ramasamy A, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. J Neurochem 2011; 119: 275–82.

Wang H-Z, Bi R, Hu Q-X, Xiang Q, Zhang C, Zhang D-F, et al. Validating GWAS-identified risk loci for Alzheimer's disease in Han Chinese populations. Mol Neurobiol 2016; 53: 379–90.

Wang Y, Lu D, Chung YJ, Xu S. Genetic structure, divergence and admixture of Han Chinese, Japanese and Korean populations. Hereditas 2018; 155: 19.

Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013; 45: 1238–43.

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010; 26: 2190–1.

Wiseman FK, Al-Janabi T, Hardy J, Karmiloff-Smith A, Nizetic D, Tybulewicz VL, et al. A genetic cause of Alzheimer disease: mechanistic insights from Down syndrome. Nat Rev Neurosci 2015; 16: 564–74.

Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. FEBS J 2005; 272: 6179–217.

Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 2017; 33: 272–9.

Zhou X, Chen Y, Mok KY, Zhao Q, Chen K, Chen Y, et al. Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer's disease pathogenesis. Proc Natl Acad Sci USA 2018; 115: 1697–706.

Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016; 48: 481–7.