frontiers
in Genetics

Identification of Alzheimer's Disease-Related Genes Based on Data Integration Method

Yang Hu[1], Tianyi Zhao[1], Tianyi Zang[1]*, Ying Zhang[2]* and Liang Cheng[3]*

[1] Department of Computer Science and Technology, School of Life Science and Technology, Harbin Institute of Technology, Harbin, China, [2] Department of Rehabilitation, Heilongjiang Province Land Reclamation Headquarters General Hospital, Harbin, China, [3] Department of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

Alzheimer disease (AD) is the fourth major cause of death in the elderly following cancer, heart disease and cerebrovascular disease. Finding candidate causal genes can help in the design of Gene targeted drugs and effectively reduce the risk of the disease. Complex diseases such as AD are usually caused by multiple genes. The Genome-wide association study (GWAS), has identified the potential genetic variants for most diseases. However, because of linkage disequilibrium (LD), it is difficult to identify the causative mutations that directly cause diseases. In this study, we combined expression quantitative trait locus (eQTL) studies with the GWAS, to comprehensively define the genes that cause Alzheimer disease. The method used was the Summary Mendelian randomization (SMR), which is a novel method to integrate summarized data. Two GWAS studies and five eQTL studies were referenced in this paper. We found several candidate SNPs that have a strong relationship with AD. Most of these SNPs overlap in different data sets, providing relatively strong reliability. We also explain the function of the novel AD-related genes we have discovered.

Keywords: Alzheimer disease, SNPs, mendelian randomization, GWAS, eQTL

INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease and has become the fourth major cause of death after cardiovascular disease, malignant tumor and stroke in the elderly (Senova et al., 2018). The cause of Alzheimer's disease remains difficult to explain. It is estimated that the impact of related genes, on the risk of AD, is nearly 70% (Cruchaga et al., 2012). Therefore, screening for candidate genes and loci associated with AD is of great significance to understand the pathogenesis and treatment of AD.

At present, the Genome-wide association study (GWAS) is widely used to identify various neurological susceptibility genes. The use of the GWAS analysis to find the risk genes of AD was only started 10 years ago (Visscher et al., 2012). At the beginning of this century, many research groups identified the susceptible locus of AD, but the actual results were not satisfactory. The only susceptibility genes found by these different research groups was SOL1 (Meng et al., 2007). The inconsistency of these research results were mainly due to the heterogeneity of the experimental samples, the complex linkage disequilibrium patterns, differences in the allele frequencies, and the size of the samples (Malhotra et al., 2012). Over the past 10 years, the emergence of high-throughput sequencing technology has allowed researchers to simultaneously detect millions of Single Nucleotide Polymorphisms (SNPs) on the genome. Through the efforts of large organizations and companies, high-throughput sequencing technologies have led to the discovery of many new pathways and susceptibility genes for AD in recent years. Researchers from

the GWAS first identified four susceptible gene loci, CLU, PICALM, CRL, and BIN1 (Kauwe et al., 2011; Barral et al., 2012). Later, other research groups (Hollingworth et al., 2011; Naj et al., 2011) found susceptibility loci such as CD33 and EPHA1, using a larger sample of the GWAS analysis. The experimental samples of these studies were mainly from Europe. The most significant SNP population attribution scores, of the above-mentioned susceptibility loci, are between 1 and 8%, and the odds ratio (OR) varies from 1.16 to 1.20. In addition, Hillburns et al. (2014) linked neurological Parkinson's disease within the GWAS and found a range of risk locus. Kauwe et al. (2008) used the GWAS for AD research and found that risk genes such as ACE and MMP3 are associated with Aβ amyloid and Tau protein levels in cerebrospinal fluid.

However, limitations in the GWAS still remain. For example, the strategy is based on the hypothesis of "disease-common variations," which misses rare variants [minor allele frequency (MAF < 0.005)] that may play a more important role in the cause of diseases; SNPs are not necessarily true pathogenic locus, but only "tags" SNPs that are associated with LD in real disease-causing locus. In particular, signals located in the so-called "desert regions" of genes make it difficult to elucidate the biological functions of genetic variation. The GWAS usually analyzes the marginal effects of a single locus, based on the most obvious principle of statistical discrepancies, while ignoring the interaction of multiple genes in complex diseases (Battle et al., 2014). Therefore, the GWAS still cannot fully reveal the genetic susceptibility factors of complex diseases, which is an important part of exploring genetic etiology mechanisms of complex diseases. How to mine the susceptible locus of the GWAS and how find the true pathogenic locus, as well as how to explore the biological mechanism of these non-coding sequences, is a major challenge in genetic research.

Expression quantitative trait loci (eQTL) are quantitative traits based on the expression level of the gene mRNA. Traditional QTL methods are used to locate the genetic loci and to locate the genetic locus of the target gene expression. This mapping method requires measuring the genotype and the gene expression levels for each individual, and then compares the association between the genotype and gene expression levels using an association analysis (outcrossing population) or a linkage analysis (pedigree or experimental hybridization population) (James Ronald, 2007; Clément-Ziza et al., 2015). It was found that about 80% of the genetic susceptibility loci detected by the GWAS were located in the non-coding region of the genome, suggesting that the pathogenic loci may have regulatory functions on gene expression. Nicolae et al. (Dan et al., 2010) compared the SNPs, found in the common complex diseases of the GWAS, with other random sampling SNPs with same typing platform matching the allelic frequency distribution, and found that the former clearly contained more eQTLs (Cheng et al., 2016, 2018a,b). Therefore, an important role of the large-scale eQTL research is to be able to prioritize the screening of possible pathogenic sites among SNP loci in the GWAS susceptible regions (Cheng et al., 2018c; Hu et al., 2018), and to speculate the possible biological mechanism through the impact of DNA polymorphism on biological traits. At present, many studies have used eQTL analysis as a very

effective tool (Libioulle et al., 2007; Moffatt et al., 2007) to interpret the results of the GWAS. By increasing the sample size, the problem of low statistical efficiency caused by the small sample size in the past, has been gradually improved (Albert and Kruglyak, 2015), and the number of eQTLs found has significantly increased.

Therefore, if we combine the GWAS with eQTL, we will find many genes and loci related to diseases. In this paper, we introduce the "Summary Mendelian Randomization, SMR" method. This method can find disease-related SNPs from the summary level of data. SMR is first proposed in the paper by Zhu et al. (2016) as the "Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets." They used this method to identify several genes which are related to five complex traits. Since the paper was published, many researchers use SMR to identify disease-related SNPs. Pavlides et al. (2016) used the SMR in 28 GWAS datasets to identify genes with expression levels associated with traits and diseases due to pleiotropy or causality. Meng et al. (2018) applied the SMR to identify bone mineral density (BMD) related genes. Du et al. (2018) also identified genes and pathways associated with Amyotrophic Lateral Sclerosis using the SMR. Fan et al. (2017) identified six genes associated with neuroticism using the SMR. Yengo et al. (2018) identified an enrichment of eQTLs amongst lead height and BMI signals, prioritizing 684 and 134 genes, respectively, using the SMR. Liu et al. (2018) used the SMR for research on Obesity and found 20 BMI associated genes. Veturi and Ritchie (2018) compared two popular methods: The MP and SMR using different datasets. Looking at these studies and the evidence, we concluded that the SMR is an effective tool.

METHODS

SMR

The GWAS has identified thousands of mutations associated with various traits and diseases. However, due to the complex linkage effects and statistical errors of the samples, the mechanisms and effects of these mutations on diseases remain unknown. If the level of a gene expression is affected by mutation, the level of gene expression will be different among individuals of different genotypes, therefore, if the level of expression of a gene affects the disease, then different genotypes will have different phenotypic effects on the disease. The idea is therefore very similar to Mendelian Randomization (MR), where SNPs can serve as instrumental variables to explore the association between genes and diseases. Therefore the new work flow of generalization of Mendel's randomized method is as **Figure 1**.

eQTL refers to regions on chromosomes that specifically regulate the expression of the mRNAs and proteins. The expression level of the mRNAs or proteins is proportional to the quantitative traits. eQTLs can be divided into cis-eQTLs and trans- eQTLs. Cis-eQTLs are the eQTLs of a gene that are located in the genomic region of the gene, indicating changes in mRNA levels that may be caused by differences in the gene itself; trans-eQTLs are the eQTLs of a gene that are located in other genomic regions, indicating other genes. These genetic differences control the different mRNA levels of the gene. The corresponding data of

FIGURE 1 | Generalization of Mendel's randomized method.

FIGURE 2 | Work flow of SMR.
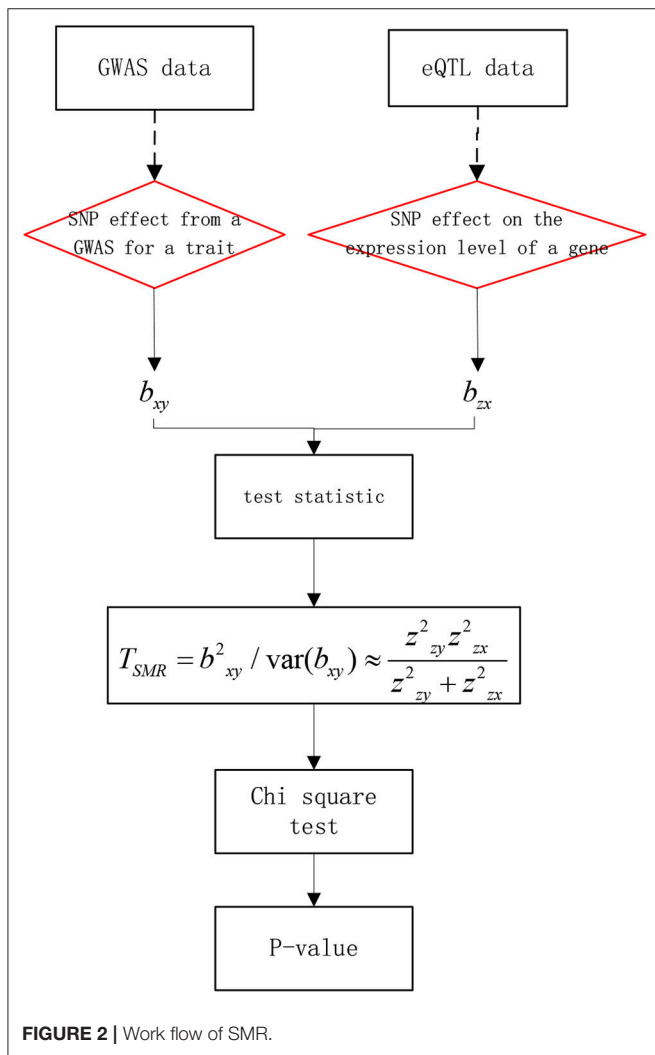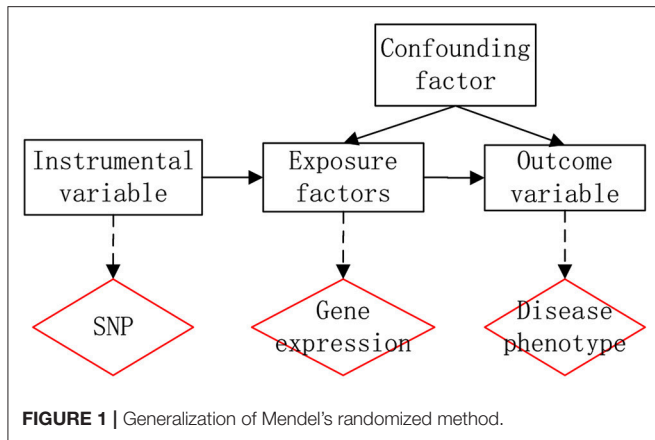
gene expression levels of individuals with different genotypes will then be obtained with eQTL.

The basic ideas are as follows: first, set y as the phenotype (output variable), x as the expression of gene (exposure factor),

and z as the gene mutation (instrumental factor). Then, b_{xy} is the effect of x on y. b_{zx} is the effect of z on x. b_{zy} is the effect of z on y. finally, b_{xy} is defined as $b_{xy} = b_{zy}/b_{zx}$, that is the effect of gene expression on the phenotype without confounding factors. The work flow of SMR is shown in **Figure 2**. First, b_{xy} can be obtained using the GWAS, as GWAS data can reflect the SNP effect on the trait. Then, b_{zx} can be obtained by the eQTL, as eQTL data can reflect the SNP effect on gene expression. Finally, we could use a Chi-square test to obtain significance.

RESULTS

As shown in **Table 1**, we totally got 2 GWAS datasets and 5 eQTL datasets. We combined these datasets for a total of 10 experiments. From these 10 experiments, we detected SNPs with statistical effects in five experiments. In total, we found 27 SNPs which are associated with AD.

Most of the SNPs are overlapped in these five experiments, evidence that our method is effective. The false positive is low.

Since one single SNP could be labeled by different probes in the eQTL, one SNP could be screened and associated with AD several times. We collate the number of times that SNP is repeatedly screened. As shown in **Figure 3**, most SNPs were screened for four times. Only three SNPs were screened once. This illustrates the high accuracy, to some extent, of the SNPs screened.

TABLE 1 | The number of SNPs obtained in five experiments.

GWAS	eQTL	Number of SNPs
IGAP_stage_1	CD4-cis-eQTLsFDR-ProbeLevel0.5	10
IGAP_stage_1	CD8-cis-eQTLsFDR-ProbeLevel0.6	9
AD_GERAD_GWAS_autosomal_additive	Hap300_CER_AD	1
AD_GERAD_GWAS_autosomal_additive	CisAssociationsProbeLevelFDR0.5	40
AD_GERAD_GWAS_autosomal_additive	associations_1e6	43
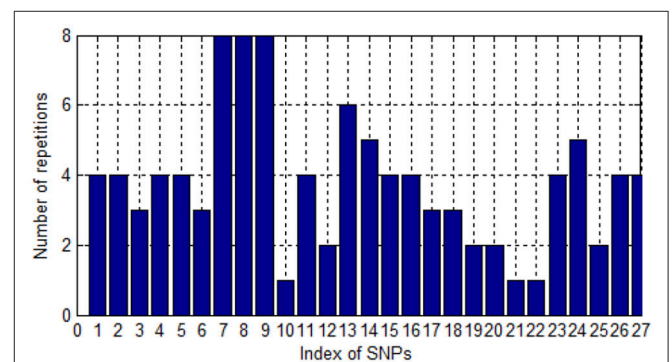
FIGURE 3 | The number of repetitions of SNPs.

TABLE 2 | Detailed information of selected SNP.

Index	SNP	GENE	P-value
1	rs10402271	MS4A2	1.59e-05
2	rs4663105	BIN1	5.63e-06
3	rs11682128	BIN1	1.65e-05
4	rs6733839	BIN1	1.75e-05
5	rs10194375	BIN1	1.79e-05
6	rs56368748	BIN1	2.13e-05
7	rs6859	PVRL2	1.22e-13
8	rs377702	PVRL2	1.56e-08
9	rs157580	PVRL2	1.64e-08
10	rs439401	PVRL2	1.38e-07
11	rs8106922	PVRL2	1.31e-06
12	rs610932	MS4A6A	2.19e-06
13	rs662196	MS4A2	7.14e-06
14	rs583791	MS4A6A	7.41e-06
15	rs2075650	PVRL2	1.12e-05
16	rs676309	MS4A6A	1.14e-05
17	rs12610605	GEMIN7	1.31e-05
18	rs1562990	MS4A4A	1.39e-05
19	rs11667640	PVRL2	2.78e-05
20	rs440277	PVRL2	2.82e-05
21	rs667897	MS4A6A	3.27e-05
22	rs405509	PVRL2	6.38e-05
23	rs540170	MS4A6A	7.40e-05
24	rs581133	MS4A6A	7.92e-05
25	rs11606287	SPI1	9.49e-05
26	rs7926344	MS4A6A	9.63e-05
27	rs744373	BIN1	2.60e-05

FIGURE 4 | Number of gene duplication.

FIGURE 5 | The number of papers found in the ALZgene.

As shown in **Table 2**, we screened 27 SNPs belonging to seven genes. Most of these genes were repeatedly chosen. **Figure 4** shows the number of times the genes were screened.
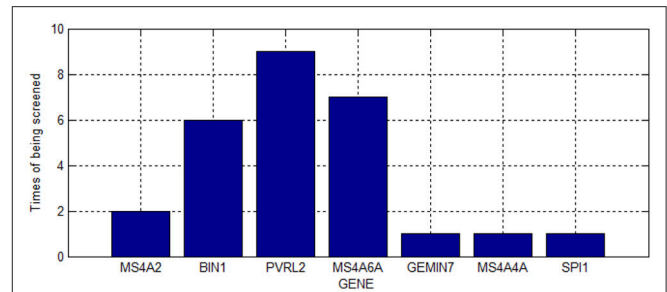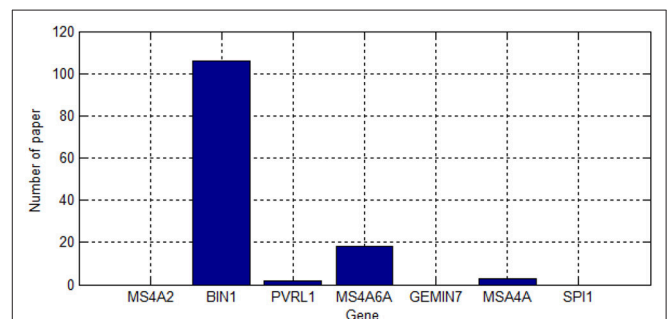
As shown in **Figure 4**, PVRL2 was screened nine times. This gene is therefore very likely associated with AD.

Gene Function

Next, we searched for known AD-related genes on different databases (as shown in **Table 1**) to identify which genes were novel and which genes are known.

First, we used the DisGeNET (Piñero et al., 2015) database to search for AD-related genes. In total 2244 AD-related genes were found. Three genes were identified in this database; MS4A2, MS4A6A and MS4A4A. Then, we used the ALZgene (Bertram et al., 2007) database to search for AD-related genes. Four novel genes were identified in this database; SPI1, GEMIN7, and MS4A2.

We searched for the seven genes on the ALZgene database, and the number of papers found in the ALZgene are shown in **Figure 5.** Since MS46A and MSA4A are both included both databases, we will not discuss them in detail below. We will discuss the functions of the following five genes: BIN1, PVRL2, GEMIN7, SPI1, and MS4A2.

BIN1

BIN1 is an encoding gene with the product name; Myc box-dependent-interacting protein 1. Several isoforms of a nucleo-cytoplasmic adaptor protein are encoded by BIN1. The isoforms expressed in the central nervous system may participate in the endocytosis of synaptic vesicles and may interact with activator proteins, synaptic proteins, endothelin and reticulin (Lee et al., 2002). Mouse studies show that this gene plays an important role in the development of myocardium. Alternatively splicing of genes leads to several transcriptional variants encoding different types of allotypes. Abnormal splicing variants expressed in tumor cell lines are also described.

Due to its ubiquitous expression in the brain, many scholars research the relationship between BIN1 and AD. Chapuis et al. (Chapuis et al., 2013) suggested that BIN1 mediates AD risk by modulating Tau pathology. Yuan et al. (Chen et al., 2018) considered that allele C at the rs744373 locus of the BINl gene was a risk factor for aMCI. The recessive model CC/CT+TT at the rs744373 locus of the BINl gene, particularly increased the risk of aMCI. Experiments were performed on 107 aMCI patients and 150 normal people.

PVRL2

PVRL2 encodes a single pass type I glycoprotein. The protein is one component of the plasma membrane. It also acts as an entry point for the herpes simplex virus and pseudorabies virus mutants and is involved in the intercellular transmission of these viruses (Almire et al., 2010). The variation of the gene is related to the severity of multiple sclerosis.

Since PVRL2 is proven to be related to multiple sclerosis causing glial fibrils to proliferate and form calcified plaques, it is very similar to the clinical symptoms of AD. Therefore, it could be a candidate AD-related gene. Yashin et al. (2017) collected data from the Framingham Heart Study (FHS), the Cardiovascular Health Study (CHS), the Health and Retirement Study (HRS) and the Late Onset Alzheimer's Disease Family Study (LOADFS). They used logistic regression and Cox's regression and found that APOE, TOMM40, PVRL2 (NECTIN2), and APOC1 are strongly associated with AD.

GEMIN7

The gene encodes the components of the core survival of motor neurons (SMN) complex, which plays a crucial role in the splicing pre-mRNAs in the nucleus. It has been found that the gene encodes three transcriptional variants of the same protein. It mainly expresses in the lymph nodes and the spleen.

Since, Baccon et al. (2002) discovered GEMIN7 as a new component of SMN in 2002, many researchers have studied the mechanism and protein structure of GEMIN7. SMN is a product of the disease gene of the common neurodegenerative disease spinal muscular atrophy. AD is a type of neurodegenerative disease and some patients also displayed symptoms of movement disorders. Therefore, we speculate that GEMIN7 may cause motor nerve lesions and then causes AD.

SPI1

The ETS domain transcription factor is encoded by SPI1. It could activate gene expression during myeloid and B lymphocyte development (Zakrzewska et al., 2010). Alternative splicing of the target gene is also regulated by its protein. It broadly expresses in bone marrow.

SPI1 is expressed by microglia which are the resident immune cells of the brain. Some researches (Huang et al., 2017; Krasemann et al., 2017) suggest that PU.1, a transcription factor encoded by the gene SPI1, is a central hub in the AD gene network and is associated with AD pathology.

MS4A2

MS4A2 (Ferreira et al., 2010) encodes the high affinity immunoglobulin epsilon receptor subunit beta, which is a member of the membrane-spanning 4A gene family. Allergic reactions involve the binding of allergens to receptor-bound IgE receptors, which are present on the surface of mast cells and basophils. Members of this new protein family share common structural characteristics and similar intron/exon splicing boundaries and exhibit unique expression patterns in hematopoietic and non-lymphoid tissues. Alternative splicing

results in multiple transcript variants encoding different proteins. It has a biased expression in lung and ball bladder.

Many researches (Lambert et al., 2010; Jones et al., 2012) have found that AD is related to the overworking of the immune system. The immune system over-clears neuronal synapses, cutting off the connections between neurons, causing AD (Blasko and Grubeckloebenstein, 2003). We suggest that MS4A2 may cause abnormalities in the immune system.

CONCLUSION

AD can cause progressive memory impairment, cognitive impairment, personality changes and language disorders and other neuropsychiatric symptoms, seriously affecting people's social, occupational and life functions. Identifying AD-related genes is of great clinical significance for the early diagnosis and treatment of AD.

In this paper, we used SMR to identify AD-related genes and locus. Using two GWAS datasets and five eQTL datasets, we completed ten experiments and collected effective results from a total of five experiments. In total, we identified 27 SNPs that are associated with AD. These SNPs correspond to seven genes: MS4A2, BIN1, PVRL2, MS4A6A, GEMIN7, MS4A4A, and SPI1. To verify the effectiveness of our method and the accuracy of our results, we compared the results with known databases. We found three of the seven genes in the DisGeNET database (BIN1, PVRL2, GEMIN7, and SPI1 is novel genes) and six of seven genes in the ALZGENE (SPI1, GEMIN7, and MS4A2 are novel genes). This confirms the accuracy of our results. To explain the mechanism of genetic pathogenesis, we described the function of the novel genes and speculated the mechanism of its pathogenesis.

AUTHOR CONTRIBUTIONS

YZ did the data preprocessing. TZh did the algorithm simulation under the direction of YH. TZa and LC helped proofreading the manuscript.

FUNDING

REFERENCES

Albert, F. W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16, 197–212. doi: 10.1038/nrg3891

Almire, C., Bertrand, P., Ruminy, P., Maingonnat, C., Wlodarska, I., Martínsubero, J. I., et al. (2010). PVRL2 is translocated to the TRA@ locus in

t(14;19)(q11;q13)-positive peripheral T-cell lymphomas. *Genes Chrom. Cancer* 46, 1011–1018. doi: 10.1002/gcc.20490

Baccon, J., Pellizzoni, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). Identification and characterization of Gemin7, a novel component of the survival of motor neuron complex. *J. Biol. Chem.* 277, 31957–31962. doi: 10.1074/jbc.M203478200

Barral, S., Bird, T., Goate, A., Farlow, M. R., Diazarrastia, R., Bennett, D. A., 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

Bertram, L., Mcqueen, M. B., Mullin, K., Blacker, D., and Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat. Genet.* 2:S23. doi: 10.1038/ng1934

Blasko, I., and Grubeckloebenstein, B. (2003). Role of the immune system in the pathogenesis, prevention and treatment of Alzheimer's disease. *Drugs Aging* 20, 101–113. doi: 10.2165/00002512-200320020-00002

Chapuis, J., Hansmannel, F., Gistelinck, M., Mounier, A., Van, C. C., Kolen, K. V., et al. (2013). Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol. Psychiatry* 18, 1225–1234. doi: 10.1038/mp.2013.1

Chen, J., Xia, Y., Gao, C. L., Wang, R. X., and Lu, Z. N. (2018). [Genes polymorphism of BIN1 and ApoE in patients with amnestic mild cognitive impairment from Enshi Tujia area]. *Zhonghua Yi Xue Za Zhi* 98,1322–1326. doi: 10.3760/cma.j.issn.0376-2491.2018.17.008

Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018a). 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. (2018b). InfAcrOnt: calculating cross-ontology term similarities using information flow by a random walk. *BMC Genomics* 19(Supp. 1):919. doi: 10.1186/s12864-017-4338-6

Cheng, L., Sun, J., Xu, W., Dong, L., Hu, Y., and Zhou, M. (2016). OAHG: an integrated resource for annotating human genes with multi-level ontologies. *Sci. Rep.* 6:34820. doi: 10.1038/srep34820

Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2018c). LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* doi: 10.1093/nar/gky1051. [Epub ahead of print].

Clément-Ziza, M., Marsellach, F. X., Codlin, S., Papadakis, M. A., Reinhardt, S., Rodríguez-López, M., et al. (2015). Natural genetic variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. *Mol. Sys. Biol.* 10:764. doi: 10.15252/msb.20145123

Cruchaga, C., Haller, G., Chakraverty, S., Mayo, K., Vallania, F. L., Mitra, R. D., et al. (2012). Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS ONE* 7:e31039. doi: 10.1371/annotation/c92e16da-7733-421d-b063-1db19488daa6

Dan, L. N., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 6:e1000888. doi: 10.1371/journal.pgen.1000888

Du, Y., Wen, Y., Guo, X., Hao, J., Wang, W., He, A., et al. (2018). A genome-wide expression association analysis identifies genes and pathways associated with amyotrophic lateral sclerosis. *Cell. Mol. Neurobiol.* 38, 635–639. 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. *Progr. Neuro Psychopharmacol. Biol. Psychiatry* 78, 149–152. doi: 10.1016/j.pnpbp.2017.05.017

Ferreira, M. A. R., Oates, N. A., Vliet, J. V., Zhao, Z. Z., Ehrich, M., Martin, N. G., et al. (2010). Characterization of the methylation patterns of MS4A2 in atopic cases and controls. *Allergy* 65, 333–337. doi: 10.1111/j.1398-9995.2009.02135.x

Hillburns, E. M., Wissemann, W. T., Hamza, T. H., Factor, S. A., Zabetian, C. P., and Payami, H. (2014). Identification of a novel Parkinson's disease locus via stratified genome-wide association study. *BMC Genomics* 15, 118–118. doi: 10.1186/1471-2164-15-118

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., et al. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* 43:429. doi: 10.1038/ng.803

Hu, Y., Zhao, T., Zhang, N., Zang, T., Zhang, J., and Cheng, L. (2018). Identifying diseases-related metabolites using random walk. *BMC Bioinform.* 19(Suppl. 5):116. doi: 10.1186/s12859-018-2098-1

Huang, K., Marcora, E., Pimenova, A., Narzo, A. D., Kapoor, M., Jin, S. C., et al. (2017). A common haplotype lowers SPI1 (PU.1) expression in myeloid cells and delays age at onset for Alzheimer's disease. *Nat. Neurosci.* 20, 1052–1061. doi: 10.1038/nn.4587

James Ronald, J. M. A. (2007). The Evolution of Gene Expression QTL in Saccharomyces cerevisiae. *PLoS ONE* 2:e678. doi: 10.1371/journal.pone.0000678

Jones, L., Holmans, P. A., Hamshere, M. L., Harold, D., Moskvina, V., Ivanov, D., et al. (2012). Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *Alzheim. Dem. J. Alzheim. Assoc.* 6:S113. doi: 10.1016/j.jalz.2010.05.350

Kauwe, J. S. K., Cruchaga, C., Karch, C. M., Sadler, B., Mo, L., Mayo, K., et al. (2011). Fine mapping of genetic variants in BIN1, CLU, CR1 and PICALM for association with cerebrospinal fluid biomarkers for Alzheimer's disease. *PLoS ONE* 6:e15918. doi: 10.1371/journal.pone.0015918

Kauwe, J. S. K., Cruchaga, C., Mayo, K., Fenoglio, C., Bertelsen, S., Nowotny, P., et al. (2008). Variation in MAPT is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition. *Proc. Natl. Acad. Sci. U. S. A.* 105, 8050–8054. doi: 10.1073/pnas.0801227105

Krasemann, S., Madore, C., Cialic, R., Baufeld, C., Calcagno, N., Fatimy, R. E., et al. (2017). The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in Neurodegenerative diseases. *Immunity* 47, 566–581. doi: 10.1016/j.immuni.2017.08.008

Lambert, J. C., Grenier-Boley, B., Chouraki, V., Heath, S., Zelenika, D., Fievet, N., et al. (2010). Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. *J. Alzheim. Dis. Jad* 20:1107. doi: 10.3233/JAD-2010-100018

Lee, E., Marcucci, M., Daniell, L., Pypaert, M., Weisz, O. A., Ochoa, G. C., et al. (2002). Amphiphysin 2 (Bin1), T-tubule biogenesis in muscle. *Science* 297, 1193–1196. doi: 10.1126/science.1071362

Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., et al. (2007). Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* 3, 538–543. doi: 10.1371/journal.pgen.0030058

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.* 2018:3848560. doi: 10.1155/2018/3848560

Malhotra, A. K., Zhang, J. P., and Lencz, T. (2012). Pharmacogenetics in psychiatry: translating research into clinical practice. *Mol. Psychiatry* 17, 760–769. doi: 10.1038/mp.2011.146

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

Meng, Y., Lee, J. H., Cheng, R., St George-Hystop, P., Mayeux, R., and Farrer, L. A. (2007). Association between SORL1 and Alzheimer's disease in a genome-wide study. *Neuroreport* 18, 1761–1764. doi: 10.1097/WNR.0b013e3282f13e7a

Moffatt, M. F., Kabesch, M., Liang, L., Dixon, A. L., Strachan, D., Heath, S., et al. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448:470. doi: 10.1038/nature06014

Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardarajan, B. N., Buros, J., et al. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* 43, 436–441. doi: 10.1038/ng.801

Pavlides, J. M. W., Zhu, Z., Gratten, J., Mcrae, A. F., Wray, N. R., and Yang, J. (2016). Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* 8:84. doi: 10.1186/s13073-016-0338-4

Piñero, J., Queraltrosinach, N., Bravo, À., Deupons, J., Bauermehren, A., Baron, M., et al. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* 2015:bav028. doi: 10.1093/database/bav028

Senova, S., Chaillet, A., and Lozano, A. M. (2018). Fornical closed-loop stimulation for Alzheimer's disease. *Trends Neurosci.* 41, 418–428 doi: 10.1016/j.tins.2018.03.015

Veturi, Y., and Ritchie, M. D. (2018). "How powerful are summary-based methods for identifying expression-trait associations under different genetic architectures?" in *Pacific Symposium on Biocomputing*, Vol. 23 (Hawaii), 228–239. doi: 10.1142/9789813235533_0021

Visscher, P. M., Brown, M. A., Mccarthy, M. I., and Yang, J. (2012). Five Years of GWAS Discovery. *Am. J. Hum. Genet.* 90, 7–24. doi: 10.1016/j.ajhg.2011.11.029

Yashin, A. I., Fang, F., Kovtun, M., Wu, D., Duan, M., Arbeev, K., et al. (2017). Hidden heterogeneity in Alzheimer's disease: insights from genetic association studies and other analyses. *Exper. Gerontol.* 107, 148–160. doi: 10.1016/j.exger.2017.10.020

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, 3641–3649. doi: 10.1101/274654

Zakrzewska, A., Cui, C., Stockhammer, O. W., Benard, E. L., Spaink, H. P., and Meijer, A. H. (2010). Macrophage-specific gene functions in Spi1-directed innate immunity. *Blood* 116, 1–11. doi: 10.1182/blood-2010-01-262873

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. doi: 10.1038/ng.3538