

PRIMARY RESEARCH

Open Access

# Convergent lines of evidence support *BIN1* as a risk gene of Alzheimer's disease



Jin Zhu<sup>1</sup>, Xia Liu<sup>2</sup>, Hongtao Yin<sup>3</sup>, Yan Gao<sup>1</sup> and Hao Yu<sup>1\*</sup>

## Abstract

Genome-wide association studies (GWAS) have identified several susceptibility loci of Alzheimer's disease (AD), which were mainly located in noncoding regions of the genome. Meanwhile, the putative biological mechanisms underlying AD susceptibility loci were still unclear. At present, identifying the functional variants of AD pathogenesis remains a major challenge. Herein, we first used summary data-based Mendelian randomization (SMR) with AD GWAS summary and expression quantitative trait loci (eQTL) data to identify variants who affects expression levels of nearby genes and contributed to the risk of AD. Using the SMR integrative analysis, we totally identified 14 SNPs significantly affected the expression level of 16 nearby genes in blood or brain tissues and contributed to the AD risk. Then, to confirm the results, we replicated the GWAS and eQTL results across multiple samples. Totally, four risk SNP (rs11682128, rs601945, rs3935067, and rs679515) were validated to be associated with AD and affected the expression level of nearby genes (*BIN1*, *HLA-DRA*, *EPHA1-AS1*, and *CR1*). Besides, our differential expression analysis showed that the *BIN1* gene was significantly downregulated in the hippocampus ( $P = 2.0 \times 10^{-3}$ ) and survived after multiple comparisons. These convergent lines of evidence suggest that the *BIN1* gene identified by SMR has potential roles in the pathogenesis of AD. Further investigation of the roles of the *BIN1* gene in the pathogenesis of AD is warranted.

**Keywords:** Alzheimer's disease, Genome-wide association study (GWAS), Expression level, Candidate gene, Mendelian randomization

## Introduction

Alzheimer's disease (AD) is the most common neurodegenerative dementia and is clinically characterized by progressive loss of memory and deficits in thinking, problem solving, and language [1]. AD is highly heritable and its estimated heritability ranges from 60 to 80% [2]. Genome-wide association studies (GWAS) have identified multiple loci containing common variant risk alleles [3–5]. A large-scale GWAS of clinically diagnosed AD and AD-by-proxy (71,880 cases and 383,378 controls) identified 29 risk loci, involving 215 potential causative genes [6]. Another GWAS of late-onset Alzheimer's disease (21,982 cases and

41,944 controls) identified five novel genome-wide loci, including *IQCK*, *ACE*, *ADAM10*, *ADAMTS1*, and *WWOX* [7]. These findings offer new routes to enhancing the diagnosis and the development of drug targets [8]. However, most of the identified risk single nucleotide polymorphisms (SNPs) are from noncoding regions [9, 10], making functional interpretation difficult.

One possible hypothesis is that the risk SNPs identified by GWAS contribute to the risk of diseases through affecting the expression level of nearby genes in different tissues [10, 11]. Consequently, to identify the functional variants from GWAS results, it is useful to integrate data of gene expression level (e.g., expression quantitative trait loci, eQTL) into GWAS data of diseases. Therefore, to prioritize candidate genes underlying GWAS hits, an integrated analysis method named summary data-based

\* Correspondence: [yuhao@mail.jnmc.edu.cn](mailto:yuhao@mail.jnmc.edu.cn)

<sup>1</sup>Department of Psychiatry, Jining Medical University, 133 He Hua Road, Jining 272067, Shandong, China

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Mendelian randomization (SMR) was developed. Using the principles of Mendelian randomization, the SMR method could examine whether the expression level of a gene and a complex phenotype caused by pleiotropy and discern pleiotropy from linkage [12]. Through the SMR analysis, several novel candidate genes underlying GWAS hits of complex diseases or traits were prioritized for follow-up functional studies [13–16]. Strikingly, through integrating different omics data, we could gain further insights into the underlying genetic mechanisms of GWAS hits and disease [17].

To prioritize AD risk genes and investigate their roles in AD pathogenesis, we first combined the AD GWAS data and eQTL using SMR test. Then, we replicate the identified risk SNPs and genes across multiple samples. For the replicated risk genes, we compared the expression patterns in AD patients with healthy controls.

## Methods

### AD GWAS data

We obtained complete summary-level of AD GWAS from the website of Complex Trait Genomics lab ([https://ctg.cncr.nl/software/summary\\_statistics](https://ctg.cncr.nl/software/summary_statistics)). The AD GWAS consisted of 71,880 cases and 383,378 controls [6]. In PGC, IGAP, and ADSP consortia, individuals were of clinically diagnosed AD case-control status. The individuals with one or two parents diagnosed with AD in UKB were defined as proxy cases, and patients with two parents were upweighted. Meanwhile, participants with two parents without AD were defined as proxy controls, and older cognitively normal parents were also upweighted [6]. Recently, the value of by-proxy phenotypes has been demonstrated [5]. More details about demographic characteristics, genotyping, and statistical analysis were in the original study [6].

### eQTL data

In the SMR analysis, we integrated the AD GWAS data with brain and blood eQTL data, respectively. (1) For blood eQTL data ( $n = 31,684$ ), the blood eQTL data was obtained from the eQTLGen consortium, which consisted of 31,684 individuals [18]. The associations between SNPs and gene expression levels were calculated using a Spearman correlation. In total, in the eQTLGen consortium, 19,960 genes that showed expression in the blood were tested and 238,340 *cis*-eQTL SNPs were identified. (2) For brain eQTL data ( $n = 1194$ ), the brain eQTL study was from a meta-analysis of brain eQTL data [19]. To increase the power of detecting brain eQTLs, Qi et al. [19] performed a meta-analysis using three brain eQTL studies, including Genotype-Tissue Expression (GTEx) [20], CommonMind Consortium (CMC) [21], and the Religious Orders Study and the Rush Memory and Aging Project (ROSMAP) [17]. To correct the overlapped sample, the MeCS approach was

used to combine the eQTL results of 10 brain regions of GTEx database [19]. In the present study, we only used the SNPs within 1 Mb distance from each gene. More details were in the original paper [18, 19].

### SMR analysis

To prioritize candidate causal genes of AD, we integrated GWAS and eQTL data through SMR method, which examine the putative pleiotropic relationships between AD and eQTL [12]. The SMR method mainly comprises of two steps. First, genetic variations are used as instrumental variables to examine for causative effect of gene expression on AD. Second, we applied the heterogeneity in dependent instruments (HEIDI) test implemented in SMR software to distinguish the causality and pleiotropy model from the linkage model. If the HEIDI test is significant ( $P_{\text{HEIDI}} < 0.05$ ), the identified genes by SMR can be a result of linkage. To account for multiple testing, we adjusted  $P_{\text{SMR}}$  values using the Bonferroni approach. The set associated genes were defined as genes with a Bonferroni-corrected  $P_{\text{SMR}} < 0.05$  and  $P_{\text{HEIDI}} > 0.05$ . The SMR software was downloaded from <https://cns.genomics.com/software/smr>.

### AD GWAS data for replication analysis

To further replicate the AD GWAS results in SMR, we investigated the associations between the identified risk SNPs and AD using the GWAS summary data of International Genomics of Alzheimer's Project (IGAP), which is a large three-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry [22]. In our study, we extracted the association results from the stage 1 results of IGAP, consisting of 21,982 AD cases and 41,944 normal controls [22]. More details of samples, quality control, imputation, and statistical analysis were in the original study [22].

### eQTL data for replication analysis

To validate the eQTL results in SMR, we examined the *cis*-eQTL effects of risk SNPs using two public databases as follows. First, we examined the blood eQTL results using the GTEx database. The genotype data used for eQTL analyses in GTEx was based on whole exome sequencing from 838 donors, which all had RNA-seq data available [23]. The associations between was performed using FastQTL. Totally, 49 tissues were tested in GTEx. Second, in the PsychENCODE database, to replicate the brain eQTL results of SMR analysis, we used the *cis*-eQTL data in the prefrontal cortex from the PsychENCODE project ( $n = 1387$ ) [24]. The eQTL analyses of PsychENCODE were performed including 100 hidden covariate factors as covariates. Only the data of SNPs in a 1-Mb window around each gene are available.

### Differential expression analysis of risk genes

To compare the expression level of the risk genes in AD cases with healthy controls, we performed the differential expression analysis using the comprehensive AlzData database (<http://www.alzdata.org/>) [25]. The AlzData database consisted of the expression data of four brain regions, including entorhinal cortex (EC), hippocampus (HIPP), temporal cortex (TC), and frontal cortex (FC). After conducting the cross-platform normalization, the normalized expression data sets were used to perform different expression analysis between AD cases and controls, using the linear regression model implemented in R package limma [25]. We used the false discovery rate (FDR) method to correct for multiple comparisons [25].

## Results

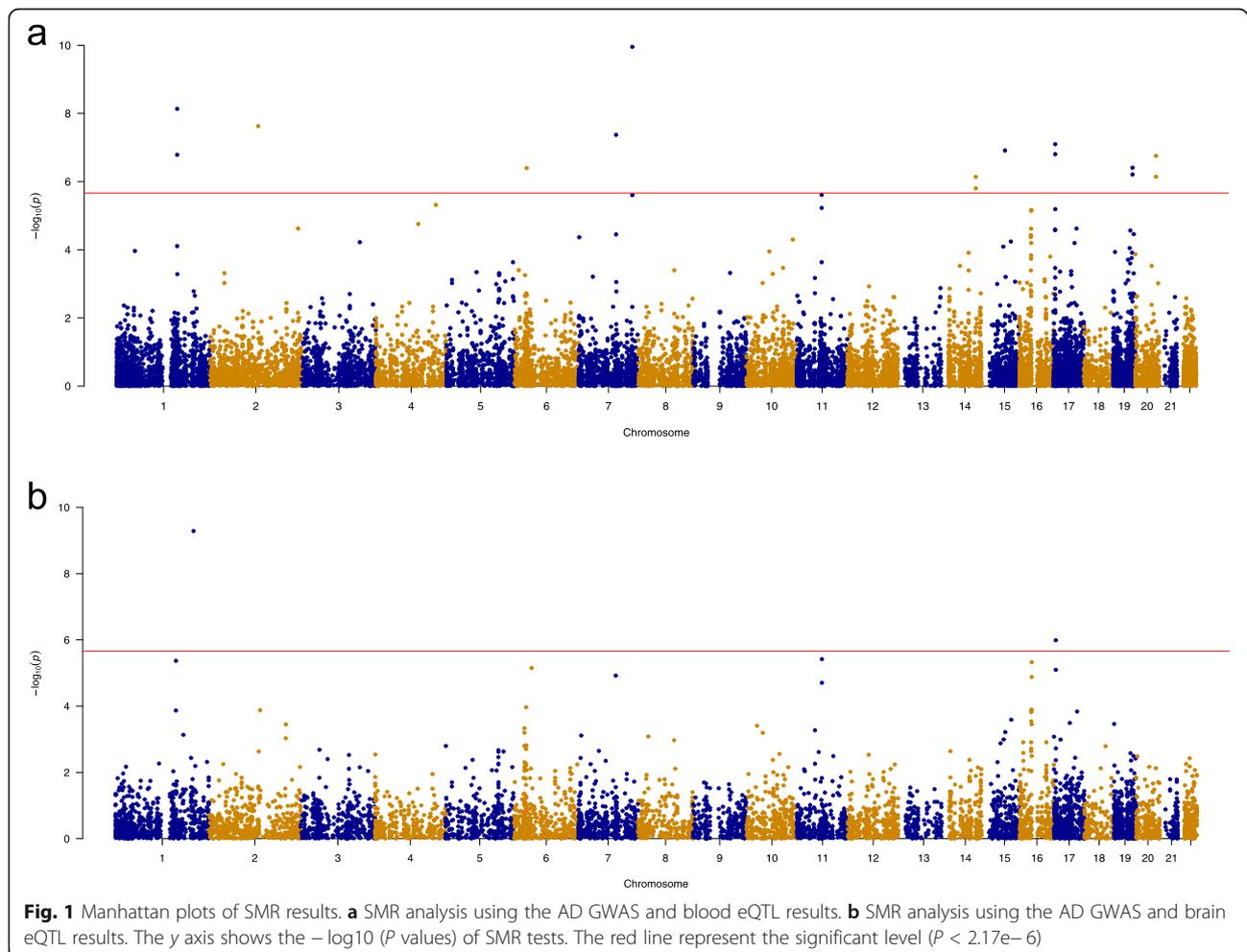
### SMR analysis identified risk variants and genes for AD

To identify functional variants related to AD, we conducted SMR analysis using the genome-wide significantly associated genetic variants as an instrumental variable to examine the association between the

expression level of each gene and AD. In the SMR analysis, we integrated AD GWAS with eQTL data from the blood and brain, respectively. Totally, 6 genes in the brain and 22 genes in blood were identified after correcting for multiple comparisons ( $P_{\text{SMR}} < 0.05/n$ ;  $n = 23048$ ;  $n$  represent the number of tests across blood and brain SMR analysis; Fig. 1 and Table 1). Then, we performed the HEIDI analysis for the identified genes to reduce the effect of potential linkage. Of the genes identified in the SMR analysis, 2 genes in the brain and 15 genes in blood were survived after the HEIDI test ( $P_{\text{HEIDI}} > 1.79 \times 10^{-3}$ , i.e.,  $0.05/n$ , with  $n = 28$  being the total number of HEIDI tests) (Fig. 1 and Table 1), with one gene in common and 16 unique genes in total. Besides, the SMR analysis identified 14 AD risk SNPs (Fig. 1 and Table 1).

### Replication analysis of GWAS and eQTL results

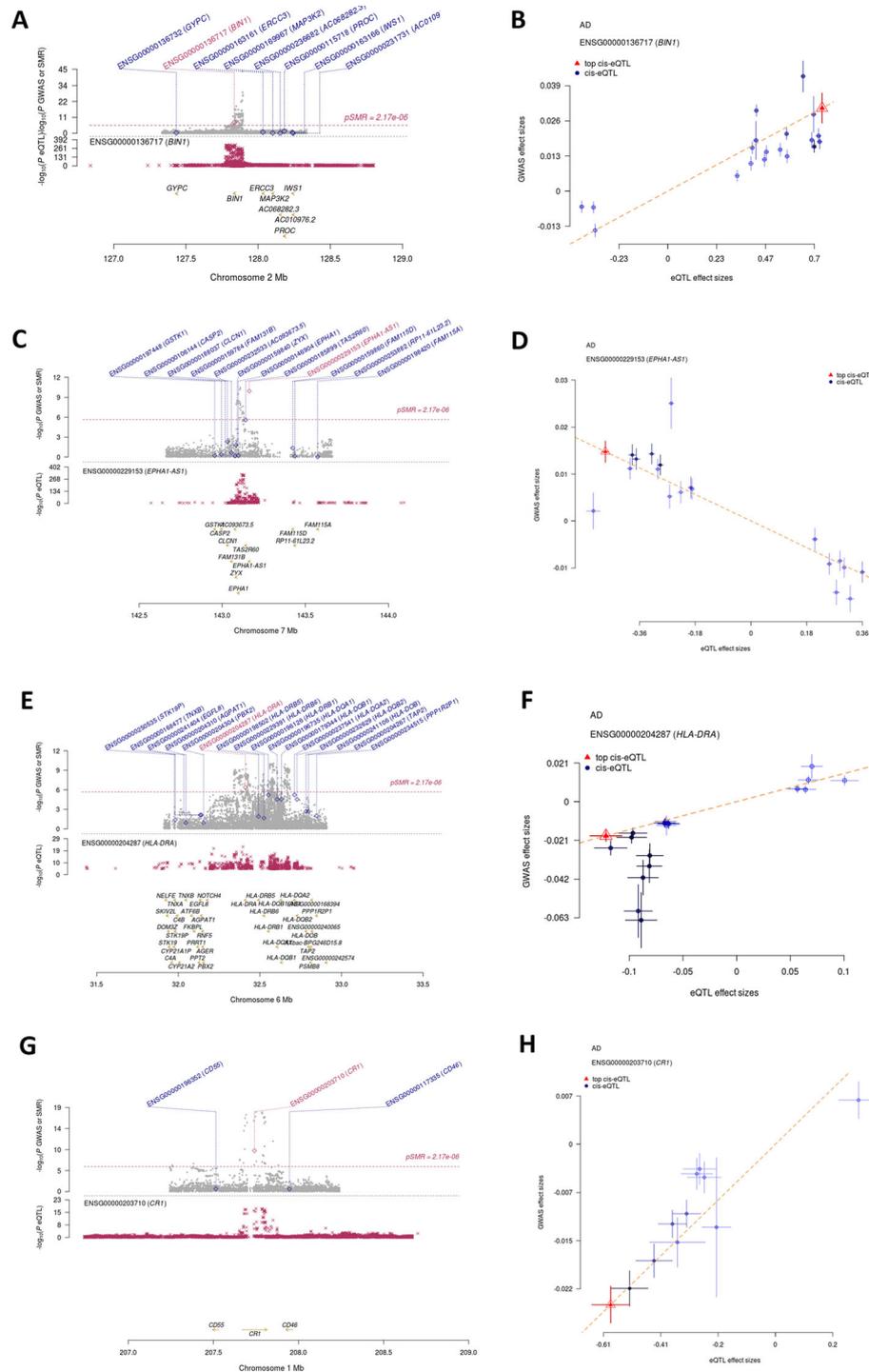
To further investigate the associations between 14 risk SNPs and AD, we replicated the SNPs results using a meta-analysis of AD GWASs. All the 14 SNPs showed



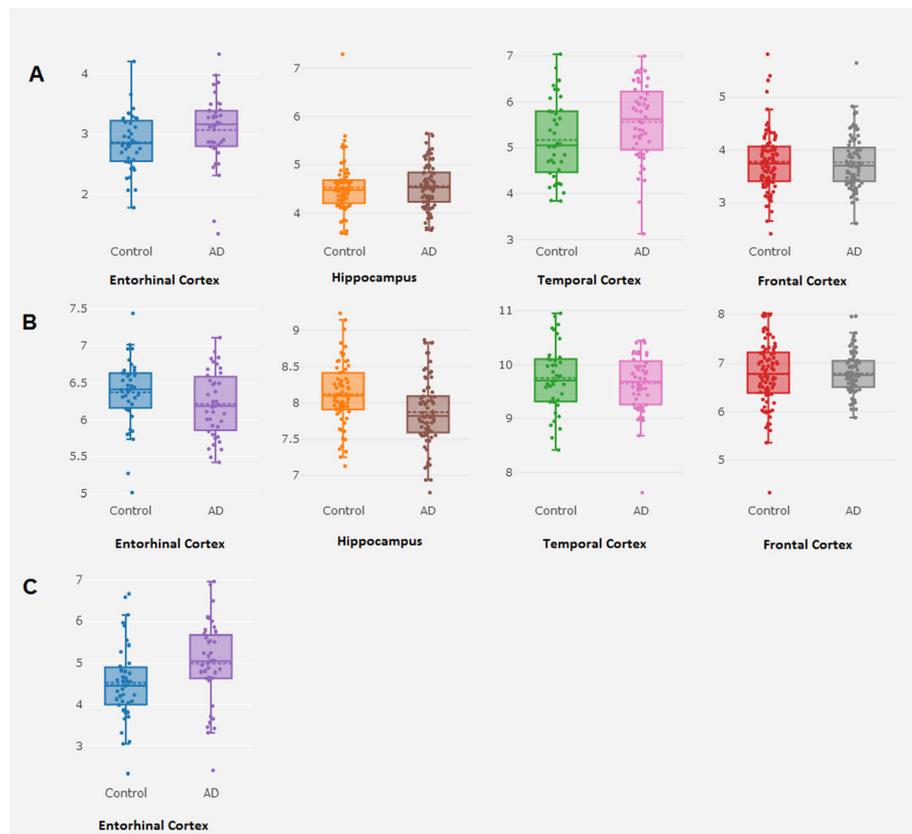
**Table 1** Integrative analysis of GWAS and eQTL identified risk genes of AD

Chr	BP	SNP	A1	A2	Gene	b_GWAS	se_GWAS	p_GWAS	b_eQTL	se_eQTL	p_eQTL	b_SMR	se_SMR	p_SMR	p_HEIDI
<b>Integrative analysis of AD GWAS and blood eQTL</b>															
1	161156033	rs11585858	A	C	B4GALT3	0.016	0.003	5.58E-10	-0.168	0.011	1.07E-57	-0.093	0.016	7.33E-09	3.33E-03
1	161186313	rs4379692	T	C	NDUFS2	0.012	0.002	8.25E-08	-0.231	0.009	4.34E-132	-0.053	0.010	1.63E-07	1.22E-01
2	127839474	rs11682128	A	G	BINI	0.031	0.006	2.21E-08	0.742	0.009	0.00E+00	0.042	0.008	2.36E-08	2.12E-03
6	32573415	rs601945	G	A	HLA-DRA	-0.019	0.003	1.38E-10	-0.117	0.014	1.39E-16	0.160	0.032	3.99E-07	4.25E-03
7	99803412	rs2950517	G	C	CASTOR3	-0.014	0.002	2.15E-09	0.124	0.009	3.30E-42	-0.114	0.021	4.26E-08	8.74E-03
7	143104331	rs3935067	C	G	EPHA1-AS1	0.015	0.002	6.52E-11	-0.466	0.011	0.00E+00	-0.031	0.005	1.11E-10	6.94E-03
14	92955385	rs17783630	C	A	SLC24A4	0.011	0.002	6.71E-07	-0.503	0.008	0.00E+00	-0.021	0.004	7.31E-07	3.28E-02
14	92955385	rs17783630	C	A	RIN3	0.011	0.002	6.71E-07	-0.165	0.009	9.12E-77	-0.064	0.013	1.58E-06	6.15E-02
15	63571820	rs75763893	T	C	APH1B	0.017	0.003	9.15E-08	0.681	0.018	0.00E+00	0.025	0.005	1.22E-07	5.00E-02
17	5014212	rs73976310	A	G	ACO12146.7	0.018	0.003	7.04E-08	-0.743	0.013	0.00E+00	-0.024	0.004	8.00E-08	1.55E-02
17	5014212	rs73976310	A	G	ZNF232	0.018	0.003	7.04E-08	-0.305	0.013	8.16E-114	-0.058	0.011	1.57E-07	7.88E-02
19	51731176	rs7245846	A	G	SIGLEC2P	-0.012	0.002	1.19E-07	0.166	0.009	2.38E-70	-0.072	0.014	3.91E-07	1.05E-01
19	51726911	rs1710398	C	A	CD33	0.011	0.002	4.46E-07	0.290	0.009	3.60E-221	0.037	0.007	6.18E-07	4.27E-03
20	54989833	rs6014722	A	T	CASS4	-0.022	0.004	2.62E-09	-0.141	0.016	3.81E-19	0.159	0.032	7.20E-07	4.18E-03
20	54987216	rs17462136	C	G	RPL39P	-0.022	0.004	6.43E-09	-0.208	0.017	5.07E-33	0.106	0.020	1.76E-07	2.50E-03
<b>Integrative analysis of AD GWAS and brain eQTL</b>															
1	207750568	rs679515	C	T	CR1	-0.025	0.003	1.10E-18	-0.587	0.067	2.10E-18	0.042	0.007	5.17E-10	4.67E-02
17	5014212	rs73976310	G	A	ACO12146.7	-0.018	0.003	7.04E-08	0.721	0.062	6.19E-31	-0.025	0.005	1.03E-06	2.73E-01

We used the blood eQTL results of eQTLGen consortium [18], and the brain eQTL data from the study by Qi et al. [19]. Chr chromosome, BP base position, Gene gene symbol, A1 reference allele, A2 alternative allele, b\_GWAS regression coefficient for AD (positive value indicates increased risk for AD in carriers of the alternative allele and negative value represents decreased risk for AD), se\_GWAS standard error in GWAS, p\_GWAS p value of GWAS, b\_eQTL beta value of eQTL (positive values indicate increased expression with alternative allele and negative value represents decreased expression level), se\_eQTL standard error of eQTL, p\_eQTL p value of eQTL, b\_SMR beta value of SMR test (positive values indicate increased expression level of target gene contributed to the risk of AD and negative values represent the opposite results), se\_SMR standard error of SMR test, p\_SMR p value of SMR test



**Fig. 2** Prioritizing genes at four loci for AD. **a, c, e, g** The brown dots at top plot represent the association between SNPs and AD in GWAS, diamonds represent the  $P$  values of SMR analysis, and triangles stand for genes without a  $P_{\text{eQTL}} < 5.0 \times 10^{-8}$ . In the bottom plot, the SNPs with  $P_{\text{eQTL}}$  of eQTL study were plotted. The genes that survived after the SMR and HEIDI tests were highlighted using red color. **b, d, f, h** We showed the effect estimates of SNPs from AD GWAS plotted against those for SNPs from the eQTL analysis. The orange lines represent the estimate of effect size at the top cis-eQTL. Error bars represent the standard errors of SNP effects size



**Fig. 3** Differential expression analysis of candidate genes using AlzData database. We examined the expression level of the risk genes in AD cases and healthy controls using the AlzData database (<http://www.alzdata.org/>). **a** Differential expression analysis of *CRI*. **b** Differential expression analysis of *BIN1*. **c** Differential expression analysis of *HLA-DRA*

nominally significant association with AD ( $P < 0.05$ ) with the same effect directions in IGAP GWAS dataset (Supplementary Table 1). Nine risk SNPs were still significant after Bonferroni correction ( $P < 0.05/14 = 3.57 \times 10^{-3}$ ; Supplementary Table 1). To further examine whether the 9 SNPs were associated with the expression level of nearby genes, we replicated the blood and brain eQTL effects identified by SMR using GTEx and PsychoENCODE datasets, respectively. Of the eight blood eQTL effects identified by SMR, 3 SNPs (rs11682128, rs601945, and rs3935067) showed genome-wide *cis*-eQTL effects in blood tissues ( $P < 5 \times 10^{-8}$ ; Supplementary Table 2). In the replication analysis of brain eQTL effects, the SNP rs679515 was replicated in PsychoENCODE database

(Supplementary Table 3). Totally, 4 SNPs (rs11682128, rs601945, rs3935067, and rs679515) are replicated in blood and brain eQTL databases, respectively. Therefore, four SNP-gene combinations, rs11682128-*BIN1*, rs601945-*HLA-DRA*, rs3935067-*EPHA1-AS1*, and rs679515-*CRI*, were strongly suggested to be promising candidates for AD risk. To better view the SMR results of these 4 SNPs, we plotted the GWAS, eQTL results, and SMR results in Fig. 2.

#### Differential expression analysis of the AD risk genes

Considering that the expression level of risk genes might change and contribute to AD risk, we further investigated whether the four risk genes are differentially expressed in AD patients compared to controls by using the AlzData

**Table 2** Expression analysis of risk genes in AD cases and controls using AlzData database

Gene	Entorhinal cortex			Hippocampus			Temporal cortex			Frontal cortex		
	FC	P	FDR	FC	P	FDR	FC	P	FDR	FC	P	FDR
CR1	0.23	0.071	0.197	-0.01	0.896	0.954	0.39	0.033	0.118	-0.04	0.655	0.906
BIN1	-0.12	0.265	0.452	-0.26	0.002	0.028	-0.09	0.458	0.67	0.11	0.13	0.268
HLA-DRA	0.46	0.042	0.142	NA	NA	NA	NA	NA	NA	NA	NA	NA

We investigated whether the identified risk genes are differentially expressed in AD patients compared to controls, using a comprehensive database AlzData (<http://www.alzdata.org/>). FC fold change, FDR false discovery rate, NA not available

database (<http://www.alzdata.org/>) [25]. Comparing AD patients with controls, the *BINI* gene was significantly downregulated in the hippocampus ( $P = 0.002$ ; Fig. 3 and Table 2), surviving after FDR correction in the original study [25]. Based on the SMR results, our differential expression analysis further supports *BINI* as an AD risk gene. However, the *HLA-DRA* and *CRI* genes showed no significant differential expression pattern between AD cases and controls (Fig. 3 and Table 2). The gene *EPHA1-AS1* was not available in the AlzData database.

## Discussion

Recently, hundreds of AD risk SNPs have been identified in GWAS [3–5]. The large majority of risk loci of AD are located in noncoding regions of the genome. How to identify the genetic mechanisms underlying risk SNPs remains a major challenge. Moreover, given that the gene density and linkage disequilibrium structure, it is difficult to identify causal SNPs for AD. Based on GWAS results alone, we could not predict whether the risk SNPs have functional consequences. In this study, by using the SMR analysis, we systematically integrate the AD GWAS and blood or brain eQTL data. Ultimately, we identified 14 risk SNPs, which affected the expression level of 16 nearby genes and contributed to risk for AD. Our results support that the gene expression might play a mediating role for effects at these risk SNPs. Our findings not only confirmed previous findings, but also highlighted new risk SNPs and genes underlying AD. Through SMR analysis, we identified eight novel risk SNPs that were not genome-wide significant in the original AD GWAS [6]. Hence, some missing heritability might be identified using SMR. To further confirm the SMR results, we replicated the GWAS and eQTL results. Totally, four genes (*BINI*, *HLA-DRA*, *EPHA1-AS1*, and *CRI*) were strongly suggested to be promising candidates for AD risk. We expect these SNPs to be detected in future genetic association studies with larger sample sizes. Then, we conducted the differential expression analysis to compare the expression level of four replicated genes in AD cases and controls. Only the *BINI* gene showed significant differential expression level. Therefore, we demonstrated that the *BINI* gene contributed to the risk of AD.

Our study provides convergent lines of evidence supporting the *BINI* gene as a candidate gene of AD. First, we identified the AD risk gene *BINI* by integrating large-scale GWAS and eQTL with SMR analysis. Second, the SMR results were replicated across GWAS and eQTL databases. Third, given that the SMR test identifies AD-associated genes with the underlying assumption that expression levels of those genes may have a role in AD pathogenesis, we explored whether AD risk genes identified by SMR were differentially

expressed in AD patients compared to controls, using the comprehensive AlzData database [25]. Comparing AD patients with controls, the *BINI* gene was also significantly downregulated in the hippocampus. However, there were no significant differences in the expression of other genes. This might be due to the lack of power and heterogeneity of different expression data sets.

Our SMR results identified that risk SNPs caused the dysregulation of the gene expression level and increased the risk for AD. However, our findings for an association between *BINI* and risk of AD are mixed, suggesting the complex role of *BINI* in AD risk. First, our SMR results in blood are consistent with previous studies. At the *BINI* locus, our SMR results suggested that the risk allele A of SNP rs11682128 could upregulate the expression level of the *BINI* gene in blood and increase the AD risk. Consistent with our results, higher *BINI* mRNA levels in blood were detected in AD patients compared with controls [26]. Next, our results of the expression level of *BINI* in brain were different from previous findings. Using AlzData database [25], we found that the *BINI* gene was significantly downregulated in AD patients compared to controls in hippocampus (Table 2). Coincidentally, the AD risk allele of *BINI* showed significant associations with memory deficits, hippocampal volume, and functional connectivity, suggesting the potential role of *BINI* in AD pathogenesis [27, 28]. However, most of previous evidence showed an increase of *BINI* expression level in the brains of patients with AD [29, 30]. Moreover, the increased *BINI* expression level has also been linked to tau pathology [29–32]. These inconsistent findings might be interpreted by the different functions of different domains in *BINI* gene. Compared to healthy controls, the amount of the largest isoform of *BINI* was found to be significantly reduced in the AD brain, and smaller *BINI* isoforms were significantly increased [31]. Third, we found inconsistency between SMR results in blood and differential expression results in brain. This phenomenon may be caused by diverse roles of *BINI* in AD pathology. Many kinds of evidence has shown that *BINI* may involve in several AD-related pathways in AD, including tau and amyloid pathology, and relevant pathways such as inflammation, apoptosis, and calcium homeostasis [33]. Additionally, though previous studies suggested that the genetic architecture underlies the regulation of gene expression across tissues, there are still some genetic differences between tissues [19]. Therefore, we inferred that the different functions of different domains and distinct tissue localizations may indicate the role of *BINI* in the pathogenesis of AD. However, adequate and reliable research on *BINI* in AD is still needed in the future.

Compared with these two previous studies, our present study has some similarities and differences.

Previous studies have demonstrated that the SMR method was helpful to prioritize novel AD-associated genes. For example, Hu et al. identified several candidate genes by integrating two AD GWASs and five eQTL studies using SMR test [34]. Then, to improve their result, Zhao et al. performed a meta-analysis using five AD GWAS and integrated the meta results with eQTL using SMR [35]. Several risk genes were identified to be associated with AD in expression levels by pleiotropy [35]. Notably, all three studies applied SMR to AD GWAS and brain eQTL data. Hu et al. used two AD GWAS (25,580 AD cases and 48,466 controls) and five eQTL to perform SMR test [34]. Zhao et al. used summary statistics from a mega-analysis of five GWAS datasets (369,957 participants) and three brain eQTL [35]. Meanwhile, our present study used GWAS data (71,880 AD cases and 383,378 controls) from the mega-analysis by Jansen et al. [6], blood eQTL data ( $n = 31,684$ ), and brain eQTL data ( $n = 1194$ ). Generally, the current study had increased the sample size compared with previous studies [34, 35] and then might improve the statistical power and accuracy of SMR statistical results. The current study identified several risk genes which were not identified by two previous SMR studies [34, 35], such as *NDUFS2*, *CASTOR3*, *APH1B*, and *B4GALT3*, extending the findings of previous studies. Second, we not only prioritized risk gene using SMR test, but also replicated the SMR results in IGAP GWAS, GTEx, and PsychENCODE databases. Besides, we also explored the functional roles of these identified SNPs using differential gene expression patterns in AD patients and controls. These identified genes using the integrated computational analyses could be prioritized based on biological relevance using follow-up laboratory-based validation using in vitro and in vivo model systems.

Our study has a number of limitations. First, in the first-stage of SMR analysis, some AD cases of the GWAS sample were defined based on the parental diagnoses. Therefore, the SNP associations might be biased. However, the strategy of AD-by-proxy was demonstrated to be robust. For example, the diagnosed case-control status and the UKB by-proxy phenotype showed high genetic correlation, and a large proportion of novel loci were replicated in the independent cohort [6]. Furthermore, we replicated the GWAS results using IGAP samples, which were clinically diagnosed. Therefore, the biases in AD associations caused by misdiagnosis might be relatively modest. Second, our study provides several lines of evidence that the *BINI* gene contributes to the risk of AD. However, the potential casual gene *BINI* was identified through using the GWAS and eQTL results of European population. These prioritized genes might not be associated with AD in other populations. Thus, these results should be validated in other populations.

## Conclusions

In this study, we combined the GWAS and eQTL datasets and identified the risk SNP rs11682128, which might contribute to AD risk through affecting the expression level of *BINI* gene. Our SMR analysis could not only identify functional genes but improve our understanding of the pathogenesis mechanism underlying AD.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-021-00307-6>.

**Additional file 1: Supplementary Table 1.** Replication analysis for the association between risk SNPs and AD. **Supplementary Table 2.** Replication analysis for the blood eQTL results in the GTEx database. **Supplementary Table 3.** Replication analysis for the brain eQTL results in the PsychENCODE database.

## Acknowledgements

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i-Select chips were funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD/PERADES was supported by the Medical Research Council (grant no. 503480), Alzheimer's Research UK (grant no. 503176), the Wellcome Trust (grant no. 082604/2/07/Z), and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant no. 01GI0102, 01GI0711, and 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01-AG-12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants, U01 AG032984, U24 AG021886, and U01 AG016976, and the Alzheimer's Association grant ADGC-10-196728.

## Authors' contributions

Hao Yu contributed to the study design. Jin Zhu, Xia Liu, Hongtao Yin, Yan Gao, and Hao Yu conducted the data analysis. Jin Zhu and Hao Yu took up the responsibility of writing this manuscript. Xia Liu, Hongtao Yin, and Hao Yu revised the manuscript. The authors read and approved the final manuscript.

## Funding

This study was funded by the National Natural Science Foundation of China (81901358), Natural Science Foundation of Shandong Province (ZR2019BH001), Medical and Health Science and Technology Development Plan of Shandong Province (2014WS0278 and 2018WS457), Young Taishan Scholars of Shandong Province (tsqn201909146), Supporting Fund for Teacher's Research of Jining Medical University (JY2016KJ020Y, JY2017JS050, and JYFC2018JS002), and Research Fund for Lin He's Academician Workstation of New Medicine and Clinical Translation in Jining Medical University (JYHL2019MS05).

## Availability of data and materials

Not applicable

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

### Competing interests

Not applicable

### Author details

<sup>1</sup>Department of Psychiatry, Jining Medical University, 133 He Hua Road, Jining 272067, Shandong, China. <sup>2</sup>Department of Psychiatry, Jining Psychiatric Hospital, Jining 272051, Shandong, China. <sup>3</sup>Department of Neurology, Zibo Central Hospital, 54 Gongqingtuan Xi Road, Zibo 255036, China.

Received: 8 November 2020 Accepted: 18 January 2021

Published online: 30 January 2021

### References

- McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263–9.
- Margaret G, Reynolds CA, Laura F, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168–74.
- Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452–8.
- Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet*. 2011;43(5):429–35.
- Liu JZ, Erlich Y, Pickrell JK. Case-control association mapping by proxy using family history of disease. *Nat Genet*. 2017;49(3):325–31.
- Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51(3):404–13.
- Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414–30.
- Cuyvers E, Sleegers K. Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol*. 2016;15(8):857–68.
- Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012;22(9):1748–59.
- Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6(4):e1000888.
- Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93(5):779–97.
- Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481–7.
- Pavlidis JMW, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med*. 2016;8(1):84.
- Meng X-H, Chen X-D, Greenbaum J, et al. Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions. *Bone*. 2018;113:41–8.
- Wang X, Cheng W, Zhu J, Yin H, Chang S, Yue W, Yu H. Integrating genome-wide association study and expression quantitative trait loci data identifies NEGR1 as a causal risk gene of major depression disorder. *J Affect Disord*. 2020;265:679–86.
- Yu H, Cheng W, Zhang X, Wang X, Yue W. Integration analysis of methylation quantitative trait loci and GWAS identify three schizophrenia risk variants. *Neuropsychopharmacology*. 2020;10.1038/s41386-41020-40605-41383.
- Ng B, White CC, Klein HU, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci*. 2017;20(10):1418–26.
- Võsa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv*. 2018:447367. <https://www.biorxiv.org/content/10.1101/447367v1>.
- Qi T, Wu Y, Zeng J, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun*. 2018;9(1):2282.
- Kubicek S, Dimond A. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;13(5):307–8.
- Fromer M, Roussos P, Sieberts SK, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci*. 2016;19(11):1442–53.
- Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414–30.
- Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580.
- Gandal MJ, Zhang P, Hadjimiichael E, Walker RL, Chen C, Liu S, Won H, van Bakel H, Varghese M, Wang Y, Shieh AW, Haney J, Parhami S, Belmont J, Kim M, Moran Losada P, Khan Z, Mleczko J, Xia Y, Dai R, Wang D, Yang YT, Xu M, Fish K, Hof PR, Warrell J, Fitzgerald D, White K, Jaffe AE; PsychENCODE Consortium, Peters MA, Gerstein M, Liu C, Iakoucheva LM, Pinto D, Geschwind DH. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):eaat8127.
- Xu M, Zhang DF, Luo R, et al. A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes as important upstream regulators in Alzheimer's disease. *Alzheimers Dement*. 2018;14(2):215–29.
- Sun L, Tan MS, Hu N, Yu JT, Tan L. Exploring the value of plasma BIN1 as a potential biomarker for Alzheimer's disease. *J Alzheimers Dis*. 2013;37(2):291–5.
- Zhang X, Yu JT, Li J, et al. Bridging Integrator 1 (BIN1) genotype effects on working memory, hippocampal volume, and functional connectivity in young healthy individuals. *Neuropsychopharmacology*. 2015;40(7):1794–803.
- Franzmeier N, Rubinski A, Neitzel J, Ewers M. The BIN1 rs744373 SNP is associated with increased tau-PET levels and impaired memory. *Nat Commun*. 2019;10(1):1766.
- Chapuis J, Hansmann F, Gistelinc M, et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol Psychiatry*. 2013;18(11):1225–34.
- De Rossi P, Andrew RJ, Musial TF, et al. Aberrant accrual of BIN1 near Alzheimer's disease amyloid deposits in transgenic models. *Brain Pathology (Zurich, Switzerland)*. 2019;29(4):485–501.
- Holler CJ, Davis PR, Beckett TL, et al. Bridging integrator 1 (BIN1) protein expression increases in the Alzheimer's disease brain and correlates with neurofibrillary tangle pathology. *J Alzheimers Dis*. 2014;42(4):1221–7.
- Voskobiyuk Y, Roth JR, Cochran JN, Rush T, Carullo NV, Mesina JS, Waqas M, Vollmer RM, Day JJ, McMahon LL, Roberson ED. Alzheimer's disease risk gene BIN1 induces Tau-dependent network hyperexcitability. *Elife*. 2020;9:e57354. <https://doi.org/10.7554/eLife.57354>.
- Gao P, Ye L, Cheng H, et al. The mechanistic role of Bridging Integrator 1 (BIN1) in Alzheimer's disease. *Cell Mol Neurobiol*. 2020. <https://doi.org/10.1007/s10571-020-00926-y>.
- Hu Y, Zhao T, Zang T, Zhang Y, Cheng L. Identification of Alzheimer's disease-related genes based on data integration method. *Front Genet*. 2019;9:703. <https://doi.org/10.3389/fgene.2018.00703>.
- Zhao T, Hu Y, Zang T, Wang Y. Integrate GWAS, eQTL, and mQTL data to identify Alzheimer's disease-related genes. *Front Genet*. 2019;10:1021.

### Publisher's Note

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

#### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

