Integrate GWAS, eQTL, and mQTL Data to Identify Alzheimer's Disease-Related Genes

Tianyi Zhao[1], Yang Hu[2], Tianyi Zang[1]*, and Yadong Wang[1]*

[1] Department of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, [2] School of Life Science and Technology, Harbin Institute of Technology, Harbin, China

It is estimated that the impact of related genes on the risk of Alzheimer's disease (AD) is nearly 70%. Identifying candidate causal genes can help treatment and diagnosis. The maturity of sequencing technology and the reduction of cost make genome-wide association study (GWAS) become an important means to find disease-related mutation sites. Because of linkage disequilibrium (LD), neither the gene regulated by SNP nor the specific SNP can be determined. Because GWAS is affected by sample size and interaction, we introduced empirical Bayes (EB) to make a meta-analysis of GWAS to greatly eliminate the bias caused by sample and the interaction of SNP. In addition, most SNPs are in the noncoding region, so it is not clear how they relate to phenotype. In this paper, expression quantitative trait locus (eQTL) studies and methylation quantitative trait locus (mQTL) studies are combined with GWAS to find the genes associated with Alzheimer disease in expression levels by pleiotropy. Summary data-based Mendelian randomization (SMR) is introduced to integrate GWAS and eQTL/mQTL data. Finally, we prioritized 274 significant SNPs, which belong to 20 genes by eQTL analysis and 379 significant SNPs, which belong to seven known genes by mQTL. Among them, 93 SNPs and 2 genes are overlapped. Finally, we did 10 case studies to prove the effectiveness of our method.

Keywords: Alzheimer's disease, Mendelian randomization, GWAS, eQTL, mQTL

INTRODUCTION

It is estimated that the impact of related genes on the risk of AD is nearly 70%. Importantly, neuronal cell death precedes the appearance of cognitive symptoms for 10 years or more, suggesting that targeted treatment needs to be performed before symptoms appear. Therefore, the identification of AD biomarkers such as genes, RNAs (Jiang et al., 2015; Cheng et al., 2018; Cheng et al., 2019), proteins, and metabolites (Cheng et al., 2019) is critical for early detection and early intervention in AD. In addition, identifying candidate genes and loci can also help us understand the pathogenesis of AD and develop drugs.

Recently, Jansen et al. (Jansen et al., 2019) published his AD GWAS study on natural genetics. The sample size is more than eight times that of Lambert et al. (Lambert et al., 2013) in 2013. Due to the increase in the number of samples, they found nine AD risk loci more than in previous studies. Jansen et al. found that most of the AD-related DNA mutations were located in the noncoding part of the genome in regions that affected gene transcription. It means that combining GWAS data with transcriptional expression data will greatly advance AD research (Cheng et al., 2016).

However, GWAS still has certain limitations. The SNP is not necessarily the true pathogenic locus, but only related to the SNP that actually causes the disease due to the LD. GWAS usually analyzes the edge effects of individual loci while ignoring the interaction of multiple genes in complex diseases (Battle et al., 2014). Therefore, GWAS still cannot fully reveal the genetic susceptibility factors of complex diseases (Cheng et al., 2018). It is only an important part of exploring the genetic etiology of complex diseases (Cheng and Hu, 2018). Therefore, using GWAS data for research, we must first start with the expression of SNP, that is, combined with data affecting gene expression, which can weaken the impact of LD on significance. Then, the interaction of multiple genes is considered, that is, the statistical values of each SNP are revised within the whole genome.

It was found that about 80% of the genetic susceptibility loci detected by GWAS were located in the noncoding region of the genome, suggesting that the pathogenic loci may have regulatory functions on gene expression. An important role of large-scale eQTL research is to be able to prioritize SNP loci (Barral et al., 2012) in GWAS susceptible regions and to infer possible biological mechanisms through the influence of DNA polymers on biological characteristics. At present, many studies have used eQTL analysis as a very effective tool to explain the results of GWAS. Hormozdiari et al. (Hormozdiari et al., 2016) present a probabilistic method named eCAVIAR, which can detect target genes by colocalization of GWAS and eQTL signals. Xu et al. purposed a more powerful method based on PrediXcan and TWAS. It can integrate single set or multiple sets of eQTL data with GWAS.

mQTL is mainly based on the analysis of cis-mQTL, that is, using Beta value of methylation level of CpG locus near a gene as dependent variable, screening all SNP variations in the chromosomal region upstream and downstream of the gene as independent variable and regressing each SNP locus S and methylation level M in this region one by one, so as to obtain SNP loci significantly related to the methylation level of a gene. There is no doubt that methylation affects gene expression. This is very similar to eQTL, both of which can cause changes in expression through mutations in a single locus. Therefore, in recent years, more and more studies have been carried out to screen genes related to traits by combining mQTL with GWAS. Hägg et al. (Hägg et al., 2015) integrated GWAS, eQTL, and mQTL to find out genes which are related to obesity. Pharoah et al. (Pharoah et al., 2013) identified three new susceptibility loci for ovarian cancer by GWAS meta-analysis and verified the result by mQTL.

In our previous paper (Hu et al., 2018), we have identified some AD-related genes by GWAS and eQTL using SMR. There are three points to be improved. Firstly, mQTL should be included to verify and improve our result. Secondly, we used several eQTL datasets in that paper, whereas a meta-analysis method should be used to integrate the datasets, which can improve the accuracy of eQTL's statistical results. Finally, GWAS datasets should also be integrated into one dataset so that can overcome the difference of statistical power caused by sample size.

METHODS

SMR

Since Zhu et al. proposed "SMR" in 2016, it has become a common way to identify the genes whose expression levels are associated with a complex trait because of pleiotropy. Using GWAS and eQTL data, SMR could screen trait-related genes. After two years, they applied SMR to mQTL data. They found 7,858 DNAm sites which are related to 14 complex traits.

The basic idea of this method is as follows. First, let y be the phenotype, which is the outcome variable. x is the gene expression, which is the exposure factor. z is the gene mutation, which is the instrumental variable. Then, b_{xy} is the effect of x on y, b_{zx} is the effect of z on x, and b_{zy} is the effect of z on y. The definition of b_{xy} is $b_{xy} = b_{zy}/b_{zx}$, which means the effect of gene expression on phenotype without confounding factors. This idea is based on the Mendelian randomization (Cheng et al., 2018; Cheng et al., 2019).

Figure 1 is a hypothetical model of a mediation mechanism tested in SMR. The blue line represents causal relationship. Methylation will cause SNP. Both SNP and methylation can affect the change of transcription. The change of transcription will cause the difference of trait. The red line denotes the relationship data represents. mQTL denotes the relationship between methylation and SNP. eQTL denotes the relationship between transcription and SNP. GWAS denotes the relationship between SNP and trait.

Based on this hypothesis, many researchers have found the genes which are related to certain traits. Diseases like bone mineral density (BMD) (Meng et al., 2018), amyotrophic lateral sclerosis (ALS) (Du et al., 2017), and neuroticism (Fan et al., 2017) have been found some potential related genes by SMR. Other traits like height, BMI (Yengo et al., 2018), and obesity (Liu et al., 2018) have also researched by SMR.

Eb-GWAS

Due to the complex linkage effects and statistical errors of the samples, the contribution of GWAS to biological research is reduced. GWAS may associate common diseases with thousands of DNA mutations, that is, every DNA region that happens to be active in diseased tissues may be associated with disease (Jiang et al., 2013). Many GWAS matches are not specifically biologically related to disease and, therefore, cannot be used as effective drug targets. In fact, these "peripheral" mutations are likely to affect the activity of "core" genes, which are more directly related to disease, through complex biochemical regulatory networks (Jiang et al., 2010).

As we discussed before in the introduction, the interaction of multiple genes is considered, that is, the statistical values of each SNP are revised within the whole genome. In this section, we will

FIGURE 1 | A hypothetical model of a mediation mechanism.
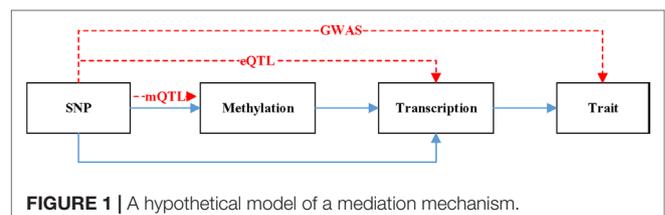
process GWAS data in two steps: 1. meta-analysis, 2. using EB, revise the statistical value of each SNP within the whole genome.

Meta-Analysis

Since SE denotes the standard error of each SNP, it represents the reliability of Beta values. Then, weight of each Beta should be:

$$w_i = 1 / SE_i^2 \qquad (1)$$

SE_i denotes the standard error for study i, w_i denotes the weight of Beta.

Then, the Beta after meta-analysis would be:

$$\beta = \sum_i \beta_i w_i / \sum_i w_i \qquad (2)$$

β_i denotes effect size estimate for study i.

Then, we could use the weight of each Beta to calculate the result of meta-analysis.

$$SE = \sqrt{1 / \sum_i w_i} \qquad (3)$$

Finally, the overall Z-score could be obtained by the original equation.

$$Z = \beta / SE \qquad (4)$$

Eb-GWAS

After meta-analysis, we could summary several GWAS datasets into one dataset. Then, we used EB to integrate all the Z scores in the whole genomic level. As we know that the SNP could interact with each other, the Z score of all SNP should have some relationship and obey normal distribution.

The overall Z-score we obtained before obeying normal distribution with standard deviation is 1. Then,

$$\hat{Z}_i \mid Z_i \overset{ind}{\sim} N(\hat{Z}_i, 1) \qquad (5)$$

$\hat{Z}_i$ denotes the Z score we obtained. It is a value with bias. Z_i denotes the real Z score.

Real Z score obeys normal distribution:

$$Z \overset{ind}{\sim} N(\theta, \sigma^2) \qquad (6)$$

Then, the marginal distribution of $\hat{Z}_i$ is

$$\hat{Z} \overset{ind}{\sim} N(\theta, \sigma^2 + 1) \qquad (7)$$

Moreover, the posterior distribution should be:

$$Z_i \mid \hat{Z}_i \overset{ind}{\sim} N(\theta + B(\hat{Z}_i - \theta), B) \qquad (8)$$

$$B = \frac{\sigma^2}{1 + \sigma^2} \qquad (9)$$

Then, we could know that $E(\hat{Z}_i) = \theta$, so the mean of $\hat{Z}_i$ can be used to estimate θ.

$$\hat{\theta} = mean(\hat{Z}_i) = \bar{\hat{Z}}_i \qquad (10)$$

$$\frac{\sum_i^N (\hat{Z}_i - \bar{\hat{Z}}_i)^2}{\sigma^2 + 1} = \frac{S}{\sigma^2 + 1} \sim \chi^2(N-1) \qquad (11)$$

Then,

$$\frac{\sigma^2 + 1}{S} \sim inverse - \chi^2(N-1) \qquad (12)$$

From the properties of inverse chi-square distribution,

$$E(\frac{\sigma^2 + 1}{S}) \sim \frac{1}{N-3} \qquad (13)$$

Then,

$$E(\frac{N-3}{S}) = \frac{1}{\sigma^2 + 1} = 1 - B \qquad (14)$$

Therefore, the EB estimation of B is

$$B = 1 - \frac{(N-3)}{S} \qquad (15)$$

Finally, we can put the (Hu et al., 2018) into (Battle et al., 2014)

$$Z_i = \bar{\hat{Z}} + (1 - \frac{(N-3)}{S})(\hat{Z}_i - Z) \qquad (16)$$

Then, we have done the meta-analysis and revised the statistical value of each SNP within the whole genome.

Dataset

As shown in **Table 1** we obtained five GWAS datasets, three eQTL dataset, and three mQTL datasets. All the eQTL and mQTL are from brain tissue. Yang Jian et al. have already meta-analysis the eQTL and mQTL datasets. Therefore, we used the data they processed.

For GWAS dataset, Scelsi M A et al. obtained the data from 1,517 Caucasian ADNI subjects. Lambert JC et al.'s dataset is

TABLE 1 | Datasets used in this paper.

Data	Name	Reference
GWAS	ADNI_DPS_GWAS	Scelsi et al. (2018)
	ADNI_amyloid_GWAS	(include three datasets)
	ADNI_hippo_GWAS	
	IGAP_stage_1	Lambert et al. (2013)
	UK_Biobank	Marioni et al. (2018)
eQTL	GTEx-brain eQTL	GTEx Consortium (2017)
	CMC	Fromer et al. (2016)
	ROSMAP	Ng et al. (2017)
mQTL	ROSMAP	Ng et al. (2017)
	Human fetal brain	Hannon et al. (2016)
	Frontal cortex	Jaffe et al. (2016)

consisted of 17,008 Alzheimer's disease cases and 37,154 controls. Marioni R E et al. obtained data from 314,278 participants.

For eQTL dataset, SNPs within 1Mb distance from each probe are available in these three datasets. After meta-analysis, the estimated effective sample size n = 1194.

For mQTL dataset, 5kb, 500kb, and 20kb are the available distance for the three datasets, respectively. After meta-analysis, the estimated effective sample size n = 1160.

RESULTS

Results of GWAS Meta-Analysis

We did a meta-analysis of five groups of GWAS data and integrated them into a GWAS file.

The blue block in **Figure 2** is P value density of GWAS after meta-analysis. The red block in **Figure 2** is P value density of GWAS after EB. As we can see in **Figure 2**, the distribution approximates uniform distribution. After using EB in all SNPs in whole dataset, the P value of the final GWAS data approximates the normal distribution.
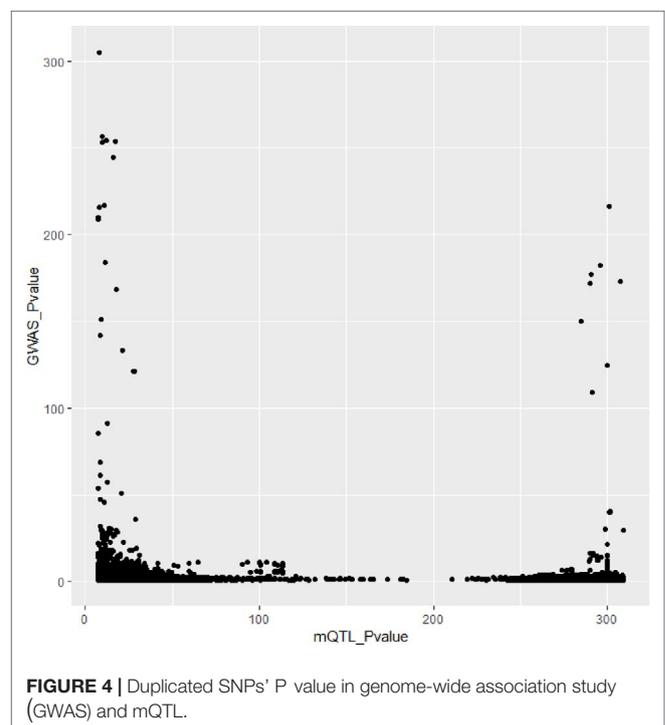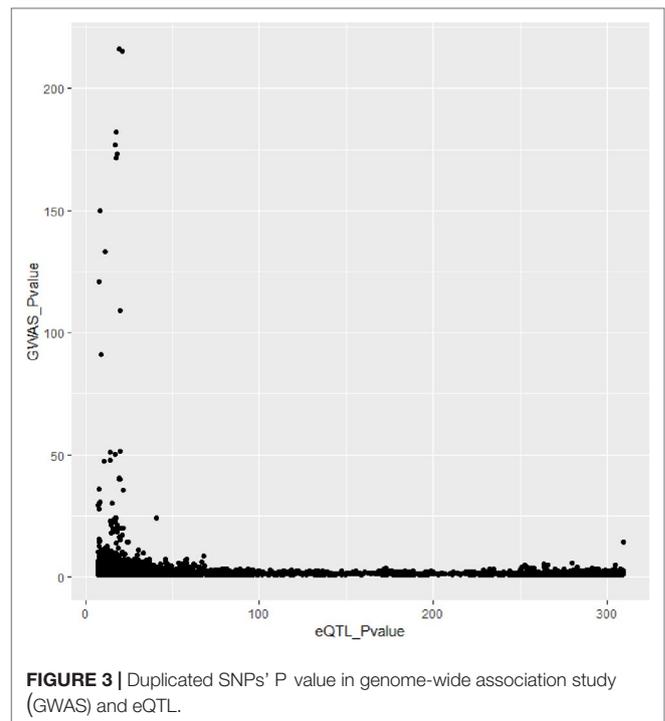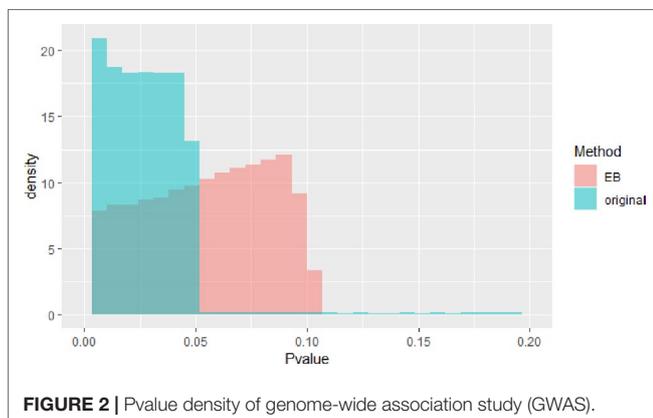
Results of SMR

GWAS included 1,474,846 SNPs, mQTL included 6,966,746, and eQTL included 1,067,443 SNPs. There are 149,326 SNPs occur in both GWAS and eQTL and 408,896 SNPs occur in both GWAS and mQTL. Therefore, we use SMR to test these repeated SNPs in data sets.

Note that some SNPs are marked by multiple probes, so one SNP may significant in more than one gene. One SNP may affect expression of multiple genes.

In **Figures 3** and **4**, we can see that SNPs' P value in GWAS are not related to eQTL and mQTL. It means that only few significant SNPs in GWAS have significance in eQTL and mQTL. Anyway, the points near the upper right corner in the images mean that the difference in expression level caused by these SNPs is related to AD and SMR can help us detect these SNPs.

We set a threshold as 0.05/(number of probers). For eQTL data, the threshold is 0.05/8362 = 5.98e-06. For mQTL data, the threshold is 0.05/97263 = 5.14e-07. The numbers of SNPs and genes identified by the two experiments are shown in **Table 2**.

FIGURE 2 | Pvalue density of genome-wide association study (GWAS).

FIGURE 3 | Duplicated SNPs' P value in genome-wide association study (GWAS) and eQTL.

FIGURE 4 | Duplicated SNPs' P value in genome-wide association study (GWAS) and mQTL.

Figure 5 shows all the SNPs' P value. The red points are the P value of GWAS SNPs. The blue points are the P value of eQTL SNPs and the green points are the P value of mQTL SNPs. There is a black line in the first picture. The line is the significant threshold of P value. It is -log10(5*10-8). The SNPs of eQTL and mQTL are already screened so each SNP's P value is less than 5*10-8.

TABLE 2 | The results of summary data-based Mendelian randomization (SMR).

Dataset	Number of SNPs	Number of Genes
GWAS&eQTL	274	20
GWAS&mQTL	379	7
Overlapped	93	2

Figure 6 shows the result of SMR by two different datasets. The first graph is the result of GWAS and eQTL and the second one is the result of GWAS and mQTL. The black line in the two graphs is significant threshold, respectively. As we can see, only few of SNPs can pass the SMR test. Some of them are not very significant in GWAS, but combined with eQTL or mQTL, they would be significant.

As we can see in **Table 3**, HLA-DQA1 and HLA-DRB5 are selected in both eQTL and mQTL datasets. The HLA complex is located in the 21.31 region (6p21.31) on the short arm of chromosome 6 and is composed of 3.6 million base pairs. It is the region with the highest gene density and the most polymorphic region in human chromosomes. Known as "chemical fingerprints in humans". Due to the complexity of HLA, the methylation level and expression level differ greatly.

Case Study

In this section, we want to confirm whether the 25 AD-related genes we found have been reported by others. In order to be precise, we only use the literature that got AD-related genes by biological experiments, rather than the bioinformatics method or GWAS method.

Zhu et al. (2017) found four CR1 SNPs showed significant associations with the Aβ deposition at the baseline level.

James et al. (2018) gathered 71 cognitively healthy women's the volumes of total gray matter, cerebrocor-tical gray matter, and subcortical gray matter by structural magnetic resonance imaging
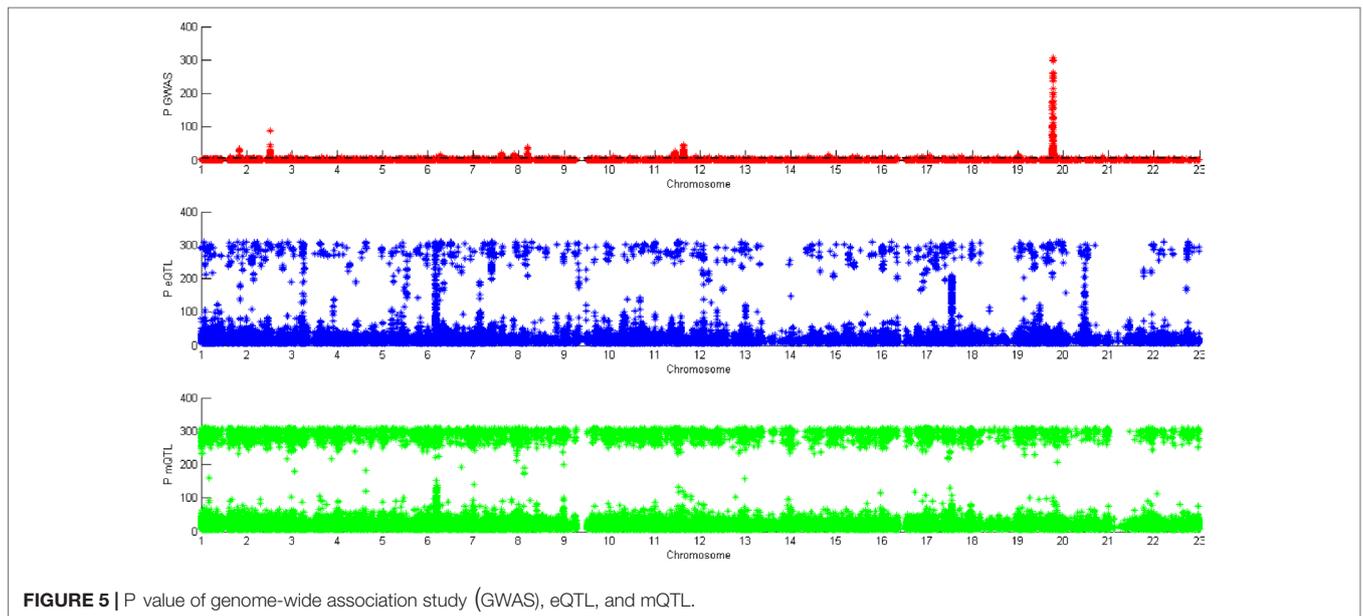
FIGURE 5 | P value of genome-wide association study (GWAS), eQTL, and mQTL.
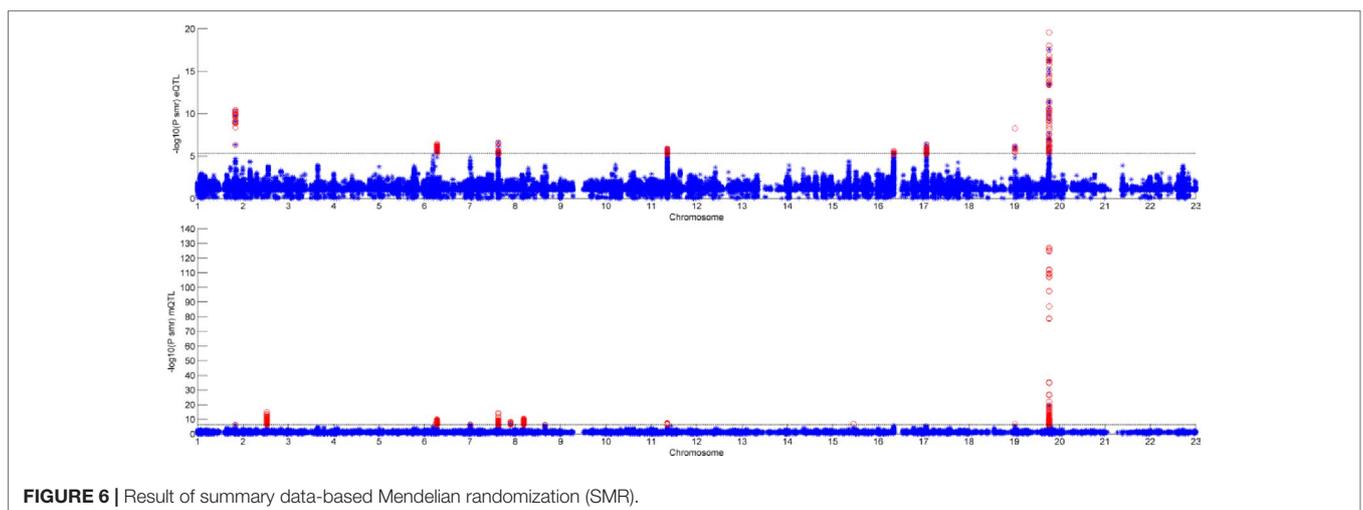
FIGURE 6 | Result of summary data-based Mendelian randomization (SMR).

TABLE 3 | The candidate genes selected by summary data-based Mendelian randomization (SMR).

	Gene	Number of SNPs
eQTL	CR1	20
	HLA-DRB1	69
	HLA-DQA1	39
	HLA-DRB5	8
	HLA-DQB1	3
	HLA-DQB1-AS1	1
	RP11-385F7.1	36
	ZSCAN21	8
	PILRB	5
	PILRA	5
	MTCH2	20
	KAT8	20
	AC012146.7	23
	ZNF232	4
	POLR2E	7
	PVR	12
	CTB-171A8.1	24
	CEACAM19	11
	TOMM40	23
	ZNF296	6
mQTL	BIN1	11
	HLA-DRB5	15
	HLA-DRB1	16
	EPHA1-AS1	3
	FAM63B	2
	APOC1	12
	EXOC3L2	24

(sMRI) scan and found that the protective effect of DRB1*13:02 is related to successful elimination of specific pathogens that would ultimately cause gradual brain atrophy.

Yu et al. (2015) found that BIN1 was associated with Aβ load and brain DNA methylation in HLA-DRB5 was associated with pathological AD by 447 participants

Lee et al. (2018) used non-Hispanic Caucasians with neuroimaging and found that HLA-DQB1 is significantly associated with entorhinal cortical thickness by controlling for multiple testing.

Yoshino et al. (2016) found that SNCA mRNA expression in 50 AD subjects was significantly higher than that in control subjects. Therefore, they inferred mRNA expression and methylation of SNCA intron 1 are altered in AD, whereas ZSCAN21 at upstream of these CpG site were reported to bind at intron 1.

Rathore et al. (2018) noted that both TREM2 and PILRB function as activating receptors and signal through DAP12. A reduction of PILRA inhibitory signals in R78 carriers could allow more microglial activation via PILRB/DAP12 signaling and reinforce the cellular mechanisms by which TREM2 is believed to protect from AD incidence.

Ruggiero et al. (2017) did biological experiments on mice and found that MTCH2 is a critical player in neuronal cell biology, controlling mitochondria metabolism, motility, and calcium buffering to regulate hippocampal-dependent cognitive functions.

De Jager et al. (2014) used a collection of 708 prospectively collected autopsied brains to assess the methylation state of the brain's DNA in relation to AD and found two SNPs associated with POLR2E are related to AD in methylation levels.

Roses et al. (2010) identified polymorphic poly-T variant rs10524523 in transposase of TOMM40 gene, which can be used to estimate the starting age of LOAD with APOE ε3 carriers.

Prendecki et al. (2018) recruited 230 individuals and found that APOC1 and TOMM40 rs2075650 polymorphisms may be independent risk factors of developing AD, whose major variants are accompanied by disruption of biothiols metabolism and inefficient removal of DNA oxidation.

We found 10 of 25 genes are reported to be related to AD by biological experiments. Some literary works may found that the other 15 genes are related to AD via other methods, but we would not discuss in this paper. This case study verified the effectiveness of our method and we hope the other 15 genes could be verified by biological experiments in future.

CONCLUSION

AD brings great burden to patients and society and identifying AD-related genes can help us known the machanism of AD then diagnose and treatment. In this paper, we used SMR to find AD-related genes by GWAS, eQTL, and mQTL. There are some overlaps between GWAS and the other two datasets, which means that some SNPs are related to AD due to the change of expression level. SMR is a method which can identify the genes whose expression levels are associated with a complex trait because of pleiotropy.

Due to the LD and interaction between genes, GWAS data has bias. In order to overcome these, we did meta-analysis on five GWAS datasets and then used EB to revise the Z-score of each SNPs in whole-SNP level.

Finally, we found 653 SNPs reached the threshold of significance and they are associated with 25 genes. Ninety-three of SNPs are significant in both GWAS&eQTL and GWAS&mQTL tests. We did 10 case studies at last, which means that the 10 of 25 genes we identified have been verified to correlated to AD by biological experiments in existing literary works.

DATA DEPOSITION

eQTL and mQTL Data
The direct link for accessing eQTL and mQTL data is as follows (origin from PMID: 29891976).

1) eQTL data: https://cnsgenomics.com/data/SMR/Brain-eMeta.tar.gz
2) mQTL data: https://cnsgenomics.com/data/SMR/Brain-mMeta.tar.gz

GWAS Dataset 1,2,3
GWAS dataset 1,2,3 are from paper PMID:29860282. The direct link is for accessing them is as following.

1) https://www.ebi.ac.uk/gwas/studies/GCST006134 & ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/ScelsiMA_29860282_GCST006134

2) https://www.ebi.ac.uk/gwas/studies/GCST006136 & ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/ScelsiMA_29860282_GCST006135

3) https://www.ebi.ac.uk/gwas/studies/GCST006135 & ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/ScelsiMA_29860282_GCST006136

GWAS Data 4

GWAS data 4 is from PMID: 24162737. The direct link is for accessing it is as following:

http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php

GWAS Data 5

GWAS data 5 is from PMID: 29777097. The direct link is for accessing it is as following:

http://datashare.is.ed.ac.uk/download/DS_10283_3364.zip

All code could be downloaded by
https://github.com/zty2009/Integrate-GWAS-eQTL-and-mQTL-data-to-identify-Alzheimer-s-Disease-related-genes

AUTHOR CONTRIBUTIONS

TZang and YW are the corresponding authors. They help to revise and support data for this data. TZhao and YH are the co-first authors. They wrote the code and write the paper.

FUNDING

REFERENCES

Barral, S., Bird, T., Goate, A., Farlow, M., Diaz-Arrastia, R., Bennett, D., et al. (2012). Genotype patterns at PICALM, CR1, BIN1, CLU, and APOE genes are associated with episodic memory. *Neurology* 78, 1464–1471. doi: 10.1212/WNL.0b013e3182553c48

Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C., et al. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* 24, 14–24. doi: 10.1101/gr.155192.113

Cheng, L., and Hu, Y. (2018). Human Disease System Biology. *Curr. Gene. Ther.* 18, 255–256. doi: 10.2174/1566523218666181010101114

Cheng, L., Sun, J., Xu, W. Y., Dong, L. X., Hu, Y., and Zhou, M. (2016). OAHG: an integrated resource for annotating human genes with multi-level ontologies. *Sci. Rep.* 6, 1–9. doi: 10.1038/srep34820

Cheng, L., Zhuang, H., Yang, S., Jiang, H., Wang, S., and Zhang, J. (2018). Exposing the causal effect of C-reactive protein on the risk of type 2 diabetes mellitus: a mendelian randomization study. *Front. Genet.* 9, 657. doi: 10.3389/fgene.2018.00657

Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018). DincRNA: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. *Bioinformatics* 34, 1953–1956. doi: 10.1093/bioinformatics/bty002

Cheng, L., Jiang, Y., Ju, H., Sun, J., Peng, J., Zhou, M., et al. (2018). InfAcrOnt: calculating cross-ontology term similarities using information flow by a random walk. *BMC Genomics* 19, 919. doi: 10.1186/s12864-017-4338-6

Cheng, L., Yang, H., Zhao, H., Pei, X., Shi, H., Sun, J., et al. (2019). MetSigDis: a manually curated resource for the metabolic signatures of diseases. *Brief Bioinform.* 20, 203–209. doi: 10.1093/bib/bbx103

Cheng, L., Zhuang, H., Ju, H., Yang, S., Han, J. W., Tan, R. J., et al. (2019). Exposing the causal effect of body mass index on the risk of type 2 diabetes mellitus: a mendelian randomization study. *Front. Genet.* 10, 10. doi: 10.3389/fgene.2019.00094

Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2019). LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* 47, D140–D144. doi: 10.1093/nar/gky1051

Consortium, G. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204. doi: 10.1038/nature24277

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., et al. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* 17, 1156. doi: 10.1038/nn.3786

Du, Y., Yan, W., Guo, X., Hao, J., Wang, W., He, A., et al. (2017). and Pathways Associated with Amyotrophic Lateral Sclerosis. *Cell. Mol. Neurobiol.* 38, 1–5. doi: 10.1007/s10571-017-0512-2

Fan, Q., Wang, W., Hao, J., He, A., Wen, Y., Guo, X., et al. (2017). Integrating genome-wide association study and expression quantitative trait loci data identifies multiple genes and gene set associated with neuroticism. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 78, 149–152. doi: 10.1016/j.pnpbp.2017.05.017

Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* 19, 1442. doi: 10.1038/nn.4399

GTEx Consortium. (2017). Genetic effects on gene expression across human tissues. *Nature* 550 (7675), 204.

Hägg, S., Ganna, A., Van Der Laan, S. W., Esko, T., Pers, T. H., Locke, A. E., et al. (2015). Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. *Hum. Mol. Genet.* 24, 6849–6860. doi: 10.1093/hmg/ddv379

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., et al. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* 19, 48. doi: 10.1038/nn.4182

Hormozdiari, F., Vandebunt, M., Segrè, A., Li, X., Joo, J. W., Bilow, M., et al. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* 99, 1245–1260. doi: 10.1016/j.ajhg.2016.10.003

Hu, Y., Zhao, T., Zang, T., Zhang, Y., and Cheng, L. (2018). Identification of Alzheimer's disease-related genes based on data integration method. *Front. Genet.* 9, 703. doi: 10.3389/fgene.2018.00703

Jaffe, A. E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T. M., Weinberger, D. R., et al. (2016). genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* 19, 40. doi: 10.1038/nn.4181

James, L. M., Christova, P., Lewis, S. M., Engdahl, B. E., Georgopoulos, A., and Georgopoulos, A. P. (2018). Protective effect of human leukocyte antigen (HLA) Allele DRB1* 13: 02 on age-related brain gray matter volume reduction in healthy women. *EBioMedicine* 29, 31–37. doi: 10.1016/j.ebiom.2018.02.005

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., et al. (2019). Genome-wide meta-analysis identifies new loci and functional

pathways influencing Alzheimer's disease risk. *Nat. Genet.* 51, 404–413. doi: 10.1038/s41588-018-0311-9

Jiang, Q., Hao, Y., Wang, G., Juan, L., Zhang, T., Teng, M., et al. (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* 4 Suppl 1, S2. doi: 10.1186/1752-0509-4-S1-S2

Jiang, Q., Wang, G., Jin, S., Li, Y., and Wang, Y. (2013). Predicting human microRNA-disease associations based on support vector machine. *Int. J. Data Min. Bioinform.* 8, 282–293. doi: 10.1504/IJDMB.2013.056078

Jiang, Q., Ma, R., Wang, J., Wu, X., Jin, S., Peng, J., et al. (2015). LncRNA2Function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data. *BMC Genomics* 16 Suppl 3, S2. doi: 10.1186/1471-2164-16-S3-S2

Lambert, J.-C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 45, 1452. doi: 10.1038/ng.2802

Lee, Y., Han, S., Kim, D., Kim, D., Horgousluoglu, E., Risacher, S. L., et al. (2018). Genetic variation affecting exon skipping contributes to brain structural atrophy in Alzheimer's disease. *AMIA Summits on Translat. Sci. Proc.* 2017, 124.

Liu, L., Fan, Q., Zhang, F., Guo, X., Liang, X., Du, Y., et al. (2018). A Genomewide Integrative Analysis of GWAS and eQTLs Data Identifies Multiple Genes and Gene Sets Associated with Obesity. *Biomed. Res. Int.* 2018. 1–5 doi: 10.1155/2018/3848560

Marioni, R. E., Harris, S. E., Zhang, Q., McRae, A. F., Hagenaars, S. P., Hill, W. D., et al. (2018). GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* 8 (1), 99. doi: 10.1038/s41398-018-0150-6

Meng, X. H., Chen, X. D., Greenbaum, J., Zeng, Q., You, S. L., Xiao, H. M., et al. (2018). Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions. *Bone* 113, 41–48. doi: 10.1016/j.bone.2018.05.012

Ng, B., White, C. C., Klein, H.-U., Sieberts, S. K., McCabe, C., Patrick, E., et al. (2017). An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* 20, 1418. doi: 10.1038/nn.4632

Pharoah, P. D., Tsai, Y.-Y., Ramus, S. J., Phelan, C. M., Goode, E. L., Lawrenson, K., et al. (2013). GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.* 45, 362. doi: 10.1038/ng.2564

Prendecki, M., Florczak-Wyspianska, J., Kowalska, M., Ilkowski, J., Grzelak, T., Bialas, K., et al. (2018). Biothiols and oxidative stress markers and polymorphisms of TOMM40 and APOC1 genes in Alzheimer's disease patients. *Oncotarget* 9 (81), 35207. doi: 10.18632/oncotarget.26184

Rathore, N., Ramani, S. R., Pantua, H., Payandeh, J., Bhangale, T., Wuster, A., et al. (2018). Paired immunoglobulin-like type 2 receptor alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. *PLoS Genet.* 14 (11), e1007427. doi: 10.1371/journal.pgen.1007427

Roses, A., Lutz, M., Amrine-Madsen, H., Saunders, A., Crenshaw, D., Sundseth, S., et al. (2010). A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J.* 10, 375. doi: 10.1038/tpj.2009.69

Ruggiero, A., Aloni, E., Korkotian, E., Zaltsman, Y., Oni-Biton, E., Kuperman, Y., et al. (2017). Loss of forebrain MTCH2 decreases mitochondria motility and calcium handling and impairs hippocampal-dependent cognitive functions. *Sci. Rep.* 7, 44401. doi: 10.1038/srep44401

Scelsi, M. A., Khan, R. R., Lorenzi, M., Christopher, L., Greicius, M. D., Schott, J. M., et al. (2018). Genetic study of multimodal imaging Alzheimer's disease progression score implicates novel loci. *Brain* 141, 2167–2180. doi: 10.1093/brain/awy141

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Hum. Mol. Genet.* 27 (20), 3641–3649. doi: 10.1101/274654

Yoshino, Y., Mori, T., Yoshida, T., Yamazaki, K., Ozaki, Y., Sao, T., et al. (2016). Elevated mRNA expression and low methylation of SNCA in Japanese Alzheimer's disease subjects. *J. Alzheimer's Dis.* 54, 1349–1357. doi: 10.3233/JAD-160430

Yu, L., Chibnik, L. B., Srivastava, G. P., Pochet, N., Yang, J., Xu, J., et al. (2015). Association of Brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol.* 72, 15–24. doi: 10.1001/jamaneurol.2014.3049

Zhu, X.-C., Wang, H.-F., Jiang, T., Lu, H., Tan, M.-S., Tan, C.-C., et al. (2017). Initiative, Effect of CR1 genetic variants on cerebrospinal fluid and neuroimaging biomarkers in healthy, mild cognitive impairment and Alzheimer's disease cohorts. *Mol. Neurobiol.* 54, 551–562. doi: 10.1007/s12035-015-9638-8