Alzheimer's & Dementia 14 (2018) 623-633

Featured Article

Genome-wide association study of Alzheimer's disease endophenotypes at prediagnosis stages

Jaeyoon Chung[a,b], Xulong Wang[c], Toru Maruyama[d,e], Yiyi Ma[b], Xiaoling Zhang[b], Jesse Mez[f], Richard Sherva[b], Haruko Takeyama[d,e,g], The Alzheimer's Disease Neuroimaging Initiative[1], Kathryn L. Lunetta[h], Lindsay A. Farrer[a,b,f,h,i,j], Gyungah R. Jun[b,c,*,2]

[a]*Bioinformatics Graduate Program, Boston University, Boston, MA, USA*
[b]*Department of Medicine (Biomedical Genetics), Boston University, Boston, MA, USA*
[c]*Neurogenetics and Integrated Genomics, Andover Innovative Medicines (AiM) Institute, Eisai Inc, Andover, MA, USA*
[d]*Department of Life Science & Medical Bioscience, Waseda University, Tokyo, Japan*
[e]*Computational Bio-Big Data Open Innovation Lab, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan*
[f]*Department of Neurology, Boston University, Boston, MA, USA*
[g]*Research Organization for Nano & Life Innovation, Waseda University, Tokyo, Japan*
[h]*Department of Biostatistics, Boston University, Boston, MA, USA*
[i]*Department of Ophthalmology, Boston University, Boston, MA, USA*
[j]*Department of Epidemiology, Boston University, Boston, MA, USA*

Abstract

Introduction: Genetic associations for endophenotypes of Alzheimer's disease (AD) in cognitive stages preceding AD have not been thoroughly evaluated.

Methods: We conducted genome-wide association studies for AD-related endophenotypes including hippocampal volume, logical memory scores, and cerebrospinal fluid $A\beta_{42}$ and total/phosphorylated tau in cognitively normal (CN), mild cognitive impairment, and AD dementia subjects from the Alzheimer's Disease Neuroimaging Initiative study.

Results: In CN subjects, study-wide significant ($P < 8.3 \times 10^{-9}$) loci were identified for total tau near *SRRM4* and *C14orf79* and for hippocampal volume near *MTUS1*. In mild cognitive impairment subjects, study-wide significant association was found with single nucleotide polymorphisms (SNPs) near *ZNF804B* for logical memory test of delayed recall scores. We found consistent expression patterns of *C14orf40* and *MTUS1* in carriers with risk alleles of expression SNPs and in brains of AD patients, compared with in the noncarriers and in brains of controls.

Discussion: Our findings for AD-related brain changes before AD provide insight about early AD-related biological processes.

Keywords: Alzheimer's disease; Genome-wide association; Endophenotypes; Cerebrospinal fluid; MRI; Logical memory; Biomarker; ADNI; Tau; Coexpression network

The authors have declared that no conflict of interest exists.

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

[2]Contributed equally to study supervision.

*Corresponding author. Tel.: (978) 722-6705; Fax: (978) 837-4864.
E-mail address: gyungah_jun@eisai.com

https://doi.org/10.1016/j.jalz.2017.11.006

1. Introduction

Alzheimer's disease (AD) is the most common type of dementia and typically occurs after the age of 65 years. It is highly heritable, but the known genetic risk factors (currently numbering more than 25 loci including apolipoprotein E [*APOE*]) account for no more than 50% of the heritability of the disorder [1]. However, genetic association findings based on AD risk do not explain the whole genetic architecture of AD because the mechanistic complexity underlying AD is not captured entirely by disease status, especially in preclinical stages [2,3]. To overcome this limitation and understand preclinical stages of AD, researchers have examined the genetic underpinnings of AD-related endophenotypes including cerebrospinal fluid (CSF) levels of amyloid β peptide (amyloid β 42 [$A\beta_{42}$]) and tau proteins, structural brain changes quantified by magnetic resonance imaging (MRI), and neuropsychological test measures of cognitive functioning, including memory loss [4,5]. Genome-wide association (GWA) studies for AD-related endophenotypes have identified novel loci in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study that enrolled appreciable numbers of subjects across three stages: AD dementia, mild cognitive impairment (MCI), and normal cognitive functioning [6]. Previous ADNI studies indicated the importance of delineating different stages of subjects [7,8]. We hypothesized that some genes may contribute to AD-related processes specifically during stages before AD dementia onset. Genes and pathways that are strongly associated with AD-related endophenotypes in early disease stages may be promising targets for developing AD biomarkers and preventive medicines. To test this hypothesis, we conducted GWA analyses for AD-related endophenotypes in the ADNI sample stratified by stage. Here, we focused on the association tests in the cognitively normal (CN) and MCI subgroups because we were interested in identifying genes that may contribute to AD-related processes before AD dementia onset.

2. Subjects and methods

2.1. Subjects

GWA and phenotype data for ADNI participants were downloaded from a public-access database (http://www.loni.usc.edu). A total of 1189 subjects before quality control (QC) were available with GWA data from two different chips (ADNI-1, n = 757 and ADNI-GO/2, n = 432). We stratified subjects by stage (CN, MCI, and AD dementia) based on diagnosis at the baseline assessment as defined by the standard ADNI protocol. Demographic information and mean endophenotype values stratified by stage as well as for the entire sample are presented in Supplementary Table 1. Age is similarly distributed in each subgroup. Sample sizes for analyses of CSF biomarkers were considerably smaller than for those of other traits.

2.2. Phenotypic evaluation

Previously suggested AD-related endophenotypes including CSF biomarkers [9], MRI brain imaging measures [10], and episodic memory tests [11] were selected for GWA analyses in this study. CSF biomarkers of $A\beta_{42}$, total tau (t-Tau), and phosphorylated tau (p-Tau), brain MRI measure for hippocampal volume (HPV), and scores for logical memory tests of immediate and delayed recall (LMiT and LMdT) which were all measured at baseline were analyzed in this study. Details about collection of CSF biomarkers, brain MRI scan data, and neuropsychological tests are reported elsewhere [12–15].

2.3. Genotyping, quality control, imputation, and population substructure analysis

Details of quality control, genotype imputation, and population substructure analyses are described in Supplementary Materials. After QC, the ADNI-1 sample with genotype data consisted of 187 CN, 329 MCI, and 163 AD dementia subjects, and the ADNI-GO/2 sample contained 118 CN, 252 MCI, and 27 AD dementia subjects with genotype data.

2.4. Statistical methods

2.4.1. GWA tests

Before the association tests, each of the six endophenotypes was adjusted for covariates using linear regression. Age and sex were used as covariates for the six endophenotypes. A term for education level was also included in the regression models for LMiT and LMdT, and the model for HPV was further adjusted for total intracranial volume. The residuals derived from the regression models were rank-transformed for normalization as previously described [16]. Analyses were conducted for all autosomal SNPs using the expected genotype dose, a quantitative measure between 0 and 2 of the number of effect alleles computed from the imputed genotype probabilities as the predictor. Association of the rank-normalized endophenotypes with each SNP was evaluated using a linear regression model including covariate terms for the first three principal components of population substructure using the R software package. The two ADNI data sets were analyzed independently for the CN and MCI subjects, and the results from the two ADNI data sets were combined by meta-analysis using inverse variance weights as implemented in the METAL program [17]. AD cases from the two ADNI data sets were analyzed as one group because the ADNI-GO/2 sample included only 27 AD subjects and regression models for this group included an extra covariate for ADNI data set. The genome-wide significant (GWS) threshold was set at 5.0×10^{-8}. We determined a conservative study-wide significant (SWS) level of 8.3×10^{-9}, which was calculated as the GWS level divided by the effective number of two independent endophenotypes and three clinical subgroups. The effective number of independent

endophenotypes was computed by counting the number of eigenvalues greater than one from the principal components analysis of all six endophenotypes. A threshold of $P < 10^{-6}$ was considered as suggestive evidence of association in the functional/pathway analysis.

2.4.2. Expression SNP analysis

We examined association of the SWS SNPs (allele counts) with transcript-level expression, that is, expression SNP (eSNP), using a publically available database via the Genotype-Tissue Expression (GTEx) portal (http://www.gtexportal.org; [18]).

2.4.3. Differential gene expression analysis

Differential gene expression was evaluated for genes containing or near significantly associated SNPs in two independent human brain expression data sets from the Eisai Bio Bank (EBB) and Mount Sinai Hospital (MSH) (which was downloaded from Gene Expression Omnibus [GEO]: GSE44772). The EBB has gene expression measures obtained from RNA sequencing from the hippocampus (HIPP) of samples collected from autopsied brains from 35 AD cases and 16 normal subjects ascertained at the University of Miami and McLean Hospital (Belmont, MA). A measure of neurofibrillary tangles (Braak stage) in the same samples was assessed following an established protocol [19]. Details about sample collection and preparation and demographic characteristics are provided in Supplementary Information. The association of $\log_2$-transformed transcript expression levels (outcome) with AD status (predictor) was evaluated using linear regression adjusting for site, age, sex, and RNA integrity number. A model testing the effect of Braak stage (0–6 stages) on transcript expression levels was also evaluated and included the same covariates. The MSH microarray expression data (GEO: GSE44772) were generated from autopsied brain tissue collected from dorsolateral prefrontal cortex (DLPFC), visual cortex (VCX), and cerebellum (CER) regions in 129 AD patients and 101 controls. Samples were profiled on a custom-made Agilent 44K array containing 40,638 probes. Gene expression data were normalized using Rosetta Resolver gene-expression analysis software as previously described [20]. The association of $\log_2$-transformed gene expression level (outcome) and AD status (predictor) was tested using a linear regression model adjusting for RNA integrity number, postmortem interval, batch, preservation method, tissue pH, age, and sex. A significance threshold to correct for the number of expression quantitative trait locus (eQTL) analysis tests was applied, which in this case was $P = .025$ since only two genes (*C14orf40* and *MTUS1*) were tested.

2.4.4. Coexpression network analysis of human brain RNA-Seq data

The top-ranked genes at or near (<50 kb) loci that achieved suggestive significance ($P < 10^{-6}$) in GWA tests of any trait in CN or MCI subjects or in the total sample were further evaluated for gene coexpression networks. We built gene coexpression networks in the EBB HIPP RNA-Seq data (41,249 transcripts) by weighted gene coexpression network analysis [21], an approach that defines modules (or subsets) of genes that are highly coexpressed (or coregulated). Details of this approach are described in Supplementary Material. In the HIPP coexpression network, we selected modules which carry both the top-ranked genes from this study and previously known AD genes [22–25]. The selected modules were functionally annotated by two enrichment analyses of "gene ontology (GO)" and "disease-associated genes" with a hypergeometric test. The disease-associated genes, genes involved in risk of diseases, were downloaded from genome-wide association studies (GWAS)db2 [26]. A P value of significance for each enrichment test was calculated along with a false discovery rate (FDR), estimated using the Benjamini-Hochberg procedure [27]. We used an FDR threshold of 0.05 to define associated GOs or diseases. The selected modules were further examined for correlations with traits (AD status and Braak stages) using the EBB sample by calculating the Pearson's correlations between the module eigengene and the traits.

2.4.5. Functional analysis and brain cell type–specific expression profiling

We evaluated predicted functions of SNPs showing suggestive evidence for association ($P < 10^{-6}$) from the GWA tests using HaploReg (http://www.broadinstitute.org/mammals/haploreg/) [28]. The ENCODE database [29] was used to evaluate potential regulatory function. We also investigated the expression profiles of the top-ranked gene from the cerebral cortex of the mouse and human brain (http://web.stanford.edu/group/barres_lab/brainseqMariko/brainseq2.html) [30]. To identify shared functions among the top-ranked genes, we performed functional analysis using Ingenuity Pathway Analysis software (IPA; QIAGEN, Redwood, CA). Ingenuity Pathway Analysis software determines which molecular/cellular function terms (e.g., top-ranked genes in association tests) are statistically overrepresented, suggesting the GWA findings capture functional mechanisms underlying disease-related biological processes. We used a nominal P value threshold of .05 to flag associated functions. We also examined neuronal cell type–specific expression for the top-ranked genes using single-cell RNA sequencing data (which was downloaded from GEO: GSE67835). Further details of this analysis are described in Supplementary Material.

3. Results

3.1. GWA results

There was slight genomic inflation in GWA results for $A\beta_{42}$ and LMdT in CN subjects ($\lambda = 1.02$ for both traits; Supplementary Figs. 1–12). Associations of the *APOE* ε4

allele were SWS with the CSF biomarkers and suggestive with the other traits in MCI subjects, whereas ε4 was significantly associated only with $A\beta_{42}$ level in CN subjects (Supplementary Table 2). Eleven of 25 other previously known AD loci [22–25]—*CR1, INPP5D, MEF2C, HBEGF, HLA region, ZCWPW1, USP6NL, MS4A region, PICALM, SLC2A4A,* and *CASS4*—were nominally significant ($P < .05$) with at least one trait in CN and/or MCI groups (Supplementary Table 3).

Novel SWS associations were observed for several endophenotypes in the CN and/or MCI groups (Table 1). Among CN subjects, t-Tau was associated with SNPs located 28 kb upstream of *SRRM4* (best SNP: rs10775009; $P = 1.6 \times 10^{-9}$; Fig. 1A) and 66.8 kb upstream of *C14orf79* (rs2819438; $P = 6.9 \times 10^{-9}$; Fig. 1B). In the same group, HPV was associated with rs4921790 in *PDGFRL* and near *MTUS1* ($P = 4.6 \times 10^{-9}$; Fig. 1C). This finding was supported by associations with many SNPs in high linkage disequilibrium (LD) ($r^2 > 0.8$ and D'>0.9) with rs4921790, which span *MTUS1*. In the MCI subgroup, LMdT was associated with three SNPs in *ZNF804B* (best SNP: rs73705514; $P = 2.9 \times 10^{-9}$; Fig. 1D). Association was also observed between HPV and an SNP near *LINC00271* and *PDE7B* ($P = 1.76 \times 10^{-8}$) in AD subjects (Supplementary Fig. 13A). Several GWS associations are also noteworthy (Supplementary Table 4) including $A\beta_{42}$ level with *GRIN2B* SNP rs74442473 ($P = 2.52 \times 10^{-8}$; Supplementary Fig. 13B) and rs2378873 near *BRIP1* and *NACA2* ($P = 2.03 \times 10^{-8}$) in CN subjects, LMdT with *DAB1* SNP rs74834332 ($P = 4.20 \times 10^{-8}$) and *PRKG1* SNP rs12268753 ($P = 2.01 \times 10^{-8}$) in CN subjects, LMdT with *ARHGAP24* SNP rs111882035 ($P = 2.74 \times 10^{-8}$) in MCI subjects, and LMiT with *NRG1* SNP rs118130881 ($P = 1.72 \times 10^{-8}$; Supplementary Fig. 13C) in MCI subjects, and $A\beta_{42}$ level with rs55644114 near *GFRA2* and *LZTS1* in the total sample.

3.2. Expression SNP association results

According to the GTEx portal database, several eSNPs under the SWS association peaks for the endophenotypes in the CN or MCI groups are significantly associated with expression of genes in those regions (Table 2). The major allele *C* of intergenic SNP rs2819438, which is associated with increasing CSF t-Tau, was significantly associated with lower expression of *C14orf79* in the brain (*P* values in the HIPP $= 3.8 \times 10^{-4}$ and CER $= 6.0 \times 10^{-3}$; Supplementary Fig. 14A), but not with expression of *PLD4* and *AHNAK2* that are located between rs2819438 and *C14orf79*. SNP rs4921790 was not associated with expression of *MTUS1* ($P > .05$); however, the major allele *G* of rs55653268 (a proxy SNP for rs4921790; $r^2 = 0.7$; D' = 1.0), which is associated with decreasing HPV, was significantly associated with increased expression of *MTUS1* in the caudate ($P = 9.7 \times 10^{-3}$; Supplementary Fig. 14B). *PDGFRL* expression was not tested because its expression in the brain is extremely low according to the GTEx database. *SRRM4* SNP rs119390525 and *ZNF804B* SNP rs73705514 were not significantly associated with expression in any brain regions ($P > .05$).

3.3. Association of gene expression with AD status and Braak stage

Further examination of *C14orf79* and *MTUS1*, whose expression levels were associated with SWS SNPs (or their proxy SNPs), in the EBB RNA-Seq and the MSH microarray data sets revealed multiple significant associations with AD status and Braak stage in several brain regions (Table 3). *C14orf79* was not differentially expressed in the HIPP ($P = .76$), but its expression was significantly lower in AD cases than controls in the CER ($P = 7.5 \times 10^{-8}$), DLPFC ($P = 3.3 \times 10^{-7}$; Fig. 2A), and VCX ($P = 2.1 \times 10^{-8}$).

Table 1
Genome-wide significant association ($P < 5.0 \times 10^{-8}$) of novel genes in CN, MCI, or AD dementia subjects with CSF protein levels ($A\beta_{42}$, t-Tau, and p-Tau), HPV, and LMiT and LMdT

Group	Traits	CHR	BP	SNP	MA	MAF	β	SE	*P* value	Closest genes
ALL	$A\beta_{42}$	8	20647323	rs55644114	A	0.16	0.41	0.07	2.54×10^{-8}	*GFRA2* and *LZTS1*
CN	$A\beta_{42}$	12	13870464	rs74442473	G	0.07	−1.02	0.18	2.53×10^{-8}	*GRIN2B*
		17	59687842	rs2378873	T	0.44	−0.53	0.10	2.03×10^{-8}	*BRIP1* and *NACA2*
	t-Tau	**12**	**119390525**	**rs10775009**	**T**	**0.34**	**0.51**	**0.09**	$\mathbf{1.59 \times 10^{-9}}$	***SRRM4***
		14	**105385352**	**rs2819438**	**A**	**0.13**	**−0.80**	**0.14**	$\mathbf{6.94 \times 10^{-9}}$	***PLD4*** and ***C14orf79***
	HPV	**8**	**17496561**	**rs4921790**	**C**	**0.12**	**0.61**	**0.10**	$\mathbf{4.58 \times 10^{-9}}$	***PDGFRL*** and ***MTUS1***
	LMdT	1	57739164	rs74834332	A	0.03	0.79	0.14	4.30×10^{-8}	*DAB1*
		10	53818149	rs12268753	C	0.21	0.30	0.05	2.01×10^{-8}	*PRKG1*
MCI	LMdT	4	86416554	rs111882035	G	0.02	−0.93	0.17	2.74×10^{-8}	*ARHGAP24*
		7	**88406552**	**rs73705514**	**C**	**0.02**	**−0.84**	**0.14**	$\mathbf{2.86 \times 10^{-9}}$	***ZNF804B***
	LMiT	8	31228770	rs118130881	G	0.04	−0.61	0.11	1.72×10^{-8}	*NRG1*
AD	HPV	6	136077929	rs79846291	T	0.02	1.85	0.31	1.76×10^{-8}	*LINC00271* and *PDE7B*

Abbreviations: CHR, chromosome; BP, basepair position; SNP, single-nucleotide polymorphism; SE, standard error; $A\beta_{42}$, amyloid β 42; AD, Alzheimer's disease; CN, cognitively normal; CSF, cerebrospinal fluid; HPV, hippocampal volume; LMdT, logical memory test of delayed recall; LMiT, logical memory test of immediate recall; MA, minor allele; MAF, minor allele frequency; MCI, mild cognitive impairment; p-Tau, phosphorylated tau; t-Tau, total tau.

NOTE. Bold SNPs denote the study-wide significant association ($P < 8.33 \times 10^{-9}$) with a trait.

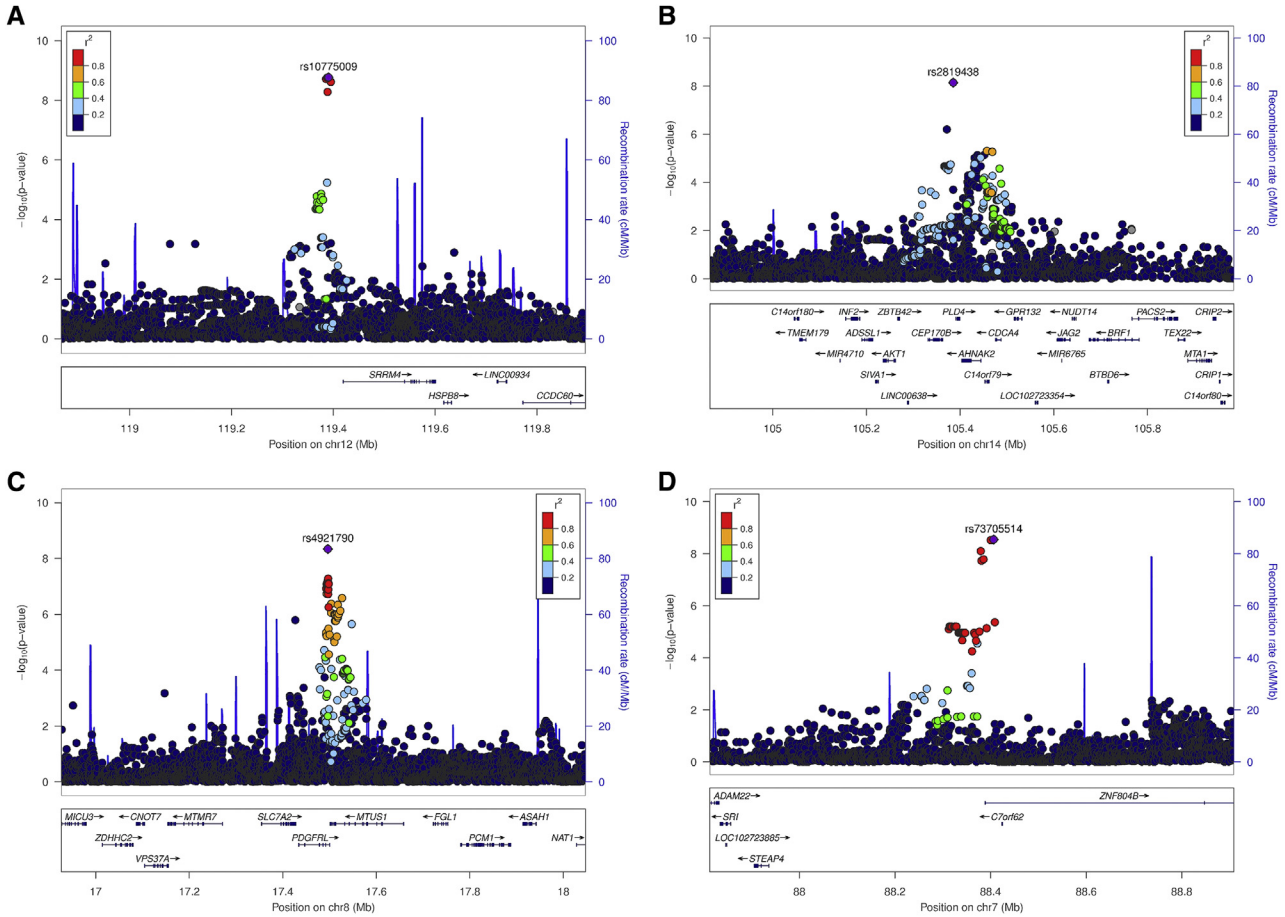

Fig. 1. Regional association plots of (A) *SRRM4* and (B) *C14orf79* for CSF total tau in the CN subjects, (C) *MTUS1* for hippocampal volume in the CN subjects, and (D) *ZNF804B* for logical memory delayed recall test in the MCI subjects. Abbreviations: CSF, cerebrospinal fluid; CN, cognitively normal; MCI, mild cognitive impairment.

The association of *C14orf79* expression with decreasing Braak stage was nearly nominally significant (β [SE] = −0.07 [0.03]; $P = .06$; Fig. 2A). *MTUS1* expression was nominally higher in AD cases than controls in the HIPP ($P = .02$), DLPFC ($P = .01$, Fig. 2B), and VCX ($P = .01$). *MTUS1* expression was also positively correlated with Braak stage (β [SE] = 0.16 [0.04]; $P = 3.9 \times 10^{-4}$; Fig. 2B). According to a publicly available murine brain expression profile data obtained from the cerebral cortex, *MTUS1* is highly expressed in microglia and myeloid cells. Moreover, expression is highest in activated microglia (Supplementary Fig. 15A). However, *MTUS1* expression in the mature human brain is ranked the second lowest in microglia/macrophage cells, whereas expression in the adult mouse brain was ranked the second highest among glia, neurons, and vascular cells. This finding suggests that the role of *MTUS1* may differ in mouse and human myeloid cells (Supplementary Fig. 15B).

3.4. Gene coexpression network analysis in the HIPP

Sixteen of 61 top-ranked genes (Supplementary Table 5) were clustered together with 17 known AD genes as coexpressed networks. Eight of 17 modules from weighted gene coexpression network analysis based on the human HIPP transcriptome (M1–M8) were enriched with the genes ($P < 1.0 \times 10^{-6}$; Supplementary Table 5) identified in our GWAS analyses with AD-related endophenotypes (Supplementary Table 6). These eight modules were also significantly correlated with AD status or Braak stages (Fig. 3A). *C14orf79* was coexpressed with *MAPT-AS1* in M1. *MTUS1* was coexpressed with *ZCWPW1* in M4. *SRRM4* and *ZNF804B* were coexpressed with previously known AD genes—*APP*, *BZRAP1*, *MAPT*, *MEF2C*, *PLXNA4*, *PTK2B*, and *PSEN2*—in M7. The M7 module is enriched for genes involved in neuronal processes including "chemical synaptic transmission" (FDR = 1.8×10^{-84}), "dendrite" (FDR = 1.7×10^{-33}), and "axon guidance" (FDR = 1.6×10^{-24}). Modules M2, M4, M5, M6, and M8 are enriched for genes involved in "protein binding", "signal transduction", and "microtubule binding" (Supplementary Table 6). Among the eight modules enriched for previously known AD risk genes and genes identified in the present study as associated with AD endophenotypes, modules M1, M2, M5, M7, and M8 were also highly enriched with genes for other neurodegenerative

Table 2
Genotype-specific effect of the expression level among study-wide significant (8.33×10^{-9}) SNPs using the GTEx portal database

				eSNP association summary		
				Hippocampus	Other brain region	
eQTL	EA	RA	*Gene*	*P* value (β)	*P* value (β)	Region
rs10775009	T	C	*SRRM4*	.09 (0.09)	.16 (0.10)	Frontal cortex
rs2819438	A	C	*PLD4*	.47 (0.12)	.10 (0.42)	Anterior cingulate cortex
			AHNAK2	.15 (0.23)	.06 (−0.28)	Cerebellar Hemisphere
			C14orf79	3.8×10^{-4} (0.44)	3.9×10^{-3} (0.31)	Cortex
					6.0×10^{-3} (0.44)	Cerebellum
rs4921790	C	A	*MTUS1*	1.0 (0.0)	.09 (−0.16)	Caudate
rs55653268	T	G		.8 (0.04)	9.7×10^{-3} (−0.33)	Caudate
					.02 (−0.41)	Nucleus accumbens

Abbreviations: eQTL, expression quantitative trait locus; GTEx, Genotype-Tissue Expression; eSNP, expression SNP; EA, effect allele; RA, reference allele.
NOTE. Positive effect (β) in eSNP association summary means that carriers with effect alleles of an SNP tend to have higher expression level of a gene; and rs55653268 is in high LD with rs4921790 ($r^2 = 0.7$; D' = 1.0) and is significantly associated with the hippocampal volume (EA: T, β [SE]: 0.6 [0.1], *P* value: 4.2×10^{-7}).

and neuropsychiatric disorders including Parkinson disease and schizophrenia, as well as for cognitive performance (Supplementary Table 6). The correlation matrix of expression levels among the genes in modules M4 and M7 (Fig. 3B) is shown in Supplementary Table 7. Further evaluation revealed that M2 and M7 contained the largest number of known AD genes among all modules coregulated in the hippocampal region (Supplementary Table 8).

3.5. Functional analysis

The top-ranked genes from the GWA analyses in the CN and MCI groups (Supplementary Table 5) are enriched in neuronal processes including synapse plasticity ($P = 3.4 \times 10^{-5}$), axon quantity ($P = 2.6 \times 10^{-4}$), microtubule dynamics ($P = 2.8 \times 10^{-4}$), abnormal morphology of dentate gyrus ($P = 3.4 \times 10^{-4}$), and neuronal development ($P = 3.5 \times 10^{-4}$). These findings indicate that genes associated with AD-related endophenotypes among participants in clinical stages preceding AD dementia have particular roles in functioning of neuronal synapses (Fig. 3C). Single-cell transcriptome analysis of the human brain confirmed that the genes in Fig. 3C were highly expressed in neurons. Some genes were also expressed in other cell types: *APOE*, *MTUS1*, *ERBIN*, and *AKAP9* in astrocytes; *MTUS1*, *ERBIN*, and *AKAP9* in oligodendrocytes; and *AKAP9* and *RCAN1* in endothelial cells. Bioinformatic analysis using HaploReg and ENCODE suggests that GWS SNPs upstream of *NRG1* may be located in a transcriptional regulatory site and rs111969453—an SNP in high linkage disequilibrium ($r^2 = 0.94$ and D' = 0.97) with the SNP (rs118130881) that is significantly associated with LMiT ($P = 5.9 \times 10^{-8}$)—is located on an enhancer histone mark in the brain.

4. Discussion

A goal of this study was to identify genes that were both associated with AD-related endophenotypes in older, nondemented individuals and coregulated with known AD genes. Using a GWA approach, we identified SWS associations in CN elders including CSF t-Tau level with *SRRM4* and *C14orf79,* HPV with *MTUS1*, and CSF $A\beta_{42}$ level with the *APOE ε4* allele. In MCI subjects, we detected SWS associations for LMdT with *ZNF804B* and for CSF levels of $A\beta_{42}$, t-Tau, and p-Tau with the *APOE ε4* allele.

Angiotensin II–interacting (AT2) protein (ATIP) isoforms encoded by *MTUS1* are highly expressed in most brain regions [31]. ATIP1 binds to AT2 proteins, mediating neuronal differentiation survival and regeneration in the

Table 3
Summary statistics of association between expression levels of the genes (*MTUS1* and *C14orf79*) and AD status as well as Braak stage

			MTUS1			*C14orf79*		
Source	Brain region	Predictor	β	SE	*P* value	β	SE	*P* value
Eisai Bio Bank (RNA-Seq)	HIPP	AD Status	0.38	0.16	.02	0.04	0.14	.76
	HIPP	Braak Stages	0.16	0.04	3.9×10^{-4}	−0.07	0.03	.06
Mt. Sinai Hospital (Microarray)	CER	AD Status	0.00	0.02	.78	−0.13	0.02	7.5×10^{-8}
	DLPFC	AD Status	0.09	0.04	.01	−0.13	0.02	3.3×10^{-7}
	VCX	AD Status	0.09	0.03	.01	−0.15	0.03	2.1×10^{-8}

Abbreviations: AD, Alzheimer's disease; CER, cerebellum; DLPFC, dorsolateral prefrontal cortex; HIPP, hippocampus; VCX, visual cortex.
NOTE. Brain regions tested were HIPP, CER, DLPFC, and VCX. Positive effect (β) for AD status means that a gene is upregulated in AD cases versus in controls. Positive effect for Braak stages means that the expression of a gene is positively correlated with the Braak stages.

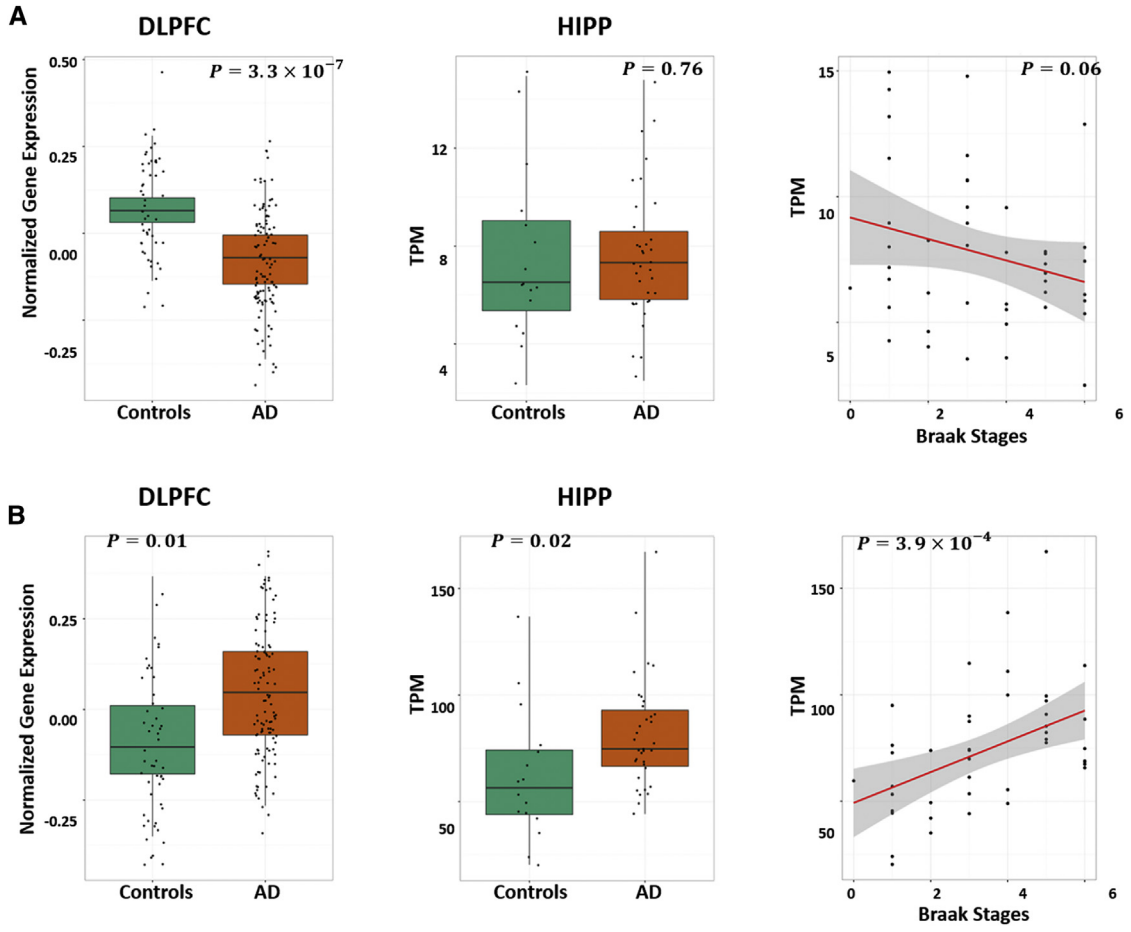

Fig. 2. Expression studies of (A) *C14orf79* and (B) *MTUS1* including box plots for differentially expressed genes in the brain DLPFC in the microarray data (GEO: GSE44772; left column) and in the HIPP in the RNA-Seq data (Eisai Bio Bank data; middle column) and regression plots of gene expression by Braak stage in the RNA-Seq data (Eisai Bio Bank data; right column). *P* values in plots were computed from linear regression models after adjusting for covariates (details in the Methods). Abbreviations: AD, Alzheimer's disease; DLPFC, dorsolateral prefrontal cortex; GEO, Gene Expression Omnibus; HIPP, hippocampus.

brain [31–33]. ATIP3 colocalizes with microtubules and regulates their polymerization, thereby regulating neuronal differentiation and neurite outgrowth [33]. Gene expression analysis revealed that the risk allele of eSNP near *MTUS1* was associated with increased expression of *MTUS1* in the caudate. Differential gene expression analyses demonstrated that expression of *MTUS1* was higher in AD cases than controls in the HIPP, DLPFC, and VCX. Also, expression of *MTUS1* was significantly greater in brains showing severe neurofibrillary tangle involvement. The ADNI GWAS findings together with these expression findings suggest that *MTUS1* expression may be related to changes in HPV before onset of cognitive impairment.

SRRM4 encodes the neural-specific Ser/Arg repeat–related protein of 100 kDa (nSR100) that promotes neurite outgrowth and alternative splicing and controls most neural microexons [34–36]. Downregulated *SRRM4* alters splicing of microexons in autism brains [34]. Association of *SRRM4* SNP rs1997111 with t-Tau level in a subset of ADNI controls (i.e., from "ADNI-1" only) was previously identified at a significance level one order of magnitude less ($P = 1.1 \times 10^{-8}$) than the result we obtained among CN subjects in our study ($P = 1.76 \times 10^{-9}$) [37]. The functional relevance of *C14orf79* and *ZNF804B,* the two other SWS but poorly characterized genes, to AD is unclear.

GRIN2B encodes the GluN2B subunit of the N-methyl-D-aspartate (NMDA) receptor (GluN2B-NMDA) that is involved in synaptic plasticity and memory function. GluN2B-NMDA is the target of memantine, a drug that provides symptomatic relief in patients with AD by antagonizing the GluN2B-NMDA receptor [38]. A suggestive association of *GRIN2B* with temporal lobe volume ($P = 1.3 \times 10^{-7}$) was reported in a previous GWA study in the ADNI-1 sample [39]. *NRG1* functions as an epidermal growth factor in the nervous system [40] and is involved in the neuregulin signal transduction pathway for synapse maturation and dendritic morphology [41,42].

Among the known AD genes, we found suggestive evidence for association with several SNPs located 381 kb upstream of *AKAP9* with p-Tau level in MCI subjects (Supplementary Table 2 and Supplementary Fig. 13D). We also found a suggestive association between rs149454736,

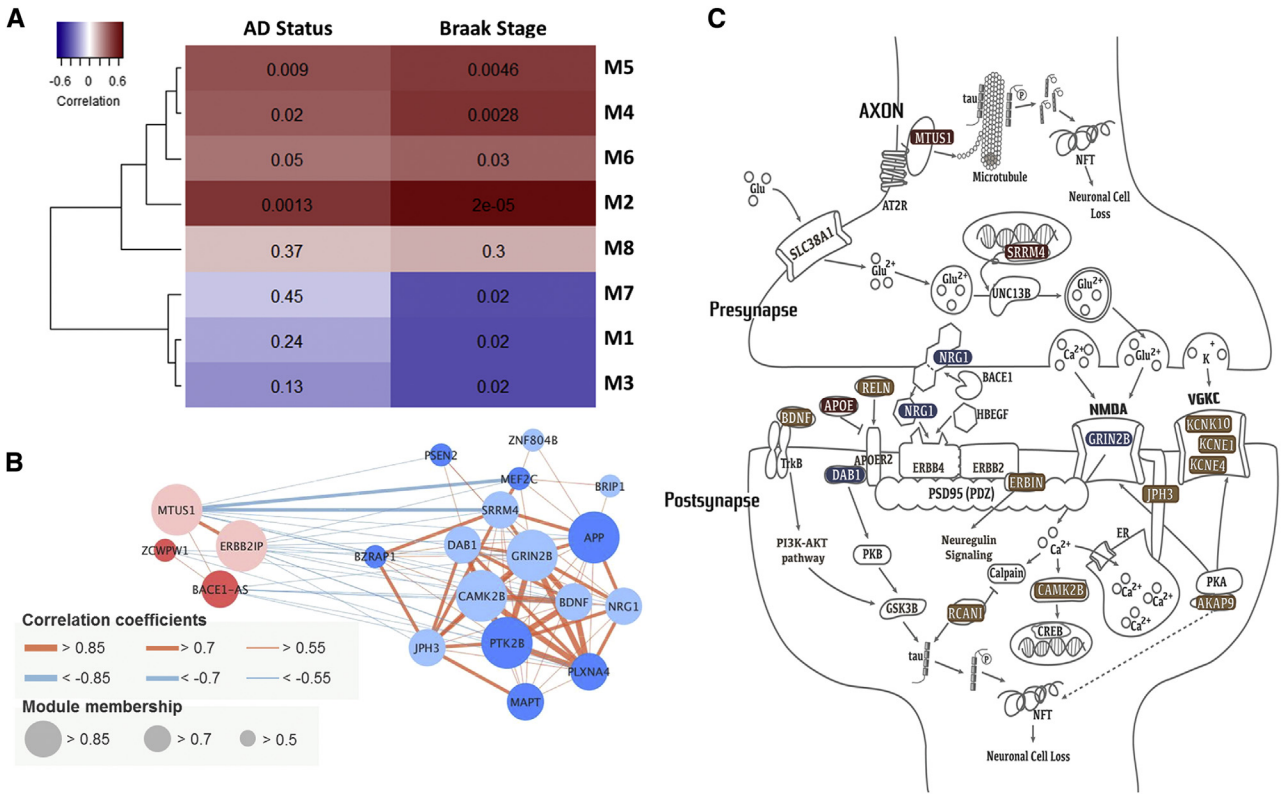

Fig. 3. (A) A heatmap of correlations between AD traits and the first principal component values of genes in modules. Deeper colors indicate higher correlation with the traits (red: positive and blue: negative). Values in the heatmap are correlation *P* values. (B) Connectivity plot of genes in modules M4 and M7, which were identified in this study or in previously reported in AD GWAS. Light colored circles indicate the genes identified in this study, and deep colored circles indicate previously identified AD genes. (C) The role of genes identified in GWAS of AD-related endophenotypes among cognitively normal and mild cognitively impaired subjects in the top-ranked canonical pathways; colors indicate the level of association significance of the genes identified in this study (red = study-wide significance; blue = genome-wide significance; and brown = $P < 1.0 \times 10^{-6}$; except for *BDNF* in Supplementary Fig. 13H). Abbreviations: AD, Alzheimer's disease; APOE, apolipoprotein E; GWAS, genome-wide association studies; NMDA, N-methyl-D-aspartate.

located between exons 45 and 46 of *AKAP9,* and HPV in AD dementia subjects ($P = 2.2 \times 10^{-7}$). Previously, we identified significant association of AD with two rare coding *AKAP9* missense mutations in exons 31 (rs144662445) and 46 (rs149979685) in African Americans [43], and our study is the first to report association with this gene for AD risk in the European ancestry individuals; rs149454736 is located 1.5 kb away from rs149979685. *AKAP9* has functional similarity with the tau protein in terms of microtubule stability and assembly [44].

With the exception of *APOE*, there is little overlap in GWS findings between our study and an investigation of genetic determinants of CSF $A\beta_{42}$, t-Tau, and p-Tau levels in a much larger sample that included 787 ADNI subjects [45]. Some of the differences may be explained by our study design, which conducted analyses within separate groups of CN, MCI, and AD subjects. The prior study combined cognitive groups and included multiple cohorts with highly variable ascertainment and distributions of subjects across cognitive groups. In addition, we extended our findings by incorporating them in analyses of coregulated transcriptional networks enriched with previously established AD genes. Further comparison of results across the two studies revealed that four of the six GWS SNPs (excluding the *APOE* region) in the Deming *et al.* report [45] were nominally significant in specific cognitive groups in our study (Supplementary Table 9). For example, the association of rs185031519 near *GLIS1* CSF $A\beta_{42}$ ($P = 2.1 \times 10^{-8}$) in the prior study was nominally significant in the CN group ($P = 8.3 \times 10^{-3}$) in our study. Several of our most significant findings (including *SRRM4*, *MTUS1*, and *GRIN2B*) are consistent with results of other studies of AD-related endophenotypes (Supplementary Table 10).

Our findings may provide important insights about the sequence of processes leading to AD. The SWS associations in CN subjects (*SRRM4* with t-Tau and *MTUS1* with HPV) implicate neuronal signaling, development, and loss, but with the exception of *APOE*, they do not involve Aβ processing in the asymptomatic stage of AD. It is noteworthy that the *APOE* ε4 allele was not associated with the CSF tau biomarkers and other endophenotypes in CN subjects. The variants associated with CSF biomarkers and HPV could also be interpreted to be markers of cognitive reserve/resilience because they predict the extent of AD pathology in CN persons, but not in MCI or AD subjects. There is extensive evidence supporting the cognitive reserve hypothesis in AD

[46]. Also, the variants identified for memory performance and HPV in CN individuals could be markers of inherent memory function and HPV, completely independent of AD.

Our study has several limitations. First, the sample sizes for analyses of all traits in each clinical group, especially AD cases, were relatively small. Thus, we had low power to identify variants having small effects. In addition, it is possible that our top findings are false positives. However, the significant SNPs were supported by evidence in both constituent data sets (ADNI-1 and ADNI-GO/2) and from expression data analysis. Second, in the transcriptomic analyses, we did not consider differences in cellular composition between AD and control brain tissue. Therefore, our results may have resulted from the excessive neuronal death in brains from AD subjects compared with controls. However, this concern would not impact our finding of increased *MTUS1* expression in AD brains. Third, the two expression data sets generated using different platforms (RNA-Seq vs. microarray) from different brain regions (HIPP vs. CER/DLPFC/VCX) limit direct comparisons between these data sets. Finally, it is necessary to repeat analyses in independent samples to confirm our findings and increase power to elevate the significance of true associations that did not attain study-wide significance.

In summary, we identified novel genes associated with AD-related endophenotypes in CN and MCI subjects. These genes had not been previously identified with AD risk, and most of them are involved in neuronal development and signaling. Our findings suggest that genes influencing AD-related processes in individuals with normal cognition or with MCI may differ from those influencing these processes in individuals with AD dementia but regulated together in the transcription level.

Acknowledgments

This study was supported in part by National Institutes of Health (NIH) grants RF1-AG057519, R01-AG048927, U01-AG032984, UF1-AG046198, and P30-AG13846. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; NIH Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. The authors thank Eisai Bio Bank, the University of Miami Brain Endowment Bank, and the Harvard Brain Tissue Resource Center, funded through NIH-NeuroBioBank HHSN-271-2013-00030C the National Institute of Mental Health (NIMH), National Institute of Neurological Diseases and Stroke (NINDS), and Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), and brain donors and their families for the tissue samples used in these studies.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jalz.2017.11.006.

RESEARCH IN CONTEXT

1. Systematic review: We reviewed previously published GWAS for AD-related endophenotypes. Few GWAS have been conducted for changes of endophenotypes in prediagnostic stages.
2. Interpretation: GWAS in the normal stage identified novel genome-wide significant associations with SNPs near SRRM4 and C14orf79 for the total tau level and MTUS1 for the hippocampal volume. In the MCI stage, GWAS of logical memory test score detected another genome-wide significant association with SNPs near ZNF804B.
3. Future directions: These results should be confirmed in independent sample including the same endophenotypes measured prior to AD diagnosis.

References

[1] Naj AC, Jun G, Beecham GW, Wang L-S, 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.

[2] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–53.

[3] Ertekin-Taner N, De Jager PL, Yu L, Bennett DA. Alternative approaches in gene discovery and characterization in Alzheimer's disease. Curr Genet Med Rep 2013;1:39–51.
[4] Glahn DC, Paus T, Thompson PM. Imaging genomics: mapping the influence of genetics on brain structure and function. Hum Brain Mapp 2007;28:461–3.
[5] Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 2003; 160:636–45.
[6] Shen L, Thompson PM, Potkin SG, Bertram L, Farrer LA, Foroud TM, et al. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. Brain Imaging Behav 2014; 8:183–207.
[7] Nho K, Corneveaux JJ, Kim S, Lin H, Risacher SL, Shen L, et al. Whole-exome sequencing and imaging genetics identify functional variants for rate of change in hippocampal volume in mild cognitive impairment. Mol Psychiatry 2013;18:781–7.
[8] Roostaei T, Nazeri A, Felsky D, De Jager PL, Schneider JA, Pollock BG, et al. Genome-wide interaction study of brain beta-amyloid burden and cognitive impairment in Alzheimer's disease. Mol Psychiatry 2017;22:287–95.
[9] Kim S, Swaminathan S, Shen L, Risacher SL, Nho K, Foroud T, et al. Genome-wide association study of CSF biomarkers Abeta1-42, t-tau, and p-tau181p in the ADNI cohort. Neurology 2011;76:69–79.
[10] Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. NeuroImage 2010;53:1051–63.
[11] Pedraza O, Allen M, Jennette K, Carrasquillo M, Crook J, Serie D, et al. Evaluation of memory endophenotypes for association with CLU, CR1, and PICALM variants in black and white subjects. Alzheimers Dement 2014;10:205–13.
[12] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol 2009; 65:403–13.
[13] Furney SJ, Simmons A, Breen G, Pedroso I, Lunnon K, Proitsi P, et al. Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. Mol Psychiatry 2011; 16:1130–8.
[14] Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27:685–91.
[15] Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement 2012; 8:S1–68.
[16] Jun G, Guo H, Klein BEK, Klein R, Wang JJ, Mitchell P, et al. *EPHA2* is associated with age-related cortical cataract in mice and humans. Plos Genet 2009;5:e1000584.
[17] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010; 26:2190–1.
[18] Ardlie KG, Deluca DS, Segrè AV, Sullivan TJ, Young TR, Gelfand ET, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648–60.
[19] Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991;82:239–59.
[20] Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov Alexei A, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 2013; 153:707–20.
[21] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 2008;9:559.
[22] 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 2013; 45:1452–8.
[23] Jun G, Asai H, Zeldich E, Drapeau E, Chen C, Chung J, et al. PLXNA4 is associated with Alzheimer disease and modulates tau phosphorylation. Ann Neurol 2014;76:379–92.
[24] 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.
[25] Jun GR, Chung J, Mez J, Barber R, Beecham GW, Bennett DA, et al. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. Alzheimers Dement 2017;13:727–38.
[26] Li MJ, Liu Z, Wang P, Wong MP, Nelson MR, Kocher JP, et al. GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res 2016; 44:D869–76.
[27] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995;57:289–300.
[28] Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930–4.
[29] An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
[30] Bennett ML, Bennett FC, Liddelow SA, Ajami B, Zamanian JL, Fernhoff NB, et al. New tools for studying microglia in the mouse and human CNS. Proc Natl Acad Sci U S A 2016;113:E1738–46.
[31] Guimond M-O, Gallo-Payet N. How does angiotensin AT2 receptor activation help neuronal differentiation and improve neuronal pathological situations? Front Endocrinol (Lausanne) 2012;3:164.
[32] Li J-M, Mogi M, Tsukuda K, Tomochika H, Iwanami J, Min L-J, et al. Angiotensin II-induced neural differentiation via angiotensin II type 2 (AT2) receptor-MMS2 cascade involving interaction between AT2 receptor-interacting protein and Src homology 2 domain-containing protein-tyrosine phosphatase 1. Mol Endocrinol 2007;21:499–511.
[33] Rodrigues-Ferreira S, le Rouzic E, Pawlowski T, Srivastava A, Margottin-Goguet F, Nahmias C. AT2 receptor-interacting proteins ATIPs in the brain. Int J Hypertens 2013;2013:6.
[34] Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell 2014;159:1511–23.
[35] Vuong CK, Black DL, Zheng S. The neurogenetics of alternative splicing. Nat Rev Neurosci 2016;17:265–81.
[36] Ohnishi T, Shirane M, Nakayama KI. SRRM4-dependent neuron-specific alternative splicing of protrudin transcripts regulates neurite outgrowth. Scientific Rep 2017;7:41130.
[37] Han MR, Schellenberg GD, Wang LSAlzheimer's Disease Neuroimaging Initiative. Genome-wide association reveals genetic effects on human Aβ42 and tau protein levels in cerebrospinal fluids: a case control study. BMC Neurol 2010;10:90.
[38] Xia P, Chen HS, Zhang D, Lipton SA. Memantine preferentially blocks extrasynaptic over synaptic NMDA receptor currents in hippocampal autapses. J Neurosci 2010;30:11246–50.
[39] Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, et al. Voxelwise genome-wide association study (vGWAS). NeuroImage 2010; 53:1160–74.
[40] Woo R-S, Lee J-H, Yu H-N, Song D-Y, Baik T-K. Expression of ErbB4 in the apoptotic neurons of Alzheimer's disease brain. Anat Cell Biol 2010;43:332–9.
[41] Tao Y, Dai P, Liu Y, Marchetto S, Xiong W-C, Borg J-P, et al. Erbin regulates NRG1 signaling and myelination. Proc Natl Acad Sci U S A 2009;106:9477–82.
[42] Wadugu B, Kühn B. The role of neuregulin/ErbB2/ErbB4 signaling in the heart with special focus on effects on cardiomyocyte proliferation. Am J Physiol Heart Circ Physiol 2012;302:H2139–47.

[43] Logue MW, Schu M, Vardarajan BN, Farrell J, Bennett DA, Buxbaum JD, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. Alzheimers Dement 2014;10:609–618.e11.
[44] Rivero S, Cardenas J, Bornens M, Rios RM. Microtubule nucleation at the cis-side of the Golgi apparatus requires AKAP450 and GM130. EMBO J 2009;28:1016–28.
[45] Deming Y, Li Z, Kapoor M, Harari O, Del-Aguila JL, Black K, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. Acta Neuropathol 2017;133:839–56.
[46] Stern Y. Cognitive reserve in ageing and Alzheimer's disease. Lancet Neurol 2012;11:1006–12.