

Identification of Key Long Non-Coding RNAs in the Pathology of Alzheimer's Disease and their Functions Based on Genome-Wide Associations Study, Microarray, and RNA-seq Data

Zhijie Han^{a,b}, Weiwei Xue^b, Lin Tao^c and Feng Zhu^{a,b,*}

^a*Innovative Drug Research and Bioinformatics Group, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China*

^b*School of Pharmaceutical Sciences and Collaborative Innovation Center for Brain Science, Chongqing University, Chongqing, China*

^c*Key Laboratory of Elemene Class Anti-cancer Chinese Medicine of Zhejiang Province, School of Medicine, Hangzhou Normal University, Hangzhou, P. R. China*

Handling Associate Editor: Jin-Tai Yu

Accepted 26 December 2018

Abstract. The pathogenesis of Alzheimer's disease (AD) is identified to be significantly regulated by long non-coding RNA (lncRNA) based on *in vivo* and clinical experiments. Single nucleotide polymorphisms (SNPs) can strongly impact expression and function of lncRNA in AD, and previous genome-wide associations studies (GWAS) have discovered substantial amount of risk SNPs associated with AD. However, current studies omit the important information about SNPs when identifying potential AD-related lncRNAs. In addition to single discovery approach and small-scale samples in these studies, the number of lncRNAs discovered as keys in AD is limited. Here, multiple computational methods were integrated to discover novel and key lncRNA of the pathology of AD. First, large-scale GWAS data involved in three ethnicities were collected from two authoritative sources, and meta-analyses were conducted to find SNPs significantly associated with AD (tag SNPs). Second, these tag SNPs together with their linkage disequilibrium information were used to discover potential lncRNAs related to AD. Third, after validation by microarray probe re-annotation of 1,282 samples and RNA-seq data analysis of 117 samples, respectively, a total of five key lncRNAs of AD were identified. Finally, possible function of these lncRNAs was predicted by genome mapping, expression quantitative trait loci, differential co-expression, and gene set enrichment analysis. Based on function prediction, four of the five key lncRNAs were identified to affect the risk of AD by regulating corresponding pathogenic genes and pathways, which are involved in regulation of amyloid- β peptide and the immune system. In summary, these findings can facilitate the discovery of potential disease-related lncRNAs and enhance understanding of the pathogenesis of AD.

Keywords: Alzheimer's disease, differential co-expression, genome-wide association study, long non-coding RNA, microarray probe re-annotation

*Correspondence to: Prof. Feng Zhu, Innovative Drug Research and Bioinformatics Group, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China. Tel.: +86 571

88208444; Fax: +86 571 88208444; E-mails: zhufeng@zju.edu.cn or prof.zhufeng@gmail.com.

INTRODUCTION

Alzheimer's disease (AD) is a chronic neurodegenerative disorder that starts slowly and worsens over time [1]. The accumulation of amyloid- β peptide ($A\beta$) is commonly recognized as a main characterization and driver of AD pathogenesis [2, 3]. So far, >35 million people around the world and 5.5 million in the United States suffer from AD [4]. Because of the lack of effective preventive measures and drugs, the number of AD patients is estimated to rapidly increase in the next forty years [5–7]. In particular, according to two independent studies in 2011 and 2015, newly diagnosed AD patients are expected to reach 10–15 million and 135 million by 2050 in the United State and around the world, respectively [8–10].

Long non-coding RNAs (lncRNAs) are a kind of considerable non-protein coding transcripts [11]. Until now, about 167,150 lncRNAs have been identified in the human genome [12], and hundreds of them are considered to be involved in many key aspects of cellular process, such as establishment of cell identity, cell cycle control, and apoptosis [13]. Recent studies have shown that the pathogenesis of many human diseases involve the dysregulations of lncRNAs [14–16], and several AD-related lncRNAs have been identified by experiments [17–19]. For example, Mus et al. found that lncRNA BC200, a translational regulator targeting eukaryotic initiation factor 4A, is significantly upregulated in AD patients compared to age-matched healthy individuals. The overexpression of BC200 can impair synaptic plasticity in AD neurons [20]. Faghihi et al. found that lncRNA BACE1-AS, a conserved antisense transcript β -site amyloid precursor protein, is a crucial regulator in AD pathophysiology. BACE1-AS concentration is significantly elevated in AD patients, and the overexpressed BACE1-AS enhances amyloid- β protein precursor ($A\beta$ PP) processing and the abundance of $A\beta_{1-42}$ in AD brain [17]. Ciarlo et al. identified an AD-related lncRNA 51A which maps in antisense configuration to intron 1 of the SORL1 gene. The overexpressed 51A in AD brain results in a decreased synthesis of SORL1 variant A (a canonical long protein associated with impaired processing of $A\beta$ PP) by affecting the alternative splicing of SORL1 gene, and further leads to increased $A\beta$ formation [18]. Moreover, according to genome-wide association study (GWAS) findings by Lambert et al., two single nucleotide polymorphisms (SNPs) rs190982 and

rs11771145 (locating in the MEF2C and EPHA1, respectively) are significantly associated with AD, and MEF2C and EPHA1 are actually located in lncRNAs [21]. These experiments identified the novel lncRNAs with great accuracy, but their significant costs in time and investment limit our understanding of the function of lncRNAs in the pathology of AD [22–26]. Compared with other complex diseases, only a few AD-related lncRNAs and their functions have been little identified [27–30]. In order to overcome the deficiencies, computational methods are applied in the identification of potential AD-related lncRNAs which can substantially enhance the discovery efficiency [31–34]. Recently, a strategy of microarray probe reannotation was adopted and 108 lncRNAs were found to be potentially related to AD in temporal cortex [31]. Another study finds 488 differential expressed lncRNAs between AD patients and healthy individuals in entorhinal cortex by a similar approach [32]. However, because of the single discovery approach and small-scale samples (16 and 20 individuals, respectively) utilized in those studies, the results are not fairly reliable making it difficult to explain the functions of these lncRNAs, and it is highly possible that some novel lncRNAs playing a key role in AD are omitted.

Recently, the SNPs were identified to contribute to the dysregulation of lncRNAs expression and play an important role in the association between lncRNAs and diseases [35–39]. Ever since the first application of GWAS in 2005 [40], a large number of SNPs significantly associated with various diseases have been identified on a genome-wide scale. For example, Harold et al. compared 3,941 AD cases with 7,848 controls and identified two susceptibility SNPs (rs11136000 and rs3851179) at the CLU and PICALM gene, respectively [41]. Naj et al. analyzed the genotyping data of 6,922 AD cases and 24,666 controls and identified four risk loci (rs4938933, rs9349407, rs11767557, and rs3865444) at the MS4A4A, CD2AP, EPHA1, and CD33 gene, respectively [42]. In 2014, a susceptibility locus (rs277470) at PLXNA4 gene was identified associated with AD among Caucasian, African American, and Japanese populations [43]. Recently, three common loci (rs3764650, rs3752246, and rs4147929) in ABCA7 gene were re-assessed and confirmed to increase the risk of AD by a meta-analysis [44]. Nevertheless, nearly all of these discoveries focus on the protein-coding genes, and the non-coding risk variants are ignored [45]. Moreover, the existence of

linkage disequilibrium (LD) in GWAS can present many SNPs associated with AD even if only one of them is the true risk variant, which may lead to neglect of the association between disease and lncRNA SNPs if they are in the LD with the nominal susceptible SNPs [46]. Therefore, the mining from GWAS data can facilitate the effective discovery of novel AD-related lncRNAs.

In this study, multiple computational approaches were integrated to identify novel lncRNAs playing a key role in the pathology of AD. First, large-scale GWAS data involving three ethnicities (European, African, and Asian ancestry) were collected from AlzGene [47] and International Genomics of Alzheimer's Project (IGAP) [21], and the meta-analyses were then conducted to eliminate the inconsistency and find the SNPs significantly associated with AD (tag SNPs). Second, the tag SNPs together with their LD information were used to discover the lncRNAs potentially related to AD. Third, 1,282 microarray samples and 117 RNA-seq samples from the Gene Expression Omnibus (GEO) database were used to validate and identify key lncRNAs of AD. Finally, a total five key lncRNAs of AD were identified, and their possible functions were predicted by applying genome mapping, expression quantitative trait loci (eQTL), differential co-expression, and gene set enrichment analysis (GSEA). The brief flow chart of this study is illustrated in Fig. 1 (details are described in Supplementary Figure 1). In summary, these findings can facilitate the discovery of novel and key lncRNAs in AD and enhance the understanding of the pathogenesis of AD.

METHODS

Collecting GWAS data and identifying AD-related tag SNP

AlzGene is a publicly available resource providing AD genetic variants. It catalogued 695 genes and 2,973 polymorphisms from 1,395 GWAS studies involved in European, African, and Asian ancestry ethnicities [47]. We used the R package 'meta' (<http://cran.r-project.org/web/packages/meta/index.html>) to conduct systemic meta-analyses. First, we extracted the detailed genotype data of each SNP both in AD patients and controls from AlzGene. Second, we used two quantities, Cochran's Q and I^2 , to measure the heterogeneity of each SNP among different studies. The Cochran's Q approximately followed a chi-squared distribution with $k-1$ degrees

of freedom (where k was the number of studies), and the I^2 value was calculated through Cochran's Q ($I^2 = \frac{Q-(k-1)}{Q} \times 100\%$), which ranged from 0 to 100%. The threshold of significant heterogeneity was set as $I^2 > 50\%$ and $p > 0.001$ according to previous studies [44, 48–52]. Finally, according to the results of the heterogeneity test, the fixed effect and random models were used for the meta-analysis when the heterogeneity was significant or not, respectively. In addition, to ensure the completeness of the tag SNPs dataset, more GWAS data were selected according to the meta-analysis results of IGAP (combined stage 1 and 2). The IGAP database includes the results of a large two-stage study involved 17,008 AD cases and 37,154 controls with European ancestry. The study performed 2,013 meta-analyses for the AD GWAS data from four consortia: 1) the Genetic and Environmental Risk in Alzheimer's Disease (GERAD), 2) the European Alzheimer's Disease Initiative (EADI), 3) the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), and 4) the Alzheimer's Disease Genetic Consortium (ADGC) [21]. All the participants of the GWAS studies in AlzGene and IGAP databases are not duplicated. The meta-analysis p -values for each SNP in the two sources are corrected by multiple testing (Benjamini–Hochberg method). In particular, the SNPs significantly correlated with AD (p -values less than 0.05 after multiple testing) were defined as tag SNPs in this study.

Identifying the non-coding SNPs in LD with the tag SNPs

HaploReg is a tool for exploring annotations about the linked variants at the regions of non-coding sequence, which is based on the LD information of SNPs from the 1000 Genomes Project [46]. HaploReg (v4) defined a core set of 52,054,804 variants mainly consisting of SNPs for the four ancestral super-populations, i.e., African (AFR), American (AMR), Asian (ASN), and European (EUR) ancestry [46]. We uploaded these tag SNPs to the HaploReg using a text file (one tag SNP ID per line), and calculated the squared correlation (r^2) between each tag SNP and the non-coding SNPs in different ethnicities (including African, American, Asian, and European). Then, according to the ethnicities of tag SNPs, we selected the non-coding SNPs in LD with them in corresponding ethnicities (the LD threshold was set as $r^2 \geq 0.8$). Parameters of HaploReg were set to the default values.

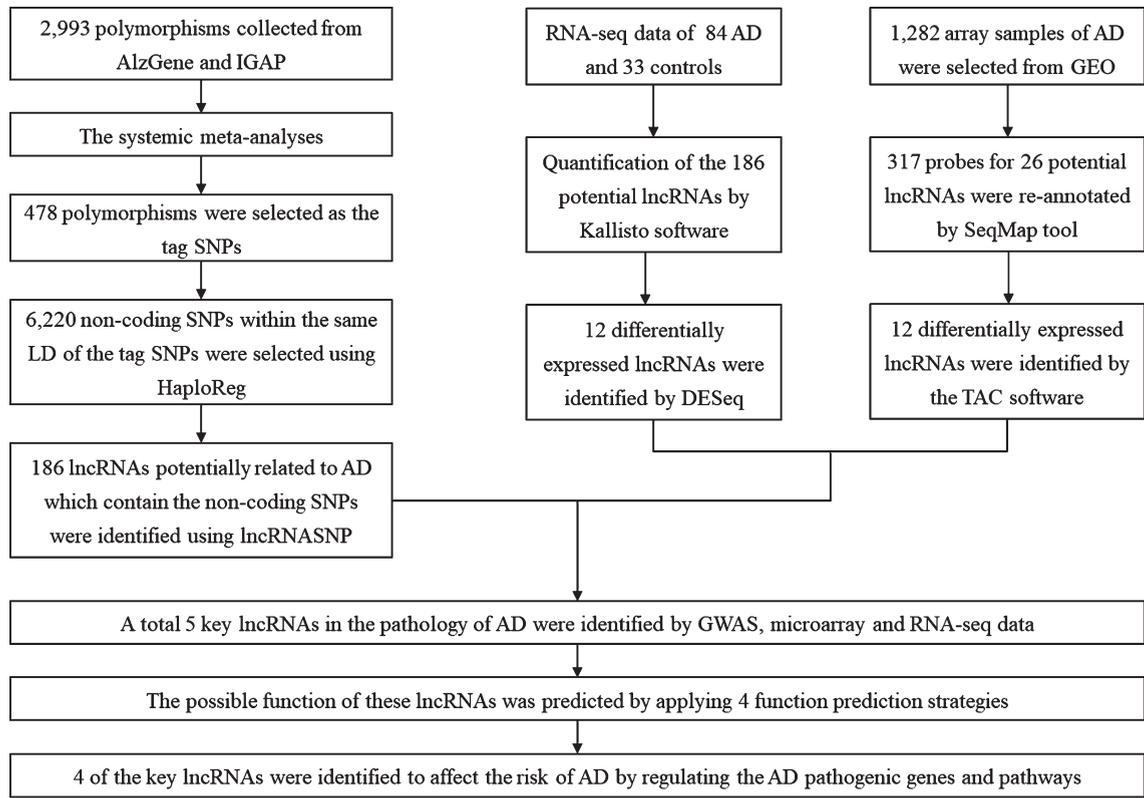


Fig. 1. The flow chart of the study design for identifying the key lncRNAs in pathology of AD and predicting their functions.

Identifying the lncRNAs potentially related to AD using non-coding SNPs

lncRNASNP database provides a comprehensive resource about lncRNAs and their related SNPs. It systematically integrates the information of SNPs and lncRNAs from dbSNP of the NCBI and LNCipedia database, respectively, and identifies 495,729 SNPs in > 30,000 human lncRNA transcripts [35]. In this study, we uploaded these non-coding SNPs in LD with the tag SNPs to lncRNASNP (v2.0) database (one SNP ID for each line), and mapped the genomic locations of these non-coding SNPs and lncRNAs. Then, we selected the lncRNAs which contain at least one of these non-coding SNP, and defined them as the potentially AD-related lncRNAs.

Validating and discovering the key lncRNAs in the pathology of AD

In order to further validate the association between AD and those lncRNAs identified in the previous step, the expression levels of the lncRNAs were compared between AD patients and healthy persons.

First, we selected all the AD-related Affymetrix Human expression profiling microarray datasets from GEO database by searching with the keyword: “Alzheimer’s Disease”. Then, we re-annotated the microarray probes to measure expression levels of the lncRNAs potentially related to AD according to the strategy in a previous study [53]. Briefly, we first downloaded the reference sequences of these potentially AD-related lncRNAs in FASTA format from NONCODE database [12]. Second, probe sets of the microarrays were aligned to the lncRNA sequences using SeqMap tool [54, 55], and the lncRNA-specific probe sets were obtained which contain at least four probes uniquely mapped to the lncRNA sequences without mismatch. Finally, quality controls for the microarrays were performed using Expression Console software, and the gene differential expression analysis between AD brains and controls were conducted using Affymetrix Transcriptome Analysis Console software. The expression difference threshold was set as fold-change (FC) ≥ 1.5 and p -values less than 0.05 after multiple testing according to previous studies [56–59].

If the samples of original microarray studies are from same brain regions and ethnicities, and the results of lncRNA expression patterns are inconsistent, a meta-analysis strategy was further performed to eliminate inconsistency [60–62]. Briefly, the meta-*FC* values were calculated using a linear amalgamation method, which combined the *FC* values by giving weights to the effect size in each study. The meta *p*-values were calculated using a Fisher's method based on the χ^2 distribution. In addition, Given that the microarray samples are mainly from European ancestry ethnicity, we further used a RNA-seq dataset of European ancestry ethnicity (GSE95587 [63]) to perform an lncRNA differential expression analysis and verify the results of microarray analysis. First, we downloaded the sequence data of GSE95587 (including 84 AD cases and 33 controls) from the NCBI Sequence Read Archive (SRA) database and converted them into FASTQ files using the SRA Toolkit software. Then, quantification of the lncRNA transcripts was performed by mapping the RNA-seq reads to their corresponding reference sequences and calculating the transcript per million (TPM) values using the Kallisto software, which is a fast and highly accurate tool to quantify transcript abundance from large-scale RNA-seq data using a k-mer lookup (instead of the traditional alignment step) [64]. Finally, we used the R package 'DESeq' to identify the differential expressed lncRNAs between AD and control according to the threshold of $FC \geq 1.5$ and p -value < 0.05 [65]. The p -values of differential expression analysis are corrected by the multiple testing (Benjamini–Hochberg method).

Inferring the function of the key lncRNAs in the pathology of AD

To explore the functions of those key lncRNAs identified in the previous step, four strategies (including genome mapping, eQTL, differential co-expression, and GSEA analysis) were applied to assess the regulations of lncRNAs to the known AD pathogenic genes both in cis and trans levels, and discover the biological pathways associated AD. Recent studies showed that many lncRNAs can effect on the expression of protein-coding genes located closely to them in genome (particularly within 2 kb), and these lncRNAs and protein-coding genes tend to be associated with the same diseases [66, 67]. So, one of the strategies to infer the functions of the key lncRNAs in the pathology of AD are to test whether some of the known AD pathogenic genes

exist within 2 kb of these lncRNAs or not. For this purpose, we first loaded all the data of the LNCipedia database into the University of California Santa Cruz (UCSC) Genome Browser [68, 69]. LNCipedia is a comprehensive database of human lncRNAs which contains the information about location, sequence, and structure of the currently known human lncRNAs [70]. Then, we used the UCSC Genome Browser to locate both these identified lncRNAs and known AD pathogenic genes in human genome. Here, we used GRCh38/hg38 as human genome references.

Other studies reported that the expression of lncRNAs can be affected by the SNPs located in them, and about 25% of these SNPs also influence the expression of protein-coding genes neighboring the lncRNAs [35], which implied that part of the cis-action of lncRNAs to protein-coding genes may be associated the SNPs located in them. To further explore the functions of those key lncRNAs in the pathology of AD, cis eQTL analysis was conducted to test whether the SNPs located in these lncRNAs affect the expression of known AD pathogenic genes or not by Braineac database [71] and Genotype-Tissue Expression (GTEx) project [72] ($p < 0.01$). Braineac and GTEx analyzed the microarray datasets of 10 (Cerebellar Cortex, Frontal Cortex, Hippocampus, Medulla, Occipital Cortex, Putamen, Substantia Nigra, Temporal Cortex, Thalamus, and Intralobular White Matter) and 13 (Amygdala, Anterior Cingulate, Caudate, Cerebellum, Cortex, Frontal Cortex, Hippocampus, Hypothalamus, Nucleus Accumbens, Putamen, Spinal Cord, and Substantia Nigra) brain regions from 131 and 175 postmortem donors using the 1000 Genomes Project as reference, respectively. More detailed information is described in the original publications [71, 72]. First, we selected all the SNPs located in these identified lncRNAs using the lncRNA-SNP database. Then, we uploaded these SNPs into the Braineac and GTEx databases, respectively, and tested whether these SNPs affected the expression of known AD pathogenic genes or not. The p -values of eQTL are corrected by multiple testing (Benjamini–Hochberg method).

Besides this cis action, it is confirmed that lncRNAs can also regulate the protein-coding genes in trans level [73]. Therefore, to further explore the functions of the key lncRNAs in the pathology of AD, the GEO data were used to conduct the differential co-expression analysis. Based on the strategy of measuring lncRNA expression level as described in the previous step, we used the same samples to identify differential expression of the protein-coding

genes between the AD patients and healthy individuals ($FC \geq 1.5$ and $p < 0.05$). Then, correlation coefficients (R) and significance level between the differential expression genes and the lncRNAs were calculated, and the co-expressed genes were selected according to the threshold ($|R| \geq 0.9$ and $p < 0.001$). If there are known AD pathogenic genes co-expressed with those lncRNAs, the functions of the corresponding lncRNA in AD are likely associated with it. Finally, besides those known pathogenic genes, other genes co-expressed with key lncRNAs in the co-expression networks can be used to identify the biological pathways associated AD. Therefore, we selected the protein-coding genes co-expressed with each of these identified lncRNAs, respectively, and used the R package ‘clusterProfiler’ to conduct the GO enrichment analysis [74]. The enrichment analysis is based on the classical Hypergeometric distribution test, and the adjusted p -values are calculated by the multiple testing (Benjamini–Hochberg method). The cutoff criterion for the enrichment analysis is set to adjusted p -values less than 0.01.

RESULTS AND DISCUSSION

Identifying the lncRNAs potentially related to AD using non-coding SNPs

We used multiple tools in the three-stage analysis to identify lncRNAs potentially related to AD. In stage 1, a total of 478 tag SNPs were selected from the European, African, and Asian ancestry ethnicities, which include 458 SNPs from AlzGene according to the results of the meta-analyses of GWAS studies and 20 SNPs from IGAP. We found that most of the tag SNPs’ allele frequencies are not significantly different between ethnicities (Supplementary Tables 1 and 2). In stage 2, we used HaploReg (v4) tool and discovered 6,220 non-coding SNPs within the LD of those 478 tag SNPs ($r^2 \geq 0.8$), and of which 4,985 non-coding SNPs are from European ancestry ethnicity, 39 are from African, and 1,472 are from Asian (Supplementary Table 3). In stage 3, by the lncRNASNP, we found that 186 lncRNAs contain at least one of the 6,220 non-coding SNPs, and defined them as potential AD-related lncRNAs (Supplementary Table 4).

Validating and discovering the key lncRNAs in the pathology of AD

By the re-annotation of microarray probes for the lncRNAs potentially related to AD, we found that

a total 317 probes can represent 26 corresponding lncRNA sequences. These probes are from five types of Affymetrix microarray, i.e., Human Genome U133A, Human Genome U133B, Human Genome U133 Plus 2.0, Human Gene 1.0 ST, and Human Gene 1.1 ST microarray (Supplementary Table 5). Then, after removing the unsuitable microarrays (the original studies are not designed by the case-control method or the ethnicity and diagnosis of the samples are unclear), e.g., Possible AD), we found a total 1,282 samples in 5 microarray datasets (GSE84422 [75], GSE43326 [76], GSE5281 [77], GSE36980 [78], GSE39420 [79]) from GEO database. Particularly, the 1,282 samples include 563 AD cases and 330 controls from European ancestry ethnicity, 134 AD cases and 56 controls from African ancestry ethnicity, and 95 AD cases and 104 controls from Asian ancestry ethnicity. Base on differential expression analysis and microarray meta-analysis between AD patients and normal people of the 1,282 samples, 12 lncRNAs were identified significantly differentially expressed between AD cases and controls ($FC \geq 1.5$ and $p < 0.05$) (Table 1).

To further validate the association between AD and the potential lncRNAs, we used the Kallisto software to quantify the transcript abundance (TPM values) of these lncRNAs from an RNA-seq dataset (including 84 AD cases and 33 controls). Then, after the differential expression analysis of these lncRNAs (based on the TPM values) by the R package ‘DESeq’, 12 lncRNAs were found significantly differentially expressed between AD cases and controls ($FC \geq 1.5$ and $p < 0.05$) (Table 2), and 5 of them are overlapped with findings of microarray analysis. We therefore defined these 5 lncRNAs (NONHSAT152299.1, NONHSAT016928.2, NONHSAT157359.1, NONHSAT160355.1, and NONHSAT018519.2) as the key lncRNAs in the pathology of AD.

In particular, we found that the expression of lncRNA NONHSAT018519.2 in European AD patients is significantly decreased compared with healthy persons in Amygdala, Nucleus Accumbens, and Primary Visual Cortex tissues ($FC = -1.90, -2.32$ and -9.42 , respectively) (Fig. 2a-c). Interestingly, according to results of the GTEx project [72], for the healthy European ancestry individuals, NONHSAT018519.2 is significantly highly expressed in brain tissues compared with 40 other human tissues (Fig. 2d). All these phenomena implied that the highly expressed NONHSAT018519.2 in brain plays an important role for healthy persons,

Table 1
The information of the differentially expressed lncRNAs identified by microarray

Tissue	Ethnicity	LncRNA (NONCODE ID)	Position (hg38)	Country	GEO Series Accession Number	Platform	Probe Set ID	Fold Change (FC)	Combined FC	<i>p</i>
Amygdala	European ancestry	NONHSAT018519.2	chr11:27657489-27677768[+]	USA	GSE84422	HG-U133.Plus.2	206382_s.at	-1.90	-1.90	0.0473
Amygdala	African ancestry	NONHSAT157359.1	chr10:95663881-95682804[-]	USA	GSE84422	HG-U133.Plus.2	212121_at	1.53	1.53	0.0037
Dorsolateral Prefrontal Cortex	European ancestry	NONHSAT015489.2	chr10:88991421-88992975[-]	USA	GSE84422	HG-U133B	237522_at	-2.52	-2.52	0.0148
Dorsolateral Prefrontal Cortex	European ancestry	NONHSAT163333.1	chr12:13540247-13549972[-]	USA	GSE84422	HG-U133B	239095_at	-1.76	-1.76	0.0071
Frontal Pole	European ancestry	NONHSAT152299.1	chr1:150996840-150998191[-]	USA	GSE84422	HG-U133A	221856_s.at	1.68	1.68	0.0052
Hippocampus	African ancestry	NONHSAT016928.2	chr10:126012386-126014767[-]	USA	GSE84422	HG-U133B	226777_at	13.68	13.68	0.0284
Inferior Frontal Gyrus	African ancestry	NONHSAT079080.2	chr20:23631825-23632316[-]	USA	GSE84422	HG-U133B	240177_at	12.59	12.59	0.0204
Inferior Temporal Gyrus	African ancestry	NONHSAT015489.2	chr10:88991421-88992975[-]	USA	GSE84422	HG-U133B	237522_at	-1.78	-1.78	0.0168
iPSC-derived neural cells	Asian ancestry	NONHSAT026096.2	chr12:6510221-6510551[+]	Japan	GSE43326	HuGene-1.0	7953383	-2.42	-2.42	0.0257
Middle Temporal Gyrus	European ancestry	NONHSAT040199.2	chr14:103696352-103697163[+]	USA	GSE5281	HG-U133.Plus.2	226579_at	2.70	2.70	0.0223
Middle Temporal Gyrus	European ancestry	NONHSAT152299.1	chr1:150996840-150998191[-]	USA	GSE5281	HG-U133.Plus.2	221856_s.at	8.78	1.90	0.0009
Nucleus Accumbens	European ancestry	NONHSAT018519.2	chr11:27657489-27677768[+]	USA	GSE84422	HG-U133A	221856_s.at	1.25	-2.32	0.0301
Parahippocampal Gyrus	European ancestry	NONHSAT015489.2	chr10:88991421-88992975[-]	USA	GSE84422	HG-U133B	237522_at	2.34	2.34	0.0045
Primary Visual Cortex	European ancestry	NONHSAT018519.2	chr11:27657489-27677768[+]	USA	GSE5281	HG-U133.Plus.2	206382_s.at	-9.42	-9.42	0.0024
Primary Visual Cortex	European ancestry	NONHSAT040199.2	chr14:103696352-103697163[+]	USA	GSE5281	HG-U133.Plus.2	226579_at	5.36	5.36	0.0013
Primary Visual Cortex	European ancestry	NONHSAT144872.2	chr17:1725747-1738585[-]	USA	GSE5281	HG-U133.Plus.2	226738_at	3.24	3.24	0.0040
Primary Visual Cortex	European ancestry	NONHSAT163333.1	chr12:13540247-13549972[-]	USA	GSE5281	HG-U133.Plus.2	239095_at	-5.35	-5.35	0.0003
Putamen	African ancestry	NONHSAT016928.2	chr10:126012386-126014767[-]	USA	GSE84422	HG-U133B	226777_at	1.60	1.60	0.0079
Superior Parietal Lobule	African ancestry	NONHSAT108933.2	chr6:32825414-32828345[-]	USA	GSE84422	HG-U133A	204770_at	-1.36	1.84	0.0250
Superior Temporal Gyrus	European ancestry	NONHSAT079080.2	chr20:23631825-23632316[-]	USA	GSE84422	HG-U133B	225973_at	2.09	1.76	0.0114
Temporal Cortex	Asian ancestry	NONHSAT160355.1	chr11:60172029-60183061[-]	Japan	GSE36980	HuGene-1.0	7948455	1.64	1.64	0.0191
Thalamus	European ancestry	NONHSAT152299.1	chr1:150996840-150998191[-]	Spain	GSE39420	HuGene-1.1	7919872	2.06	2.06	0.0074
Thalamus	European ancestry	NONHSAT160355.1	chr11:60172029-60183061[-]	Spain	GSE39420	HuGene-1.1	7948455	2.85	2.85	0.0023

Table 2
The information of the differentially expressed lncRNAs identified by RNA-seq

Tissue	Ethnicity	lncRNA (NONCODE ID)	Position (hg38)	Country	GEO Series Accession Number	Platform	Fold Change (FC)	P
Fusiform Gyrus	European ancestry	NONHSAT016928.2	chr10:126012390-126012642[-]	USA	GSE95587	Illumina HiSeq 2500	1.55	0.04607
Fusiform Gyrus	European ancestry	NONHSAT018494.2	chr11:27506848-27698174[+]	USA	GSE95587	Illumina HiSeq 2500	-2.81	0.00204
Fusiform Gyrus	European ancestry	NONHSAT018519.2	chr11:27657489-2767768[+]	USA	GSE95587	Illumina HiSeq 2500	-2.79	0.00004
Fusiform Gyrus	European ancestry	NONHSAT040198.2	chr14:103695415-103697123[-]	USA	GSE95587	Illumina HiSeq 2500	1.58	0.01613
Fusiform Gyrus	European ancestry	NONHSAT108832.2	chr6:31764135-31765588[-]	USA	GSE95587	Illumina HiSeq 2500	-1.96	0.02013
Fusiform Gyrus	European ancestry	NONHSAT113034.2	chr6:47475413-47477692[-]	USA	GSE95587	Illumina HiSeq 2500	-3.05	0.00077
Fusiform Gyrus	European ancestry	NONHSAT152299.1	chr1:150996840-150998191[-]	USA	GSE95587	Illumina HiSeq 2500	1.56	0.02409
Fusiform Gyrus	European ancestry	NONHSAT157359.1	chr10:95663881-95682804[-]	USA	GSE95587	Illumina HiSeq 2500	-1.64	0.01018
Fusiform Gyrus	European ancestry	NONHSAT160355.1	chr11:60172029-60183061[-]	USA	GSE95587	Illumina HiSeq 2500	2.26	0.00204
Fusiform Gyrus	European ancestry	NONHSAT185498.1	chr2:127104196-127104528[-]	USA	GSE95587	Illumina HiSeq 2500	2.65	0.02736
Fusiform Gyrus	European ancestry	NONHSAT200278.1	chr4:108810720-109302367[-]	USA	GSE95587	Illumina HiSeq 2500	-2.77	0.00407
Fusiform Gyrus	European ancestry	NONHSAT211955.1	chr7:37650300-37802383[+]	USA	GSE95587	Illumina HiSeq 2500	-1.51	0.00862

and the decreased expression of NONHSAT018519.2 may be associated with the pathology of AD.

Genome mapping analysis using UCSC Genome Browser and LNCipedia

To infer the function of those lncRNAs identified as key to AD pathology, both the lncRNAs and known AD pathogenic genes were located in genome using UCSC Genome Browser and LNCipedia. Finally, a total of two pathogenic genes with known functions, BDNF and ADAM12, were located within 2 kb of two key lncRNAs, NONHSAT018519.2 and NONHSAT016928.2, respectively. Moreover, the genes, BDNF and ADAM12, are the only protein-coding gene within 2 kb of their corresponding lncRNAs. The detailed information was described in Fig. 3.

Based on the results of previous studies, BDNF (brain-derived neurotrophic factor) has a neuroprotective effect against AD by facilitating neurotransmitter release and reducing A β peptide toxicity, and shows a lower expression in AD brains [80, 81]. Similarly, the expression of ADAM12 (a proteolytic enzyme) is reduced in AD brain, and it provides a broad neuroprotection by mediating neurotoxic effect of A β peptide [82]. In view of the downregulated ($FC = -1.90, -2.32, -9.42$ and -2.79), in the Amygdala, Nucleus Accumbens, Primary Visual Cortex, and Fusiform Gyrus, respectively) and upregulated expression ($FC = 13.68, 1.60$ and 1.55 in Hippocampus, Putamen, and Fusiform Gyrus, respectively) of NONHSAT018519.2 and NONHSAT016928.2 in AD patients compared with the healthy persons, respectively (Tables 1 and 2), the function of them in AD may be associated with regulation of the expression of BDNF and ADAM12 in the opposite directions.

The eQTL analysis based on the data of the Braineac and GTEx projects

The Braineac and GTEx projects involve 10 and 13 brain regions, respectively, and 6 of which are consistent with this study (Amygdala, Hippocampus, Nucleus Accumbens, Putamen, Temporal Cortex, and Thalamus). To further explore the functions of the key lncRNAs in the pathology of AD, cis eQTL analysis was performed to test whether the SNPs located in the lncRNAs affect the expression of the known AD pathogenic genes in the 6 brain regions. As shown in Fig. 4a, the SNP rs7232 ($p = 0.0011$) and rs12453

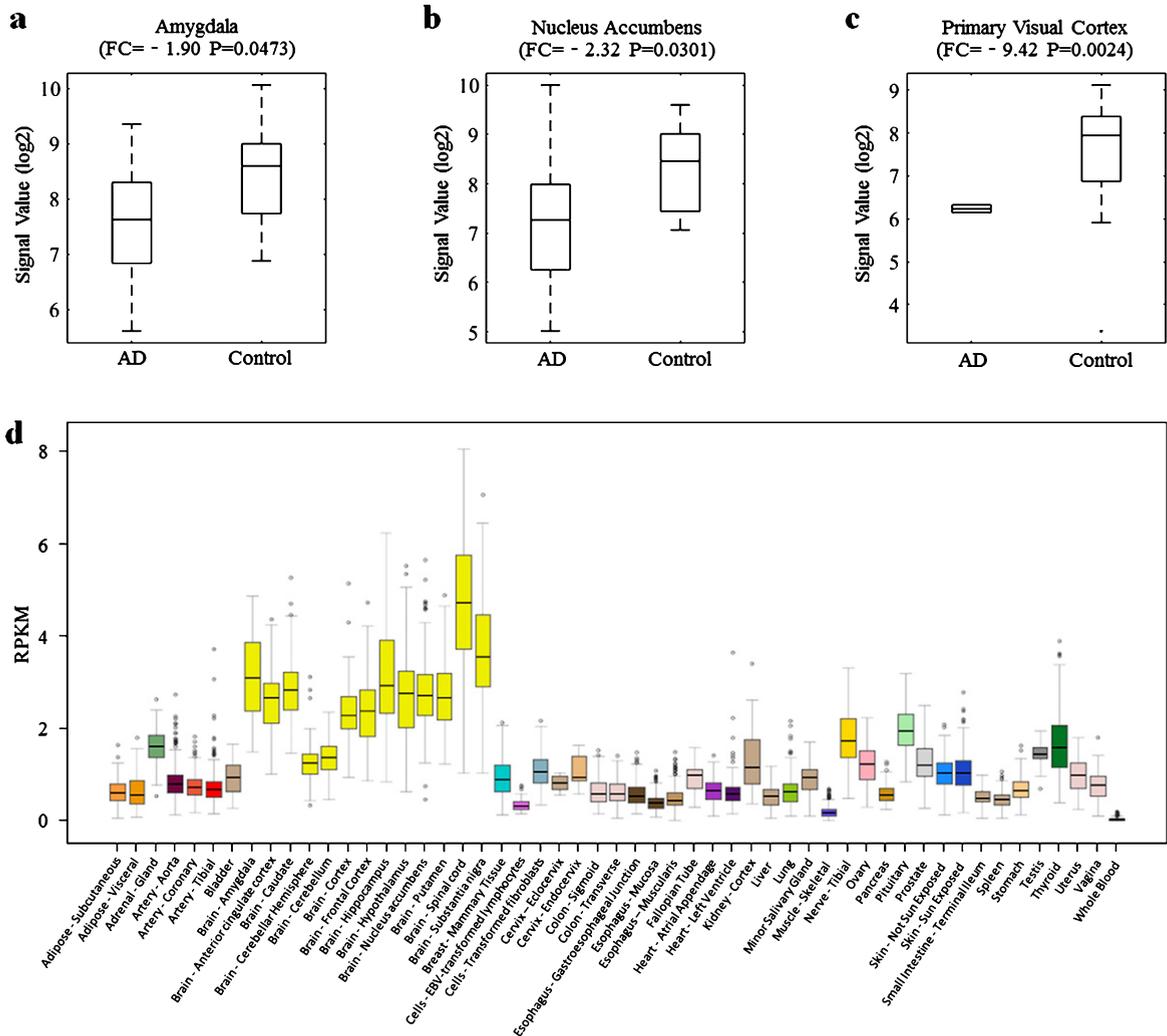


Fig. 2. The expression levels of lncRNA NONHSAT018519.2 in AD patients and healthy persons (a-c). To AD patients, NONHSAT018519.2 expression significantly decreases compared with healthy persons in the Amygdala (a), Nucleus Accumbens (b), and Primary Visual Cortex (c), and the differentially expressed NONHSAT018519.2 is only identified in these three brain tissues. d) Results of Genotype-Tissue Expression (GTEx) project show that NONHSAT018519.2 is highly expressed in brain tissues of healthy individuals compared with other 40 human tissues. All the samples are collected from European ancestry populations.

($p=0.00057$) in lncRNA NONHSAT160355.1 significantly downregulate the expression of a known AD pathogenic gene TCN1 in Temporal Cortex, which participates in the regulation of homocysteine in brain to increase risk of AD [83]. Then, as seen in Fig. 4b and c, rs12453 is in LD with the SNP rs983392 ($r^2=0.821568$), which is identified to be associated with a reduced risk of AD ($OR=0.9$; $95\% CI=0.87-0.92$; $p=1.6E-16$) [21]. Moreover, the expression of NONHSAT160355.1 in AD patients is significantly increased compared with healthy persons ($FC=1.64$, 2.85 and 2.26 in Temporal Cortex, Thalamus, and Fusiform Gyrus,

respectively) (Tables 1 and 2). Therefore, the possible functions of NONHSAT160355.1 in the pathology of AD can be inferred as promoting the expression of TCN1.

The differential co-expression analysis and enrichment analysis

To further explore the trans regulation of 5 key lncRNAs and their functions acting upon AD pathways, we performed the differential co-expression analysis and enrichment analysis. After the differential co-expression analysis, a total 555

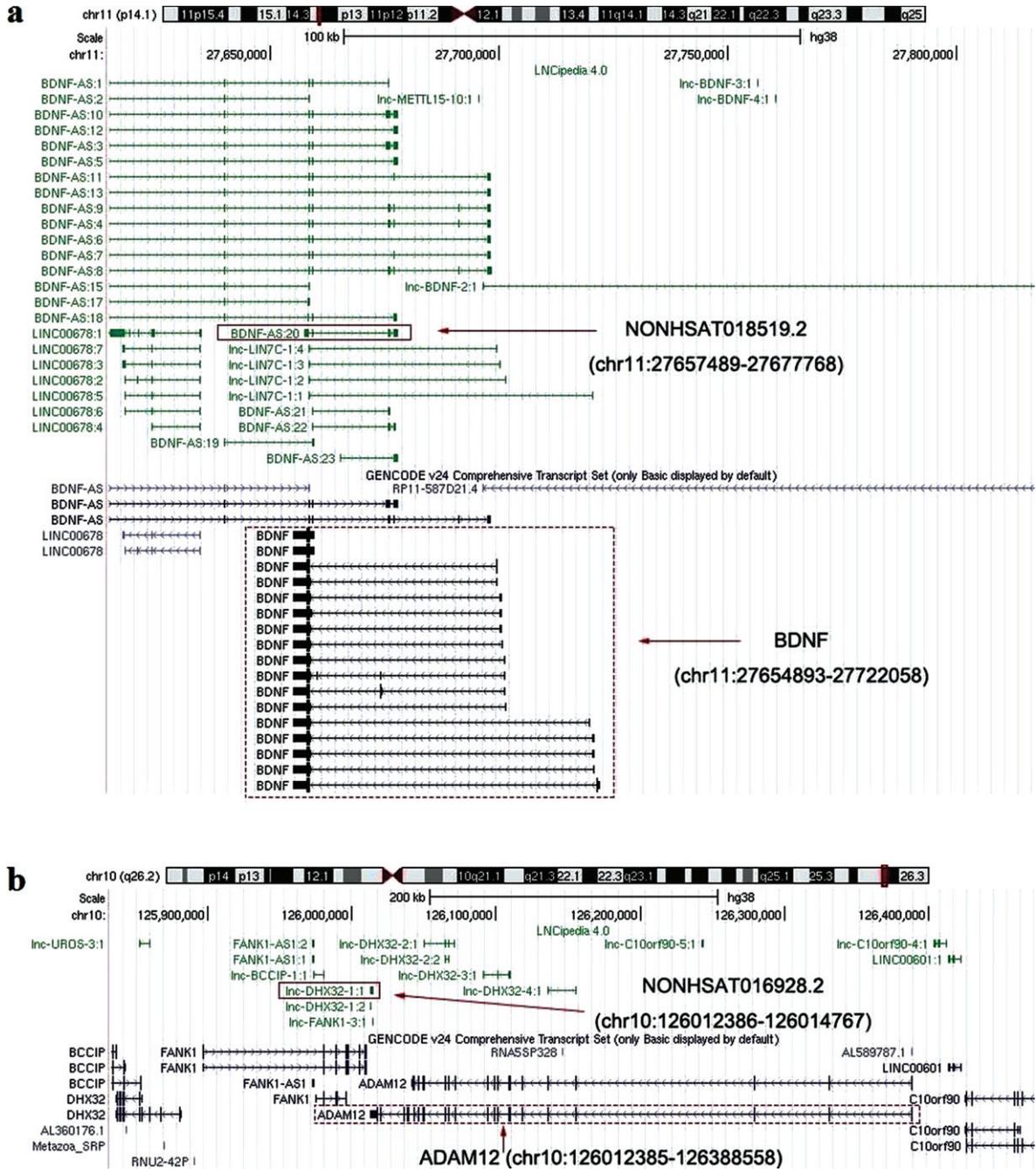


Fig. 3. The location of lncRNAs and protein-coding genes in human genome. a) The locations (GRCh38/hg38) of lncRNA NONHSAT018519.2 and the protein-coding gene BDNF near it. b) The locations of lncRNA NONHSAT016928.2 and the protein-coding gene ADAM12 near it. The green lines with a solid red box were lncRNAs according to LNCipedia database. The LNCipedia ID of these lncRNAs was in one-to-one correspondence with NONCODE ID. The black lines with a dashed box were the protein-coding genes.

protein-coding genes with significantly differential expression between AD cases and controls were discovered to co-express with 3 key lncRNAs, NONHSAT152299.1, NONHSAT157359.1, and NONHSAT160355.1 ($|R| \geq$ and $p < 0.001$).

Detail information is provided in Supplementary Table 6. Among these protein-coding genes, we found that four known AD susceptibility genes (C4A, C4B, TCF4, GRIP1) co-expressed with the lncRNA NONHSAT152299.1 (Table 3). Particularly,

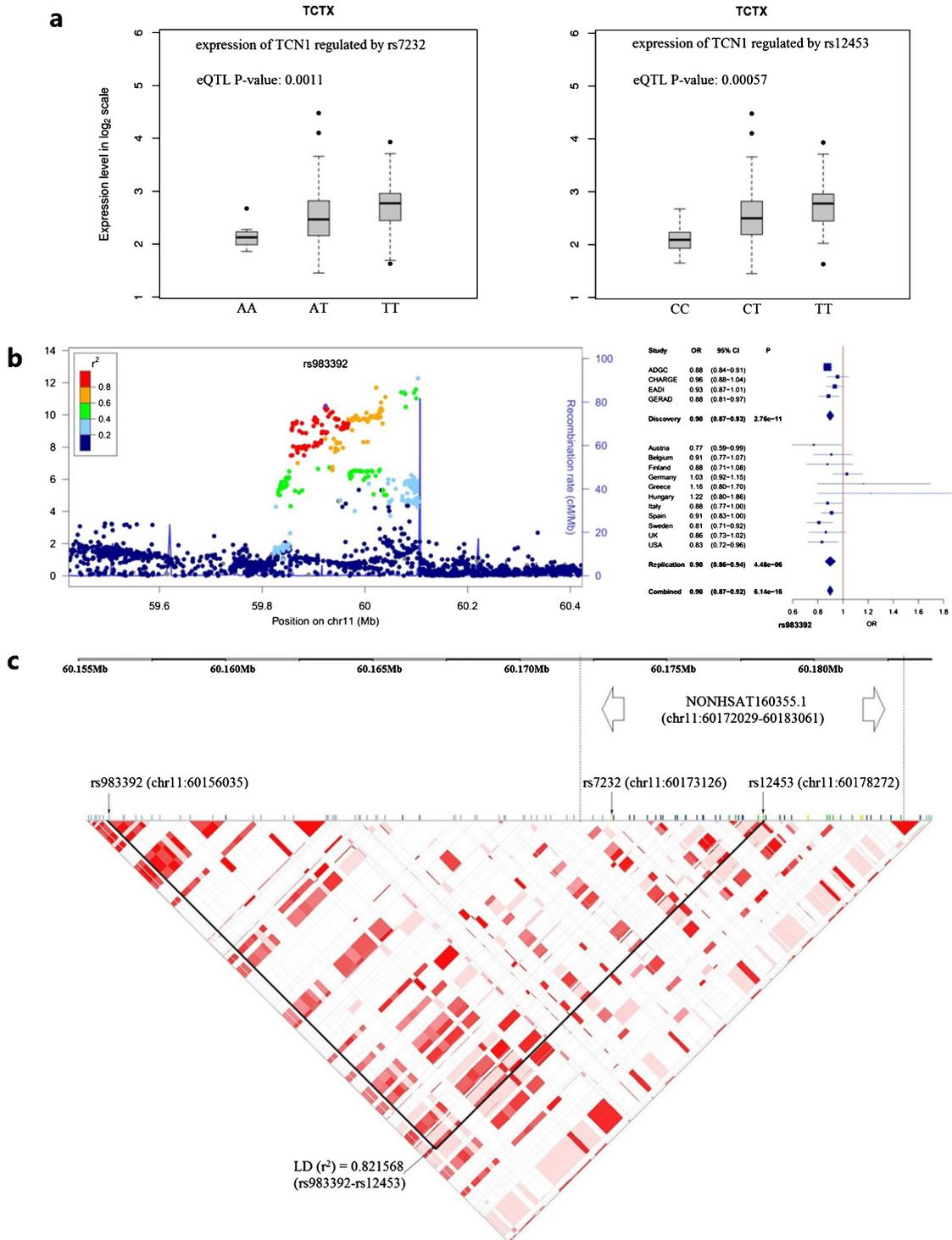


Fig. 4. The rs7232 and rs12453 in NONHSAT160355.1 show an eQTL effect on FADS1 expression. a) The SNP rs7232 (chr11:60173126 T>A) and rs12453 (chr11:60178272 T>C) affect the expression of gene TCN1 (chr11:59852808-59866568) by an eQTL pattern in the Temporal Cortex (TCTX). b) As the result of Lambert et al.'s study [21], rs983392 (chr11:60156035 A>G) is a susceptibility locus associated with AD, and it significantly decreases the risk of AD ($OR=0.9$; $95\% CI=0.87-0.92$). c) The rs7232 and rs12453 are located in the region of lncRNA NONHSAT160355.1, and the rs12453 is in the LD with the SNP rs983392 ($r^2=0.821568$) based on 1000 Genomes Project. All the location information is according to the human reference genome hg38.

Table 3
Co-expression between the known AD pathogenic genes and the lncRNA NONHSAT152299.1

lncRNA (NONCODE ID)	Gene Symbol	Gene ID		<i>p</i>	Tissue	Ethnicity
NONHSAT152299.1	C4A	720	0.92999061	1.83E-07	Middle Temporal Gyrus	European ancestry
NONHSAT152299.1	C4B	721	0.92999061	1.83E-07	Middle Temporal Gyrus	European ancestry
NONHSAT152299.1	TCF4	6925	0.93547919	1.05E-07	Middle Temporal Gyrus	European ancestry
NONHSAT152299.1	GRIP1	10499	-0.90025420	2.73E-08	Thalamus	European ancestry

the number of copies for both C4A and C4B are increased in AD patients, which significantly affect the expression of complement component 4 protein (a key complement factor in A β peptide metabolism) [84]. TCF4 functions as a transcriptional repressor of BACE1 gene by binding BACE1 promoter, and further represses Wnt/ β -catenin stimulation in AD [85]. The GRIP1 affects the functions of PFA1 and PFA2 in recognizing A β monomers, protofibrils, and fibrils and the structures of their antigen binding fragments in AD [86]. Moreover, the polymorphisms of C4A, C4B, TCF4, and GRIP1 are proved to be significantly associated with the pathology of AD. In addition, these four susceptibility genes locate in different chromosomes than the lncRNAs co-expressed with them, and they may be regulated by lncRNAs in trans level.

As the results of the gene set enrichment analysis, we found that 11 enriched GO terms related to the pathology of AD are significantly associated with the lncRNA NONHSAT160355.1 and NONHSAT152299.1 ($p < 0.01$) (Table 4). Particularly, three enriched GO terms associated with the lncRNA NONHSAT160355.1 are involved in the immune response. As reported, immune response participates in the regulation of metabolic processes related to AD [87]. In a recent study, the same pathways (GO:0019886, GO:0002495, and GO:0002504) are also identified to be associated the pathology of AD by an AD-related variants enrichment in GO database [88]. Interestingly, immune responses associated with NONHSAT160355.1 are related to antigen processing and presentation via the major histocompatibility complex (MHC) class II, which is a cell surface protein to bind and present antigen peptide fragments to T cell and markedly high expressed in AD brain [89]. As a result of eQTL analysis in a previous step, NONHSAT160355.1 was found to significantly affect the expression of a known AD pathogenic gene (TCN1) in the Temporal Cortex (Fig. 4). As reported, the TCN1 participates in the regulation of homocysteine in brain to increase risk of AD, and homocysteine can oxidize the HLA antigens to affect the antigen processing and presentation

related to MHC class II [83, 90]. Combined with the higher expression of the NONHSAT160355.1 in AD patients compared with controls ($FC = 1.64, 2.85$ and 2.26 in Temporal Cortex, Thalamus, and Fusiform Gyrus, respectively) (Tables 1 and 2), the lncRNA may increase the risk of AD by promoting the expression of TCN1 to affect the antigen processing and presentation related to MHC class II.

The remaining 8 enriched GO terms associated with the lncRNA NONHSAT152299.1 are involved the development of nervous system. Moreover, GRIP1 is negatively co-expressed with NONHSAT152299.1 in Thalamus ($R = -0.9003$; $p = 2.73E - 08$) (Table 3), and it is demonstrated to play a critical role for the development of nervous system in brain [91–93]. Given that the higher expression of NONHSAT152299.1 in AD patients compared with controls ($FC = 2.06$ in Thalamus) (Table 1), the lncRNA may increase the risk of AD by inhibiting the expression of GRIP1 to affect the development of nervous system in brain.

Conclusion

Based on the GWAS, microarray, and RNA-seq data, this study identified a total of five key lncRNAs in the pathology of AD, which contain the SNPs significantly associated with AD and are differentially expressed between AD cases and controls. By combining the results of function prediction by genome mapping, eQTL, differential co-expression, and GSEA analysis, four of the key lncRNAs were identified to increase or decrease the risk of AD by regulating disease-related pathogenic genes and pathways. Particularly, the lncRNA NONHSAT160355.1, NONHSAT152299.1, and NONHSAT016928.2 were identified as accelerating the risk of AD by 1) promoting TCN1 expression to affect the antigen processing and presentation related to MHC class II, 2) inhibiting GRIP1 expression to affect the development of nerve system in brain, and 3) inhibiting ADAM12 expression to aggravate neurotoxic effect of A β peptide, respectively. The NONHSAT018519.2 demonstrates a neuroprotective

Table 4
GO enrichment results of the protein-encoding genes co-expressed with lncRNAs

lncRNA (NONCODE ID)	Accession	p	Description	Tissue	Ethnicity
NONHSAT160355.1	GO:0019886	4.80E-04	Antigen processing and presentation of exogenous peptide antigen via MHC II	Temporal Cortex	Asian ancestry
NONHSAT160355.1	GO:0002495	4.80E-04	Antigen processing and presentation of peptide antigen via MHC class II	Temporal Cortex	Asian ancestry
NONHSAT160355.1	GO:0002504	4.80E-04	Antigen processing and presentation of peptide or polysaccharide antigen via MHC II	Temporal Cortex	Asian ancestry
NONHSAT152299.1	GO:0071560	7.00E-03	Cellular response to transforming growth factor beta stimulus	Middle Temporal Gyrus	European ancestry
NONHSAT152299.1	GO:0071559	7.00E-03	Response to transforming growth factor beta	Middle Temporal Gyrus	European ancestry
NONHSAT152299.1	GO:0050885	7.25E-04	Neuromuscular process controlling balance	Thalamus	European ancestry
NONHSAT152299.1	GO:0048667	9.48E-04	Cell morphogenesis involved in neuron differentiation	Thalamus	European ancestry
NONHSAT152299.1	GO:0010975	1.23E-03	Regulation of neuron projection development	Thalamus	European ancestry
NONHSAT152299.1	GO:0050808	1.23E-03	Synapse organization	Thalamus	European ancestry
NONHSAT152299.1	GO:0061564	2.16E-03	Axon development	Thalamus	European ancestry
NONHSAT152299.1	GO:0051962	4.89E-03	Positive regulation of nervous system development	Thalamus	European ancestry

effect against AD by promoting BDNF expression to facilitate neurotransmitter release and reducing A β peptide toxicity. These findings can help to discover novel key lncRNAs in the pathology of AD and improve the understanding of the pathogenesis of AD.

ACKNOWLEDGMENTS

The National Natural Science Foundation of China (81872798), National Key Research and Development Program of China (2018YFC0910500), Innovation Project on Industrial Generic Key Technologies of Chongqing (cstc2015zdcy-ztx120003), and Fundamental Research Funds for the Central Universities (2018QNA7023, 10611CDJXZ238826, 2018CDQYSG0007, CDJZR14468801, CDJKXB14011).

Authors' disclosures available online (<https://www.j-alz.com/manuscript-disclosures/18-1051r2>).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/JAD-181051>.

REFERENCES

- [1] Burns A, Iliffe S (2009) Alzheimer's disease. *BMJ* **338**, b158.
- [2] Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**, 353-356.
- [3] Yang H, Qin C, Li YH, Tao L, Zhou J, Yu CY, Xu F, Chen Z, Zhu F, Chen YZ (2016) Therapeutic target database update 2016: Enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res* **44**, D1069-D1074.
- [4] Querfurth HW, LaFerla FM (2010) Alzheimer's disease. *N Engl J Med* **362**, 329-344.
- [5] Hebert LE, Weuve J, Scherr PA, Evans DA (2013) Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology* **80**, 1778-1783.
- [6] Li XX, Yin J, Tang J, Li Y, Yang Q, Xiao Z, Zhang R, Wang Y, Hong J, Tao L, Xue W, Zhu F (2018) Determining the balance between drug efficacy and safety by the network and biological system profile of its therapeutic target. *Front Pharmacol* **9**, 1245.
- [7] Zhu F, Han B, Kumar P, Liu X, Ma X, Wei X, Huang L, Guo Y, Han L, Zheng C, Chen Y (2010) Update of TTD: Therapeutic target database. *Nucleic Acids Res* **38**, D787-D791.
- [8] Brookmeyer R, Evans DA, Hebert L, Langa KM, Heeringa SG, Plassman BL, Kukull WA (2011) National estimates of the prevalence of Alzheimer's disease in the United States. *Alzheimers Dement* **7**, 61-73.

- [9] Vaudano E, Vannieuwenhuysse B, Van Der Geyten S, van der Lei J, Visser PJ, Streffer J, Ritchie C, McHale D, Lovestone S, Hofmann-Apitius M, Truyen L, Goldman M (2015) Boosting translational research on Alzheimer's disease in Europe: The Innovative Medicine Initiative AD research platform. *Alzheimers Dement* **11**, 1121-1122.
- [10] Zhu F, Han L, Zheng C, Xie B, Tammi MT, Yang S, Wei Y, Chen Y (2009) What are next generation innovative therapeutic targets? Clues from genetic, structural, physicochemical, and systems profiles of successful targets. *J Pharmacol Exp Ther* **330**, 304-315.
- [11] Ulitsky I (2016) Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* **17**, 601-614.
- [12] Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, Chen R (2016) NONCODE 2016: An informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res* **44**, D203-D208.
- [13] Ulitsky I, Bartel DP (2013) lincRNAs: Genomics, evolution, and mechanisms. *Cell* **154**, 26-46.
- [14] Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* **136**, 629-641.
- [15] Patrick E, Rajagopal S, Wong HA, McCabe C, Xu J, Tang A, Imboywa SH, Schneider JA, Pochet N, Krichevsky AM, Chibnik LB, Bennett DA, De Jager PL (2017) Dissecting the role of non-coding RNAs in the accumulation of amyloid and tau neuropathologies in Alzheimer's disease. *Mol Neurodegener* **12**, 51.
- [16] Wang P, Zhang X, Fu T, Li S, Li B, Xue W, Yao X, Chen Y, Zhu F (2017) Differentiating physicochemical properties between addictive and nonaddictive ADHD drugs revealed by molecular dynamics simulation studies. *ACS Chem Neurosci* **8**, 1416-1428.
- [17] Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G 3rd, Kenny PJ, Wahlestedt C (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* **14**, 723-730.
- [18] Ciarlo E, Massone S, Penna I, Nizzari M, Gigoni A, Dieci G, Russo C, Florio T, Cancedda R, Pagano A (2013) An intronic ncRNA-dependent regulation of SORL1 expression affecting A β formation is upregulated in post-mortem Alzheimer's disease brain samples. *Dis Model Mech* **6**, 424-433.
- [19] Han ZJ, Xue WW, Tao L, Zhu F (2018) Identification of novel immune-relevant drug target genes for Alzheimer's Disease by combining ontology inference with network analysis. *CNS Neurosci Ther* **24**, 1253-1263.
- [20] Mus E, Hof PR, Tiedge H (2007) Dendritic BC200 RNA in aging and in Alzheimer's disease. *Proc Natl Acad Sci U S A* **104**, 10679-10684.
- [21] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thornton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, Lin CF, Gerrish A, Schmidt H, Kunkle B, Dunstan ML, Ruiz A, Bioreau MT, Choi SH, Reitz C, Pasquier F, Cruchaga C, Craig D, Amin N, Berr C, Lopez OL, De Jager PL, Deramecourt V, Johnston JA, Evans D, Lovestone S, Letenneur L, Morón FJ, Rubinsztein DC, Eiriksdottir G, Sleegers K, Goate AM, Fiévet N, Huentelman MW, Gill M, Brown K, Kamboh MI, Keller L, Barberger-Gateau P, McGuinness B, Larson EB, Green R, Myers AJ, Dufouil C, Todd S, Wallon D, Love S, Rogava E, Gallacher J, St George-Hyslop P, Clarimon J, Lleo A, Bayer A, Tsuang DW, Yu L, Tsolaki M, Bossú P, Spalletta G, Proitsi P, Collinge J, Sorbi S, Sanchez-Garcia F, Fox NC, Hardy J, Deniz Naranjo MC, Bosco P, Clarke R, Brayne C, Galimberti D, Mancuso M, Matthews F; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology, Moebus S, Mecocci P, Del Zompo M, Maier W, Hampel H, Pilotto A, Bullido M, Panza F, Caffarra P, Nacmias B, Gilbert JR, Mayhaus M, Lannfelt L, Hakonarson H, Pichler S, Carrasquillo MM, Ingelsson M, Beekly D, Alvarez V, Zou F, Valladares O, Younkin SG, Coto E, Hamilton-Nelson KL, Gu W, Razquin C, Pastor P, Mateo I, Owen MJ, Faber KM, Jonsson PV, Combarros O, O'Donovan MC, Cantwell LB, Soininen H, Blacker D, Mead S, Mosley TH Jr, Bennett DA, Harris TB, Fratiglioni L, Holmes C, de Bruijn RF, Passmore P, Montine TJ, Bettens K, Rotter JJ, Brice A, Morgan K, Foroud TM, Kukull WA, Hannequin D, Powell JF, Nalls MA, Ritchie K, Lunetta KL, Kauwe JS, Boerwinkle E, Riemenschneider M, Boada M, Hiltunen M, Martin ER, Schmidt R, Rujescu D, Wang LS, Dartigues JF, Mayeux R, Tzourio C, Hofman A, Nöthen MM, Graff C, Psaty BM, Jones L, Haines JL, Holmans PA, Lathrop M, Pericak-Vance MA, Launer LJ, Farrer LA, van Duijn CM, Van Broeckhoven C, Moskvina V, Seshadri S, Williams J, Schellenberg GD, Amouyel P (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458.
- [22] Ayana R, Singh S, Pati S (2017) Decoding crucial lncRNAs implicated in neurogenesis and neurological disorders. *Stem Cells Dev* **26**, 541-553.
- [23] Zheng G, Yang F, Fu T, Tu G, Chen Y, Yao X, Xue W, Zhu F (2018) Computational characterization of the selective inhibition of human norepinephrine and serotonin transporters by an escitalopram scaffold. *Phys Chem Chem Phys* **20**, 29513-29527.
- [24] Xue W, Wang P, Tu G, Yang F, Zheng G, Li X, Li X, Chen Y, Yao X, Zhu F (2018) Computational identification of the binding mechanism of a triple reuptake inhibitor amitridine for the treatment of major depressive disorder. *Phys Chem Chem Phys* **20**, 6606-6616.
- [25] Zhu F, Li XX, Yang SY, Chen YZ (2018) Clinical success of drug targets prospectively predicted by in silico study. *Trends Pharmacol Sci* **39**, 229-231.
- [26] Zhu F, Shi Z, Qin C, Tao L, Liu X, Xu F, Zhang L, Song Y, Liu X, Zhang J, Han B, Zhang P, Chen Y (2012) Therapeutic target database update 2012: A resource for facilitating target-oriented drug discovery. *Nucleic Acids Res* **40**, D1128-D1136.
- [27] Zhang Z (2016) Long non-coding RNAs in Alzheimer's disease. *Curr Top Med Chem* **16**, 511-519.
- [28] Li C, Chen J, Zhang K, Feng B, Wang R, Chen L (2015) Progress and prospects of long noncoding RNAs (lncRNAs) in hepatocellular carcinoma. *Cell Physiol Biochem* **36**, 423-434.
- [29] Riquelme I, Ili C, Roa JC, Brebi P (2016) Long non-coding RNAs in gastric cancer: Mechanisms and potential applications. *Oncotarget* **1**, 1-1.
- [30] Yang F, Zheng G, Fu T, Li X, Tu G, Li YH, Yao X, Xue W, Zhu F (2018) Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H. *Phys Chem Chem Phys* **20**, 23873-23884.

- [31] Zhou X, Xu J (2015) Identification of Alzheimer's disease-associated long noncoding RNAs. *Neurobiol Aging* **36**, 2925-2931.
- [32] Wang LK, Chen XF, He DD, Li Y, Fu J (2017) Dissection of functional lncRNAs in Alzheimer's disease by construction and analysis of lncRNA-mRNA networks based on competitive endogenous RNAs. *Biochem Biophys Res Commun* **485**, 569-576.
- [33] Zhu F, Zheng CJ, Han LY, Xie B, Jia J, Liu X, Tammi MT, Yang SY, Wei YQ, Chen YZ (2008) Trends in the exploration of anticancer targets and strategies in enhancing the efficacy of drug targeting. *Curr Mol Pharmacol* **1**, 213-232.
- [34] Li B, Tang J, Yang Q, Li S, Cui X, Li Y, Chen Y, Xue W, Li X, Zhu F (2017) NORENA: Normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res* **45**, W162-W170.
- [35] Gong J, Liu W, Zhang J, Miao X, Guo AY (2015) lncRNASNP: A database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res* **43**, D181-D186.
- [36] Li YH, Yu CY, Li XX, Zhang P, Tang J, Yang Q, Fu T, Zhang X, Cui X, Tu G, Zhang Y, Li S, Yang F, Sun Q, Qin C, Zeng X, Chen Z, Chen YZ, Zhu F (2018) Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res* **46**, D1121-D1127.
- [37] Wang P, Fu T, Zhang X, Yang F, Zheng G, Xue W, Chen Y, Yao X, Zhu F (2017) Differentiating physicochemical properties between NDRIs and sNRIs clinically important for the treatment of ADHD. *Biochim Biophys Acta* **1861**, 2766-2777.
- [38] Li B, Tang J, Yang Q, Cui X, Li S, Chen S, Cao Q, Xue W, Chen N, Zhu F (2016) Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Sci Rep* **6**, 38881.
- [39] Zheng G, Xue W, Yang F, Zhang Y, Chen Y, Yao X, Zhu F (2017) Revealing vilazodone's binding mechanism underlying its partial agonism to the 5-HT1A receptor in the treatment of major depressive disorder. *Phys Chem Chem Phys* **19**, 28885-28896.
- [40] Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-389.
- [41] Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskva V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schurmann B, Heun R, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frolich L, Hampel H, Hull M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Muhleisen TW, Nothen MM, Moebs S, Jockel KH, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41**, 1088-1093.
- [42] Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buross J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JS, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, Green RC, Rogava E, St George-Hyslop P, Arnold SE, Barber R, Beach T, Bigio EH, Bowen JD, Boxer A, Burke JR, Cairns NJ, Carlson CS, Carney RM, Carroll SL, Chui HC, Clark DG, Corneveaux J, Cotman CW, Cummings JL, DeCarli C, DeKosky ST, Diaz-Arrastia R, Dick M, Dickson DW, Ellis WG, Faber KM, Fallon KB, Farlow MR, Ferris S, Frosch MP, Galasko DR, Ganguli M, Gearing M, Geschwind DH, Ghetti B, Gilbert JR, Gilman S, Giordani B, Glass JD, Growdon JH, Hamilton RL, Harrell LE, Head E, Honig LS, Hulette CM, Hyman BT, Jicha GA, Jin LW, Johnson N, Karlawish J, Karydas A, Kaye JA, Kim R, Koo EH, Kowall NW, Lah JJ, Levey AI, Lieberman AP, Lopez OL, Mack WJ, Marson DC, Martiniuk F, Mash DC, Masliah E, McCormick WC, McCurry SM, McDavid AN, McKee AC, Mesulam M, Miller BL, Miller CA, Miller JW, Parisi JE, Perl DP, Peskind E, Petersen RC, Poon WW, Quinn JF, Rajbhandary RA, Raskind M, Reisberg B, Ringman JM, Roberson ED, Rosenberg RN, Sano M, Schneider LS, Seeley W, Shelanski ML, Slifer MA, Smith CD, Sonnen JA, Spina S, Stern RA, Tanzi RE, Trojanowski JQ, Troncoso JC, Van Deerlin VM, Vinters HV, Vonsattel JP, Weintraub S, Welsh-Bohmer KA, Williamson J, Woltjer RL, Cantwell LB, Dombroski BA, Beekly D, Lunetta KL, Martin ER, Kambou MI, Saykin AJ, Reiman EM, Bennett DA, Morris JC, Montine TJ, Goate AM, Blacker D, Tsuang DW, Hakonarson H, Kukull WA, Foroud TM, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436-441.
- [43] Jun G, Asai H, Zeldich E, Drapeau E, Chen C, Chung J, Park JH, Kim S, Haroutunian V, Foroud T, Kuwano R, Haines JL, Pericak-Vance MA, Schellenberg GD, Lunetta KL, Kim JW, Buxbaum JD, Mayeux R, Ikezu T, Abraham CR, Farrer LA (2014) PLXNA4 is associated with Alzheimer disease and modulates tau phosphorylation. *Ann Neurol* **76**, 379-392.
- [44] Ma FC, Wang HF, Cao XP, Tan CC, Tan L, Yu JT (2018) Meta-analysis of the association between variants in ABCA7 and Alzheimer's disease. *J Alzheimers Dis* **63**, 1261-1267.
- [45] Cuyvers E, Sleegers K (2016) Genetic variations underlying Alzheimer's disease: Evidence from genome-wide association studies and beyond. *Lancet Neurol* **15**, 857-868.
- [46] Ward LD, Kellis M (2016) HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877-D881.
- [47] Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat Genet* **39**, 17-23.
- [48] Han Z, Jiang Q, Zhang T, Wu X, Ma R, Wang J, Bai Y, Wang R, Tan R, Wang Y (2015) Analyzing large-scale samples confirms the association between the rs1051730 polymorphism and lung cancer susceptibility. *Sci Rep* **5**, 15642.

- [49] Zhu F, Han LY, Chen X, Lin HH, Ong S, Xie B, Zhang HL, Chen YZ (2008) Homology-free prediction of functional class of proteins and peptides by support vector machines. *Curr Protein Pept Sci* **9**, 70-95.
- [50] Tu G, Fu T, Yang F, Yao L, Xue W, Zhu F (2018) Prediction of GluN2B-CT1290-1310/DAPK1 interaction by protein(-)peptide docking and molecular dynamics simulation. *Molecules* **23**, 3018.
- [51] Xue W, Fu T, Zheng G, Tu G, Zhang Y, Yang F, Tao L, Yao L, Zhu F (2018) Recent advances and challenges of the drugs acting on monoamine transporters. *Curr Med Chem*, doi: 10.2174/0929867325666181009123218
- [52] Han Z, Qu J, Zhao J, Zou X (2018) Analyzing 74,248 samples confirms the association between CLU rs11136000 polymorphism and Alzheimer's disease in Caucasian but not Chinese population. *Sci Rep* **8**, 11062.
- [53] Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y, Liu XS (2013) Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol* **20**, 908-913.
- [54] Jiang H, Wong WH (2008) SeqMap: Mapping massive amount of oligonucleotides to the genome. *Bioinformatics* **24**, 2395-2396.
- [55] Zheng G, Xue W, Wang P, Yang F, Li B, Li X, Li Y, Yao X, Zhu F (2016) Exploring the inhibitory mechanism of approved selective norepinephrine reuptake inhibitors and reboxetine enantiomers by molecular dynamics study. *Sci Rep* **6**, 26883.
- [56] Peart MJ, Smyth GK, van Laar RK, Bowtell DD, Richon VM, Marks PA, Holloway AJ, Johnstone RW (2005) Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. *Proc Natl Acad Sci U S A* **102**, 3697-3702.
- [57] Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, Iscove N, Jones S, McKinney S, Emerman J, Aparicio S, Marra M, Eaves C (2008) Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell* **3**, 109-118.
- [58] Tang J, Zhang Y, Fu J, Wang Y, Li Y, Yang Q, Yao L, Xue W, Zhu F (2018) Computational advances in the label-free quantification of cancer proteomics data. *Curr Pharm Des*, doi: 10.2174/1381612824666181102125638
- [59] Fu J, Tang J, Wang Y, Cui X, Yang Q, Hong J, Li X, Li S, Chen Y, Xue W, Zhu F (2018) Discovery of the consistently well-performed analysis chain for SWATH-MS based pharmacoproteomic quantification. *Front Pharmacol* **9**, 681.
- [60] Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM (2002) Meta-analysis of microarrays: Interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* **62**, 4427-4433.
- [61] Tseng GC, Ghosh D, Feingold E (2012) Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res* **40**, 3785-3799.
- [62] Li YH, Li XX, Hong JJ, Wang YX, Fu JB, Yang H, Yu CY, Li FC, Hu J, Xue WW, Jiang YY, Chen YZ, Zhu F (2019) Clinical trials, progression-speed differentiating features, and swiftness rule of the innovative targets of first-in-class drugs. *Brief Bioinform*, doi: 10.1093/bib/bby130
- [63] Friedman BA, Srinivasan K, Ayalon G, Meilandt WJ, Lin H, Huntley MA, Cao Y, Lee SH, Haddick PCG, Ngu H, Modrusan Z, Larson JL, Kaminker JS, van der Brug MP, Hansen DV (2018) Diverse brain myeloid expression profiles reveal distinct microglial activation states and aspects of Alzheimer's disease not evident in mouse models. *Cell Rep* **22**, 832-847.
- [64] Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527.
- [65] Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**, R106.
- [66] Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452-455.
- [67] Tang J, Fu JB, Wang YX, Li B, Li YH, Yang QX, Cui XJ, Hong JJ, Li XF, Chen YZ, Xue WW, Zhu F (2019) ANPELA: Analysis and performance-assessment of the label-free quantification workflow for metaproteomic studies. *Brief Bioinform*, doi: 10.1093/bib/bby127
- [68] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. *Genome Res* **12**, 996-1006.
- [69] Li X, Li X, Li Y, Yu C, Xue W, Hu J, Li B, Wang P, Zhu F (2018) What makes species productive of anti-cancer drugs? Clues from drugs' species origin, druglikeness, target and pathway. *Anticancer Agents Med Chem*, doi: 10.2174/1871520618666181029132017
- [70] Volders PJ, Helsen K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P (2013) LNCipedia: A database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* **41**, D246-251.
- [71] Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, UK Brain Expression Consortium, North American Brain Expression Consortium, Coin L, de Silva R, Cookson MR, Singleton AB, Hardy J, Ryten M, Weale ME (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* **17**, 1418-1428.
- [72] Consortium GT (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660.
- [73] Pauli A, Rinn JL, Schier AF (2011) Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* **12**, 136-149.
- [74] Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287.
- [75] Wang M, Roussos P, McKenzie A, Zhou X, Kajiwara Y, Brennan KJ, De Luca GC, Crary JF, Casaccia P, Buxbaum JD, Ehrlich M, Gandy S, Goate A, Katsel P, Schadt E, Haroutunian V, Zhang B (2016) Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med* **8**, 104.
- [76] Kondo T, Asai M, Tsukita K, Kutoku Y, Ohsawa Y, Sunada Y, Imamura K, Egawa N, Yahata N, Okita K, Takahashi K, Asaka I, Aoi T, Watanabe A, Watanabe K, Kadoya C, Nakano R, Watanabe D, Maruyama K, Hori O, Hibino S, Choshi T, Nakahata T, Hioki H, Kaneko T, Naitoh M, Yoshikawa K, Yamawaki S, Suzuki S, Hata R, Ueno S, Seki T, Kobayashi K, Toda T, Murakami K, Irie K, Klein WL, Mori H, Asada T, Takahashi R, Iwata N, Yamanaka S, Inoue H (2013) Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Abeta and differential drug responsiveness. *Cell Stem Cell* **12**, 487-496.
- [77] Liang WS, Reiman EM, Valla J, Dunckley T, Beach TG, Grover A, Niedzielko TL, Schneider LE, Mastromei D, Caselli R, Kukull W, Morris JC, Hulette CM, Schmechel D, Rogers J, Stephan DA (2008) Alzheimer's disease is associated with reduced expression of energy metabolism genes

- in posterior cingulate neurons. *Proc Natl Acad Sci U S A* **105**, 4441-4446.
- [78] Hokama M, Oka S, Leon J, Ninomiya T, Honda H, Sasaki K, Iwaki T, Ohara T, Sasaki T, LaFerla FM, Kiyohara Y, Nakabeppu Y (2014) Altered expression of diabetes-related genes in Alzheimer's disease brains: The Hisayama study. *Cereb Cortex* **24**, 2476-2488.
- [79] Antonell A, Llado A, Altirriba J, Botta-Orfila T, Balasa M, Fernandez M, Ferrer I, Sanchez-Valle R, Molinuevo JL (2013) A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease. *Neurobiol Aging* **34**, 1772-1778.
- [80] Zuccato C, Cattaneo E (2009) Brain-derived neurotrophic factor in neurodegenerative diseases. *Nat Rev Neurol* **5**, 311-322.
- [81] Tao L, Zhu F, Xu F, Chen Z, Jiang YY, Chen YZ (2015) Co-targeting cancer drug escape pathways confers clinical advantage for multi-target anticancer drugs. *Pharmacol Res* **102**, 123-131.
- [82] Malinin NL, Wright S, Seubert P, Schenk D, Griswold-Prenner I (2005) Amyloid-beta neurotoxicity is mediated by FISH adapter protein and ADAM12 metalloprotease activity. *Proc Natl Acad Sci U S A* **102**, 3058-3063.
- [83] Anello G, Gueant-Rodriguez RM, Bosco P, Gueant JL, Romano A, Namour B, Spada R, Caraci F, Pourie G, Daval JL, Ferri R (2004) Homocysteine and methylenetetrahydrofolate reductase polymorphism in Alzheimer's disease. *Neuroreport* **15**, 859-861.
- [84] Zorzetto M, Datturi F, Divizia L, Pistono C, Campo I, De Silvestri A, Cuccia M, Ricevuti G (2017) Complement C4A and C4B gene copy number study in Alzheimer's disease patients. *Curr Alzheimer Res* **14**, 303-308.
- [85] Parr C, Mirzaei N, Christian M, Sastre M (2015) Activation of the Wnt/beta-catenin pathway represses the transcription of the beta-amyloid precursor protein cleaving enzyme (BACE1) via binding of T-cell factor-4 to BACE1 promoter. *FASEB J* **29**, 623-635.
- [86] Gardberg AS, Dice LT, Ou S, Rich RL, Helmbrecht E, Ko J, Wetzel R, Myszka DG, Patterson PH, Dealwis C (2007) Molecular basis for passive immunotherapy of Alzheimer's disease. *Proc Natl Acad Sci U S A* **104**, 15659-15664.
- [87] Jevtic S, Sengar AS, Salter MW, McLaurin J (2017) The role of the immune system in Alzheimer disease: Etiology and treatment. *Ageing Res Rev* **40**, 84-94.
- [88] Jiang Q, Jin S, Jiang Y, Liao M, Feng R, Zhang L, Liu G, Hao J (2017) Alzheimer's disease variants with the genome-wide significance are significantly enriched in immune pathways and active in immune cells. *Mol Neurobiol* **54**, 594-600.
- [89] Tooyama I, Kimura H, Akiyama H, McGeer PL (1990) Reactive microglia express class I and class II major histocompatibility complex antigens in Alzheimer's disease. *Brain Res* **523**, 273-280.
- [90] Gao XM, Wordsworth P, McMichael AJ, Kyaw MM, Seifert M, Rees D, Dougan G (1996) Homocysteine modification of HLA antigens and its immunological consequences. *Eur J Immunol* **26**, 1443-1450.
- [91] Bladt F, Tafuri A, Gelkop S, Langille L, Pawson T (2002) Epidermolysis bullosa and embryonic lethality in mice lacking the multi-PDZ domain protein GRIP1. *Proc Natl Acad Sci U S A* **99**, 6816-6821.
- [92] Xue W, Yang F, Wang P, Zheng G, Chen Y, Yao X, Zhu F (2018) What contributes to serotonin-norepinephrine reuptake inhibitors' dual-targeting mechanism? The key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem Neurosci* **9**, 1128-1140.
- [93] Fu T, Zheng G, Tu G, Yang F, Chen Y, Yao X, Li X, Xue W, Zhu F (2018) Exploring the binding mechanism of metabotropic glutamate receptor 5 negative allosteric modulators in clinical trials by molecular dynamics simulations. *ACS Chem Neurosci* **9**, 1492-1502.