IDENTIFYING THE KEY PATHOGENIC FACTORS OF NEUROLOGICAL DISORDERS BY INTEGRATING MULTI-OMICS DATA

EDITED BY: Zhijie Han, Guiyou Liu, Liang Cheng, Shicheng Guo and Andrea Legati

PUBLISHED IN: Frontiers in Genetics and Frontiers in Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-786-5
DOI 10.3389/978-2-83250-786-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

IDENTIFYING THE KEY PATHOGENIC FACTORS OF NEUROLOGICAL DISORDERS BY INTEGRATING MULTI-OMICS DATA

Topic Editors:
Zhijie Han, Chongqing Medical University, China
Guiyou Liu, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences (CAS), China
Liang Cheng, Harbin Medical University, China
Shicheng Guo, University of Wisconsin-Madison, United States
Andrea Legati, IRCCS Carlo Besta Neurological Institute Foundation, Italy

Citation: Han, Z., Liu, G., Cheng, L., Guo, S., Legati, A., eds. (2022). Identifying the Key Pathogenic Factors of Neurological Disorders by Integrating Multi-omics Data. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-786-5

Table of Contents

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fgene.2021.751999

Proteomic Response of Rat Pituitary Under Chronic Mild Stress Reveals Insights Into Vulnerability and Resistance to Anxiety or Depression

Fenfang Tian[1†], Dan Liu[1†], Jin Chen[1,2†], Wei Liao[1], Weibo Gong[1], Rongzhong Huang[3,4], Liang Xie[1,5], Faping Yi[1]* and Jian Zhou[1]**

[1]Institute of Neuroscience, Basic Medical College, Chongqing Medical University, Chongqing, China, [2]Department of Neurology, The First Affiliated Hospital of Nanchang University, Nanchang, China, [3]Statistics Laboratory, ChuangXu Institute of Life Science, Chongqing, China, [4]Chongqing Institute of Life Science, Chongqing, China, [5]Department of Neurology, The Second Affiliated Hospital of Nanchang University, Nanchang, China

OPEN ACCESS

Edited by:
Zhijie Han,
Chongqing Medical University, China

Reviewed by:
Meichun Deng,
Central South University, China
Zhen Zhong,
Zhejiang University, China

***Correspondence:**
Liang Xie
xl580122@163.com
Faping Yi
100506@cqmu.edu.cn
Jian Zhou
zhoujian@cqmu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: 02 August 2021
Accepted: 01 September 2021
Published: 17 September 2021

Citation:
Tian F, Liu D, Chen J, Liao W, Gong W, Huang R, Xie L, Yi F and Zhou J (2021) Proteomic Response of Rat Pituitary Under Chronic Mild Stress Reveals Insights Into Vulnerability and Resistance to Anxiety or Depression. Front. Genet. 12:751999. doi: 10.3389/fgene.2021.751999

Chronic stress as one of the most significant risk factor can trigger overactivity of hypothalamic-pituitary-adrenal (HPA) axis in depression as well as anxiety. Yet, the shared and unique neurobiological underpinnings underlying the pituitary abnormality in these two disorders have not been made clear. We previously have established depression-susceptible, anxiety-susceptible and insusceptible groups using a valid chronic mild stress (CMS) model. In this work, the possible protein expression changes in the rat pituitary of these three groups were continuously investigated through the use of the comparative quantitative proteomics and bioinformatics approaches. The pituitary-proteome analysis identified totally 197 differential proteins as a CMS response. These deregulated proteins were involved in diverse biological functions and significant pathways potentially connected with the three different behavioral phenotypes, likely serving as new investigative protein targets. Afterwards, parallel reaction monitoring-based independent analysis found out that expression alterations in *Oxct1*, *Sec24c*, *Ppp1cb*, *Dock1*, and *Coq3*; *Lama1*, *Glb1*, *Gapdh*, *Sccpdh*, and *Renbp*; *Sephs1*, *Nup188*, *Spp1*, *Prodh1*, and *Srm* were specifically linked to depression-susceptible, anxiety-susceptible and insusceptible groups, respectively, suggesting that the same CMS had different impacts on the pituitary protein regulatory system. Collectively, the current proteomics research elucidated an important molecular basis and furnished new valuable insights into neurochemical commonalities and specificities of the pituitary dysfunctional mechanisms in HPA axis underlying vulnerability and resistance to stress-induced anxiety or depression.

Keywords: anxiety, chronic mild stress, depression, proteome, rat pituitary

INTRODUCTION

Anxiety and depression are two severe and chronic neuropsychiatric illnesses. The prevalences of these disorders are increasing, potentially representing a significant clinical challenge. Mounting evidence suggests that many risk factors are shared between the anxiety and depression disorders such as chronic life stress (Krishnan et al., 2007; Zhou et al., 2016; Jefferson et al., 2020). Chronic

stress can result in the adverse health impacts when it increases beyond a certain level, thereby causing anxiety and depression (Chang and Grace, 2014; Tian et al., 2020). However, many individuals can manage the psychological and physical effects of the stressful situations and do not have the disease symptoms (Henningsen et al., 2012). To model the adverse environment factors that affect humans, chronic mild stress (CMS) protocol has been commonly employed to induce anxious-like and depressive-like behaviors in rodent animals (Chang and Grace, 2014; Zhou et al., 2016). To identify the potential biological relationships between CMS and pathological changes, it may be useful to focus on the neurobiological components and processes reflecting adaptive and maladaptive responses to the stress-caused anxiety and depression.

Generally, the clinical symptoms of anxiety and depression are different. However, they are frequently presented simultaneously (Liu et al., 2021; Thorp et al., 2021). Importantly, there are lots of overlaps with respect of the pathophysiology and comorbidity of these two disorders. Considerable data in many clinical and animal researches are usually mixed, thereby confusing our knowledge of the underlying causes and effects of anxiety and depression (Chiba et al., 2012; Lucassen et al., 2016; Oh et al., 2020). In recent years, researchers have attempted to separately investigate non-comorbid individuals to unravel the specificities and commonalities of the two disorders (Lotan et al., 2014; Hamilton et al., 2015; Zhao et al., 2017; Chen et al., 2018). Many studies have demonstrated that the activity of hypothalamic-pituitary-adrenal (HPA) axis is perturbed in these stress-related disorders (Borrow et al., 2016; Delvecchio et al., 2017; Lee and Rhee, 2017). As an integral part of the HPA axis, the pituitary synthesizes and secretes a variety of hormones to mediate a series of biological functions (Yelamanchi et al., 2018). It may be one of the areas most impacted by stress dysregulation in anxiety and depression (Stelzhammer et al., 2015). An increase in the size of the pituitary has also been found in subjects with depression and anxiety through magnetic resonance imaging (Tsigos and Chrousos, 2002; Lorenzetti et al., 2009; Krishnamurthy et al., 2017). To some extent, this reflects an increase in the size and number of corticotropin-releasing hormone (CRH) cells that produce and secrete higher levels of hormones, such as CRH and adrenocorticotrophic hormone (ACTH) (Tsigos and Chrousos, 2002; Krishnamurthy et al., 2017). Despite the morphological and functional abnormalities of the pituitary have been implicated in stress-related anxiety and depression, the corresponding neurobiological molecular basis may remain difference and need to be extensively explored.

Our previous study has demonstrated the three different subpopulations induced by CMS including depression-susceptible (Dep-Sus), anxiety-susceptible (Anx-Sus), and insusceptible (Insus) groups and carried out the comparative proteomic analysis of the rat hippocampal tissues (Tang et al., 2019). In this work, the pituitary tissues from the identical batch of CMS-exposed rats were used to continuously study stress-caused anxiety and depression (Tang et al., 2019). A proteomic approach based on isobaric tags for relative and absolute quantitation (iTRAQ) was utilized to gain unbiased profiling data. Enrichments of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were conducted to analyze the main function and the significant pathways of the identified abnormally-expressed proteins. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database and Cytoscape were employed to map protein-protein interaction (PPI) networks. The results help elucidate commonalities and differences of the complex molecular mechanisms that underlie stress resistance and stress-caused anxiety or depression.

METHODS

Animals and Ethics Statement

Healthy adult male Sprague-Dawley rats (weight, about 250 g; Animal Center of Chongqing Medical University) were used in the present study. All the rats were individually housed in standard laboratory conditions (55 ± 5% relative humidity, 12/12 h light/dark cycle, 21–22°C) with ad libitum feeding. The study protocol was approved by the local Ethics Committee (2017013). All animals were treated according to the National Institutes of Health protocols for the use and care of laboratory animals.

CMS Rat Model

As previously described (Tang et al., 2019), the 8-weeks CMS protocol was employed to build the rat model. Following exposure to the CMS, the stressed rats were divided into the three groups: 1) Dep-Sus group [assessed by sucrose preference (SP) test and forced swimming (FS) test]; 2) Anx-Sus group [assessed by elevated plus-maze (EPM) test]; and 3) Insus group. Additional non-handled rats acted as the control (Ctrl) group. For a more detailed description, please refer to our previous study (Tang et al., 2019).

Tissue Isolation and Lysis

After the behavioral assessment, the animals were anesthetized and decapitated and their whole brains were carefully removed on ice. The pituitary tissue was isolated from the rat brain and frozen rapidly in liquid nitrogen and then stored at −80°C in a refrigerator prior to use. For protein extraction, a sample of the pituitary of each animal was added to an SDT buffer composed of 4% SDS, 0.1 M dithiothreitol, 0.1 M Tris–HCl, pH 8.0, and protease inhibitors. The tissues were homogenized and lyzed, the extracted proteins were boiled for 5 min. After centrifugation at 4°C and 40,000 × *g* for 15 min, the supernatants were collected and the protein concentrations were quantified using Pierce bicinchoninic acid assay kit.

Digestion of Pituitary Proteins and iTRAQ Labeling

Following our previously described procedure (Gong et al., 2021), the protein samples were in parallel digested using filter-aided sample preparation (FASP). In this method, an ultrafiltration filter (10 kD cutoff) was used for effective digestion. In brief, UA buffer (8 M urea, 0.15 M Tris-HCl, pH 8.0) was added to each sample. The sample was transferred to an ultrafiltration

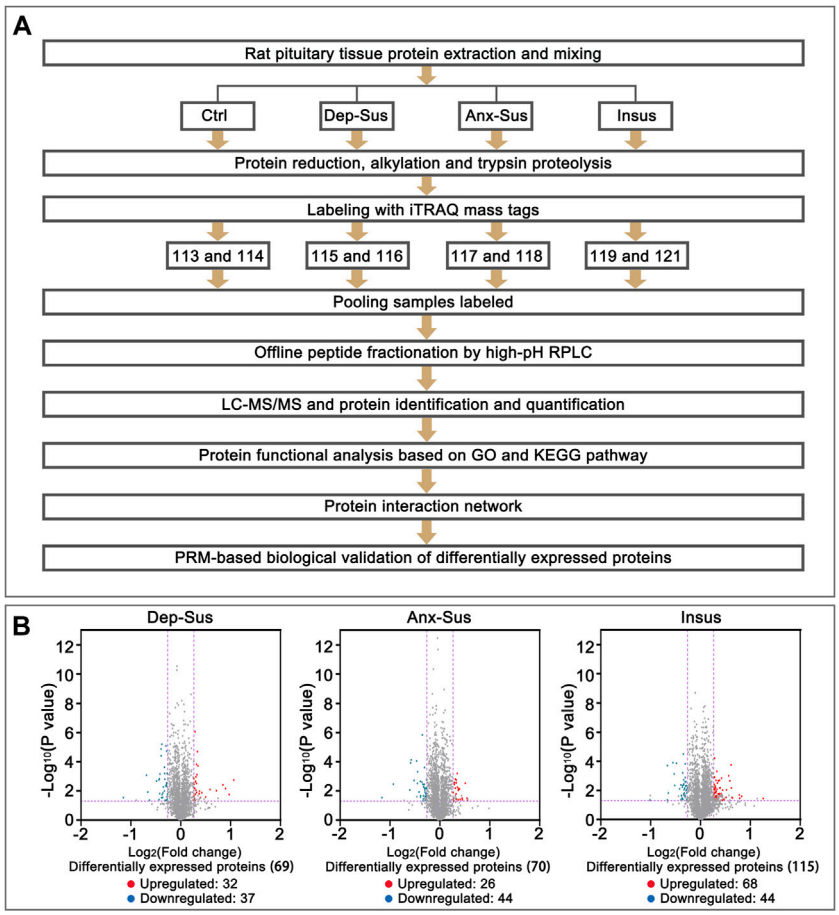

FIGURE 1 | Comparative analysis of the pituitary proteomic response of the rats under chronic mild stress (CMS). **(A)** Schematic representation of quantitative proteomics analysis of the control (Ctrl), depression-susceptible (Dep-Sus), anxiety-susceptible (Anx-Sus) and insusceptible (Insus) groups. **(B)** Volcano plot of the protein expression changes in the three groups. In volcano plot, the red plot represented up-regulated proteins, and the blue plot represented down-regulated proteins. The x-axis shows the log2-transformed average fold change. The y-axis shows the negative log10-transformed *p*-value.

centrifuge tube and then centrifuged, and washed again with UA buffer. Subsequently, 0.05 M iodoacetamide in UA buffer was added to the filter. The protein mixture was incubated and alkylated for 30 min at room temperature in the dark. The filter unit was centrifuged and then washed twice with UA buffer. Finally, trypsin solution was added and digested at 37°C overnight. The resulting peptides were collected as a filtrate and then dried in a Speed Vac.

High-pH Reversed-Phase Liquid Chromatography (RPLC) Fractionation and Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

The tryptic peptides were labeled with eight-plex iTRAQ reagents according to the protocol of the manufacturer. The reagents 113–121 were used to label the eight samples from the three stressed and the Ctrl groups, as depicted in **Figure 1A**. Each used sample was obtained from 2 to 3 rats in each group (Lenselink et al., 2015). Subsequently, the eight labeled samples were pooled and preliminarily separated using high-pH RPLC. Briefly, the peptides were dissolved with buffer A (5% acetonitrile, 0.01 M ammonium formate, pH 10.0) and fractionated through linear elution in a gradient of 5–38% buffer B (90% acetonitrile (ACN), 0.01 M ammonium formate, pH 10.0) for 80 min at 300 µL/min. A total of sixteen fractions were collected, desalted and dried for the subsequent LC-MS/MS analysis.

The peptides in each fraction were re-dissolved in 0.1% formic acid, and delivered into Thermo Scientific Easy-nLC 1200 system coupled with a nanoViper C18 trap column (3 µm, 100 Å). The peptide mixtures were trapped and then desalted using 100% solvent A (0.1% formic acid). Afterward, the peptides were eluted with 8–38% solvent B (80% ACN/0.1% formic acid) for 50 min, and separated with an analytical column (50 µm × 150 mm, 3 µm-C18 100 Å). Q-Exactive Orbitrap mass spectrometer equipped with a Nano Flex ion source (ThermoFisher) was used for the MS analysis (interface heater temperature, 275°C; ion spray voltage, 1.9 kV). The tandem MS data were acquired through the use of a data-dependent acquisition mode along with full MS scans. The acquisition range was 350–1,200 m/z for the

MS1 and 110–1,200 m/z for the MS2. For the information acquisition, survey scans were acquired in 250 ms and up to 14 product ion scans (50 ms) were collected. Those MS spectra along with charge state 2–4 were selected and subjected to fragmentation using higher-energy collision dissociation, and dynamic exclusion for selected precursor ions was set to 25 s.

Protein Identification and Quantification

Raw files were processed and searched using the Sequest HT search engine embedded into Proteome Discoverer software 2.1 (ThermoFisher) against the UniProt Rat database. The following search parameters were set: monoisotopic mass values, fragment mass tolerance at 0.05 Da and precursor mass tolerance ± 10 ppm, trypsin as the enzyme, and allowing up to 2 missed cleavages. Fixed modifications were defined as iTRAQ labeling and carbamidomethylation of Cys; Oxidation on Met, acetylation on protein N-term, deamidation on Asn and Gln, and Pyro-Glu were specified as a variable modification. The decoy database pattern was set as the reversed version of the target database. All reported data were based on 99% confidence for peptide identification as determined by a false discovery rate (FDR) of lower than 1%. Relative ratios of identified peptides among labeled samples were computed using relative peak intensities of released iTRAQ reporter ions in each of the MS/MS spectra, and introduced into Excel spreadsheet for manual treatment. Then, the ratios of all identified proteins were analyzed *via* a two-tailed Student's t-test. Those proteins with 1.2-fold expression alterations and *p*-values lower than 0.05 could be considered as significantly different. The raw data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the iProX partner repository with the dataset identifier PXD025429 (Ma et al., 2019).

Bioinformatics

GO analyses including biological processes (GO-BP), molecular functions (GO-MF), and cellular components (GO-CC) were conducted through the use of the OmicsBean tool (http://www.omicsbean.cn/). KEGG (http://www.genome.jp/kegg/) was used to identify the significant pathways with *p*-values of lower than 0.1 following the previously described procedure (Yu et al., 2017). Moreover, the STRING database and Cytoscape were used to construct PPI networks following the previously reported protocol (Gong et al., 2021).

Parallel Reaction Monitoring (PRM) MS Assay

Following the iTRAQ-based proteomics experiment, extraction and digestion of the pituitary proteins were performed. The resulting peptides were analyzed using the Q-Exactive Orbitrap mass spectrometer. A normalized collision energy of 28 was used for the fragmentation of the peptides, and the resulting fragments were analyzed at a resolution of 35,000. The acquired raw data were analyzed *via* the Proteome Discoverer tool. The MS data were further processed using the analysis software Skyline 19.1 (ThermoFisher). The statistical analysis of the data were performed using Student's t tests of SPSS software. The data were presented as means ± standard error (SE). The difference was considered to be statistically significant when *p*-values lower than 0.05.

RESULTS

iTRAQ-Based Proteomics Analysis of the Rat Pituitary Under the CMS

In the present work, our used pituitary tissue samples were from the identical batch of the stressed animals in our recently published paper (Tang et al., 2019). Briefly, the stress-induced depressive-like behavior including anhedonia and behavioral despair were firstly assessed through the use of the SP and FS tests. Meanwhile, we also utilized the EPM test for indexing the anxious-like symptom. Based on these testing data, a subset of the Dep-Sus, Anx-Sus, Insus, and Ctrl groups was finally obtained. Overall, these results indicated that we could effectively utilized the CMS model to investigate the neurobiological processes associated with the resistance and vulnerability of stress-related anxious or depressive disorders.

Next, we investigated the effects of CMS on the expression of the rat pituitary proteins through the use of iTRAQ-based quantitative proteomics analyses (**Figure 1A**). In this experiment five animals per group were used, and the pituitary proteins from 2 to 3 rats were equally pooled for each sample (Lenselink et al., 2015). Matching to the UniProt database, within the Ctrl, Dep-Sus, Anx-Sus and Insus groups, totally 3,601 non-redundant proteins were identified and quantified based on the FDR lower than 0.01. The iTRAQ-based protein expressions that changed greater than 1.2-fold and *p*-values lower than 0.05 *versus* the values for the Ctrl group were deemed to be significantly different. Overall, 197 proteins were found to exhibit a significant differential expression in the three groups (**Supplementary Table S1**). Here, the proteome profile of the pituitary was contrasted with that of the hypothalamus from our previous work (Gong et al., 2021) (**Supplementary Figure S1**). Despite the profile of hypothalamus and pituitary was similar based on a 60–70% overlap of the total quantified proteins, the differential protein sets in each area exhibited considerably divergent. This suggested that there were different proteome responses to stress in these two areas.

Functional and Network Characterization of CMS-Responsive Differential Proteins

The pituitary site-specific proteome signature of the CMS-exposed rats unraveled 37 downregulated and 32 upregulated proteins in the Dep-Sus group, 44 downregulated and 26 upregulated proteins in the Anx-Sus group, and 44 downregulated and 68 upregulated in the Insus group (**Figure 1B**). In the two stress-susceptible cohorts, 30 proteins were seen to be similarly deregulated, potentially representing the commonality of stress-induced anxiety and depression. Among the susceptible and the insusceptible groups, 27 similarly deregulated proteins were seen and might sever as a

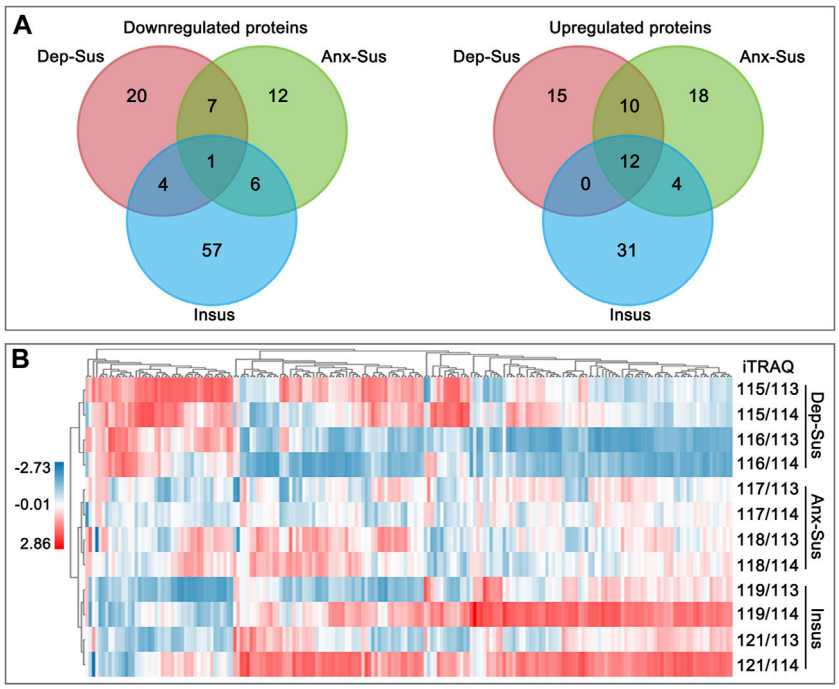

FIGURE 2 | Analysis of the deregulated pituitary proteins identified in the depression-susceptible (Dep-Sus), anxiety-susceptible (Anx-Sus), and insusceptible (Insus) groups. **(A)** Venn diagrams of the deregulated proteins in the three stressed groups. **(B)** Heatmap of the deregulated proteins in the three groups. Higher expressions were indicated by red and lower by blue. The expression levels were shown with various color intensities. In the color bar the log2 scale was used.

consequence of stress exposure (**Figure 2A**). To sum up, as many as 78% of these deregulated proteins were uniquely connected with the three phenotypes, which demonstrated that the three stressed cohorts had specific protein expression disturbances as a response of stress. Further, based on the unsupervised hierarchical clustering analysis, the expression profile of the 197 deregulated proteins were divided into three different units, to some extent suggesting the three specific CMS responses (**Figure 2B**).

We carried out GO classification and pathway enrichment of the deregulated proteins through the use of the OmicsBean software, for a better understanding of the significant protein functions and biological pathways correlated with the three behavioral phenotypes. The 69 deregulated proteins in the Dep-Sus group were subjected to enrichment analyses of the GO and KEGG pathways. Total 377, 85, 82, and 9 terms in the GO-BP, GO-CC, GO-MF, and KEGG pathways were significantly overrepresented (**Supplementary Table S2**). The ten top enriched GO terms are shown in **Figure 3A**. According to the GO-BP annotations, many deregulated proteins were associated with acute-phase and inflammatory responses, tRNA modification and processing, interferon-alpha, type 1 interferon and protein secretion and regulation. The GO-CC annotations showed that these differential proteins were mainly located in blood microparticle, extracellular region and organelle, and membrane-bounded organelle and vesicle. According to the GO-MF annotations, most of proteins were involved in enzyme inhibitor, peptidase inhibitor and regulator activity, and RNA binding. In pathway enrichment analyses, the deregulated proteins were mainly related to complement and coagulation cascades, RNA transport, mRNA surveillance pathway, synthesis and degradation of ketone bodies, metabolism, apoptosis and SNARE interactions in vesicular transport (**Figure 3B**).

At the same time, enrichments of GO annotations and KEGG pathways of the 70 deregulated proteins in the Anx-Sus group were carried out. There were 388 GO-BP, 84 GO-CC, 82 GO-MF, and 9 KEGG pathway terms overrepresented. The top 10 enriched GO terms are indicated in **Figure 3C**. The GO-BP annotations displayed that the majority of proteins were associated with coagulation, hemostasis, and amino acid and glutathione metabolic processes. According to the GO-CC annotations, the differential proteins were mainly found in membrane-bounded and intracellular organelle, protein, supraspliceosomal and macromolecular complex, and nucleoplasm and cytoplasm parts. The GO-MF annotations indicated that most proteins were involved in enzyme activity, protein and thyroid hormone receptor binding. According to pathway enrichment analyses, the deregulated proteins were primarily implicated in complement and coagulation cascades, metabolism and biosynthesis, cytosolic DNA-sensing pathway, and endocytosis (**Figure 3D**).

Afterwards, GO annotation and KEGG pathway enrichments of the 115 deregulated proteins in the Insus group were also conducted. There were 516 GO-BP, 116 GO-CC, 104 GO-MF, and 10 KEGG pathway terms overrepresented. The top ten

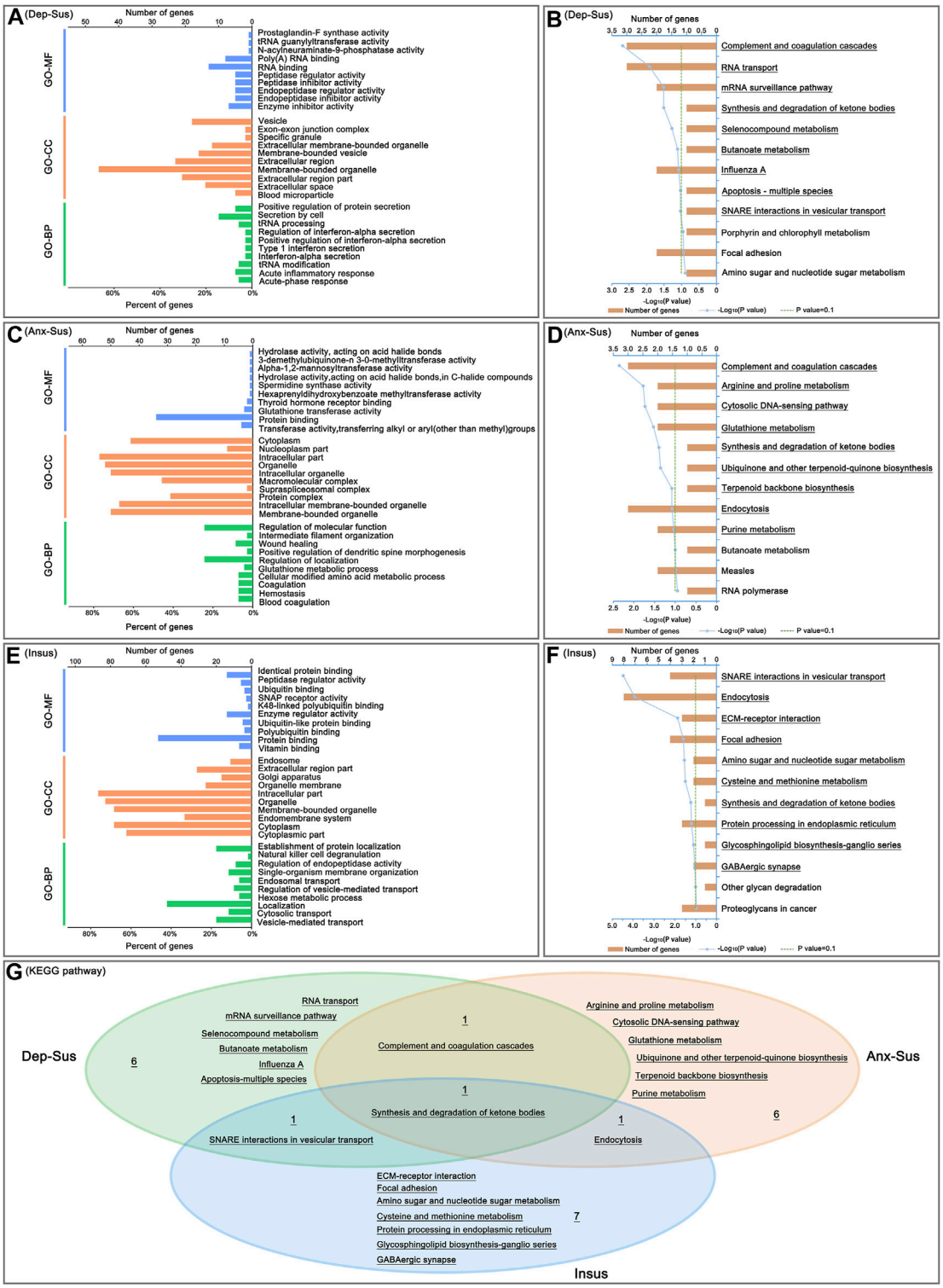

FIGURE 3 | Analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments. The top ten enriched GO biological process (GO-BP), cellular component (GO-CC) and molecular function (GO-MF) terms of the deregulated pituitary proteins from the depression-susceptible (Dep-Sus, **A**), anxiety-susceptible (Anx-Sus, **C**) and insusceptible (Insus, **E**) groups were indicated. Meanwhile, the significantly overrepresented KEGG pathway terms from the Dep-Sus **(B)**, Anx-Sus **(D)**, and Insus **(F)** groups were shown with underscores. The x-axis represented the negative log10-transformed *p*-value. **(G)** Venn diagram displaying unique and common significantly-enriched pathways among the three groups.

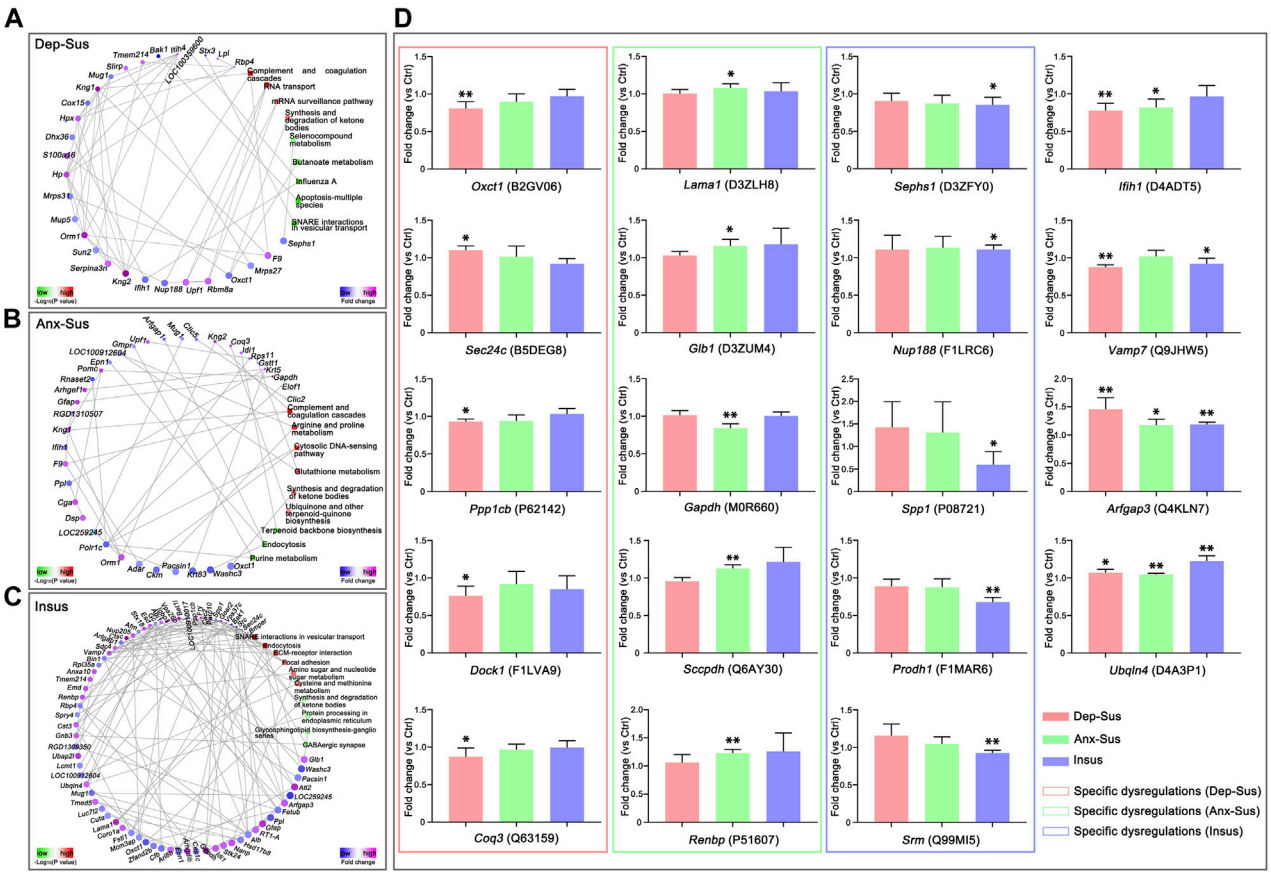

FIGURE 4 | Protein–protein interaction (PPI) network and parallel reaction monitoring (PRM) analyses of the deregulated pituitary proteins of the three groups. The PPI networks of depression-susceptible (Dep-Sus, **A**), anxiety-susceptible (Anx-Sus, **B**), and insusceptible (Insus, **C**) were built based on fold changes of protein expression, PPIs and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments. Circular nodes represented proteins/genes and rectangles represented KEGG pathway terms. Lower *p*-value was indicated in blue and higher *p*-value in yellow. **(D)** PRM analysis of the deregulated proteins in the three stressed groups when compared to the control (Ctrl). *Oxct1*, *Sec24c*, *Ppp1cb*, *Dock1*, *Coq3*, *Lama1*, *Glb1*, *Gapdh*, *Sccpdh*, *Renbp*, *Sephs1*, *Nup188*, *Spp1*, *Prodh1*, *Srm*, *Ifih1*, *Vamp7*, *Arfgap3*, and *Ubqln4* were determined on the pituitary protein extracts of the rats. The relative abundance of target proteins among sample groups were compared based on the abundance of the corresponding peptides. $n = 5$ per group, $^{*}p < 0.05$, $^{**}p < 0.01$.

enriched GO terms are shown in **Figure 3E**. According to the GO-BP annotations, most of the differential proteins were involved in vesicle-mediated, cytosolic and endosomal transport and regulation, localization and metabolic process, and GO-CC category analysis showed that the majority of the deregulated proteins located in the cytoplasmic and intracellular parts, endomembrane system, organelle and endosome. The GO-MF annotations predicted that most of the proteins were engaged in vitamin, protein and ubiquitin binding, and enzyme and peptidase regulator, SNAP receptor activities. The pathway enrichment analyses uncovered that the deregulated proteins were mainly enriched in SNARE interactions in vesicular transport, endocytosis, ECM-receptor interaction, focal adhesion, metabolism and biosynthesis, and GABAergic synapse (**Figure 3F**).

Interestingly, of these significantly-enriched KEGG pathways, there were one shared terms among the three cohorts (**Figure 3G**). Meanwhile, we could see the two common pathways between the two susceptible cohorts. Importantly, the 6, 6 and 7 pathways were seen to be uniquely related to the Dep-Sus, Anx-Sus and Insus groups, respectively, potentially suggesting the three different neurobiological response to the identical CMS.

Furthermore, we also focused on the proteome-inferred PPI networks in the Dep-Sus, Anx-Sus and Insus groups, as shown in **Figures 4A–C**. The PPI network maps of the three stressed groups were built through the use of the deregulated proteins correlated with the significant pathways. 29, 36, and 71 deregulated proteins were identified to be several important factors based on the unified conceptual framework of the three networks from the Dep-Sus, Anx-Sus and Insus groups, respectively. As expected, these networks unraveled close relationships between the deregulated proteins and the significantly enriched pathways, thereby furnishing a useful and valuable interactome unit connected with the three different behavioral phenotypes.

PRM Analysis of CMS-Response Proteins

In this work, the PRM technique was used to further independently validate nineteen abnormally-expressed proteins

of interest involved in the significant biological functions and pathways. On the whole, the PRM data mirrored the iTRAQ results (**Supplementary Figure S2**). As illustrated in other proteomics work (Abdi et al., 2006; Cheng et al., 2011; Xu et al., 2012; Wu et al., 2019), some discrepancies existed between the iTRAQ and PRM data. Except as the assay difference of these two approaches, another probable reason was the additional mixing step in the iTRAQ experiment (Xu et al., 2012; Wu et al., 2019). Compared with the Ctrl group, the expressions of *Oxct1*, *Ppp1cb*, *Dock1*, and *Coq3* were significantly down-regulated while *Sec24c* was up-regulated in the Dep-Sus group; the expressions of *Gapdh* was significantly down-regulated whereas *Lama1*, *Glb1*, *Sccpdh*, and *Renbp* were up-regulated in the Anx-Sus group; the expressions of *Sephs1*, *Spp1,* and *Srm* were significantly down-regulated whereas *Nup188* and *Prodh1* were up-regulated in the Insus group (**Figure 4D**). In addition, the expression level of *Ifih1* was displayed to be significantly reduced in both the Dep-Sus and Anx-Sus groups as compared to the Ctrl group. The reduced level of *Vamp7* in both the Dep-Sus and Insus groups, and the elevated level of *Arfgap3* and *Ubqln4* in the three stressed groups were observed as contrasted with the Ctrl group.

DISCUSSION

Chronic stress is the most major factor among many factors that may cause psychiatric illnesses, including anxiety and depression (Henningsen et al., 2012; Chang and Grace, 2014). This is largely due to the means through which the stress affects the function of the HPA axis (Tsigos and Chrousos, 2002). A valid CMS paradigm was thus commonly employed to cause anxious-like and depressive-like behaviors of rats (Henningsen et al., 2012; Chang and Grace, 2014). Previously we constructed the CMS model to gain the three different phenotypes (Dep-Sus, Anx-Sus, and Insus) of the rats through assessment the behavior performance (Tang et al., 2019). This stress model provided a useful means for assay of common and specific neurochemical characteristics of resistance and susceptibility to anxiety or depression. Profiling the phenotype-related protein expressions may lead to new molecular insights into translational research of depression and anxiety.

To discover the phenotype-related protein deregulations, we compared the expression of proteins in the pituitary of the rats exposed to CMS using iTRAQ-based proteomics analyses. There is a total number of 197 deregulated proteins found in the pituitary of Dep-Sus, Anx-Sus, and Insus rats. The overlapped protein deregulations between the Dep-Sus and Anx-Sus groups likely reflected the shared protein expression patterns of anxiety and depression. Those similar deregulations between the Insus and Dep/Anx-Sus groups could be considered to be a general response to CMS. Interestingly, the specifically deregulated protein expressions in the three stressed cohorts suggested potential differences among the stress-induced behavioral phenotypes. The specific protein dysfunctional profiles were further exhibited and evidenced through the use of the clustering analysis.

Subsequently, some potentially affected biological processes and pathways in the pituitary tissue uniquely associated to stress-induced depressive-like and anxious-like behaviors and stress resistance were found through integrated analysis of the proteomics and bioinformatics. The analysis of biological pathways indicated that the deregulated proteins were significantly enriched for complement and coagulation, ketone bodies, vesicular transport, and metabolism dysfunctions in the Dep-Sus group, complement and coagulation, metabolism and endocytosis deregulations in the Anx-Sus group, and vesicular transport, endocytosis, metabolism and synapse repercussions in the Insus group. Importantly, many significant pathways were found to be distinctly connected with the three phenotypes, which reflected differences in active biological processes and events that happened in these stressed cohorts. The further network mapping unraveled the protein deregulation systems and likely offered some useful clues correlated with resistance and susceptibility to stress-caused anxiety or depression.

In this work, we further utilized PRM-based quantitative method to independently validate the nineteen abnormally-expressed proteins involved with the remarkable biological functions and pathways. The results indicated that *Oxct1*, *Sec24c*, *Ppp1cb*, *Dock1*, and *Coq3* were distinctly deregulated in the pituitary of the Dep-Sus group, whereas *Lama1*, *Glb1*, *Gapdh*, *Sccpdh*, and *Renbp* were distinctly deregulated in the Anx-Sus group. These specific alterations suggested that the same stimuli could lead to the different molecular response and neurobiological processes in the pituitary, thereby triggering the depression and anxiety behaviors. Meanwhile, we observed that *Sephs1*, *Nup188*, *Spp1*, *Prodh1*, and *Srm* were distinctly deregulated in the Insus group, suggesting a potential positive way to dealing with the stress-caused pituitary protein deregulations for stress protection and behavioral adaptation (Krishnan et al., 2007; Zhou et al., 2016).

We found that these PRM-determined phenotype-specific deregulated proteins were mainly involved in the metabolism, focal adhesion, protein processing and RNA transport. *Oxct1*, *Coq3*, *Glb1*, *Sccpdh*, *Gapdh*, *Renbp*, *Sephs1*, *Prodh1*, and *Srm* are involved in a wide range of principal metabolic pathways. In the Dep-Sus group, specific dysregulations of *Oxct1* and *Coq3* would result in the abnormalities of synthesis and degradation of ketone bodies, and ubiquinone and other terpenoid-quinone biosynthesis. In the Anx-Sus group, the aberrations of *Glb1* and *Sccpdh* were important for glycosphingolipid and glycolipid biosynthetic processes, and potentially affected the formation of lipid (Dakik et al., 2021). In the Insus group, *Sephs1*, *Prodh1,* and *Srm* have also been reported to participate in some critical metabolic pathways, such as amino sugar and nucleotide sugar metabolism, selenocompound metabolism, and arginine and proline metabolism. More importantly, these metabolisms were generally considered as a significant source of energy supply. The regulatory abnormality of multiple metabolic processes in the pituitary would lead to a negative or positive energy balance of the HPA axis (Nieuwenhuizen and Rutters, 2008; Harris, 2015). Furthermore, *Ppp1cb*, *Dock1*, *Lama1*, and *Spp1* were found to be involved in focal adhesion pathway. These neural cell adhesion molecule might be vital for the neuronal plasticity of stress-induced disorders (Cotman et al., 1998; Ditlevsen et al., 2008).

Moreover, dysregulation of *Sec24c* in the Dep-Sus group might affect cell surface levels of the serotonin transporters and thus be linked to depression (Sucic et al., 2011). Meanwhile, we also noted the underexpression of *Ifih1* in both the two susceptible groups, which probably was an important pathological clue for depression and anxiety. In our present work, changes in the expressions of proteins involved in multiple significant biological functions and pathways especially metabolism were identified in the pituitary of the stressed rats, it would be interesting to further explore the possible complex mechanisms behind these stress-induced deregulations pointing to the HPA axis dysfunction in depression and anxiety.

CONCLUSION

In this study, we determined the impacts of CMS on the rat pituitary proteome *via* iTRAQ-based and PRM-based quantitative approaches. We found out some candidate pituitary proteins that were likely linked to resistance and susceptibility to CMS-induced depression or anxiety and thus furnished new valuable insights into the stress-affected molecular deregulations in the chronically stressed groups. The current proteomic research can serve as an important molecular underpinning, and help to better understand similarities and differences of the pituitary dysfunction mechanisms in the HPA axis that underlie stress resistance and stress-caused anxiety or depression.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by The experimental protocol was approved by the Ethical Committee of Chongqing Medical University (2017013). Animals were treated in accordance with the National Institutes of Health Guidelines for the use and care of laboratory animals.

AUTHOR CONTRIBUTIONS

FY, LX, and JZ designed the project. FT, DL, and JC performed experiments. FT, DL, JC, WL, WG, and RH analyzed the data. FY, LX, and JZ prepared the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (grant numbers 31770890, 31860276, and 31570826).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.751999/full#supplementary-material

REFERENCES

Abdi, F., Quinn, J. F., Jankovic, J., McIntosh, M., Leverenz, J. B., Peskind, E., et al. (2006). Detection of Biomarkers with a Multiplex Quantitative Proteomic Platform in Cerebrospinal Fluid of Patients with Neurodegenerative Disorders. *Jad* 9, 293–348. doi:10.3233/jad-2006-9309

Borrow, A. P., Stranahan, A. M., Suchecki, D., and Yunes, R. (2016). Neuroendocrine Regulation of Anxiety: Beyond the Hypothalamic-Pituitary-Adrenal Axis. *J. Neuroendocrinol.* 28. doi:10.1111/jne.12403

Chang, C.-h., and Grace, A. A. (2014). Amygdala-ventral Pallidum Pathway Decreases Dopamine Activity after Chronic Mild Stress in Rats. *Biol. Psychiatry* 76, 223–230. doi:10.1016/j.biopsych.2013.09.020

Chen, J.-j., Bai, S.-J., Li, W.-w., Zhou, C.-j., Zheng, P., Fang, L., et al. (2018). Urinary Biomarker Panel for Diagnosing Patients with Depression and Anxiety Disorders. *Transl. Psychiatry* 8, 192. doi:10.1038/s41398-018-0245-0

Cheng, P.-J., Wang, T.-H., Huang, S.-Y., Kao, C.-C., Lu, J.-H., Hsiao, C.-H., et al. (2011). Differential Proteomics Analysis of Amniotic Fluid in Pregnancies of Increased Nuchal Translucency with normal Karyotype. *Prenat. Diagn.* 31, 274–281. doi:10.1002/pd.2719

Chiba, S., Numakawa, T., Ninomiya, M., Richards, M. C., Wakabayashi, C., and Kunugi, H. (2012). Chronic Restraint Stress Causes Anxiety- and Depression-like Behaviors, Downregulates Glucocorticoid Receptor Expression, and Attenuates Glutamate Release Induced by Brain-Derived Neurotrophic Factor in the Prefrontal Cortex. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* 39, 112–119. doi:10.1016/j.pnpbp.2012.05.018

Cotman, C. W., Hailer, N. P., Pfister, K. K., Soltesz, I., and Schachner, M. (1998). Cell Adhesion Molecules in Neural Plasticity and Pathology: Similar Mechanisms, Distinct Organizations?. *Prog. Neurobiol.* 55, 659–669. doi:10.1016/s0301-0082(98)00025-2

Dakik, H., Mantash, S., Nehme, A., Kobeissy, F., Zabet-Moghaddam, M., Mirzaei, P., et al. (2021). Analysis of the Neuroproteome Associated with Cell Therapy after Intranigral Grafting in a Mouse Model of Parkinson Disease. *Front. Neurosci.* 15, 621121. doi:10.3389/fnins.2021.621121

Delvecchio, G., Altamura, A. C., Soares, J. C., and Brambilla, P. (2017). Pituitary Gland in Bipolar Disorder and Major Depression: Evidence from Structural MRI Studies. *J. Affective Disord.* 218, 446–450. doi:10.1016/j.jad.2017.03.066

Ditlevsen, D. K., Povlsen, G. K., Berezin, V., and Bock, E. (2008). NCAM-induced Intracellular Signaling Revisited. *J. Neurosci. Res.* 86, 727–743. doi:10.1002/jnr.21551

Gong, W., Liao, W., Fang, C., Liu, Y., Xie, H., Yi, F., et al. (2021). Analysis of Chronic Mild Stress-Induced Hypothalamic Proteome: Identification of Protein Dysregulations Associated with Vulnerability and Resiliency to Depression or Anxiety. *Front. Mol. Neurosci.* 14, 633398. doi:10.3389/fnmol.2021.633398

Hamilton, J. P., Chen, M. C., Waugh, C. E., Joormann, J., and Gotlib, I. H. (2015). Distinctive and Common Neural Underpinnings of Major Depression, Social Anxiety, and Their Comorbidity. *Soc. Cogn. Affect. Neurosci.* 10, 552–560. doi:10.1093/scan/nsu084

Harris, R. B. S. (2015). Chronic and Acute Effects of Stress on Energy Balance: Are There Appropriate Animal Models?. *Am. J. Physiology-Regulatory, Integr. Comp. Physiol.* 308, R250–R265. doi:10.1152/ajpregu.00361.2014

Henningsen, K., Palmfeldt, J., Christiansen, S., Baiges, I., Bak, S., Jensen, O. N., et al. (2012). Candidate Hippocampal Biomarkers of Susceptibility and Resilience to Stress in a Rat Model of Depression. *Mol. Cell Proteomics* 11, M111–1. 016428. doi:10.1074/mcp.M111.016428

Jefferson, S. J., Feng, M., Chon, U., Guo, Y., Kim, Y., and Luscher, B. (2020). Disinhibition of Somatostatin Interneurons Confers Resilience to Stress in Male but Not Female Mice. *Neurobiol. Stress* 13, 100238. doi:10.1016/j.ynstr.2020.100238

Krishnamurthy, D., Rahmoune, H., and Guest, P. C. (2017). Proteomic Profiling of the Pituitary Gland in Studies of Psychiatric Disorders. *Adv. Exp. Med. Biol.* 974, 313–319. doi:10.1007/978-3-319-52479-5_30

Krishnan, V., Han, M.-H., Graham, D. L., Berton, O., Renthal, W., Russo, S. J., et al. (2007). Molecular Adaptations Underlying Susceptibility and Resistance to Social Defeat in Brain Reward Regions. *Cell* 131, 391–404. doi:10.1016/j.cell.2007.09.018

Lee, S., and Rhee, D.-K. (2017). Effects of Ginseng on Stress-Related Depression, Anxiety, and the Hypothalamic-Pituitary-Adrenal axis. *J. Ginseng Res.* 41, 589–594. doi:10.1016/j.jgr.2017.01.010

Lenselink, A. M., Rotaru, D. C., Li, K. W., van Nierop, P., Rao-Ruiz, P., Loos, M., et al. (2015). Strain Differences in Presynaptic Function. *J. Biol. Chem.* 290, 15635–15645. doi:10.1074/jbc.M114.628776

Liu, C., Dai, J., Chen, Y., Qi, Z., Xin, F., Zhuang, Q., et al. (2021). Disorder- and Emotional Context-specific Neurofunctional Alterations during Inhibitory Control in Generalized Anxiety and Major Depressive Disorder. *NeuroImage: Clin.* 30, 102661. doi:10.1016/j.nicl.2021.102661

Lorenzetti, V., Allen, N. B., Fornito, A., Pantelis, C., De Plato, G., Ang, A., et al. (2009). Pituitary Gland Volume in Currently Depressed and Remitted Depressed Patients. *Psychiatry Res. Neuroimaging* 172, 55–60. doi:10.1016/j.pscychresns.2008.06.006

Lotan, A., Fenckova, M., Bralten, J., Alttoa, A., Dixson, L., Williams, R. W., et al. (2014). Neuroinformatic Analyses of Common and Distinct Genetic Components Associated with Major Neuropsychiatric Disorders. *Front. Neurosci.* 8, 331. doi:10.3389/fnins.2014.00331

Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., et al. (2019). iProX: an Integrated Proteome Resource. *Nucleic Acids Res.* 47, D1211–D1217. doi:10.1093/nar/gky869

Nieuwenhuizen, A. G., and Rutters, F. (2008). The Hypothalamic-Pituitary-Adrenal-axis in the Regulation of Energy Balance. *Physiol. Behav.* 94, 169–177. doi:10.1016/j.physbeh.2007.12.011

Oh, J.-Y., Liu, Q. F., Hua, C., Jeong, H. J., Jang, J.-H., Jeon, S., et al. (2020). Intranasal Administration of Melanin-Concentrating Hormone Reduces Stress-Induced Anxiety- and Depressive-like Behaviors in Rodents. *Exp. Neurobiol.* 29, 453–469. doi:10.5607/en20024

Stelzhammer, V., Alsaif, M., Chan, M. K., Rahmoune, H., Steeb, H., Guest, P. C., et al. (2015). Distinct Proteomic Profiles in post-mortem Pituitary Glands from Bipolar Disorder and Major Depressive Disorder Patients. *J. Psychiatr. Res.* 60, 40–48. doi:10.1016/j.jpsychires.2014.09.022

Sucic, S., El-Kasaby, A., Kudlacek, O., Sarker, S., Sitte, H. H., Marin, P., et al. (2011). The Serotonin Transporter Is an Exclusive Client of the Coat Protein Complex II (COPII) Component SEC24C. *J. Biol. Chem.* 286, 16482–16490. doi:10.1074/jbc.M111.230037

Tang, M., Huang, H., Li, S., Zhou, M., Liu, Z., Huang, R., et al. (2019). Hippocampal Proteomic Changes of Susceptibility and Resilience to Depression or Anxiety in a Rat Model of Chronic Mild Stress. *Transl. Psychiatry* 9, 260. doi:10.1038/s41398-019-0605-4

Thorp, J. G., Campos, A. I., Campos, A. I., Grotzinger, A. D., Gerring, Z. F., An, J., et al. (2021). Symptom-level Modelling Unravels the Shared Genetic Architecture of Anxiety and Depression. *Nat. Hum. Behav.*. doi:10.1038/s41562-021-01094-9

Tian, P., O'Riordan, K. J., Lee, Y.-k., Wang, G., Zhao, J., Zhang, H., et al. (2020). Towards a Psychobiotic Therapy for Depression: Bifidobacterium Breve CCFM1025 Reverses Chronic Stress-Induced Depressive Symptoms and Gut Microbial Abnormalities in Mice. *Neurobiol. Stress* 12, 100216. doi:10.1016/j.ynstr.2020.100216

Tsigos, C., and Chrousos, G. P. (2002). Hypothalamic-pituitary-adrenal axis, Neuroendocrine Factors and Stress. *J. Psychosomatic Res.* 53, 865–871. doi:10.1016/s0022-3999(02)00429-4

Wu, X., Yan, J., Wu, Y., Zhang, H., Mo, S., Xu, X., et al. (2019). Proteomic Analysis by iTRAQ-PRM Provides Integrated Insight into Mechanisms of Resistance in Pepper to Bemisia Tabaci (Gennadius). *BMC Plant Biol.* 19, 270. doi:10.1186/s12870-019-1849-0

Xu, H.-B., Zhang, R.-F., Luo, D., Zhou, Y., Wang, Y., Fang, L., et al. (2012). Comparative Proteomic Analysis of Plasma from Major Depressive Patients: Identification of Proteins Associated with Lipid Metabolism and Immunoregulation. *Int. J. Neuropsychopharm.* 15, 1413–1425. doi:10.1017/s1461145712000302

Yelamanchi, S. D., Tyagi, A., Mohanty, V., Dutta, P., Korbonits, M., Chavan, S., et al. (2018). Proteomic Analysis of the Human Anterior Pituitary Gland. *OMICS: A J. Integr. Biol.* 22, 759–769. doi:10.1089/omi.2018.0160

Yu, R., Liu, D., Yang, Y., Han, Y., Li, L., Zheng, L., et al. (2017). Expression Profiling-Based Clustering of Healthy Subjects Recapitulates Classifications Defined by Clinical Observation in Chinese Medicine. *J. Genet. Genomics* 44, 191–197. doi:10.1016/j.jgg.2017.01.001

Yun, S., Donovan, M. H., Ross, M. N., Richardson, D. R., Reister, R., Farnbauch, L. A., et al. (2016). Stress-Induced Anxiety- and Depressive-like Phenotype Associated with Transient Reduction in Neurogenesis in Adult Nestin-CreERT2/Diphtheria Toxin Fragment A Transgenic Mice. *PLoS One* 11, e0147256. doi:10.1371/journal.pone.0147256

Zhao, Y., Chen, L., Zhang, W., Xiao, Y., Shah, C., Zhu, H., et al. (2017). Gray Matter Abnormalities in Non-comorbid Medication-Naive Patients with Major Depressive Disorder or Social Anxiety Disorder. *EBioMedicine* 21, 228–235. doi:10.1016/j.ebiom.2017.06.013

Zhou, J., Liu, Z., Yu, J., Han, X., Fan, S., Shao, W., et al. (2016). Quantitative Proteomic Analysis Reveals Molecular Adaptations in the Hippocampal Synaptic Active Zone of Chronic Mild Stress-Unsusceptible Rats. *Ijnppy* 19, pyv100. doi:10.1093/ijnp/pyv100

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a shared affiliation with several of the authors FT, DL, JC, WL, WG, LX, YF, and JZ at time of review.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 December 2021
doi: 10.3389/fnins.2021.774638

White Matter Characteristics of Cognitive Impairment in Tap-Test Positive Idiopathic Normal Pressure Hydrocephalus: A Diffusion Tensor Tract-Based Spatial Study

Yufeng Tang[1†], Xiaoqin Yuan[1†], Jinfeng Duan[1], Xianwen Zhang[1], Jiao Chen[2], Ying Zhou[2], Fangzhou Song[3] and Dong Zhou[4*]*

[1] Department of Neurology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, China, [2] Department of Radiology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, China, [3] Basic Medicine College, Chongqing Medical University, Chongqing, China, [4] Department of Neurology, West China Hospital, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:
Guiyou Liu,
Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences (CAS), China

Reviewed by:
Yuyan Cheng,
UCLA Health System, United States
Jiaxing Wang,
Emory University, United States

***Correspondence:**
Fangzhou Song
fzsongcq@163.com
Dong Zhou
dongzhou66@yahoo.de

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: 12 September 2021
Accepted: 16 November 2021
Published: 03 December 2021

Citation:
Tang Y, Yuan X, Duan J, Zhang X, Chen J, Zhou Y, Song F and Zhou D (2021) White Matter Characteristics of Cognitive Impairment in Tap-Test Positive Idiopathic Normal Pressure Hydrocephalus: A Diffusion Tensor Tract-Based Spatial Study. Front. Neurosci. 15:774638. doi: 10.3389/fnins.2021.774638

The present study was designed to systemically evaluate changes in the diffusion tensor imaging (DTI)-derived parameters of iNPH (idiopathic normal pressure hydrocephalus) patients with different responses to the tap test (TT), and to correlate cognitive impairment with white matter (WM) degeneration. This study included 22 iNPH patients and 14 healthy controls with structural magnetic resonance imaging (MRI) and DTI scanning. DTI was used to explore the differences in fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) for all participants. DTI parameters were evaluated using an ROI (region of interest)-based and tract-based spatial statistics (TBSS) approach. Neuropsychological assessments and the idiopathic normal pressure hydrocephalus grading scoring scale (iNPHGS) were performed. Compared to the TT non-responders, the TT responders group had significantly lower FA values in the corpus callosum, cingulum cingulate gyrus, superior longitudinal fasciculus, and lower AD values in the right cingulum cingulate gyrus and the left posterior thalamic radiation. Besides, the MD values were significantly increased in the corpus callosum, left anterior corona radiata, and the RD values in the corpus callosum and cingulum cingulate gyrus. In addition, the cognitive improvement was negatively correlated with FA of the corpus callosum, cingulum cingulate gyrus, and MD values of the genu of corpus callosum. While, the cognitive improvement was positively related to the AD of the cingulum cingulate gyrus, superior longitudinal, and RD values of the corpus callosum, cingulum cingulate gyrus and uncinate fasciculus. The ROI specific WM lesions in iNPH patients are the underlying basis for cognitive impairment.

Keywords: idiopathic normal pressure hydrocephalus (iNPH), tap test, diffusion tensor imaging, tract-based spatial statistics (TBSS), cognitive impairment

INTRODUCTION

Idiopathic normal pressure hydrocephalus (iNPH) is a complex clinical disease with an undetermined etiology. The clinical characteristics of iNPH include gait disorders, cognitive impairment and urinary incontinence. Ventriculomegaly on neuroimaging and cerebrospinal fluid pressures ranging from 70 to 200 mm H_2O (1 mm H_2O = 0.0098 kPa) are primary diagnostic criteria for iNPH (Williams and Malm, 2016). iNPH is one of the few etiologies of reversible dementia. Ventriculo-peritoneal shunting (VPS) is an effective treatment for iNPH (Marmarou et al., 2005) that can significantly improve cognitive function in patients (Klinge et al., 2005; Liu et al., 2016).

The increased aging population across the world has resulted in dementia becoming a major global public health problem. As iNPH is a reversible form of dementia, the disease has become the focus of intense research efforts. The symptoms and neuroimaging findings of iNPH are similar to other neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) (Kang et al., 2013). All of these clinical entities mainly occur in elderly patients and so iNPH is often found along with other neurodegenerative diseases. According to the uniform diagnostic criteria (Marmarou et al., 2005), the postoperative effects in different iNPH vary significantly. The accurate prediction of the shunt response can distinguish patients with reversible dementia from other forms of the disease.

The tap test (TT) is the most widely used and effective method for the preoperative evaluation of iNPH (Martín-Láez et al., 2016). Patients diagnosed with iNPH show differential responses to the cerebrospinal fluid (Ko et al., 2017). Patients with a positive TT response can obtain obvious improvements in cognitive function after shunt surgery, whilst most TT negative patients usually experience very poor postoperative effects often with no change in cognitive deficits (McKhann and Mayeux, 2010; Wolfsegger and Topakian, 2017). These observations suggest that different mechanisms of cognitive impairment may occur between TT responders and non-responders and could potentially be used to predict cognitive function outcomes after surgery in iNPH patients.

The mechanism of cognitive impairment in iNPH patients remains unclear. The cognitive network is highly complex and its dysfunction in cognitive disorders is an area of intense research interest. Diffusion tensor imaging (DTI) is a magnetic resonance (MR) technique that has recently been used to study white matter (WM) degeneration in patients with iNPH. Amongst the DTI parameters, fractional anisotropy (FA) and mean diffusivity (MD) have been demonstrated as a useful index of WM impairment in iNPH patients (Kanno et al., 2011; Nicot et al., 2014; Radovnický et al., 2016). FA is the most widely used DTI parameter, which reflects the integrity of the axon and is highly sensitive to change in microstructure. MD quantifies cellular and membrane density whereas an increase in MD indicates cellularity, edema, and necrosis of WM (Tae et al., 2018). Previous studies observed lower FA and higher MD within various supratentorial regions including the corticospinal tract (CST), the corpus callosum (CC), and some subcortical WM (Hattori et al., 2011, 2012; Koyama et al., 2013; Daouk et al., 2014). However, few studies have systemically analyzed whole-brain WM microstructures and explored the relationship between the integrity of WM and cognitive decline. The DTI parameters of axial diffusivity (AD) and radial diffusivity (RD) have rarely been reported in previous iNPH studies (Scheel et al., 2012; Jurcoane et al., 2014). RD is a putative myelin marker and increases with demyelination. AD is related to axonal injury and thus decreases in cases of axonal damage (Tae et al., 2018). Furthermore, few studies have compared the differences between TT responders and non-responders in iNPH patients.

This study aimed to systemically evaluate the WM changes in iNPH patients with different responses to the TT, and to correlate cognitive impairment and WM microstructural damage in iNPH patients.

Abbreviations: DTI, diffusion tensor imaging; TBSS, tract-based spatial statistics; WM, white matter; FA, fractional anisotropy; MD, mean diffusivity; AD, axial diffusivity; RD, radial diffusivity; TT, tap test; TT-R, TT responsive group; TT-nR, TT non-responsive group; MMSE, Mini-Mental State Examination; DST, digit span forward; VFT-A, Verbal Fluency Test –ANIMAL; TMT-A, Trail Making Test A; CDT, Clock Drawing Test; CWT-B, Stroop Color Word Test- card B; L-ATR, anterior thalamic radiation L; R-ATR, anterior thalamic radiation R; L-PTR, posterior thalamic radiation include optic radiation L; R-PTR, posterior thalamic radiation include optic radiation R; L-ACR, anterior corona radiata R; R-ACR, anterior corona radiata L; L-SCR, superior corona radiata R; R-SCR, superior corona radiata L; L-PCR, posterior corona radiata R; R-PCR, posterior corona radiata L; F-major, forceps major; F-minor, forceps minor; GCC, genu of corpus callosum; BCC, body of corpus callosum; SCC, splenium of corpus callosum; L-TAP, tapetum L; R-TAP, tapetum R; FN, fornix (column and body of fornix); L-CgC, cingulum cingulate gyrus L; R-CgC, cingulum cingulate gyrus R; L-CgH, cingulum hippocampus L; R-CgH, cingulum hippocampus R; L-SFOF, superior fronto-occipital fasciculus L; R-SFOF, superior fronto-occipital fasciculus R; L-IFOF, inferior frontooccipital fasciculus L; R-IFOF, inferior frontooccipital fasciculus R; L-ILF, inferior longitudinal fasciculus L; R-ILF, inferior longitudinal fasciculus R; L-SLF, superior longitudinal fasciculus L; R-SLF, superior longitudinal fasciculus R; L-SLFT, superior longitudinal fasciculus temporal part L; R-SLFT, superior longitudinal fasciculus temporal part R; L-SS, sagittal stratum (include ILF and IFOF) L; R-SS, sagittal stratum (include ILF and IFOF) R; L-UF, uncinate fasciculus L; R-UF, uncinate fasciculus R.

MATERIALS AND METHODS

Participants

A total of 22 patients diagnosed with iNPH in the Neurology Department of Mianyang Central Hospital from May 2016 to December 2019 were included in this study. Before lumbar puncture and at 8, 24, 48, and 72 h after the drainage, gait disturbance, mini-mental state examination (MMSE) score, and the idiopathic normal pressure hydrocephalus grading scoring scale (iNPHGS) were assessed (Tarnaris et al., 2007). Gait improvements at any observation times after drainage, improvements in the MMSE score of $\geq$3 points, or improvements in the iNPHGS of $>$1 point were considered a positive criterion for the cerebrospinal fluid discharge test. The twenty-five iNPH patients consisted of 12 patients in the TT responsive group and 13 patients in the TT non-responsive group. A total of 14 control subjects with no cognitive impairments were included in the study across the same period.

Demographic and Clinical Data Collection

Cognitive function was assessed using the following tests:

1. The MMSE was used to test the subjects' overall cognitive level including orientation, immediate and short-term memory function, language function, and computational power (Folstein et al., 1975).
2. The digit span test (DST) was used to assess attention and immediate memory in memory function (Richardson, 2007).
3. The verbal fluency test animal (VFT-A) was used to assess working memory and vocabulary storage memory in executive functions, and long-term memory in memory function and semantic smooth function (Carlesimo et al., 1996).
4. The trail-making test A (TMT-A) was used to assess performance functions and attention (O'Leary et al., 1977).
5. The Stroop color-word test-card B (CWT-B) was used to assess attention (Jensen and Rohwer, 1966).
6. The clock drawing test (CDT, Huashan version) was used to assess multiple cognitive functions including the task plan in the executive function, the spatial mechanism function, the semantic and digital memory in the memory function, the abstract thinking ability, and the anti-interference ability (Olazarán et al., 2016).

The iNPHGS is a clinician-rated scale to evaluate the severity of core symptoms of iNPH (cognitive impairment, gait disturbance, and urinary disturbance). The score of each domain ranges from 0 to 4, with higher scores indicating worse symptoms (Tarnaris et al., 2007).

All the subjects were scored at the baseline before the tap test. All of the iNPH patients were scored at 8, 24, 48, and 72 after the tap test.

Magnetic Resonance Imaging Acquisition and Image Processing

Magnetic resonance imaging was performed on a 3.0T Siemens MAGNETON Skyra using a 12-channel head matrix radio frequency receive coil. The MR imaging protocol included a T1-weighted sequence (TR = 700 ms, TE = 11 ms, 0.9 mm slice separation, giving a voxel size 0.9 mm × 0.9 mm × 0.9 mm), a T2-weighted sequence (TR = 4,910 ms, TE = 99 ms, 5 mm slice separation, giving a voxel size 0.6 mm × 0.6 mm × 5 mm), and a fluid attenuated inversion recovery (FLAIR) sequence (TR = 8000, TE = 99 ms, 5 mm slice separation, giving a voxel size 0.9 mm × 0.9 mm × 5 mm). The DTI data set was acquired by using a spin echo diffusion weighted echo planar imaging sequence with the following parameters: TR = 10,400 ms; TE = 89 ms; FOV = 256 mm × 256 mm; acquisition matrix = 128 × 128; voxel size 2 mm × 2 mm × 2 mm; 75 axial slices; 4 images without (b0) and 60 images with diffusion weighting (b = 1,000 s/mm^{-2}) uniformly distributed across 60 gradient directions.

DTI data was processed using several approaches as follows:

a) Tract-based spatial statistics (TBSS): PANDA [Pipeline for Analyzing braiN Diffusion imAges, a MATLAB toolbox which consists of FMRIB Software Library (FSL) and several established packages] was used for the processing of the DTI raw data[1] (Cui et al., 2013). All of the DTI data of the subjects were automatically processed by TBSS to achieve the DTI scalars FA, MD, AD, and RD used in the analysis.
b) ABA-TBSS: FSL was used to generate a WM map (JHU DTI-based white-matter atlases) by separating all of the whole WM. This approach was used to automatically calculate the average skeleton value of each brain region. The outputs were saved in Excel file format.

Comparison of whole brain WM skeleton (TBSS): quantitative analysis of the whole brain WM skeleton was performed using the built-in TBSS randomize statistical tool in FSL. The statistical results were displayed using the xjview and fslview software packages. The regions of interest (ROIs) were mapped using the JHU DTI-based white-matter atlases: anterior thalamic radiation (ATR), posterior thalamic radiation include optic radiation (PTR), anterior corona radiata (ACR), superior corona radiata (SCR), posterior corona radiata (PCR), tapetum (TAP), cingulum cingulate gyrus (CgC), cingulum hippocampus (CgH), superior fronto-occipital fasciculus (SFOF), inferior frontooccipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), superior longitudinal fasciculus temporal part (SLFT), sagittal stratum (include ILF and IFOF) (SS), and uncinate fasciculus (UF) within each hemisphere, and forceps major (F-major), forceps minor (F-minor), genu of corpus callosum (GCC), body of corpus callosum (BCC), splenium of corpus callosum (SCC), and fornix (column and body of fornix) (FN) across hemispheres (shown in **Supplementary Figure 1**) (Alexander et al., 2007).

Statistical Analysis

Statistical analysis was performed using SPSS 20.0 software. A p-value threshold of <0.05 was used to determine the level of statistical significance. The demographic data, neuropsychological scores, and the iNPHGS scores were presented as the mean ± standard deviation. One-way analysis of variance (ANOVA) was used to compare demographic data and baseline cognitive scores among the control group and the iNPH patients (TT responsive and TT non-responsive groups). A Mann–Whitney U test was used to compare the maximum improvement scores (time duration) of the neuropsychological performance in the two iNPH groups after the TT. Comparison of the average skeletal values (FA, MD, AD, and RD) in the ROIs was performed using ANOVA among different groups. Bonferroni correction was used to control for multiple comparisons, while uncorrected results are also presented because Bonferroni's correction is quite conservative (Narum, 2006). Correlation analysis was performed between the DTI parameters (FA, MD, AD, and RD) and the MMSE scores, the total cognitive scores, and the improvement of

[1] http://www.nitrc.org/projects/panda/

TABLE 1 | Demographic and characteristics of all subjects.

Parameter	TT-R	TT-nR	Controls	*P*-value[1]	*P*-value[2]
Number, *n*	10	12	14	–	–
Age, y	76.10 ± 4.15	74.41 ± 7.53	75.18 ± 5.76	0.93	0.52
Gender, M/F	10/0	11/1	11/3	0.36	1
Education, y	6.50 ± 6.10	7.50 ± 5.93	7.05 ± 4.69	0.74	0.70
iNPHGS	7.90 ± 2.08	5.50 ± 1.31		–	0.00
MMSE	16.30 ± 7.45	19.42 ± 5.50	24.57 ± 2.59	0.00	0.27
DST	5.90 ± 1.66	6.17 ± 1.75	9.79 ± 1.48	0.00	0.72
VFT-A	6.90 ± 2.08	7.08 ± 2.35	13.36 ± 2.56	0.00	0.85
CWT-B	38.40 ± 8.49	42.67 ± 6.96	47.14 ± 2.60	0.00	0.21
TMT-A	11.58 ± 8.84	8.23 ± 8.37	0.93 ± 1.27	0.00	0.46
CDT	8.95 ± 8.76	11.15 ± 9.87	22.36 ± 7.82	0.00	0.55

[1]*Comparison between all iNPH patients (TT-R and TT-nR) and control subjects.*
[2]*Comparison between TT-R and TT-nR patients.*
TT, tap test; TT-R, TT responsive group; TT-nR, TT non-responsive group; MMSE, Mini-Mental State Examination; DST, Digit Span forward; VFT-A, Verbal Fluency Test –ANIMAL; TMT-A, Trail Making Test A; CDT, Clock Drawing Test; CWT-B, Stroop Color Word Test- card B.

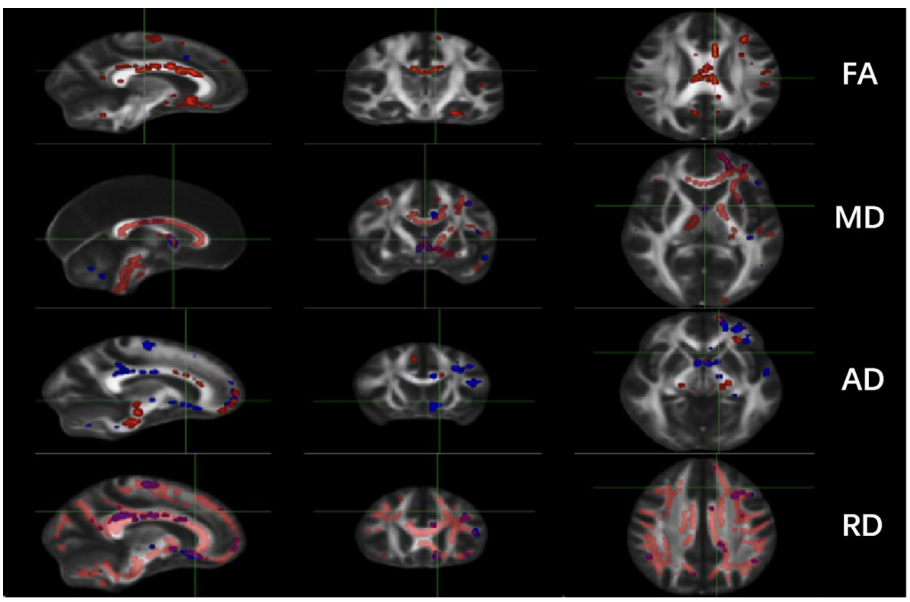

FIGURE 1 | Comparisons of DTI imaging in group analyses of TBSS. Results of TBSS between TT responsive group and TT non-responsive group. Significant region ($P < 0.05$) illustrated in warm colors for decreased values and in cool colors for increased values on mean WM skeleton. DTI, diffusion tensor imaging; TBSS, tract-based spatial statistics; FA, fractional anisotropy; MD, mean diffusivity; AD, axial diffusivity; RD, radial diffusivity.

cognitive scores by Pearson correlation analysis. The Pearson coefficient (*r*-value) > 0.4 and $P < 0.05$ were set to define moderate correlation.

RESULTS

Demographic and Clinical Profiles

The detailed demographic and clinical information from the patients is presented in **Table 1**. This study involved 22 patients with iNPH who met the inclusion criteria and 14 healthy controls (HCs). In the iNPH patient group, 20 patients were male and 2 patients were female. The average age of the patients in this group was 75.40 ± 5.83 years and the average education period was 6.92 ± 5.72 years. The control group consisted of 11 males and 3 females with an average age of 75.36 ± 5.76 years and an average education period of 6.43. ± 4.69 years. There were no significant differences in gender, age, and years of schooling between the iNPH patients and the HCs. The iNPH patients had significantly poorer performance in the MMSE, DST, VFT-A, CWT-B, TMT-A, and CDT compared to the HCs ($P < 0.05$). According to the improvements after the TT, 12 patients were classified in the TT responsive group (TT-R) and 13 patients were classified in the TT non-responsive group (TT-nR). No significant differences were

found between the TT-R and TT-nR groups in age, sex, years of schooling, and baseline cognitive levels (MMSE, DST, VFT-A, CWT-B, TMT-A, and CDT scores). The iNPHGS score of the TT-R group was higher than the TT-nR group suggesting that the clinical symptoms were more severe in the TT-R group.

Tract-Based Spatial Statistics Whole Brain White Matter Skeleton Comparison

Significant changes were observed in the TT-R group compared to the TT-nR group. These included decreases in the FA skeleton values in specific areas (GCC, BCC, SCC, F-major, FN, CgC, L-CgH, L-ATR, L-IFOF, SLF, L-UF) ($P < 0.05$), and decreased in the MD skeleton values in areas of the CC, FN, B-CgC, B-PTR, L-CgH, L-IFOF, SLF, L-UF ($P < 0.05$), AD values in areas of the GCC, BCC, SCC, F-major, CgC, PTR, IFOF ($P < 0.05$) and RD values in areas of the CC, CgC, CgH, PTR, IFOF, ILF, ACR, UF ($P < 0.05$) (shown in **Figure 1**).

The Region of Interest Average Skeleton Values Based on ABA-Tract-Based Spatial Statistics Comparison

Comparison of the TT-R and TT nR groups showed that the average FA skeleton values in the areas of the GCC, BCC, SCC, F-major, CgC, SLF, L-SLFT were significantly lower in the TT-R group ($P < 0.05$). The average MD skeleton values in the areas of the GCC, BCC, L-ACR, and the average RD skeleton values in the areas of GCC, BCC, SCC, CgC, L-SLFT, L-ACR were significantly increased in the TT-R group ($P < 0.05$). Also, the average AD skeleton values in the areas of the R-CgC, L-PTR, L-SLFT were significantly reduced in the TT-R group ($P < 0.05$) (Shown in **Table 2** and **Supplementary Table 1**).

Correlation Analysis Between the Region of Interest Average Skeleton Values and Cognitive Performance in Idiopathic Normal Pressure Hydrocephalus Patients

Correlation Analysis Between Region of Interest Average Skeleton Values (Fractional Anisotropy, Mean Diffusivity, Axial Diffusivity, and Radial Diffusivity) and Baseline Total Cognitive Scores in Idiopathic Normal Pressure Hydrocephalus Patients

The total cognitive scores of the iNPH patients were positively correlated with the average FA values of the GCC, BCC, SCC, F-major, F-minor, CgC, and the ILF ($r > 0.4$, $P < 0.05$). The total cognitive scores were negatively correlated with the average MD values of the GCC, BCC, SCC, F-major, F-minor, ACR, and SCR ($r > 0.4$, $P < 0.05$), the total cognitive scores were negatively correlated with the average AD values of the GCC and SCR ($r > 0.4$, $P < 0.05$) and positively correlated with the average AD values of the CgC ($r > 0.4$, $P < 0.05$). The total cognitive scores were negatively correlated with the average RD values of the GCC, BCC, SCC, F-major, F-minor, CgC, SLF, and ACR ($r > 0.4$, $P < 0.05$) (shown in **Table 3** and **Supplementary Table 2**).

TABLE 2 | ABA-TBSS analysis results.

	TT-R vs. HC				TT-nR vs. HC				TT-R vs. TT-nR			
ROI	**FA**	**MD**	**AD**	**RD**	**FA**	**MD**	**AD**	**RD**	**FA**	**MD**	**AD**	**RD**
ATR-L												
ATR-R									↓			
CgC-L	↓			↑					↓			↑
CgC-R	↓		↓	↑	↓		↓	↑	↓		↓	↑
CgH-L				↑								
CgH-R												
F-major	↓			↑	↓			↑				
GCC	↓	↑	↑	↑	↓	↑	↑	↑	↓	↑		↑
BCC	↓	↑		↑					↓	↑		↑
SCC	↓	↑		↑	↓	↑		↑	↓			↑
FN	↓											
ACR-R	↓	↑	↑	↑		↑	↑	↑		↑		
ACR-L	↓	↑	↑	↑		↑	↑			↑	↑	↑
F-minor	↓	↑		↑	↓	↑	↑	↑	↓			
IFOF-L	↓	↑		↑	↓	↑		↑				
IFOF-R	↓			↑	↓	↑		↑				
ILF-L	↓				↓							
ILF-R	↓			↑	↓							
SLF-L	↓			↑	↓	↑			↑		↓	
SLF-R	↓			↑	↓				↑	↓		
UF-L	↓	↑		↑								
UF-R												
SLFT-L	↓			↑						↓		
SLFT-R											↓	
SCR-L		↑	↑	↑			↑					↑
SCR-R		↑	↑				↑					
PCR-R						↑	↑					
PCR-LPTR-R												
PTR-L	↓		↓		↓				↑		↓	
SS-R	↓			↑		↑						
SS-L	↓			↑	↓	↑			↑			
SFOF-R	↓		↑				↑					
SFOF-L			↑		↑		↑					
TAP-R												
TAP-L												

Up arrows (↑) indicate higher values, and down arrows (↓) indicate lower values in the former group compared with the later group.

TT, tap test; TT-R, TT responsive group; TT-nR, TT non-responsive group; HC, healthy controls; TBSS, tract-based spatial statistics; FA, fractional anisotropy; MD, mean diffusivity; AD, axial diffusivity; RD, radial diffusivity; L-ATR, anterior thalamic radiation L; R-ATR, anterior thalamic radiation R; L-PTR, posterior thalamic radiation include optic radiation L; R-PTR, posterior thalamic radiation include optic radiation R; L-ACR, anterior corona radiata R; R-ACR, anterior corona radiata L; L-SCR, superior corona radiata R; R-SCR, superior corona radiata L; L-PCR, posterior corona radiata R; R-PCR, posterior corona radiata L; F-major, forceps major; F-minor, forceps minor; GCC, genu of corpus callosum; BCC, body of corpus callosum; SCC, splenium of corpus callosum; L-TAP, tapetum L; R-TAP, tapetum R; FN, fornix (column and body of fornix); L-CgC, cingulum cingulate gyrus L; R-CgC, cingulum cingulate gyrus R; L-CgH, cingulum hippocampus L; R-CgH, cingulum hippocampus R; L-SFOF, superior fronto-occipital fasciculus L; R-SFOF, superior fronto-occipital fasciculus R; L-IFOF, inferior frontooccipital fasciculus L; R-IFOF, inferior frontooccipital fasciculus R; L-ILF, inferior longitudinal fasciculus L; R-ILF, inferior longitudinal fasciculus R; L-SLF, superior longitudinal fasciculus L; R-SLF, superior longitudinal fasciculus R; L-SLFT, superior longitudinal fasciculus temporal part L; R-SLFT, superior longitudinal fasciculus temporal part R; L-SS, sagittal stratum (include ILF and IFOF) L; R-SS, sagittal stratum (include ILF and IFOF) R; L-UF, uncinate fasciculus L; R-UF, uncinate fasciculus R.

TABLE 3 | Correlation analysis between the ROI average skeleton values and cognitive performance in iNPH patients.

	TS				MMSE				CI			
ROI	**FA**	**MD**	**AD**	**RD**	**FA**	**MD**	**AD**	**RD**	**FA**	**MD**	**AD**	**RD**
ATR-L												
ATR-R												
CgC-L	0.353			−0.380					−0.680			0.547
CgC-R	0.533		0.427	−0.479	0.441			−0.412	−0.591		−0.472	
CgH-L		−0.368		−0.395								
CgH-R												
F-major	0.546	−0.451		−0.522	0.436	−0.360		−0.418				
F-minor	0.412	−0.456	−0.351	−0.458		−0.373		−0.377	−0.531			
IFOF-L												
IFOF-R												
LIF-L												
ILF-R	0.411											
SLF-L	0.389	−0.380		−0.420					−0.473		−0.519	
SLF-R				−0.353								
UF-L												
UF-R												
SLFT-L									−0.527			
SLFT-R											−0.557	
GCC	0.446	−0.576	−0.493	−0.540	0.389	−0.524	−0.439	−0.495	−0.608			0.531
BCC	0.430	−0.508	−0.362	−0.509	0.433	−0.507	−0.346	−0.514	−0.606	0.475		0.574
SCC	0.512	−0.537		−0.541	0.394	−0.398		−0.413	−0.496			
FN												
ACR-R		−0.400	−0.380	−0.400								
ACR-L		−0.354	−0.386								0.463	
SCR-R												
SCR-L		−0.446	−0.525	−0.368		−0.401	−0.447	−0.345				
PCR-R		−0.343	−0.360									
PCR-L												
PTR-R												
PTR-L	0.348										−0.503	
SS-R			−0.149	−0.336							−0.497	
SS-L												
SFOF-R												
SFOF-L			−0.396									
TAP-R												
TAP-L												

Only significant correlations were displayed, and more detailed information were shown in the ***Supplementary Materials****.*
TS, Total scores; CI, cognitive improvement; MMSE, Mini-Mental State Examination; FA, fractional anisotropy; MD, mean diffusivity; AD, axial diffusivity; RD, radial diffusivity; L-ATR, anterior thalamic radiation L; R-ATR, anterior thalamic radiation R; L-PTR, posterior thalamic radiation include optic radiation L; R-PTR, posterior thalamic radiation include optic radiation R; L-ACR, anterior corona radiata R; R-ACR, anterior corona radiata L; L-SCR, superior corona radiata R; R-SCR, superior corona radiata L; L-PCR, posterior corona radiata R; R-PCR, posterior corona radiata L; F-major, forceps major; F-minor, forceps minor; GCC, genu of corpus callosum; BCC, body of corpus callosum; SCC, splenium of corpus callosum; L-TAP, tapetum L; R-TAP, tapetum R; FN, fornix (column and body of fornix); L-CgC, cingulum cingulate gyrus L; R-CgC, cingulum cingulate gyrus R; L-CgH, cingulum hippocampus L; R-CgH, cingulum hippocampus R; L-SFOF, superior fronto-occipital fasciculus L; R-SFOF, superior fronto-occipital fasciculus R; L-IFOF, inferior frontooccipital fasciculus L; R-IFOF, inferior frontooccipital fasciculus R; L-ILF, inferior longitudinal fasciculus L; R-ILF, inferior longitudinal fasciculus R; L-SLF, superior longitudinal fasciculus L; R-SLF, superior longitudinal fasciculus R; L-SLFT, superior longitudinal fasciculus temporal part L; R-SLFT, superior longitudinal fasciculus temporal part R; L-SS, sagittal stratum (include ILF and IFOF) L; R-SS, sagittal stratum (include ILF and IFOF) R; L-UF, uncinate fasciculus L; R-UF, uncinate fasciculus R.

Correlation Between the Region of Interest Average Skeleton Values (Fractional Anisotropy, Mean Diffusivity, Axial Diffusivity, and Radial Diffusivity) and Cognitive Improvements After the Tap Test in Idiopathic Normal Pressure Hydrocephalus Patients

A moderate negative correlation was observed between the cognitive improvement and the mean FA of GCC, BCC, SCC, F-minor, CgC, SLF, and SLFT in iNPH patients after lumbar puncture ($r > 0.4$, $P < 0.05$), and also with the mean MD of GCC ($r > 0.4$, $P < 0.05$). The degree of cognitive improvement in iNPH patients was positively related to the mean AD of the CgC, SLF, SLFT, and SS ($r > 0.4$, $P < 0.05$), and also with the average RD values of the GCC, BCC, SCC, CgC, ATR, and UF ($r > 0.4$, $P < 0.05$) (shown in **Table 3** and **Supplementary Material 2**).

DISCUSSION

The underlying mechanisms of cognitive impairment have been the major research efforts in the field of iNPH research. The TT is the most widely used and effective method for preoperative evaluation of iNPH and the test is used to clinically classify patients into two groups. Patients who respond to the TT can achieve improvements in cognitive performance after shunt surgery, whilst patients who are not responsive to the TT usually experience very poor postoperative effects. These observations suggest different mechanisms of cognitive impairment between TT-R and TT-nR patients.

The present study used the TBSS method and a quantitative ROI analysis of skeletonized brain maps to compare differences in the cognitive-related WM microstructure of iNPH patients with different TT responses. Our data showed that the microstructural WM damage in TT-R patients was significantly more severe than in TT-nR patients. Furthermore, we assessed the associations between FA, MD, AD, and RD values and the cognitive performance of iNPH patients.

The mean FA in the areas of the GCC, BCC, SCC, F-major, FN, B-CgC, L-CgH, L-ATR, L-IFOF, SLF, and L-UF were significantly lower in the TT responsive compared to the TT non-responsive group ($P < 0.05$). Also, the MD and RD in the area of the GCC, BCC, SCC, B-CgC, and ATR were significantly increased ($P < 0.05$). These results indicated that WM edema or the destruction of myelin sheath were more severe in the TT-R group than in the TT-nR group.

In comparison to the subjects in the HC group, patients in the TT-R and TT-nR groups had more extensive microstructural damage presenting with lower FA, higher MD, and RD in CC, CgC, ATR, ACR, SLF, and UF. However, no significant difference was found in preoperative cognitive performance between the TT-R and TT-nR groups. These data may indicate that the cognitive dysfunction in the TT-R group is mostly caused by WM injury, whilst cortical volume atrophy plays a more contributing role in the cognitive decline of patients in the TT-nR group. This hypothesis is supported by previous studies. Kang et al. (2013) found that CSFTT non-responders had statistically significant cortical thinning in the left superior frontal gyrus compared to responders suggesting that comorbid AD pathology might be related to the cortical thinning patterns found in CSFTT non-responders. Also, biopsy studies found that iNPH patients with pathological evidence of AD exhibited more severe initial symptoms and had lower shunt responsiveness compared to patients without AD (Golomb et al., 2000; Savolainen et al., 2002; Picascia et al., 2016).

Correlation analysis showed significant associations between lower FA, higher MD, higher AD, higher RD and poor executive performance in the GCC, BCC, SCC, F-major, and F-minor. The data indicated that the corpus callosum and cingulate gyrus are involved in both memory and executive function, furthermore, different parts of the corpus callosum may participate in different cognitive functions.

The data presented in this study are compatible with other previous studies (Bettcher et al., 2016). Our data showed that WM damage is dominated by the anterior and superior part of the lateral ventricle in iNPH patients, particularly in the frontal lobe, which may account for the prominent executive impairment of iNPH patients. Also, our results showed that more severe damage in the anterior, outer, and upper regions of the periventricular WM obtains more obvious cognitive improvement after cerebrospinal fluid drainage in patients with iNPH. Particularly, the RD values in the corpus callosum and cingulate gyrus were significantly associated with cognitive improvement suggesting that the edema and WM degeneration in the area of anterior and superior lateral ventricles were reversible. WM demyelination, wallerian degeneration, and late axonal degeneration are short-term irreversible processes. In contrast, edema and early axonal degeneration of WM are reversible pathologies. We speculate that the TT rapidly reduces the extravasation pressure of the cerebrospinal fluid by releasing cerebrospinal fluid, and subsequently reduces edema in the periventricular WM. The changes significantly improve cognitive function in iNPH patients. However, the extent of WM fiber stretching in this area is not related to the degree of cognitive improvement after drainage. Our findings may suggest that edema in the area of anterior and superior lateral ventricles contributed mostly to the reversible cognitive impairment.

Overall, these observations confirm the role of the hydrocephalus effect in the occurrence of reversible cognitive impairment in iNPH patients.

This study had several limitations as the analysis was performed on a relatively small sample size and further validation is required in larger patient cohorts. Also, due to the low acceptance rate of shunt surgery for iNPH patients in the Chinese population, our study lacked postoperative follow-up data to further validate the longer-term responses of patients. Prospective cohort studies need to be designed to confirm the values of DTI parameters as a non-invasive imaging biomarker to predict post-operative cognitive improvement in iNPH patients. At last, combining more scales of information, such as radiomics features, might lead to more fine-grained findings in the future.

CONCLUSION

1. The extensive microstructural damage of cognitive WM in iNPH patients is the material basis for the development of cognitive impairment.
2. The microstructural damage of the anterior superior ventricle and the WM in the frontal ventricle in TT positive iNPH patients was greater than TT negative patients.
3. The microstructural changes of the CC, the cingulate ligament and the adjacent radiant fibers can affect the memory and executive functions in the cognitive field of iNPH patients, whilst microstructural changes of the anterior subcortical WM in the frontal lobe mainly affect the executive features.
4. The more severe the edema degeneration of WM in the anterior superior region of the lateral ventricle, the more obvious the cognitive improvement in the iNPH patients after the TT. The decrease in the WM FA value and increase

of RD value in this region has diagnostic and prognostic value in iNPH patients.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of Mianyang Central Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YT and DZ: conceptualization. JC, YZ, and YT: methodology. XZ and JD: project administration. FS and DZ: supervision. XY: writing – original draft. YT: writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This manuscript received funding from the Mianyang Central Hospital for open access publication fees.

ACKNOWLEDGMENTS

The authors thank the patients for their participation in the study and thank financial support from Mianyang Central Hospital.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.774638/full#supplementary-material

REFERENCES

Alexander, A. L., Lee, J. E., Lazar, M., and Field, A. S. (2007). Diffusion tensor imaging of the brain. *Neurotherapeutics* 4, 316–329.

Bettcher, B. M., Mungas, D., Patel, N., Elofson, J., Dutt, S., Wynn, M., et al. (2016). Neuroanatomical substrates of executive functions: beyond prefrontal structures. *Neuropsychologia* 85, 100–109. doi: 10.1016/j.neuropsychologia.2016.03.001

Carlesimo, G. A., Caltagirone, C., and Gainotti, G. (1996). The mental deterioration battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. the group for the standardization of the mental deterioration battery. *Eur. Neurol.* 36, 378–384. doi: 10.1159/000117297

Cui, Z., Zhong, S., Xu, P., He, Y., and Gong, G. (2013). PANDA: a pipeline toolbox for analyzing brain diffusion images. *Front. Hum. Neurosci.* 7:42. doi: 10.3389/fnhum.2013.00042

Daouk, J., Chaarani, B., Zmudka, J., Capel, C., Fichten, A., Bouzerar, R., et al. (2014). Relationship between cerebrospinal fluid flow, ventricles morphology, and DTI properties in internal capsules: differences between Alzheimer's disease and normal-pressure hydrocephalus. *Acta Radiol.* 55, 992–999. doi: 10.1177/0284185113508112

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Golomb, J., Wisoff, J., Miller, D. C., Boksay, I., Kluger, A., Weiner, H., et al. (2000). Alzheimer's disease comorbidity in normal pressure hydrocephalus: prevalence and shunt response. *J. Neurol. Neurosurg. Psychiatry* 68, 778–781. doi: 10.1136/jnnp.68.6.778

Hattori, T., Ito, K., Aoki, S., Yuasa, T., Sato, R., Ishikawa, M., et al. (2012). White matter alteration in idiopathic normal pressure hydrocephalus: tract-based spatial statistics study. *AJNR Am. J. Neuroradiol.* 33, 97–103.

Hattori, T., Yuasa, T., Aoki, S., Sato, R., Sawaura, H., Mori, T., et al. (2011). Altered microstructure in corticospinal tract in idiopathic normal pressure hydrocephalus: comparison with Alzheimer disease and Parkinson disease with dementia. *AJNR Am. J. Neuroradiol.* 32, 1681–1687. doi: 10.3174/ajnr.A2570

Jensen, A. R., and Rohwer, W. D. Jr. (1966). The stroop color-word test: a review. *Acta Psychol.* 25, 36–93. doi: 10.1016/0001-6918(66)90004-7

Jurcoane, A., Keil, F., Szelenyi, A., Pfeilschifter, W., Singer, O. C., and Hattingen, E. (2014). Directional diffusion of corticospinal tract supports therapy decisions in idiopathic normal-pressure hydrocephalus. *Neuroradiology* 56, 5–13. doi: 10.1007/s00234-013-1289-8

Kang, K., Yoon, U., Lee, J. M., and Lee, H. W. (2013). Idiopathic normal-pressure hydrocephalus, cortical thinning, and the cerebrospinal fluid tap test. *J. Neurol. Sci.* 334, 55–62. doi: 10.1016/j.jns.2013.07.014

Kanno, S., Abe, N., Saito, M., Takagi, M., Nishio, Y., Hayashi, A., et al. (2011). White matter involvement in idiopathic normal pressure hydrocephalus: a voxel-based diffusion tensor imaging study. *J. Neurol.* 258, 1949–1957.

Klinge, P., Marmarou, A., Bergsneider, M., Relkin, N., and Black, P. M. (2005). Outcome of shunting in idiopathic normal-pressure hydrocephalus and the value of outcome assessment in shunted patients. *Neurosurgery* 57, S40–S52.

Ko, P. W., Lee, H. W., and Kang, K. (2017). Frontal assessment battery and cerebrospinal fluid tap test in idiopathic normal-pressure hydrocephalus. *Eur. Neurol.* 77, 327–332. doi: 10.1159/000472712

Koyama, T., Marumoto, K., Domen, K., and Miyake, H. (2013). White matter characteristics of idiopathic normal pressure hydrocephalus: a diffusion tensor tract-based spatial statistic study. *Neurol. Med. Chir.* 53, 601–608. doi: 10.2176/nmc.oa2012-0307

Liu, A., Sankey, E. W., Jusué-Torres, I., Patel, M. A., Elder, B. D., Goodwin, C. R., et al. (2016). Clinical outcomes after ventriculoatrial shunting for idiopathic normal pressure hydrocephalus. *Clin. Neurol. Neurosurg.* 143, 34–38. doi: 10.1016/j.clineuro.2016.02.013

Marmarou, A., Black, P., Bergsneider, M., Klinge, P., and Relkin, N. (2005). Guidelines for management of idiopathic normal pressure hydrocephalus: progress to date. *Acta Neurochir. Suppl.* 95, 237–240. doi: 10.1007/3-211-32318-x_48

Martín-Láez, R., Caballero-Arzapalo, H., Valle-San Román, N., ópez-Menéndez, L. L., Arango-Lasprilla, J. C., and ázquez-Barquero, A. V. (2016). Incidence of idiopathic normal-pressure hydrocephalus in Northern Spain. *World Neurosurg.* 87, 298–310. doi: 10.1016/j.wneu.2015.10.069

McKhann, G., and Mayeux, R. (2010). Brain drain: a bottom-up approach to normal pressure hydrocephalus. *Anna. Neurol.* 68, 415–417. doi: 10.1002/ana.22212

Narum, S. R. J. C. G. (2006). Beyond bonferroni: less conservative analyses for conservation genetics. *Conserv. Genet.* 7, 783–787.

Nicot, B., Bouzerar, R., Gondry-Jouet, C., Capel, C., Peltier, J., Fichten, A., et al. (2014). Effect of surgery on periventricular white matter in normal pressure hydrocephalus patients: comparison of two methods

of DTI analysis. *Acta Radiol.* 55, 614–621. doi: 10.1177/0284185113504570

Olazarán, J., Hoyos-Alonso, M. C., del Ser, T., Garrido Barral, A., Conde-Sala, J. L., Bermejo-Pareja, F., et al. (2016). Practical application of brief cognitive tests. *Neurologia* 31, 183–194.

O'Leary, M. R., Radford, L. M., Chaney, E. F., and Schau, E. J. (1977). Assessment of cognitive recovery in alcoholics by use of the trail-making test. *J. Clin. Psychol.* 33, 579–582.

Picascia, M., Minafra, B., Zangaglia, R., Gracardi, L., Pozzi, N. G., Sinforiani, E., et al. (2016). Spectrum of cognitive disorders in idiopathic normal pressure hydrocephalus. *Funct. Neurol.* 31, 143–147.

Radovnický, T., Adámek, D., Derner, M., and Sameš, M. (2016). Fractional anisotropy in patients with disproportionately enlarged subarachnoid space hydrocephalus. *Acta Neurochir.* 158, 1495–1500.

Richardson, J. T. (2007). Measures of short-term memory: a historical review. *Cortex* 43, 635–650.

Savolainen, S., Hurskainen, H., Paljärvi, L., Alafuzoff, I., and Vapalahti, M. (2002). Five-year outcome of normal pressure hydrocephalus with or without a shunt: predictive value of the clinical signs, neuropsychological evaluation and infusion test. *Acta Neurochir.* 144, 515–523; discussion523.

Scheel, M., Diekhoff, T., Sprung, C., and Hoffmann, K. T. (2012). Diffusion tensor imaging in hydrocephalus–findings before and after shunt surgery. *Acta Neurochir.* 154, 1699–1706. doi: 10.1007/s00701-012-1377-2

Tae, W. S., Ham, B. J., Pyun, S. B., Kang, S. H., and Kim, B. J. (2018). Current clinical applications of diffusion-tensor imaging in neurological disorders. *J. Clin. Neurol.* 14, 129–140.

Tarnaris, A., Stephenson, R. F., and Cipolotti, L. (2007). Diagnosis, treatment, and analysis of long-term outcomes in idiopathic normal-pressure hydrocephalus. *Neurosurgery* 60:E208.

Williams, M. A., and Malm, J. (2016). Diagnosis and treatment of idiopathic normal pressure hydrocephalus. *Continuum* 22, 579–599.

Wolfsegger, T., and Topakian, R. (2017). Cognitive impairment predicts worse short-term response to spinal tap test in normal pressure hydrocephalus. *J. Neurol. Sci.* 379, 222–225. doi: 10.1016/j.jns.2017.06.028

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 December 2021
doi: 10.3389/fgene.2021.757014

Fam20C Overexpression Predicts Poor Outcomes and is a Diagnostic Biomarker in Lower-Grade Glioma

Jing Feng[1,2†], Jinping Zhou[3†], Lin Zhao[4], Xinpeng Wang[5], Danyu Ma[5], Baoqing Xu[6], Feilai Xie[6], Xingfeng Qi[6], Gang Chen[7], Hu Zhao[8] and Junxin Wu[9,10]**

[1]Department of Radiation Oncology, The Third Clinical Medical College of Fujian Medical University, Fuzhou, China, [2]The 900th Hospital of Joint Logistic Support Force, PLA, Fuzhou, China, [3]Department of Clinical Quality Management, The 900th Hospital of Joint Logistic Support Force, PLA, Fuzhou, China, [4]Department of Neurosurgery, The 900th Hospital of Joint Logistic Support Force, PLA, Fuzhou, China, [5]Department of Radiation Oncology, The 900th Hospital of Joint Logistic Support Force, PLA, Fuzong Clinical Medical College of Fujian Medical University, Fuzhou, China, [6]Department of Pathology, The 900th Hospital of Joint Logistic Support Force, PLA, Fuzhou, China, [7]Department of Pathology, The 900th Hospital of Joint Logistic Support Force, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou, China, [8]Department of General Surgery, The 900th Hospital of Joint Logistic Support Force, PLA, Fujian Medical University, Fuzhou, China, [9]Department of Radiation Oncology, Fujian Medical University Cancer Hospital, The Third Clinical Medical College of Fujian Medical University, Fuzhou, China, [10]Fujian Key Laboratory of Translational Cancer Medical, Fujian Cancer Hospital, Fujian Provincial Clinical Research Center for Cancer Radiotherapy and Immunotherapy, Fuzhou, China

Glioma is a relatively low aggressive brain tumor. Although the median survival time of patients for lower-grade glioma (LGG) was longer than that of patients for glioblastoma, the overall survival was still short. Therefore, it is urgent to find out more effective molecular prognostic markers. The role of the Fam20 kinase family in different tumors was an emerging research field. However, the biological function of Fam20C and its prognostic value in brain tumors have rarely been reported. This study aimed to evaluate the value of Fam20C as a potential prognostic marker for LGG. A total of 761 LGG samples (our cohort, TCGA and CGGA) were included to investigate the expression and role of Fam20C in LGG. We found that Fam20C was drastically overexpressed in LGG and was positively associated with its clinical progression. Kaplan-Meier analysis and a Cox regression model were employed to evaluate its prognostic value, and Fam20C was found as an independent risk factor in LGG patients. Gene set enrichment analysis also revealed the potential signaling pathways associated with Fam20C gene expression in LGG; these pathways were mainly enriched in extracellular matrix receptor interactions, cell adhesion, cell apoptosis, NOTCH signaling, cell cycle, etc. In summary, our findings provide insights for understanding the potential role of Fam20C and its application as a new prognostic biomarker for LGG.

Keywords: FAM20C, lower-grade gliomas, LGG, biomarker, prognosis, bioinformatics

OPEN ACCESS

Edited by:
Zhijie Han,
Chongqing Medical University, China

Reviewed by:
Aijun Shan,
Jinan University, China
Winnie Liang,
Translational Genomics Research Institute (TGen), United States

****Correspondence:***
Hu Zhao
zhaohubear@163.com
Junxin Wu
junxinwufj@aliyun.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: *11 August 2021*
Accepted: *18 October 2021*
Published: *14 December 2021*

Citation:
Feng J, Zhou J, Zhao L, Wang X, Ma D, Xu B, Xie F, Qi X, Chen G, Zhao H and Wu J (2021) Fam20C Overexpression Predicts Poor Outcomes and is a Diagnostic Biomarker in Lower-Grade Glioma. Front. Genet. 12:757014. doi: 10.3389/fgene.2021.757014

INTRODUCTION

Malignant central nervous system tumors account for 31.5% of nervous system tumors, and gliomas account for 80.7% of malignant central nervous system tumors (Goodenberger and Jenkins 2012; Ostrom et al., 2018). Global cancer statistics in 2018 showed that nervous system cancer was the 19th most common cancer in the world, with 296,851 new cases, accounting for 1.6% of the total cancer incidence, and 241,037 deaths each year, accounting for 2.5% of the total case mortality (Bray et al., 2018). According to the World Health Organization (WHO) 2016 version of the central nervous

system classification, diffuse gliomas include WHO grade II and grade III astrocytic tumors, grade II and III oligodendrogliomas, and grade IV glioblastomas (Louis et al., 2016; Wesseling and Capper 2018).

At present, the standard treatment of glioma includes surgical resection to the maximum safety range followed by postoperative radiotherapy and chemotherapy (Stupp et al., 2005). However, the prognosis of glioma is still poor. Glioblastoma (GBM) is the most aggressive type of brain tumor in adults. Despite the improvement of current treatment methods, the median survival time is only 17–23 months (Xu et al., 2017; Jiang et al., 2019). The median survival time of WHO grade II-III glioblastoma is longer than that of WHO grade IV glioblastoma, with a median survival time of 1.7–13.3 years (Buckner et al., 2016; Mair et al., 2021; van den Bent 2014). There is extensive heterogeneity among lower-grade glioma patients. Some patients could survive for many years without any treatments; however, other patients progress quickly after active treatment. Therefore, it is very important to find more effective molecular prognostic markers for the treatment of patients with LGG. Understanding the pathogenesis and etiology of LGG may assist in discovering advanced treatment methods and effective biomarkers for diagnosis and prognosis.

The Fam20 kinase family is a newly discovered class of secreted kinases that can phosphorylate secreted proteins and proteoglycans. This family includes Fam20A, Fam20B, and Fam20C (Nalbant et al., 2005; Zhang et al., 2018). Fam20C is a casein kinase protein enriched in the Golgi that can phosphorylate a variety of secreted proteins (Tagliabracci et al., 2014; Cozza et al., 2018). Protein phosphorylation modification refers to the process of transferring the phosphate group of ATP or GTP to the amino acid residue of the substrate protein through the catalytic effect of a protein kinase (Fischer 2013). This process mediates most of the signal transduction in eukaryotic cells and it regulates many cellular processes, including metabolic regulation, transcription regulation, cell cycle, cytoskeleton rearrangement, apoptosis, and differentiation (Manning et al., 2002; Sreelatha et al., 2015). Abnormal protein phosphorylation is the leading cause of many diseases, including cancer, diabetes, Alzheimer's disease, and Parkinson's disease (Fischer 2013; Klement and Medzihradszky 2017).

Fam20C is located inside the cell, but it may also play an important role outside the cell (Wang et al., 2013; Tagliabracci et al., 2015). Fam20C has been shown to phosphorylate secreted proteins by recognizing the protein motif "Ser-x-Glu/phospho-Ser," thereby being involved in biomineralization, lipid homeostasis, cell adhesion and migration. More importantly, many Fam20C substrates are related to tumor cell apoptosis and metastasis, including insulin-like growth factor binding proteins, osteopontin, and serine protease inhibitors (Rangaswami et al., 2006; Baxter 2014; Zhang et al., 2020). Insulin-like growth factor binding protein 7 (IGFBP7), which depends on Fam20C phosphorylation, could induce cell migration (Bieche et al., 2004; Georges et al., 2011). However, the utility of Fam20C as a potential tumor diagnostic and prognostic marker has not been fully elucidated.

In this study, we found that Fam20C was overexpressed in a variety of cancers, including LGG. High expression of Fam20C was associated with tumor progression. Therefore, Fam20C may serve as a potential biomarker for the diagnosis and prognosis of LGG. Moreover, the transcriptional expression of Fam20C in LGG patients may be an independent risk factor for survival. In addition, pathway and function enrichment indicated that the mechanism of Fam20C-mediated tumorigenesis involves extracellular matrix receptor interactions, cell adhesion, and the cell cycle. Our results clarified the important role of Fam20C in the prognosis of LGG and provided a reliable biomarker for the diagnosis and prognosis of LGG.

MATERIALS AND METHODS

Data Acquisition and Processing

LGG gene expression data and clinical information were obtained from The Cancer Genome Atlas TCGA database (http://cancergenome.nih.gov/) and the Chinese Glioma Genome Atlas CGGA database (http://www.cgga.org.cn). From the TCGA database, we obtained the original mRNAseq data of 529 LGG samples, which were normalized using the edge R package in R (version 4.0.2). A total of 132 LGG samples were obtained from the CGGA database, and the gene expression profile of each sample and the corresponding clinical data were sorted (**Supplementary Table S1**). The RNA-seq data from the CGGA database were generated from total RNA and directly expressed as fragment values per thousand bases per million mapped reads (FPKM). In CGGA database, a rapid hematoxylin and eosin-stain for frozen sections was applied to each sample to assess the tumor cell proportion before RNA extraction. In addition, the RNA was extracted from only those samples with >80% tumor cells (Zhao Z. et al., 2021).

Patient Information and Ethics

This study was approved by the ethics committee of 900th Hospital of Joint Logistics Support Force. Between January 2016 and November 2020, a cohort assessment of 100 patients who underwent neurosurgery was conducted. According to the WHO 2007 and 2016 standards, all patients were newly diagnosed with grade II and III gliomas. Patients younger than 16 years old at the time of diagnosis were excluded from this study (**Supplementary Table S1**). Clinical data and detailed follow-up data were obtained from all patients. Sanger sequencing was then employed to investigate the mutation status of isocitrate dehydrogenase (IDH). In addition, we also studied the 1p/19q deletion and the heterozygosity status of LGG using fluorescence *in situ* hybridization.

Immunohistochemistry Analysis

One hundred patients with LGG and three normal brain tissues from grade 1 glioma patients in the 900th Hospital of Joint Logistics Support Force were collected. The adjacent brain tissues to the three cases of grade 1 glioma patients were used as normalized data. The surgical specimens were fixed with 40 g/L formaldehyde solution, routinely embedded in paraffin, cut into

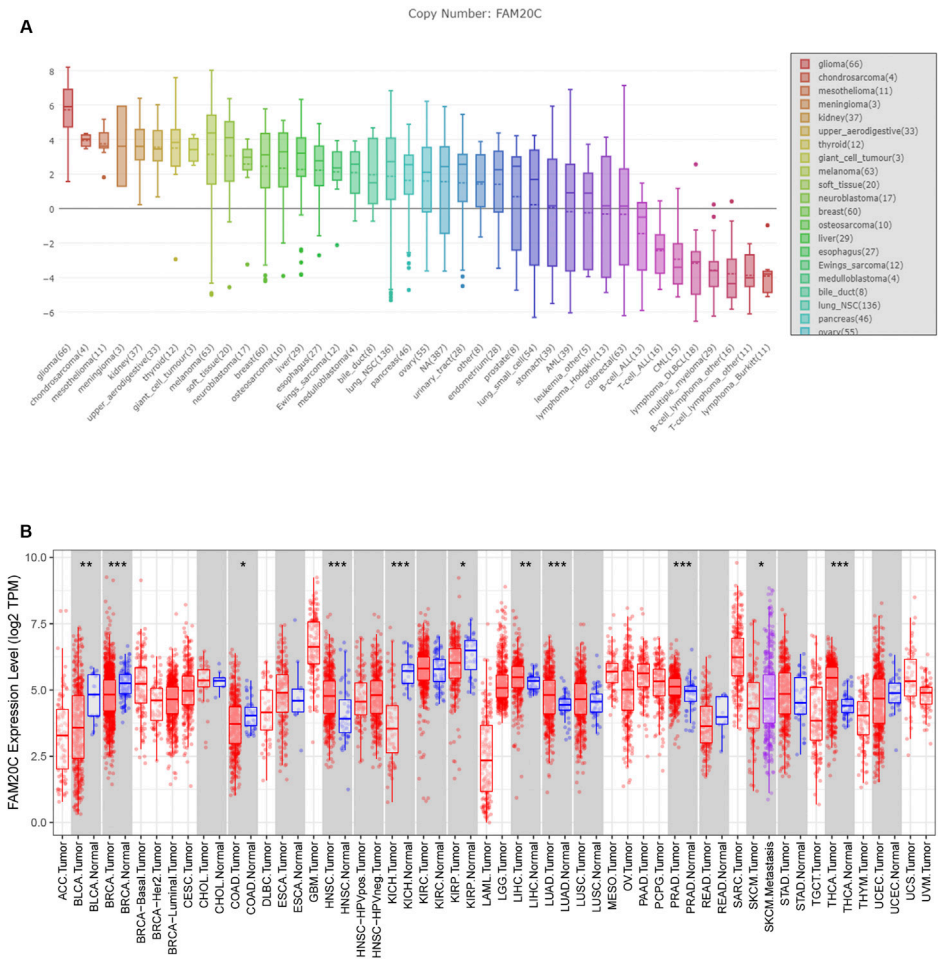

FIGURE 1 | The expression of FAM20C in different types of cancer, including glioma. **(A)** The expression of FAM20C in different types of cancer cells was obtained from the CCLE database, including glioma (n = 66), chondrosarcoma (n = 4), mesothelioma (n = 11), meningioma (n = 3), kidney (n = 37), upper aerodigestive (n = 33), thyroid (n = 12), giant cell tumour (n = 3), melanoma (n = 63), soft tissue (n = 20), neuroblastoma (n = 17), breast (n = 60), osteosarcoma (n = 10), liver (n = 29), esophagus (n = 27), Ewing's sarcoma (n = 12), medulloblastoma (n = 4), bile duct (n = 8), lung NSC (n = 136), pancreas (n = 46), ovary (n = 55), urinary tract (n = 28), endometrium (n = 28), prostate (n = 8), lung small cell (n = 54), stomach (n = 39), acute myeloid leukemia (n = 39), leukemia other (n = 5), lymphoma Hodgkin (n = 13), colorectal (n = 63), B cell acute lymphoblastic leukemia (n = 13), T cell acute lymphoblastic leukemia (n = 16), chronic myelogenous leukemia (n = 15), lymphoma DLBCL (n = 18), multiple myeloma (n = 29), B cell lymphoma other (n = 16), T cell lymphoma other (n = 11), and lymphoma Burkitt (n = 11); **(B)** the expression of FAM20C in different types of cancer was obtained from Tumor Immune Estimation Resource database, including ACC (n = 77), BLCA (n = 423), BRCA (n = 1197), CESC (n = 309), CHOL (n = 45), COAD (n = 316), DLBC (n = 47), ESCA (n = 195), GBM (n = 163), HNSC (n = 563), KICH (n = 91), KIRC (n = 595), KIRP (n = 318), LAML (n = 173), brain LGG (n = 518), LIHC (n = 419), LUAD (n = 542), LUSC (n = 542), MESO (n = 87), OV (n = 426), PAAD (n = 183), PCPG (n = 185), PRAD (n = 544), READ (n = 102), SARC (n = 264), SKCM (n = 462), STAD (n = 444), TGCT (n = 137), THCA (n = 571), THYM (n = 120), UCEC (n = 187), UCS (n = 57), and UVM (n = 79). *p < 0.05; ***p < 0.001.

4 μm-thick sections, and stained with HE. The EliVision method was used for Fam20C immunohistochemical staining and the results were observed through light microscopy. Anti-Fam20C polyclonal antibody, was purchased from Abcam, UK (product number ab154740). Non-biotin universal two-step immunohistochemistry kit (mouse/rabbit enhanced polymer detection system) was purchased from Beijing Zhongshan Jinqiao Biotechnology Co., Ltd. The positive control tissue in this experiment was glioblastoma tumor tissues (Du et al., 2020). Results interpretation criteria: Fam20C positive expression means brown-yellow particles in the nucleus and cytoplasm. Dark brown in the nucleus and cytoplasm of the cells was defined as a strong cell; Cells with yellow or brown nucleus and cytoplasm were defined as medium-strength cells; The nucleus and cytoplasm of the cells were light yellow or had faintly visible staining, which was defined as a weak intensity cell. No staining of nucleus and cytoplasm was negative. The histochemical score (H-score) was employed to quantify the expression of Fam20C. H-score = (percentage of weak intensity cells×1) + (percentage of medium intensity cells 2) + (percentage of strong cells×3).

Gene Set Enrichment Analysis (GSEA)

GSEA was conducted to detect whether a set of a priori defined genes showed statistically significant differential expression between the high and low Fam20C expression groups during the MSigDB set enrichment process, with 1000 genome

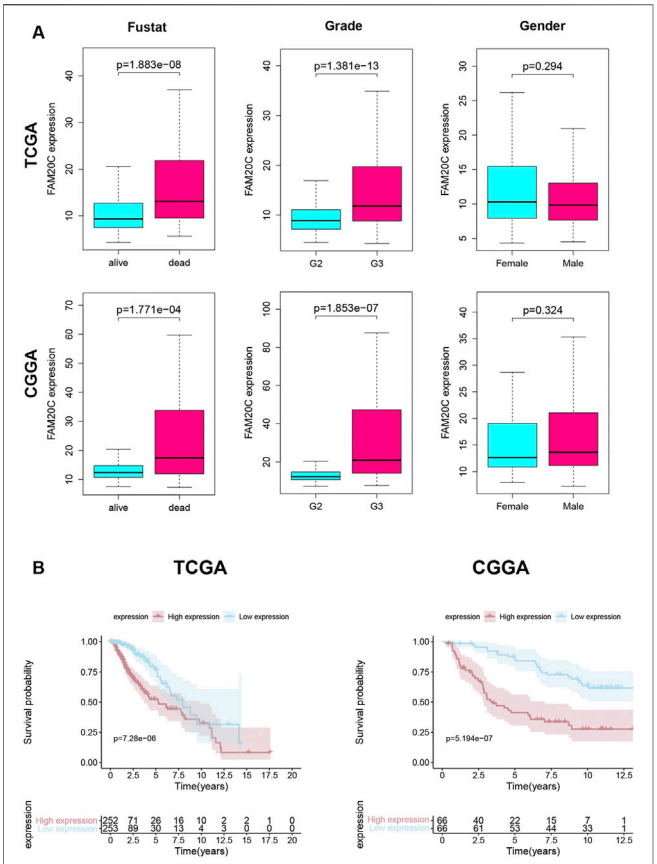

FIGURE 2 | Association with FAM20C expression and clinicopathological characteristics. **(A)** Clinical in TCGA database, including grade (grade 2 $n = 248$, and grade 3 $n = 261$); fustat (alive $n = 400$, and dead $n = 109$); gender (male $n = 281$, and female $n = 228$); Clinical in CGGA database, including grade (grade 2 $n = 87$, and grade 3 $n = 45$); fustat (alive $n = 68$, and dead $n = 64$); gender (male $n = 81$, and female $n = 51$); TCGA, The Cancer Genome Atlas. CGGA, Chinese Glioma Genome Atlas. **(B)** Kaplan–Meier curves for OS in LGG Higher FAM20C expression was remarkably associated with poorer OS in TCGA database; Higher FAM20C expression was remarkably associated with poorer OS in CGGA database. OS, overall survival. The fustat means the patients' survival status.

permutations performed per analysis. In this study, GSEA first generated an ordered list of all genes based on the correlation between the genes and Fam20C expression. Then, GSEA was performed to clarify the significance of the difference in survival between the high and low Fam20C expression groups. The expression level of Fam20C was used as the phenotype label. The phenotypic enrichment pathways were ranked by the nominal p value and normalized enrichment score. The calculation results were given using the ggplot2 R packages.

Functional Enrichment Analysis

Gene Ontology (GO) was employed to detect the function of the differentially expressed genes. The analysis gained a new understanding of the biological effects of Fam20C. The genes related to Fam20C expression (absolute Pearson correlation coefficient>0.5 and $p < 0.05$) were regarded as risk score-related genes, and their potential biological functions and pathways were determined. The Ggplot2 software package in R software was employed to analyze the GO pathways. The enrichment analysis of GO was based on a p-value and a q-value threshold <0.05.

Statistical Analysis

The Wilcoxon signed-rank test was used to detect the expression of Fam20C. The correlation between the clinicopathological characteristics and Fam20C expression was tested with the Wilcoxon signed-rank test. The survival ROC software package in R software was used to generate receiver operating characteristic (ROC) curves to evaluate the diagnostic value of Fam20C expression. The area under the curve represents the diagnostic value. Using the Survival package in R, the overall survival (OS) rates of the high expression group and the low expression group were compared by Kaplan-Meier analysis. Univariate Cox analysis was used to determine the potential survival rate, and multivariate Cox analysis was used to determine whether Fam20C expression was an independent risk factor for OS in LGG patients. $p < 0.05$ was considered statistically significant. All data were processed using R software (version 4.0.2) and Adobe Photoshop CC.

RESULTS

Fam20C Was Overexpressed in LGG

Data from the Cancer Cell Line Encyclopedia (CCLE) database showed that Fam20C was highly expressed in multiple cancer cell lines, especially glioma (**Figure 1A**). At present, there are few studies on the relationship between Fam20C and tumorigenesis. To determine the expression of Fam20C in other tumors, we conducted a comprehensive analysis of 33 tumors in TCGA. Among them, there were five cancer types in which Fam20C was overexpressed (**Figure 1B**).

Overexpressed Fam20C Was Associated With Advanced LGG

Next, we analyzed the correlation between the level of Fam20C mRNA in LGG patients and their clinicopathological parameters. The TCGA database includes the patient's tumor grade, sex, and survival status. The CGGA database includes the patient's tumor grade, sex, survival status, IDH mutation/wild-type, and 1p19q joint deletion status. As shown in **Figure 2A**, the higher the grade of the tumor, the higher the Fam20C expression level. In addition, in the TCGA database, high expression of the Fam20C gene was positively related to grade and survival status but not to sex. In the CGGA database, higher Fam20C expression was related to grade, survival status, IDH wild-type, and 1p19q nonjoint deletion but not to sex (**Figure 2A** and **Supplementary Figure S1**).

Since high expression of Fam20C in LGG patients was related to tumor grade, we further tried to determine whether this overexpression of Fam20C in LGG patients was related to a poor prognosis through the use of Kaplan-Meier curves. As shown in **Figure 2B**, higher Fam20C expression levels were significantly correlated with a worse OS in both the TCGA and CGGA datasets (**Figure 2B**). In general, the results showed that the expression of Fam20C was significantly

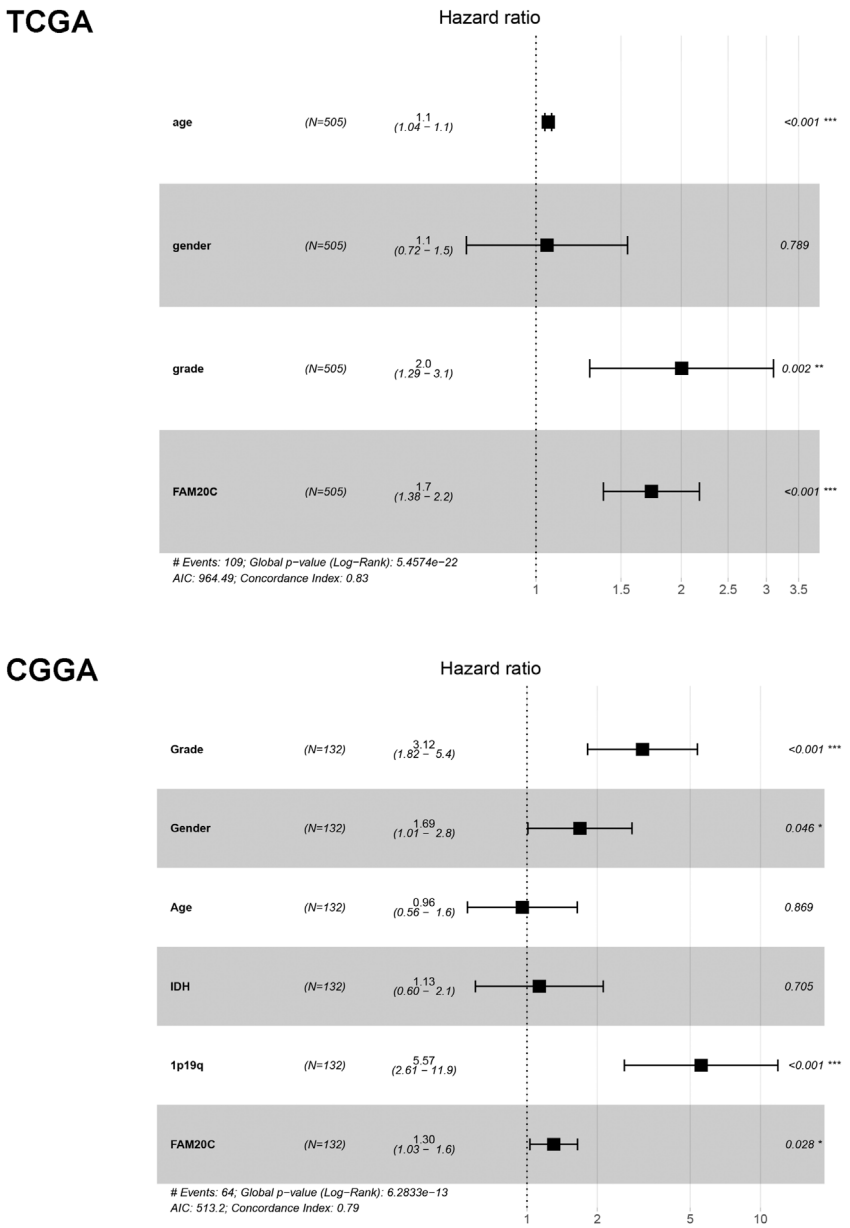

FIGURE 3 | Multivariate Cox analysis evaluating independently predictive ability of Fam20c for OS in TCGA and CGGA database. $^{**}p < 0.01$; $^{***}p < 0.001$.

related to the prognosis of LGG patients and could be used as a biomarker to predict the survival of LGG patients.

High Fam20C Expression Served as an Independent Risk Factor Among LGG Patients

Univariate and multivariate Cox analyses were utilized to evaluate the independent prognostic values of Fam20C expression in LGG patients. The univariate analysis results showed that high Fam20C expression was significantly correlated with a shorter OS (HR = 1.02, 95% CI: 1.01–1.03, $p < 0.001$; HR = 1.01, 95% CI: 1.00–1.01, $p = 0.001$) in TCGA and CGGA. Other variables related to poor survival included age and grade in TCGA (**Supplementary Table S2**). In CGGA, variables related to poor survival that including grade IDH and 1p19q (**Supplementary Table S3**). Multivariate analysis showed that high expression of Fam20C in LGG patients was independently associated with a significant decrease in OS (**Figure 3** and **Supplementary Tables S2, S3**).

Fam20C Expression Is a Novel Diagnostic Biomarker for LGG

To evaluate the diagnostic value of Fam20C for LGG, TCGA RNA-seq data were employed to draw the ROC curve. The area

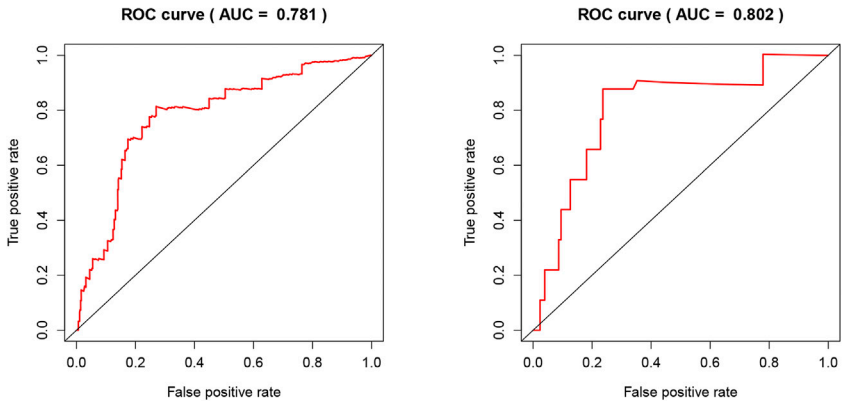

FIGURE 4 | Diagnosis value of FAM20C expression in LGG analysis. **(A)** ROC curve for FAM20C expression in LGG tissues in TCGA database; **(B)** validation of FAM20C diagnosis value in CGGA database. ROC, receiver operating characteristic.

under the ROC curve was 0.690, which had high diagnostic value (**Figure 4A**). This result was further verified with the CGGA data set, and the area under the ROC curve was 0.778 (**Figure 4B**).

Functional Enrichment Analysis

To clarify the functions and signaling pathways of genes co-expressed with Fam20C, we performed GO and KEGG enrichment analyses. GO analysis results showed that co-expressed genes were mainly closely related to the biological process of extracellular matrix remodeling (**Figure 5A**). KEGG analysis showed that co-expressed genes were mainly enriched in extracellular matrix receptor interactions, cell adhesion, apoptosis, cancer pathways, P53 signaling pathways, NOTCH signaling pathways, and cell cycle signaling pathways (**Figure 5B**).

Fam20C Was Also Overexpressed in Our Cohort

To further verify the expression of Fam20C in our cohort, we detected its expression in our clinical samples and found that Fam20C was significantly overexpressed in grade 3 tumors (**Figure 6A**). Higher Fam20C expression levels were also correlated with a worse OS in our cohort (**Figure 6B**).

Univariate and multivariate Cox analyses were utilized to evaluate the independent prognostic values of Fam20C expression in LGG patients. The univariate analysis results showed that high Fam20C expression was significantly correlated with a shorter OS (HR = 6.39, 95% CI: 1.86–21.86, p = 0.003). Other variables related to poor survival included IDH 1p19q and extent of resection (**Supplementary Table S4**). Multivariate analysis showed that high expression of Fam20C in LGG patients was independently associated with a significant decrease in OS (**Figure 6C** and **Supplementary Table S4**).

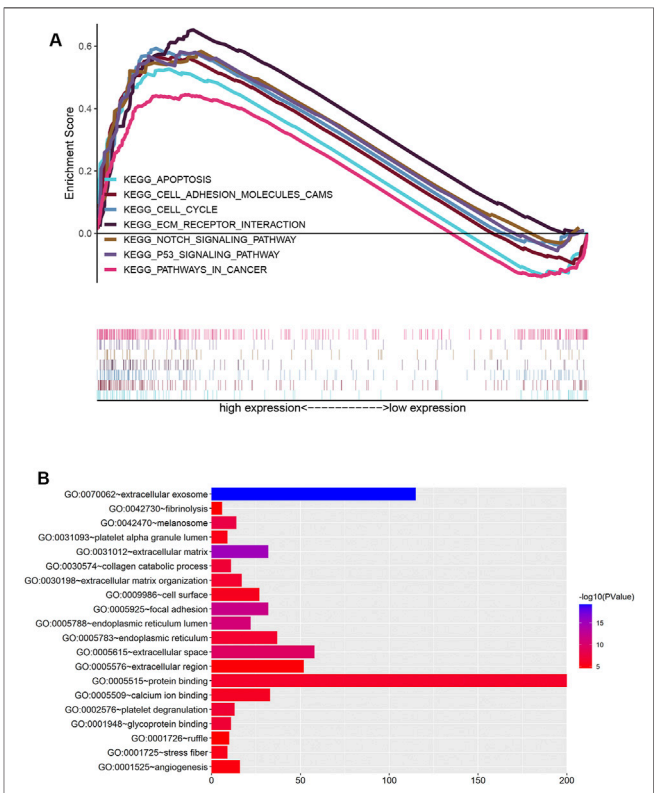

FIGURE 5 | Functional enrichment analysis of Fam20c in LGG. **(A)** Gene Ontology enrichment analysis; **(B)** enrichment plots from GSEA. KEGG, Kyoto Encyclopedia of Genes and Genomes.

DISCUSSION

Glioma is one of the most common primary malignant tumors in the nervous system. It arises from active glial cells in the brain, including astrocytes, oligodendrocytes, and ependymal cells.

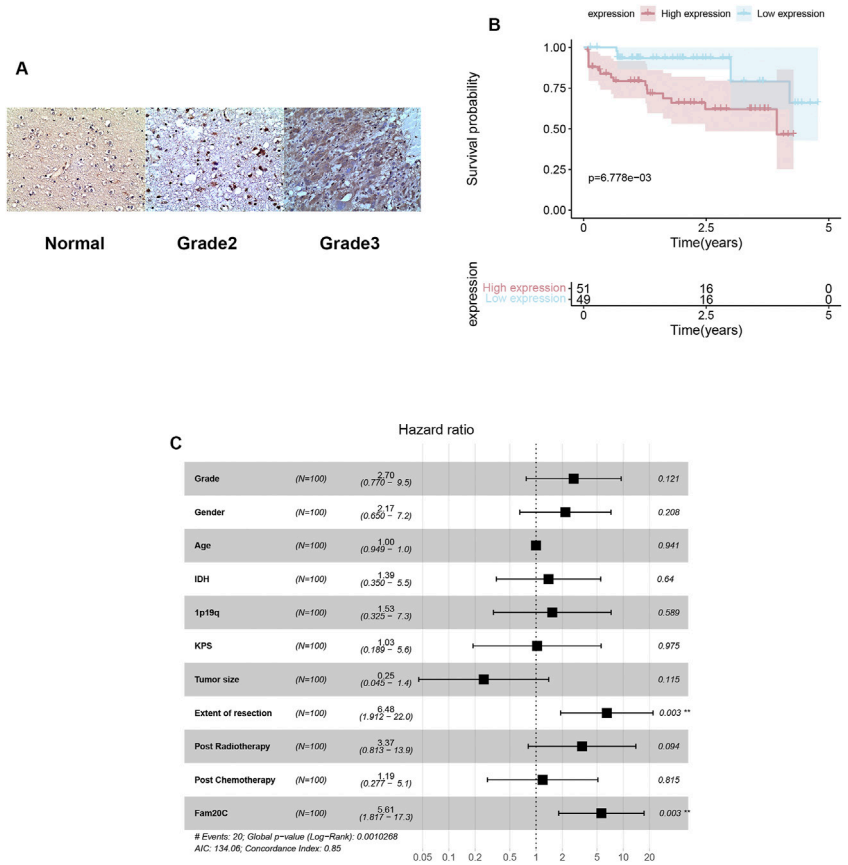

FIGURE 6 | Expressions, immunohistochemistry and multivariate Cox analysis of Fam20c in our cohort. **(A)** Representative figures of FAM20C immune-staining in our clinical LGG samples (200X; grade II: $n = 60$, grade III: $n = 40$, normal: $n = 3$); **(B)** Kaplan–Meier curve evaluating the correlation between FAM20C protein expression and LGG patients' survival (FAM20C low vs high, low $n = 51$, high $n = 49$, $p < 0.001$; Log rank test). **(C)** Multivariate Cox analysis evaluating independently predictive ability of Fam20c for OS.

Although the prognosis of lower-grade glioma is better than that of glioblastoma, there are still some lower-grade gliomas with a poor prognosis and a short survival time, and 70% of low-grade patients undergo a high-grade transformation within 10 years. Therefore, early diagnosis and accurate prognostic biomarkers are essential for improving the prognosis of patients with LGG.

In recent years, a class of secreted kinases have been newly discovered that are involved in the regulation of many important physiological reactions. The Fam20 kinase family includes Fam20A, Fam20B, and Fam20C (Nalbant et al., 2005; Zhang et al., 2018). Fam20C is a casein kinase enriched in the Golgi apparatus that modulates many downstream substrates through protein phosphorylation and plays an important role in the formation of the secretome of tumor cells. However, its diagnostic and prognostic value in cancer is still unclear. Our results provide insights for further understanding the pathological role of Fam20C in promoting tumor growth and invasion and its potential value as a diagnostic and prognostic marker for LGG.

Fam20C protein kinase has a significant promotion effect on the metastasis and invasion of triple-negative breast cancer (Tagliabracci et al., 2015). Fam20 is also a potential target gene related to the pathogenesis of early lung adenocarcinoma (Kang 2013). Therefore, we speculate that the expression of Fam20C may affect the survival of patients through promoting the progression of tumor cells. However, the expression of Fam20C in cancer and its effects on other important aspects, such as tumor cell metastasis, still lack consensus. It has been previously reported that insulin-like growth factor binding protein 7 (IGFBP-7) regulates the migration of glioma cells through the AKT-ERK pathway, thereby playing an important role in the growth and migration of gliomas (Jiang et al., 2008). Adult diffuse glial tumor GWAS contains variants of D2HGDH and Fam20C in different molecular subtypes. In IDH mutant gliomas, the nine variants located on chromosome two of D2HGDH and those in its vicinity are all significant genome-wide (Eckel-Passow et al., 2020).

In this study, we systematically detected the expression level of Fam20C in different types of cancer in the TCGA database. Based on the available evidence, our results indicated that Fam20C expression was elevated in breast cancer. In addition, Fam20C was also overexpressed in five other cancers, such as glioma, meningioma, and kidney cancer, and Fam20C overexpression was associated with higher-grade gliomas.

At present, the biological functions of Fam20C and its mechanism of action in tumorigenesis have rarely been reported. The phosphorylated substrate of Fam20C is related to tumor cell apoptosis and migration and can accelerate the process of tumor metastasis by activating matrix metalloproteinases

(MMPs). In this study, the Fam20C-related signaling pathways activated in LGG were mainly enriched in extracellular matrix receptor interactions, cell adhesion, apoptosis, cancer pathways, P53 signaling pathways, NOTCH signaling pathways, and the cell cycle, which further stimulated tumor proliferation and invasion.

Biomarkers are biological characteristics that could be objectively measured or evaluated. They may be employed as indicators of biological and pathological processes, or reflect the results of treatment methods, which are mainly used for disease prevention, diagnosis, treatment, prognosis, and drug development. Fam20C was an effective target for the treatment of triple-negative breast cancer. Fam20C inhibitor induced apoptosis of TNBC cell line (MDA-MB-468) and potentially inhibited cell migration (Qin et al., 2016). In our study, Fam20C expression was detected in postoperative pathological specimens of resectable glioma patients. Our present data has demonstrated that Fam20c may be a protentional prognostic marker for LGG. There is no research on whether Fam20C was highly expressed in the serum of LGG patients, we will reconsider and complete this topic in the future.

A Fam20C inhibitor induced cell apoptosis through the mitochondrial pathway and had the potential to inhibit cell migration (Qin et al., 2016; Zhao R. et al., 2021). Shaonan Du et al. found that Fam20C may serve as a predictive protein and a therapeutic target for GBM (Du et al., 2020). However, there have been few studies on the Fam20C gene in LGG. Hence, we further investigated whether Fam20C could be used as a diagnostic and prognostic marker for LGG. The ROC curve showed that the expression of Fam20C had a high diagnostic value for LGG. In addition, the Kaplan-Meier curves showed that high expression of Fam20C mRNA in LGG patients was significantly associated with a poor OS. In addition, univariate and multivariate Cox analyses showed that high Fam20C expression was an independent risk factor for a poor OS of LGG patients. In summary, our research showed that Fam20C was over-expressed in LGG and was correlated with more aggressive tumors and a worse prognosis. Our results showed that Fam20C is a promising biomarker for LGG diagnosis and prognosis.

CONCLUSION

In conclusion, we established a potential prognostic and diagnostic signature for LGG patients based on two databases (TCGA and CGGA) and clinical samples. This biomarker could efficiently stratify the LGG patients into two groups with distinct survival differences. Moreover, we identified the potential signaling pathways of Fam20C in LGG patients. Overexpression of Fam20C was correlated with progressive malignancy and poor survival of LGG patients and was associated with significant enrichment of extracellular matrix receptor interactions, cell adhesion and apoptosis in LGG. Taken together, our results suggest that Fam20C inhibition could be a potential therapeutic target to prevent LGG progression.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories at: http://github.com/fengjing0314/jing. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JF performed conception and design, provision of study materials or patients, collection and assembly of data, data analysis and interpretation, manuscript writing, and final approval of article. JZ and LZ performed provision of study materials or patients, collection and assembly of data, data analysis and interpretation, article writing, and final approval of manuscript. XW performed collection and assembly of data, data analysis and interpretation, article writing, and final approval of manuscript. DM performed provision of study materials or patients, collection and assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript. BX, FX, and XQ performed provision of study materials or patients, collection and assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript. GC performed conception and design, administrative support, collection and assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript. HZ and JW performed conception and design, administrative support, provision of study materials or patients, collection and assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript.

FUNDING

This work was supported by the National Clinical Key Specialty Construction Program (No. 2020Y2012), the Science and Technology Program of Fujian Province (No. 2019L3018 and 2019YZ016006) and the Research Project of the 900th Hospital of Joint Logistic Support Force (No. 2018J18).

ACKNOWLEDGMENTS

We would like to acknowledge Weidong Wu for the help of computer software.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.757014/full#supplementary-material

REFERENCES

Baxter, R. C. (2014). IGF Binding Proteins in Cancer: Mechanistic and Clinical Insights. *Nat. Rev. Cancer* 14, 329–341. doi:10.1038/nrc3720

Bièche, I., Lerebours, F., Tozlu, S., Espie, M., Marty, M., and Lidereau, R. (2004). Molecular Profiling of Inflammatory Breast Cancer. *Clin. Cancer Res.* 10, 6789–6795. doi:10.1158/1078-0432.CCR-04-0306

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68, 394–424. doi:10.3322/caac.21492

Buckner, J. C., Shaw, E. G., Pugh, S. L., Chakravarti, A., Gilbert, M. R., Barger, G. R., et al. (2016). Radiation Plus Procarbazine, CCNU, and Vincristine in Low-Grade Glioma. *N. Engl. J. Med.* 374, 1344–1355. doi:10.1056/NEJMoa1500925

Cozza, G., Moro, E., Black, M., Marin, O., Salvi, M., Venerando, A., et al. (2018). The Golgi 'casein Kinase' Fam20C Is a Genuine 'phosvitin Kinase' and Phosphorylates Polyserine Stretches Devoid of the Canonical Consensus. *FEBS J.* 285, 4674–4683. doi:10.1111/febs.14689

Du, S., Guan, S., Zhu, C., Guo, Q., Cao, J., Guan, G., et al. (2020). Secretory Pathway Kinase FAM20C, A Marker for Glioma Invasion and Malignancy, Predicts Poor Prognosis of Glioma. *Ott* 13, 11755–11768. doi:10.2147/OTT.S275452

Eckel-Passow, J. E., Drucker, K. L., Kollmeyer, T. M., Kosel, M. L., Decker, P. A., Molinaro, A. M., et al. (2020). Adult Diffuse Glioma GWAS by Molecular Subtype Identifies Variants in D2HGDH and FAM20C. *Neuro Oncol.* 22, 1602–1613. doi:10.1093/neuonc/noaa117

Fischer, E. H. (2013). Cellular Regulation by Protein Phosphorylation. *Biochem. Biophysical Res. Commun.* 430, 865–867. doi:10.1016/j.bbrc.2012.10.024

Georges, R. B., Adwan, H., Hamdi, H., Hielscher, T., Linnemann, U., and Berger, M. R. (2011). The Insulin-like Growth Factor Binding Proteins 3 and 7 Are Associated with Colorectal Cancer and Liver Metastasis. *Cancer Biol. Ther.* 12, 69–79. doi:10.4161/cbt.12.1.15719

Goodenberger, M. L., and Jenkins, R. B. (2012). Genetics of Adult Glioma. *Cancer Genet.* 205, 613–621. doi:10.1016/j.cancergen.2012.10.009

Jiang, H., Zeng, W., Ren, X., Cui, Y., Li, M., Yang, K., et al. (2019). Super-early Initiation of Temozolomide Prolongs the Survival of Glioblastoma Patients Without Gross-Total Resection: A Retrospective Cohort Study. *J. Neurooncol.* 144, 127–135. doi:10.1007/s11060-019-03211-1

Jiang, W., Xiang, C., Cazacu, S., Brodie, C., and Mikkelsen, T. (2008). Insulin-like Growth Factor Binding Protein 7 Mediates Glioma Cell Growth and Migration. *Neoplasia* 10, 1335–1342. doi:10.1593/neo.08694

Kang, J. U. (2013). Characterization of Amplification Patterns and Target Genes on the Short Arm of Chromosome 7 in Early-Stage Lung Adenocarcinoma. *Mol. Med. Rep.* 8, 1373–1378. doi:10.3892/mmr.2013.1686

Klement, E., and Medzihradszky, K. F. (2017). Extracellular Protein Phosphorylation, The Neglected Side of the Modification. *Mol. Cell Proteomics* 16, 1–7. doi:10.1074/mcp.O116.064188

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary. *Acta Neuropathol.* 131, 803–820. doi:10.1007/s00401-016-1545-1

Mair, M. J., Geurts, M., van den Bent, M. J., and Berghoff, A. S. (2021). A Basic Review on Systemic Treatment Options in WHO Grade II-III Gliomas. *Cancer Treat. Rev.* 92, 102124. doi:10.1016/j.ctrv.2020.102124

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. *Science* 298, 1912–1934. doi:10.1126/science.1075762

Nalbant, D., Youn, H., Nalbant, S. I., Sharma, S., Cobos, E., Beale, E. G., et al. (2005). FAM20: An Evolutionarily Conserved Family of Secreted Proteins Expressed in Hematopoietic Cells. *BMC Genomics* 6, 11. doi:10.1186/1471-2164-6-11

Ostrom, Q. T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., and Barnholtz-Sloan, J. S. (2018). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. *Neuro Oncol.* 20, iv1–iv86. doi:10.1093/neuonc/noy131

Qin, Z., Wang, P., Li, X., Zhang, S., Tian, M., Dai, Y., et al. (2016). Systematic Network-Based Discovery of a Fam20C Inhibitor (FL-1607) with Apoptosis Modulation in Triple-Negative Breast Cancer. *Mol. Biosyst.* 12, 2108–2118. doi:10.1039/c6mb00111d

Rangaswami, H., Bulbule, A., and Kundu, G. C. (2006). Osteopontin: Role in Cell Signaling and Cancer Progression. *Trends Cel Biol.* 16, 79–87. doi:10.1016/j.tcb.2005.12.005

Sreelatha, A., Kinch, L. N., and Tagliabracci, V. S. (2015). The Secretory Pathway Kinases. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1854, 1687–1693. doi:10.1016/j.bbapap.2015.03.015

Stupp, R., Mason, W. P., van den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J. B., et al. (2005). Radiotherapy Plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.* 352, 987–996. doi:10.1056/NEJMoa043330

Tagliabracci, V. S., Engel, J. L., Wiley, S. E., Xiao, J., Gonzalez, D. J., Nidumanda Appaiah, H., et al. (2014). Dynamic Regulation of FGF23 by Fam20C Phosphorylation, GalNAc-T3 Glycosylation, and Furin Proteolysis. *Proc. Natl. Acad. Sci.* 111, 5520–5525. doi:10.1073/pnas.1402218111

Tagliabracci, V. S., Wiley, S. E., Guo, X., Kinch, L. N., Durrant, E., Wen, J., et al. (2015). A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* 161, 1619–1632. doi:10.1016/j.cell.2015.05.028

van den Bent, M. J. (2014). Practice Changing Mature Results of RTOG Study 9802: Another Positive PCV Trial Makes Adjuvant Chemotherapy Part of Standard of Care in Low-Grade Glioma. *Neuro-Oncology* 16, 1570–1574. doi:10.1093/neuonc/nou297

Wang, S.-K., Samann, A. C., Hu, J. C.-C., and Simmer, J. P. (2013). FAM20C Functions Intracellularly Within Both Ameloblasts and Odontoblasts In Vivo. *J. Bone Miner Res.* 28, 2508–2511. doi:10.1002/jbmr.1990

Wesseling, P., and Capper, D. (2018). WHO 2016 Classification of Gliomas. *Neuropathol. Appl. Neurobiol.* 44, 139–150. doi:10.1111/nan.12432

Xu, W., Li, T., Gao, L., Zheng, J., Shao, A., and Zhang, J. (2017). Efficacy and Safety of Long-Term Therapy for High-Grade Glioma with Temozolomide: A Meta-Analysis. *Oncotarget* 8, 51758–51765. doi:10.18632/oncotarget.17401

Zhang, H., Zhu, Q., Cui, J., Wang, Y., Chen, M. J., Guo, X., et al. (2018). Structure and Evolution of the Fam20 Kinases. *Nat. Commun.* 9, 1218. doi:10.1038/s41467-018-03615-z

Zhang, N., Li, F., Gao, J., Zhang, S., and Wang, Q. (2020). Osteopontin Accelerates the Development and Metastasis of Bladder Cancer via Activating JAK1/STAT1 Pathway. *Genes Genom* 42, 467–475. doi:10.1007/s13258-019-00907-6

Zhao, R., Fu, L., Yuan, Z., Liu, Y., Zhang, K., Chen, Y., et al. (2021a). Discovery of a Novel Small-Molecule Inhibitor of Fam20C that Induces Apoptosis and Inhibits Migration in Triple Negative Breast Cancer. *Eur. J. Med. Chem.* 210, 113088. doi:10.1016/j.ejmech.2020.113088

Zhao, Z., Zhang, K.-N., Wang, Q., Li, G., Zeng, F., Zhang, Y., et al. (2021b). Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients. *Genomics, Proteomics & Bioinformatics* 19, 1–12. doi:10.1016/j.gpb.2020.10.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/fgene.2021.773725

Multigenomics Reveals the Causal Effect of Herpes Simplex Virus in Alzheimer's Disease: A Two-Sample Mendelian Randomization Study

Yuwei Zhang[1,2], *Jiaojiao Qu*[3], *Li Luo*[1,2], *Zhongshun Xu*[1,2] *and Xiao Zou*[1,2]*

[1]College of Life Sciences, Institute of Fungal Resources, Guizhou University, Guiyang, China, [2]Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Institute of Agro-Bioengineering, Guizhou University, Guiyang, China, [3]College of Tea Sciences, Guizhou University, Guiyang, China

In recent years, the herpes virus infectious hypothesis for Alzheimer's disease (AD) has gained support from an increasing number of researchers. Herpes simplex virus (HSV) is a potential risk factor associated with AD. This study assessed whether HSV has a causal relationship with AD using a two-sample Mendelian randomization analysis model. Six single-nucleotide polymorphisms (SNPs) associated with HSV-1 and thirteen SNPs associated with HSV-2 were used as instrumental variables in the MR analysis. We estimated MR values of relevance between exposure and the risk of AD using inverse-variance weighted (IVW) method, MR-Egger regression (Egger), and weighted median estimator (WME). To make the conclusion more robust and reliable, sensitivity analyses and RadialMR were performed to evaluate the pleiotropy and heterogeneity. We found that anti-HSV-1 IgG measurements were not associated with risk of AD (OR, 0.96; 95% CI, 0.79–1.18; $p = 0.736$), and the same was true for HSV-2 (OR, 1.03; 95% CI, 0.94–1.12; $p = 0.533$). The findings indicated that any HSV infection does not appear to be a genetically valid target of intervention in AD.

Keywords: HSV, genome, Mendelian randomization, Alzheimer's disease, causality

OPEN ACCESS

Edited by:
Guiyou Liu,
Tianjin Institute of Industrial Biotechnology (CAS), China

Reviewed by:
Sezen Vatansever,
Icahn School of Medicine at Mount Sinai, United States
Yang Zhang,
Hebei Medical University, China

***Correspondence:**
Xiao Zou
xzou@gzu.edu.cn

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: *10 September 2021*
Accepted: *25 November 2021*
Published: *05 January 2022*

Citation:
Zhang Y, Qu J, Luo L, Xu Z and Zou X (2022) Multigenomics Reveals the Causal Effect of Herpes Simplex Virus in Alzheimer's Disease: A Two-Sample Mendelian Randomization Study. Front. Genet. 12:773725. doi: 10.3389/fgene.2021.773725

1 INTRODUCTION

Alzheimer's disease (AD) is a complex chronic progressive degenerative disorder of the central nervous system, affecting primarily the elderly, which severely reduces the quality of life (Calabrò et al., 2021). According to the 2015 World Alzheimer Report, the number of AD patients is expected to double every 20 years, reaching up to 131.5 million by 2050 (Prince et al., 2015; Du et al., 2018) with the incidence rate of AD increasing exponentially after 65 years of age (Hou et al., 2019). AD is diagnosed after age 65 as late-onset AD (LOAD) and before age 65 as early-onset AD (EOAD). LOAD accounts for about 95% of AD cases. EOAD is essentially an inherited disease, with a 92%–100% heritability. In contrast, there are multiple factors influencing LOAD, which are sporadic (Laval and Enquist, 2021). AD has two central pathological features: the extracellular deposition of

Abbreviations: AD, Alzheimer's disease; APOE-ε4, apolipoprotein e gene; CI, confidence interval; GWAS, genome-wide association studies; HSV, herpes simplex virus; IGAP, International Genomics of Alzheimer's Project; IV, instrumental variables; IVW, inverse-variance weighted; LD, linkage disequilibrium; MR, Mendelian randomization; OR, odds ratio; WME, weighted median estimator.

amyloid plaques and intracellular accumulation of neurofibrillary tangles (NFTs). Amyloid plaques are mainly composed of amyloid-β (Aβ) protein and NFTs are composed of hyperphosphorylated tau proteins. Hence, there have been contrasting theories proposed about the underlying pathogenesis of AD, such as amyloid cascade hypothesis, Tau protein hypothesis, and oxidative stress. Nonetheless, to date, current therapies have failed to delay disease progression. In recent years, the herpes virus infection hypothesis has received a renewed interest by scientists who believe that infection is the main cause of AD.

In the 1980s, herpes simplex virus (HSV) was first proposed to be associated with AD after viral genetic material was discovered in the human brain as well as virus-induced lesions present in the limbic system were associated with AD (Ball, 1982). The viruses belong to the *Alphaherpesviridae* subfamily of the Herpesviridae family, including HSV-1 and HSV-2, which are ubiquitous human pathogens (Piret and Boivin, 2020). Previous studies (Wozniak et al., 2010) found that HSV-1 DNA was present in the brains of both AD patients and normal elderly people; however, in the brains of AD patients, HSV-1 DNA was found within 90% of the plaques and 72% of HSV-1 DNA was associated with plaques, while in the brains of normal elderly people, only 24% of HSV-1 DNA was associated with plaques. Thus, it was proposed that the HSV-1 infects infants and remains latent in the peripheral nervous system. Reactivation of latent HSV-1 infections may cause local neuronal damage and inflammation, which over time may lead to the deposition of Aβ and abnormal phosphorylation of tau in the brain. A recent study proposed that Aβ deposition and abnormal phosphorylation of tau were the brain's immune response to HSV-1 (Eimer et al., 2018). However, another recent study showed that AD associated β-amyloid does not protect against HSV-1 infection in the mouse brain (Bocharova et al., 2021).

To date, the precise molecular events, and biological pathways underlying the disease have yet to be identified and the existing evidence does not definitively support the herpesviruses hypothesis of AD. The deposition of Aβ and abnormal phosphorylation of tau are not necessarily the cause of AD, but may be the result of other risk factors leading to AD. Meanwhile, given the existence of unmeasured confounding variables and reverse causation, previous epidemiological studies have demonstrated a correlation but no direct causal relationship between HSV and AD, which allows for a re-evaluation of the theory as a possible strategy.

Multi-omics research probes the interaction between multiple factors in biological systems, including genomics, epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics. These factors jointly affect phenotypes and physiological traits. With the development of high-throughput sequencing technology, omics research continues to provide more extensive data. Through high-throughput sequencing, omics, and data integration studies, we can comprehensively and systematically understand the relationship between various factors in the fields of basic research, molecular biology, clinical diagnosis, and drug discovery. (Hasin et al., 2017).

Genomics is the earliest discipline stemming from histology, and focuses on the study of the entire genome, and is currently the most established discipline in the field. Genomics focuses on the identification of genetic variants associated with disease, treatment response, or patient prognosis (Hasin et al., 2017). With the successful development of next-generation sequencing (NGS) technology and the completion of the human genome project and the International Human Genome HapMap project (HapMap), genome-wide association studies (GWAS) have become a method for identifying millions of genetic variants related to complex diseases (GWAS catalog https://www.ebi.ac.uk/gwas/home) in different human populations. In such studies, millions of individuals are genotyped for many genetic markers, and the genotypes and phenotypes are subjected to statistical analysis at a population level. Significant differences in minor allele frequencies (MAF) between cases and controls are thought to be markers affecting the trait. GWAS studies provide an invaluable contribution to our understanding of complex phenotypes (Hasin et al., 2017).

Mendelian randomization (MR) is a strategy for evaluating the causality of risk factors of a disease using genetic variants from the GWAS as instrumental variables (IV) (Lawlor et al., 2008). It is based on the Mendelian inheritance law of "random allocation of parental alleles to offspring" in meiosis, which is equivalent to a randomized controlled trial using genotypes. MR analysis can remove the limitations of traditional epidemiology. As alleles were randomly allocated at conception, confounders cannot influence the result of the allocated alleles. Because the disease cannot alter genetic variants, reverse causation may be avoided.

IVs should satisfy three major hypotheses (**Figure 1**), which have been widely described in recent studies (Liu et al., 2018; Liu et al., 2021).

1) The IV is associated with the exposure ($\gamma \neq 0$, strong IVs).
2) The IV is not associated with the confounders ($\varphi = 0$).
3) The IV does not influence the outcome through some pathways other than the exposure ($\alpha = 0$, no directional pleiotropy).

Scepanovic et al. (2018) measured quantitative IgG responses to HSV-1 and HSV-2 infection in humoral immunity to explore the influence of genetic factors on the variability of humoral responses. After genome-wide genotyping of single-nucleotide polymorphisms (SNPs) and imputation, they examined associations between genetic variants and HSV-1 and HSV-2 IgG and performed two genome-wide association analyses. The International Genomics of Alzheimer's Project Consortium (IGAP) (Lambert et al., 2013) conducted a meta-analysis using genotyped and imputed data on four previously published GWAS datasets and obtained a novel genome-wide association analysis demonstrating the relationship of genetic variants with AD. In the present study, we used many SNPs of multi-genome association analysis as IVs to perform two-sample MR analysis (Gibran et al., 2018).

2 MATERIALS AND METHODS

2.1 Data Sources

The exposure risk factors considered in this study were HSV-1 and HSV-2. The genetic variations for both exposures were anti-

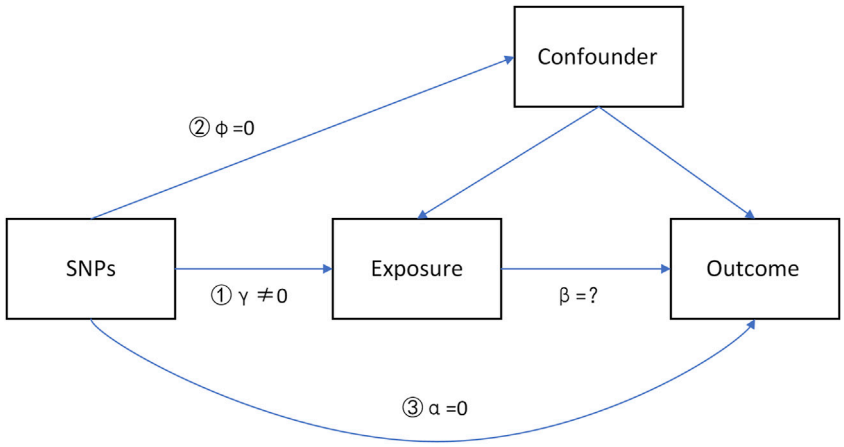

FIGURE 1 | Directed acyclic graph (DAG) model of instrumental variables in causal associations.

TABLE 1 | Description of consortium datasets for IGAP stage 1.

Consortium	*N* (cases/controls)	Percent women (cases/controls)	Case mean AAO (s.d.)	Control mean AAE (s.d.)
ADGC	10,273/10,892	59.4/58.6	74.7 (7.7)	76.3 (8.1)
CHARGE	1,315/12,968	63.6/57.8	82.7 (6.8)	72.8 (8.6)
EADI	2,243/6,017	64.9/60.7	68.5 (8.9)	74.0 (5.4)
GERAD	3,177/7,277	64.0/51.8	73.0 (8.5)	51.0 (11.8)
N	17,008/37,154			

AAO, age at onset; AAE, age at examination.

HSV-1 IgG measurement and anti-HSV-2 IgG measurement, which were downloaded from a GWAS study of Scepanovic et al. (2018), which was the basis of the summary data published in the NHGRI-EBI GWAS (https://www.ebi.ac.uk/gwas). The sample was derived from The French Milieu Interieur cohort, which was stratified by sex (500 men, 500 women) and age (200 individuals from each decade of life, between 20 and 70 years of age). The HSV-2 datasets contained 208 cases and 792 controls, and HSV-1 datasets contained 645 cases and 355 controls.

The summary data of AD derived from the International Genomics of Alzheimer's Project Consortium (IGAP), which was a sizeable two-stage research based on GWASs of AD in 74,046 diseased and normal individuals of European ancestry (Lambert et al., 2013). In stage 1, the IGAP performed a meta-analysis of four previously published GWAS datasets containing 17,008 AD patients and 37,154 controls, using genotyped and imputed data on 7,055,881 SNPs. The outcome data from IGAP stage 1 results were from the study of Kunkle et al. (2019). **Table 1** shows the detailed descriptions of IGAP stage 1 data.

2.2 Methods

All the analyses were performed using R version 4.1.0 software.

2.2.1 Selection of Instrumental Variables

The most critical step in MR design is to identify suitable genetic variants as IVs. First, we extracted SNPs that had significant ($p < 1 \times 10^{-5}$) associations with HSV-1 and HSV-2. Then, we performed a linkage disequilibrium (LD) analysis to exclude mutual linkage SNPs and to discard non-biallelic SNPs. LD ($r^2 < 0.001$, kb > 10,000) was applied to select IVs of HSV-1 and HSV-2. The samples used to estimate the LD effect derived from individuals of European ancestry from the 1,000 Genome Project. Correlated SNPs in LD were excluded using the "clump_data" function of the "TwoSampleMR" R package. As a result, 7 SNPs were identified for HSV-1 and 13 SNPs for HSV-2.

2.2.2 Harmonize

A summary set can generate errors if the effect alleles for the SNP effects in the exposure and outcome datasets are different. We aligned the effect alleles for exposure and outcome based on reported effect alleles and effect allele frequencies using the "harmonise_data" function of the "TwoSampleMR" R package (Gibran et al., 2018). Furthermore, we used *F*-statistics (Bowden et al., 2016) to measure the strength of the selected IVs. If the *F*-statistic was more than ten, genetic variants were generally deemed to be a strong IV.

2.2.3 Mendelian Randomization

We conducted the MR analysis using inverse-variance weighted (IVW) regression analysis, MR-Egger regression analysis, and weighted median estimator (WME). IVW can provide accurate estimates when the IV satisfies the MR assumptions that there are no invalid IVs (Burgess et al., 2013). The mean effect estimate of IVW is derived from a random effect IVW meta-analysis of the Wald ratios (SNP-outcome associations divided by SNP-exposure associations)

TABLE 2 | SNPs significantly (p-value < 1×10^{-5}) and independently ($r^2 < 0.001$) associated with herpes simplex virus type 1 (HSV-1) infection (SNPs = 6).

Exposure	SNP	Effect allele	Other allele	EAF	Beta	SE	p-value	F-statistic
HSV-1	rs10977313	T	G	0.11	−0.12	0.02	2.97×10^{-7}	26.9
	rs1446553	A	G	0.25	0.07	0.02	9.85×10^{-6}	19.9
	rs34018815	A	G	0.10	−0.11	0.02	8.35×10^{-6}	20.2
	rs58599785	T	C	0.16	0.08	0.02	4.91×10^{-6}	21.3
	rs78421079	T	C	0.08	−0.12	0.03	6.49×10^{-6}	20.7
	rs8020017	G	A	0.40	−0.06	0.01	9.85×10^{-6}	19.9

TABLE 3 | Association of six SNPs for HSV-1 infection with AD using MR with different methods.

Outcome	SNPs		OR	95% CI	p-value	Pleiotropy	Heterogeneity		I^2
						Intercept p-value	Cochran's Q-statistic	p-value	
AD	6	IVW	0.96	[0.79–1.18]	0.736		5.836	0.323	
		WME	0.97	[0.77–1.23]	0.833				
		MR-Egger	0.79	[0.32–1.99]	0.653	0.694	5.586	0.232	0.958

estimated for each IV (Staley and Burgess, 2017). MR-Egger regression is robust for invalid instruments, and can be used to test for directional pleiotropy, providing an estimate of the causal effect adjusted for a variable'presence. In MR-Egger, an intercept that differs from zero estimates the average pleiotropy effect across the genetic variants, which indicates that the IVW estimate is biased (Bowden et al., 2015). However, MR-Egger regression is more easily influenced by regression dilution, so that it should be approximated using the I^2 statistic. If I^2 is high ($I^2 > 0.9$), Egger regression can be considered an unbiased estimation (Bowden et al., 2016). The WME provides a consistent, valid estimate if at least half of the IVs are valid (Verbanck et al., 2018). MR analyses were performed using the R-based package "TwoSampleMR".

2.2.4 Sensitivity Analysis

The three methods described above were applied to analyze causal estimation, and we performed the following additional analyses and assessments to examine the robustness of the results. First, we used Egger intercept to test the pleiotropy of SNPs (Burgess and Thompson, 2017). Then, we calculated the heterogeneity among SNPs using Cochran's Q-statistic to assess the robustness of IVs (Kippersluis and Rietveld, 2017). Furthermore, to evaluate whether the MR estimate was driven or biased by a single SNP that might have an enormous pleiotropic effect, RadialMR was applied to present a more straightforward detection of outliers and to correct horizontal pleiotropy by removing outliers (Bowden et al., 2018). All sensitivity analyses were performed using the R-based package "TwoSampleMR" and "RadialMR".

3 RESULTS AND DISCUSSION

3.1 The Causality of HSV-1 and AD

After removing the palindrome SNP (rs1738233), six SNPs for HSV-1 infection were identified, which were significant ($p < 1 \times 10^{-5}$) and independent ($r^2 < 0.001$). The F-statistics for the six SNPs were all more than 10, which indicated that all six IVs were strong instruments (**Table 2**).

Table 3; **Figure 2** showed the estimated associations of HSV-1 risk factor with AD from MR analysis. Genetically predicted HSV-1 infection was not associated with AD risk using IVW (OR = 0.96, p = 0.736), WME (OR = 0.97, p = 0.833), and MR-Egger (OR = 0.79, p = 0.653). The MR-Egger intercept indicated no directional pleiotropy (intercept = 0.018, p = 0.694), suggesting that horizontal pleiotropy was unlikely to influence the IVW estimate. The I^2 statistics was 0.958, indicating that relative bias did not materially affect the standard MR-Egger analysis. Cochran's Q test showed no existence of heterogeneity of SNPs (Cochran's Q-statistic = 5.83, p = 0.322), while RadialMR showed that there were no outliers in the six SNPs.

3.2 The Causality of HSV-2 and AD

Thirteen SNPs for HSV-2 infection were identified, which were both significant ($p < 1 \times 10^{-5}$) and independent ($r^2 < 0.001$) (**Table 4**). The F-statistics for the thirteen SNPs were more than 10, which indicated that they were strong IVs.

Table 5; **Figure 3** showed the estimated associations of HSV-2 risk factors with AD from the MR analysis. Genetically predicted HSV-2 infection was not associated with the AD using IVW (OR

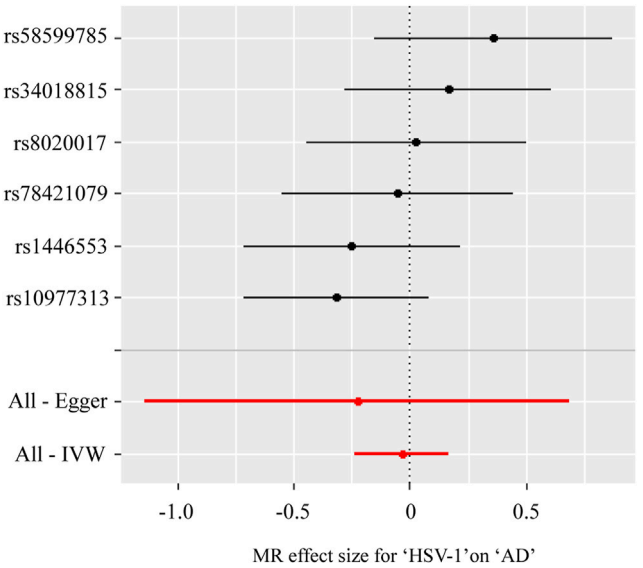

FIGURE 2 | Forest plot for two-sample Mendelian randomization effect size between Herpes simplex virus type 1 (HSV-1) and Alzheimer's disease (AD).

= 1.03, $p = 0.533$), WME (OR = 1.08, $p = 0.121$), and MR-Egger (OR = 0.95, $p = 0.764$). The MR-Egger intercept indicated that there was no directional pleiotropy (intercept = 0.017, $p = 0.646$). Furthermore, the Cochran's Q-statistic indicated the existence of heterogeneity of SNPs (Cochran's Q-statistic = 18.8, $p = 0.04$). Meanwhile, no outliers were detected using RadialMR.

4 DISCUSSION

We found that both HSV-1 and HSV-2 were not causally associated with an increased risk of AD using genetic variation as instrumental variables. Kwok and Schooling (2021) used the GWAS summary statistics data from the French Milieu Interieur cohort, the United Kingdom biobank, and the US 23 and Me Study, pointing out that HSV-1 and HSV-2 were not associated with AD. SY et al. (2021) used the GWAS summary statistics data from the 23 and Me cohort, indicating the same result.

4.1 The Result of HSV and AD

Although the causality of the association is unclear, many studies have proven that HSV is not unrelated to AD. HSV-1 virus was detected in the brains of both AD patients and elderly normal people. However, most of the AD patients were APOE-ε4 gene carriers. The herpesvirus hypothesis proposes that HSV-1 enters the brains of APOE-ε4 carriers, where it remains a latent life with limited transcription and low protein synthesis. In response to immunosuppression, peripheral infection, and inflammation, HSV-1 reactivates, creating a combination of viral action and inflammatory effects that are poorly repaired by APOE-ε4 carriers, ultimately leading to the development of AD (Itzhaki, 2018). In addition, a recent study pointed out novel molecular mechanisms through which recurrent HSV-1 infection may affect neuronal aging, likely contributing to neurodegeneration (Napoletani et al., 2021).

We inferred that our results may have occurred mainly due to several reasons. The major reason is that reactivation after latent HSV-1 infection may be responsible for a pathogenetic mechanism of AD, and IgM is a marker of activation of primary infection. Our study used anti-HSV-1 IgG antibodies rather than IgM as a proxy for HSV-1 infection, implicitly demonstrating that previous HSV-1 infection is not associated with AD risk. Another reason is the speculation that HSV-1 infection is not a risk factor for cognitive decline but rather a phenomenon that co-occurs with neuro-inflammation or as a result of neuro-inflammation.

Meanwhile, we found that HSV-2 was not causally associated with an increased risk of AD using genetic variation as an instrumental variable. This is probably because that according to the available epidemiological observations, HSV-2 mainly invades the genitalia and the area from the waist down and is not associated with the brain.

Future studies should perform MR analyses using anti-HSV-1 IgM antibodies as an IV for HSV-1 infection. What we can conclude, however, is that AD is not simply a single factor disease caused by HSV, but that it encompasses complex disease mechanisms.

TABLE 4 | SNPs significantly (p-value < 1×10^{-5}) and independently ($r^2 < 0.001$) associated with herpes simplex virus type 2 (HSV-2) infection (SNPs = 13).

Exposure	SNP	Effect allele	Other allele	EAF	Beta	SE	p-value	F-statistic
HSV-2	rs10888851	G	C	0.11	−0.24	0.05	3.27×10^{-6}	23.2
	rs10782620	G	T	0.39	0.16	0.03	2.60×10^{-6}	23.7
	rs12042287	C	T	0.27	−0.16	0.03	7.67×10^{-6}	21.3
	rs10174926	C	T	0.13	−0.24	0.04	9.72×10^{-7}	25.9
	rs72804080	G	A	0.13	0.26	0.05	1.92×10^{-7}	29.5
	rs355547	C	T	0.39	0.17	0.04	2.00×10^{-6}	24.3
	rs35213774	G	A	0.11	0.26	0.05	1.10×10^{-6}	25.6
	rs6826994	A	G	0.06	−0.30	0.06	6.76×10^{-6}	21.6
	rs113043839	A	G	0.06	0.32	0.07	6.94×10^{-6}	21.5
	rs10100854	T	C	0.37	−0.15	0.03	7.00×10^{-6}	21.5
	rs10964023	T	G	0.19	−0.19	0.04	3.58×10^{-6}	23.0
	rs17802723	G	C	0.08	−0.28	0.06	7.90×10^{-6}	21.3
	rs10790877	A	G	0.47	−0.16	0.03	7.82×10^{-7}	26.4

TABLE 5 | Association of thirteen SNPs for HSV-2 infection with AD using MR with different methods.

Outcome	SNPs		OR	95% CI	*p*-value	Pleiotropy	Heterogeneity		I^2
						Intercept *p*-value	Cochran's *Q*-statistic	*p*-value	
AD	13	IVW	1.03	[0.94–1.12]	0.533		18.802	0.043	
		WME	1.08	[0.98–1.18]	0.121				
		MR-Egger	0.95	[0.67–1.34]	0.764	0.017	18.343	0.031	0.959

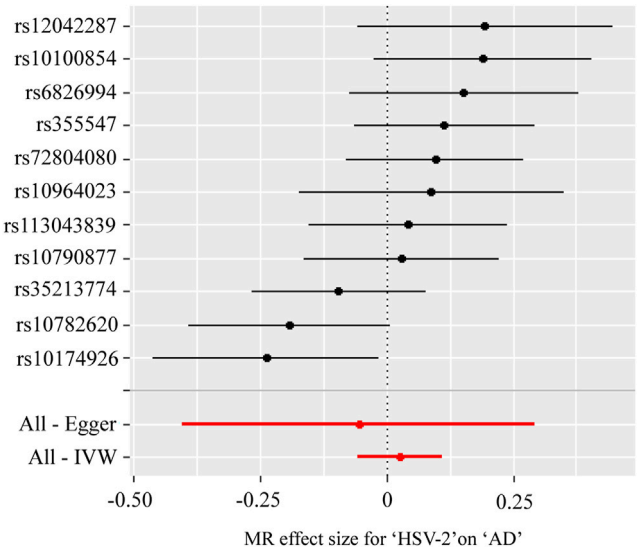

FIGURE 3 | Forest plot for two-sample Mendelian randomization effect size between Herpes simplex virus type 2 (HSV-2) and Alzheimer's disease (AD).

4.2 Advantages and Challenges of MR Analysis

In the investigation of risk factors for AD, traditional research methods present many challenges in discovering the cause of the disease. Observational studies can only demonstrate a correlation rather than causality between exposure and outcome due to confounding factors and reverse causality. Cohort studies can make causal arguments but waste time. Random control trials (RCT) are considered the gold standard for clinical diagnosis and have a solid causal view. However, when applied by researchers, they are difficult to practice due to medical ethics and the many limitations of the design process. For these reasons, MR analysis has become a more convenient and effective way of exploring the causal links between risk factors and AD.

The application of MR analysis in this study has several advantages. First, reverse causality can be avoided, and second, it can prevent the interference of confounding factors. MR analysis can also address situations where an intervention experiment cannot be performed because of ethical restrictions (Zheng et al., 2017). Our exposure data were obtained from a publicly available GWAS database published with credibility. Our outcome data derived from a study conducted by the IGAP with a large sample population.

Nonetheless, our study also has some limitations. First, our data samples were based on individuals of European ancestry, so the results are not representative of all races. Second, the sample size of the exposure data was not sufficiently large, leading to the low power of statistics and false negatives. However, a significant number of IVs can lead to high power but inevitable heterogeneity and pleiotropy of IVs. This is where the general challenge of MR.

5 CONCLUSION

We implemented a two-sample MR to demonstrate the causal relationship between HSV infection and AD risk. The SNPs were independent and strong instrumental variables, and the result was robust and reliable. Our findings indicated the negative association between any HSV IgG and AD. Further research is needed to investigate whether HSV IgM is corelated with AD, and whether HSV infections that co-occur with neuro-inflammation are more relevant.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The AD data can be found here: https://www.niagads.org/datasets/ng00075. The HSV data can be found here: https://www.ebi.ac.uk/gwas/efotraits/EFO_0009349 and https://www.ebi.ac.uk/gwas/efotraits/EFO_0009350.

AUTHOR CONTRIBUTIONS

XZ contributed to the conception and designed the study. LL and ZX organized the database. YZ performed the statistical analysis and wrote the article. JQ supervised the project and acquired the funding. XZ reviewed the article. All the authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (31860037) and Science and Technology Program of Guizhou Province (qiankehe supported (2019) 2405).

REFERENCES

Ball, M. J. (1982). "Limbic Predilection in Alzheimer Dementia: Is Reactivated Herpesvirus Involved?". *Can. J. Neurol. Sci.* 9, 303–306. doi:10.1017/s0317167100044115

Bocharova, O., Pandit, N. P., Molesworth, K., Fisher, A., Mychko, O., Makarava, N., et al. (2021). Alzheimer's Disease-Associated β-amyloid Does Not Protect against Herpes Simplex Virus 1 Infection in the Mouse Brain. *J. Biol. Chem.* 297, 100845. doi:10.1016/j.jbc.2021.100845

Bowden, J., Spiller, W., Del Greco M, F., Sheehan, N., Thompson, J., Minelli, C., et al. (2018). Improving the Visualization, Interpretation and Analysis of Two-Sample Summary Data Mendelian Randomization via the Radial Plot and Radial Regression. *Int. J. Epidemiol.* 47, 2100. doi:10.1093/ije/dyy265

Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection through Egger Regression. *Int. J. Epidemiol.* 44, 512–525. doi:10.1093/ije/dyv080

Bowden, J., Del Greco M., F., Minelli, C., Davey Smith, G., Sheehan, N. A., and Thompson, J. R. (2016). Assessing the Suitability of Summary Data for Two-Sample Mendelian Randomization Analyses Using MR-Egger Regression: The Role of the I2 Statistic. *Int. J. Epidemiol.* 45, dyw220–1974. doi:10.1093/ije/dyw220

Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian Randomization Analysis with Multiple Genetic Variants Using Summarized Data. *Genet. Epidemiol.* 37, 658–665. doi:10.1002/gepi.21758

Burgess, S., and Thompson, S. G. (2017). Interpreting Findings from Mendelian Randomization Using the MR-Egger Method. *Eur. J. Epidemiol.* 32, 377–389. doi:10.1007/s10654-017-0255-x

Calabrò, M., Rinaldi, C., Rinaldi, C., Santoro, G., and Crisafulli, C. (2021). The Biological Pathways of Alzheimer Disease: A Review. *AIMS Neurosci.* 8, 86–132. doi:10.3934/Neuroscience.2021005

Du, X., Wang, X., and Geng, M. (2018). Alzheimer's Disease Hypothesis and Related Therapies. *Transl Neurodegener* 7, 2. doi:10.1186/s40035-018-0107-y

Eimer, W. A., Vijaya Kumar, D. K., Navalpur Shanmugam, N. K., Rodriguez, A. S., Mitchell, T., Washicosky, K. J., et al. (2018). Alzheimer's Disease-Associated β-Amyloid Is Rapidly Seeded by Herpesviridae to Protect against Brain Infection. *Neuron* 100, 1527–1532. doi:10.1016/j.neuron.2018.11.043

Gibran, H., Zheng, J., Benjamin, E., Wade, K. H., Valeriia, H., Baird, D., et al. (2018). The MR-Base Platform Supports Systematic Causal Inference across the Human Phenome. *ELife* 7, e34408.

Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics Approaches to Disease. *Genome Biol.* 18, 83. doi:10.1186/s13059-017-1215-1

Hou, Y., Dan, X., Babbar, M., Wei, Y., Hasselbalch, S. G., Croteau, D. L., et al. (2019). Ageing as a Risk Factor for Neurodegenerative Disease. *Nat. Rev. Neurol.* 15, 565–581. doi:10.1038/s41582-019-0244-7

Huang, S.-Y., Yang, Y.-X., Kuo, K., Li, H.-Q., Shen, X.-N., Chen, S.-D., et al. (2021). Herpesvirus Infections and Alzheimer's Disease: A Mendelian Randomization Study. *Alz Res. Ther.* 13, 158. doi:10.1186/s13195-021-00905-5

Itzhaki, R. F. (2018). Corroboration of a Major Role for Herpes Simplex Virus Type 1 in Alzheimer's Disease. *Front. Aging Neurosci.* 10, 324. doi:10.3389/fnagi.2018.00324

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., et al. (2019). Author Correction: Genetic Meta-Analysis of Diagnosed Alzheimer's Disease Identifies New Risk Loci and Implicates Aβ, Tau, Immunity and Lipid Processing. *Nat. Genet.* 51, 1423–1424. doi:10.1038/s41588-019-0358-210.1038/s41588-019-0495-7

Kwok, M. K., and Schooling, C. M. (2021). Herpes Simplex Virus and Alzheimer's Disease: A Mendelian Randomization Study. *Neurobiol. Aging* 99, e11–101. doi:10.1016/j.neurobiolaging.2020.09.025

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, 1452–1458. doi:10.1038/ng.2802

Laval, K., and Enquist, L. W. (2021). The Potential Role of Herpes Simplex Virus Type 1 and Neuroinflammation in the Pathogenesis of Alzheimer's Disease. *Front. Neurol.* 12, 458. doi:10.3389/fneur.2021.658695

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., and Davey Smith, G. (2008). Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology. *Statist. Med.* 27, 1133–1163. doi:10.1002/sim.3034

Liu, G., Zhoa, Y., Jin, S., Hu, Y., Wang, T., Tian, R., et al. (2018). Circulating Vitamin E Levels and Alzheimer's Disease: A Mendelian Randomization Study. *Neurobiol. Aging* 72, 181–189. doi:10.1016/j.neurobiolaging.2018.08.008

Liu, H., Zhang, Y., Zhang, H., Wang, L., Wang, T., Han, Z., et al. (2021). Effect of Plasma Vitamin C Levels on Parkinson's Disease and Age at Onset: A Mendelian Randomization Study. *J. Transl. Med.* 19, 221. doi:10.1186/s12967-021-02892-5

Napoletani, G., Protto, V., Marcocci, M. E., Nencioni, L., Palamara, A. T., and De Chiara, G. (2021). Recurrent Herpes Simplex Virus Type 1 (HSV-1) Infection Modulates Neuronal Aging marks in *In Vitro* and *In Vivo* Models. *Ijms* 22, 6279. doi:10.3390/ijms22126279

Piret, J., and Boivin, G. (2020). Immunomodulatory Strategies in Herpes Simplex Virus Encephalitis. *Clin. Microbiol. Rev.* 33. doi:10.1128/CMR.00105-19

Prince, M., Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., and Prima, M. (2015). World Alzheimer Report 2015:the Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends. London, United Kingdom: Alzheimer's Disease International, 84. Available at: https://kclpure.kcl.ac.uk/portal/en/publications/world-alzheimer-report-2015-the-global-impact-of-dementia(ae525fda-1938-4892-8daa-a2222a672254)/export.html.

Scepanovic, P., Alanio, C., Alanio, C., Hammer, C., Hodel, F., Bergstedt, J., et al. (2018). Human Genetic Variants and Age Are the Strongest Predictors of Humoral Immune Responses to Common Pathogens and Vaccines. *Genome Med.* 10, 59. doi:10.1186/s13073-018-0568-8

Staley, J. R., and Burgess, S. (2017). Semiparametric Methods for Estimation of a Nonlinear Exposure-Outcome Relationship Using Instrumental Variables with Application to Mendelian Randomization. *Genet. Epidemiol.* 41, 341–352. doi:10.1002/gepi.22041

van Kippersluis, H., and Rietveld, C. A. (2017). Pleiotropy-robust Mendelian Randomization. *Int. J. Epidemiol.* 47, 1279–1288. doi:10.1093/ije/dyx002

Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization between Complex Traits and Diseases. *Nat. Genet.* 50, 693–698. doi:10.1038/s41588-018-0099-7

Wozniak, M. A., Mee, A. P., and Itzhaki, R. F. (2010). Herpes Simplex Virus Type 1 DNA Is Located within Alzheimer's Disease Amyloid Plaques. *J. Pathol.* 217, 131–138. doi:10.1002/path.2449

Zheng, J., Baird, D., Borges, M.-C., Bowden, J., Hemani, G., Haycock, P., et al. (2017). Recent Developments in Mendelian Randomization Studies. *Curr. Epidemiol. Rep.* 4, 330–345. doi:10.1007/s40471-017-0128-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/fnins.2022.823741

Identification of Ferroptosis-Related Genes in Alzheimer's Disease Based on Bioinformatic Analysis

Ying Wang†, Guohua Chen and Wei Shao†*

Department of Neurology, Wuhan Hospital of Traditional Chinese and Western Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Introduction: Alzheimer's disease (AD) is the most prevalent cause of dementia, and emerging evidence suggests that ferroptosis is involved in the pathological process of AD.

Materials and Methods: Three microarray datasets (GSE122063, GSE37263, and GSE140829) about AD were collected from the GEO database. AD-related module genes were identified through a weighted gene co-expression network analysis (WGCNA). The ferroptosis-related genes were extracted from FerrDb. The apoptosis-related genes were downloaded from UniProt as a control to show the specificity of ferroptosis. The overlap was performed to obtain the module genes associated with ferroptosis and apoptosis. Then the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses and the protein-protein interaction (PPI) were conducted. Cytoscape with CytoHubba was used to identify the hub genes, and the Logistic regression was performed to distinguish the AD patients from controls.

Results: 53 ferroptosis-related module genes were obtained. The GO analysis revealed that response to oxidative stress and starvation, and multicellular organismal homeostasis were the most highly enriched terms. The KEGG analysis showed that these overlapped genes were enriched not only in renal cell carcinoma pathways and central carbon metabolism in cancer, but also in autophagy-related pathways and ferroptosis. Ferroptosis-related hub genes in AD (JUN, SLC2A1, TFRC, ALB, and NFE2L2) were finally identified, which could distinguish AD patients from controls ($P < 0.05$). The area under the ROC curve (AUC) was 0.643. Apoptosis-related hub genes in AD (STAT1, MCL1, and BCL2L11) were also identified and also could distinguish AD patients from controls ($P < 0.05$). The AUC was 0.608, which was less than the former AUC value, suggesting that ferroptosis was more special than apoptosis in AD.

Conclusion: We identified five hub genes (JUN, SLC2A1, TFRC, ALB, and NFE2L2) that are closely associated with ferroptosis in AD and can differentiate AD patients from controls. Three hub genes of apoptosis-related genes in AD (STAT1, MCL1, and BCL2L11) were also identified as a control to show the specificity of ferroptosis. JUN, SLC2A1, TFRC, ALB, and NFE2L2 are thus potential ferroptosis-related biomarkers for disease diagnosis and therapeutic monitoring.

Keywords: Alzheimer's disease (AD), WGCNA, ferroptosis, apoptosis, GEO

OPEN ACCESS

Edited by:
Guiyou Liu,
Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences (CAS), China

Reviewed by:
Zhihao Wu,
Southern Methodist University, United States
Anandhan Annadurai,
University of Arizona, United States

***Correspondence:**
Wei Shao
shaowei74@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: *28 November 2021*
Accepted: *06 January 2022*
Published: *07 February 2022*

Citation:
Wang Y, Chen G and Shao W (2022) Identification of Ferroptosis-Related Genes in Alzheimer's Disease Based on Bioinformatic Analysis. Front. Neurosci. 16:823741. doi: 10.3389/fnins.2022.823741

INTRODUCTION

Alzheimer's disease (AD) is the most prevalent cause of dementia, accounting for approximately 60–80% of all cases (Gbd 2016 Dementia Collaborators, 2019). The exact pathogenesis of AD is still not fully elucidated (Zhang et al., 2021). Ferroptosis is an iron-dependent lipid peroxidation-driven cell death, and emerging evidence suggests that it is involved in the pathological process of AD (Lane et al., 2018; Weiland et al., 2019). In addition, several characteristics of the pathogenesis of AD were consistent

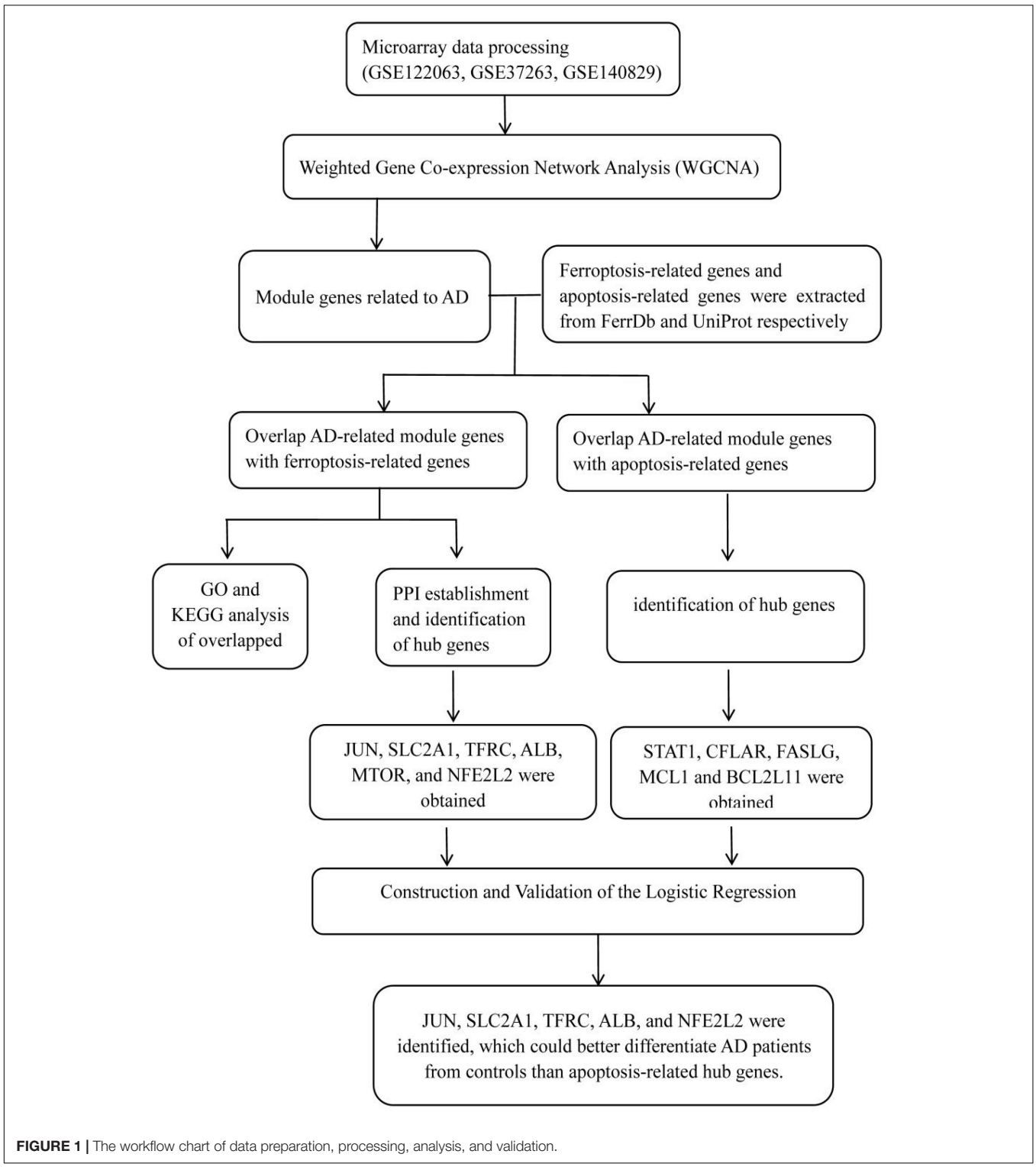

FIGURE 1 | The workflow chart of data preparation, processing, analysis, and validation.

with those of ferroptosis, such as excess iron accumulation, elevated lipid peroxides (Zhang et al., 2012; Hambright et al., 2017; Ayton et al., 2019). Therefore, ferroptosis is increasingly being recognized as a unique cell death mechanism participating in the pathogenesis of AD. However, more direct evidence is needed to be presented (Chen et al., 2021). Apoptosis is the

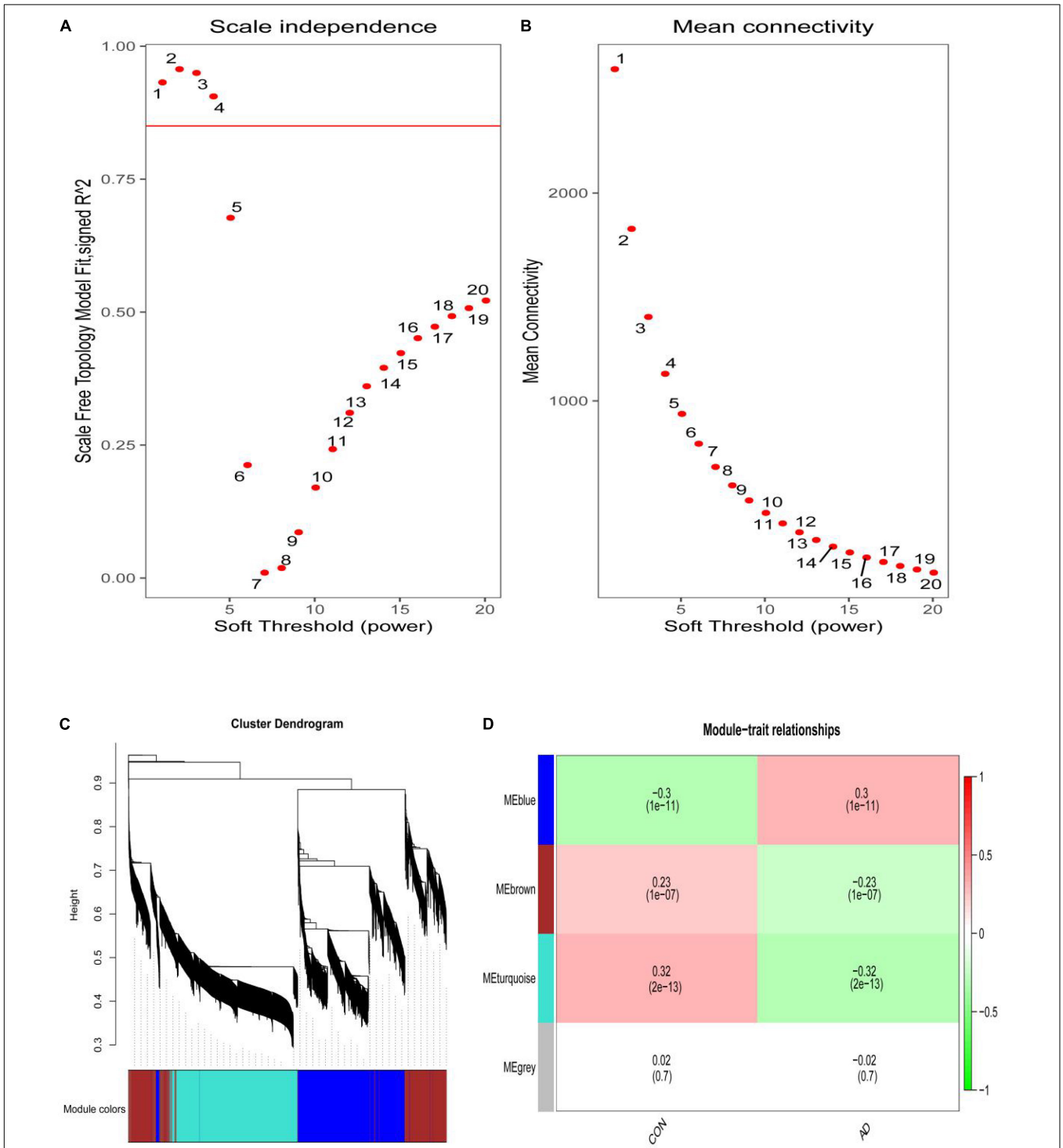

FIGURE 2 | (A) Analysis of the scale-free index for various soft-threshold powers (β). **(B)** Analysis of the mean connectivity for various soft-threshold powers. **(C)** Identification of co-expression gene modules. The branches of the dendrogram cluster into 4 modules and each one was labeled in a unique color. **(D)** A heatmap showing the correlation between each module eigengene and phenotype. Two modules were correlated with AD-namely, turquoise and blue modules.

spontaneous and orderly death of cells, which involves the activation, expression and regulation of a series of genes, and it is a biological process that plays an essential role in normal physiology (Obulesu and Lakshmi, 2014). It is now generally accepted that massive neuronal death due to apoptosis is a common characteristic in the brains of patients suffering from neurodegenerative diseases, and apoptotic cell death has been found in neurons and glial cells in AD (Shimohama, 2000; Sharma et al., 2021).

Current studies on ferroptosis and AD are mainly focused on two aspects: one is the mechanism of ferroptosis in the pathological process of AD, mainly discussing how ferroptosis participates in the AD (Masaldan et al., 2019; Jakaria et al., 2021); the second is the clinical efficacy study of ferroptosis inhibitors in AD, mainly to explore whether ferroptosis as a drug target of AD can effectively delay the progression of AD (Yan and Zhang, 2019; Plascencia-Villa and Perry, 2021; Vitalakumar et al., 2021). The purpose of this study is to investigate the association between ferroptosis-related genes and AD with the gene level, which is a supplement to existing studies and also a reference for ferroptosis as a therapeutic target for AD. These hub genes identified by this study could also serve as the ferroptosis-related biomarkers for disease diagnosis and therapeutic monitoring.

MATERIALS AND METHODS

Microarray Data Processing

Three microarray datasets (GSE122063, GSE37263, and GSE140829) of AD were collected from the GEO database[1]. GSE122063 was based on the platforms of the GPL16699 (Mckay et al., 2019); GSE37263 was based on the platforms of the GPL5175 (Tan et al., 2010); and GSE140829 was based on the platforms of the GPL15988. Data for 56 AD patients and 44 control samples from GSE122063, 8 AD patients and 8 control

[1]http://www.ncbi.nlm.nih.gov/geo

TABLE 1 | Details for FerrDb.

Data set	Category	Annotated from	Count	Annotations
Driver	Regulator	Gene	108	150
Suppressor	Regulator	Gene	69	109
Marker	Marker	Gene	111	123
Inducer	Regulator	Small molecule	35	54
Inhibitor	Regulator	Small molecule	41	46
Ferroptosis aggravates disease	Ferroptosis-diseaseassociation	Ferroptosis and disease	49	58
Ferroptosis alleviates disease	Ferroptosis-diseaseassociation	Ferroptosis and disease	46	77

The number of "Count" and "Annotations" is inconsistent, because one gene can have multiple annotations.

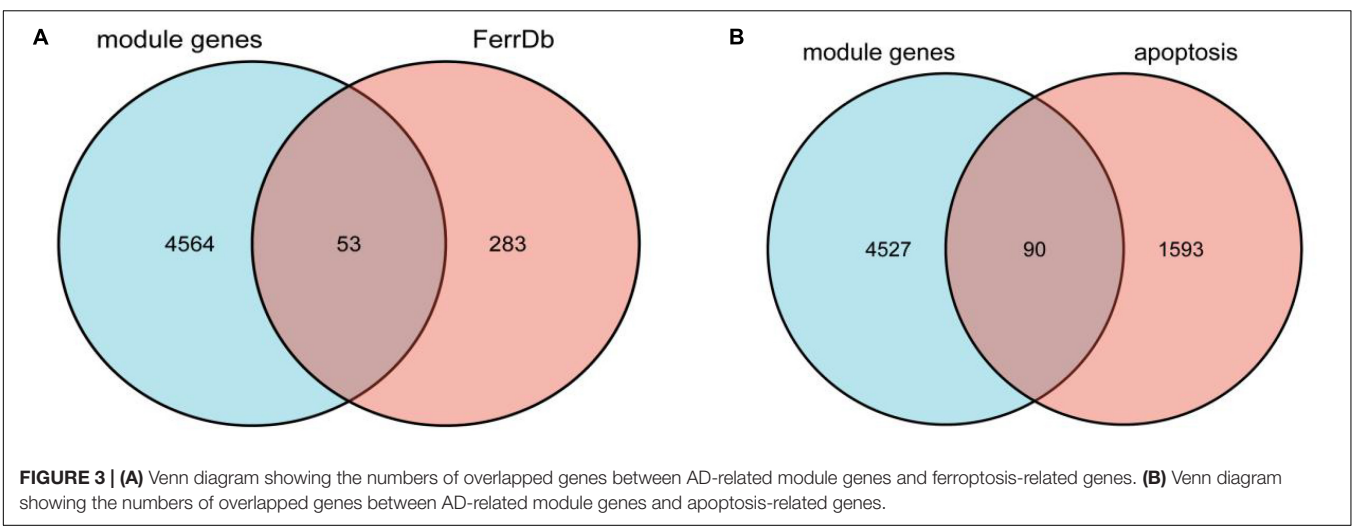

FIGURE 3 | (A) Venn diagram showing the numbers of overlapped genes between AD-related module genes and ferroptosis-related genes. **(B)** Venn diagram showing the numbers of overlapped genes between AD-related module genes and apoptosis-related genes.

TABLE 2 | Ferroptosis-related module genes obtained through the Venn diagram.

Type	Genes
Driver	PGD, YY1AP1, ATG3, ATG7, DPP4, NRAS, LPIN1, FBXW7, SCP2, EPAS1, TF, ATG16L1, IDH1, TFRC, BAP1, SNX4, PIK3CA, ATF3, PRKAA2
Suppressor	SQSTM1, SLC40A1, MTOR, FANCD2, MUC1, TP63, FTMT, PRDX6, NFE2L2, ACSL3, JUN, SLC7A11, FH, CISD2, SESN2, PROM2
Marker	TXNIP, HSD17B11, NCF2, PTGS2, ALB, STEAP3, SLC1A4, RRM2, CXCL2, ANGPTL7, PRDX1, SLC2A1, STMN1, RGS4, OXSR1, KLHL24, CAPG, DRD5

samples from GSE37263, and 182 AD patients and 207 control samples from GSE140829 were analyzed in our study. A flow diagram of the study is shown in **Figure 1**.

Weighted Gene Co-expression Network Analysis

Firstly, the expression profiles of three datasets were removed from the batch effect for further analysis. The gene co-expression network was constructed with an R package termed "weighted gene co-expression network analysis (WGCNA)" (Langfelder and Horvath, 2008, 2012). The Adjacency matrix was constructed by a weighted correlation coefficient. Subsequently, the adjacency matrix was transformed into a topological overlap matrix (TOM). Then, hierarchical clustering was performed to identify modules, and the eigengene was calculated. Finally, we assessed the correlation between phenotype (i.e., AD or control samples) and each module by Pearson's correlation analysis and identified AD-related modules. The genes in these modules were considered as AD-related module genes.

The Extraction of Ferroptosis-Related Genes From FerrDb and Apoptosis-Related Genes From UniProt

FerrDb[2] is an artificial ferroptosis database for the management and identification of ferroptosis-related markers and regulatory

[2]http://www.zhounan.org/ferrdb

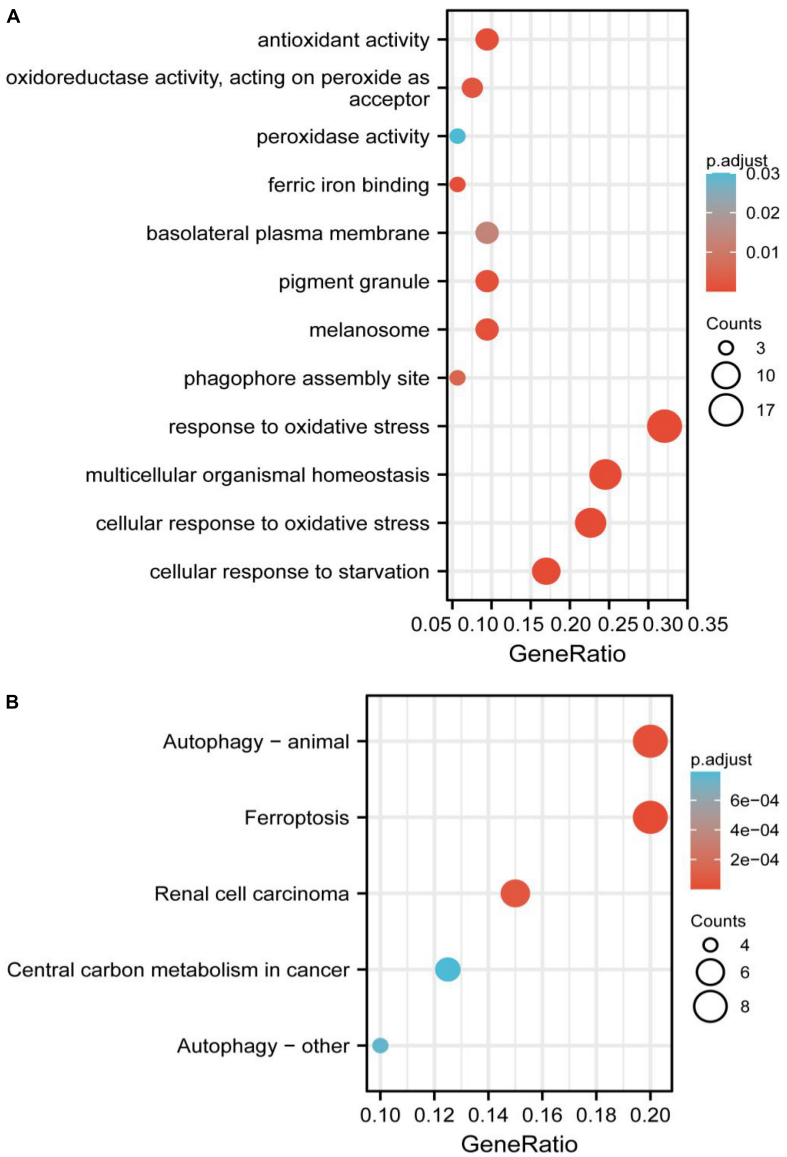

FIGURE 4 | (A) Gene Ontology (GO) functional analysis showing enrichment of ferroptosis-related module genes. **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of ferroptosis-related module genes.

factors, as well as ferroptosis-related diseases (Zhou and Bao, 2020). Therefore, ferroptosis-related genes were downloaded from this database for further analysis. The UniProt Knowledgebase is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation, and thus apoptosis-related genes were extracted from UniProt[3].

[3]https://www.uniprot.org/

Overlap Alzheimer's Disease-Related Module Genes With Ferroptosis-Related Genes and Apoptosis-Related Genes, Respectively

Ferroptosis-related genes were downloaded from FerrDb and apoptosis-related genes were downloaded from UniProt. We overlapped these genes with AD-related module genes derived from WGCNA, respectively. The

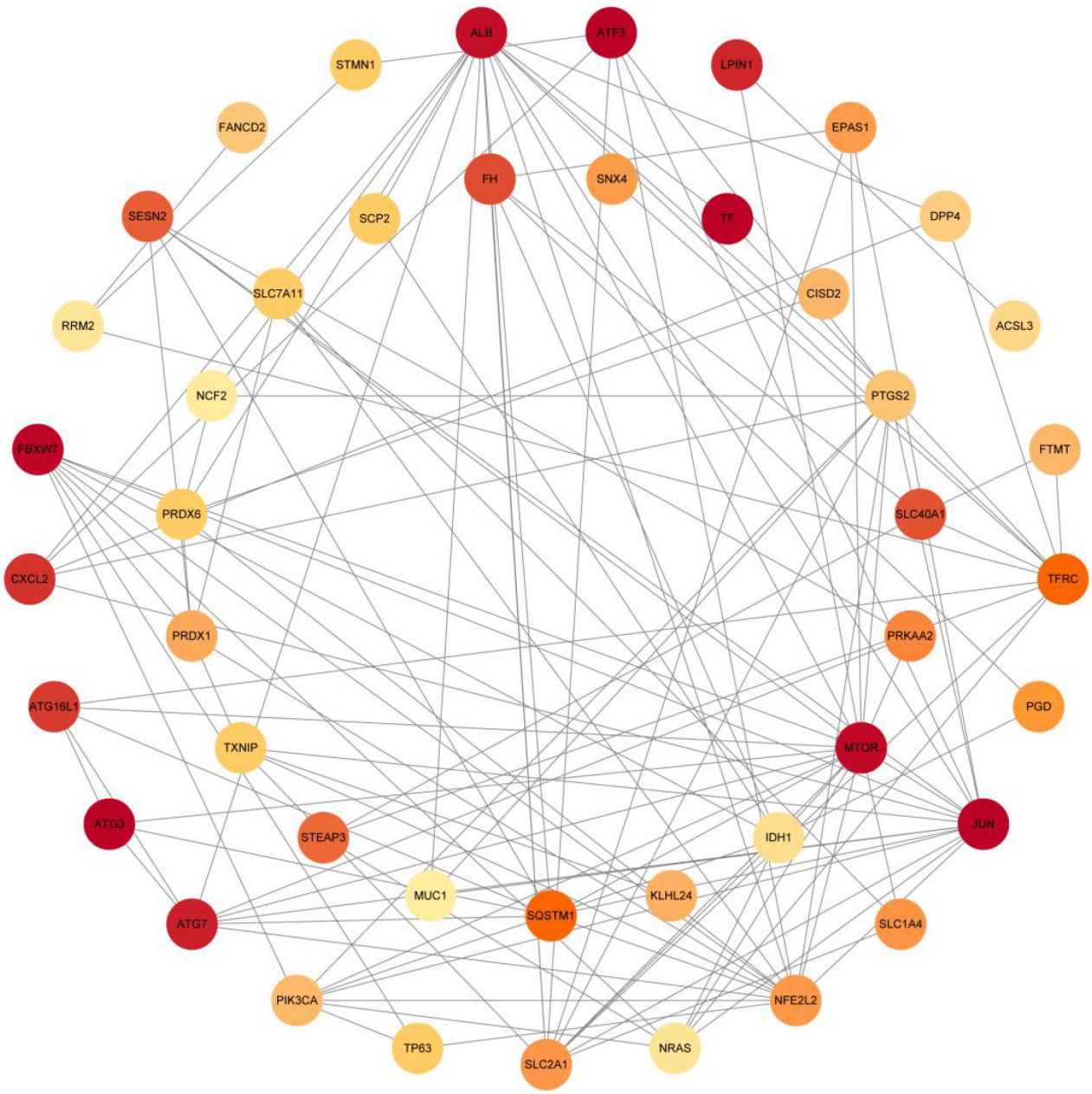

FIGURE 5 | Protein-protein interaction network of 53 ferroptosis-related module genes were analyzed using Cytoscape software. The network includes 44 nodes and 120 edges (The disconnected nodes were hided). The edges between 2 nodes represent the gene-gene interactions. The size and color of the nodes corresponding to each gene were determined according to the degree of interaction. Color gradients represent the variation of the degrees of each gene from high to low.

Venn diagram was used to describe the details of the overlapped genes.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis of Overlapped Genes

Functional enrichment analysis was performed in three domains of GO, including biological process (BP), cellular component (CC), and molecular function (MF). The KEGG database contains datasets of pathways involving biological functions, diseases, chemicals, and drugs. The enrichment analysis was carried out by clusterProfiler R package to determine the biological functions of the genes and associated pathways (Yu et al., 2012).

Protein-Protein Interaction Establishment and Identification of Hub Genes

An online tool (Search Tools for the Retrieval of Interacting Genes, STRING[4]) was used to analyze protein interactions. The PPI pairs were screened by confidence score (>0.40), and the PPI network was visualized by the Cytoscape V3.9.0 software (Shannon et al., 2003). Three indicators (Degree, closeness, and Betweenness) were calculated through CytoHubba to evaluate the importance of each node, and the top 10 nodes were selected. The hub genes were their common nodes.

Construction and Validation of the Logistic Regression

To effectively differentiate the AD patients from controls, the logistic regression was constructed, and to evaluate the performance of the logistic regression model for predicting the occurrence of AD, we performed receiver operating characteristic (ROC) curve analyses using the pROC package of R (Robin et al., 2011). We selected the statistically significant genes from hub genes ($P < 0.05$) and used the nomogram to predict the occurrence of AD. The expression level of the hub genes was shown by the violin plot.

RESULTS

Weighted Co-expression Network Construction and Identification of Core Modules

The scale-free network was constructed with the soft threshold set to 4 ($R^2 = 0.905$) (**Figures 2A,B**). Then, the adjacency matrix and topological overlap matrix were built. We then calculated the module eigengenes representing the overall gene expression level of each module; these were clustered based on their correlation. A total of 4 modules were identified and labeled with a unique color (**Figure 2C**). We analyzed the correlations of each eigengene with phenotype (AD or control samples), and found two modules were correlated with AD-namely, the turquoise (cor = −0.32, P = 2e-13), and blue (cor = 0.30, P = 1e-11) modules (**Figure 2D**). The 4,617 genes in these modules-which are associated with AD-were retained for further analysis.

The Extraction of Ferroptosis-Related Genes From FerrDb and Apoptosis-Related Genes From UniProt

The ferroptosis-related genes were downloaded and summarized from the FerrDb (Zhou and Bao, 2020; **Table 1**). 253 regulatory factors (including 108 drivers, 69 suppressors, 35 inducers, and 41 inhibitors), 111 markers, and 95 ferroptosis-related diseases were collated by FerrDb. We have extracted 2,130 genes from Uniprot, which is related to apoptosis.

Overlap Alzheimer's Disease-Related Module Genes With Ferroptosis-Related Genes and Apoptosis-Related Genes, Respectively

We overlapped the AD-related module genes derived from WGCNA with ferroptosis-related genes extracted from FerrDb, 53 overlapped genes were obtained, namely ferroptosis-related module genes, which was shown by the Venn diagram (**Figure 3A**). The details of overlapped genes, including 19 drivers, 16 suppressors, and 18 markers, were shown in **Table 2**. We also overlapped the AD-related module genes with apoptosis-related genes to obtain apoptosis-related module genes as a control for further analysis, and 90 overlapped genes were obtained, which was also shown by the Venn diagram (**Figure 3B**).

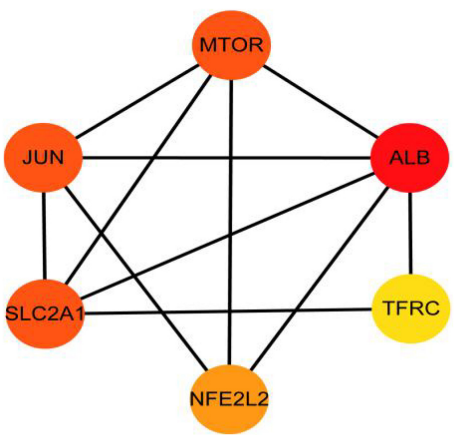

FIGURE 6 | Protein–protein interaction network for the six hub genes. Three indicators (degree, closeness and betweenness) were, respectively, calculated to evaluate the importance of each node and the top 10 nodes were selected. The six hub genes were their common nodes.

[4]https://string-db.org/

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis of Overlapped Genes

The significant GO functional terms of the 53 ferroptosis-related module genes, including BP, MF, and CC, were illustrated in **Figure 4A**. The significant terms of GO-BP were principally associated with the response to stress, such as the response to oxidative stress. The pathways enriched by GO-MF were principally associated with the activity of peroxidase, oxidoreductase, and antioxidant. The ferric iron-binding was also enriched by the GO-MF. The analysis of GO-CC indicated that overlapped genes were significantly enriched in basolateral plasma membrane, phagophore assembly site, pigment granule, and melanosome. The KEGG analysis showed that these overlapped genes were enriched not only in renal cell carcinoma pathways and central carbon metabolism in cancer, but also in autophagy-related pathways and ferroptosis (**Figure 4B**). The

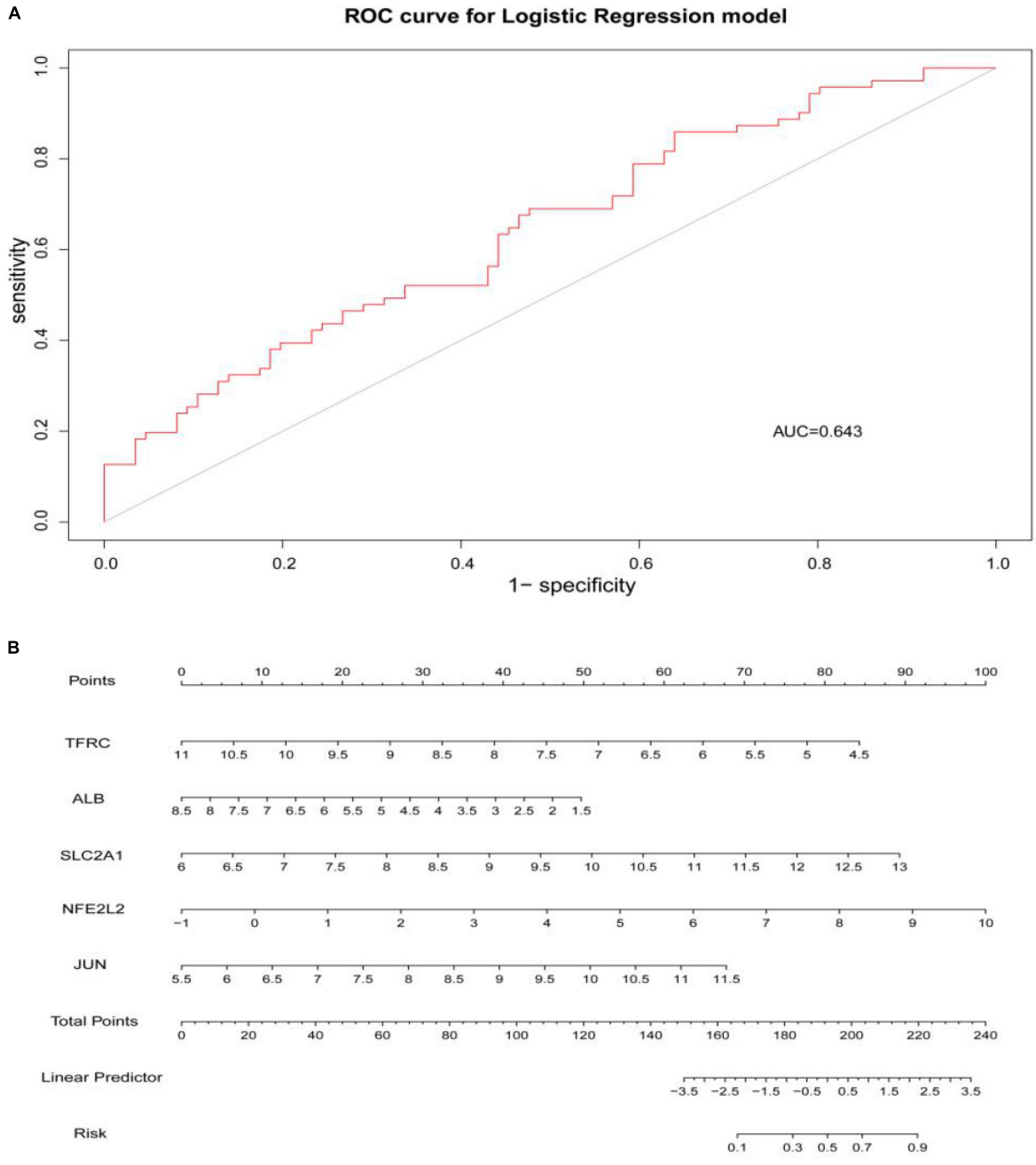

FIGURE 7 | (A) ROC curve was used to evaluate the performance of the logistic regression model. The area under the curve (AUC) was 0.643. **(B)** The nomogram was used to predict the occurrence of AD. Ferroptosis-related hub genes, JUN, SLC2A1, TFRC, ALB, and NFE2L2 ($P < 0.05$), were included in this nomogram.

pathway of ferroptosis was enriched by KEGG, suggesting that these overlapped genes were significant for our study and could be used for further analysis.

Protein-Protein Interaction Establishment and Identification of Hub Genes

The PPI analysis of 53 ferroptosis-related module genes was performed through the STRING database and visualized by Cytoscape V3.9.0 (**Figure 5**). JUN, SLC2A1, TFRC, ALB, MTOR, and NFE2L2 were taken as potential hub genes based on Degree, closeness, and betweenness. The hub genes were their common top ten nodes. The PPI network of the hub genes was presented in **Figure 6**. Similarly, the identification of hub genes of apoptosis-related module genes was also conducted, and STAT1, CFLAR, FASLG, MCL1 and BCL2L11 were obtained from the 90 overlapped genes.

Construction and Validation of the Logistic Regression

Through constructing the logistic regression, JUN, SLC2A1, TFRC, ALB, and NFE2L2 were selected, which could effectively differentiate AD patients from controls ($P < 0.05$). The P-value of MTOR was more than 0.05, which was not statistically significant. We used the ROC curve to evaluate the performance of the logistic regression model (the area under the ROC curve of the model was 0.643), and the nomogram was used for predicting the occurrence of AD (**Figures 7A,B**). The expression level of the five hub genes is shown in **Figure 8**. Similarly, the logistic regression was also constructed for apoptosis-related hub genes, and STAT1, MCL1, and BCL2L11 were selected and could distinguish AD patients from controls ($P < 0.05$). The AUC was 0.608, which was less than the former AUC value, suggesting that ferroptosis was more special than apoptosis in AD. The ROC curve and nomogram are shown in **Figures 9A,B**.

DISCUSSION

The pathological process of ferroptosis has some characteristics in common with AD, such as excess iron accumulation and elevated lipid peroxides. It has been reported that the pathological process of ferroptosis could be directly induced by iron overload (Wang et al., 2017; Fang et al., 2019). Clinically, lipid peroxidation metabolites were highly correlated with the progression of AD (Benseny-Cases et al., 2014). Besides, it has also been reported that reactive oxygen species (Wang et al., 2016) and reduced glutathione (Chiang et al., 2017) were found in the pathological process of AD. However, how does ferroptosis mediate AD? Some ferroptosis-related signaling pathways were found in AD, such as iron metabolism pathway, redox homeostasis pathway,

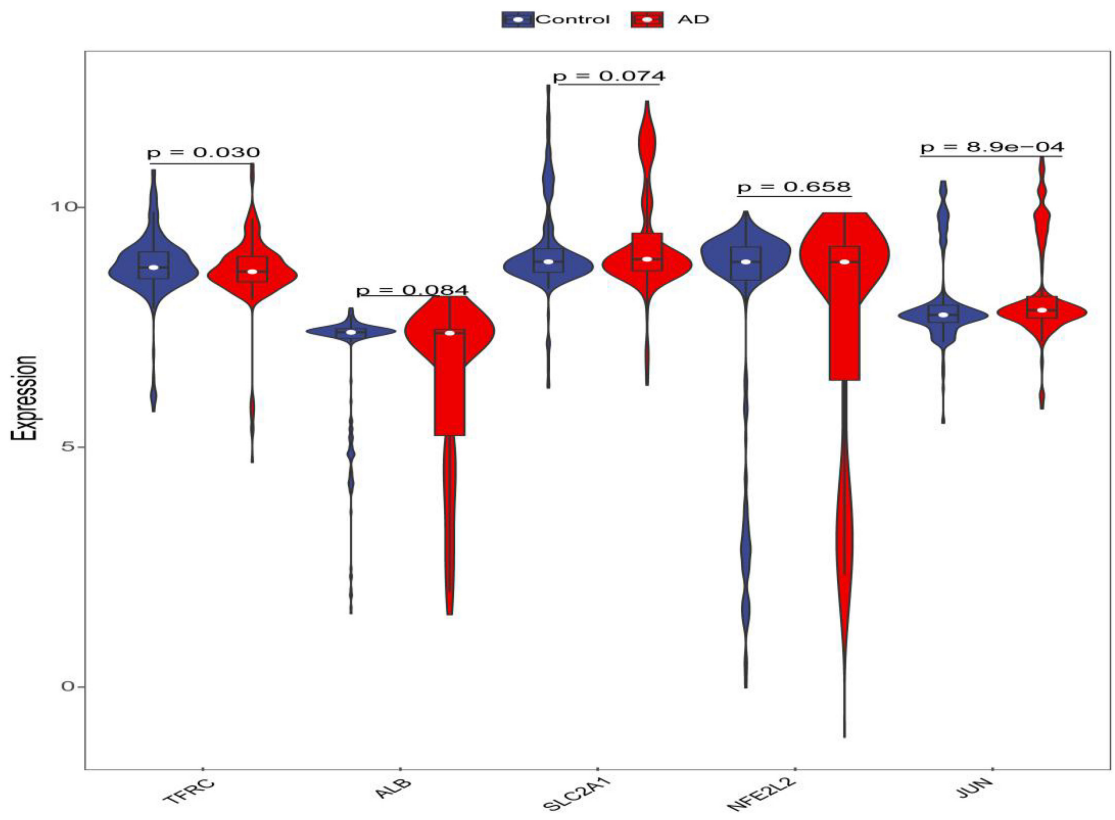

FIGURE 8 | Violin plot of the expression level of five hub genes. The red violin reflects the AD group, and the blue violin reflects the control group.

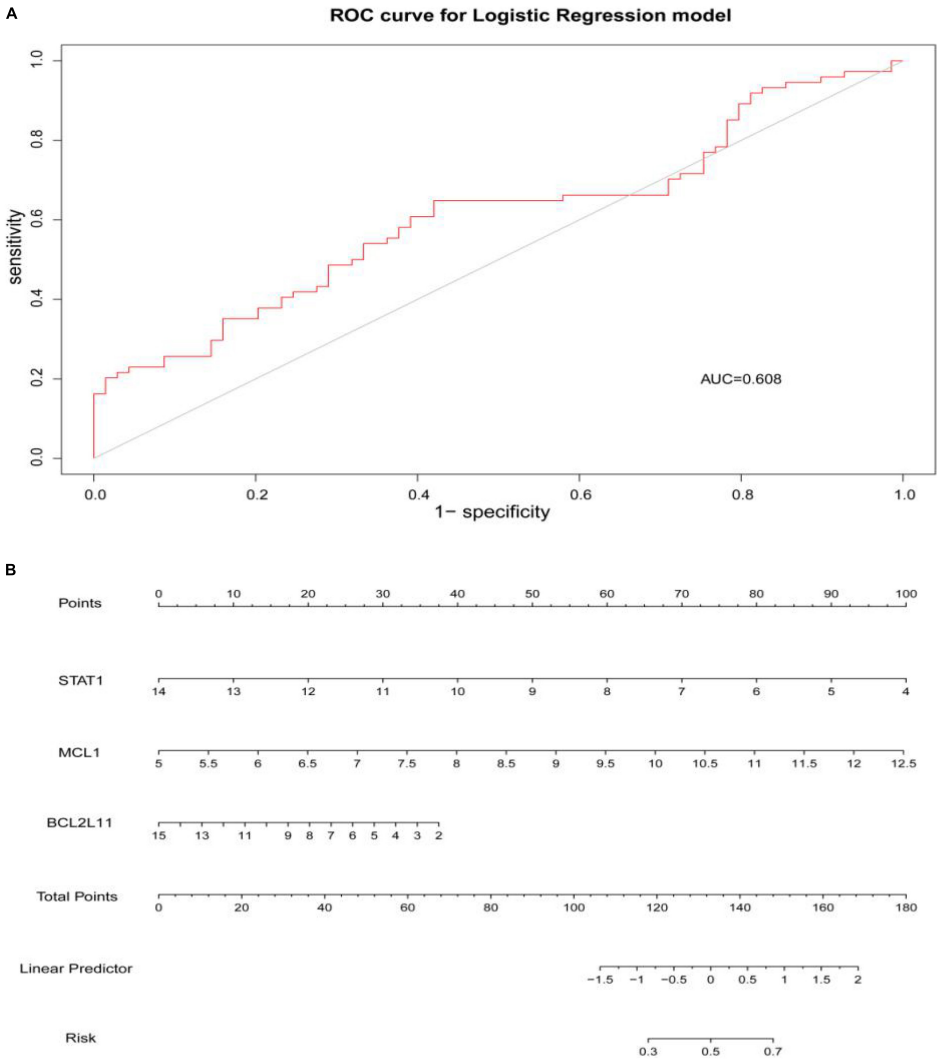

FIGURE 9 | (A) The area under the curve (AUC) was 0.608. **(B)** The nomogram of apoptosis-related hub genes, STAT1, MCL1, and BCL2L11 ($P < 0.05$).

and lipid metabolism pathway (Chen et al., 2021). Exploring of the mechanism of ferroptosis in AD could provide a novel therapeutic target for the treatment of AD and possibly, other neurodegenerative diseases (Ashraf and So, 2020). This study identified five hub genes that may participate in the pathologic processes associated with ferroptosis in AD. The possible pathways of these five genes involved in ferroptosis are shown in **Figure 10** (see text footnote 2) (Gao et al., 2016; Shin et al., 2018; Chen et al., 2019).

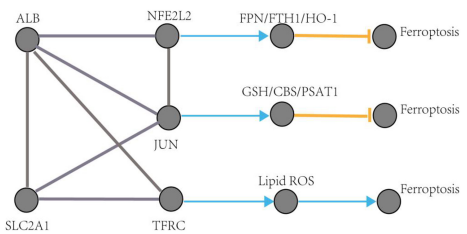

FIGURE 10 | The regulation pathways of JUN, SLC2A1, TFRC, ALB, and NFE2L2 participating in ferroptosis in AD. The gray edge represents the gene-gene interactions. The orange T-shaped edge denotes suppression, and the blue arrow denotes promotion.

Emerging evidence has demonstrated that ferroptosis could be a therapeutic target for AD (Gleason and Bush, 2021). Some ferroptosis inhibitors, such as iron-chelators and vitamin E, have shown clinical efficacy in treating AD. Deiprone is a brain osmotic iron-chelating agent currently in phase II clinical trials to treat AD (Nikseresht et al., 2019). Antioxidant vitamin E could delay decline in function and relieve caregiver burden in patients with AD (Dysken et al., 2014a,b). Collectively, Patients with AD may benefit from ferroptosis as a therapeutic target. Unlike targeting β-amyloid, the clinical trials of ferroptosis inhibitors are still in the exploratory stage and need to be dose-optimized and replicated on a larger scale (Nikseresht et al., 2019). The clinical

efficacy of ferroptosis inhibitors in the treatment of AD also needs to be further improved.

There were some limitations to this study. Firstly, while selecting datasets for differentially expressed analysis, it was found that some datasets had fewer or no differentially expressed genes (DEGs, correcting P-value < 0.05 and $|\log FC| \geq 1.0$), such as GSE48350 (Berchtold et al., 2013) and GSE131617 (Miyashita et al., 2014; Kikuchi et al., 2020). Therefore, the datasets and related AD patients we can choose are still limited. In addition, if the DEGs further overlaps with the ferroptosis-related module genes, the number of available genes are limited and could not be used for further analysis. Secondly, the potential ferroptosis-related biomarkers identified by this study still need further literature support and laboratory evidence verification. Thirdly, the ferroptosis-related genes are derived from FerrDb, which is being updated continuously, and more genes are yet to be discovered.

CONCLUSION

We identified five hub genes (JUN, SLC2A1, TFRC, ALB, and NFE2L2) that are closely associated with ferroptosis in AD and can differentiate AD patients from controls, and are thus potential ferroptosis-related biomarkers for disease diagnosis and therapeutic monitoring. Three hub genes of apoptosis-related genes in AD (STAT1, MCL1, and BCL2L11) were also identified as a control to show the specificity of Ferroptosis. JUN, SLC2A1, TFRC, ALB, and NFE2L2 are thus potential ferroptosis-related biomarkers for disease diagnosis and therapeutic monitoring.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

AUTHOR CONTRIBUTIONS

YW, GC, and WS contributed equally to this work. All authors contributed toward data analysis, drafted and critically revised the manuscript, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

We would like to thank the data support provided by the GEO databases, FerrDb and Uniprot.

REFERENCES

Ashraf, A., and So, P. W. (2020). Spotlight on Ferroptosis: iron-Dependent Cell Death in Alzheimer's Disease. *Front. Aging Neurosci.* 12:196. doi: 10.3389/fnagi.2020.00196

Ayton, S., Wang, Y., Diouf, I., Schneider, J. A., Brockman, J., Morris, M. C., et al. (2019). Brain iron is associated with accelerated cognitive decline in people with Alzheimer pathology. *Mol. Psychiatry* 25, 2932–2941. doi: 10.1038/s41380-019-0375-7

Benseny-Cases, N., Klementieva, O., Cotte, M., Ferrer, I., and Cladera, J. (2014). Microspectroscopy (muFTIR) reveals co-localization of lipid oxidation and amyloid plaques in human Alzheimer disease brains. *Anal. Chem.* 86, 12047–12054. doi: 10.1021/ac502667b

Berchtold, N. C., Coleman, P. D., Cribbs, D. H., Rogers, J., Gillen, D. L., and Cotman, C. W. (2013). Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol. Aging* 34, 1653–1661. doi: 10.1016/j.neurobiolaging.2012.11.024

Chen, K., Jiang, X., Wu, M., Cao, X., Bao, W., and Zhu, L. Q. (2021). Ferroptosis, a Potential Therapeutic Target in Alzheimer's Disease. *Front. Cell Dev. Biol.* 9:704298. doi: 10.3389/fcell.2021.704298

Chen, Y., Zhu, G., Liu, Y., Wu, Q., Zhang, X., and Bian, Z. (2019). O-GlcNAcylated c-Jun antagonizes ferroptosis via inhibiting GSH synthesis in liver cancer. *Cell Signal* 63:109384. doi: 10.1016/j.cellsig.2019.109384

Chiang, G. C., Mao, X., Kang, G., Chang, E., Pandya, S., Vallabhajosula, S., et al. (2017). Relationships among cortical glutathione levels, brain amyloidosis, and memory in healthy older adults investigated *in vivo* with H-1-MRS and pittsburgh compound-B PET. *Am. J. Neuroradiol.* 38, 1130–1137. doi: 10.3174/ajnr.a5143

Dysken, M. W., Guarino, P. D., Vertrees, J. E., Asthana, S., Sano, M., Llorente, M., et al. (2014a). Vitamin E and memantine in Alzheimer's disease: clinical trial methods and baseline data. *Alzheimer's Dement.* 10, 36–44. doi: 10.1016/j.jalz.2013.01.014

Dysken, M. W., Sano, M., Asthana, S., Vertrees, J. E., Pallaki, M., Llorente, M., et al. (2014b). Effect of vitamin E and memantine on functional decline in Alzheimer disease: the TEAM-AD VA cooperative randomized trial. *JAMA* 311, 1161–1161. doi: 10.1001/jama.2013.282834

Fang, X. X., Wang, H., Han, D., Xie, E. J., Yang, X., Wei, J. Y., et al. (2019). Ferroptosis as a target for protection against cardiomyopathy. *Proc. Natl. Acad. Sci. U S A.* 116, 2672–2680. doi: 10.1073/pnas.1821022116

Gao, M., Monian, P., Pan, Q., Zhang, W., Xiang, J., and Jiang, X. (2016). Ferroptosis is an autophagic cell death process. *Cell Res.* 26, 1021–1032. doi: 10.1038/cr.2016.95

Gbd 2016 Dementia Collaborators. (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 18, 88–106. doi: 10.1016/S1474-4422(18)30403-4

Gleason, A., and Bush, A. I. (2021). Iron and Ferroptosis as Therapeutic Targets in Alzheimer's Disease. *Neurotherapeutics* 18, 252–264. doi: 10.1007/sl3311-020-00954-y

Hambright, W. S., Fonseca, R. S., Chen, L., Na, R., and Ran, Q. (2017). Ablation of ferroptosis regulator glutathione peroxidase 4 in forebrain neurons promotes cognitive impairment and neurodegeneration. *Redox Biol.* 12, 8–17. doi: 10.1016/j.redox.2017.01.021

Jakaria, M., Bush, A. I., and Ayton, S. (2021). Ferroptosis as a mechanism of neurodegeneration in Alzheimer's disease. *J. Neurochem.* 159, 804–825. doi: 10.1111/jnc.15519

Kikuchi, M., Sekiya, M., Hara, N., Miyashita, A., Kuwano, R., Ikeuchi, T., et al. (2020). Disruption of a RAC1-centred network is associated with Alzheimer's disease pathology and causes age-dependent neurodegeneration. *Hum. Mol. Genet.* 29, 817–833. doi: 10.1093/hmg/ddz320

Lane, D. J. R., Ayton, S., and Bush, A. I. (2018). Iron and Alzheimer's Disease: an Update on Emerging Mechanisms. *J. Alzheimers Dis.* 64, S379–S395. doi: 10.3233/JAD-179944

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-559

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. *J. Statist. Softw.* 46:11.

Masaldan, S., Bush, A. I., Devos, D., Rolland, A. S., and Moreau, C. (2019). Striking while the iron is hot: Iron metabolism and ferroptosis in neurodegeneration. *Free Radic. Biol. Med.* 133, 221–233. doi: 10.1016/j.freeradbiomed.2018.09.033

Mckay, E. C., Beck, J. S., Khoo, S. K., Dykema, K. J., Cottingham, S. L., and Winn, M. E. (2019). Peri-Infarct Upregulation of the Oxytocin Receptor in Vascular Dementia. *J. Neuropathol. Exp. Neurol.* 78, 436–452. doi: 10.1093/jnen/nlz023

Miyashita, A., Hatsuta, H., Kikuchi, M., Nakaya, A., Saito, Y., Tsukie, T., et al. (2014). Genes associated with the progression of neurofibrillary tangles in Alzheimer's disease. *Transl. Psychiatry* 4:e396. doi: 10.1038/tp.2014.35

Nikseresht, S., Bush, A. I., and Ayton, S. (2019). Treating Alzheimer's disease by targeting iron. *Br. J. Pharmacol.* 176, 3622–3635. doi: 10.1111/bph.14567

Obulesu, M., and Lakshmi, M. J. (2014). Apoptosis in Alzheimer's disease: an understanding of the physiology, pathology and therapeutic avenues. *Neurochem. Res.* 39, 2301–2312. doi: 10.1007/s11064-014-1454-4

Plascencia-Villa, G., and Perry, G. (2021). Preventive and Therapeutic Strategies in Alzheimer's Disease: focus on Oxidative Stress. *Antioxid. Redox Signal.* 34, 591–610. doi: 10.1089/ars.2020.8134

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S + to analyze and compare ROC curves. *BMC Bioinform.* 12:77. doi: 10.1186/1471-2105-12-77

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sharma, V. K., Singh, T. G., Singh, S., Garg, N., and Dhiman, S. (2021). Apoptotic Pathways and Alzheimer's Disease: probing Therapeutic Potential. *Neurochem. Res.* 46, 3103–3122. doi: 10.1007/s11064-021-03418-7

Shimohama, S. (2000). Apoptosis in Alzheimer's disease–an update. *Apoptosis* 5, 9–16. doi: 10.1023/a:1009625323388

Shin, D., Kim, E. H., Lee, J., and Roh, J. L. (2018). Nrf2 inhibition reverses resistance to GPX4 inhibitor-induced ferroptosis in head and neck cancer. *Free Radi. Biol. Med.* 129, 454–462. doi: 10.1016/j.freeradbiomed.20180.10.426

Tan, M. G., Chua, W. T., Esiri, M. M., Smith, A. D., Vinters, H. V., and Lai, M. K. (2010). Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease. *J. Neurosci. Res.* 88, 1157–1169. doi: 10.1002/jnr.22290

Vitalakumar, D., Sharma, A., and Flora, S. J. S. (2021). Ferroptosis: a potential therapeutic target for neurodegenerative diseases. *J. Biochem. Mol. Toxicol.* 35:e22830. doi: 10.1002/jbt.22830

Wang, H., An, P., Xie, E., Wu, Q., Fang, X., Gao, H., et al. (2017). Characterization of ferroptosis in murine models of hemochromatosis. *Hepatology* 66, 449–465. doi: 10.1002/hep.29117

Wang, X., Hu, X., Yang, Y., Takata, T., and Sakurai, T. (2016). Nicotinamide mononucleotide protects against β-amyloid oligomer-induced cognitive impairment and neuronal death. *Brain Res.* 1643, 1–9. doi: 10.1016/j.brainres.2016.04.060

Weiland, A., Wang, Y., Wu, W., Lan, X., Han, X., Li, Q., et al. (2019). Ferroptosis and Its Role in Diverse Brain Diseases. *Mol. Neurobiol.* 56, 4880–4893. doi: 10.1007/s12035-018-1403-3

Yan, N., and Zhang, J. (2019). Iron Metabolism. *Front. Neurosci.* 13:1443. doi: 10.3389/fnins.2019.01443

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, C., Rodriguez, C., Spaulding, J., Aw, T. Y., and Feng, J. (2012). Age-dependent and tissue-related glutathione redox status in a mouse model of Alzheimer's disease. *J. Alzheimers Dis.* 28, 655–666. doi: 10.3233/JAD-2011-111244

Zhang, G., Zhang, Y., Shen, Y., Wang, Y., Zhao, M., and Sun, L. (2021). The Potential Role of Ferroptosis in Alzheimer's Disease. *J. Alzheimers Dis.* 80, 907–925. doi: 10.3233/JAD-201369

Zhou, N., and Bao, J. (2020). FerrDb: a manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. *Database J. Biol. Databases Curation* 2020:baaa021. doi: 10.1093/database/baaa021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 March 2022
doi: 10.3389/fnins.2022.817067

Mendelian Randomization Analysis Suggests No Associations of Herpes Simplex Virus Infections With Multiple Sclerosis

Wan Zhang[1,2†], Pengfei Wu[2,3†], Rui Yin[4], Meichen Sun[5], Rongsen Zhang[6], Xiaoyao Liao[7], Yuhong Lin[8] and Hui Lu[5]*

[1] Department of Biology, Boston University, Boston, MA, United States, [2] Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, United States, [3] School of Life Sciences, Central South University, Changsha, China, [4] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States, [5] Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing, China, [6] Department of Ultrasound, Second Xiangya Hospital, Central South University, Changsha, China, [7] College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, [8] Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

OPEN ACCESS

Edited by:
Liang Cheng,
Harbin Medical University, China

Reviewed by:
Wei Zhong Ji,
Qinghai Provincial People's Hospital, China
Chunpeng Wang,
Northeastern University, United States

***Correspondence:**
Hui Lu
erjihuixin@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: 17 November 2021
Accepted: 27 January 2022
Published: 01 March 2022

Citation:
Zhang W, Wu P, Yin R, Sun M, Zhang R, Liao X, Lin Y and Lu H (2022) Mendelian Randomization Analysis Suggests No Associations of Herpes Simplex Virus Infections With Multiple Sclerosis. Front. Neurosci. 16:817067. doi: 10.3389/fnins.2022.817067

Previous studies have suggested an association between infection with herpes simplex virus (HSV) and liability to multiple sclerosis (MS), but it remains largely unknown whether the effect is causal. We performed a two-sample Mendelian randomization (MR) study to explore the relationship between genetically predicted HSV infection and MS risk. Genetic instrumental variables for diagnosed infections with HSV ($p < 5 \times 10^{-6}$) were retrieved from the FinnGen study, and single nucleotide polymorphisms associated with circulating immunoglobulin G (IgG) levels of HSV-1 and HSV-2 and corresponding summary-level statistics of MS were obtained from genome-wide association studies of the European-ancestry. Inverse-variance weighted MR was employed as the primary method and multiple sensitivity analyses were performed. Genetically proxied infection with HSV was not associated with the risk of MS (odds ratio [OR], 0.96; 95% confidence interval [CI], 0.90–1.02; $p = 0.22$) per one-unit increase in log-OR of herpes viral infections. MR results provided no evidence for the relationship between circulating HSV-1 IgG levels and MS risks (OR = 0.91; 95% CI, 0.81–1.03; $p = 0.37$), and suggested no causal effect of HSV-2 IgG (OR = 1.04; 95% CI, 0.96–1.13; $p = 0.32$). Additional sensitivity analyses confirmed the robustness of these null findings. The MR study did not support the causal relationship between genetic susceptibly to HSV and MS in the European population. Further studies are still warranted to provide informative knowledge, and triangulating evidence across multiple lines of evidence are necessary to plan interventions for the treatment and prevention of MS.

Keywords: multiple sclerosis, herpes simplex virus, genetic epidemiology, Mendelian randomization, causal risk factors

INTRODUCTION

Multiple sclerosis (MS) is the most common chronic demyelinating and neurodegenerative disease of the central nervous system (CNS) (Hauser and Cree, 2020). It is the leading cause of non-traumatic neurological disability in young adults, affecting more than 2 million people worldwide (GBD, 2017). The symptoms of MS usually follow relapsing or progressive path, eventually leading to impaired mobility or cognition (Reich et al., 2018). MS is currently incurable though therapeutic advances have remarkably improved the long-term outcome for patients at this time (Hauser and Cree, 2020; Iqubal et al., 2020). The etiology of MS has not been fully elucidated. Early infections with herpes simplex virus (HSV) infection are constantly proposed to be involved in the pathogenesis of MS. HSV-1 and HSV-2 infections usually occur in the early years of life, mostly latent and asymptomatic (Koyuncu et al., 2013). HSV viruses lurk in the sensory ganglion of the trigeminal nerve, remain exist lifelong, and could invade CNS (Kimberlin et al., 2001). Post-mortem results have also confirmed the presence of HSV in brain demyelinating plaques of MS patients (Sanders et al., 1996).

Based on retrospective data in Sarajevo, the positive incidence of HSV immunoglobulin G (IgG) antibodies was 93.2% in 110 newly diagnosed MS patients (Djelilovic-Vranic and Alajbegovic, 2012). In another study, the prevalence of HSV-1 mRNA and DNA in the peripheral blood mononuclear cells (PBMC) of acute MS patients is significantly higher compared to controls (Ferrante et al., 2000). They also suggested that HSV-1 reactivate in the acute attack and might trigger MS relapses (Ferrante et al., 2000). Data addressing pediatric MS showed that HSV-1 IgG antibodies in serum was associated with increased risk of pediatric MS (Waubant et al., 2011; Nourbakhsh et al., 2018). Waubant et al. (2011) recruited 189 pediatric MS patients and found that HSV-1 was associated with an increased risk of MS in those negative for HLA DRB1*1501. Another multi-center research suggested that sero-positivity for HSV-1 was significantly increased in pediatric MS patients, but the increase was only seen in Caucasian people and those without a DRB1*15 allele (Nourbakhsh et al., 2018). Pooled results of a recent meta-analysis has implicated a statistical difference in the serum prevalence of IgG against HSV-2 between patients with MS and controls (Xu et al., 2021).

However, other studies reported conflicting results, and did not find any relationship between HSV infection and MS risk. Data in several studies showed that the prevalence of antibodies against HSV-1 or HSV-2 had no statistical associations with adult MS (Wandinger et al., 2000; Kiriyama et al., 2010; Sotelo et al., 2014; Etemadifar et al., 2019). By testing HSV DNA in cerebrospinal fluid or in PBMC, Koros et al. (2014) and Sotelo et al. (2014) reported no significant difference of HSV DNA between adult MS and healthy controls. Another pediatric study found no difference in the association of prior HSV infections with the onset of pediatric MS (Mowry et al., 2011).

Those equivocal results might be caused by methodological shortcomings of observational studies, such as residual confounding and reverse causality. Confined by these limitations, observational research is unable to deduce the causal role of HSV infection in the development of MS. With the exponential growth in and widespread availability of genotype data, Mendelian randomization (MR) approach as an epidemiologic study designed to establish causality between exposures and outcomes has gained its popularity in the last two decades (Zhuang et al., 2019; Huang et al., 2021; Kwok and Schooling, 2021; Zhang et al., 2021). MR utilizes germline genetic variants as proxies. Since genetic variants are unaffected by environmental factors or disease process, MR can diminish confounding, strengthen exposure-outcome associations and avoid reverse causalities (Smith and Ebrahim, 2003). In this study, we leveraged the MR approach to infer the associations of HSV infection with risk of MS.

MATERIALS AND METHODS

The schematic for the MR design was shown in **Figure 1** and datasets underlying the study was summarized in **Supplementary Table 1**. This study was built upon summary-level statistics which were publicly accessible. Informed consent from participants and approval by ethical committees had been completed by consortia involved in original studies.

Instrumental Variables for Herpes Simplex Virus

Genetic instrumental variants for HSV infection were obtained from the FinnGen study (FinnGen, 2021). Diagnosed infections with HSV were defined by International Classification of Diseases (ICD) from the Finnish registries of inpatient, outpatient and cause of death. In the R5 release, there were 1,595 cases (ICD-10, B00; ICD-9, and ICD-8, 054) and 211,856 participating controls of Finnish ancestry. Sex, age, 10 principal components and genotyping batch were included as logistic regression covariates. Eight instrumental single-nucleotide polymorphisms (SNPs) were selected at a suggestive genome-wide significance threshold ($p < 5 \times 10^{-6}$) as previous studies did (Kodali et al., 2018; Bae and Lee, 2020; Kwok and Schooling, 2021). The effect size was presented in a unit of log- odds ratio (OR) using the additive model (**Supplementary Table 2**).

Instrumental variables for circulating HSV-1 and HSV-2 IgG levels were selected from one genome-wide association study (GWAS) conducted in the Milieu Intérieur cohort (Scepanovic et al., 2018). Total IgG levels and antigen specific seropositivity was tested in 1,000 individuals, and in seropositive donors, serum IgGs specific for HSV-1 (n = 645) and HSV-2 (n = 208) were further measured using the BioPlex^{TM} 2200 HSV-1 and HSV-2 IgG kit (Bio-Rad, Hercules, CA, United States). After log10-transformed of IgG levels, genetic association analyses were performed with the additive regression adjusted for age, sex, total IgG and the first two principal components incorporated as covariates. Four and eight SNPs were utilized as instrumental variables for HSV-1 IgG (**Supplementary Table 3**) and HSV-2 IgG (**Supplementary Table 4**), respectively.

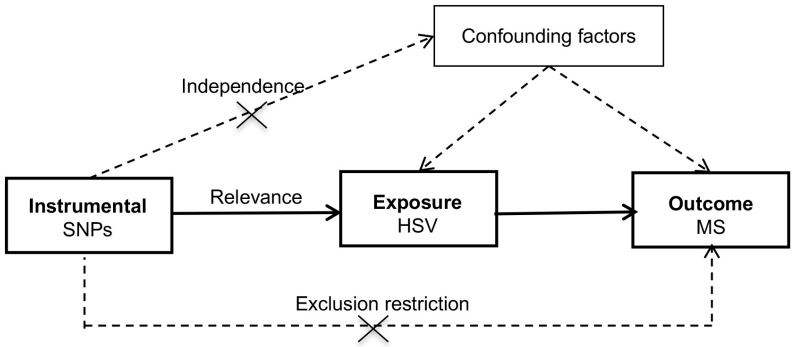

FIGURE 1 | Schematic diagram of the Mendelian randomization study. HSV, herpes simplex virus; MR, Mendelian randomization; MS, multiple sclerosis; SNP, single nucleotide polymorphism.

TABLE 1 | Association of genetically predicted herpes simplex virus infection with the risk of multiple sclerosis by different Mendelian randomization approaches.

MR methods	HSV infection			HSV-1 IgG			HSV-2 IgG		
	OR	95% CI	*P*-value	OR	95% CI	*P*-value	OR	95% CI	*P*-value
Inverse variance weighted	0.96	0.90–1.02	0.22	0.75	0.35–1.60	0.45	1.04	0.96–1.13	0.32
Weighted median	0.98	0.90–1.06	0.58	0.92	0.67–1.27	0.62	1.03	0.94–1.14	0.49
MR-Egger regression slope	0.95	0.80–1.13	0.59	0.45	0.004–44.82	0.76	0.89	0.58–1.36	0.60
MR-PRESSO raw estimate	0.96	0.92–1.01	0.12	0.75	0.35–1.60	0.51	1.04	0.96–1.13	0.35
MR-PRESSO outlier corrected	–	–	–	0.91	0.81–1.03	0.37	–	–	–

CI, confidence interval; HSV, herpes simplex virus; MR, Mendelian randomization; MR-PRESSO, Mendelian randomization pleiotropy residual sum and outlier; OR, odds ratio.

TABLE 2 | Results from Mendelian randomization sensitivity analyses between herpes simplex virus (HSV) and multiple sclerosis (MS).

Exposures	MR-Egger regression			Heterogeneity test		MR-PRESSO global test	
	Intercept	SE	*P*-value	*Q* statistic	*P*-value	*RSSobs*	*P*-value
HSV infection	0.002	0.02	0.91	4.05	0.85	4.92	0.88
HSV-1 IgG	0.05	0.21	0.84	33.98	<0.001	58.55	<0.001
HSV-1 IgG (excluding rs3132935)	0.04	0.11	0.79	5.06	0.08	–	–
HSV-2 IgG	0.03	0.04	0.47	8.77	0.27	11.42	0.30

MR-PRESSO, Mendelian randomization pleiotropy residual sum and outlier; RSSobs, observed residual sum of squares; SE, standard error.
MR-PRESSO global test was not available when examining the association of HSV-IgG (excluding rs3132935) with multiple sclerosis due to insufficient number of genetic instrumental variables.

Summary-Level Datasets of Multiple Sclerosis

Summary-level GWAS results of MS were released by the International Multiple Sclerosis Genetics Consortium (2019). In total, 14,802 individuals diagnosed with MS and 26,703 healthy controls of the European-ancestry were recruited in the discovery stage with 8,589,719 SNPs being meta-analyzed. Diagnostic criteria and demographic descriptions in each contributing cohort were summarized in the published GWAS (International Multiple Sclerosis Genetics Consortium, 2019). Effect estimates were adjusted for age, sex, batch effects and ten principal components in the logistic regression, and *Beta* represented one-unit increase in log-OR of MS per additional effect allele (**Supplementary Table 5**). Here, the following formulae were employed in transforming variables: $Beta = log(OR)$ and $Standard\ error = Beta/abs((qnorm(P-value/2))$. We kept instrumental SNPs which were present in the MS dataset, or whose proxied SNPs ($r^2 > 0.8$ or $D' > 0.8$, EUR panel 1000 Genomes Phase 3) were available. We harmonized the exposure and outcome effect size in terms of the effect allele and merged datasets were used for subsequent analyses.

Statistical Analysis

We performed MR analyses in the R language, version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) with the TwoSampleMR and MR-PRESSO packages (Hemani et al., 2018; Verbanck et al., 2018). Effect of HSV related exposures on the risk of MS contributed by individual instrumental variable was first given by Wald ratio: Y_k/X_k with its standard error ó$\acute{o}_{Y_k}/X_k$, where the SNP-effect on HSV was denoted with X_k and its standard error σ_{X_k}, and the SNP-MS association statistics denoted with Y_k and σ_{Y_k}. Then the primary MR method,

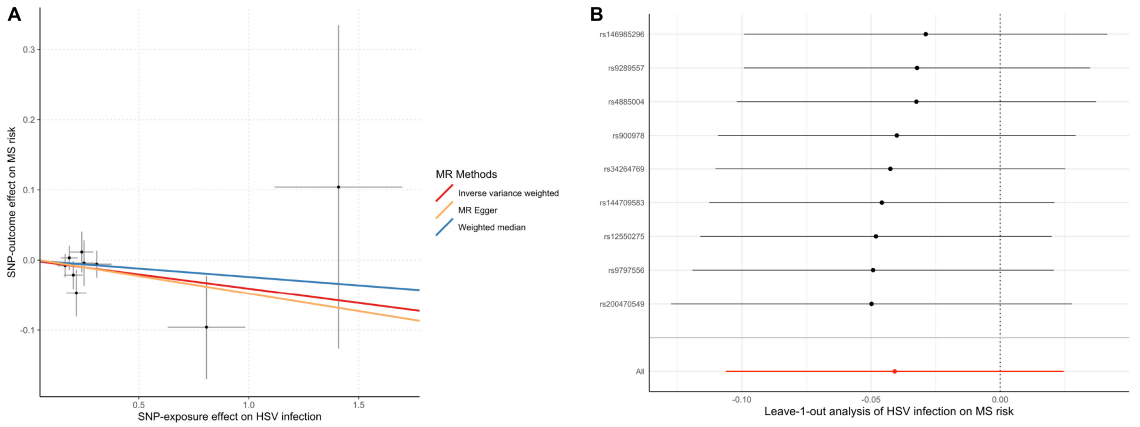

FIGURE 2 | Scatter plot **(A)** and leave-one-out plot **(B)** in the Mendelian randomization analysis of HSV infection on MS risk. HSV, herpes simplex virus; MR, Mendelian randomization; MS, multiple sclerosis.

the inverse-variance weighted (IVW) model combined ratio estimates for each exposure and yielded an overall estimate: $\widehat{\beta}_{MR} = \sum X_k Y_k \sigma_{Y_k}^{-2} / \sum X_k^2 \sigma_{Y_k}^{-2}$ with $\widehat{\sigma}_{MR} = \sqrt{1/\sum X_k^2 \sigma_{Y_k}^{-2}}$. Based on stringent prerequisites, IVW estimates would be biased if not all variants are valid or unbalanced pleiotropy exists (Burgess et al., 2013). Three additional approaches were implemented. Weighted median method effectively pooled individual estimate if less than half instrumental SNPs were invalid (Bowden et al., 2016). MR-Egger regression identified horizontal pleiotropic effects with p for intercept <0.05, meanwhile the regression slope provided a causal estimate corrected for unbalanced pleiotropy (Bowden et al., 2015). MR-PRESSO also examined outlier SNPs with potential pleiotropy by the global test and computed both a raw estimate and an outlier-adjusted estimate (Verbanck et al., 2018). We conducted Cochran's Q test and leave-one-out analysis to identify individual SNP which exerted an extremely heterogenous effect. As a measure of causal associations between HSV-related exposures and the risk of MS, we reported OR and 95% confidence interval (CI) per one unit increase in log-OR of diagnosed HSV infection or one SD elevation in circulating IgG levels of HSV-1 or HSV-2. Associations with $P < 0.05/3$, using the Bonferroni correction, were deemed as significant.

RESULTS

Association of Herpes Simplex Virus Infection With Multiple Sclerosis Risk

In the MR analysis investigating the relationship between infections with HSV and MS risk, nine instrumental SNPs were utilized and they collectively explained 0.09% variances of HSV (**Supplementary Table 2**). MR results suggested that diagnosed infections with HSV were not associated with the risk of MS (**Table 1**). By the IVW method, OR of MS was 0.96 (95% CI, 0.90–1.02; p = 0.22) per one-unit increase in log-OR of herpes viral infections. Sensitivity analyses by weighted median, MR-Egger regression slope and MR-PRESSO provided similar and consistent results. There was no evidence of pleiotropy by MR-Egger regression intercept (p = 0.85) or MR-PRESSO global test (p = 0.91). Besides, Cochran's Q test (**Table 2**) and leave-one-out analysis (**Figure 2**) indicated no heterogeneity among the instrumental SNPs.

Effect of Circulating Herpes Simplex Virus-1 and Herpes Simplex Virus-2 Immunoglobulin G Levels on Multiple Sclerosis Risk

Genetically predicted HSV-1 IgG was not associated with the risk of MS (OR = 0.75; 95% CI, 0.35–1.60; p = 0.45) by the IVW method. Notably, rs3132935 was associated with MS at genome-wide significance ($p = 3.40 \times 10^{-9}$). MR-PRESSO global test, Cochran's Q test (**Table 2**) and leave-one-out analysis (**Figure 3**) all indicated that rs3132935 might have pleiotropic effects and was an outlier variant in the MR analysis. Nevertheless, the MR-PRESSO corrected estimate with the removal of rs3132935 suggested no causal effect of circulating HSV-1 IgG levels on MS risks (OR = 0.91; 95% CI, 0.81–1.03; p = 0.37), either.

The MR analyses did not support the causal effect of HSV-2 IgG on MS (OR = 1.04; 95% CI, 0.96–1.13; p = 0.32) per one SD increase in HSV-2 IgG levels. Additional MR methods provided consistent results (**Table 1**). Furthermore, no unbalanced horizontal pleiotropy or evident heterogeneity was identified through multiple sensitivity analysis (**Figure 4**).

DISCUSSION

Seroprevalence of HSV is ubiquitous in populations where MS is prevalent (50–100% in adult members) (Lycke, 2017). In response to exposure to HSV-1, persistent lymphocytic cells would infiltrate in the CNS, levels of cytokine transcripts would elevate, and amounts of chemokine mRNAs would increase, which suggest that latent HSV-1 infection might trigger a chronic inflammatory process in brain tissue (Theil et al., 2003; Menendez et al., 2016). Meanwhile, levels of matrix metalloproteinases 2

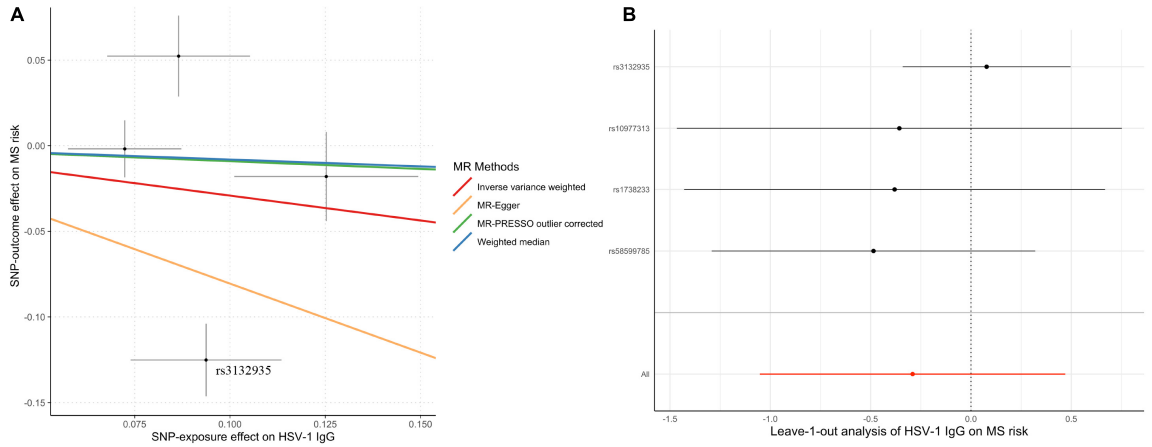

FIGURE 3 | Scatter plot **(A)** and leave-one-out plot **(B)** in the Mendelian randomization analysis of HSV-1 IgG levels on multiple sclerosis. MR-PRESSO outlier-corrected estimate was calculated with the removal of rs3132935 as an outlying variant, while the raw estimate was not delineated since it was nearly the same as the value given by inverse-variance-weighted method. HSV, herpes simplex virus; MR, Mendelian randomization; MR-PRESSO, Mendelian randomization pleiotropy residual sum and outlier; MS, multiple sclerosis.

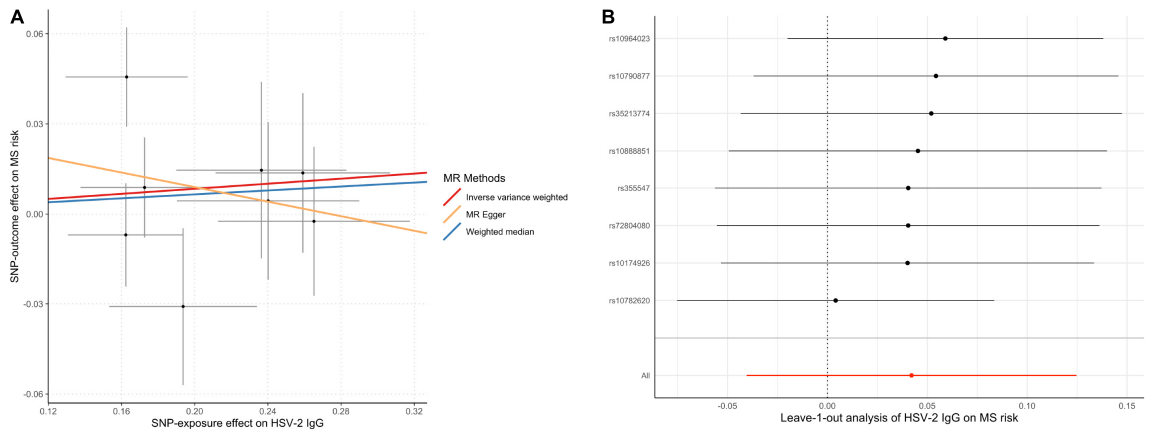

FIGURE 4 | Scatter plot **(A)** and leave-one-out plot **(B)** in the Mendelian randomization analysis of HSV-2 IgG on MS risk. HSV, herpes simplex virus; MR, Mendelian randomization; MS, multiple sclerosis.

and 9 would increase in infected CNS, and contribute to a partial breakdown of the blood brain barrier which is crucial in MS (Martínez-Torres et al., 2004). After exposure to HSV-1, plasmacytoid dendritic cells (DC) produce a great deal of Type I IFNs, including IFN-α and β (Soumelis and Liu, 2006). Plasmacytoid DC can promote naïve T cells to produce IL-10, which leads to anti-inflammatory reactions (Rissoan et al., 1999). Type I IFNs can inhibit the production of IL-12 and increase IL-10 production, which act on myeloid DC and switch pro- to anti-inflammation (Sanna et al., 2008). PBMC of MS patients showed increased production of IFN-α, IL-6, and IL-10 but decreased production of IL-4 (Sanna et al., 2008) and productions of IL-6 and IL-10 by PBMC and plasmacytoid DC were lower in MS patients compared with healthy controls (Sanna et al., 2008). The above data implied impaired anti-inflammatory response after HSV-1 infection in MS. Animal experiments have further proved that previous exposure to HSV-1 can cause an earlier onset of symptoms and more severe experimental autoimmune encephalomyelitis compared to uninfected control mice (Duarte et al., 2021).

Clinical trials of antiviral treatments in MS (Lycke et al., 1996; Bech et al., 2002; Friedman et al., 2005) were limited when compared with the development of other therapies (Lizak et al., 2017; Islam et al., 2020). There are three phase II clinical trials of acyclovir or valacyclovir in MS patients (Lycke et al., 1996; Bech et al., 2002; Friedman et al., 2005). One trial showed 34% reduction of annualized relapse rate in acyclovir-treated patients and a significant reduction in the relapse rate in favor of acyclovir treatment (Bech et al., 2002). In a high-activity group of another trial, valacyclovir-treated patients had significant reduction of new lesions compared to placebo-treated patients (Friedman et al., 2005). Although the above research suspected HSV as a candidate for the etiology of MS, the fact that HSV infection is far more prevalent in human populations compared to MS

argues against this viewpoint. HSV DNA in 77 demyelinated plaques from 23 MS patients revealed that HSV-1 DNA was amplified from only one plaque and HSV-2 DNA was amplified from none of the plaques (Nicoll et al., 1992). The infection of HSV in the CNS might be insufficient for the development of MS which requires other genetic and environmental triggers. Further investigations are warranted to detangle the role of HSV in disease onset or disease progression of MS.

The major strength of this study is the multivariable MR method, which explicated the roles of HSV infection in MS and exempted the result from residual confounding or reverse causality. Also, up-to-date genetic instruments for HSV infection traits and the largest GWAS dataset for MS were used to boost the power. There are several limitations for this study. Firstly, instrumental SNPs collectively explained small proportions of variance for HSV infection, and especially for circulating IgG levels of HSV-1 and HSV-2 due to inadequate sample size. Hence, we had restricted power to identify small causal effects. Secondly, we used a relaxed significance level ($p < 5 \times 10^{-6}$) rather than the classical GWAS threshold ($p < 5 \times 10^{-8}$) to choose instrumental variables. Distortion to the overall estimate might occur in the scenario, albeit no weak instrument was identified in the present study. Thirdly, biological implications for most SNPs are yet to be explored; thus, the suitability of current instrumental sets would be disputed by the possibility of pleiotropy, although no pleiotropic effects (except for rs3132935) were indicated through our sensitivity analyses. Lastly, this study was based on genome-wide association data only from the Europeans and we should be cautious with the interpretation and generalization when it comes to other populations.

CONCLUSION

In conclusion, we failed to provide evidence for the effect of HSV on the risk of MS. Further studies triangulating evidence from observational cohorts, clinical trials and genetic-epidemiological biobanks are still warranted to elucidate whether targeting HSV is an effective intervention for MS.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

WZ, PW, and HL conceptualized the study. PW, RY, RZ, and HL contributed to the data analysis and interpretation. WZ, MS, XL, and YL contributed to the manuscript drafting and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Scientific Research and Cultivation Plan of Beijing Municipal Hospital (PX2021036). PW received a visiting Ph.D. stipend from the China Scholarship Council (201906370155).

ACKNOWLEDGMENTS

We want to acknowledge the participants and investigators of FinnGen study, the Milieu Intérieur Consortium and International Multiple Sclerosis Genetics Consortium. We authors are grateful to them for making summary-level association statistics of HSV and MS possible and accessible. We are grateful for technical support and resources from the High Performance Computing Center of Central South University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.817067/full#supplementary-material

REFERENCES

Bae, S. C., and Lee, Y. H. (2020). Causal association between periodontitis and risk of rheumatoid arthritis and systemic lupus erythematosus: a Mendelian randomization. *Z. Rheumatol.* 79, 929–936. doi: 10.1007/s00393-019-00742-w

Bech, E., Lycke, J., Gadeberg, P., Hansen, H. J., Malmeström, C., Andersen, O., et al. (2002). A randomized, double-blind, placebo-controlled MRI study of anti-herpes virus therapy in MS. *Neurology* 58, 31–36. doi: 10.1212/wnl.58.1.31

Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44, 512–525. doi: 10.1093/ije/dyv080

Bowden, J., Davey Smith, G., Haycock, P. C., and Burgess, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40, 304–314. doi: 10.1002/gepi.21965

Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* 37, 658–665. doi: 10.1002/gepi.21758

Djelilovic-Vranic, J., and Alajbegovic, A. (2012). Role of early viral infections in development of multiple sclerosis. *Med. Arch.* 66, 37–40.

Duarte, L. F., Altamirano-Lagos, M. J., Tabares-Guevara, J. H., Opazo, M. C., Díaz, M., Navarrete, R., et al. (2021). Asymptomatic herpes simplex virus type 1 infection causes an earlier onset and more severe experimental autoimmune encephalomyelitis. *Front. Immunol.* 12:635257. doi: 10.3389/fimmu.2021.635257

Etemadifar, M., Izadi, A., Sabeti, F., and Noorshargh, P. (2019). Anti-HSV-2 antibody in patients with MS and NMO. *Mult. Scler. Relat. Disord.* 28, 286–289. doi: 10.1016/j.msard.2019.01.004

Ferrante, P., Mancuso, R., Pagani, E., Guerini, F. R., Calvo, M. G., Saresella, M., et al. (2000). Molecular evidences for a role of HSV-1 in multiple sclerosis clinical acute attack. *J. Neurovirol.* 6(Suppl. 2), S109–S114.

FinnGen (2021). *FinnGen Documentation of R5 Release*. Available online at: https://www.finngen.fi/en/access_results (accessed May 11, 2021).

Friedman, J. E., Zabriskie, J. B., Plank, C., Ablashi, D., Whitman, J., Shahan, B., et al. (2005). A randomized clinical trial of valacyclovir in multiple sclerosis. *Mult. Scler.* 11, 286–295. doi: 10.1191/1352458505ms1185oa

GBD (2017). Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Neurol.* 16, 877–897. doi: 10.1016/s1474-4422(17)30299-5

Hauser, S. L., and Cree, B. A. C. (2020). Treatment of multiple sclerosis: a review. *Am. J. Med.* 133, 1380–1390.e1382. doi: 10.1016/j.amjmed.2020.05.049

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 7:e34408. doi: 10.7554/eLife.34408

Huang, S. Y., Yang, Y. X., Kuo, K., Li, H. Q., Shen, X. N., Chen, S. D., et al. (2021). Herpesvirus infections and Alzheimer's disease: a Mendelian randomization study. *Alzheimers Res. Ther.* 13:158. doi: 10.1186/s13195-021-00905-5

International Multiple Sclerosis Genetics Consortium (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 365:eaav7188. doi: 10.1126/science.aav7188

Iqubal, A., Iqubal, M. K., Khan, A., Ali, J., Baboota, S., and Haque, S. E. (2020). Gene therapy, a novel therapeutic tool for neurological disorders: current progress, challenges and future prospective. *Curr. Gene Ther.* 20, 184–194. doi: 10.2174/1566523220999200716111502

Islam, M. A., Kundu, S., and Hassan, R. (2020). Gene therapy approaches in an autoimmune demyelinating disease: multiple sclerosis. *Curr. Gene Ther.* 19, 376–385. doi: 10.2174/1566523220666200306092556

Kimberlin, D. W., Lin, C. Y., Jacobs, R. F., Powell, D. A., Corey, L., Gruber, W. C., et al. (2001). Safety and efficacy of high-dose intravenous acyclovir in the management of neonatal herpes simplex virus infections. *Pediatrics* 108, 230–238. doi: 10.1542/peds.108.2.230

Kiriyama, T., Kataoka, H., Kasai, T., Nonomura, A., and Ueno, S. (2010). Negative association of Epstein-Barr virus or herpes simplex virus-1 with tumefactive central nervous system inflammatory demyelinating disease. *J. Neurovirol.* 16, 466–471. doi: 10.1007/bf03210852

Kodali, H. P., Pavilonis, B. T., and Schooling, C. M. (2018). Effects of copper and zinc on ischemic heart disease and myocardial infarction: a Mendelian randomization study. *Am. J. Clin. Nutr.* 108, 237–242. doi: 10.1093/ajcn/nqy129

Koros, C., Ioannidis, A., Acquaviva, T., Zoga, M., Nikolaou, C., Chatzipanagiotou, S., et al. (2014). HSV1 and 2 detection in the CSF of multiple sclerosis patients by real-time PCR. *In Vivo* 28, 1201–1205.

Koyuncu, O. O., Hogue, I. B., and Enquist, L. W. (2013). Virus infections in the nervous system. *Cell Host Microbe* 13, 379–393. doi: 10.1016/j.chom.2013.03.010

Kwok, M. K., and Schooling, C. M. (2021). Herpes simplex virus and Alzheimer's disease: a Mendelian randomization study. *Neurobiol. Aging* 99, 101.e111–101.e113. doi: 10.1016/j.neurobiolaging.2020.09.025

Lizak, N., Lugaresi, A., Alroughani, R., Lechner-Scott, J., Slee, M., Havrdova, E., et al. (2017). Highly active immunomodulatory therapy ameliorates accumulation of disability in moderately advanced and advanced multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 88, 196–203. doi: 10.1136/jnnp-2016-313976

Lycke, J. (2017). Trials of antivirals in the treatment of multiple sclerosis. *Acta Neurol. Scand.* 136(Suppl. 201), 45–48. doi: 10.1111/ane.12839

Lycke, J., Svennerholm, B., Hjelmquist, E., Frisén, L., Badr, G., Andersson, M., et al. (1996). Acyclovir treatment of relapsing-remitting multiple sclerosis. A randomized, placebo-controlled, double-blind study. *J. Neurol.* 243, 214–224. doi: 10.1007/bf00868517

Martínez-Torres, F. J., Wagner, S., Haas, J., Kehm, R., Sellner, J., Hacke, W., et al. (2004). Increased presence of matrix metalloproteinases 2 and 9 in short- and long-term experimental herpes simplex virus encephalitis. *Neurosci. Lett.* 368, 274–278. doi: 10.1016/j.neulet.2004.06.064

Menendez, C. M., Jinkins, J. K., and Carr, D. J. (2016). Resident T cells are unable to control herpes simplex virus-1 activity in the brain ependymal region during latency. *J. Immunol.* 197, 1262–1275. doi: 10.4049/jimmunol.1600207

Mowry, E. M., James, J. A., Krupp, L. B., and Waubant, E. (2011). Vitamin D status and antibody levels to common viruses in pediatric-onset multiple sclerosis. *Mult. Scler.* 17, 666–671. doi: 10.1177/1352458510394398

Nicoll, J. A., Kinrade, E., and Love, S. (1992). PCR-mediated search for herpes simplex virus DNA in sections of brain from patients with multiple sclerosis and other neurological disorders. *J. Neurol. Sci.* 113, 144–151. doi: 10.1016/0022-510x(92)90242-d

Nourbakhsh, B., Rutatangwa, A., Waltz, M., Rensel, M., Moodley, M., Graves, J., et al. (2018). Heterogeneity in association of remote herpesvirus infections and pediatric MS. *Ann. Clin. Transl. Neurol.* 5, 1222–1228. doi: 10.1002/acn3.636

Reich, D. S., Lucchinetti, C. F., and Calabresi, P. A. (2018). Multiple sclerosis. *N. Engl. J. Med.* 378, 169–180. doi: 10.1056/NEJMra1401483

Rissoan, M. C., Soumelis, V., Kadowaki, N., Grouard, G., Briere, F., de Waal Malefyt, R., et al. (1999). Reciprocal control of T helper cell and dendritic cell differentiation. *Science* 283, 1183–1186. doi: 10.1126/science.283.5405.1183

Sanders, V. J., Waddell, A. E., Felisan, S. L., Li, X., Conrad, A. J., and Tourtellotte, W. W. (1996). Herpes simplex virus in postmortem multiple sclerosis brain tissue. *Arch. Neurol.* 53, 125–133. doi: 10.1001/archneur.1996.00550020029012

Sanna, A., Huang, Y. M., Arru, G., Fois, M. L., Link, H., Rosati, G., et al. (2008). Multiple sclerosis: reduced proportion of circulating plasmacytoid dendritic cells expressing BDCA-2 and BDCA-4 and reduced production of IL-6 and IL-10 in response to herpes simplex virus type 1. *Mult. Scler.* 14, 1199–1207. doi: 10.1177/1352458508094401

Scepanovic, P., Alanio, C., Hammer, C., Hodel, F., Bergstedt, J., Patin, E., et al. (2018). Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med.* 10:59. doi: 10.1186/s13073-018-0568-8

Smith, G. D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22. doi: 10.1093/ije/dyg070

Sotelo, J., Ordoñez, G., Pineda, B., and Flores, J. (2014). The participation of varicella zoster virus in relapses of multiple sclerosis. *Clin. Neurol. Neurosurg.* 119, 44–48. doi: 10.1016/j.clineuro.2013.12.020

Soumelis, V., and Liu, Y. J. (2006). From plasmacytoid to dendritic cell: morphological and functional switches during plasmacytoid pre-dendritic cell differentiation. *Eur. J. Immunol.* 36, 2286–2292. doi: 10.1002/eji.200636026

Theil, D., Derfuss, T., Paripovic, I., Herberger, S., Meinl, E., Schueler, O., et al. (2003). Latent herpesvirus infection in human trigeminal ganglia causes chronic immune response. *Am. J. Pathol.* 163, 2179–2184. doi: 10.1016/s0002-9440(10)63575-4

Verbanck, M., Chen, C. Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698. doi: 10.1038/s41588-018-0099-7

Wandinger, K., Jabs, W., Siekhaus, A., Bubel, S., Trillenberg, P., Wagner, H., et al. (2000). Association between clinical disease activity and Epstein-Barr virus reactivation in MS. *Neurology* 55, 178–184. doi: 10.1212/wnl.55.2.178

Waubant, E., Mowry, E. M., Krupp, L., Chitnis, T., Yeh, E. A., Kuntz, N., et al. (2011). Common viruses associated with lower pediatric multiple sclerosis risk. *Neurology* 76, 1989–1995. doi: 10.1212/WNL.0b013e31821e552a

Xu, L., Zhang, L. J., Yang, L., Yang, C. S., Yi, M., Zhang, S. N., et al. (2021). Positive association of herpes simplex virus-IgG with multiple sclerosis: a systematic review and meta-analysis. *Mult. Scler. Relat. Disord.* 47:102633. doi: 10.1016/j.msard.2020.102633

Zhang, Y., Qu, J., Luo, L., Xu, Z., and Zou, X. (2021). Multigenomics reveals the causal effect of herpes simplex virus in Alzheimer's disease: a two-sample Mendelian randomization study. *Front. Genet.* 12:773725. doi: 10.3389/fgene.2021.773725

Zhuang, H., Zhang, Y., Yang, S., Cheng, L., and Liu, S. L. (2019). A Mendelian randomization study on infant length and type 2 diabetes mellitus risk. *Curr. Gene Ther.* 19, 224–231. doi: 10.2174/1566523219666190925115535

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 11 March 2022
doi: 10.3389/fnins.2022.828382

Integrated Metabolomics and Proteomics Analysis of Urine in a Mouse Model of Posttraumatic Stress Disorder

Daxue Zhou[1], Chengyan Long[2], Yan Shao[1], Fei Li[1], Wei Sun[1], Zihan Zheng[1], Xiaoyang Wang[1], Yiwei Huang[1], Feng Pan[1], Gang Chen[1,3,4], Yanlei Guo[2]* and Yi Huang[1]**

[1] Biomedical Analysis Center, Army Medical University, Chongqing, China, [2] Chongqing Academy of Chinese Materia Medica, Chongqing, China, [3] Key Laboratory of Extreme Environmental Medicine, Ministry of Education of China, Chongqing, China, [4] Chongqing Key Laboratory of Cytomics, Chongqing, China

Posttraumatic stress disorder (PTSD) is a serious stress disorder that occurs in individuals who have experienced major traumatic events. The underlying pathological mechanisms of PTSD are complex, and the related predisposing factors are still not fully understood. In this study, label-free quantitative proteomics and untargeted metabolomics were used to comprehensively characterize changes in a PTSD mice model. Differential expression analysis showed that 12 metabolites and 27 proteins were significantly differentially expressed between the two groups. Bioinformatics analysis revealed that the differentiated proteins were mostly enriched in: small molecule binding, transporter activity, extracellular region, extracellular space, endopeptidase activity, zymogen activation, hydrolase activity, proteolysis, peptidase activity, sodium channel regulator activity. The differentially expressed metabolites were mainly enriched in Pyrimidine metabolism, D-Glutamine and D-glutamate metabolism, Alanine, aspartate and glutamate metabolism, Arginine biosynthesis, Glutathione metabolism, Arginine, and proline metabolism. These results expand the existing understanding of the molecular basis of the pathogenesis and progression of PTSD, and also suggest a new direction for potential therapeutic targets of PTSD. Therefore, the combination of urine proteomics and metabolomics explores a new approach for the study of the underlying pathological mechanisms of PTSD.

Keywords: posttraumatic stress disorder (PTSD), PTSD model, urine, metabolomics, proteomics, mice

OPEN ACCESS

Edited by:
Zhijie Han,
Chongqing Medical University, China

Reviewed by:
Shayne Mason,
North-West University, South Africa
Yubo Li,
Tianjin University of Traditional Chinese Medicine, China

****Correspondence:***
Gang Chen
cg5237@tmmu.edu.cn
Yanlei Guo
guoyanlei1210@cqacmm.com
Yi Huang
huangyi@tmmu.edu.cn

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: *03 December 2021*
Accepted: *23 February 2022*
Published: *11 March 2022*

Citation:
Zhou D, Long C, Shao Y, Li F, Sun W, Zheng Z, Wang X, Huang Y, Pan F, Chen G, Guo Y and Huang Y (2022) Integrated Metabolomics and Proteomics Analysis of Urine in a Mouse Model of Posttraumatic Stress Disorder. Front. Neurosci. 16:828382. doi: 10.3389/fnins.2022.828382

INTRODUCTION

Post-Traumatic Stress Disorder (PTSD) is a persistent stress disorder type that may be delayed or imminent following major psychological trauma (Kessler et al., 1995; Breslau et al., 1998). PTSD can be caused by a variety of major events, including diseases (Kangas et al., 2005; Bush, 2010), war incidents (Owens et al., 2005), natural disasters (Wu et al., 2009), etc. PTSD has four core symptoms according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5); the re-experiencing of traumatic event(s), continuous avoidance of trauma-related stimuli, negative emotions related to cognitive trauma, and continued increase in alertness (Mahan and Ressler, 2012; Tandon, 2014; Tanaka et al., 2019). Several of these aspects can be captured using situational reminder programming in animal models, leading it to become a common model for studying the

symptoms and mechanisms of PTSD. However, the precise molecular changes occurring in these models remains incompletely understood.

Assessments for compositional changes in urine have been demonstrated to have considerable potential for monitoring bodily health (Gao, 2013; Nicholas, 2020). In comparison with blood, urine has the advantages of being non-invasive, convenient to sample repeatedly, biochemical stability, and so on (Wu and Gao, 2015; Jing and Gao, 2018). In addition, urine may not be as strongly regulated by homeostatic mechanisms (Wang et al., 2014; Huang and Lo, 2018). The detection of blood biomarkers usually reflects the relatively stable state in the middle and late stages of the disease (Li, 2015), but misses the signals of short-term changes in the early stage of the disease. In opposite urine, as a blood filter, will collect all the body's metabolites, thereby detecting more differentiated factors (Li et al., 2014; Gao, 2015). Moreover, recent reports have shown that urine can provide a lot of non-urogenital information, including regarding neuropsychiatric disorders (Emanuele et al., 2010; Marc et al., 2011).

In the past few years, advances in "omics" technology have yielded powerful new tools for biomarker screening, disease mechanism identification, and diagnostic modeling (Petricoin et al., 2006). Cutting-edge "omics" technology has already been deployed to study PTSD. Diverse epigenetic phenomena have enabled researchers to discover conserved molecular mechanisms involved in chromatin modification (Goldberg et al., 2007), especially non-coding RNAs, which play an important role in multiple epigenetic phenomena (Bernstein and Allis, 2005). Studies on PTSD and miRNA have revealed several key contributors to the underlying pathophysiological basis of PTSD (Wingo et al., 2015; Bam et al., 2016a,b; Martin et al., 2017). Genomics research can be used to analyze DNA and RNA sequences by second-generation sequencing and third-generation sequencing techniques to discover new transcripts or exon single nucleotide polymorphisms (SNPS) (Girgenti and Duman, 2018). However, researchers thus far have primarily used blood and postmortem brain tissue to identify biomarkers for PTSD (Thomson et al., 2014; Breen et al., 2015; Bharadwaj et al., 2018). Glycomics studies can analyze the biological functions of all glycans by studying the unique pond group of organisms (Miura and Endo, 2016). Compared with genome sequence discovery, glycomics can better reflect the biological state of complex diseases (Zoldos et al., 2013; Lauc et al., 2016). It has been reported that there are significant changes in the N-glycomic group in psychiatric and neurodegenerative diseases (Vanhooren et al., 2010; Lundstrom et al., 2014; Park et al., 2018). Proteomics studies the complete set of proteins in a biological system (cell, tissue or organism) in a given state at a given time, analyzing changes in protein expression, post-translational modifications, and protein-protein interactions (Wilkins et al., 1996; Jensen, 2006). It is more complex than genomics, but can reflect the precise functional characteristics of proteins (Baloyianni and Tsangaris, 2009). Metabolomics mainly analyses final or intermediate small molecule metabolites produced by gene regulation and can evaluate metabolites altered by treatment or disease (Kaddurah-Daouk and Krishnan, 2009; Mastrangelo et al., 2016). It is reported that metabolomics plays an important role in analyzing the metabolic profile, inflammatory mechanisms and biomarker identification of PTSD (De Bellis et al., 2000; Pace and Heim, 2011; Karabatsiakis et al., 2015; Bersani et al., 2016; de Vries et al., 2016; Hemmings et al., 2017; Mellon et al., 2018; Nedic Erjavec et al., 2018). Therefore, Omics technologies can be further improved study the underlined mechanisms of PTSD and identify diagnostic and prognostic biosignatures.

Despite the promising features of urine biomarkers, the biological interpretation of single typology data is very challenging due to the complexity of urine samples. Therefore, in this study, the analytical capabilities of proteomics and metabolomics were combined to obtain more comprehensive data on mice in the normal group and PTSD group, aiming to discover new potential biomarkers.

MATERIALS AND METHODS

Induction of Electric Foot Shock Stress

Twelve healthy male 8–10 weeks old C57BL/6 mice were purchased from Laboratory Animal Centre at the Army Medical University. All mice were housed in individual cages under a reversed 12 h light/12 h dark cycle (light on at 6 AM) and standard laboratory conditions (21 ± 1°C, 55 ± 5% relative humidity). Food and water were provided *ad libitum*. This study was approved by the Ethics Committee of Army Medical University (Animal Ethics Statement: AMUWEC20211605). As shown in **Figure 1**, after a 14-day adaptation phase, the mice were divided into plantar foot shock group (PTSD group, $n = 6$) and non-foot shock group (control group, $n = 6$). Mice were subjected to electric foot shock in a Plexiglas chamber (27 × 20 × 300 cm3) with a grid floor made of stainless-steel rods (0.3 cm diameter, spaced 1.0 cm apart) connected to a shock generator. After a habituation period of 2 min, the mice in the foot shock group received a series of foot shocks of medium (0.15 mA) intensity of 10 s duration with foot shock interval of 10 s being delivered for 5 min to produce acute stress (Rabasa et al., 2011). The mice in the control group were placed in the chamber for a similar period without receiving a foot shock. Thereafter, the mice in the foot shock group were subjected to the same moderate electric foot shock stressor for 12 days (twice a day) to induce stress adaptation (Van den Berg et al., 1998; Daniels et al., 2008).

Behavioral Test

All the behavioral tests were performed in daytime from 8 AM to 3 PM. Mice were given two tests a day to avoid the potential interference from the other tests. Animal cages were moved to a testing room at least 0.5 h before each test. After completion of the test session, the behavioral apparatus and chamber were cleaned with 70% ethanol and then completely hand-fan dried.

Elevated Plus Maze Test

The apparatus consisted of four arms (28 cm × 5.8 cm width), with two arms open and two closed by gray walls (15.5 cm height) arranged on the opposite side of the same type. The platform was

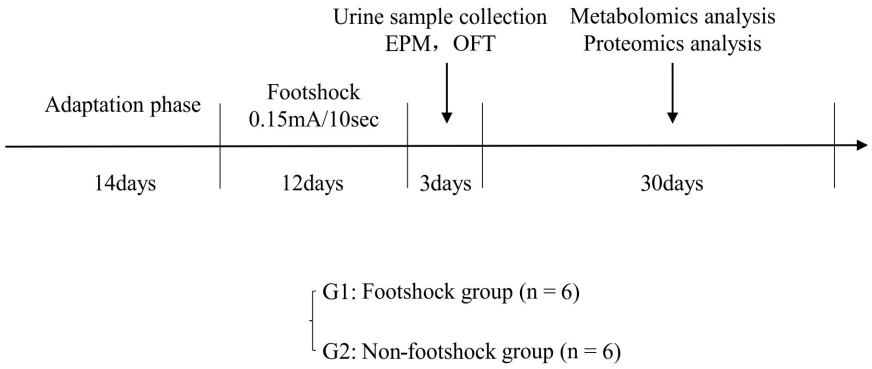

FIGURE 1 | The experimental procedure for PTSD. EPM, elevated plus maze; OFT, open field test.

located 55 cm above the floor of the testing room illuminated and four arms were connected in the center platform (5 cm × 5 cm), where the animal was placed facing a closed arm. The position and movement of the animal were monitored for 5 min by a video camera. An entry was defined as more than half of the animal's full body entering the open arm. The time spent in the open arms and the number of visits to the open arms were analyzed.

Open Field Test

The periphery and bottom of the test space were made of black opaque metal sheets, with length 72 cm, width 72 cm, and height 60 cm. The floor area was divided into 16 squares of the same size. During the test, each mouse was placed in the center of the area and was allowed to freely explore the area for 5 min. The numbers of crossing and standing were recorded during the last 4 min by technicians. After each test, the open area was washed with 70% ethanol to avoid any olfactory cues.

Urine Sample Collection

Urine was collected on ice using metabolic cages at the end of the experiment from 9:00 pm to 9:00 am in the next day. The collected urine was centrifuged at 13,000 *g* for 20 min at 4°C to obtain the supernatant sample. The average sample size was 2 mL. The urine sample was stored at −80°C before analysis.

Metabolomics Analysis

Urine samples (50 μL) were thawed on ice and immediately mixed with 200 μL of ice-cold acetonitrile. After mixing by vortex for 1 min, the mixture was centrifuged at 13,000 *g* for 15 min at 4°C. A supernatant aliquot of 10 μL was used for liquid chromatography-mass spectrometry (LC-MS/MS) analysis. Quality control (QC) samples were prepared by supernatant aliquot with an equal amount (15 μL) and were periodically analyzed throughout the complete run to monitor signal drift.

The LC 30A UHPLC system (Shimadzu, Kyoto, Japan) was linked to a Triple TOF 4600 system (SCIEX, Framingham, MA, United States). The separation step was conducted using the hydrophilic interaction liquid chromatography (HILIC) and the reversed-phase liquid chromatography (RPLC) methods. A Kinetex C_{18} column (2.1 mm × 100 mm, 2.6 μm, 100 Å, Phenomenex) was used with a binary gradient method. Solvent A was 0.1% formic acid in water (vol/vol), and solvent B was 0.1% formic acid in acetonitrile (vol/vol). A flow rate of 0.35 mL/min was used, and the injection volume was 2 μL. The gradient program used was 15% B at 0 min to 85% B at 10 min, with a total running time of 15 min. A TSK gel NH_2-100 column (2.1 mm × 100 mm, 3.0 μm, TOSOH) was also used with a binary gradient method. Solvent A was 5 mmol/L ammonium acetate, and solvent B was acetonitrile. A flow rate of 0.25 mL/min was used, and the injection volume was 2 μL. The gradient program used was 100% B at 2 min to 15% B at 15 min, and at 20 min to 100% B, with a total running time of 25 min.

Proteomics Analysis

1 mL urine sample was thawed and transferred to a centrifuge tube, and then centrifuged at 12,000 *g* at 4°C for 30 min to remove impurities. The samples were six times mixed with the volume of acetone, fully mixed, and precipitated overnight at – 20°C. The mixture was removed and centrifuged at 12,000 *g* at 4°C for 30 min to remove the supernatant. The precipitate was dissolved in pyrolysis buffer solution (8 mol/L urea, 2 mol/L Thiourea, 50 mmol/L Tris, and 25 mmol/L DTT), and completely dissolved, centrifuged at 12,000 *g* for 30 min at 4°C, and then the supernatant was saved. Protein concentration was determined using the Bradford method. 100 ug protein was added to each sample in a 30 KDa filter (millipore, MRCF0R030), Urea buffer solution (UA, 8 mol/L, 0.1 mol/L Tris–HCl, pH 8.5), and 25 mmol/L NH_4HCO_3 solutions were in turn washed several times. Protein samples were reduced with 20 mmol/L dithiothreitol (DTT, Sigma) at 37°C for 1 h, followed by 50 mmol/L iodoacetamide (IAA, Sigma) in darkness for 30 min. Then, the samples were centrifuged at 18°C for 30 min at 14,000 *g*, washed with UA and NH_4HCO_3, with trypsin being added (enzyme protein ratio 1:50) and digested overnight at 37°C. The peptide mixture was desalted using a C_{18} column (Thermo, 84850), concentrated, dried in vacuum, and stored at −80°C.

AU3000 UHPLC system (Thermo Fisher Scientific, Waltham, MA, United States) was used to separate the peptides. Peptides

were loaded onto an analytical column (Acclaim™ PepMap™ 100, 75 μm × 15 cm, C_{18}, 3 μm, 100 Å, Thermo Fisher Scientific, Waltham, MA, United States) with a Trap Column (Acclaim™ PepMap™ 100, 75 μm × 2 cm, C18, 3 μm, 100 Å, Thermo Fisher Scientific, Waltham, MA, United States) and separated by reversed-phase chromatography (U3000nano, Thermo Fisher Scientific, Waltham, MA, United States) using a 106 min gradient. The gradient was composed of Solvent A (0.1% formic acid in water) and Solvent B (0.1% formic acid in 80% acetonitrile) elution gradient: 1% B for 13 min, 1–30% B in 70 min, 30–90% B in 10 min, 90% B for 2 min and 90–1% B in 1 min, 1% B for 10 min. The eluted peptides were analyzed using the Data Dependent Acquisition (DDA) method applying one full MS scan (350.00–1800.00 m/z) in the Orbitrap at a resolution of 60,000 M/ΔM, followed by consecutive MS/MS (profile) scans in the ion trap by product ion scans (relative CID energy 35) of the 16 most abundant ions in each survey scan. The product ion scans were acquired with a 2.0 unit isolation width and a normalized collision energy of 35 in an LTQ-Orbitrap Velos Pro MS spectrometer (Thermo Fisher Scientific, Waltham, MA, United States).

Statistical Analysis

Statistical analyses were performed using SPSS 20.0 software, values are expressed as mean ± standard deviation (X ± SD, $n = 6$ per group), the graphics were generated using GraphPad Prism 8.0.1 software. Metabolomics and Proteomics data analyses were performed in MetaboAnalyst 5.0 and Proteome Discover Daemon 2.5. The metabolite peaks of the urine samples were normalized, analyses performed using SIMCA-P 14.1 multivariate statistical analysis software. All variables were tested and found to be normally distributed, an independent-samples student's *t*-test was used to compare differences between the two groups, and identify differentially expressed metabolites and proteins, and then we used Ingenuity Pathway Analysis to analyze the significantly altered canonical pathways and molecular interaction networks. A *p*-value threshold of 0.05 was used to infer statistically significant findings, and a more strict *p*-value threshold of 0.01 was used to infer highly statistically significant changes.

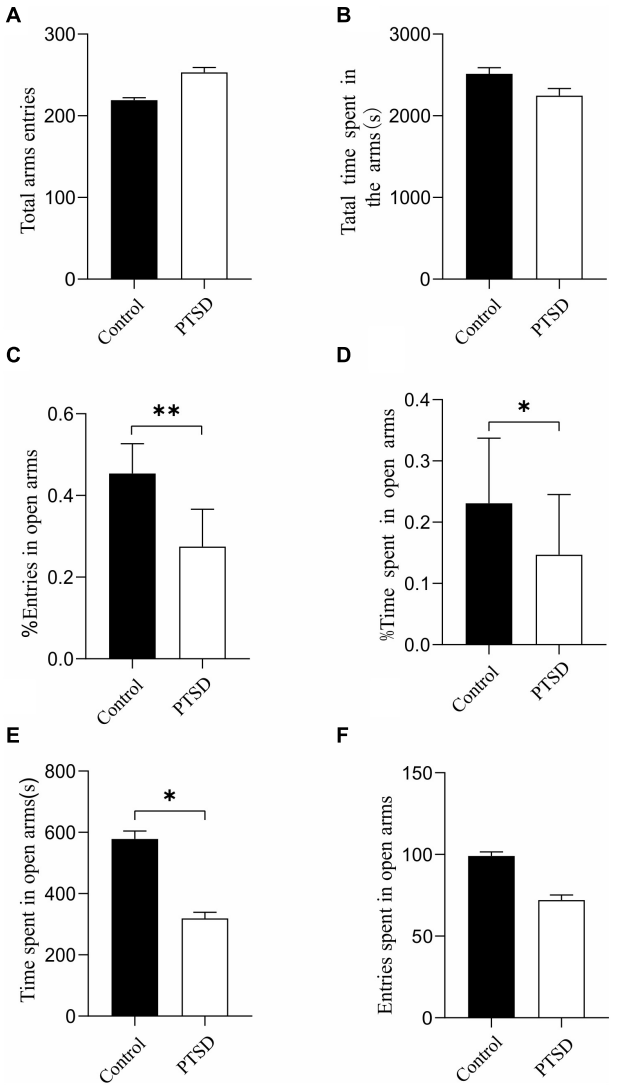

FIGURE 2 | The effect of the induction of electric foot shock on anxiety in the elevated plus-maze test in mice compared to the control group ($n = 6$). **(A)** Total arms entries; **(B)** Total time spent in the arms; **(C)** The percent of entries in open arms; **(D)** The percent of time spent in open arms; **(E)** The actual times spent in open arms; **(F)** The actual amount of entries in open arms. Results are provided in the form of mean ± SD. *$p < 0.05$; **$p < 0.01$.

RESULTS

The Results of the Behavioral Test

Elevated Plus Maze Test

Elevated plus maze test was deployed to explore the potential anxiety actions of the induction of electric foot shock stress. There was no significant difference in terms of total arms entries (**Figure 2A**) and total time spent in the arms (**Figure 2B**) between the control and the PTSD groups.

The induction of the electric foot shock stress caused a significant reduction in the percent of open arm entries (open arm/total × 100) with the ones of the control and the PTSD group being 45.4 and 27.5%, respectively (**Figure 2C**). The percent of time spent in open arms was also significantly reduced when applying induction of electric foot shock stress with control and PTSD group values being 23.1 and 14.7%, respectively (**Figure 2D**). The actual times spent in open arms in the control group and the PTSD group were 578 s and 319 s, respectively (**Figure 2E**). The actual amount of entries in open arms in the control group and the PTSD group were 99 and 72, respectively (**Figure 2F**).

Open Field Test

The anxiety-like behavior of the induction of electric foot shock stress was measured with an open-field test. The overall distance was significantly reduced when applying induction of electric foot shock stress with the values of the control and PTSD groups being 1511.18 and 1292.94 cm, respectively (**Figure 3A**). The number

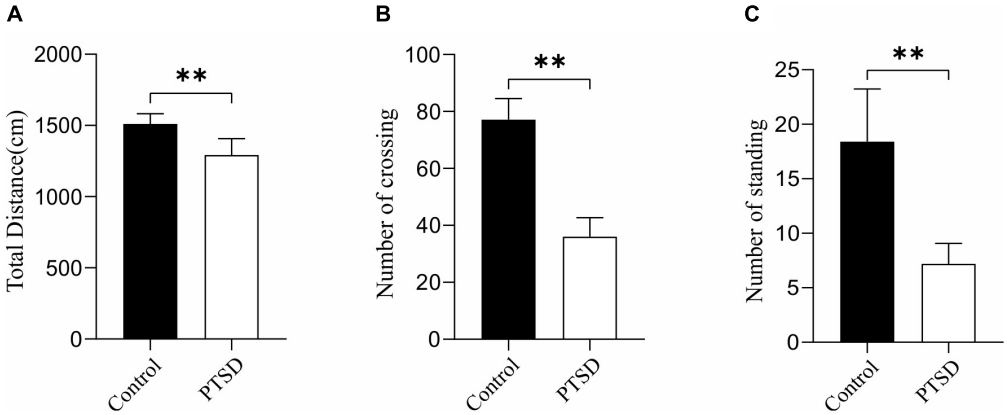

FIGURE 3 | The effect of the induction of electric foot shock on anxiety in the open-field test in mice compared to the control group (n = 6). **(A)** Overall distance; **(B)** Numbers of crossing; **(C)** Numbers of standing. Data are expressed as mean ± SD. $^{**}p < 0.01$.

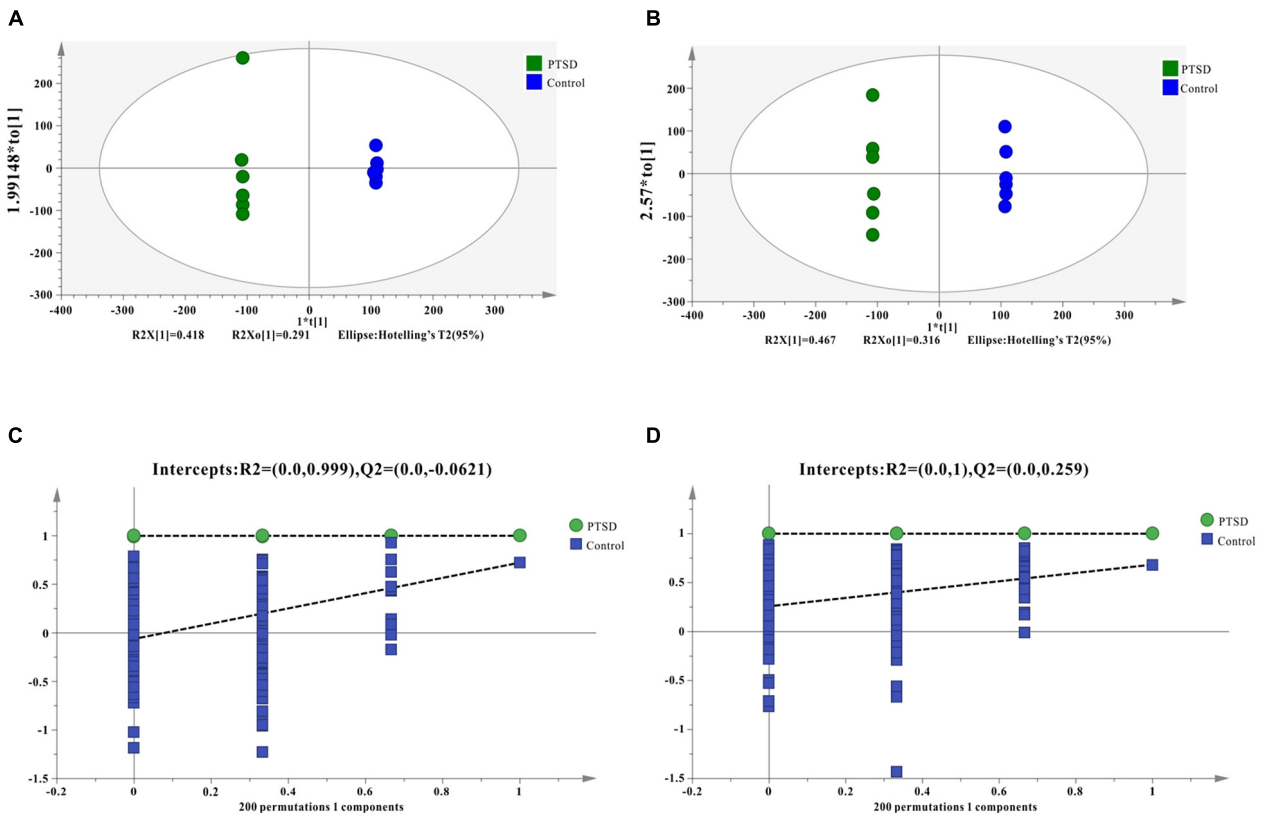

FIGURE 4 | OPLS-DA score results and OPLS-DA quality control figure of mice urine samples. **(A)** Positive ion mode OPLS-DA scores; **(B)** Negative ion mode OPLS-DA scores; **(C)** Positive ion mode OPLS-DA permutation test; **(D)** Negative ion mode OPLS-DA permutation test; intercepts: R2 and Q2 represent y-intercept of R2 and Q2 regression lines.

of crossing and standing was also significantly reduced by the Induction of electric foot shock stress. The numbers of crossing in the control group and the PTSD group were 77.1 and 36.0 times, respectively (**Figure 3B**). The amount of times of standing in the control group and the PTSD group were 18.4 and 7.2 times, respectively (**Figure 3C**).

Metabolomics Analysis

Quality control results pinpointed that the variation caused by instrument error is small and the data quality is reliable (**Supplementary Figure 1**). The PCA results plot does not show clear segregation between the PTSD group and the control group (**Supplementary Figure 2**). In addition, The OPLS-DA

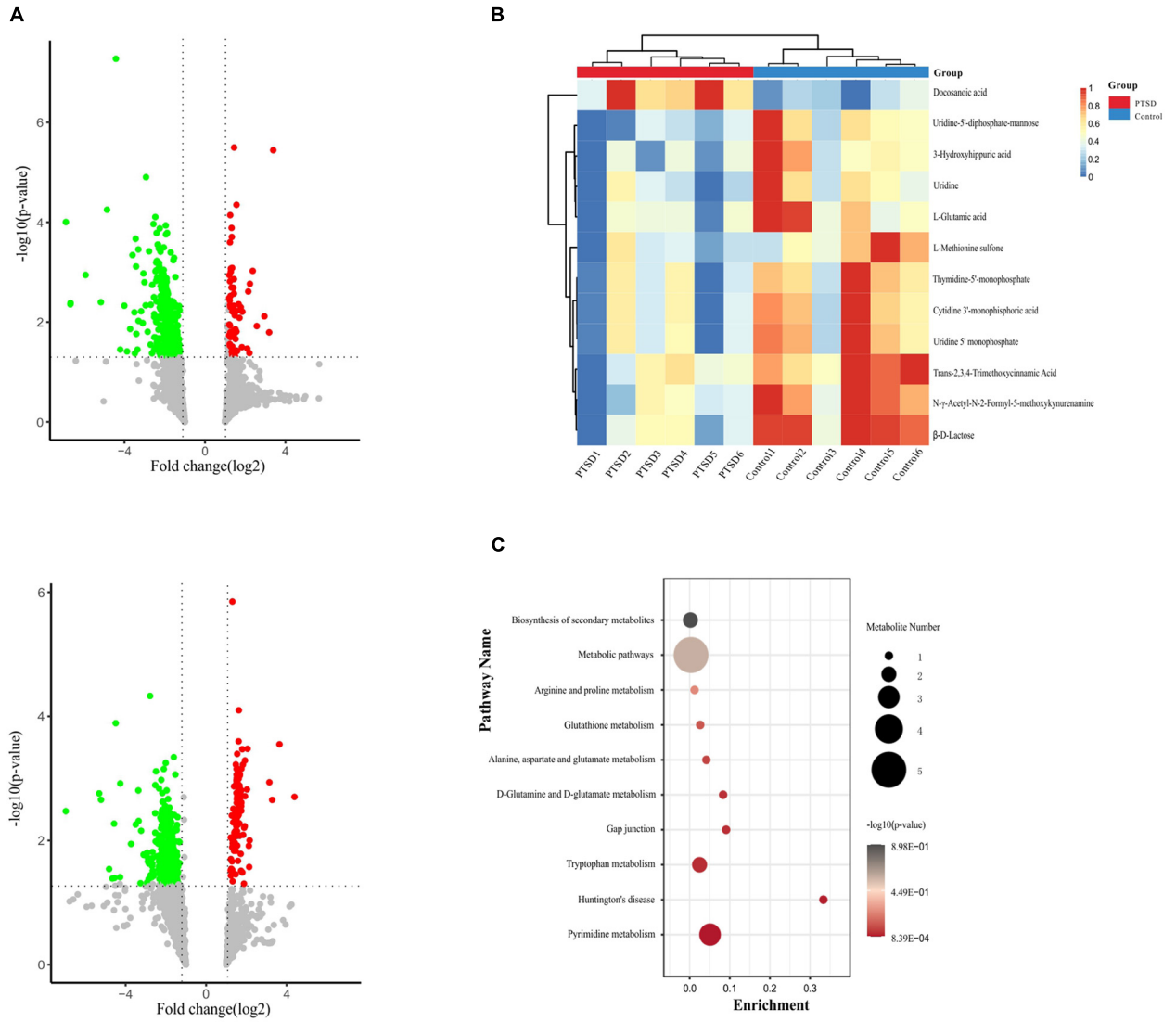

FIGURE 5 | Metabolites profile shift during of mice urine samples between the PTSD group and the control group. **(A)** Volcano plot showing altered metabolites, Positive (up), negative (down) ion mode. The red dots indicate significantly upregulated metabolites (fold change > 1.2), while the green dots indicate significantly downregulated metabolites (fold change < 0.8). **(B)** Heat map of the differentially expressed metabolites. The red band indicates the upregulation of metabolites levels (fold change > 1.2), while the blue band indicates the downregulation of metabolites levels (fold change < 0.8). **(C)** KEGG pathway enrichment category of the differentially expressed metabolites, $n = 6$.

model was established and a permutation test of the OPLS-DA model was performed (positive mode: R2X = 0.919, R2Y = 1.0, Q2 = 0.723; negative mode: R2X = 0.854, R2Y = 0.998, Q2 = 0.657; **Figures 4A,B**). The results of the permutation test showed the absence of overfitting (positive mode: R2 = 0.999, Q2 = −0.0431; negative mode: R2 = 0.995, Q2 = −0.0933; **Figures 4C,D**). In conclusion, the model presented good reliability and predictability.

Differential Metabolites

Metabolomics profiling of urine from the C57BL/6 normal group mice and the PTSD mouse-model group detected a total of 559 metabolite components and revealed 12 differentially expressed metabolites between the PTSD group and the Control group using as criteria to infer significant findings the VIP > 1 and $p < 0.05$ (**Figure 5A** and **Table 1**). These differentially expressed metabolites possess different characteristics (**Figure 5B**) and were enriched for several KEGG pathways associated with amino acid and nucleic acid metabolism, including Pyrimidine metabolism, D-Glutamine and D-glutamate metabolism, Alanine, aspartate and glutamate metabolism, Arginine biosynthesis, Glutathione metabolism, Arginine, and proline metabolism (**Figure 5C**).

Proteomics Analysis

The number of peptide-spectral matches, unique peptide number, and quantified proteins, were 88,734, 4,125, and 691

TABLE 1 | The significantly differentiated metabolites in control vs. post-traumatic stress disorder (PTSD) groups.

No	Metabolites	VIP	FC	*P*-value	Trend
1	L-Methionine sulfone	1.27643	0.39592211	0.00128436	↓
2	Docosanoic acid	1.11139	1.24292539	0.00153343	↑
3	Uridine	2.21959	0.54969523	0.00295338	↓
4	3-Hydroxyhippuric acid	1.10166	0.58905271	0.00886306	↓
5	L-Glutamic acid	1.04933	0.42636368	0.00901154	↓
6	Uridine 5′ monophosphate(UMP)	2.42806	0.47128583	0.02161141	↓
7	*Trans*-2,3,4-Trimethoxycinnamic Acid	1.05031	0.46980814	0.02427128	↓
8	β-D-Lactose	1.65882	0.51933915	0.02631537	↓
9	Cytidine 3′-monophisphoric acid	3.48481	0.49014935	0.02692515	↓
10	N-γ-Acetyl-N-2-Formyl-5-methoxykynurenamine	1.5343	0.52282757	0.02905089	↓
11	Thymidine-5′-monophosphate(dTMP)	7.26583	0.50788472	0.03078474	↓
12	Uridine-5′-diphosphate-mannose(UDP-Gal)	1.16583	0.49116102	0.03410945	↓

for both PTSD and control groups. 27 proteins exhibited significantly differentiated expression between the two groups using the criteria of p-value < 0.05 and fold change > 1.20 or <0.80. A total of 18 proteins among these were upregulated and 9 downregulated in the PTSD group compared to the Control group (**Figure 6A** and **Table 2**). These altered features were subjected to clustering, and the heat map revealed clusters with the ability to discriminate between control and PTSD samples (**Figure 6B**). Gene Ontology (GO) function annotation analysis showed that these differentially expressed proteins were mainly involved in biological processes, such as small molecule binding, transporter activity, extracellular region, extracellular space, endopeptidase activity, zymogen activation, hydrolase activity, proteolysis, peptidase activity and sodium channel regulator activity (**Figure 6C**). Based on the KEGG database, the significantly enriched pathways ($P < 0.05$) were Endocrine and other factor-regulated calcium reabsorption, Lysosome, Renin-angiotensin system, Carbohydrate digestion and absorption, Thyroid hormone synthesis, Metabolic pathways, Proximal tubule bicarbonate reclamation, Galactose metabolism and Starch and sucrose metabolism (**Figure 6D**).

Integrative Analysis of the Metabolomics and Proteomics

A total of 12 differential expression metabolites and 27 differential expression proteins that were submitted to Ingenuity Pathway Analysis (IPA) for significantly altered canonical pathways analysis. As shown in **Table 3**, We found three pathways significantly expressed proteins and metabolites. They were Pyrimidine Metabolism, Metabolic pathways, and Small Molecule Biochemistry. These significantly differential metabolic pathways were selected for more detailed analysis (**Figure 7**). In these pathways, L-Glutamic acid(L-Glu), Uridine5-monophosphate(UMP), Thymidine 5-monophosphate(dTMP), Uridine, URIDINE-5′-DIPHOSPHATE-MANNOSE(UDP-Gal), CTSH and CTS6 were downregulated, and UMOD, Fxyd2, AHCY, ACY3, Hamp2, CTSE, SCLT1, WFDC2 were upregulated.

DISCUSSION

Currently, diagnosis of PTSD is primarily based on subjective symptom representation and patient self-reporting, and the molecular mechanism remains unclear. As such, rates of PTSD in the general population may be significantly underestimated. In the present manuscript, we established a mouse model of PTSD to investigate some of its qualitative biomarkers and potential mechanistic contributors. The elevated cross maze and open-field test were evaluated based on the fact that the plantar shock can continuously produce traumatic stimulation. The mice with the PTSD group were observed to have reduced movement, weakened active exploration ability, and showed negative avoidance and anxiety in comparison to the mice of the control group (Montgomery, 1955; Pellow et al., 1985), indicating significant stress disorder. Combined proteomics and metabolomics analysis was performed revealing 27 significantly dysregulated proteins and 12 significantly dysregulated metabolites.

In this study, urinary uridine levels in the mice model were significantly reduced, suggesting that PTSD can cause metabolic abnormalities of uridine in urine. It has been reported that uridine has a protective effect on mental disorders (Mironova et al., 2018) and can improve neurophysiological functions (Connolly and Duley, 1999). Uridine excretion is mainly achieved through renal and pyrimidine metabolism, producing uracil and β -alanine, which can enter the tricarboxylic acid (TCA) cycle (Gonzalez and Fernandez-Salguero, 1995; Connolly et al., 1996). The homeostasis and metabolic abnormalities of uridine can be accurately monitored by the detection of uridine in urine.

In mammals, in pyrimidine metabolism, uridine (UR) is involved in the de initio synthesis of uridine monophosphate (UMP) to form uridine 5′ -diphosphate (UDP), which can be combined with UDP-galactose and plays an important role in the glycosylation of protein (Connolly and Duley, 1999). It has been reported that pyrimidines are mainly recovered from uridine, which synthesizes RNA and biofilms through pyrimidine nucleotide – lipid conjugates (Yamamoto et al., 2011). In the study, the urine metabolism of uridine (UR), uridine monophosphate (UMP), and UDP-galactose in mice of the PTSD

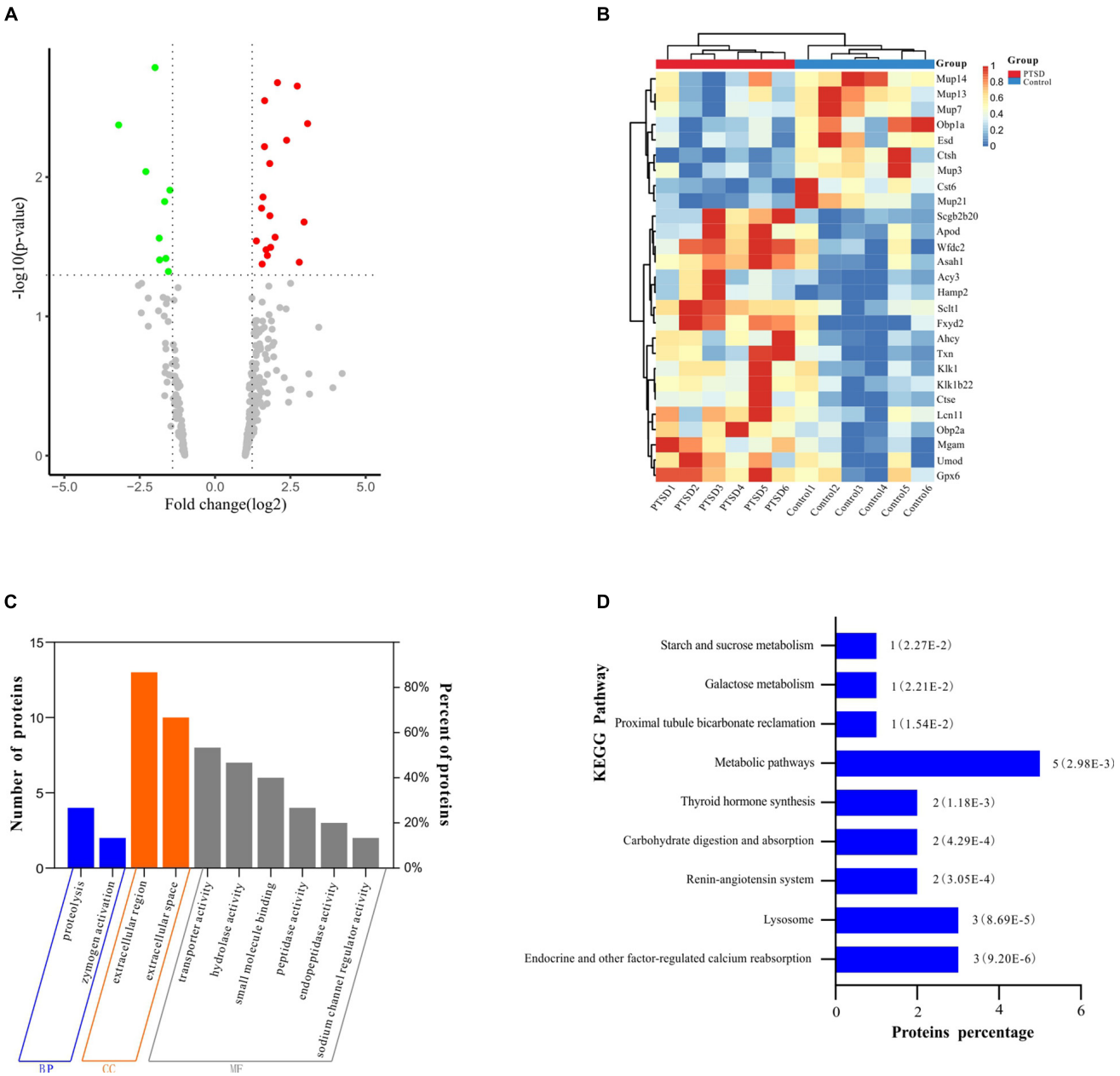

FIGURE 6 | Proteomic profile shift of mice urine samples between the PTSD group and the control group. **(A)** Volcano plot showing dysregulated proteins. The red dots indicate significantly upregulated proteins (fold change > 1.2), while the green dots indicate significantly downregulated proteins (fold change < 0.8); **(B)** Heat map of the differentially expressed proteins. The red band indicates the upregulated proteins (fold change > 1.2), while the blue band indicates the downregulated proteins (fold change < 0.8); **(C)** Differentially expressed protein GO function enrichment diagram; **(D)** KEGG pathway enrichment of the differentially expressed proteins, $n = 6$.

model group showed decreased expression, suggesting that the PTSD mice induced by plantar electric shock exhibit disorder of pyrimidine metabolism.

Glutamate (L-glutamate) is a major excitatory neurotransmitter, and glutamate disorder in the brain is often observed in depression models (Hemanth Kumar et al., 2012; Liu et al., 2016). In this study, the expression of L-glutamate in urine metabolism was decreased in the PTSD model group, while glutamate can provide a nitrogen source for pyrimidine synthesis (Vincenzetti et al., 2016). The pyrimidine metabolism disorder can directly reflect abnormal glutamate metabolism. Studies have reported that patients with PTSD and alcohol use disorder (AUD) have significantly reduced glutamate in the anterior cingulate cortex (ACC; Pennington et al., 2014). Glutamate is the basis of synaptic plasticity and memory formation, and stress response significantly affects glutamate transmission and plays a

TABLE 2 | The differentially expressed proteins in control vs. PTSD groups.

No	UniProt accession	Gene symbol	Protein name	FC	*P*-value	Trend
1	P49935	Ctsh	Pro-cathepsin H	−1.99023949	0.00163517	↓
2	G5E861	Sclt1	Sodium channel and clathrin linker 1	2.07219424	0.00209911	↑
3	A0A1L1SQP8	Fxyd2	FXYD domain-containing ion transport regulator	2.72899236	0.00222118	↑
4	Q3TWT5	Asah1	Ceramidase	1.64659341	0.00282754	↑
5	J3QK77	Scgb2b20	ABPBG20	3.07210788	0.00413095	↑
6	Q9D1B1	Cst6	Cystatin E/M	−3.19900397	0.00422517	↓
7	A0A571BF69	Mgam	Maltase-glucoamylase	2.37524214	0.00543735	↑
8	Q4FZJ6	Wfdc2	WAP four-disulfide core domain 2	1.64128707	0.00604615	↑
9	Q8K1H9	Obp2a	Odorant-binding protein 2a	1.81533979	0.00799517	↑
10	Q80YX8	Mup21	Major urinary protein 21	−2.29692495	0.00913460	↓
11	Q3KQQ2	Mup3	Major urinary protein 25	−1.49751001	0.01240960	↓
12	Q91WR8	Gpx6	Glutathione peroxidase 6	1.59143645	0.01390856	↑
13	Q9D3H2	Obp1a	Odorant-binding protein 1a	−1.67221913	0.01499012	↓
14	A2BHR2	Lcn11	Lipocalin 11	1.54440618	0.01672011	↑
15	Q3TF14	Ahcy	Adenosylhomocysteinase	1.82072680	0.01895520	↑
16	A0A0U1RPF4	Hamp2	Hepcidin-2 (Fragment)	2.95442297	0.02102362	↑
17	P10639	Txn	Thioredoxin	1.993343846	0.027021861	↑
18	A2CEK6	Mup13	Major urinary protein 11	−1.85239538	0.02745480	↓
19	P15947	Klk1	Kallikrein-1	1.37162847	0.02874193	↑
20	P70269	Ctse	Cathepsin E	1.84456216	0.03194560	↑
21	P51910	Apod	Apolipoprotein D	1.69316293	0.03324157	↑
22	P15948	Klk1b22	Kallikrein 1-related peptidase b22	1.73666194	0.03657697	↑
23	H3BKH6	Esd	S-formylglutathione hydrolase	−1.63242181	0.03836365	↓
24	L7MUC7	Mup7	Major urinary protein 7 (Fragment)	−1.84000481	0.03930886	↓
25	Q91XE4	Acy3	N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming)	2.79646484	0.04085258	↑
26	Q91 × 17	Umod	Uromodulin	1.56345644	0.04212900	↑
27	B8JI96	Mup14	Major urinary protein 14 (Fragment)	−1.54699463	0.04765555	↓

TABLE 3 | Significantly altered pathways with differentially expressed proteins and metabolites.

No	Pathway name	Proteins	Metabolites
1	Pyrimidine metabolism		Uridine, UMP, dTMP, UDP-Gal
2	Metabolic pathways	UMOD,Fxyd2,AHCY,ACY3	L-Glu
3	Small Molecule Biochemistry	Hamp2,CTSE,SCLT1,WFDC2	UMP

key role in PTSD (Chambers et al., 1999; Lamprecht and LeDoux, 2004; Popoli et al., 2011; Kelmendi et al., 2016). Urine collects all metabolites of the body and is not regulated by the homeostasis mechanism. Abnormal l-glutamate metabolism detected in urine directly reflects PTSD.

In addition to metabolomic changes, significant proteomics differences were also identified. GO analysis of the urine proteome data showed that proteins with differential expression were mainly located in the extracellular space and extracellular region. It mainly binds to small-molecule, and it is involved in hydrolase activity, endopeptidase activity, and sodium channel regulator activity. Pathway enrichment analysis showed that these proteins are mainly involved in Endocrine and other factor-regulated calcium reabsorption, Lysosome, Renin-angiotensin system, Carbohydrate digestion and absorption, Metabolic pathways, etc.

Hepcidin is a circulating antimicrobial peptide involved in iron homeostasis, inflammation, infection, and metabolic signaling (Lu et al., 2015), There are two murine hepcidin genes: hepcidin-1 (Hamp1) and hepcidin-2 (Hamp2) (Truksa et al., 2007). Studies have shown that in addition to liver level, inflammation can increase the expression level of iron modulin

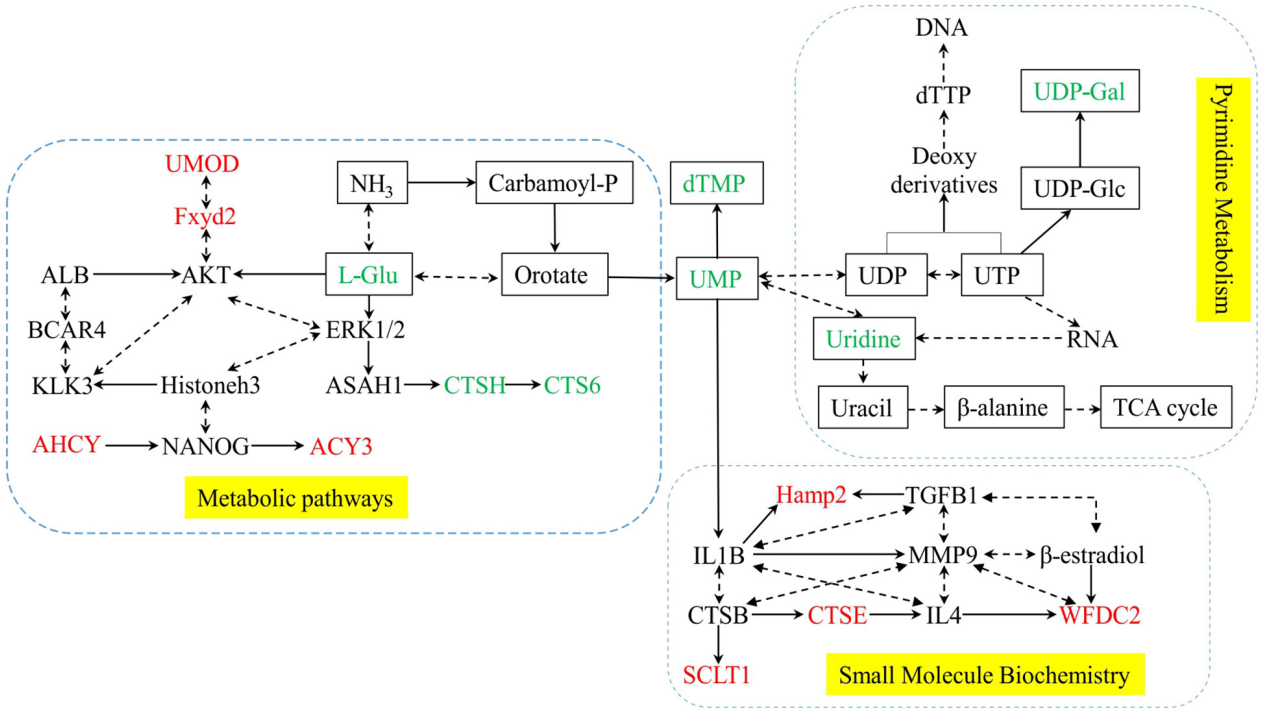

FIGURE 7 | Network of significantly differential metabolic pathways for Posttraumatic stress disorder. Yellow dashed areas represent the pathways. metabolites are shown as rectangles, and Proteins are shown without no rectangles. Red represents significant upregulation in the PTSD group compared to the control group, green represents significant downregulation.

(Kanamori et al., 2017; Silva et al., 2019). In this study, hepcidin-2 (Hamp2) expression increased. It has been reported that people with PTSD show elevated levels of pro-inflammatory cytokines, including IL1B (Dinarello, 2011; Tursich et al., 2014; Passos et al., 2015). In animal studies, IL1B expression level in hippocampus of depressed animal model was increased (Goshen et al., 2008). Therefore, Hamp2 expression may be induced by inflammatory factors in mice with PTSD. At the same time, IPA analysis showed that increased Hamp2 expression was correlated with IL1B, and the relationship between Hamp2 and IL1B in THE urine of PTSD will be further discussed in subsequent studies.

There are some limitations in the present study. First, the abundance of metabolites and proteins in the urine itself is small, and removing the peak degree through database construction ends up in data-loss, suggesting a potential data loss in urine protein-metabolism combined analysis. Second, only 6 eligible mice from each group were used for protein-metabolic analysis, and the results of the discovery omics study were not validated by targeted methods (e.g., western blotting). Therefore, further studies are required to validate these findings.

CONCLUSION

In this study, based on urine protein-metabolomics combined analysis, we found that the differentially expressed proteins of PTSD in mice were mainly in the extracellular space and region, and showed dysfunction of pyrimidine metabolism. Furthermore, Uridine and L-glutamate can be used as key urine biomarkers to provide a reference for subsequent studies on PTSD.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Committee of Army Medical University.

AUTHOR CONTRIBUTIONS

DZ, YH, and FL contributed to the design. XW, YWH, WS, and CL participated in the animal experiments. FP, YG, ZZ, and YS assisted to analyze the data. DZ drafted and modified the

manuscript. YH and GC provide to financial support. All authors have read and agreed to the manuscript for submission.

FUNDING

This work was supported by the Foundation of Army Medical University (2018XQN15) and Open Cooperation Program from Key Laboratory of Extreme Environmental Medicine, Ministry of Education of China (PF-KL2020-007).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.828382/full#supplementary-material

REFERENCES

Baloyianni, N., and Tsangaris, G. T. (2009). The audacity of proteomics: a chance to overcome current challenges in schizophrenia research. *Expert Rev. Proteomics* 6, 661–674. doi: 10.1586/epr.09.85

Bam, M., Yang, X., Zhou, J., Ginsberg, J. P., Leyden, Q., Nagarkatti, P. S., et al. (2016a). Evidence for epigenetic regulation of pro-inflammatory cytokines, interleukin-12 and interferon gamma, in peripheral blood mononuclear cells from PTSD patients. *J. Neuroimmune Pharmacol.* 11, 168–181. doi: 10.1007/s11481-015-9643-8

Bam, M., Yang, X., Zumbrun, E. E., Zhong, Y., Zhou, J., Ginsberg, J. P., et al. (2016b). Dysregulated immune system networks in war veterans with PTSD is an outcome of altered miRNA expression and DNA methylation. *Sci. Rep.* 6:31209. doi: 10.1038/srep31209

Bernstein, E., and Allis, C. D. (2005). RNA meets chromatin. *Genes Dev.* 19, 1635–1655. doi: 10.1101/gad.1324305

Bersani, F. S., Wolkowitz, O. M., Lindqvist, D., Yehuda, R., Flory, J., Bierer, L. M., et al. (2016). Global arginine bioavailability, a marker of nitric oxide synthetic capacity, is decreased in PTSD and correlated with symptom severity and markers of inflammation. *Brain Behav. Immun.* 52, 153–160. doi: 10.1016/j.bbi.2015.10.015

Bharadwaj, R. A., Jaffe, A. E., Chen, Q., Deep-Soboslay, A., Goldman, A. L., Mighdoll, M. I., et al. (2018). Genetic risk mechanisms of posttraumatic stress disorder in the human brain. *J. Neurosci. Res.* 96, 21–30. doi: 10.1002/jnr.23957

Breen, M. S., Maihofer, A. X., Glatt, S. J., Tylee, D. S., Chandler, S. D., Tsuang, M. T., et al. (2015). Gene networks specific for innate immunity define post-traumatic stress disorder. *Mol. Psychiatry* 20, 1538–1545. doi: 10.1038/mp.2015.9

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., and Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: the 1996 Detroit area survey of trauma. *Arch. Gen. Psychiatry* 55, 626–632. doi: 10.1001/archpsyc.55.7.626

Bush, N. J. (2010). Post-traumatic stress disorder related to cancer: hope, healing, and recovery. *Oncol. Nurs. Forum* 37, E331–E343. doi: 10.1188/10.ONF.E331-E343

Chambers, R. A., Bremner, J. D., Moghaddam, B., Southwick, S. M., Charney, D. S., and Krystal, J. H. (1999). Glutamate and post-traumatic stress disorder: toward a psychobiology of dissociation. *Semin. Clin. Neuropsychiatry* 4, 274–281. doi: 10.153/SCNP00400274

Connolly, G. P., and Duley, J. A. (1999). Uridine and its nucleotides: biological actions, therapeutic potentials. *Trends Pharmacol. Sci.* 20, 218–225. doi: 10.1016/s0165-6147(99)01298-5

Connolly, G. P., Simmonds, H. A., and Duley, J. A. (1996). Pyrimidines and CNS regulation. *Trends Pharmacol. Sci.* 17, 106–107. doi: 10.1016/0165-6147(96)20001-x

Daniels, W. M., de Klerk Uys, J., van Vuuren, P., and Stein, D. J. (2008). The development of behavioral and endocrine abnormalities in rats after repeated exposure to direct and indirect stress. *Neuropsychiatr. Dis. Treat.* 4, 451–464. doi: 10.2147/ndt.s2450

De Bellis, M. D., Keshavan, M. S., Spencer, S., and Hall, J. (2000). N-acetylaspartate concentration in the anterior cingulate of maltreated children and adolescents with PTSD. *Am. J. Psychiatry* 157, 1175–1177. doi: 10.1176/appi.ajp.157.7.1175

de Vries, G. J., Mocking, R., Lok, A., Assies, J., Schene, A., and Olff, M. (2016). Fatty acid concentrations in patients with posttraumatic stress disorder compared to healthy controls. *J. Affect. Disord.* 205, 351–359. doi: 10.1016/j.jad.2016.08.021

Dinarello, C. A. (2011). Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood* 117, 3720–3732. doi: 10.1182/blood-2010-07-273417

Emanuele, E., Colombo, R., Martinelli, V., Brondino, N., Marini, M., Boso, M., et al. (2010). Elevated urine levels of bufotenine in patients with autistic spectrum disorders and schizophrenia. *Neuro Endocrinol. Lett.* 31, 117–121.

Gao, Y. (2013). Urine-an untapped goldmine for biomarker discovery? *Sci. China Life Sci.* 56, 1145–1146. doi: 10.1007/s11427-013-4574-1

Gao, Y. (2015). Urine is a better biomarker source than blood especially for kidney diseases. *Adv. Exp. Med. Biol.* 845, 3–12. doi: 10.1007/978-94-017-9523-4_1

Girgenti, M. J., and Duman, R. S. (2018). Transcriptome alterations in posttraumatic stress disorder. *Biol. Psychiatry* 83, 840–848. doi: 10.1016/j.biopsych.2017.09.023

Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128, 635–638. doi: 10.1016/j.cell.2007.02.006

Gonzalez, F. J., and Fernandez-Salguero, P. (1995). Diagnostic analysis, clinical importance and molecular basis of dihydropyrimidine dehydrogenase deficiency. *Trends Pharmacol. Sci.* 16, 325–327. doi: 10.1016/s0165-6147(00)89065-3

Goshen, I., Kreisel, T., Ben-Menachem-Zidon, O., Licht, T., Weidenfeld, J., Ben-Hur, T., et al. (2008). Brain interleukin-1 mediates chronic stress-induced depression in mice via adrenocortical activation and hippocampal neurogenesis suppression. *Mol. Psychiatry* 13, 717–728. doi: 10.1038/sj.mp.4002055

Hemanth Kumar, B. S., Mishra, S. K., Rana, P., Singh, S., and Khushu, S. (2012). Neurodegenerative evidences during early onset of depression in CMS rats as detected by proton magnetic resonance spectroscopy at 7 T. *Behav. Brain Res.* 232, 53–59. doi: 10.1016/j.bbr.2012.03.011

Hemmings, S. M. J., Malan-Muller, S., van den Heuvel, L. L., Demmitt, B. A., Stanislawski, M. A., Smith, D. G., et al. (2017). The microbiome in posttraumatic stress disorder and trauma-exposed controls: an exploratory study. *Psychosom. Med.* 79, 936–946. doi: 10.1097/PSY.0000000000000512

Huang, T. L., and Lo, L. H. (2018). Proteomics approach for biomarker research in major depression: antidepressant effects. *Curr. Drug Metab.* 19, 502–512. doi: 10.2174/1389200219666180404094609

Jensen, O. N. (2006). Interpreting the protein language using proteomics. *Nat. Rev. Mol. Cell Biol.* 7, 391–403. doi: 10.1038/nrm1939

Jing, J., and Gao, Y. (2018). Urine biomarkers in the early stages of diseases: current status and perspective. *Discov. Med.* 25, 57–65.

Kaddurah-Daouk, R., and Krishnan, K. R. (2009). Metabolomics: a global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology* 34, 173–186. doi: 10.1038/npp.2008.174

Kanamori, Y., Murakami, M., Sugiyama, M., Hashimoto, O., Matsui, T., and Funaba, M. (2017). Interleukin-1beta (IL-1beta) transcriptionally activates hepcidin by inducing CCAAT enhancer-binding protein delta (C/EBPdelta) expression in hepatocytes. *J. Biol. Chem.* 292, 10275–10287. doi: 10.1074/jbc.M116.770974

Kangas, M., Henry, J. L., and Bryant, R. A. (2005). Predictors of posttraumatic stress disorder following cancer. *Health Psychol.* 24, 579–585. doi: 10.1037/0278-6133.24.6.579

Karabatsiakis, A., Hamuni, G., Wilker, S., Kolassa, S., Renu, D., Kadereit, S., et al. (2015). Metabolite profiling in posttraumatic stress disorder. *J. Mol. Psychiatry* 3:2. doi: 10.1186/s40303-015-0007-3

Kelmendi, B., Adams, T. G., Yarnell, S., Southwick, S., Abdallah, C. G., and Krystal, J. H. (2016). PTSD: from neurobiology to pharmacological

treatments. *Eur. J. Psychotraumatol.* 7:31858. doi: 10.3402/ejpt.v7.31858

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., and Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Arch. Gen. Psychiatry* 52, 1048–1060. doi: 10.1001/archpsyc.1995.03950240066012

Lamprecht, R., and LeDoux, J. (2004). Structural plasticity and memory. *Nat. Rev. Neurosci.* 5, 45–54. doi: 10.1038/nrn1301

Lauc, G., Pezer, M., Rudan, I., and Campbell, H. (2016). Mechanisms of disease: the human N-glycome. *Biochim. Biophys. Acta* 1860, 1574–1582. doi: 10.1016/j.bbagen.2015.10.016

Li, M. (2015). Urine reflection of changes in blood. *Adv. Exp. Med. Biol.* 845, 13–19. doi: 10.1007/978-94-017-9523-4_2

Li, M., Zhao, M., and Gao, Y. (2014). Changes of proteins induced by anticoagulants can be more sensitively detected in urine than in plasma. *Sci. China Life Sci.* 57, 649–656. doi: 10.1007/s11427-014-4661-y

Liu, L., Zhou, X., Zhang, Y., Liu, Y., Yang, L., Pu, J., et al. (2016). The identification of metabolic disturbances in the prefrontal cortex of the chronic restraint stress rat model of depression. *Behav. Brain Res.* 305, 148–156. doi: 10.1016/j.bbr.2016.03.005

Lu, S., Seravalli, J., and Harrison-Findik, D. (2015). Inductively coupled mass spectrometry analysis of biometals in conditional Hamp1 and Hamp1 and Hamp2 transgenic mouse models. *Transgenic Res.* 24, 765–773. doi: 10.1007/s11248-015-9879-3

Lundstrom, S. L., Yang, H., Lyutvinskiy, Y., Rutishauser, D., Herukka, S. K., Soininen, H., et al. (2014). Blood plasma IgG Fc glycans are significantly altered in Alzheimer's disease and progressive mild cognitive impairment. *J. Alzheimers Dis.* 38, 567–579. doi: 10.3233/JAD-131088

Mahan, A. L., and Ressler, K. J. (2012). Fear conditioning, synaptic plasticity and the amygdala: implications for posttraumatic stress disorder. *Trends Neurosci.* 35, 24–35. doi: 10.1016/j.tins.2011.06.007

Marc, D. T., Ailts, J. W., Campeau, D. C., Bull, M. J., and Olson, K. L. (2011). Neurotransmitters excreted in the urine as biomarkers of nervous system activity: validity and clinical applicability. *Neurosci. Biobehav. Rev.* 35, 635–644. doi: 10.1016/j.neubiorev.2010.07.007

Martin, C. G., Kim, H., Yun, S., Livingston, W., Fetta, J., Mysliwiec, V., et al. (2017). Circulating miRNA associated with posttraumatic stress disorder in a cohort of military combat veterans. *Psychiatry Res.* 251, 261–265. doi: 10.1016/j.psychres.2017.01.081

Mastrangelo, A., Panadero, M. I., Perez, L. M., Galvez, B. G., Garcia, A., Barbas, C., et al. (2016). New insight on obesity and adipose-derived stem cells using comprehensive metabolomics. *Biochem. J.* 473, 2187–2203. doi: 10.1042/BCJ20160241

Mellon, S. H., Gautam, A., Hammamieh, R., Jett, M., and Wolkowitz, O. M. (2018). Metabolism, metabolomics, and inflammation in posttraumatic stress disorder. *Biol. Psychiatry* 83, 866–875. doi: 10.1016/j.biopsych.2018.02.007

Mironova, G. D., Khrenov, M. O., Talanov, E. Y., Glushkova, O. V., Parfenyuk, S. B., Novoselova, T. V., et al. (2018). The role of mitochondrial KATP channel in anti-inflammatory effects of uridine in endotoxemic mice. *Arch. Biochem. Biophys.* 654, 70–76. doi: 10.1016/j.abb.2018.07.006

Miura, Y., and Endo, T. (2016). Glycomics and glycoproteomics focused on aging and age-related diseases–glycans as a potential biomarker for physiological alterations. *Biochim. Biophys. Acta* 1860, 1608–1614. doi: 10.1016/j.bbagen.2016.01.013

Montgomery, K. C. (1955). The relation between fear induced by novel stimulation and exploratory behavior. *J. Comp. Physiol. Psychol.* 48, 254–260. doi: 10.1037/h0043788

Nedic Erjavec, G., Konjevod, M., Nikolac Perkovic, M., Svob Strac, D., Tudor, L., Barbas, C., et al. (2018). Short overview on metabolomic approach and redox changes in psychiatric disorders. *Redox Biol.* 14, 178–186. doi: 10.1016/j.redox.2017.09.002

Nicholas, S. B. (2020). Use of urinary proteomics in diagnosis and monitoring of diabetic kidney disease. *Lancet Diabetes Endocrinol.* 8, 261–262. doi: 10.1016/S2213-8587(20)30067-X

Owens, G. P., Baker, D. G., Kasckow, J., Ciesla, J. A., and Mohamed, S. (2005). Review of assessment and treatment of PTSD among elderly American armed forces veterans. *Int. J. Geriatr. Psychiatry* 20, 1118–1130. doi: 10.1002/gps.1408

Pace, T. W., and Heim, C. M. (2011). A short review on the psychoneuroimmunology of posttraumatic stress disorder: from risk factors to medical comorbidities. *Brain Behav. Immun.* 25, 6–13. doi: 10.1016/j.bbi.2010.10.003

Park, D. I., Stambuk, J., Razdorov, G., Pucic-Bakovic, M., Martins-de-Souza, D., Lauc, G., et al. (2018). Blood plasma/IgG N-glycome biosignatures associated with major depressive disorder symptom severity and the antidepressant response. *Sci. Rep.* 8:179. doi: 10.1038/s41598-017-17500-0

Passos, I. C., Vasconcelos-Moreno, M. P., Costa, L. G., Kunz, M., Brietzke, E., Quevedo, J., et al. (2015). Inflammatory markers in post-traumatic stress disorder: a systematic review, meta-analysis, and meta-regression. *Lancet Psychiatry* 2, 1002–1012. doi: 10.1016/S2215-0366(15)00309-0

Pellow, S., Chopin, P., File, S. E., and Briley, M. (1985). Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *J. Neurosci. Methods* 14, 149–167. doi: 10.1016/0165-0270(85)90031-7

Pennington, D. L., Abe, C., Batki, S. L., and Meyerhoff, D. J. (2014). A preliminary examination of cortical neurotransmitter levels associated with heavy drinking in posttraumatic stress disorder. *Psychiatry Res.* 224, 281–287. doi: 10.1016/j.pscychresns.2014.09.004

Petricoin, E. F., Belluco, C., Araujo, R. P., and Liotta, L. A. (2006). The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat. Rev. Cancer* 6, 961–967. doi: 10.1038/nrc2011

Popoli, M., Yan, Z., McEwen, B. S., and Sanacora, G. (2011). The stressed synapse: the impact of stress and glucocorticoids on glutamate transmission. *Nat. Rev. Neurosci.* 13, 22–37. doi: 10.1038/nrn3138

Rabasa, C., Munoz-Abellan, C., Daviu, N., Nadal, R., and Armario, A. (2011). Repeated exposure to immobilization or two different footshock intensities reveals differential adaptation of the hypothalamic-pituitary-adrenal axis. *Physiol. Behav.* 103, 125–133. doi: 10.1016/j.physbeh.2011.02.022

Silva, I., Peccerella, T., Mueller, S., and Rausch, V. (2019). IL-1 beta-mediated macrophage-hepatocyte crosstalk upregulates hepcidin under physiological low oxygen levels. *Redox Biol.* 24:101209. doi: 10.1016/j.redox.2019.101209

Tanaka, M., Li, H., Zhang, X., Singh, J., Dalgard, C. L., Wilkerson, M., et al. (2019). Region- and time-dependent gene regulation in the amygdala and anterior cingulate cortex of a PTSD-like mouse model. *Mol. Brain* 12:25. doi: 10.1186/s13041-019-0449-0

Tandon, R. (2014). Schizophrenia and other psychotic disorders in diagnostic and statistical manual of mental disorders (DSM)-5: clinical implications of revisions from DSM-IV. *Indian J. Psychol. Med.* 36, 223–225. doi: 10.4103/0253-7176.135365

Thomson, C. A., McColl, A., Cavanagh, J., and Graham, G. J. (2014). Peripheral inflammation is associated with remote global gene expression changes in the brain. *J. Neuroinflammation* 11:73. doi: 10.1186/1742-2094-11-73

Truksa, J., Lee, P., and Beutler, E. (2007). The role of STAT, AP-1, E-box and TIEG motifs in the regulation of hepcidin by IL-6 and BMP-9: lessons from human HAMP and murine Hamp1 and Hamp2 gene promoters. *Blood Cells Mol. Dis.* 39, 255–262. doi: 10.1016/j.bcmd.2007.06.014

Tursich, M., Neufeld, R. W., Frewen, P. A., Harricharan, S., Kibler, J. L., Rhind, S. G., et al. (2014). Association of trauma exposure with proinflammatory activity: a transdiagnostic meta-analysis. *Transl. Psychiatry* 4:e413. doi: 10.1038/tp.2014.56

Van den Berg, C. L., Lamberts, R. R., Wolterink, G., Wiegant, V. M., and Van Ree, J. M. (1998). Emotional and footshock stimuli induce differential long-lasting behavioural effects in rats; involvement of opioids. *Brain Res.* 799, 6–15. doi: 10.1016/s0006-8993(98)00397-7

Vanhooren, V., Dewaele, S., Libert, C., Engelborghs, S., De Deyn, P. P., Toussaint, O., et al. (2010). Serum N-glycan profile shift during human ageing. *Exp. Gerontol.* 45, 738–743. doi: 10.1016/j.exger.2010.08.009

Vincenzetti, S., Polzonetti, V., Micozzi, D., and Pucciarelli, S. (2016). Enzymology of pyrimidine metabolism and neurodegeneration. *Curr. Med. Chem.* 23, 1408–1431. doi: 10.2174/0929867323666160411125803

Wang, Y., Chen, J., Chen, L., Zheng, P., Xu, H. B., Lu, J., et al. (2014). Urinary peptidomics identifies potential biomarkers for major depressive disorder. *Psychiatry Res.* 217, 25–33. doi: 10.1016/j.psychres.2014.02.029

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., et al. (1996). Progress with proteome projects: why

all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* 13, 19–50. doi: 10.1080/02648725.1996.10647923

Wingo, A. P., Almli, L. M., Stevens, J. S., Klengel, T., Uddin, M., Li, Y., et al. (2015). DICER1 and microRNA regulation in post-traumatic stress disorder with comorbid depression. *Nat. Commun.* 6:10106. doi: 10.1038/ncomms10106

Wu, C. H., Chen, S. H., Weng, L. J., and Wu, Y. C. (2009). Social relations and PTSD symptoms: a prospective study on earthquake-impacted adolescents in Taiwan. *J. Trauma. Stress* 22, 451–459. doi: 10.1002/jts.20447

Wu, J., and Gao, Y. (2015). Physiological conditions can be reflected in human urine proteome and metabolome. *Expert Rev. Proteomics* 12, 623–636. doi: 10.1586/14789450.2015.1094380

Yamamoto, T., Koyama, H., Kurajoh, M., Shoji, T., Tsutsumi, Z., and Moriwaki, Y. (2011). Biochemistry of uridine in plasma. *Clin. Chim. Acta* 412, 1712–1724. doi: 10.1016/j.cca.2011.06.006

Zoldos, V., Horvat, T., and Lauc, G. (2013). Glycomics meets genomics, epigenomics and other high throughput omics for system biology studies. *Curr. Opin. Chem. Biol.* 17, 34–40. doi: 10.1016/j.cbpa.2012.12.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fgene.2022.840577

A *de Novo* ZMIZ1 Pathogenic Variant for Neurodevelopmental Disorder With Dysmorphic Facies and Distal Skeletal Anomalies

Guanting Lu[1†], Liya Ma[2†], Pei Xu[1], Binqiang Xian[2], Lianying Wu[1], Jianying Ding[2], Xiaoyan He[1], Huiyun Xia[2], Wuwu Ding[1], Zhirong Yang[1] and Qiongling Peng[2]**

[1]Deyang Key Laboratory of Tumor Molecular Research, Department of Pathology, Translational Medicine Research Center, Deyang People's Hospital, Deyang, China, [2]Department of Child Healthcare, Shenzhen Baoan Women's and Children's Hospital, Jinan University, Shenzhen, China

OPEN ACCESS

Edited by:
Shicheng Guo,
University of Wisconsin-Madison,
United States

Reviewed by:
Xianli Shen,
Dana-Farber Cancer Institute and Harvard Medical School, United States
Yajie Gu,
University of California, San Diego, United States
Wentao Huang,
Massachusetts Institute of Technology, United States
Yiran Guo,
Children's Hospital of Philadelphia, United States

***Correspondence:**
Qiongling Peng
314633621@qq.com
Zhirong Yang
1324986817@qq.com

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: 21 December 2021
Accepted: 22 February 2022
Published: 31 March 2022

Citation:
Lu G, Ma L, Xu P, Xian B, Wu L, Ding J, He X, Xia H, Ding W, Yang Z and Peng Q (2022) A de Novo ZMIZ1 Pathogenic Variant for Neurodevelopmental Disorder With Dysmorphic Facies and Distal Skeletal Anomalies. Front. Genet. 13:840577. doi: 10.3389/fgene.2022.840577

Background: Neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA) is a rare syndromic disorder characterized by global neurodevelopmental delay, early-onset hypotonia, poor overall growth, poor speech/language ability, and additional common phenotypes such as eye anomalies, joint hypermobility, and skeletal anomalies of the hands and feet. NEDDFSA is caused by heterozygous pathogenic variants in the *ZMIZ1* gene on chromosome 10q22.3 with autosomal dominant (AD) mode of inheritance. All the 32 reported cases with variants in *ZMIZ1* gene had a genetic background in Caucasian, Hispanic, North African, and Southeastern Asian. Until now, there are no reports of Chinese patients with *ZMIZ1* pathogenic variants.

Methods: A 5-year-old girl was found to have the characteristic phenotypes of NEDDFSA. Array-Comparative Genomic Hybridization (array-CGH) and whole exome sequencing (WES) were applied for the trio of this female patient. Sanger sequencing was used to verify the selected variants. A comprehensive molecular analysis was carried out by protein structure prediction, evolutionary conservation, motif scanning, tissue-specific expression, and protein interaction network to elucidate pathogenicity of the identified ZMIZ1 variants.

Results: The karyotype was 46, XX with no micro-chromosomal abnormalities identified by array-CGH. There were 20 variants detected in the female patient by WES. A *de novo* heterozygous missense variant (c.2330G > A, p.Gly777Glu, G777E) was identified in the exon 20 of *ZMIZ1*. No variants of *ZMIZ1* were identified in the non-consanguineous parents and her healthy elder sister. It was predicted that G777E was pathogenic and detrimental to the spatial conformation of the MIZ/SP-RING zinc finger domain of ZMIZ1.

Conclusion: Thus far, only four scientific articles reported deleterious variants in *ZMIZ1* and most of the cases were from Western countries. This is the first report about a Chinese patient with *ZMIZ1* variant. It will broaden the current knowledge of ZMIZ1 variants and variable clinical presentations for clinicians and genetic counselors.

Keywords: Zmiz1, NEDDFSA, Chinese, low-complexity region, whole-exome sequencing

INTRODUCTION

Neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA; OMIM #618659) is a rare syndromic disorder characterized by global neurodevelopmental delay, hypotonia, poor overall growth, poor speech/language ability, and other common phenotypes such as eye anomalies, joint hypermobility, and distal skeletal anomalies of the hands and feet (Carapito et al., 2019). A balanced translocation t (10; 19) (q22.3; q13.33) was first reported in 2015, involving zinc finger MIZ-type containing 1 (*ZMIZ1*, OMIM #607159) and proline-rich protein 12 (*PRR12*, OMIM #616633). It produced two types of fusion genes, ZMIZ1-PRR12 and PRR12-ZMIZ1, which might be related to the occurrence of intellectual disability (ID) and neuropsychiatric alterations (Córdova-Fletes et al., 2015). Later, in 2019, pathogenic variants involving the gene ZMIZ1 were identified in a cohort of 19 NEDDFSA cases from a transatlantic collaborative effort (Carapito et al., 2019). In the same year of 2019, an affected father and his two sons were identified to be suffering from the ZMIZ1-related neurodevelopmental disorder in Florida (Latchman et al., 2020). In 2021, Phetthong et al. reported a 5-year-old Thai girl with developmental delay, facial phenotypes resembling Williams syndrome, and cardiac defects. She carried three types of compound variants, a heterozygous *ZMIZ1* variant (c.1497+2T > C), a heterozygous frameshift variant of *OTUD6B* (OMIM #612021) (c.873delA, p.Lys291AsnfsTer3), and a 0.118 Mb 8q21.3 microdeletion involving *OTUD6B* (Phetthong et al., 2021).

The gene ZMIZ1 was mapped to chromosome 10q22.3 and it contains 21 exons to produce a 1067-amino acid protein with a calculated molecular mass of 123 kDa (Sharma et al., 2003). According to the Conserved Domain database (CDD) (Lu et al., 2020), ZMIZ1 contains a Zmiz1 N-terminal tetratricopeptide repeat domain (Zmiz1_N, 8–100), Med15 domain (184–557), a nuclear localization signal (NLS, 697–711), a SP-RING zine finger domain (SP-RING_ZMIZ, 739–786), and a transactivation domain (TAD, 837–1067). In 1999, Nagase et al. identified the gene *ZMIZ1* (previously called KIAA1224) from a fetal brain cDNA library (Nagase et al., 1999). The encoded protein is a transcriptional co-activator, which belongs to the Protein Inhibitor of Activated STAT (PIAS) family. As a member of the PIAS family, ZMIZ1 has a highly conserved MIZ (Msx-interacting zinc finger) domain which is important for protein-protein interaction and SUMOylation (Sharma et al., 2003; Beliakoff and Sun, 2006). It had been reported that ZMIZ1 could regulate the activity of many transcription factors, such as androgen receptor (AR) (Beliakoff and Sun, 2006), SMAD3 (Li et al., 2006), SMAD4 (Li et al., 2006), and p53 (Lee et al., 2007). As an ortholog of ZMIZ1, tonalli (tna) was identified in *Drosophila melanogaster* and interacted with the ATP-dependent SWI/SNF complexes, which suggested a potential role in chromatin remodeling (Gutiérrez et al., 2003). Recently, ZMIZ1 was identified to be interacted with BRG1 (SMARCA4) (Li et al., 2011), BAF57 (SMARCE1) (Li et al., 2011), or SATB1 (Pinnell et al., 2015) to regulate the chromatin remodeling in humans. Chromatin remodeling complex could regulate the expression of genes which were essential for the normal dendrite development, synaptic plasticity, and synapse formation (Wu et al., 2007; Vogel-Ciernia et al., 2013; Vogel-Ciernia and Wood, 2014). It has been reported that *in utero* electroporation of ZMIZ1 pathogenic variants into the progenitor cells in the ventricular zone (VZ) of mice cortices (E14.5) resulted in impaired neuronal positioning with an accumulation in the ventricular and subventricular zones (VZ/SVZ) and intermediate zone (IZ) and a corresponding depletion in the upper cortical plate (CP). Therefore, *ZMIZ1* variants were regarded as the causal genetic factors for NEDDFSA (Carapito et al., 2019).

Thus far, no patients with *ZMIZ1* variants have been reported in Chinese. In order to decipher the genetic factors for neurodevelopmental disorder or intellectual disability (NEDD/ID) in China, array-CGH and WES were carried out for a cohort of 54 patients with NEDD/ID living in Shenzhen, Guangdong Province, China. After comprehensive bioinformatic analysis, a *de novo* missense variant (c.2330G > A, p.Gly777Glu, or p.G777E) was identified in the exon 20 of *ZMIZ1* in a 5-year-old girl with mild development delay, mild intellectual disability, bilateral hip dysplasia, joint hypermobility, amblyopia in both eyes, strabismus in the right eye, and dysmorphic facial features. According to the criteria proposed by the American College of Medical Genetics and Genomics (ACMG) (Richards et al., 2015), this variant was classified as PS2 + PM1 + PM2 + PP2 + PP3 and annotated as "Pathogenic." After comparing the clinical phenotypes described for NEDDFSA with the clinical phenotypes of our current Chinese patient, this girl was diagnosed as NEDDFSA. This variant is located in the highly conserved zf-MIZ domain and affected the three-dimensional conformation which might be detrimental for the binding of ZMIZ1 to its partners.

To our knowledge, this is the first case of Chinese NEDD/ID caused by a *ZIMI1* variant. Due to the huge population, more patients with ZMIZ1-related disorder will be found in the near future.

METHODS

Sample Collection

This study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. This study was approved by the Ethics Committee of the Shenzhen Baoan Women's and Children's Hospital. Written informed consent was obtained from each individual.

Peripheral venous blood was collected from the 54 NEDD/ID patients and their parents. Genomic DNA was extracted using the TIANamp Blood DNA Kit (DP348, Tiangen Biotech, Beijing, China) according to the manufacturer's instructions.

Array-Comparative Genomic Hybridization

Array-CGH was performed using the Fetal DNA Chip (Version 1.2) designed by The Chinese University of Hong Kong (CUHK) (Leung et al., 2011; Huang et al., 2014). The chip contains a total

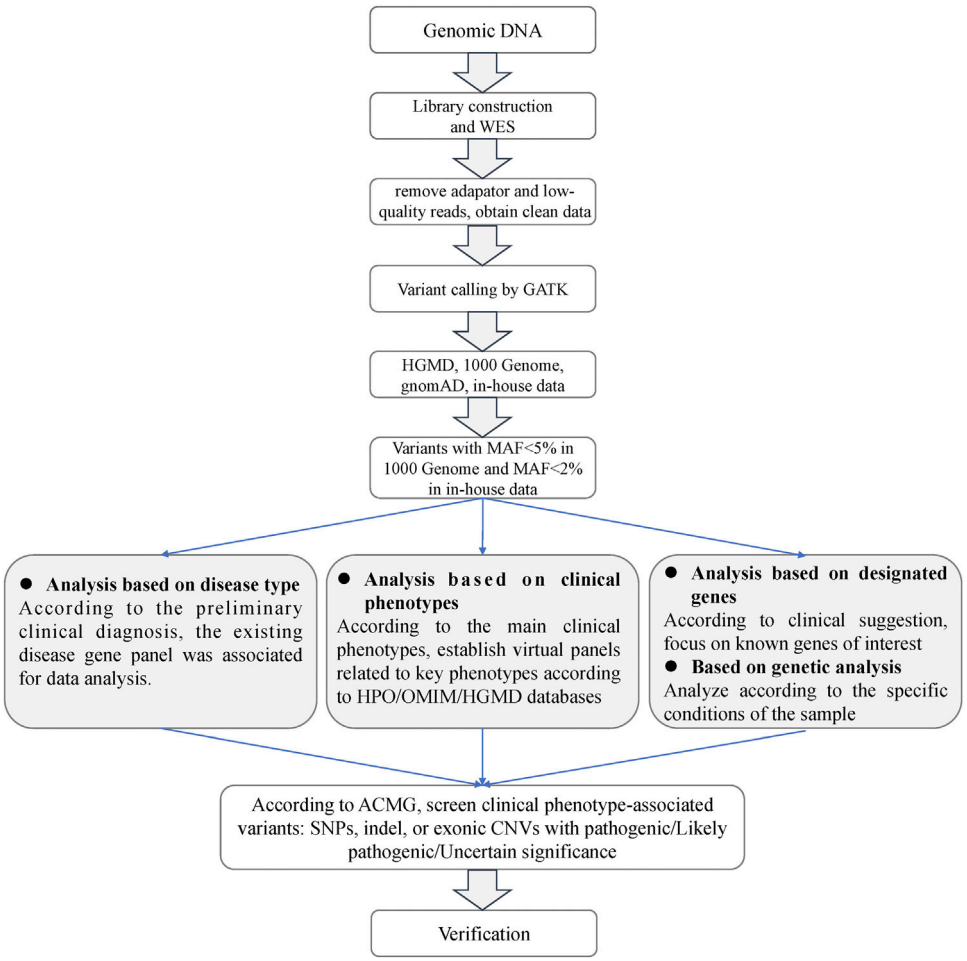

FIGURE 1 | Analysis flowchart of the whole exome sequencing data.

of 60,000 probes for more than 100 diseases caused by known microduplication/microdeletions. It does not include small-size chromosomal abnormalities, copy number polymorphism, chimerism, or chromosomal rearrangement (Iafrate et al., 2004). The experimental procedures were carried out according to the standard Agilent protocol (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis, version 3.5). Hybridized slides were scanned with SureScan High-Resolution Microarray Scanner (G2505B, Agilent Technologies, Santa Clara, CA), and the image data were extracted and converted to text files using Agilent Feature Extraction software (Version 10.5.1.1). The data were graphed and analyzed using Agilent CGH Analytics software.

Only gains or losses that were encompassed by at least three consecutive oligomers on the array were considered. Then, the clinical relevance of observed chromosomal aberrations was estimated according to data found in the scientific literature and databases for each of the regions and genes involved, using the DECIPHER database (Swaminathan et al., 2012) for known microdeletion and microduplication syndromes and the Online Mendelian Inheritance in Man (OMIM) (Sayers et al., 2021) for known disease-causing genes, gene functions, and inheritance patterns. Copy number variations were considered as "likely pathogenic/pathogenic" when they involved regions known to be associated with microdeletion or microduplication syndromes.

High-Throughput Whole Exome Sequencing

WES was performed for family trios (trio-WES) without chromosomal abnormalities at MyGenostics or BerryGenomics Co. LTD. Briefly, the fragmented genomic DNAs were ligated with the 3′ end of the Illumina adapters and amplified by polymerase chain reaction (PCR). The amplified DNA was captured with Gencap Human whole Exon Kit (52M) at MyGenostics or with xGen Exome Research Panel v2.0 (Integrated DNA Technologies, Coralville, IA) at BerryGenomics. The capture procedure was performed in accordance with the manufacturer's protocol. Finally, the generated libraries were sequenced on Illumina HiSeq 2500 platform for paired-end sequencing.

The sequencing depth of each sample was about 100. Sequencing reads were aligned with the human reference genome (UCSC hg19). The workflow of the screening for causal variants was depicted in **Figure 1**. Briefly, clean reads were obtained after removal of adaptors and low-quality reads. GATK (Genome Analysis Toolkit) was used to trim the variant calling in the trimmed WES clean data (https://gatk.broadinstitute.org/hc/en-us). ANNOVAR was applied to annotate the generated VCF file (Wang et al., 2010). Deleted variants with a minor allele frequency (MAF) > 5% in the 1000 Genome Project, MAF >2% in in-house data, or synonymous single nucleotide variants (SNVs) were removed. SNVs that caused splicing, frameshift, stop gain, or stop loss were retained for subsequent analysis. A position was called as heterozygous if 25% or more of the reads identify the minor allele.

The chromosomal location and type of the identified variants were retrieved in UCSC Genome Browser (Navarro Gonzalez et al., 2021) and NCBI dbSNP (Bhagwat, 2010). The MAFs of the variants were screened in several public databases with a large number of human samples, such as 1000 Genome Project (n = 2504) (Siva, 2008), NHLBI Exome Sequencing Project (GO-ESP) (n = 6503) (Amendola et al., 2015), The Exome Aggregation Consortium (ExAC) (n = 60,706) (Lek et al., 2016), gnomAD (n = 15,708) (Scheps et al., 2020), and NHLBI Trans-Omics for Precision Medicine (TOPMED) (n = 60,000) (Taliun et al., 2021). The function prediction of these variants was carried out by online software, PolyPhen-2 (Adzhubei et al., 2010), and PROVEAN (Choi et al., 2012). Pathogenicity of the variants was evaluated according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015). The selected variants were verified by Sanger sequencing using the ABI 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA).

Computational Analysis for the G777E Variant of ZMIZ1

Protein sequences of ZMIZ1 in 34 species were downloaded from NCBI GenBank, including five primates (*Homo sapiens*, *Pan troglodytes*, Gorilla *Gorilla gorilla*, *Hylobates moloch*, and *Macaca fascicularis*), one cattle (*Bos taurus*), one horse (*Equus caballus*), one dog (*Canis lupus* familiar), three carnivores (Neogale vison, *Panthera tigris*, and Halichoerus grypus), three rodents (*Eptesicus fuscus*, *Mus musculus*, and *Rattus norvegicus*), five reptiles (Crotalus tigris, *Varanus komodoensis*, Dermochelys coriacea, Chelonoidis abingdonii, and Mauremys mutica), two birds (*Falco rusticolus*, *Gallus gallus*), three amphibians (*Bufo bufo*, *Xenopus* tropicalis, and *Oryzias latipes*), three fish (*Takifugu rubripes*, Hippocampus comes, and *Danio rerio*), three arthropods (Limulus polyphemus, *Penaeus monodon*, and Ceratitis capitata), and four mollusks (Acropora millepora, *Crassostrea gigas*, Octopus sinensis, and Exaiptasia diaphana). The protein sequences were aligned by the ClustalW alignment algorithms of MEGA X (Kumar et al., 2018) (gap opening penalty and gap extension penalty for pairwise alignment and multiple alignment were set as 10.00, 0.10 and 10.00, 0.20, respectively; the delay divergent cutoff was 30%).

The intrinsically disordered regions of ZMIZ1 protein (NP_065071) were analyzed using the online web server IUPred2A (https://iupred2a.elte.hu/) with long disorder setting to identify probable disordered regions using the IUPred2 model and disordered binding regions using the ANCHOR2 model (Erdos and Dosztanyi, 2020). The distinct motifs of ZMIZ1 were analyzed using the online software Motif Scan (https://myhits.sib.swiss/cgi-bin/motif_scan) under default settings to search all known motifs in HAMAP (Pedruzzi et al., 2015), PROSITE (Hulo et al., 2006), Pfam (Mistry et al., 2021), and InterPro databases (Mitchell et al., 2019). The possible phosphorylation sites of ZMIZ1 were predicted by Disorder Enhanced Phosphorylation Predictor (DEEP, http://www.pondr.com/cgi-bin/depp.cgi) using 0.50 as the cutoff value (Iakoucheva et al., 2004).

The effect of G777E on the structural change was predicted by the online protein structure and function prediction tool, I-TASSER (Iterative Threading ASSEmbly Refinement) under default parameters (Yang and Zhang, 2015) for the whole second globular region (aa575-820) and visualized using Mol* 3D Viewer (Sehnal et al., 2021). The gene expression data of ZMIZ1 were evaluated according to the normalized signal intensity of probe 212124 at which were extracted from a gene atlas of human protein-encoding transcriptomes for 79 human tissues (NCBI GEO #GSE1133) (Su et al., 2004). The protein interaction network with ZMIZ1 (PPI enrichment p value = 1.51E-03) was generated by STRING (version 11.5, https://string-db.org/) under default settings. Gene Ontology (GO) analysis was performed on the nine members of the network in the GO knowledgebase (http://geneontology.org/) under default parameters.

RESULTS

Sample Description

There were 54 cases in our current NEDD/ID cohort collected from southern China. There were 48.15% (26/54) women and 51.85% (28/54) men. The mean age of women and men patients was 2.45 ± 1.15 and 2.67 ± 1.71, respectively. In these 54 samples, pathogenic variants were found in 33 patients, 5 with microdeletions and 28 with variants in protein-coding genes (**Supplementary Table S1**). The positive rate was 61.11% (33/54). In one patient, *ZMIZ1* was detected to have a pathogenic missense variant (c.2330G > A, p.G777E). This patient was a 5-year-old girl who was referred to our department because of psychomotor developmental delay. She was the second child of a non-consanguineous couple (**Figure 2A**). The proband was delivered at term to a 36-year-old mother by Cesarean section due to breech position at 2016-10. Her birth weight was 3000 g, and there was no history of asphyxia at birth. At the sixth month after birth, asymmetric dermatoglyphs were found on both of her lower limbs after a physical examination and later diagnosed as dysplasia of bilateral hip joints. At 1 year old, the patient had chronic constipation. Since 2017, she has been sent to the

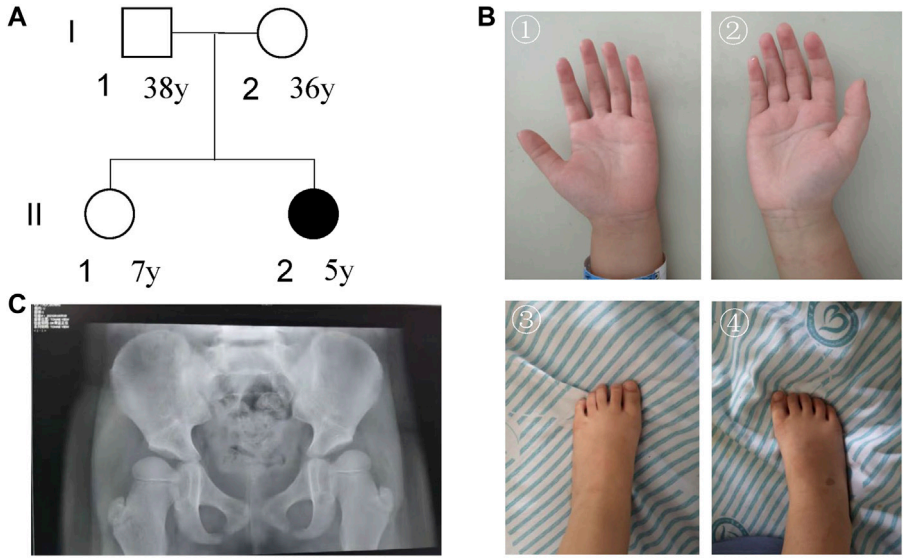

FIGURE 2 | Characterization of the patient's information. **(A)** Pedigree; **(B)** pictures of hands and foot; **(C)** DR X-ray film for bilateral hip joints.

ophthalmology department several times due to binocular weak eyesight and strabismus in the right eye. After 1 year old, she was still unable to speak and walk without support and was sent to the rehabilitation center for special training. Until 2 years old, she was able to speak simple words and walk, and was finally diagnosed as "developmental delay".

For **facial features**, the patient had epicanthus, ptosis, up-slanting eyelid fissure, wide eye distance, wide nose bridge, Cupid lip arch and low-set ear. Regarding **skeletal abnormalities**, she had short fingers and toes, conical fingers (**Figure 2B**), and excessive joint mobility. As for the **gross motor**, the patient could walk alone. She could not stand on one foot for more than 3 s and jump on one foot. The trotting posture was slightly abnormal, and easy to fall. In terms of **fine motor**, she could draw a straight line, pull a zipper, unbutton buckles, cut paper inflexibly with scissors, eat with spoons and chopsticks, but couldn't draw circles, squares, and triangles. In terms of the **language**, she could speak simple and long sentences with clear pronunciation, understand a few Chinese characters, recite numbers from 1 to 20, answer simple questions, but sometimes with confused word order and logic. For the **social adaptive capacity**, she could wear and take off clothes, shoes, and socks, could go to the bathroom, and eat by herself. She had the initiative to share and express her needs but without the initiative to say hello and goodbye. Besides, she had poor name calling response, poor sitting quietly ability, and hyperactive behavior.

The visually evoked potential (VEP) test showed that after blink flash stimulation for both eyes, N75, P100, and N145 waves could be induced with good repeatability. However, the latency of the P100 and N145 waves on both sides was prolonged, which was slightly abnormal. The evaluations for audiology system, heart, and urinary system were normal. DR X ray film for hip joint anteroposterior projection at 4-years-7-months old showed that the left and right acetabular angles were about 22 and 25°, respectively (**Figure 2C**). She was diagnosed with congenital dysplasia of the hip by an orthopedic surgeon at a tertiary children's hospital. She received 2 brain MRI scans (March 28, 2019and October 26, 2021) and 3 electroencephalogram (EEG) examinations (March 15, 2019, August 20, 2020, and June 23, 2021). No obvious irregularities were identified.

Neuropsychological development assessment was performed for the patient at 5-years-1-month old using the Wechsler Preschool and Primary Scale of Intelligence Fourth Edition (WPPSI-IV) and the parent-rated Adaptive Behavior Assessment System II (ABAS-II) infant version. Her score on the full-scale intelligence quotient of WPPSI-IV was 75 (95% CI: 70-82, P5). The verbal comprehension index, visual spatial index, perceptual reasoning index, working memory index, and processing speed index of WPPSI-IV were 77 (95% CI: 71-86, P6), 83 (95% CI 76-94, P13), 79 (95% CI 73-87, P8), 76 (95% CI 76-94, P5), and 71 (95% CI 66-85, P3), respectively. The overall adaptive function score of ABAS-II was 77 (95% CI: 73-81, P6). The scores of social skills, conceptual skills, and practical skills in the three composite areas of adaptive function were 71 (95% CI 64-78, P3), 84 (95% CI 77-91, P14), and 80 (95% CI 74-86, P9), respectively. According to the clinical evaluation, she was at the edge level of intellectual development.

Trio-WES Identified a *de Novo* Missense Variant of *ZMIZ1* Gene

Whole exome sequencing was performed for the trio to identify possible genetic factors of the proband. After removal of adaptors and low-quality reads, the obtained total clean data obtained for the trio were 11,959.04 (Mb) for the proband, 15,681.80 (Mb) for the father, and 12,202.51 (Mb) for the mother (**Table 1**). The

TABLE 1 | Characterization of the whole exome sequencing for the trio.

Items	Proband	Father	Mother
Total clean data (Mb)	11,959.04	15,681.80	12,202.51
Target coverage	98.00%	98.31%	98.04%
Average depth of target region (X)	110.73	138.49	114.47
Ratio of average depth of target region (>4X)	97.59%	97.89%	97.64%
Ratio of average depth of target region (>10X)	97.29%	97.62%	97.36%
Ratio of average depth of target region (>20X)	96.77%	97.29%	96.85%
Ratio of average depth of target region (>30X)	95.64%	96.76%	95.71%
On target ratio	39.66%	37.83%	40.19%
Total SNVs	175,809	197,560	179,722
Percentage of pathogenic variants	3.57%	4.46%	3.47%
Total small insertion (and duplications)	17,011	19,694	17,254
Total small deletions	19,441	23,004	19,562

target coverage was at least or more than 98%. The average depth of target region was more than 100X. The on-target ratio was more than 35%. In these samples, the total numbers of identified SNVs were 175,809 for the proband, 197,560 for the father, and 179,722 for the mother, respectively. The percentages of pathogenic variants were around 4%. The total number of small insertions and deletions were 36,452, 42,698, and 36,816 for the proband, father, and mother, respectively.

In this proband, 20 specific variants were selected. Twelve of them were heterozygous in 10 genes (*ANKRD36C, MUC2, MUC4, HRCT1, KLHL29, MYO15B, PER3, RPTN, TWIST1,* and *ZMIZ1*) and 8 homozygous in 8 genes (*AGAP3, CCDC177, CGN, DSPP, ESX1, FOXN4, MUC4, POTEB3,* and *SLC35E2A*) (**Table 2**). According to the criteria of ACMG guidelines, 10 heterozygous and 8 homozygous variants were annotated as variants of uncertain significance (VUS). Most of these variants were predicted to be "neutral" by Provean or "benign" by Polyphen. *PER3* (OMIM #603427), *TWIST1* (OMIM #601622), and *DSPP* (OMIM #125485) were also recorded in the OMIM database as disease-causing genes. However, the phenotypes caused by these genes were not in line with our female proband. Besides, two rare heterozygous variants (c.148C > T, p.R50W in *KLHL29* and c.2330G > A, p.G777E in *ZMIZ1*) were annotated as "likely pathogenic" (PS2 + PM2 + PP2) and "pathogenic" (PS2 + PM1 + PM2 + PP2 + PP3), respectively.

Since there was no experimental evidence for *KLHL29* leading to neurodevelopmental disorder, *ZMIZ1* was considered as the most potential disease-causing gene. The c.2330G > A (p.G777E) was a heterozygous SNV in the exon 20 of *ZMIZ1* gene (NM_020338) (**Figures 3A,B**) and confirmed by Sanger sequencing (**Figure 3C**) only in the patient, but not in her healthy parents or her elder sister. Thus, it was a *de novo* variant. The protein sequences of ZMIZ1 from more than 34 species (Mollusca, crabs, fish, amphibians, insects, reptiles, rodents, dogs, cats, cattle, and primates) were downloaded from NCBI GenBank and aligned by the ClustalW alignment algorithms of MEGA 11, the G777 was highly conserved in the animals during evolution (**Figure 3D**). G777E was localized in the functional MSX-interacting zinc finger (zf-MIZ) domain (**Figure 3E**) and predicted to be "deleterious" with a score of 0.536, "deleterious" with a score of -7.736 (Provean) and "probably damaging" with a score of 0.992 (PolyPhen-2). In addition, this SNV has not been detected in multiple public genome databases, such as 1000 Genome Project (n = 2504), NHLBI Exome Sequencing Project (GO-ESP) (n = 6503), the Exome Aggregation Consortium (ExAC) (n = 60,706), Genome Aggregation database (gnomAD) (n = 15,708), and NHLBI Trans-Omics for Precision Medicine (TOPMED) (n = 60,000). Since this amino acid changing variant was only identified in the patient, and not in her parents, it was regarded as "spontaneous." According to the Probability of Loss-of-function Intolerance (pLI) analysis, the pLI value of ZMIZ1 was 1.000, which indicated *ZMIZ1* being a haploinsufficient gene. It has been reported that *ZMIZ1* could cause the occurrence of a rare neurodevelopmental disorder, neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA). Based on the recorded clinical phenotypes (**Supplementary Table S2**), this patient was finally diagnosed as NEDDFSA. Besides, four rare variants in *ZMIZ1* were also detected in another four NEDD patients, namely, c.3096 + 15C > T, c.1024A > G (p.M342V), c.540 + 20T > C and c.679G > A (p.A227T) (**Supplementary Table S3**).

The intrinsically disordered regions of ZMIZ1 protein (NP_065071) were analyzed using the online web server IUPred2A. Two functional globular regions were identified at two portions (aa2-110 and 575-820), which overlapped with the two important functional domains, Zmiz1 N-terminal tetratricopeptide repeat domain (Zmiz1_N, aa8-100) and MIZ/SP-RING zinc finger (zf-MIZ, aa739-786), respectively. The predicted two globular regions and two intrinsically disordered regions are displayed in **Figure 3E**. The possible phosphorylation sites of ZMIZ1 were predicted using DEEP under default parameters. Interestingly, most of the phosphorylation sites were located in the two long disordered regions (**Figure 3F**). As for G777E, it was localized in the second globular regions containing zf-MIZ domain and might affect the probable tertiary structures as predicted by I-TASSER (**Supplementary Figure S1**).

Analysis of the Distinct Regions of ZMIZ1

The distinct regions of ZMIZ1 were analyzed using the online software Motif Scan under default settings. Seven distinct

TABLE 2 | Identified variants in the five-year-old proband.

No	Location (GRCH37)	Genes	Ref genes	Variants	dbSNP ID	Zygosity (P/F/M)	ACMG annotation	1000 genomes	ExAC	gnomAD exome	PROVEAN (score)	Polyphen2 (score)	Phenotype OMIM	Inheritance and phenotype
1	2: 96521280	ANKRD36C	NM_001310154	c.5827A > C (p.I1943L)	rs112858216	Het/ WT/WT	VUS: PS2	—	3.26E-02	—	Neutral −0.876	—	—	—
2	9: 35906601	HRCT1	NM_001039792	c.317C > A (p.P106H)	rs112212538	Het/ WT/WT	VUS: PS2	—	—	—	Neutral 0.071	Benign 0.146	—	—
3	2: 23785214	KLHL29	NM_052920	c.148C > T (p.R50W)	rs558454968	Het/ WT/WT	Likely pathogenic: PS2+PM2+PP2	—	1.52E-04	9.87E-05	Neutral −1.567	Damaging 0.988	—	—
4	11: 1093349	MUC2	NM_002457	c.6863C > T (p.P2288L)	rs1382972456	Het/ WT/WT	VUS: PM2	—	—	—	Neutral −0.461	—	—	—
5	3: 195507271	MUC4	NM_018406	c.11180C > G (p.T3727S)	rs868067409	Het/ WT/WT	VUS: PM2	—	1.67E-04	1.76E-04	Neutral 0.217	Benign 0.301	—	—
6	3: 195508523	MUC4	NM_018406	c.9928G > A (p.A3310T)	rs879281830	Het/ WT/WT	VUS: NA	—	2.79E-03	4.42E-04	Neutral 0.500	Damaging 0.494	—	—
7	3: 195508526	MUC4	NM_018406	c.9925C > G (p.H3309D)	rs1424606542	Het/ WT/WT	VUS: NA	—	2.43E-03	3.27E-04	Neutral 0.083	Benign 0.234	—	—
8	17: 73585468	MYO15B	NM_001309242	c.1330C > T (p.R444C)	rs185791490	Het/ WT/WT	VUS: PM2	—	—	—	Neutral −0.446	—	—	—
9	1:7890053	PER3	NM_016831	c.3019G > A (p.A1007T)	rs1776342	Het/ WT/WT	VUS: PM2+PP3+BP4	—	—	9.47E-06	Neutral −0.848	Benign 0.004	#616882	AD: Advanced sleep phase syndrome, familial, 3
10	1: 152129100	RPTN	NM_001122965	c.475G > A (p.G159S)	rs200003389	Het/ WT/WT	VUS: PM2+BP4	—	—	—	Neutral −1.883	Benign 0.275	—	—
11	7: 19156668	TWIST1	NM_000474	c.256_276dup (p.G86_G92dup)	—	Het/ WT/WT	VUS: PS2+BP3	—	0	1.58E-05	—	—	#123100	AD: Craniosynostosis 1
													#180750	AD: Robinow-Sorauf syndrome
													#101400	AD: Saethre-Chotzen syndrome with or without eyelid anomalies
													#617746	AD: Sweeney-Cox syndrome
12	10: 81064964	ZMIZ1	NM_020338	c.2330G > A (p.G777E)	—	Het/ WT/WT	Pathogenic: PS2+PM1+PM2+PP2+PP3	—	—	—	Deleterious -7.736	Damaging 0.992	#618659	AD: Neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA)
13	7: 150783920	AGAP3	NM_031946	c.92T > G (p.V31G)	rs1171186819	Hom/ WT/WT	VUS: PM2	—	—	—	Neutral −0.091	—	—	—
14	14: 70039807-70039809	CCDC177	NM_001271507	c.534_536del (p.A180del)	—	Hom/ Het/Het	VUS: PM2+ PM3_supporting + BP3	—	—	—	—	—	—	—
15	1: 151491411	CGN	NM_020770	c.416C > T (p.A139V)	rs181435993	Hom/ Het/Het	VUS: NA	9.98E-04	7.44E-04	6.33E-04	Neutral −1.470	Damaging 0.937	—	—
16	4: 88535832	DSPP	NM_014208	c.2018A > G (p.D673G)	rs201553143	Hom/ Het/Het	VUS: PM2	—	1.98E-04	—	Neutral −1.162	Benign 0.004	#605594	AD: Deafness, autosomal dominant 39, with dentinogenesis
													#125420	AD: Dentin dysplasia, type II

(Continued on following page)

TABLE 2 | (*Continued*) Identified variants in the five-year-old proband.

No	Location (GRCH37)	Genes	Ref genes	Variants	dbSNP ID	Zygosity (P/F/M)	ACMG annotation	1000 genomes	ExAC	gnomAD exome	PROVEAN (score)	Polyphen2 (score)	Phenotype OMIM	Inheritance and phenotype
													#125490	AD: Dentinogenesis imperfecta, Shields type II
													#125500	AD: Dentinogenesis imperfecta, Shields type III
17	X: 103495090	ESX1	NM_153448	c.1040C > G (p.P347R)	rs200088361	Hom/ Hemi/Het	VUS: NA	–	1.67E-03	1.04E-03	Neutral −0.236	Damaging 0.915	–	–
18	12: 109719311	FOXN4	NM_213596	c.1195G > A (p.A399T)	rs146550988	Hom/ Het/Het	VUS: NA	2.20E-03	1.35E-03	1.49E-03	Neutral −0.233	Benign 0.083	–	–
19	3: 195506197	MUC4	NM_018406	c.12254A > G (p.D4085G)	rs148307810	Hom/ Het/Hom	VUS: NA	7.39E-03	–	2.28E-03	Neutral −1.433	Damaging 0.553	–	–
20	15: 22053725	POTEB3	NM_207355	c.1531A > G (p.K511E)	rs1949282	Hom/ Het/WT	VUS: NA	–	–	–	Neutral 0.706	Benign 0.000	–	–

P, proband; F, father; M, mother; WT, wild type; Het, heterozygous; Hom, homozygous; Hemi, hemizygous; VUS, variants of uncertain significance; PS, strong pathogenic; PM, moderate pathogenic; PP, pathogenic supporting; BP, benign supporting; NA, not available; AD, autosomal dominant.

regions were identified, one alanine-rich region (aa280-305, E-score = 2.1E-06), two proline-rich regions (aa334-555, E-score = 3.9E-16; aa867-1002, E-score = 3.8E-07), one bipartite nuclear localization signal (NLS, aa697-711, E-score = 2.1E+04), MIZ/SP-RING zinc finger (aa738-787, E-score = 1.1E-33), and one copper binding octapeptide (aa947-954, E-value = 1.5). All variants of ZMIZ1 were also recruited from the DECIPHER database (Swaminathan et al., 2012) and the four published articles (Córdova-Fletes et al., 2015; Carapito et al., 2019; Latchman et al., 2020; Phetthong et al., 2021). A total of 33 patients with ZMIZ1 pathogenic variants were collected, 1 from our current cohort (**Figures 4A,B**), 8 from the DECIPHER database (**Figure 4C**), and 24 from published articles (**Figure 4D**). Except for K91R, H581R, and H683Y, other variants were localized in the low-complexity regions, such as the alanine-rich region and the proline-rich region of the Med15 (mediator complex subunit 15) domain, and the proline-rich region in the C-terminal transactivation domain (TAD). There were nine amino acid-changing variants, which were strongly conserved during evolution (**Figure 4E**). Six of them were in the alanine-rich region, accounting for 66.67% (6/9). From the phosphorylation prediction by DEEP, except for T300M, other variants could distinctly change the phosphorylation pattern of the alanine-rich region (**Figure 4F**).

Interaction Network of ZMIZ1

The gene expression data of 79 human tissues showed that *ZMIZ1* was expressed in the heart, thyroid, immune cells, ovary, retina, and brain, with the highest in the pineal (**Figure 5A**). The protein interaction network with ZMIZ1 indicated that ZMIZ1 could interact with SMAD3, SMAD4, MYC, NOTCH1, RBPJ, SMARCA4, ETS1, and UBE2I (**Figure 5B**). According to the GO analysis for the 9 members (**Figure 5C**), the network was involved significantly in mesenchyme morphogenesis, hypoxia, tube morphogenesis, regulation of transcription, response to stimulus, endocardium development, epithelial to mesenchymal transition, and cardiac left ventricle morphogenesis in GO term "biological process." In "molecular function," transcription, SMAD binding, and SUMOylation were significantly enriched. As for "cellular component" and "subcellular localization," members of this network were localized in nuclear to form multiple protein complexes, mainly MAML1-RBP-Jκ-ICN1 (Intracellular Notch1) complex and SMAD protein complex to regulate the expression of target genes.

DISCUSSION

Pathogenic variants of the zinc finger MIZ-type containing 1 (*ZMIZ1*, OMIM#607159) could cause the occurrence of a rare syndromic disease, neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA) with an autosomal dominant (AD) mode of inheritance. Currently, 32 patients with neurodevelopmental disorders have reported carrying pathogenic variants in the protein

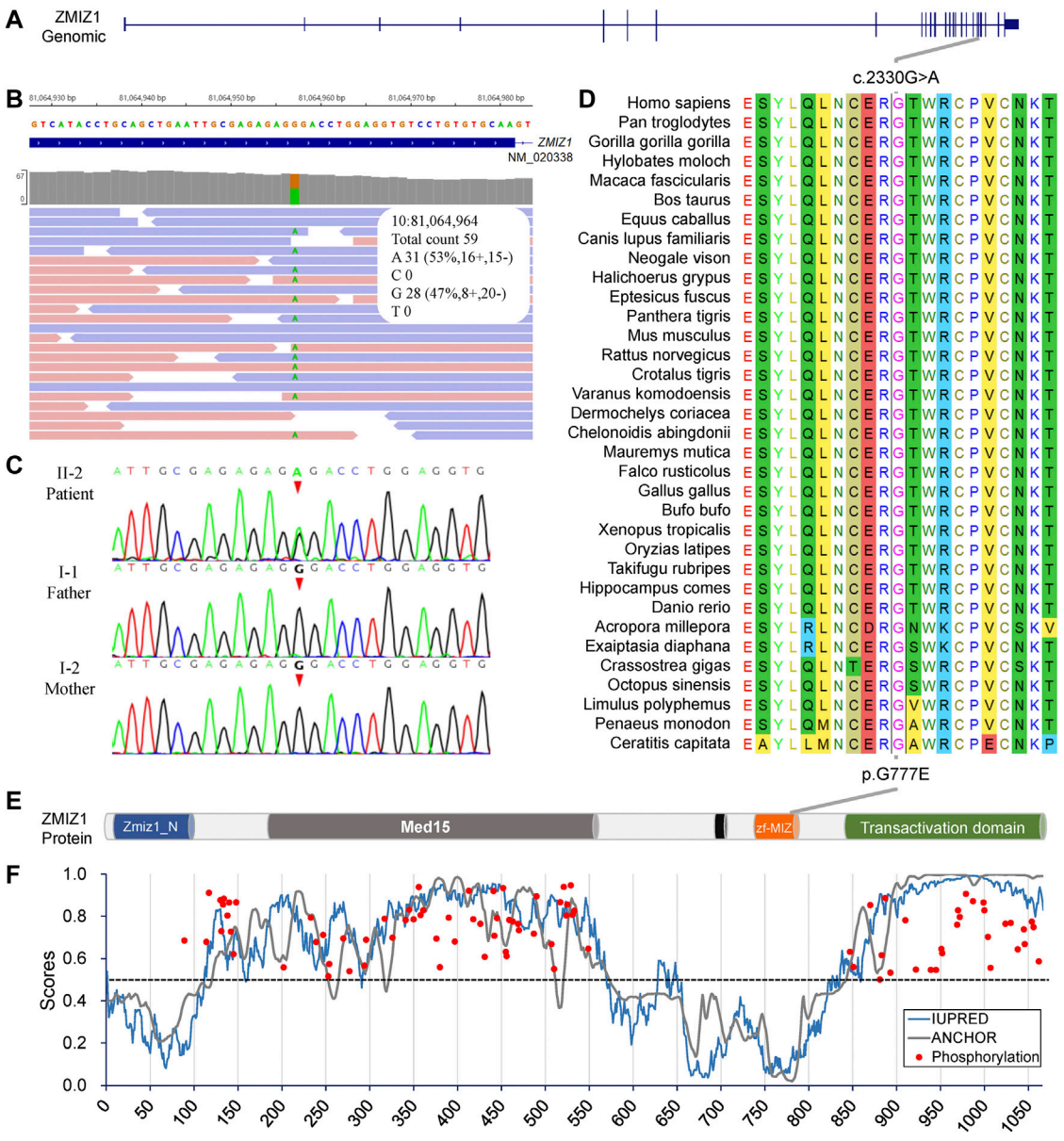

FIGURE 3 | Molecular analysis of the c.2330G > A (p.G777E) in ZMIZ1 gene. **(A)** Gene structure of ZMIZ1; **(B)** IGV view of the c.2330G > A identified by WES; **(C)** Sanger sequencing of the c.2330G > A variant; **(D)** evolutionary conservation analysis; **(E)** protein structure of ZMIZ1; **(F)** analysis for intrinsically disordered regions and phosphorylation sites.

coding sequences of *ZMIZ1* (n = 29) and chromosomal translocations involving ZMIZ1 (n = 3) (**Supplementary File S2**). Among these patients, except for c.1491+2T > C identified in a Thai female, the remaining 31 variants were detected in patients with Caucasian origin in Western countries. In our small cohort of NEDD/ID cases in China, a *de novo* missense pathogenic variant c.2330G > A (p.G777E) was detected in a 5-year-old girl. This patient presented the characteristic clinical phenotypes of NEDDFSA, such as neurodevelopmental delay, mild intellectual disability, hypotonia, language delay, dysmorphic facial features, joint hypermobility, and hand and foot anomalies, which were the common features of NEDDFSA (**Figure 5D**, **Supplementary Table S2**). As far as we know, this was the first report of *ZMIZ1* variant in Chinese. Besides, we also identified four other rare variants in the ZMIZ1 gene (c.540 + 20T > C, c.679G > A, c.1024A > G, and c.3096 + 15C > T) (**Supplementary Table S3**). Although predicted as "neutral" or "benign" to the function of ZMIZ1, it still could not rule out their pathogenicity. Cellular and animal experiments should be taken to verify the function of these variants, including the c.2330G > A (p.G777E).

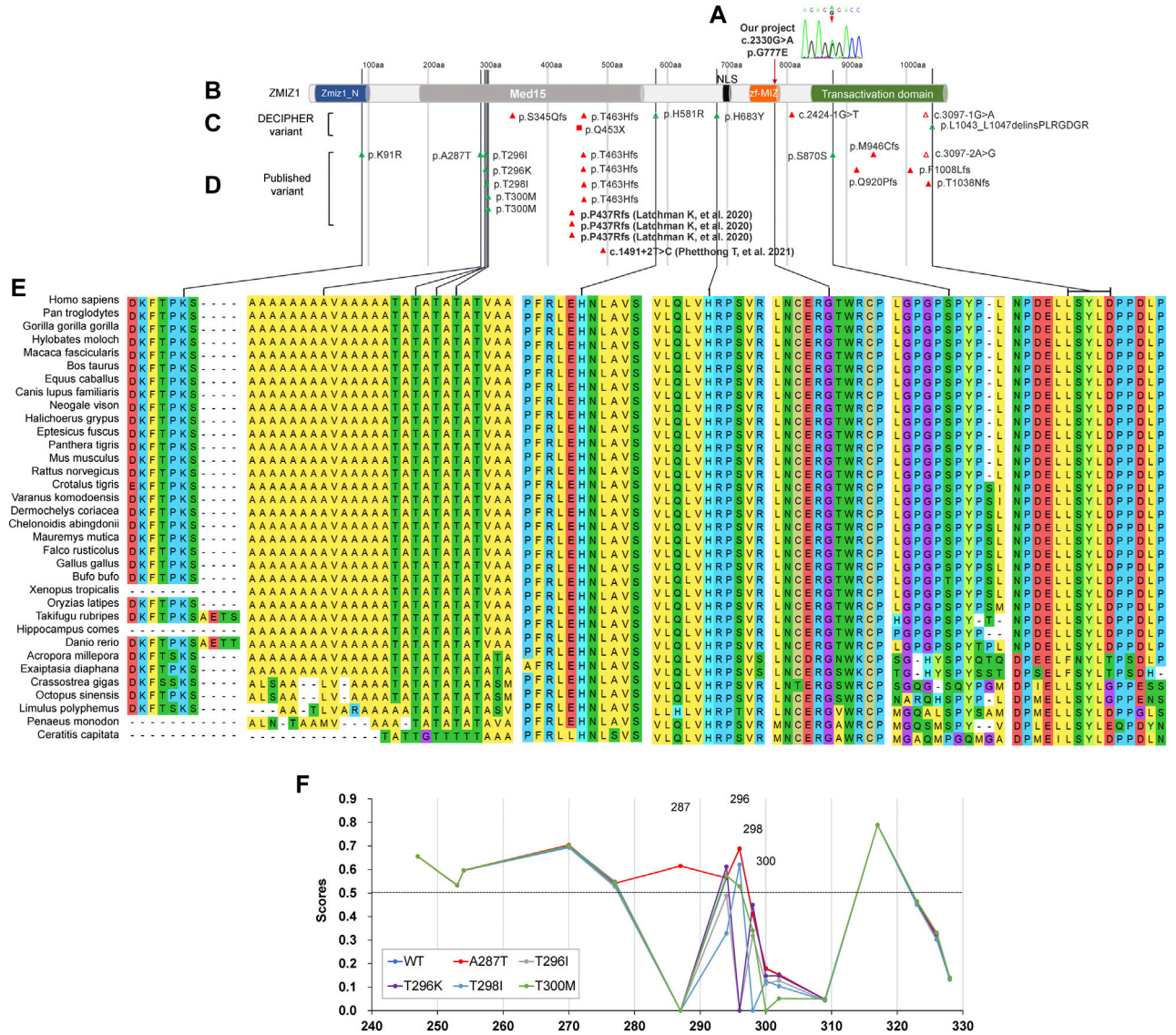

FIGURE 4 | Molecular analysis of variants of ZMIZ1. **(A)** Variant in our cohort; **(B)** diagram of ZMIZ1 protein; **(C)** variants in DECIPHER; **(D)** variants in reported articles; **(E)** evolutionary conservation; **(F)** phosphorylation analysis in the alanine-rich region.

ZMIZ1 was previously known as ZIMP10, RAI17, or KIAA1224. In 1999, Nagase et al. identified the gene *ZMIZ1* (previously called KIAA1224) from a fetal brain cDNA library (Nagase et al., 1999). According to the human tissue-specific transcriptomes, it was expressed in the heart, thyroid, immune cells, ovary, retina, and brain, with the highest in the pineal gland (Su et al., 2004). The encoded protein is a transcriptional co-activator, which belongs to the Protein Inhibitor of Activated STAT (PIAS) family. As a member of the PIAS family, ZMIZ1 has a highly conserved MIZ (Msx-interacting zinc finger) domain which is important for protein-protein interaction and SUMOylation (Sharma et al., 2003; Beliakoff and Sun, 2006). It had been reported that ZMIZ1 could regulate the activity of many transcription factors, such as androgen receptor (AR) (Beliakoff and Sun, 2006), SMAD3 (Li et al., 2006), SMAD4 (Li et al., 2006), and p53 (Lee et al., 2007). As an ortholog of ZMIZ1, tonalli (tna) was identified in *Drosophila melanogaster* and interacted with the ATP-dependent SWI/SNF complexes, which suggested a potential role in chromatin remodeling (Gutiérrez et al., 2003). Recently, ZMIZ1 was identified to be interacted with BRG1 (SMARCA4) (Li et al., 2011), BAF57 (SMARCE1) (Li et al., 2011), or SATB1 (Pinnell et al., 2015) to regulate the chromatin remodeling in humans. The protein-protein interaction network showed that ZMIZ1 could interact with SMAD3, SMAD4, MYC, NOTCH1, RBPJ, SMARCA4, ETS1, and UBE2I. According to the GO analysis for the 9 members of the protein network containing ZMIZ1, the network was significantly involved in mesenchyme

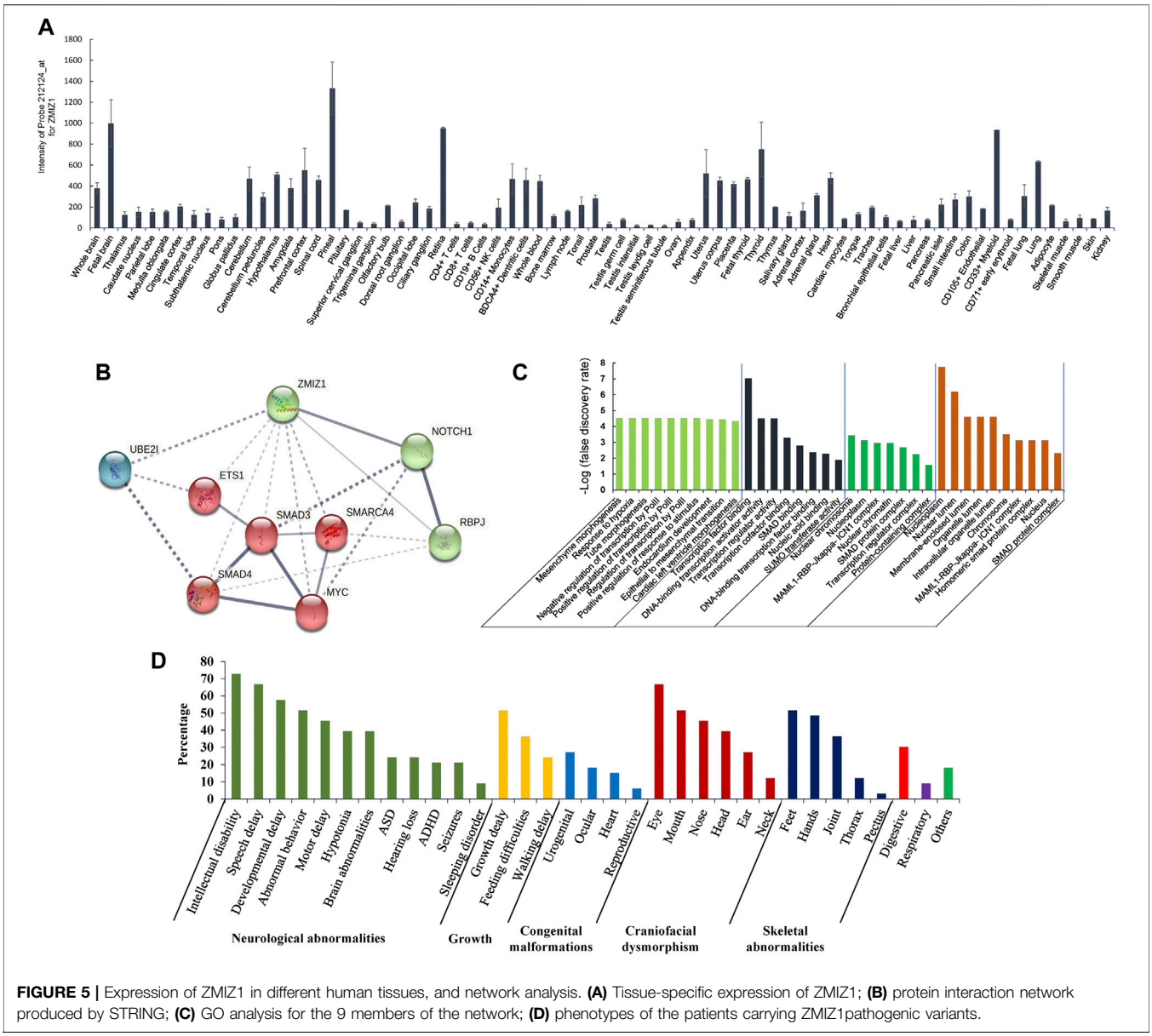

FIGURE 5 | Expression of ZMIZ1 in different human tissues, and network analysis. **(A)** Tissue-specific expression of ZMIZ1; **(B)** protein interaction network produced by STRING; **(C)** GO analysis for the 9 members of the network; **(D)** phenotypes of the patients carrying ZMIZ1pathogenic variants.

morphogenesis, hypoxia, tube morphogenesis, regulation of transcription, response to stimulus, endocardium development, epithelial to mesenchymal transition, and cardiac left ventricle morphogenesis. This explained why *ZMIZ1* pathogenic variant could affect the normal development of multiple systems, such as nerve, heart, and bones. GO analysis also showed that members of this network were localized in the nucleus to form two multiple protein complexes, mainly MAML1-RBP-Jκ-ICN1 complex and SMAD protein complex, to regulate the expression of target genes. The proper expression of *ZMIZ1* was essential for the standard embryonic development. It has been revealed in mice embryos at different stages that *ZMIZ1* was expressed dynamically in the neural tissues, craniofacial tissues, mandibular, foregut, limb buds, optic vesicle and otic pit, and somite (Beliakoff et al., 2008; Rodriguez-Magadán et al., 2008). This was consistent with the above-mentioned clinical features produced by the mutant ZMIZ1.

After compiling all the *ZMIZ1* variants in the DECIPHER database, published articles, and our cohort (**Supplementary Table S2**), 12 patients were found to carry amino-acid changing variants, and half of them (6/12) had variants in the alanine-rich sequence. The alanine-rich low-complexity region (LCR) was localized in the N-terminal intrinsic disordered region of ZMIZ1. The alanine-rich sequences were extremely conserved in different species during evolution, suggesting its importance for the proper function of ZMIZ1. According to the reports, many transcription factors or transcription mediators, such as FUS (FUS RNA binding protein), EWSR1 (EWS RNA binding protein 1), TAF15 (TATA-box binding protein associated factor 15), Sp1 (Sp1 transcription factor), and

AR could interact with ZMIZ1 at the transcriptional start sites via their extremely low-complexity regions (LCRs) to form local phase-separated condensates (or called droplets) to stabilize DNA binding, recruit RNA polymerase II (RNA Pol II), and activate transcription (Chong et al., 2018; Zamudio et al., 2019). These special condensates were a trade-off between proper functionality and risk of abnormal aggregation. The aberrant phase transitions within liquid-like droplets lie at the heart of many kinds of diseases, such as TATA box-binding protein (TBP, OMIM#600075) for spinocerebellar ataxia 17 (SCA17, OMIM#607136) (Friedman et al., 2007), FUS (OMIM#137070) for amyotrophic lateral sclerosis 6 (ALS6, OMIM#608030) (Patel et al., 2015), and androgen receptor (AR, OMIM#313700) for spinal and bulbar muscular atrophy (SBMA, OMIM#313200). As predicted by IUPred2A, ZMIZ1 contained three low-complexity regions (one alanine-rich and two proline-rich regions). It is reasonable that the alanine-rich region might be indispensable for the phase separation of ZMIZ1 to carry out the transcription mediator function. As predicted, the variants could change the phosphorylation pattern in the alanine-rich region, which might affect the local conformation. This might be the underlying molecular mechanism for the alanine-rich region being the variation hotspot of ZMIZ1. However, this has not yet been experimentally verified.

CONCLUSION

In conclusion, a *de novo* missense variant was first discovered in a Chinese female with a rare heterozygous syndromic disease, neurodevelopmental disorder with dysmorphic facies, and distal skeletal anomalies (NEDDFSA). Currently, a total of 32 patients with 27 types of variants of *ZMIZ1* (24 in protein-coding sequences and 3 translocations) have been identified globally. However, the underlying molecular mechanism of these variants has not been elucidated. Further experimental studies should be carried out to clarify these unknown fields to determine potential drug targets for the treatment of NEDDFSA.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding authors.

ETHICS STATEMENT

This study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. This study was approved by the Ethical Committee of the Shenzhen Baoan Women's and Children's Hospital. We obtained written informed consent from all individual members of the study. For minors, written informed consent has been obtained from their parents or legal guardians.

AUTHOR CONTRIBUTIONS

GL, LM, ZY, and QP conceived of the project and wrote the manuscript. GL analyzed the data. PX, BX, LW, JD, HX, and WD collected samples, consulted with patients, and extracted genomic DNAs. QP conducted genetic counseling, analyzed the clinical data, and revised the manuscript.

FUNDING

This work was funded by grants from the Shenzhen Science and Technology Innovation Commission (JCYJ20180305164359668 to QP), Shenzhen Municipal Health Commission (SZBC2018020 to QP), Project of Sichuan Provincial Department of Science and Technology (2019YFS0443), Key Research and Development Project of Deyang City's Science and Technology Bureau (2021SZ003 to GL), and Special Fund for Incubation Projects of Deyang People's Hospital (FHG202004 to GL).

ACKNOWLEDGMENTS

The authors would like to thank the patient and parents for their cooperation and consent to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.840577/full#supplementary-material

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* 7, 248–249. doi:10.1038/nmeth0410-248

Amendola, L. M., Dorschner, M. O., Robertson, P. D., Salama, J. S., Hart, R., Shirts, B. H., et al. (2015). Actionable Exomic Incidental Findings in 6503 Participants: Challenges of Variant Classification. *Genome Res.* 25, 305–315. doi:10.1101/gr.183483.114

Beliakoff, J., and Sun, Z. (2006). Zimp7 and Zimp10, Two Novel PIAS-like Proteins, Function as Androgen Receptor Coregulators. *Nucl. Recept Signal.* 4, e017. doi:10.1621/nrs.04017

Beliakoff, J., Lee, J., Ueno, H., Aiyer, A., Weissman, I. L., Barsh, G. S., et al. (2008). The PIAS-like Protein Zimp10 Is Essential for Embryonic Viability and Proper Vascular Development. *Mol. Cel Biol* 28, 282–292. doi:10.1128/mcb.00771-07

Bhagwat, M. (2010). Searching NCBI's dbSNP Database. *Curr. Protoc. Bioinformatics* Chapter 1, Unit 1–19. doi:10.1002/0471250953.bi0119s32

Carapito, R., Ivanova, E. L., Morlon, A., Meng, L., Molitor, A., Erdmann, E., et al. (2019). ZMIZ1 Variants Cause a Syndromic Neurodevelopmental Disorder. *Am. J. Hum. Genet.* 104, 319–330. doi:10.1016/j.ajhg.2018.12.007

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One* 7, e46688. doi:10.1371/journal.pone.0046688

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., et al. (2018). Imaging Dynamic and Selective Low-Complexity Domain Interactions that Control Gene Transcription. *Science* 361, eaar2555. doi:10.1126/science.aar2555

Córdova-Fletes, C., Domínguez, M. G., Delint-Ramirez, I., Martínez-Rodríguez, H. G., Rivas-Estilla, A. M., Barros-Núñez, P., et al. (2015). A De Novo t(10; 19)(q22.3;q13.33) Leads to ZMIZ1/PRR12 Reciprocal Fusion Transcripts in a Girl with Intellectual Disability and Neuropsychiatric Alterations. *Neurogenetics* 16, 287–298. doi:10.1007/s10048-015-0452-2

Erdos, G., and Dosztanyi, Z. (2020). Analyzing Protein Disorder with IUPred2A. *Curr. Protoc. Bioinformatics* 70, e99. doi:10.1002/cpbi.99

Friedman, M. J., Shah, A. G., Fang, Z.-H., Ward, E. G., Warren, S. T., Li, S., et al. (2007). Polyglutamine Domain Modulates the TBP-TFIIB Interaction: Implications for its normal Function and Neurodegeneration. *Nat. Neurosci.* 10, 1519–1528. doi:10.1038/nn2011

Gutiérrez, L., Zurita, M., Kennison, J. A., and Vázquez, M. (2003). The Drosophila Trithorax Group Gene Tonalli (Tna) Interacts Genetically with the Brahma Remodeling Complex and Encodes an SP-RING finger Protein. *Development* 130, 343–354. doi:10.1242/dev.00222

Huang, J., Poon, L. C., Akolekar, R., Choy, K. W., Leung, T. Y., and Nicolaides, K. H. (2014). Is High Fetal Nuchal Translucency Associated with Submicroscopic Chromosomal Abnormalities on Array CGH. *Ultrasound Obstet. Gynecol.* 43, 620–624. doi:10.1002/uog.13384

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., et al. (2006). The PROSITE Database. *Nucleic Acids Res.* 34, D227–D230. doi:10.1093/nar/gkj063

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., et al. (2004). Detection of Large-Scale Variation in the Human Genome. *Nat. Genet.* 36, 949–951. doi:10.1038/ng1416

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., et al. (2004). The Importance of Intrinsic Disorder for Protein Phosphorylation. *Nucleic Acids Res.* 32, 1037–1049. doi:10.1093/nar/gkh253

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., and Mega, X. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096

Latchman, K., Calder, M., Morel, D., Rhodes, L., Juusola, J., and Tekin, M. (2020). Autosomal Dominant Inheritance in a Recently Described ZMIZ1- Related Neurodevelopmental Disorder: Case Report of Siblings and an Affected Parent. *Am. J. Med. Genet.* 182, 548–552. doi:10.1002/ajmg.a.61446

Lee, J., Beliakoff, J., and Sun, Z. (2007). The Novel PIAS-like Protein hZimp10 Is a Transcriptional Co-activator of the P53 Tumor Suppressor. *Nucleic Acids Res.* 35, 4523–4534. doi:10.1093/nar/gkm476

Lek, M., Karczewski, K. J., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., et al. (2016). Analysis of Protein-Coding Genetic Variation in 60,706 Humans. *Nature* 536, 285–291. doi:10.1038/nature19057

Leung, T. Y., Vogel, I., Lau, T. K., Chong, W., Hyett, J. A., Petersen, O. B., et al. (2011). Identification of Submicroscopic Chromosomal Aberrations in Fetuses with Increased Nuchal Translucency and Apparently normal Karyotype. *Ultrasound Obstet. Gynecol.* 38, 314–319. doi:10.1002/uog.8988

Li, X., Thyssen, G., Beliakoff, J., and Sun, Z. (2006). The Novel PIAS-like Protein hZimp10 Enhances Smad Transcriptional Activity. *J. Biol. Chem.* 281, 23748–23756. doi:10.1074/jbc.m508365200

Li, X., Zhu, C., Tu, W. H., Yang, N., Qin, H., and Sun, Z. (2011). ZMIZ1 Preferably Enhances the Transcriptional Activity of Androgen Receptor with Short Polyglutamine Tract. *PLoS One* 6, e25040. doi:10.1371/journal.pone.0025040

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the Conserved Domain Database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi:10.1093/nar/gkz991

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi:10.1093/nar/gkaa913

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: Improving Coverage, Classification and Access to Protein Sequence Annotations. *Nucleic Acids Res.* 47, D351–D360. doi:10.1093/nar/gky1100

Nagase, T., Ishikawa, K., Kikuno, R., Hirosawa, M., Nomura, N., and Ohara, O. (1999). Prediction of the Coding Sequences of Unidentified Human Genes.XV. The Complete Sequences of 100 New cDNA Clones from Brain Which Code for Large Proteins *In Vitro*. *DNA Res.* 6, 337–345. doi:10.1093/dnares/6.5.337

Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., et al. (2021). The UCSC Genome Browser Database: 2021 Update. *Nucleic Acids Res.* 49, D1046–D1057. doi:10.1093/nar/gkaa1070

Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., et al. (2015). A Liquid-To-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162, 1066–1077. doi:10.1016/j.cell.2015.07.047

Pedruzzi, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., de Castro, E., et al. (2015). HAMAP in 2015: Updates to the Protein Family Classification and Annotation System. *Nucleic Acids Res.* 43, D1064–D1070. doi:10.1093/nar/gku1002

Phetthong, T., Khongkrapan, A., Jinawath, N., Seo, G. H., and Wattanasirichaigoon, D. (2021). Compound Heterozygote of Point Variant and Chromosomal Microdeletion Involving OTUD6B Coinciding with ZMIZ1 Variant in Syndromic Intellectual Disability. *Genes (Basel)* 12, 1583. doi:10.3390/genes12101583

Pinnell, N., Yan, R., Cho, H. J., Keeley, T., Murai, M. J., Liu, Y., et al. (2015). The PIAS-like Coactivator Zmiz1 Is a Direct and Selective Cofactor of Notch1 in T Cell Development and Leukemia. *Immunity* 43, 870–883. doi:10.1016/j.immuni.2015.10.007

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi:10.1038/gim.2015.30

Rodriguez-Magadán, H., Merino, E., Schnabel, D., Ramírez, L., and Lomelí, H. (2008). Spatial and Temporal Expression of Zimp7 and Zimp10 PIAS-like Proteins in the Developing Mouse Embryo. *Gene Expr. Patterns* 8, 206–213. doi:10.1016/j.modgep.2007.10.005

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., et al. (2021). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 49, D10–D17. doi:10.1093/nar/gkaa892

Scheps, K. G., Hasenahuer, M. A., Parisi, G., Targovnik, H. M., and Fornasari, M. S. (2020). Curating the gnomAD Database: Report of Novel Variants in the Globin-coding Genes and Bioinformatics Analysis. *Hum. Mutat.* 41, 81–102. doi:10.1002/humu.23925

Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., et al. (2021). Mol* Viewer: Modern Web App for 3D Visualization and Analysis of Large Biomolecular Structures. *Nucleic Acids Res.* 49, W431–W437. doi:10.1093/nar/gkab314

Sharma, M., Li, X., Wang, Y., Zarnegar, M., Huang, C. Y., Palvimo, J. J., et al. (2003). hZimp10 Is an Androgen Receptor Co-activator and Forms a Complex with SUMO-1 at Replication Foci. *EMBO J.* 22, 6101–6114. doi:10.1093/emboj/cdg585

Siva, N. (2008). 1000 Genomes Project. *Nat. Biotechnol.* 26, 256. doi:10.1038/nbt0308-256b

Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., et al. (2004). A Gene Atlas of the Mouse and Human Protein-Encoding Transcriptomes. *Proc. Natl. Acad. Sci.* 101, 6062–6067. doi:10.1073/pnas.0400782101

Swaminathan, G. J., Bragin, E., Chatzimichali, E. A., Corpas, M., Bevan, A. P., Wright, C. F., et al. (2012). DECIPHER: Web-Based, Community Resource for Clinical Interpretation of Rare Variants in Developmental Disorders. *Hum. Mol. Genet.* 21, R37–R44. doi:10.1093/hmg/dds362

Taliun, D., Harris, D. N., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., et al. (2021). Sequencing of 53,831 Diverse Genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299. doi:10.1038/s41586-021-03205-y

Vogel-Ciernia, A., Matheos, D. P., Barrett, R. M., Kramár, E. A., Azzawi, S., Chen, Y., et al. (2013). The Neuron-specific Chromatin Regulatory Subunit BAF53b Is Necessary for Synaptic Plasticity and Memory. *Nat. Neurosci.* 16, 552–561. doi:10.1038/nn.3359

Vogel-Ciernia, A., and Wood, M. A. (2014). Neuron-specific Chromatin Remodeling: a Missing Link in Epigenetic Mechanisms Underlying Synaptic Plasticity, Memory, and Intellectual Disability Disorders. *Neuropharmacology* 80, 18–27. doi:10.1016/j.neuropharm.2013.10.002

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Wu, J. I., Lessard, J., Olave, I. A., Qiu, Z., Ghosh, A., Graef, I. A., et al. (2007). Regulation of Dendritic Development by Neuron-specific Chromatin Remodeling Complexes. *Neuron* 56, 94–108. doi:10.1016/j.neuron.2007.08.021

Yang, J., and Zhang, Y. (2015). I-TASSER Server: New Development for Protein Structure and Function Predictions. *Nucleic Acids Res.* 43, W174–W181. doi:10.1093/nar/gkv342

Zamudio, A. V., Dall'Agnese, A., Henninger, J. E., Manteiga, J. C., Afeyan, L. K., Hannett, N. M., et al. (2019). Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol. Cel* 76, 753–766. doi:10.1016/j.molcel.2019.08.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 21 April 2022
doi: 10.3389/fgene.2022.880997

Identifying Key MicroRNA Signatures for Neurodegenerative Diseases With Machine Learning Methods

ZhanDong Li[1†], Wei Guo[2†], ShiJian Ding[3†], Lei Chen[4], KaiYan Feng[5], Tao Huang[6,7] and Yu-Dong Cai[3]**

[1]College of Food Engineering, Jilin Engineering Normal University, Changchun, China, [2]Key Laboratory of Stem Cell Biology, Shanghai Jiao Tong University School of Medicine (SJTUSM) and Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China, [3]School of Life Sciences, Shanghai University, Shanghai, China, [4]College of Information Engineering, Shanghai Maritime University, Shanghai, China, [5]Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou, China, [6]Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, [7]CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China*

OPEN ACCESS

Edited by:
Liang Cheng,
Harbin Medical University, China

Reviewed by:
Jing Lu,
Yantai University, China
Yue Zhang,
Harbin Engineering University, China

***Correspondence:**
Tao Huang
tohuangtao@126.com
Yu-Dong Cai
cai_yud@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: *22 February 2022*
Accepted: *30 March 2022*
Published: *21 April 2022*

Citation:
Li Z, Guo W, Ding S, Chen L, Feng K, Huang T and Cai Y-D (2022) Identifying Key MicroRNA Signatures for Neurodegenerative Diseases With Machine Learning Methods. Front. Genet. 13:880997. doi: 10.3389/fgene.2022.880997

Neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson's disease, and many other disease types, cause cognitive dysfunctions such as dementia via the progressive loss of structure or function of the body's neurons. However, the etiology of these diseases remains unknown, and diagnosing less common cognitive disorders such as vascular dementia (VaD) remains a challenge. In this work, we developed a machine-leaning-based technique to distinguish between normal control (NC), AD, VaD, dementia with Lewy bodies, and mild cognitive impairment at the microRNA (miRNA) expression level. First, unnecessary miRNA features in the miRNA expression profiles were removed using the Boruta feature selection method, and the retained feature sets were sorted using minimum redundancy maximum relevance and Monte Carlo feature selection to provide two ranking feature lists. The incremental feature selection method was used to construct a series of feature subsets from these feature lists, and the random forest and PART classifiers were trained on the sample data consisting of these feature subsets. On the basis of the model performance of these classifiers with different number of features, the best feature subsets and classifiers were identified, and the classification rules were retrieved from the optimal PART classifiers. Finally, the link between candidate miRNA features, including hsa-miR-3184-5p, has-miR-6088, and has-miR-4649, and neurodegenerative diseases was confirmed using recently published research, laying the groundwork for more research on miRNAs in neurodegenerative diseases for the diagnosis of cognitive impairment and the understanding of potential pathogenic mechanisms.

Keywords: neurodegenerative disease, microRNA, feature selection, expression pattern, classification algorithm

1 INTRODUCTION

Dementia is one kind of cognitive impairment that is characterized by difficulties in memory, language, and behavior. Of all chronic diseases, dementia has become one of the most important contributors to dependence and disability (Iliffe et al., 2009). With an increasing number of morbidity, dementia has become a great concern worldwide (Prince et al., 2016). Unfortunately, there is no cure for this disease at present, and earlier diagnosis and interventions to slow down the disease progress are needed (Iliffe et al., 2009). Therefore, researchers have focused on searching effective diagnostic methods, including the identification of new biomarkers for diagnosis, and interventions for dementia.

Although young-onset cases are increasingly recognized, dementia is typically a condition that affects older people. Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of intellectual deficit in populations older than 65 years. More than 20% of individuals over 80 years of age are affected by AD, and epidemiological data predict that there will be over 35 million AD patients by 2050 (Danborg et al., 2014). Other less common causes of cognitive impairment include vascular dementia (VaD) whose definition and distinction remain controversial, mixed dementia, and dementia with Lewy bodies (DLB) (Mckeith et al., 1996). Diagnosing dementia is markedly difficult due to its insidious onset and diversity of other presenting symptoms such as difficulty in making decisions (Kostopoulou et al., 2008). Recent studies have reported that certain protein biomarkers in cerebrospinal fluid (CSF) can be applied in the clinical diagnosis of AD with a high predictive accuracy (De Meyer et al., 2010). However, such biomarkers have their limitations in differentiating AD from other types of dementia. In addition, biomarkers in CSF require an invasive collection process; thus, new methods through less invasive procedures are needed. Considering that the diagnosis of dementia subtypes is important to manage different therapies, disease courses, and outcomes for different dementias (Robinson et al., 2015), development of better biomarkers for AD and other dementias will contribute to more accurate diagnosis for an early and specialized treatment.

For a better clinical care in disease prevention and treatment, several computational models have been developed to predict dementia risk or subtypes (Stephan et al., 2010). For example, Licher et al. (2019) reported a dementia risk model using optimism-corrected C-statistics, which can be used to identify individuals with high risk of dementia with an accuracy of 0.86. This model was based on comprehensive clinical information such as age, cognitive impairment, and lifestyle factors. Interestingly, a novel machine learning prediction model for dementia risk identification using the voice data from daily conversations was proposed by Shimoda et al. (2021). They applied three strategies including extreme gradient boosting, random forest (RF), and logistic regression methods in developing models, which had AUCs of 0.86, 0.88, and 0.89, respectively. Li et al. (2019) reported a deep learning model for the early prediction of AD using hippocampal magnetic resonance imaging data, which achieved a concordance index of 0.762. In addition, genetic data were taken into account to improve the ability of the prediction model given that many genes were confirmed to be associated with AD (Seshadri et al., 2010). So far, models in dementia prediction lack molecular signatures such as transcriptional expression, which can reflect the underlying pathogenic mechanisms.

MicroRNAs (miRNAs) are small non-coding RNA molecules of approximately 22 nucleotides in length, which have been shown to regulate gene expression by binding to complementary regions of messenger transcripts (Lagos-Quintana et al., 2001). The detection of circulating miRNA levels has been proposed to be a potential diagnostic tool for a number of diseases (Gilad et al., 2008). MiRNAs play a crucial role in the control of neuronal cell development (Mistur et al., 2009). The alteration of the expression of some miRNAs has been shown to relate to various neurological diseases including AD. For example, miR-137, miR-181c, and miR-29a/b were reported to be involved in AD by modulating ceramide levels (Geekiyanage and Chan, 2011). The downregulation of miR-16, miR-195, and miR-103 was observed in the brain of AD patients, and these miRNAs were shown to target the β-site amyloid precursor protein cleaving enzyme 1 (*BACE1*), which is involved in amyloid plaque formation (Bekris et al., 2013). Cogswell et al. found significantly decreased expression of miR-9, which regulates neuronal differentiation, in the human hippocampus of AD patients (Cogswell et al., 2008; Coolen et al., 2013). Different expression patterns of miRNAs have also been found between AD and other neurodegenerative diseases; for example, miR-15a is uniquely elevated in the plasma of AD patients (Bekris et al., 2013). Therefore, miRNAs in the blood or serum are easily accessible and noninvasive biomarkers for diagnosing dementia. In addition, some miRNAs can be used to distinguish different subtypes of dementia for more precise treatment.

In this study, on the basis of the miRNA expression profiles from 1601 serum samples (Shigemizu et al., 2019a), including AD cases, VaD cases, DLB cases, mild cognitive impairment (MCI) cases, and normal controls (NC), we computationally analyzed such expression data. The data was first analyzed by Boruta (Kursa and Rudnicki, 2010), irrelevant miRNA features were excluded. Remaining miRNA features were evaluated by minimum redundancy maximum relevance (mRMR) (Peng et al., 2005) and Monte Carlo feature selection (MCFS) (Dramiński et al., 2007), respectively. Two feature lists were generated, which were fed into incremental feature selection (IFS) (Liu and Setiono, 1998), incorporating random forest (RF) (Breiman, 2001) or PART (Frank and Witten, 1998). As a result, we identified the crucial miRNAs that show the most relevance to the distinction of four different types of dementia and NC, suggesting that these selected miRNAs may play crucial roles in neuronal development. Furthermore, we also identified interesting classification rules, which suggested different miRNA expression patterns on different dementia subtypes and NC. These results can guide further research about the interaction between miRNAs and neurodegenerative diseases. Finally, we constructed two optimal classifiers with high accuracy to group individuals into the corresponding categories (four dementia

TABLE 1 | Sample size for normal control and four neurodegenerative diseases.

Disease case	Sample size
Alzheimer's disease (AD)	1,021
Vascular dementia (VaD)	91
Dementia with lewy bodies (DLB)	169
Mild cognitive impairment (MCI)	32
Normal control (NC)	288

subtypes and NC). They can be useful tools for the precise diagnosis of dementia subtypes. Our study highlights the potential application of miRNAs in dementia subtype diagnosis, indicating that the prediction framework using serum miRNA expression data can provide feasible therapeutic and diagnostic targets for dementia.

2 MATERIALS AND METHODS

2.1 Dataset

In this study, the miRNA expression profiles were obtained from the Gene Expression Omnibus database under the accession code GSE120584 (Shigemizu et al., 2019a; Shigemizu et al., 2019b; Asanomi et al., 2021). These expression profiles include 1,601 samples, which are composed of AD cases, VaD cases, DLB cases, MCI cases, and NC. The sample sizes of different cases are provided in **Table 1**. A total of 2547 miRNAs were identified in the expression profiles. Subsequently, we performed a computational workflow to detect key miRNA features and expression patterns in the expression profiles.

2.2 Boruta Feature Filtering

Aside from the time and energy costs of dealing with a high number of features, most machine learning algorithms work better when the number of predicting features employed is kept as small as possible. We thus applied a Boruta analysis on the miRNA expression profiles to reduce feature dimension and retain important miRNA features (Kursa and Rudnicki, 2010). Boruta is a feature selection approach based on the RF model to access feature importance (Z-score) by comparing the relevance of real features with shadow features, which are randomly shuffled from original features. The python application from https://github.com/scikit-learn-contrib/boruta_py with default parameters was used for Boruta feature selection in this analysis.

2.3 Feature Ranking

2.3.1 Minimum Redundancy Maximum Relevance

The mRMR algorithm (Peng et al., 2005) is an entropy-based feature selection method that calculates the mutual information (MI) between a group of features and class variable. The MI is defined as follows:

$$I(X;Y) = \iint p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \quad (1)$$

where $p(x, y)$ is the joint probability density function of X and Y, $p(x)$ and $p(y)$ are the marginal probability density functions of X and Y, respectively. In the mRMR method, the correlation (D) between features and target label and the redundancy (R) between features and other features are computed as follows:

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c), \quad (2)$$

where S is the selected features and $I(x_i; c)$ is the MI between feature x_i and the target label c.

$$R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j), \quad (3)$$

where $I(x_i, x_j)$ is the MI between feature x_i and feature x_j. To repeatedly add a new feature to a feature subset S, the following objective function is optimized:

$$\max \Phi(D, R), \ \Phi = D - R, \quad (4)$$

In this study, we used the mRMR program acquired from http://home.penglab.com/proj/mRMR/ to rank all the features obtained by Boruta analysis, resulting in an mRMR feature list.

2.3.2 Monte Carlo Feature Selection

The MCFS method (Dramiński et al., 2007) evaluates the feature importance by creating numerous decision trees. More specifically, for a dataset with M features, MCFS first randomly constructs s feature subsets with m features ($m << M$). For each feature subset, t decision trees are constructed using the bootstrap sampling method. Finally, $s \times t$ classification trees are constructed and evaluated. The RI score of feature g based on these classification trees is defined as follows:

$$RI_g = \sum_{\tau=1}^{st} (wAcc)^u \sum_{n_g(\tau)} IG\left(n_g(\tau)\right) \left(\frac{no.in\, n_g(\tau)}{no.in\, \tau} \right)^v \quad (5)$$

where $wAcc$ is the weight accuracy of the decision tree τ; $IG(n_g(\tau))$ denotes the gain information of node $n_g(\tau)$; $(no.in\, n_g(\tau))$ and $(no.in\, \tau)$ represent the number of samples of node $n_g(\tau)$ and the number of samples in tree τ, respectively; and u and v are parameters that are recommended to be 1. After MCFS processing, all features are ranked in a feature list in descending order of RI values. In this study, we applied the MCFS program developed by Draminski et al., which can be accessed at http://www.ipipan.eu/staff/m.draminski/mcfs.html, for feature sorting, and the parameters were set to default values. The obtained feature list was called MCFS feature list.

2.4 Incremental Feature Selection

In the previous analysis, the mRMR and MCFS feature ranking lists were obtained, but it was not possible to determine the optimal feature subsets for classifying disease cases. Thus, the IFS method (Liu and Setiono, 1998) was used in this study to identify the best number of features in a feature list for a specific classification algorithm. IFS first generates a series of feature subsets on the basis of a step size. For example, if the step size

equals to 1, the first feature subset includes one top-ranked feature, the second feature subset is made up of two top-ranked features, and so on. Then, the sample datasets represented by these feature subsets are trained by one classification algorithm (RF or PART in this study). The classifiers are evaluated by using 10-fold cross-validation (Kohavi, 1995; Tang and Chen, 2022; Yang and Chen, 2022). The evaluation metrics (e.g., Matthews correlation coefficient [MCC]) for each classifier with different number of features are obtained and used to plot IFS curves, where the X-axis is the number of features and the Y-axis is the evaluation metrics. In the end, the optimal feature subsets that achieves the best classification results are identified, and the optimal classifiers are built.

2.5 Classification Algorithms

2.5.1 RF

The RF (Breiman, 2001) is an ensemble learning algorithm that takes decision trees as the base learner. It first produces a number of training sets from the original dataset using a bootstrapping method with randomized put-back sampling. These training sets are then used to train the decision tree model individually, and the generated decision trees are formed into a forest. Lastly, the final result is determined by aggregating the voting results of many tree classifiers. As RF is powerful, it is always an important candidate for constructing efficient classifiers (Chen et al., 2017; Zhao et al., 2018; Chen et al., 2021; Li X. et al., 2022; Li Z. et al., 2022; Chen et al., 2022; Ding et al., 2022). In this study, the RF program in Weka (Frank et al., 2004) was employed with default parameters.

2.5.2 PART

In contrast to black-box models, such as RF, rule learning models may learn rules from data to make discriminations on unknown data, and these rules are commonly expressed in an IF–THEN structure, which clearly expresses the patterns existing in the data. PART is a rule-generating method that combines the Ripper and C4.5 approaches without the need for global optimization (Frank and Witten, 1998). It uses a separate-and-conquer technique to develop several partial decision trees, in which a rule is constructed each time. Then, the instances it covers are eliminated, and rules are created recursively for the remaining instances until the end. The PART program in WEKA was used with the default parameters in this investigation.

2.6 SMOTE

The distribution of samples under five cases is uneven, which may lead to the poor performance of the established classifiers. To address this issue, we applied SMOTE methods to increase the sample size of the minority class, which is an oversampling technique presented by Chawla et al. (2002). SMOTE generates synthetic samples randomly between samples of a minority class and their neighbors on the basis of the k-nearest neighbor concept. The SMOTE algorithm in Weka software was used to process the miRNA expression profiles in this investigation, resulting in an equal number of samples in each class. It was necessary to pointed out that SMOTE was only used in evaluating the performance of classifiers in the IFS method. Pseudo samples generated by SMOTE did not participate in the mRMR or MCFS methods as they can influence the feature selection results.

2.7 Performance Measurement

For the 10-fold cross-validation, we used the MCC as a predictive metric for the evaluation of classifiers. In this study, considering that the analyzed miRNA dataset includes multiple disease cases, the multi-categorical version of MCC (Gorodkin, 2004) was applied and calculated as follows:

$$MCC = \frac{cov(X,Y)}{\sqrt{cov(X,X)cov(Y,Y)}} \tag{6}$$

where the binary matrix X represents the prediction results, the binary matrix Y indicates the real class label, and $cov(X,Y)$ stands for the covariance of the two matrices. The MCC ranges from −1 to 1, with a value closer to 1 indicating stronger model performance.

To fully display the performance of classification models, we also calculated other measurements, including individual accuracy on each class and overall accuracy (ACC). For one class, its individual accuracy was defined as the proportion of correctly predicted samples in this class. The ACC was defined as the proportion of correctly predicted samples.

3 RESULTS

3.1 Feature Selection Results on miRNA Expression Profiles

A flow chart of the present study is illustrated in **Figure 1**. We started by removing unnecessary features using the Boruta feature selection method, and the 108 retained features are listed in **Supplementary Table S1**.

Then, using mRMR and MCFS, remaining 108 features were ranked according to feature importance, yielding two ranked feature lists (mRMR feature list and MCFS feature list), as shown in **Supplementary Table S1**. Top ten miRNA features in these two lists were investigated, as shown in **Figure 2**. Four miRNAs, including hsa-miR-3184-5p, hsa-miR-1227-5p, hsa-miR-3181, and hsa-miR-6088, appeared in the top 10 features yielded by two methods, highlighting their visibility and importance. The biological roles of these miRNA features will be explored in **Section 4**.

3.2 IFS Results on the mRMR Feature List

Based on the mRMR feature list, it was fed into the IFS method with a step size of 1, returning 108 feature subsets. For example, the first feature subset includes the first feature, the second feature subset includes the first two features, and so on. The RF and PART classifiers were trained using the sample set consisting of these feature subsets, and the performance was assessed using 10-fold cross-validation. Obtained measurements are provided in **Supplementary Table S2**. To clearly display the performance of classifiers on different feature subsets, an IFS curve was plotted

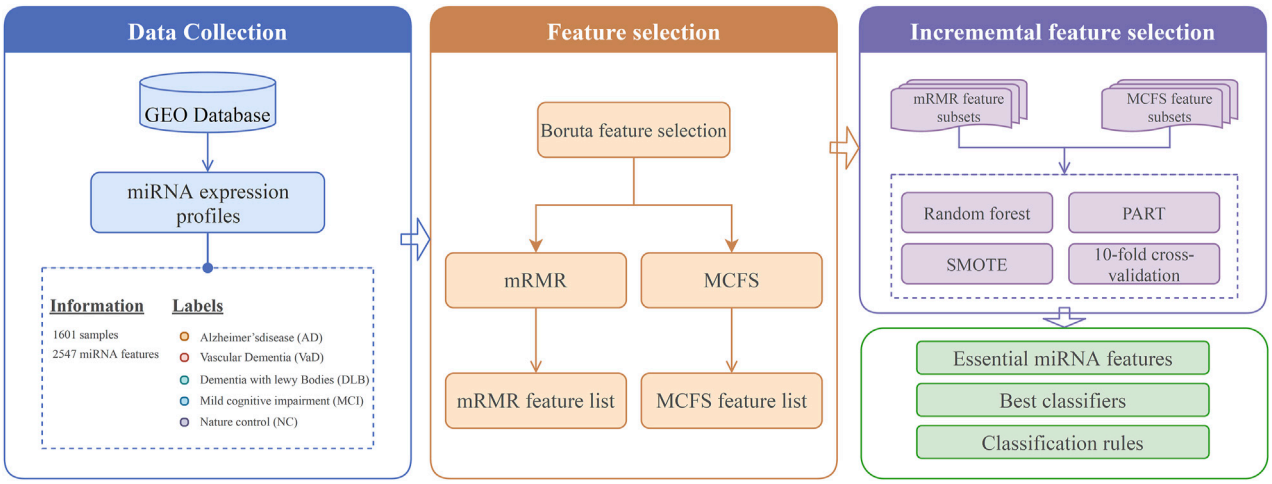

FIGURE 1 | Analysis flowchart for this study, which consists of three main steps: 1) miRNA dataset collection; 2) filtering and ranking of miRNA features in the dataset using Boruta, mRMR, and MCFS; 3) determining the essential miRNA features and building the best classifiers and classification rules using IFS method with RF and PART algorithms.

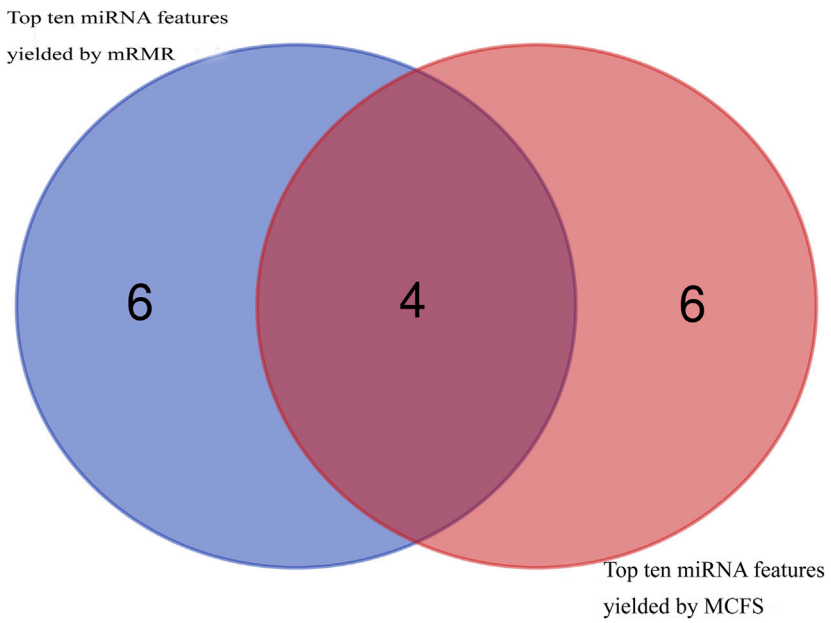

FIGURE 2 | Venn diagram to show top ten miRNA features obtained by mRMR and MCFS methods. Four miRNA features are commonly identified.

for each classification algorithm, which is shown in **Figure 3A**. When RF was selected as the classification algorithm in the IFS method, the highest MCC was 0.683, which was obtained by using top 106 features. Accordingly, the optimal RF classifier can be built with these features. The ACC of such classifier was 0.802, as listed in **Table 2**. As for PART, the highest MCC was 0.359. It was obtained by using top 72 features, with which the optimal PART classifier can be built. The ACC of such PART classifier was 0.570, as listed in **Table 2**. Clearly, the optimal PART classifier was much inferior to the optimal RF classifier. As for their performance on five classes, individual accuracies are shown in **Figure 4A**. Evidently, the optimal RF classifier provided better performance than the optimal PART classifier on all classes. Both MCI and VaD have an individual accuracy of over 0.900 in the optimal RF classifier.

Although the optimal RF classifier gave good performance, it was not very proper to do large-scale tests because lots of miRNA features involved. In view of this, we carefully checked the IFS

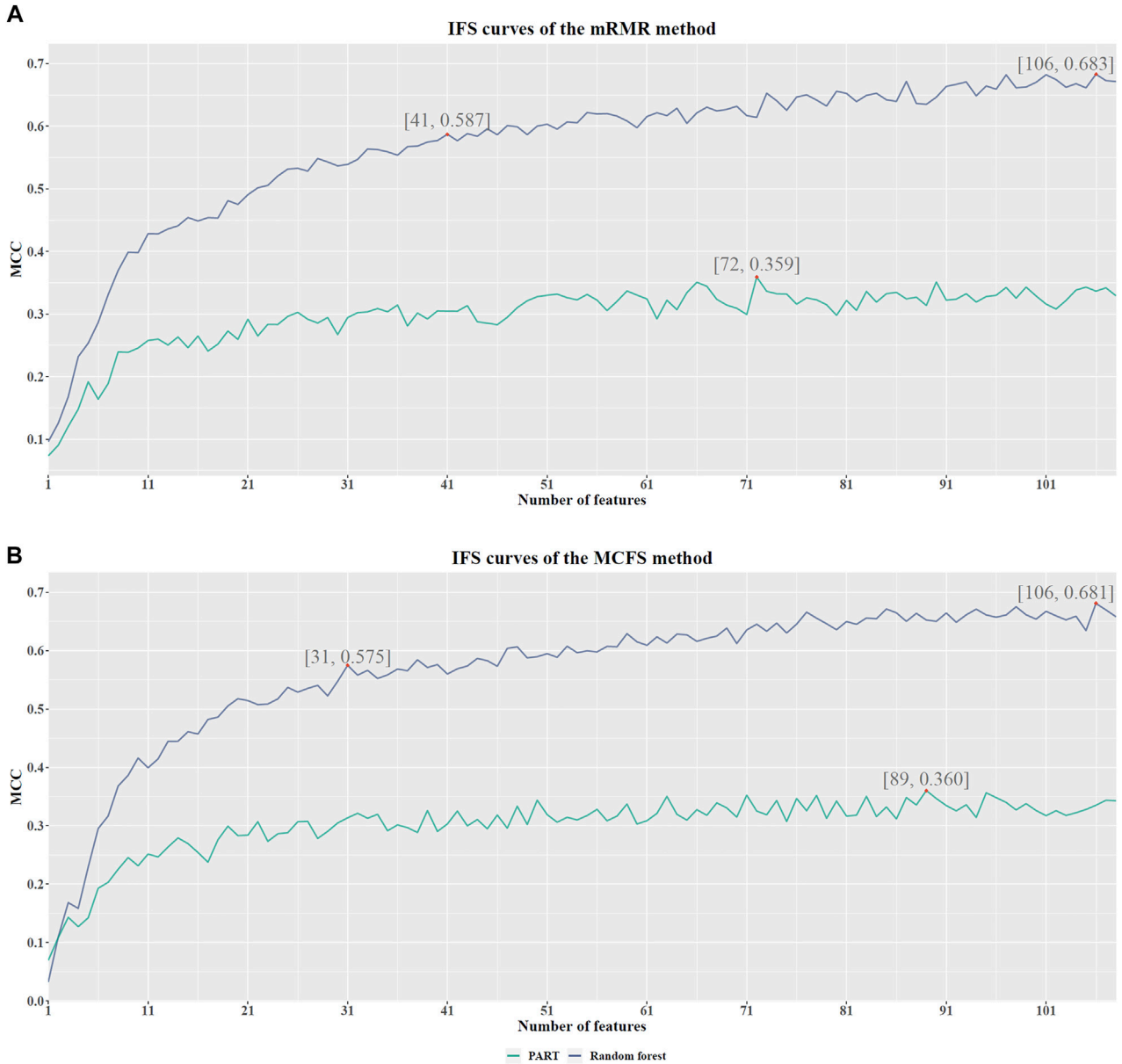

FIGURE 3 | IFS curves with different number of features in RF and PART under the mRMR and MCFS feature lists. **(A)**. With the mRMR feature list, RF reaches the highest point (MCC = 0.683) with the top 106 features, and PART obtains the highest MCC (0.359) when using the top 72 features. The RF with top 41 features also provides high performance (MCC = 0.587). **(B)**. With the MCFS feature list, RF and PART reach the highest points (MCC = 0.681 and 0.360, respectively) at the top 106 and 89 features. The RF with top 31 features also yields high performance (MCC = 0.575).

TABLE 2 | Performance of key classifiers with different algorithms based on the mRMR feature list.

Classification algorithm	Number of features	ACC	MCC
Random forest	106	0.802	0.683
Random forest	41	0.743	0.587
PART	72	0.570	0.359

results with RF and found that RF provided the MCC of 0.587 when top 41 features were used (**Figure 3A**). This classifier yielded the ACC of 0.743 (**Table 2**). Its performance on five classes is shown in **Figure 4A**. Although it provided lower performance than the optimal RF classifier, it was much faster as much less miRNA features were needed. This classifier can be an efficient tool to identify four dementia subtypes and NC.

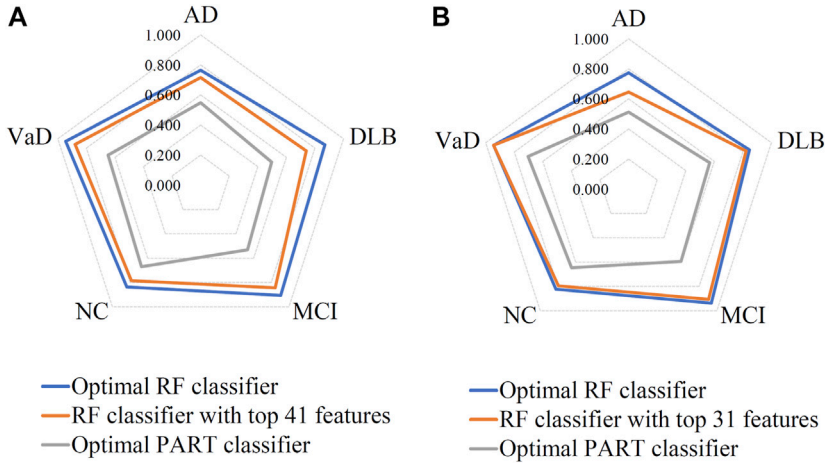

FIGURE 4 | Performance of the key RF and PART classifiers on each class based on mRMR **(A)** and MCFS **(B)** feature lists. AD, VaD, DLB, MCI, and NC stand for Alzheimer's disease, Vascular dementia, Dementia with Lewy bodies, Mild cognitive impairment and Normal control, respectively.

3.3 IFS Results on the MCFS Feature List

For the MCFS feature list, the same procedures were conducted. Detailed performance of RF and PART on different number of features is listed in **Supplementary Table S3**. Likewise, an IFS curve was plotted for each classification algorithm to display the performance of them on different feature subsets, as illustrated in **Figure 3B**. It can be observed that the highest MCC for RF was 0.681, which was obtained by using top 106 features. Thus, we can build the optimal RF classifier with these features. The ACC of such classifier was 0.803, as listed in **Table 3**. Its performance on each class is shown in **Figure 4B**. Compared with the performance of the optimal RF classifier in **Section 3.2**, their performance was almost equal. As for PART, its highest MCC was 0.360. It was obtained by using top 89 miRNA features. Accordingly, the optimal PART classifier was built using these features. The ACC of this classifier was 0.555 (**Table 3**). The performance of this classifier on each class is shown in **Figure 4B**. Evidently, this PART classifier provided equal performance to the optimal PART classifier in **Section 3.2**. However, they were all inferior to the optimal RF classifiers.

Similar to the optimal RF classifier in **Section 3.2**, this optimal RF classifier also need several features. It was necessary to discover another RF classifier with a higher efficiency. After careful checking, we found that RF classifier with top 31 features can produce the MCC of 0.575 (**Figure 3B**) and ACC of 0.713 (**Table 3**). Its performance on five classes is shown in **Figure 4B**. Clearly, it was inferior to the optimal RF classifier. However, it had a higher efficiency because it used much less features. Thus, it can be a useful tool to identify four dementia subtypes and NC. Furthermore, the performance of such RF classifier and RF classifier with top 41 features yielded by mRMR method was almost equal.

TABLE 3 | Performance of key classifiers with different algorithms based on the MCFS feature list.

Classification algorithm	Number of features	ACC	MCC
Random forest	106	0.803	0.681
Random forest	31	0.713	0.575
PART	89	0.555	0.360

3.4 miRNA Expression Patterns Extracted From the Optimal PART Classifiers

Although the performance of two optimal PART classifier was much lower than two optimal RF classifiers, they can give interpretable rules, which can help us uncover the difference between four dementia subtypes and NC at miRNA level. For the mRMR feature list, the optimal PART classifier used top 72 features. With these features, PART was applied to all samples, resulting in 245 rules. These rules are provided in **Supplementary Table S4**. Likewise, for the MCFS feature list, top 89 features were adopted in the optimal PART classifier. 251 decision rules were obtained by applying PART on these features, which are also available in **Supplementary Table S4**. Accordingly, we accessed two groups of decision rules. For each group, each class received some rules. The number of rules for each class on each group is shown in **Figure 5**. With the exception of MCI, which has a relatively small number of rules, the numbers of rules of other classes were quite considerable. Some key expression rules are listed in **Tables 4**, **5** and the relevance of these rules in differentiating neurological disorders will be reviewed in **Section 4.1**.

3.5 Comparison of Optimal Classifiers Without SMOTE

In the IFS method, we employed SMOTE to reduce the influence of imbalanced problem. To elaborate the utility of SMOTE, the

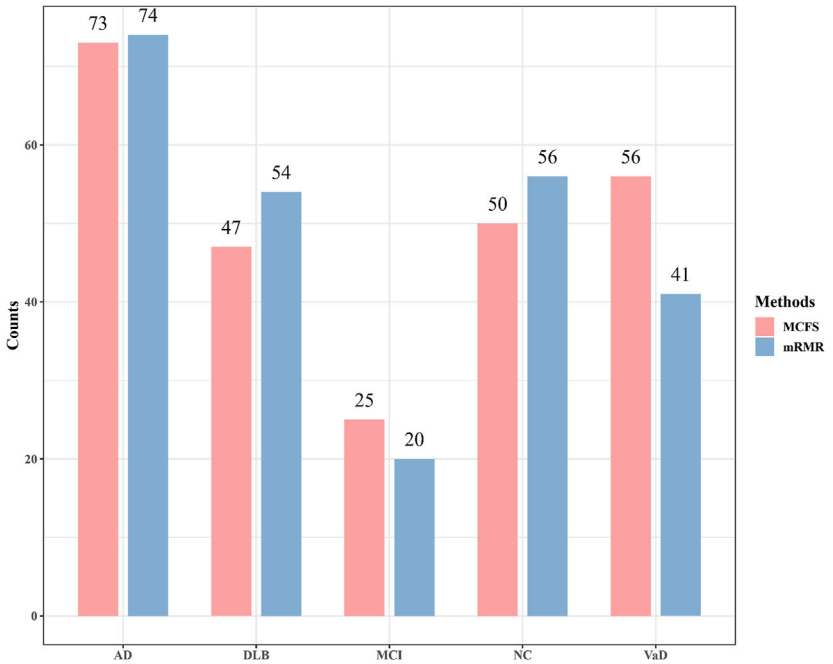

FIGURE 5 | Number of rules generated by the optimal PART classifiers based on mRMR and MCFS feature lists. AD, VaD, DLB, MCI, and NC stand for Alzheimer's disease, Vascular dementia, Dementia with Lewy bodies, Mild cognitive impairment and Normal control, respectively.

RF and PART classifiers mentioned in **Sections 3.2**, **3.3** were tested when SMOTE was not adopted. All classifiers were assessed by 10-fold cross-validation. The ACCs and MCCs of these classifiers are listed in **Table 6**. Compared with the ACCs and MCCs listed in **Tables 2**, **3**, MCC greatly decreased by at least 19%, even over 30% for the optimal RF classifiers. The ACC also decreased, but the degree was much smaller than that of the MCC. As the dataset was imbalanced, classifiers directly built on such dataset may be apt to the major classes (AD and NC in this study). Individual accuracies on these classes may be high, whereas individual accuracies on other classes may be low. The individual accuracies shown **Figure 6** confirmed this fact. The individual accuracies on AD were very high, followed by those on NC, whereas the individual accuracies on other three classes were very low, even zero. By employing SMOTE, the individual accuracies on AD decreased and those on other classes greatly increased, improving the entire performance of the classifiers. All these indicated the utility of the SMOTE.

4 DISCUSSION

The alteration of miRNA expression has been shown to relate with many pathological processes, including nervous system disorders. In this study, using the expression data of serum miRNAs, two optimal classifiers were constructed with high accuracy to identify the expression features of miRNAs through mRMR and MCFS method. We identified several putative miRNA biomarkers, which displayed strong relevance to the classification, suggesting that these miRNAs have specific effect in different types of neurodegenerative diseases. Additionally, the optimal PART classifiers yielded by mRMR and MCFS feature lists were then applied to generate 245 and 251 decision rules, respectively, which can classify each sample into one of five categories, namely, AD, VaD, DLB, MCI, and NC. In this section, we mainly focused on several optimal and common features identified both by mRMR and MCFS methods, considering that common features are much more important in the classification. We examined the selected features and decision rules and searched for the function and target genes of each miRNA using miRBase, an online database of miRNA sequences and annotation (Kozomara et al., 2018). For some miRNAs that have never been reported, we conducted bioinformatic analysis using miRDB for miRNA target prediction and functional annotation (Liu and Wang, 2019). Through literature review, several pieces of experimental evidence have been found to support the reliability of our prediction.

4.1 Analysis of Decision Rules Identified by mRMR and MCFS Methods

The most impactful feature in our computational analysis is miR-3184-5p, the mature miRNA product originating from the stem–loop precursor miRNA through cleavage by ribonuclease. As demonstrated by miRNA array experiment in multiple system atrophy disorders, a downregulated expression of miR-3184-5p was found in the FFPE sample of pons compared with controls, which indicates that this miRNA molecule plays an important role in normal brain development and may contribute

TABLE 4 | Some important rules extracted by the optimal PART classifier under the mRMR feature list.

Index	Decision Rules	Class
1	(hsa-miR-6088 ≤ 10.1065) & (hsa-miR-520f-5> 2.0854) & (hsa-miR-6836–3 ≤8.6821) & (hsa-miR-6811–5>1.8782) & (hsa-miR-4667–5> 6.1925) & (hsa-miR-6823–5 > 1.8811) & (hsa-miR-7851–3> 5.1826) & (hsa-miR-4667–5 ≤7.1244) & (hsa-miR-6756–5 ≤8.7714)	Normal control
2	(hsa-miR-6088 ≤ 9.9516) & (hsa-miR-4327 > 7.8591) & (hsa-miR-6861–5> 6.5728) & (hsa-miR-4485–5 ≤6.5037) & (hsa-miR-3622a-3> 4.5067) & (hsa-miR-6875–5 ≤10.0546) & (hsa-miR-7854–3 ≤4.8701)	Alzheimer's disease
3	(hsa-miR-208a-5> 5.8741) & (hsa-miR-548f-3 ≤2.1097) & (hsa-miR-4667–5> 6.7261) & (hsa-miR-6761–3> 4.7880) & (hsa-miR-520f-5 ≤1.8849)	Vascular dementia
4	(hsa-miR-208a-5> 5.8741) & (hsa-miR-548f-3 ≤2.1097) & (hsa-miR-4649–5>10.8160) & (hsa-miR-3622a-3 ≤4.4907) & (hsa-miR-6070 > 1.8843) & (hsa-miR-663b ≤ 8.7018)	Dementia with lewy bodies
5	(hsa-miR-520f-5 ≤1.8945) & (hsa-miR-6840–3 ≤7.6738) & (hsa-miR-185–5 ≤2.9551)	Mild cognitive impairment

TABLE 5 | Some important rules extracted by the optimal PART classifier under the MCFS feature list.

Index	Decision rules	Class
1	(hsa-miR-6088 ≤ 10.1065) & (hsa-miR-520f-5> 2.0854) & (hsa-miR-6836–3 ≤8.6821) & (hsa-miR-6811–5> 1.8782) & (hsa-miR-4667–5> 6.1925) & (hsa-miR-4746–3 ≤7.4409) & hsa-miR-3917 > 5.1453) & (hsa-miR-6070 ≤ 2.9233) & (hsa-miR-6869–3>1.8805)	Normal control
2	(hsa-miR-6088 ≤ 9.9516) & (hsa-miR-4327 > 7.8591) & (hsa-miR-1292–3> 4.0332) & (hsa-miR-6861–5> 6.5728) & (hsa-miR-125b-1–3 ≤ 4.7145) & (hsa-miR-128-1-5> 7.0405) & (hsa-miR-7854–3 ≤4.8762) & (hsa-miR-6088 ≤ 9.7663) & (hsa-miR-4506 ≤ 3.6756)	Alzheimer's disease
3	(hsa-miR-520f-5 ≤1.8945) & (hsa-miR-4485–3> 1.8928) & (hsa-miR-3184–5 ≤8.4938) & (hsa-miR-4496 > 1.8938) & (hsa-miR-6756–5 ≤8.5013) & (hsa-miR-548f-3 ≤1.8935) & (hsa-miR-6822–5> 3.4091) & (hsa-miR-4472 ≤ 6.3202) & (hsa-miR-1914–5 ≤4.1342) & (hsa-miR-6776–3> 4.0568) & (hsa-miR-548o-3> 1.8798)	Vascular dementia
4	(hsa-miR-208a-5> 5.8741) & (hsa-miR-548f-3 ≤2.1097) & (hsa-miR-4667–5 ≤6.7261) & (hsa-miR-4649–5>10.8290) & (hsa-miR-195–3> 1.8967)	Dementia with lewy bodies
5	(hsa-miR-520f-5 ≤1.8945) & (hsa-miR-4485–3> 1.8928) & (hsa-miR-1254 > 6.9170) & (hsa-miR-197–5> 7.3729)	Mild cognitive impairment

in the prevention of neurodegenerative disorders (Wakabayashi et al., 2016). In another research of spinocerebellar ataxia type 3 (SCA3), which is known as a highly heterogeneous neurodegenerative disorder, significantly downregulated expression of miR-3184 was observed in plasma from SCA3 patients compared with healthy controls (Hou et al., 2019). Therefore, we concluded that miR-3184-5p is necessary for the normal function of the brain, and the depletion of this molecule will lead to certain neurodegenerative disorders. Consistent with this finding, several decision rules in which miR-3184-5p is implicated show similar prediction that low expression levels of miR-3184-5p indicate AD and VaD categories, while relatively high expression levels indicate healthy controls.

In many decision rules that indicate the AD category, a relatively high expression of miR-6088 is required for the classification. Although little has been known about this miRNA, we found a report that miR-6088 displays a significantly upregulated expression in patients with stroke compared with NC (Gui et al., 2019). Considering that stroke is a brain disease induced by deficient blood supply and will lead to nervous system injury, we inferred that miR-6088 may also participate in the process of neurodegeneration. Additionally, miR-6088 was identified as one of the differentially methylated genes with high relevance to Parkinson's disease and neurodegeneration (Marsh et al., 2016), which provides strong support for the crucial role of miR-6088 in pathological processes of the nervous system.

Another important miRNA (miR-4327) is significantly associated with dementia, especially AD, through literature review. In the decision rules, we found that high expression of miR-4327 will lead to the classification of dementia, while relatively low expression indicates the normal cohort. As demonstrated by a miRNA expression profile experiment with Down syndrome, the expression level of miR-4327 was significantly higher in the case group than in the control group, suggesting that dysregulated miR-4327 may be related to abnormal development (Karaca et al., 2018). Individuals with Down syndrome usually show characteristics of damaged brain and intellectual disability, suggesting that miR-4327 affects brain development and results in several pathological processes including neurodegeneration. Moreover, using miRDB website tools, we found that the *OTUD1* gene is predicted as one of the target genes of miR-4327. *OTUD1* encodes a deubiquitinase, and mutations in this gene were reported to be associated with the

TABLE 6 | Performance of key classifiers without SMOTE.

Feature selection method	Classification algorithm	Number of features	ACC	MCC
mRMR	Random forest	106	0.691	0.323
	Random forest	41	0.690	0.313
	PART	72	0.550	0.158
MCFS	Random forest	106	0.690	0.319
	Random forest	31	0.691	0.317
	PART	89	0.547	0.162

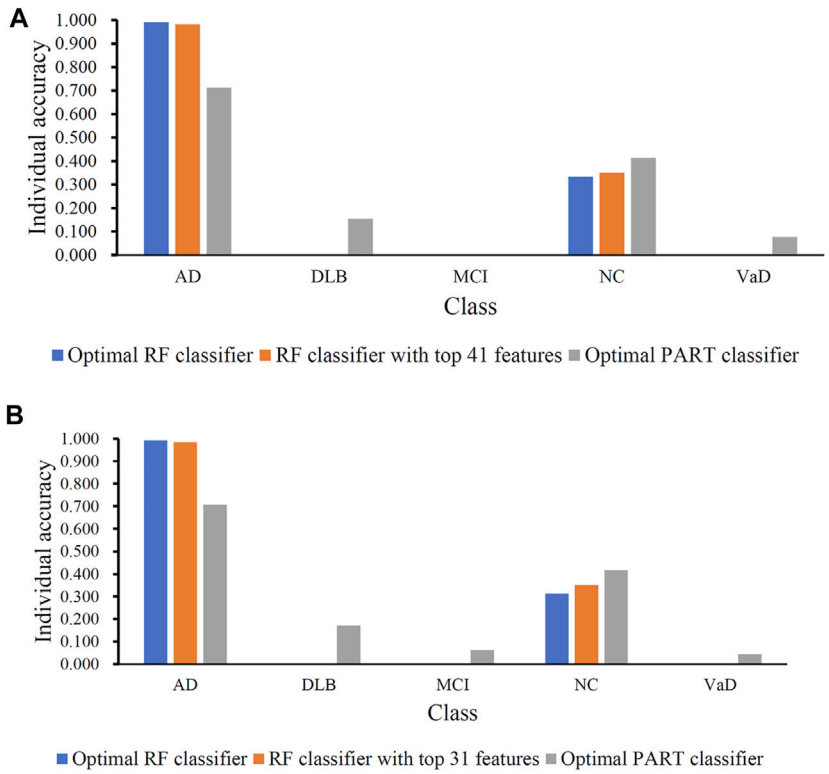

FIGURE 6 | Performance of the key RF and PART classifiers without SMOTE. **(A)**. Classifiers obtained by using mRMR feature list; **(B)**. Classifiers obtained by using MCFS feature list. AD, VaD, DLB, MCI, and NC stand for Alzheimer's disease, Vascular dementia, Dementia with Lewy bodies, Mild cognitive impairment and Normal control, respectively.

development of neurological phenotypes including ataxia with cerebellar atrophy and dementia (De Roux et al., 2016). On the basis of this finding, *OTUD1* is necessary for the normal neurological function, while excessive miR-4327 levels may inhibit *OTUD1* transcription and break the normal expression status. Therefore, the high level of miR-4327 is a risk indicator of dementia, which is consistent with our prediction model.

The high expression levels of miR-208a-5p display a strong indication to the categories of dementia in decision rules, suggesting that this miRNA plays a potential role in the associated processes. Several studies have described the role of miR-208a in cardiovascular diseases; for example, circulating levels of miR-208a are significantly elevated in patients with acute coronary syndrome (De Rosa et al., 2011). MiR-208a was undetectable in the blood from healthy individuals, while upregulated expression was observed in the plasma of patients with acute myocardial infarction (Wang et al., 2010). Transgenic overexpression of miR-208a in heart tissue led to hypertrophic growth and arrhythmias in mice (Callis et al., 2009), providing reliable experimental evidence regarding the key function of miR-208a in cardiovascular diseases. Healthy brain functioning is dependent on adequate blood supply, while certain vascular diseases will cause brain injury such as VaD. We inferred that high expression of miR-208a first induces disorders in the vascular system that gradually develop into VaD, which is consistent with the decision rules. Our study is the first to present the role of miR-208a in neurodegenerative diseases, and this will contribute to the clinical diagnosis of dementia.

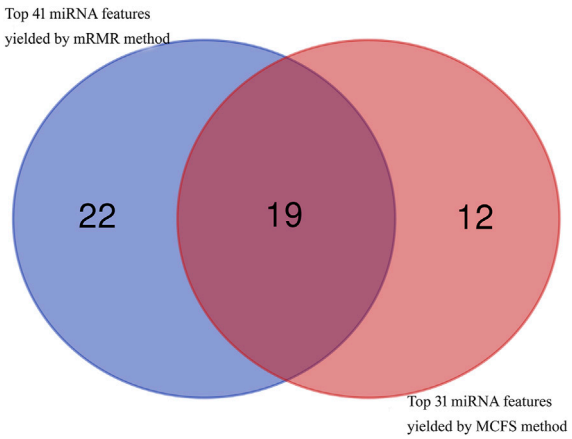

FIGURE 7 | Venn diagram to show top 41 miRNA features obtained by mRMR method and top 31 miRNA features obtained by MCFS method. Nineteen miRNA features are commonly identified.

The high expression of miR-520f, one of the identified features implicated in both decision rules, indicates dementia. MiR-520f was found to be significantly increased in the CSF of patients with Huntington's disease compared with controls, suggesting that miR-520f can be used as a CSF biomarker for evaluating treatments (Reed et al., 2018). Huntington's disease is a neurodegenerative disease typically diagnosed in midlife, and this disease shares similar neuropathologic phenotypes to dementia. Thus, we inferred that an elevated level of miR-520f may also influence the pathologic processes of dementia. In addition, miR-520f is also significantly upregulated in multiple system atrophy, and its expression is negatively correlated with the target gene *AKT3* (Kim et al., 2019). *AKT3* has been reported to be related to neuronal insulin resistance in neurodegenerative diseases (Schubert et al., 2004). Taken together, we concluded that miR-520f acts as a transcriptional inhibitor of *AKT3*, and *AKT3* reduction will cause the neuropathologic processes of dementia.

The expression level of miR-1227 can be efficiently used to distinguish the types of dementia and NC in the prediction model, which suggests that miR-1227 is another important functional molecule involved in neurodegeneration. On the basis of a rabbit AD model, the specific expression pattern of miR-1227 was observed, which showed similar profiles to those observed in human AD samples (Liu et al., 2014), indicating the potential role of miR-1227 in AD and other dementia diseases. A recent study reported that *LINC00639*, the target gene of miR-1227, was downregulated in HIV-associated dementia (HAD), a kind of cognitive impairment induced by HIV infection (Li et al., 2018). Even though the pathogenesis of HAD remains unclear, the aberration of certain miRNAs such as miR-1227 can provide novel direction for further research. Similarly, increased expression of miR-1227 was detected in CSF from patients with intracerebral hemorrhage (Shi et al., 2018). In summary, miR-1227 displays distinct expression profiles in many brain injury disorders or dementia, suggesting that it may be an auxiliary diagnostic biomarker for these diseases. These findings confirmed the reliability of our decision rules and implied that the expression criteria of identified miRNAs can be used in disease risk classification and clinical diagnostic.

4.2 Analysis of the Top Features Identified by mRMR and MCFS Methods

In addition to the quantitative analysis discussed above, we have also identified many miRNAs that can be used as indicators for dementia. As the RF classifier with less features provided slight lower performance than the corresponding optimal RF classifier, miRNA features used in these two RF classifiers with less features were investigated in this section. Based on the mRMR feature list, 41 miRNA features were obtained, whereas 31 miRNA features were accessed from the MCFS feature list. After taking the union of these two feature subsets, 53 different miRNA features were obtained, which are listed in **Supplementary Table S5**. A Venn diagram was plotted to show the distribution of these miRNA features in two feature sets, as shown in **Figure 7**. It can be observed that nineteen miRNA features were commonly identified. These features were thought to be more reliable than others. Some of them were discussed as follows.

MiR-4649-5p exhibits an upregulated expression profile in neurodegenerative disorders (Viswambharan et al., 2017). In amyotrophic lateral sclerosis (ALS), which is a fatal neurodegenerative disease, increasing concentration of miR-4649-5p was observed in the plasma of ALS patients, suggesting that this miRNA can be used in the diagnosis of ALS (Takahashi et al., 2015). On the basis of the miRDB database, we found that miR-4649-5p can target *INSYN2*, a protein coding gene implicated in inhibitory synapses. This synaptic inhibition is fundamental for the functioning of the central nervous system, shaping and orchestrating the flow of information through neuronal networks to generate a precise neural code (Uezu et al., 2016). Therefore, miR-4649-5p plays an important role in neural development, which confirms the reliability of our computational analysis.

MiR-3181 is one of the most related features in our computational analysis, and many studies indicate the close association between this miRNA and vascular diseases. Significantly upregulated miR-3181 was detected in endothelial cells treated with acrolein, which is a component of cigarette smoke and has been implicated in the development of vascular disease, suggesting that this miRNA may improve the diagnosis of vascular disease induced by environmental pollutants (Lee et al., 2015). As discussed previously, the development of vascular disease may be accompanied by brain injury such as VaD, suggesting the role of miR-3181 in dementia. The *TCL1B* gene, which is predicted as one target of miR-3181, showed significant differential expression between Parkinson's disease patients and NC (Infante et al., 2015). *TCL1B* is also an activator of Akt, a kinase involved in neuron survival (Hashimoto et al., 2013), and abnormal Akt signaling has been reported to induce dopamine neuron degeneration (Greene et al., 2011).

The expression profile of miR-128-1-5p is also a strong indicator for the classification in our analysis. MiR-128 is a neuronally enriched miRNA that plays a crucial role in neuronal differentiation and survival (Guidi et al., 2010). The

expression of miR-128 is increased in the hippocampus of AD patients (Lukiw and Pogue, 2007). In addition, upregulated miR-128 can cause a decreased expression of *SNAP25* and lead to the perturbation of neuronal activity (Eletto et al., 2008). These results support the role of miR-128 in neurodegenerative disease. Using RNA sequencing techniques, miR-128 showed decreased expression in Huntington's disease (Martí et al., 2010). MiR-128 displays distinct expression patterns in different neurodegenerative diseases, indicating its potential capability of distinguishing varied disease subtypes and confirming the ability of our prediction model to classify different dementias.

Besides above commonly identified miRNAs, some miRNAs identified by exact one feature selection method (mRMR or MCFS) were also quite essential. For example, miR-185-5p is identified as one of the most relevant features that contribute to the classification. MiR-185 has been suggested to participate in the pathogenesis of major depression, a psychosocial impairment, and finally lead to suicide. It was thought to influence neuronal and circuit formation by regulating target downstream gene, *TrkB-T1*, which has been associated with suicidal behavior (Serafini et al., 2014). This finding suggests the key role of miR-185-5p involved in nervous system development, physiology, and diseases.

In this section, we discussed the verified or speculative functions of miRNAs identified by our computational analysis. All these miRNAs have been confirmed to contribute to distinguishing patients with dementia from healthy and varied disease subtypes. Strikingly, many miRNAs related to vascular diseases usually play a putative role in neurodegenerative diseases. This finding suggests the interaction between these two distinct disease types. In summary, this study presented a novel computational approach to identify potential biomarkers for diagnosis and therapy, and also set up a basic research foundation for further studies on the detailed pathological mechanism of miRNAs in neurodegenerative diseases.

5 CONCLUSION

We employed a computational analysis approach to discovery key miRNA properties that differentiate normal and neurodegenerative disease subgroups in this work. The Boruta feature selection method was utilized to exclude unnecessary miRNA features, and then mRMR and MCFS were used to rank the remaining features. A series of feature subsets was generated from these ranked feature lists using the IFS method, and the sample data containing these feature subsets was used to train the RF and PART classifiers. As a result, the optimal miRNA biomarker set was identified on the basis of the evaluation metrics of classifiers under varying number of features, and the classification rules were extracted from the optimal PARTs. Finally, the relationship between candidate features including hsa-miR-3184-5p, has-miR-6088, and has-miR-4649 and neurodegenerative diseases was validated in recent studies, confirming the efficacy of our methods and establishing the groundwork for further investigation into the underlying pathogenic mechanisms of miRNAs in neurodegenerative illnesses.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120584.

AUTHOR CONTRIBUTIONS

TH and Y-DC designed the study. SD, LC and KF performed the experiments. ZL and WG analyzed the results. ZL, WG and SD wrote the manuscript. All authors contributed to the research and reviewed the manuscript.

FUNDING

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA26040304, XDB38050200), National Key R&D Program of China (2018YFC0910403), the Fund of the Key Laboratory of Tissue Microenvironment and Tumor of Chinese Academy of Sciences (202002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.880997/full#supplementary-material

Supplementary Table S1 | Ranked feature lists analyzed by Boruta feature selection method, mRMR, and MCFS.

Supplementary Table S2 | Performance of different classifiers with different numbers of features obtained by IFS method for mRMR feature list.

Supplementary Table S3 | Performance of different classifiers with different numbers of features obtained by IFS method for MCFS feature list.

Supplementary Table S4 | Rule sets extracted from the optimal PART classifiers based on mRMR and MCFS feature lists.

Supplementary Table S5 | Different miRNAs identified by IFS method on mRMR or MCFS feature lists.

REFERENCES

Asanomi, Y., Shigemizu, D., Akiyama, S., Sakurai, T., Ozaki, K., Ochiya, T., et al. (2021). Dementia Subtype Prediction Models Constructed by Penalized Regression Methods for Multiclass Classification Using Serum microRNA Expression Data. *Sci. Rep.* 11, 20947. doi:10.1038/s41598-021-00424-1

Bekris, L. M., Lutz, F., Montine, T. J., Yu, C. E., Tsuang, D., Peskind, E. R., et al. (2013). MicroRNA in Alzheimer's Disease: an Exploratory Study in Brain,

Cerebrospinal Fluid and Plasma. *Biomarkers* 18, 455–466. doi:10.3109/1354750x.2013.814073

Breiman, L. (2001). Random Forests. *Machine Learn.* 45, 5–32. doi:10.1023/a:1010933404324

Callis, T. E., Pandya, K., Seok, H. Y., Tang, R.-H., Tatsuguchi, M., Huang, Z.-P., et al. (2009). MicroRNA-208a Is a Regulator of Cardiac Hypertrophy and Conduction in Mice. *J. Clin. Invest.* 119, 2772–2786. doi:10.1172/jci36154

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *jair* 16, 321–357. doi:10.1613/jair.953

Chen, L., Li, Z., Zhang, S., Zhang, Y. H., Huang, T., and Cai, Y. D. (2022). Predicting RNA 5-methylcytosine Sites by Using Essential Sequence Features and Distributions. *Biomed. Res. Int.* 2022, 4035462. doi:10.1155/2022/4035462

Chen, L., Wang, S., Zhang, Y.-H., Li, J., Xing, Z.-H., Yang, J., et al. (2017). Identify Key Sequence Features to Improve CRISPR sgRNA Efficacy. *IEEE Access* 5, 26582–26590. doi:10.1109/access.2017.2775703

Chen, W., Chen, L., and Dai, Q. (2021). iMPT-FDNPL: Identification of Membrane Protein Types with Functional Domains and a Natural Language Processing Approach. *Comput. Math. Methods Med.* 2021, 7681497. doi:10.1155/2021/7681497

Cogswell, J. P., Ward, J., Taylor, I. A., Waters, M., Shi, Y., Cannon, B., et al. (2008). Identification of miRNA Changes in Alzheimer's Disease Brain and CSF Yields Putative Biomarkers and Insights into Disease Pathways. *Jad* 14, 27–41. doi:10.3233/jad-2008-14103

Coolen, M., Katz, S., and Bally-Cuif, L. (2013). miR-9: a Versatile Regulator of Neurogenesis. *Front. Cel. Neurosci.* 7, 220. doi:10.3389/fncel.2013.00220

Danborg, P. B., Simonsen, A. H., Waldemar, G., and Heegaard, N. H. H. (2014). The Potential of microRNAs as Biofluid Markers of Neurodegenerative Diseases - a Systematic Review. *Biomarkers* 19, 259–268. doi:10.3109/1354750x.2014.904001

De Meyer, G., Shapiro, F., Vanderstichele, H., Vanmechelen, E., Engelborghs, S., De Deyn, P. P., et al. (2010). Diagnosis-independent Alzheimer Disease Biomarker Signature in Cognitively normal Elderly People. *Arch. Neurol.* 67, 949–956. doi:10.1001/archneurol.2010.179

De Rosa, S., Fichtlscherer, S., Lehmann, R., Assmus, B., Dimmeler, S., and Zeiher, A. M. (2011). Transcoronary Concentration Gradients of Circulating microRNAs. *Circulation* 124, 1936–1944. doi:10.1161/circulationaha.111.037572

De Roux, N., Carel, J.-C., and Léger, J. (2016). "Congenital Hypogonadotropic Hypogonadism: a Trait Shared by Several Complex Neurodevelopmental Disorders," in *Puberty from Bench to Clinic* (Karger Publishers), 72–86.

Ding, S., Wang, D., Zhou, X., Chen, L., Feng, K., Xu, X., et al. (2022). Predicting Heart Cell Types by Using Transcriptome Profiles and a Machine Learning Method. *Life* 12, 228. doi:10.3390/life12020228

Draminski, M., Rada-Iglesias, A., Enroth, S., Wadelius, C., Koronacki, J., and Komorowski, J. (2007). Monte Carlo Feature Selection for Supervised Classification. *Bioinformatics* 24, 110–117. doi:10.1093/bioinformatics/btm486

Eletto, D., Russo, G., Passiatore, G., Del Valle, L., Giordano, A., Khalili, K., et al. (2008). Inhibition of SNAP25 Expression by HIV-1 Tat Involves the Activity of Mir-128a. *J. Cel. Physiol.* 216, 764–770. doi:10.1002/jcp.21452

Frank, E., Hall, M., Trigg, L., Holmes, G., and Witten, I. H. (2004). Data Mining in Bioinformatics Using Weka. *Bioinformatics* 20, 2479–2481. doi:10.1093/bioinformatics/bth261

Frank, E., and Witten, I. H. (1998). "Generating Accurate Rule Sets without Global Optimization," in *Computer Science Working Papers* (Hamilton, New Zealand: University of Waikato, Department of Computer Science).

Geekiyanage, H., and Chan, C. (2011). MicroRNA-137/181c Regulates Serine Palmitoyltransferase and in Turn Amyloid , Novel Targets in Sporadic Alzheimer's Disease. *J. Neurosci.* 31, 14820–14830. doi:10.1523/jneurosci.3883-11.2011

Gilad, S., Meiri, E., Yogev, Y., Benjamin, S., Lebanony, D., Yerushalmi, N., et al. (2008). Serum microRNAs Are Promising Novel Biomarkers. *PloS one* 3, e3148. doi:10.1371/journal.pone.0003148

Gorodkin, J. (2004). Comparing Two K-Category Assignments by a K-Category Correlation Coefficient. *Comput. Biol. Chem.* 28, 367–374. doi:10.1016/j.compbiolchem.2004.09.006

Greene, L. A., Levy, O., and Malagelada, C. (2011). Akt as a Victim, Villain and Potential Hero in Parkinson's Disease Pathophysiology and Treatment. *Cell Mol Neurobiol* 31, 969–978. doi:10.1007/s10571-011-9671-8

Gui, Y., Xu, Z., Jin, T., Zhang, L., Chen, L., Hong, B., et al. (2019). Using Extracellular Circulating microRNAs to Classify the Etiological Subtypes of Ischemic Stroke. *Transl. Stroke Res.* 10, 352–361. doi:10.1007/s12975-018-0659-2

Guidi, M., Muiños-Gimeno, M., Kagerbauer, B., Martí, E., Estivill, X., and Espinosa-Parrilla, Y. (2010). Overexpression of miR-128 Specifically Inhibits the Truncated Isoform of NTRK3 and Upregulates BCL2 in SH-Sy5y Neuroblastoma Cells. *BMC Mol. Biol* 11, 95. doi:10.1186/1471-2199-11-95

Hanchuan Peng, H., Fuhui Long, F., and Ding, C. (2005). Feature Selection Based on Mutual Information Criteria of max-dependency, max-relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Machine Intell.* 27, 1226–1238. doi:10.1109/tpami.2005.159

Hashimoto, M., Suizu, F., Tokuyama, W., Noguchi, H., Hirata, N., Matsuda-Lennikov, M., et al. (2013). Protooncogene TCL1b Functions as an Akt Kinase Co-activator that Exhibits Oncogenic Potency *In Vivo*. *Oncogenesis* 2, e70. doi:10.1038/oncsis.2013.30

Hou, X., Gong, X., Zhang, L., Li, T., Yuan, H., Xie, Y., et al. (2019). Identification of a Potential Exosomal Biomarker in Spinocerebellar Ataxia Type 3/Machado-Joseph Disease. *Epigenomics* 11, 1037–1056. doi:10.2217/epi-2019-0081

Iliffe, S., Robinson, L., Brayne, C., Goodman, C., Rait, G., Manthorpe, J., et al. (2009). Primary Care and Dementia: 1. Diagnosis, Screening and Disclosure. *Int. J. Geriat. Psychiatry* 24, 895–901. doi:10.1002/gps.2204

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., et al. (2015). Identification of Candidate Genes for Parkinson's Disease through Blood Transcriptome Analysis in LRRK2-G2019s Carriers, Idiopathic Cases, and Controls. *Neurobiol. Aging* 36, 1105–1109. doi:10.1016/j.neurobiolaging.2014.10.039

Karaca, E., Aykut, A., Ertürk, B., Durmaz, B., Güler, A., Büke, B., et al. (2018). Diagnostic Role of MicroRNA Expression Profile in the Prenatal Amniotic Fluid Samples of Pregnant Women with Down Syndrome. *Balkan Med. J.* 35, 163–166. doi:10.4274/balkanmedj.2017.0511

Kim, T., Valera, E., and Desplats, P. (2019). Alterations in Striatal microRNA-mRNA Networks Contribute to Neuroinflammation in Multiple System Atrophy. *Mol. Neurobiol.* 56, 7003–7021. doi:10.1007/s12035-019-1577-3

Kohavi, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *International Joint Conference on Artificial Intelligence* (Lawrence Erlbaum Associates), 1137–1145.

Kostopoulou, O., Delaney, B. C., and Munro, C. W. (2008). Diagnostic Difficulty and Error in Primary Care-Aa Systematic Review. *Fam. Pract.* 25, 400–413. doi:10.1093/fampra/cmn071

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2018). miRBase: from microRNA Sequences to Function. *Nucleic Acids Res.* 47, D155–D162. doi:10.1093/nar/gky1141

Kursa, M. B., and Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *J. Stat. Softw.* 36, 1–13. doi:10.18637/jss.v036.i11

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. *Science* 294, 853–858. doi:10.1126/science.1064921

Lee, S. E., Yang, H., Son, G. W., Park, H. R., Cho, J.-J., Ahn, H.-J., et al. (2015). Identification and Characterization of MicroRNAs in Acrolein-Stimulated Endothelial Cells: Implications for Vascular Disease. *Biochip J.* 9, 144–155. doi:10.1007/s13206-015-9303-3

Li, D., Bao, P., Yin, Z., Sun, L., Feng, J., He, Z., et al. (2018). Exploration of the Involvement of LncRNA in HIV-Associated Encephalitis Using Bioinformatics. *PeerJ* 6, e5721. doi:10.7717/peerj.5721

Li, H., Habes, M., Wolk, D. A., Fan, Y., and Initiative, A. S. D. N. (2019). A Deep Learning Model for Early Prediction of Alzheimer's Disease Dementia Based on Hippocampal Magnetic Resonance Imaging Data. *Alzheimer's & Demen.* 15, 1059–1070. doi:10.1016/j.jalz.2019.02.007

Li, X., Lu, L., Lu, L., and Chen, L. (2022a). Identification of Protein Functions in Mouse with a Label Space Partition Method. *Mbe* 19, 3820–3842. doi:10.3934/mbe.2022176

Li, Z., Wang, D., Liao, H., Zhang, S., Guo, W., Chen, L., et al. (2022b). Exploring the Genomic Patterns in Human and Mouse Cerebellums via Single-Cell

Sequencing and Machine Learning Method. *Front. Genet.* 13, 857851. doi:10.3389/fgene.2022.857851

Licher, S., Leening, M. J. G., Yilmaz, P., Wolters, F. J., Heeringa, J., Bindels, P. J. E., et al. (2019). Development and Validation of a Dementia Risk Prediction Model in the General Population: an Analysis of Three Longitudinal Studies. *Ajp* 176, 543–551. doi:10.1176/appi.ajp.2018.18050566

Liu, H., and Setiono, R. (1998). Incremental Feature Selection. *Appl. Intelligence* 9, 217–230. doi:10.1023/a:1008363719778

Liu, Q. Y., Chang, M. N., Lei, J. X., Koukiekolo, R., Smith, B., Zhang, D., et al. (2014). Identification of microRNAs Involved in Alzheimer's Progression Using a Rabbit Model of the Disease. *Am. J. Neurodegener Dis.* 3, 33–44.

Liu, W., and Wang, X. (2019). Prediction of Functional microRNA Targets by Integrative Modeling of microRNA Binding and Target Expression Data. *Genome Biol.* 20, 18. doi:10.1186/s13059-019-1629-z

Lukiw, W. J., and Pogue, A. I. (2007). Induction of Specific Micro RNA (miRNA) Species by ROS-Generating Metal Sulfates in Primary Human Brain Cells. *J. Inorg. Biochem.* 101, 1265–1269. doi:10.1016/j.jinorgbio.2007.06.004

Marsh, A. G., Cottrell, M. T., and Goldman, M. F. (2016). Epigenetic DNA Methylation Profiling with MSRE: A Quantitative NGS Approach Using a Parkinson's Disease Test Case. *Front. Genet.* 7, 191. doi:10.3389/fgene.2016.00191

Martí, E., Pantano, L., Bañez-Coronel, M., Llorens, F., Miñones-Moyano, E., Porta, S., et al. (2010). A Myriad of miRNA Variants in Control and Huntington's Disease Brain Regions Detected by Massively Parallel Sequencing. *Nucleic Acids Res.* 38, 7219–7235.

Mckeith, I. G., Galasko, D., Kosaka, K., Perry, E. K., Dickson, D. W., Hansen, L. A., et al. (1996). Consensus Guidelines for the Clinical and Pathologic Diagnosis of Dementia with Lewy Bodies (DLB). *Neurology* 47, 1113–1124. doi:10.1212/wnl.47.5.1113

Mistur, R., Mosconi, L., Santi, S. D., Guzman, M., Li, Y., Tsui, W., et al. (2009). Current Challenges for the Early Detection of Alzheimer's Disease: Brain Imaging and CSF Studies. *J. Clin. Neurol.* 5, 153–166. doi:10.3988/jcn.2009.5.4.153

Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., and Karagiannidou, M. (2016). World Alzheimer Report 2016: Improving Healthcare for People Living with Dementia: Coverage, Quality and Costs Now and in the Future. *Alzheimer's Dis. Int.* 1, 1. doi:10.13140/RG.2.2.22580.04483

Reed, E. R., Latourelle, J. C., Bockholt, J. H., Bregu, J., Smock, J., Paulsen, J. S., et al. (2018). MicroRNAs in CSF as Prodromal Biomarkers for Huntington Disease in the PREDICT-HD Study. *Neurology* 90, e264–e272. doi:10.1212/wnl.0000000000004844

Robinson, L., Tang, E., and Taylor, J.-P. (2015). Dementia: Timely Diagnosis and Early Intervention. *Bmj* 350, h3029. doi:10.1136/bmj.h3029

Schubert, M., Gautam, D., Surjo, D., Ueki, K., Baudler, S., Schubert, D., et al. (2004). Role for Neuronal Insulin Resistance in Neurodegenerative Diseases. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3100–3105. doi:10.1073/pnas.0308724101

Serafini, G., Pompili, M., Hansen, K. F., Obrietan, K., Dwivedi, Y., Shomron, N., et al. (2014). The Involvement of microRNAs in Major Depression, Suicidal Behavior, and Related Disorders: a Focus on miR-185 and miR-491-3p. *Cel Mol Neurobiol* 34, 17–30. doi:10.1007/s10571-013-9997-5

Seshadri, S., Fitzpatrick, A. L., Ikram, M. A., Destefano, A. L., Gudnason, V., Boada, M., et al. (2010). Genome-wide Analysis of Genetic Loci Associated with Alzheimer Disease. *Jama* 303, 1832–1840. doi:10.1001/jama.2010.574

Shi, Q., Ge, D., Yang, Q., Wang, L., and Fu, J. (2018). MicroRNA Profiling of Cerebrospinal Fluid from Patients with Intracerebral Haemorrhage. *Front. Lab. Med.* 2, 141–145. doi:10.1016/j.flm.2019.07.001

Shigemizu, D., Akiyama, S., Asanomi, Y., Boroevich, K. A., Sharma, A., Tsunoda, T., et al. (2019a). Risk Prediction Models for Dementia Constructed by Supervised Principal Component Analysis Using miRNA Expression Data. *Commun. Biol.* 2, 77. doi:10.1038/s42003-019-0324-7

Shigemizu, D., Akiyama, S., Asanomi, Y., Boroevich, K. A., Sharma, A., Tsunoda, T., et al. (2019b). A Comparison of Machine Learning Classifiers for Dementia with Lewy Bodies Using miRNA Expression Data. *BMC Med. Genomics* 12, 150. doi:10.1186/s12920-019-0607-3

Shimoda, A., Li, Y., Hayashi, H., and Kondo, N. (2021). Dementia Risks Identified by Vocal Features via Telephone Conversations: A Novel Machine Learning Prediction Model. *PloS one* 16, e0253988. doi:10.1371/journal.pone.0253988

Stephan, B. C. M., Kurth, T., Matthews, F. E., Brayne, C., and Dufouil, C. (2010). Dementia Risk Prediction in the Population: Are Screening Models Accurate? *Nat. Rev. Neurol.* 6, 318–326. doi:10.1038/nrneurol.2010.54

Takahashi, I., Hama, Y., Matsushima, M., Hirotani, M., Kano, T., Hohzen, H., et al. (2015). Identification of Plasma microRNAs as a Biomarker of Sporadic Amyotrophic Lateral Sclerosis. *Mol. Brain* 8, 67. doi:10.1186/s13041-015-0161-7

Tang, S., and Chen, L. (2022). iATC-NFMLP: Identifying Classes of Anatomical Therapeutic Chemicals Based on Drug Networks, Fingerprints and Multilayer Perceptron. *Curr. Bioinformatics* 1, 1. doi:10.2174/1574893617666220318093000

Uezu, A., Kanak, D. J., Bradshaw, T. W. A., Soderblom, E. J., Catavero, C. M., Burette, A. C., et al. (2016). Identification of an Elaborate Complex Mediating Postsynaptic Inhibition. *Science* 353, 1123–1129. doi:10.1126/science.aag0821

Viswambharan, V., Thanseem, I., Vasu, M. M., Poovathinal, S. A., and Anitha, A. (2017). miRNAs as Biomarkers of Neurodegenerative Disorders. *Biomarkers Med.* 11, 151–167. doi:10.2217/bmm-2016-0242

Wakabayashi, K., Mori, F., Kakita, A., Takahashi, H., Tanaka, S., Utsumi, J., et al. (2016). MicroRNA Expression Profiles of Multiple System Atrophy from Formalin-Fixed Paraffin-Embedded Samples. *Neurosci. Lett.* 635, 117–122. doi:10.1016/j.neulet.2016.10.034

Wang, G.-K., Zhu, J.-Q., Zhang, J.-T., Li, Q., Li, Y., He, J., et al. (2010). Circulating microRNA: a Novel Potential Biomarker for Early Diagnosis of Acute Myocardial Infarction in Humans. *Eur. Heart J.* 31, 659–666. doi:10.1093/eurheartj/ehq013

Yang, Y., and Chen, L. (2022). Identification of Drug-Disease Associations by Using Multiple Drug and Disease Networks. *Cbio* 17, 48–59. doi:10.2174/1574893616666210825115406

Zhao, X., Chen, L., and Lu, J. (2018). A Similarity-Based Method for Prediction of Drug Side Effects with Heterogeneous Information. *Math. Biosciences* 306, 136–144. doi:10.1016/j.mbs.2018.09.010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fgene.2022.883810

Revealing Potential Spinal Cord Injury Biomarkers and Immune Cell Infiltration Characteristics in Mice

Liang Cao[1,2] *and Qing Li*[3]*

[1]Department of Traumatic Orthopedics, The Second Affiliated Hospital, University of South China, Hengyang, China, [2]School of Clinical Medicine, Guizhou Medical University, Guiyang, China, [3]Department of Orthopedics Traumatic, The Affiliated Hospital of Guizhou Medical University, Guiyang, China

Spinal cord injury (SCI) is a disabling condition with significant morbidity and mortality. Currently, no effective SCI treatment exists. This study aimed to identify potential biomarkers and characterize the properties of immune cell infiltration during this pathological event. To eliminate batch effects, we concurrently analyzed two mouse SCI datasets (GSE5296, GSE47681) from the GEO database. First, we identified differentially expressed genes (DEGs) using linear models for microarray data (LIMMA) and performed functional enrichment studies on those DEGs. Next, we employed bioinformatics and machine-learning methods to identify and define the characteristic genes of SCI. Finally, we validated them using immunofluorescence and qRT-PCR. Additionally, this study assessed the inflammatory status of SCI by identifying cell types using CIBERSORT. Furthermore, we investigated the link between key markers and infiltrating immune cells. In total, we identified 561 robust DEGs. We identified Rab20 and Klf6 as SCI-specific biomarkers and demonstrated their significance using qRT-PCR in the mouse model. According to the examination of immune cell infiltration, M0, M1, and M2 macrophages, along with naive CD8, dendritic cell-activated, and CD4 Follicular T cells may have a role in the progression of SCI. Therefore, Rab20 and Klf6 could be accessible targets for diagnosing and treating SCI. Moreover, as previously stated, immune cell infiltration may significantly impact the development and progression of SCI.

Keywords: spinal cord injury, immune cells, key genes, bioinformatics analysis, machine learning strategies

OPEN ACCESS

Edited by:
Guiyou Liu,
Tianjin Institute of Industrial Biotechnology (CAS), China

Reviewed by:
Valentine Svensson,
Independent researcher, United States
Klaus G. PETRY,
Institut National de la Santé et de la Recherche Médicale (INSERM), France

***Correspondence:**
Qing Li
liq168@gmc.edu.cn

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: *25 February 2022*
Accepted: *29 April 2022*
Published: *30 May 2022*

Citation:
Cao L and Li Q (2022) Revealing Potential Spinal Cord Injury Biomarkers and Immune Cell Infiltration Characteristics in Mice. Front. Genet. 13:883810. doi: 10.3389/fgene.2022.883810

INTRODUCTION

Spinal cord injury (SCI) is a devastating injury that frequently results in total or partial impairment of motor, sensory, and sphincter function (Lago et al., 2018). Moreover, whether classified as traumatic or non-traumatic, SCI always causes significant lifelong disability. SCI is becoming more common as vehicle accidents and extreme sports increase. As a result, this condition has disastrous impacts on patients, families, and society (Ropper and Ropper, 2017).

The pathological process of SCI is generally divided into two stages (Alizadeh et al., 2019). The primary injury causes hemorrhage, ischemia, edema, anoxia, and neuron and glial cell necrosis. The secondary injury involves complex pathophysiologic mechanisms, including ionic imbalance, free radical stress, inflammatory responses, and glial scars. Although the creation of glial scars can slow secondary damage spread, it also inhibits axon regrowth. Secondary injuries impair nerve plasticity and functional recovery. The main challenge in SCI treatment development is the difficulty of

repairing injured neurons and restoring the conducting function of axons. Currently, no effective drugs or therapeutic approaches exist for SCI (Badhiwala et al., 2019). As many patients experience severe physical and psychological consequences, SCI has become a global issue. Thus, elucidating the specific molecular mechanisms underlying the pathophysiology of SCI is crucial.

Recently, an increasing number of articles revealed that immune cell infiltration plays a pivotal role in SCI healing. For example, microRNA-151-3p is abundant in microglia-derived exosomes and has neuroprotective properties during SCI healing (Li et al., 2021). The chemokine (C-C motif) ligand 28 (CCL28) acts as a protective factor after SCI by recruiting C-C chemokine receptor 10 (CCR10)-positive and immunosuppressive regulatory T cells (Wang et al., 2019). After SCI, interleukin 19 (IL-19) enhances locomotor function recovery and decreases motor neuron loss, as well as microglial and glial activation (Guo et al., 2018). C3 is a novel Th2 interleukin reducing neurite outgrowth and neuronal survival *in vitro* and axon regeneration *in vivo* (Peterson et al., 2017). Chronic SCI can impair CD8 T cell function by up-regulating programmed cell death-1 expression (Zha et al., 2014). γδ T cells are recruited to the SCI site, promoting the inflammatory response and exacerbating neurological impairment. CCL2/CCR2 signaling is critical for T cell recruitment to the SCI site and may be used as a novel therapeutic target in the future (Xu P. et al., 2021). Nonetheless, it is necessary to elucidate the molecular mechanism by which diverse immune cells influence SCI progression. As previously stated, assessing immune cell infiltration and dissecting the components of invading immune cells is crucial for unraveling the SCI molecular system and identifying novel immunotherapeutic targets (Ahmed et al., 2018; Al Mamun et al., 2021). CIBERSORT is a computational method for quantifying cell composition using gene expression data. This approach may help characterize immune cell infiltration (Newman et al., 2015).

We used the GEO database to obtain microarray datasets and conduct differential expression gene analyses. Additionally, we combined bioinformatics analysis and machine-learning techniques to thoroughly screen and identify key SCI genes. Next, we used CIBERSORT to compare immune cell infiltration in 25 immune cell subsets between SCI and sham samples. Additionally, we explored the relationships between the key genes and immune cells to better understand the molecular immunological mechanisms during SCI development.

MATERIALS AND METHODS

Data Source

We downloaded two SCI datasets (GSE5296, GSE47681) from the NCBI Gene Expression Omnibus (GEO) (Clough and Barrett, 2016). These two datasets were gene expression arrays generated using GPL1261 [Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array (Affymetrix, Santa Clara, CA, United States) (Barrett et al., 2013). We selected 18 samples with SCI and 12 sham samples from the GSE5296 dataset. Similarly, we selected 17 spinal cord tissue samples from the GSE47681 dataset (Wu et al., 2013), including 13 samples with SCI and 4 sham spinal cord tissue samples.

Data Normalization and Differentially Expressed Genes Screening

We processed the two SCI datasets using the R package "affy," notably for normalization and log2 transformation (Irizarry et al., 2003). Here, we considered the average value as the expression value when a group of probes corresponded to the same gene. Moreover, we eliminated the batch effects between two datasets using the surrogate variable analysis (SVA) package from Bioconductor (Leek et al., 2012). Finally, we screened the DEGs using the LIMMA package with a p-value < 0.05 and | log2 Fold change (FC)| > 1 (Ritchie et al., 2015).

GO, Kyoto Encyclopedia of Genes and Genomes, and GSEA Analysis of the Differentially Expressed Genes

We performed the analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) for the DEGs using DAVID 6.8 (https://david-d.ncifcrf.gov/). To understand the function of the DEGs, we uploaded them to the DAVID (Ritchie et al., 2015) and KOBAS databases (http://kobas.cbi.pku.edu.cn/) (Wu et al., 2006). We used a p-value < 0.05 and count ≥ 2 as the significant enrichment threshold. To provide a more intuitive understanding of the gene expression levels associated with significantly enriched functional pathways, we performed a gene set enrichment analysis (GSEA) using the R software (Subramanian et al., 2005).

Screening and Validation of Characteristic Genes

We screened for key genes associated with SCI using three algorithms: least absolute shrinkage and selection operator (LASSO) regression analysis (Tibshirani, 1996), random forests analysis (Strobl et al., 2007; Wang et al., 2016), and support vector machine-recursive feature elimination (SVM-RFE) analysis (Suykens and Vandewalle, 1999). For the random forest method, we used the R package "randomForest". We performed the LASSO logistic regression using the R package "glmnet," and a lambda of zero was considered optimal. We constructed the SVM classifier with tenfold cross-validation using the R package "e1071." We also used the RFE function within the "caret" package to select the featured gene using tenfold cross-validation. Then, we selected the genes from the three classification models for further analysis. The GSE45006 dataset was used as a validation dataset (Niu et al., 2021).

Spinal Cord Injury Procedure and Immunofluorescence

6–8 weeks C57BL/6 mice were obtained from the Experimental Animal Center of Guizhou Medical University [license no. SCXK (Qian) 2018-0001]. All animal experiments were approved by the

TABLE 1 | mRNA-specific primers of key genes.

Gene	Primer	Sequence (5′-3′)	PCR products
Klf6	Forward	GTTTCTGCTCGGACTCCTGAT	108bp
	Reverse	TTCCTGGAAGATGCTACACATTG	
Rab20	Forward	GGGAGCAGTTTCATGGTCTGG	143bp
	Reverse	GCAGTCATTGTTGGCTGTTTC	
β-tubulin	Forward	CTGTCCGTCCATCAGTTGGT	122bp
	Reverse	TGGTTCAGGTCTCCGTAGGT	

Animal Care and Use Committee of Guizhou Medical University. We divided the mice into an SCI group and a sham group. All mice were anesthetized with 1.25% Avertin. First, we performed a 1-cm dorsal incision and performed a laminectomy of the T10 vertebra. We crushed the spinal cord with vessel clamping for 15 s. Paralysis of both lower limbs indicated successful modeling. For the sham group, we isolated the skin and tissue to expose the spinal cord without injuring the animals. After surgery, we returned the mice to their home cages and performed manual bladder expression three times a day. We sacrificed the animals 7 days later. We injected 100 ml of phosphate-buffered saline (PBS) from the apex with a syringe to remove the blood, followed by 100 ml of 4% paraformaldehyde for tissue fixation until the mouse body was rigid. Next, we fixed the spinal cord with formalin and embedded it in paraffin before transversely cutting 20-μm-thick tissue sections using a Cryotome. We washed the sections three times with PBS for 5 min and added blocking buffer (10% goat serum and 0.3% TritonX-100) for 1 h. We then incubated the sections with the primary antibody overnight at 4°C, washed them three times with PBS for 5 min, incubated them with secondary antibodies (goat anti-rabbit Alexa Fluor 488,1:500,CST) in the dark for 2 h, and stained cells nuclei with DAPI (4',6-diamidino-2-phenylindole). Finally, we photographed the sections with a laser confocal microscope.

We used Rab20 (YT3922, 1:200, Immunoway) and Klf6 (14716-1-AP, 1:200, Proteintech) as primary antibodies.

Quantitative PCR Analysis

To summarize, we extracted total RNA from the spinal cord of mice using the TRIZOL reagent (TIANGEN BIOTECH Corp, Beijing, China), then polyadenylated and reverse-transcribed it into cDNA using a poly(T) adapter following the manufacturer's instructions. We performed real-time PCR using a thermal cycler with the following parameters: a 5 min initial denaturation step at 95°C; 44 cycles at 95°C for 15 s; 55°C for 30 s, and 72°C for 20 s. We subjected each sample to the entire experimental procedure in triplicate. **Table 1** lists the primers specific for mRNA.

Immune Cell Infiltration Analysis

We measured the relative proportions of immune cells in SCI mouse tissue using the CIBERSORT method to annotate merged expression data and calculate immune cell infiltrations based on mouse tissue expression profiles (Chen et al., 2017). Next, we compared the relative levels of 25 immune cells between the SCI and sham groups. A correlation heatmap, produced using the "corrplot" package, revealed the relationships between 25 types of infiltrating immune cells. Finally, we analyzed and visualized the Spearman correlation between key biomarkers and immune infiltrating cells using the "ggstatplot" and "ggplot2" packages.

RESULTS

Identification of Differentially Expressed Genes

Figure 1 shows the workflow of this study. We integrated two SCI datasets (GSE5296, GSE47681), including 16 sham samples and 31 SCI samples (**Figure 2A**). We found a total of 561 DEGs—536 up-regulated genes and 25 downregulated genes. **Figures 1B**, **2B** display the DEGs heatmap and volcano plot, respectively.

Function Enrichment Analysis

The GO and KEGG analyses revealed that the DEGs were mainly involved in the biological processes of leukocyte migration, cytokine-mediated signaling pathway, positive regulation of cytokine production, positive regulation of defense response, tumor necrosis factor superfamily cytokine production, response to molecule of bacterial origin, regulation of inflammatory response, and cell chemotaxis. Regarding the cellular components, these DEGs were mainly associated with the membrane raft receptor complex, endocytic vesicle, Golgi apparatus sub-compartment, membrane microdomain, phagocytic vesicle, collagen trimer, collagen-containing extracellular matrix, inflammasome complex, and NADPH oxidase complex (**Figure 3A**). The KEGG pathway analysis showed that the DEGs were involved in lipid metabolism, cytokine–cytokine receptor interaction, atherosclerosis, osteoclast differentiation, tuberculosis, phagosome, TNF signaling pathway, rheumatoid arthritis, Leishmaniasis, viral protein interaction with cytokine and cytokine receptor, and IL-17 signaling pathway (**Figure 3B**). Additionally, GSEA data indicated that certain pathways were enriched (**Figure 3C**). These results suggested that the immune system plays a vital role in SCI.

Key Biomarkers Screening and Validation

We used the LASSO logistic regression method to find 14 important biomarkers from the DEGs (**Figure 4A**). With the SVM-RFE method, we identified four genes qualifying as key biomarkers among the DEGs (**Figure 4B**). Additionally, we identified 30 genes as significant biomarkers using the random forest strategy (**Figures 4C,D**). The Rab20 and Klf6 genes were overlapped genes. Thus, we selected Rab20 and Klf6 as key biomarkers for further validation (**Figure 4E**). To verify the relationship between key genes and SCI vulnerability, we selected GSE45006 as the training data and Rab20 and Klf6 as the test genes. We compared the expression of hub genes during the SCI process. The SCI group had higher Rab20 and Klf6 expression levels than the sham group (**Figures 4F,G**). To confirm this, we performed immunofluorescence staining experiments using mouse spinal cord tissue. SCI tissues had significantly higher Rab20 and Klf6 levels than those from the sham group (**Figures 5A,B**). Finally, we quantified Rab20 and

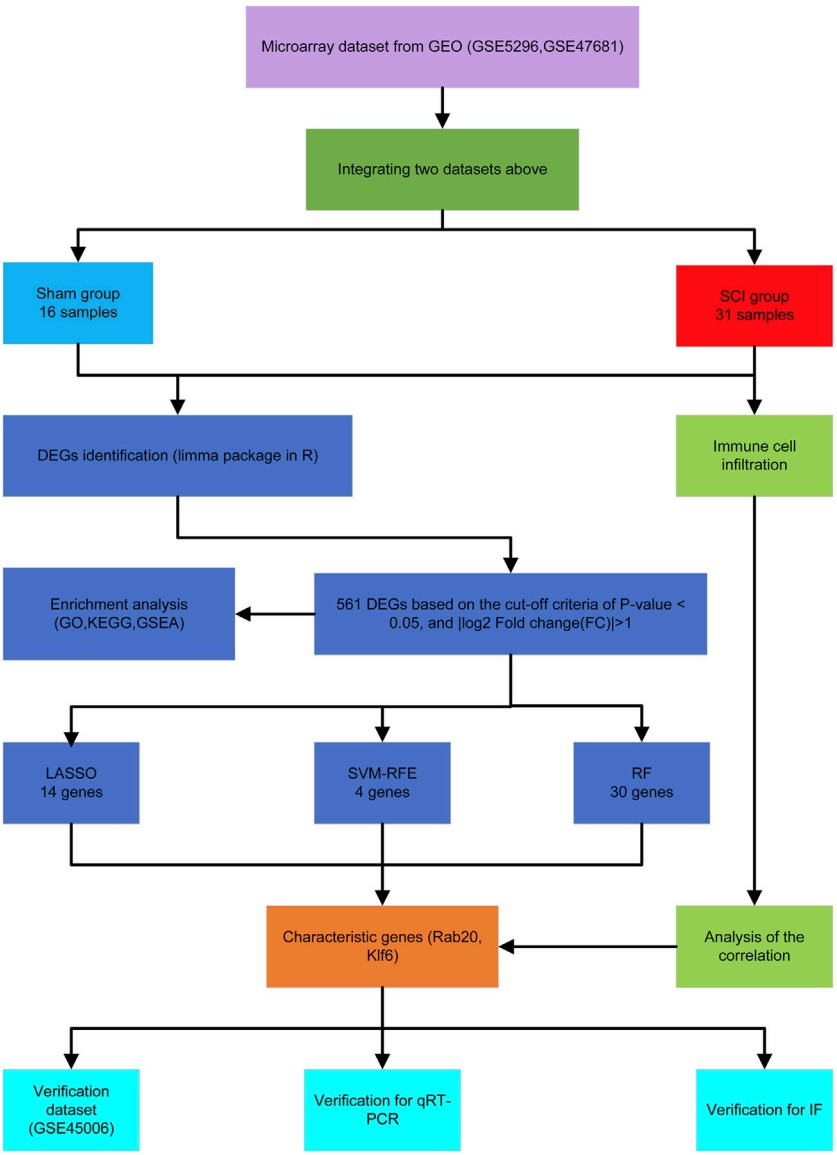

FIGURE 1 | The flowchart of the analysis process.

Klf6 expression in mouse samples using qRT-PCR. The SCI group had considerably higher levels of these two biomarkers (Rab20 and Klf6) than the sham group (**Figure 5C**).

Immune Cell Infiltration Analysis

To further investigate the association between SCI and immune cells during the development of SCI, we predicted immune cell infiltration using the CIBERSORT method. **Figure 6A** is a bar plot depicting the percentages of the 25 different kinds of immune cells. As revealed by the correlation heatmap for the 25 different types of immune cells, memory B cells and M0 macrophages, we found that neutrophil cells and M0 macrophages exhibited a substantial negative association. We also found substantial positive associations between monocytes and memory CD4 T cells, γδ T cells and naive CD4 T cells, γδ T cells and mast cells, eosinophils and neutrophils, natural killer resting cells and plasma cells, and memory CD8 T cells and plasma cells (**Figure 6B**). Furthermore, the SCI group had significantly higher proportions of naive CD8 T Cells, CD4 Follicular T cells, M0 macrophages, M1 macrophages, M2 macrophages, DC-activated cells than the sham group, and markedly lower proportions of memory B cells, plasma cells, memory CD8 T cells, memory CD4 T cells, naive CD4 T cells, Th17 Cells, and γδ T cells (**Figure 6C**).

Immune cell infiltration gradually played an important role after SCI. Thus, we selected different time points (1, 3, 7, 28 days of post-injury) to show the relationship between SCI and immune cells. We found no obvious difference between the SCI group and the sham group on the first day. However, the SCI group had significantly higher proportions of M0 macrophages, M2

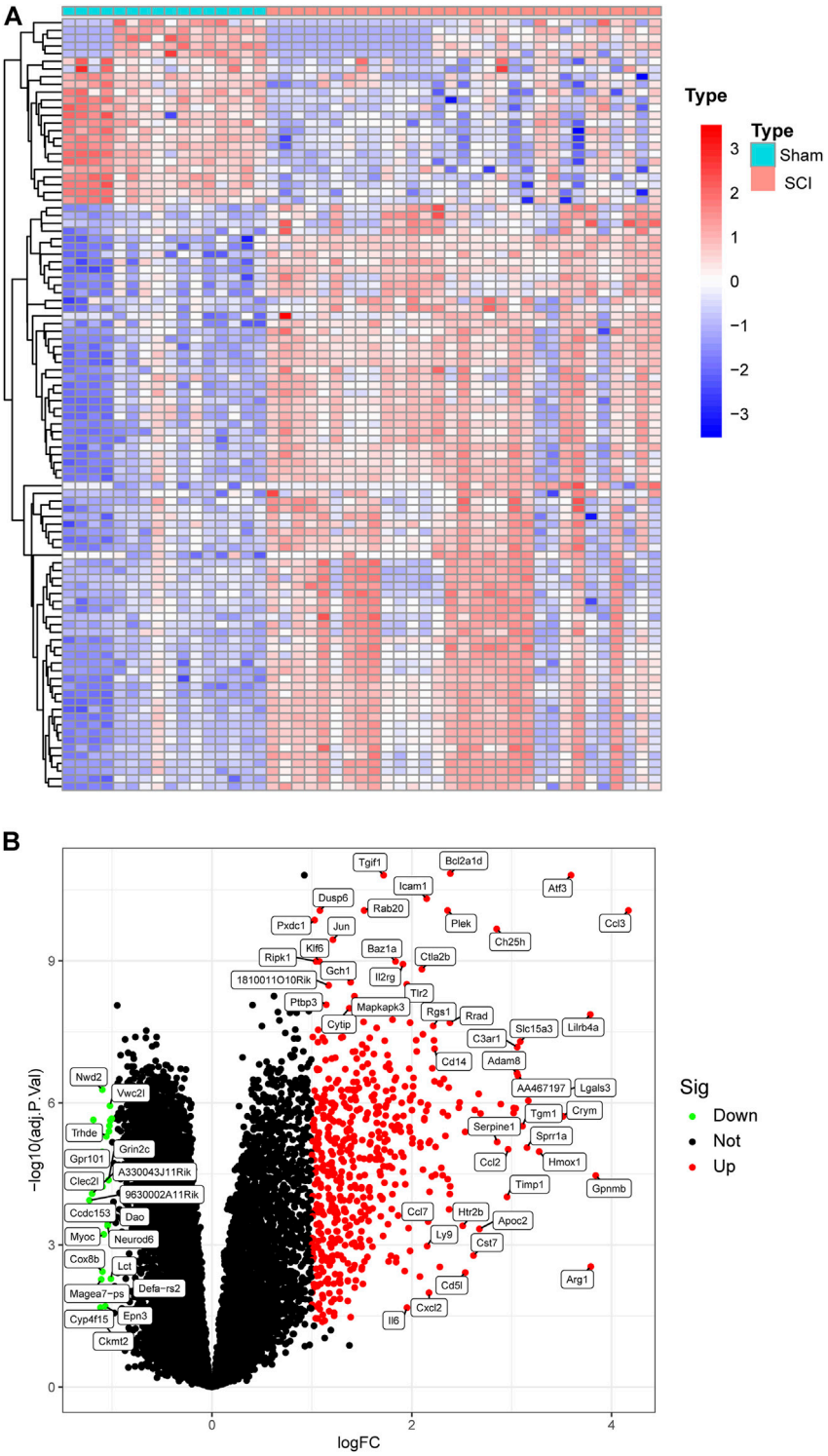

FIGURE 2 | Heat map and Volcano plot of the DEGs. **(A)** Each row of the heat map represented one DEG, and each column represents one sample. The red and blue colors represent upregulated and downregulated DEGs, respectively. **(B)** Red points represented upregulated DEGs, and green points displayed downregulated DEGs.

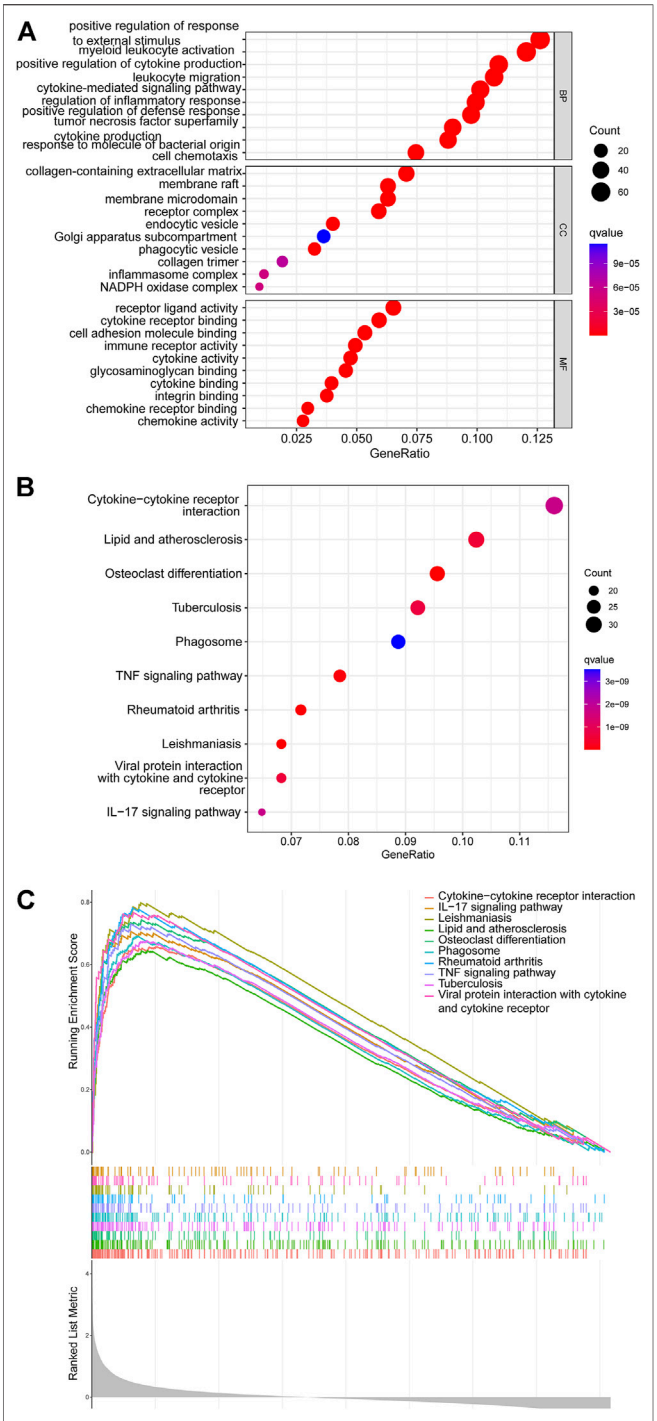

FIGURE 3 | The results of functional enrichment analyses. **(A)** GO analysis results of DEGs. **(B)** KEGG analysis results of DEGs. **(C)** GSEA profiles showed the ten significant GSEA sets.

macrophages than the sham group, and markedly lower proportions of memory B cells, plasma cells, naive CD4 T cells, and NK resting cells at day 3. At 7 days, the SCI group displayed higher ratios of M0 macrophages, M1 macrophages, M2 macrophages than the sham group, and noticeably lower ratios of memory B cells, plasma cells, memory CD8 T cells, and naive CD4 T cells at day 7. On day 28, there was no apparent difference between the SCI group and the sham group (**Figure 7**).

Based on the correlation analysis, Rab20 was positively correlated with M0 macrophages, M2 macrophages, and naive CD8 T cells and negatively correlated with memory CD4 T cells, naive CD4 T cells, plasma cells, memory B cells, γδ T Cells, and memory CD8 T cells (**Figure 8A**). In addition, Klf6 was positively correlated with naive CD8 T cells, DC-activated cells, M2 macrophages, and activated CD8 T cells and negatively correlated with plasma cells, memory CD4 T cells, naive CD4 T cells, γδ T cells, and memory CD8 T Cells (**Figure 8B**).

DISCUSSION

SCI frequently results in permanent functional deficits below the affected spinal cord region. The pathology of SCI is generally divided into two processes, named the primary injury and secondary injury. The secondary injury plays a crucial role in SCI onset and progression, leading to acute and chronic inflammation, tissue architecture damage, and motor and sensory dysfunction (Kong and Gao, 2017). Additionally, current research indicates that immune cell infiltration noticeably affects SCI development and progression (Al Mamun et al., 2021). Therefore, this study aimed to discover relevant SCI biomarkers and to investigate the role of immune cell infiltration in SCI.

Based on the GEO database, researchers can easily access Spinal cord injury (SCI) related datasets. Two SCI datasets (GSE5296, GSE47681) were included in this study (Zhao et al., 2018; Liu et al., 2019; Wei et al., 2019). We identified 561 DEGs in total. Among them, 536 were up-regulated and 25 were downregulated. Next, we performed functional enrichment analysis on these DEGs and found potential associations with immune responses and inflammatory signals (e.g., regulation of inflammatory response, leukocyte migration, positive regulation of cytokine production, and cytokine-mediated signaling pathway). Furthermore, The top 10 pathways of these DEGs according to *p* value were screened. Cytokines are crucial for immune response, pro-inflammatory cytokines influence the progression of disease (Bass et al., 2008). Cytokine–cytokine receptor interaction can be activated by neuroinflammation after SCI (Baek et al., 2017). Phagocytosis has important functions in immunity. Innate immune cells recognize and degrade microbes and debris by phagosomes. Macrophages process the debris of the spinal cord and promote the neurological function recovery after SCI (Zhou et al., 2020). TNF signaling pathway involves in the modulation of immune response and triggering the activation of T cells to induce cell death. The suppression of the TNF-α signaling pathway promotes function restoration after SCI (Wang N. et al., 2018). IL-17 is a pro-inflammatory cytokine and generated by T helper 17 cells (Torchinsky and Blander, 2010). IL-17 exacerbates the neuroinflammation of the spinal cord after SCI in the rat (Zong et al., 2014). Atherosclerosis is a chronic inflammatory

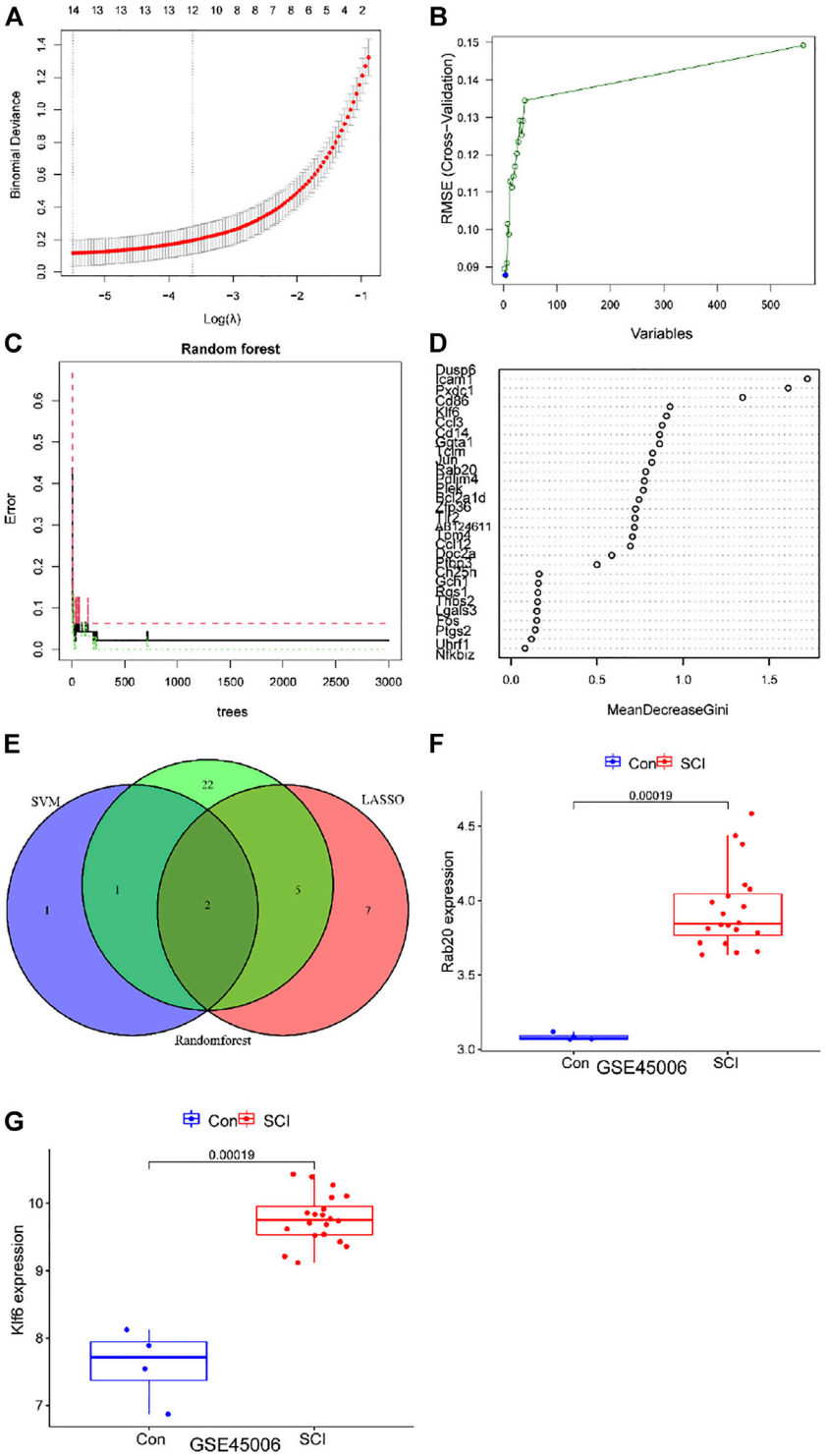

FIGURE 4 | Screening of key genes via the comprehensive strategy. **(A)** screening key markers using Least absolute shrinkage and selection operator (LASSO) logistic regression method. **(B)** screening key markers through support vector machine recursive feature elimination (SVMRFE) method. **(C,D)** random forest (RF) strategy to screen biomarkers. **(E)**Venn diagram displayed the intersection of key markers obtained by the three methods. **(F,G)** The expression levels of Rab20 and Klf6 in GSE45006.

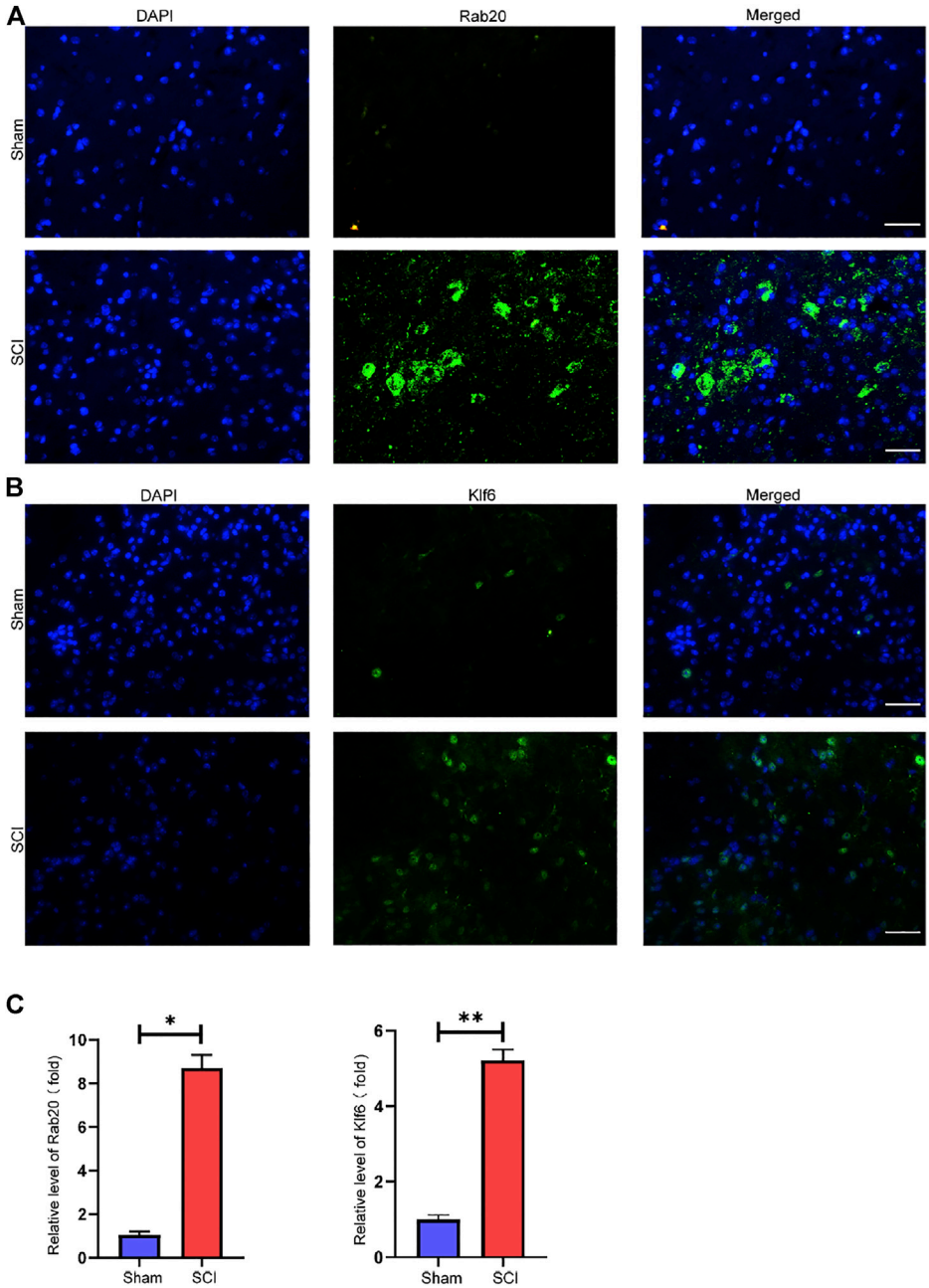

FIGURE 5 | Key genes validation. **(A)** Representative images showed that Rab20 expression was examined using immunofluorescence. **(B)** Representative images showed that Klf6 expression was examined using immunofluorescence. **(C)** qRT-PCR verification of Rab20 and Klf6 in SCI samples of mice and sham samples of mice. The results were represented as mean average ± SE with $p < 0.05$. scale bar represents 100 μM in **(A,B)**.

disease. Monocytes and macrophages contribute to the initiation and development of atherosclerosis (Moroni et al., 2019). Osteoclasts originated from hematopoietic monocyte-macrophage lineage. Osteoclast differentiation is mainly regulated by receptor activators of NF-κB and immune receptors (Park-Min et al., 2009). The pathology of tuberculosis is closely related to immune cells. Innate immune cells determined the inflammatory environment against Mycobacterium tuberculosis infection and induced adaptive immune responses (de Martino et al., 2019). Rheumatoid arthritis (RA) is a systemic inflammatory disorder. Immune cells (like T-cells, B cells, and macrophages) played a crucial role in the pathogenesis of RA (Yap et al., 2018). Viruses have produced many mechanisms to escape detection and destruction from the immune system by copying and repurposing host cytokine and cytokine receptor genes. Viral protein interaction with cytokine and cytokine receptor activates downstream cytokine signaling and affects different immune processes

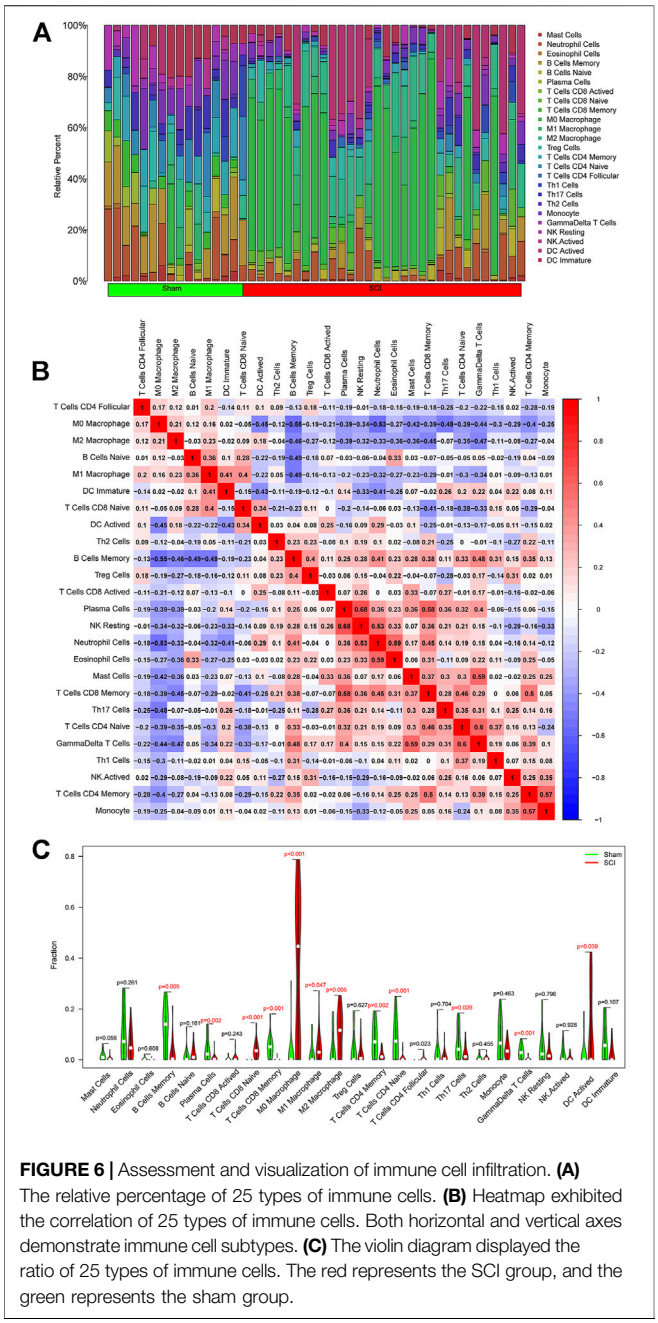

FIGURE 6 | Assessment and visualization of immune cell infiltration. **(A)** The relative percentage of 25 types of immune cells. **(B)** Heatmap exhibited the correlation of 25 types of immune cells. Both horizontal and vertical axes demonstrate immune cell subtypes. **(C)** The violin diagram displayed the ratio of 25 types of immune cells. The red represents the SCI group, and the green represents the sham group.

(Bruggeman, 2007). Immunity and leishmaniasis are also closely related. Leishmania first infected macrophages in the host. Then neutrophils secreted chemokine (C-C motif) ligand 3 (CCL3) to recruit dendritic cells. The interleukin (IL) 12 was produced by dendritic cell, which induced the differentiation of T helper type (Th) 1 cell to produce more IFN-γ to control the infection of Leishmania (de Freitas and von Stebut, 2021).

To explore the potential biomarkers during the development of SCI, we integrated and analyzed two mouse SCI datasets. We only selected SCI-related data. Thus, 47 samples were included. Based on the differentially expressed genes (DEGs), three machinery learning methods were applied for screening important genes. The random forest (RF) is a non-parametric approach for achieving classification under supervision. The term "random forest" refers to decision trees constructed from a subdivided data set. This method does not generate overfitting phenomena readily and exhibits strong anti-noise properties (Yang et al., 2020). Thus, the RF method has been employed

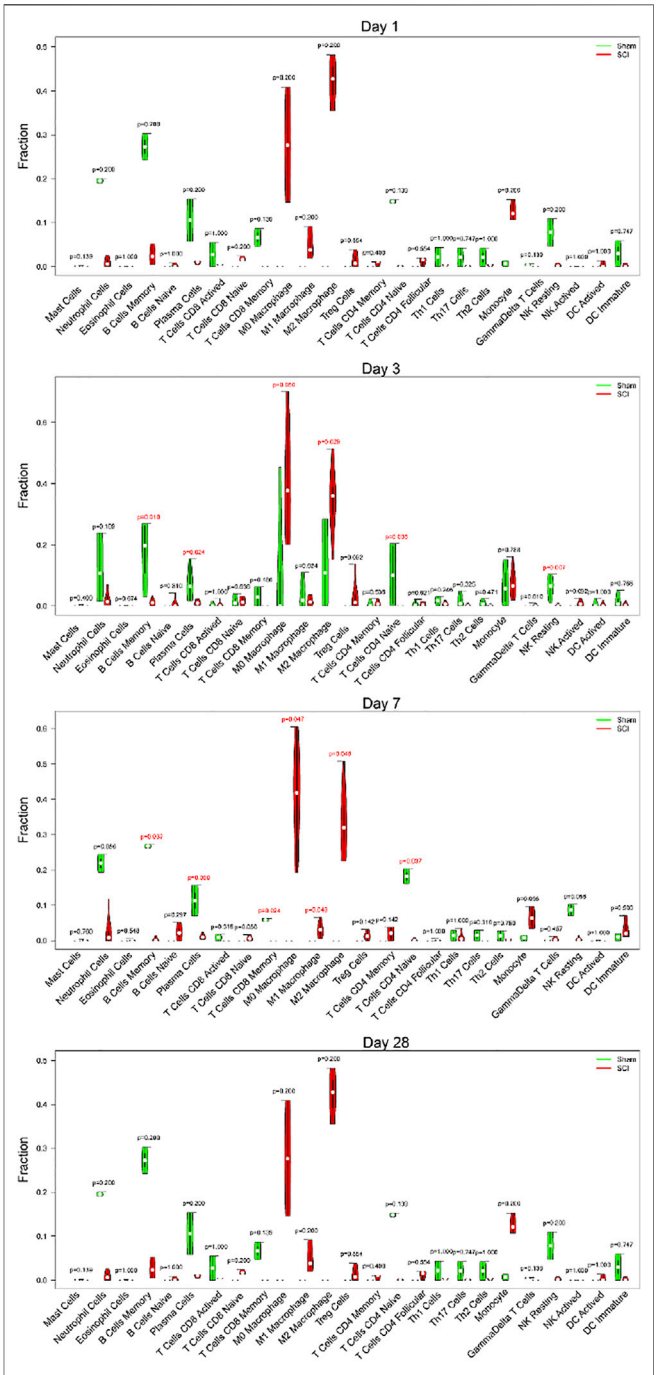

FIGURE 7 | Different time points of immune cell infiltration. The violin diagram exhibited the ratio of immune cells at different time points (1, 3, 7, 28 days of post-injury). The red represents the SCI group, and the green represents the sham group.

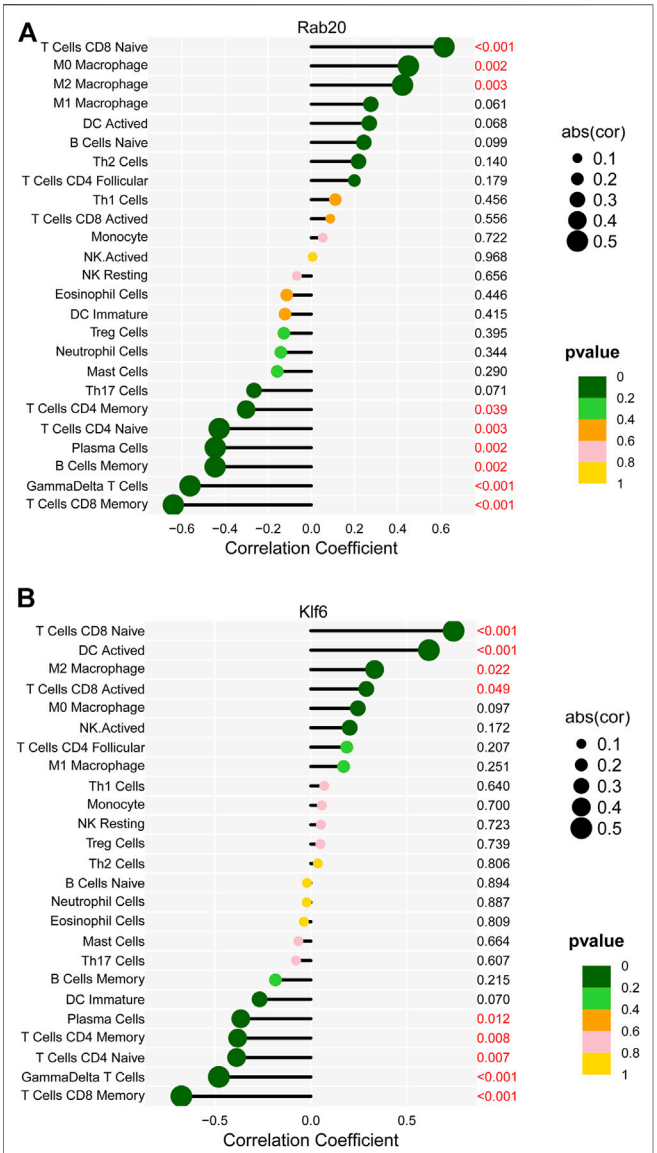

FIGURE 8 | The Correlation between key markers and infiltrating immune cells. **(A)** Correlation between Rab20 and infiltrating immune cells. **(B)** Correlation between Klf6 and infiltrating immune cells.

widely in recent years for prediction. ASSO logistic regression is a comprehensive machine-learning method for selecting variables by identifying those with the lowest chance of classification error. SVM-RFE is a machine learning approach for ranking and selecting the most significant features for classification. Every method obtained some essential genes. This study integrated these three distinct methods. We picked Rab20 and Klf6 because they were overlapped genes. In a previous study, researchers found a possible link with the immune and inflammatory functions, neuronal function, and synaptic transmission based on the functional enrichment analysis (GO and KEGG) of differentially expressed genes (DEGs) from GSE5296. Then they defined and collected these Neuronal function and synaptic transmission-associated genes and inflammation-associated genes from the literature review and investigated their expression in trauma site (R), adjacent rostral (M), and caudal (C) regions at different time points after SCI (Chen et al., 2015; Zhao et al., 2018). Another study aimed to explore the critical genes with genes expression of SCI from trkB.T1 knockout mice. This study identified the top four modules genes from GSE47681 using Weighted correlation network analysis (WGCNA). These module genes were used to construct the Protein-protein interaction (PPI) network. Finally, protein tyrosine phosphatase, receptor type C (PTPRC), coagulation factor II, thrombin (F2), plasminogen (PLG) were the most significant nodes in the PPI network (Wei et al., 2019). Compared to these two studies, we confirmed the differential expression of Rab20 and Klf6 with validation experiments, whereas other did not. Secondly, we used different screening methods to obtain SCI-related biomarkers. Importantly, our combined approach is more innovative as it points straightforward to relevant SCI markers Rab20 and Klf6, which are still not that much investigated.

Rab20 is a member of the Rab GTPase family, associating with macropinosomes at stages that overlap with those of Rab5, Rab21, Rab7, and Lamp1. Rab20 up-regulation may contribute to plaque destabilization *via* increased autophagy and cell death (Cederstrom et al., 2020). High Rab20 levels promote B cell activation and facilitate rheumatoid arthritis development (Tseng et al., 2019). Rab20's expression was increased during B cell transformation by a polymorphism associated with Crohn's disease and vaccination (Mehta et al., 2017). Additionally, Rab20 is an interferon-regulated Rab GTPase that promotes the homotypic fusion of early endosomes and directs endosomal cargo to lysosomes for degradation (Pei et al., 2015).

The Klf family of zinc finger transcription factors participates in various processes, including development, cellular differentiation, and stem cell biology. Klf6 promotes corticospinal tract sprouting and regeneration after SCI (Kramer et al., 2021). Alternatively, Klf6 is required for chronic pain maintenance, emphasizing its potential as a therapeutic target in chronic pain management (Mamet et al., 2017). When expressed ectopically in the adult injured central nervous system, Klf6 can promote axon growth (Wang Z. et al., 2018). Therefore, the identification of Rab20 and Klf6 together may imply that Rab20-mediated phagosomes cause cell death and KLf6 promotes nerve regeneration during SCI.

To more precisely assess the impact of immune cell infiltration in SCI, we analyzed immune cell infiltration through mice tissue expression profiles using CIBERSORT (Chen et al., 2017). The immune cell infiltration of M1, M0, and M2 macrophages, naive CD8 T cells, follicular CD4 T Cells, and DC-activated cells increased, indicating possible associations with SCI development and progression. Additionally, immune cell infiltration may have the property of dynamic changes at different time points after SCI. Microglia are well-known as the central nervous system's resident immune cells. After traumatic SCI, microglia/macrophages and neutrophils are recruited to the damaged location (Xu L. et al., 2021). Macrophages, microglia, and other antigen-presenting cells (APCs) activate T lymphocytes. SCI inhibits B cell activation

and antibody production (Ankeny and Popovich, 2010). Autoantibodies aggravate post-SCI complications such as cardiovascular, renal, and reproductive failure (Alizadeh et al., 2019). Although this has been mentioned previously, more investigation into the molecular mechanisms and effects of immune cell infiltration in SCI seems required.

Regarding the associations between immune cells and key genes, Rab20 was positively correlated with M0 and M2 macrophages and naive CD8 T cells during activation and negatively correlated with memory CD4 T cells, naive CD4 T cells, plasma cells and memory B cells, γδ T cells, and memory CD8 T cells. Further, Klf6 was positively correlated with naive CD8 T cells, DC-activated cells, M2 macrophages, and activated CD8 T cells, but negatively correlated with plasma cells, memory CD4 T cells, naive CD4 T cells, γδ T cells, and memory CD8 T cells. According to one study, the expression of a dominant-negative Rab20 mutant may impair macrophage phagosome maturation (Pei et al., 2014). KLF6 promotes HIF1 expression in macrophages, regulating inflammatory and hypoxic responses (Kim et al., 2019). Because there is no additional information about the sophisticated interaction mechanisms between key genes and immune cells, they should be thoroughly investigated based on the assumption mentioned above.

We used novel and scientific approaches (e.g., LASSO logistic regression, random forest, and SVM-RFE algorithm) to identify characteristic SCI makers. Additionally, we used CIBERSORT to investigate immune cell infiltration. Nonetheless, this study has some limitations. First, it is the result of the second round of data mining and analysis. Additionally, we did not obtain clinical specimens for this study and had to rely on mouse tissue to confirm our predictions. Finally, the results' reliability should be thoroughly validated using large samples.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by The Animal Care and Use Committee of Guizhou Medical University. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

QL designed the experiments. LC performed the experiments. LC performed the informatics analysis. LC wrote the manuscript.

FUNDING

We would like to express our gratitude to the providers of the datasets GSE5296, GSE45006, GSE47681 for sharing the data online.

REFERENCES

Ahmed, A., Patil, A.-A., and Agrawal, D. K. (2018). Immunobiology of Spinal Cord Injuries and Potential Therapeutic Approaches. *Mol. Cell Biochem.* 441 (1-2), 181–189. doi:10.1007/s11010-017-3184-9

Al Mamun, A., Monalisa, I., Tul Kubra, K., Akter, A., Akter, J., Sarker, T., et al. (2021). Advances in Immunotherapy for the Treatment of Spinal Cord Injury. *Immunobiology* 226 (1), 152033. doi:10.1016/j.imbio.2020.152033

Alizadeh, A., Dyck, S. M., and Karimi-Abdolrezaee, S. (2019). Traumatic Spinal Cord Injury: An Overview of Pathophysiology, Models and Acute Injury Mechanisms. *Front. Neurol.* 10, 282. doi:10.3389/fneur.2019.00282

Ankeny, D. P., and Popovich, P. G. (2010). B Cells and Autoantibodies: Complex Roles in CNS Injury. *Trends Immunol.* 31 (9), 332–338. doi:10.1016/j.it.2010.06.006

Badhiwala, J. H., Wilson, J. R., and Fehlings, M. G. (2019). Global Burden of Traumatic Brain and Spinal Cord Injury. *Lancet Neurology* 18 (1), 24–25. doi:10.1016/s1474-4422(18)30444-7

Baek, A., Cho, S.-R., and Kim, S. H. (2017). Elucidation of Gene Expression Patterns in the Brain after Spinal Cord Injury. *Cell Transpl.* 26 (7), 1286–1300. doi:10.1177/0963689717715822

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets--Update. *Nucleic Acids Res.* 41 (Database issue), D991–D995. doi:10.1093/nar/gks1193

Bass, W., Buescher, E., Hair, P., White, L., Welch, J., and Burke, B. (2008). Proinflammatory Cytokine-Receptor Interaction Model Improves the Predictability of Cerebral White Matter Injury in Preterm Infants. *Amer J. Perinatol.* 25 (4), 211–218. doi:10.1055/s-2008-1064931

Bruggeman, L. A. (2007). Viral Subversion Mechanisms in Chronic Kidney Disease Pathogenesis. *Clin. J. Am. Soc. Nephrol.* 2 (Suppl. 1), S13–S19. doi:10.2215/CJN.04311206

Cederström, S., Lundman, P., Folkersen, L., Paulsson-Berne, G., Karadimou, G., Eriksson, P., et al. (2020). New Candidate Genes for ST -elevation Myocardial Infarction. *J. Intern Med.* 287 (1), 66–77. doi:10.1111/joim.12976

Chen, G., Fang, X., and Yu, M. (2015). Regulation of Gene Expression in Rats with Spinal Cord Injury Based on Microarray Data. *Mol. Med. Rep.* 12 (2), 2465–2472. doi:10.3892/mmr.2015.3670

Chen, Z., Huang, A., Sun, J., Jiang, T., Qin, F. X.-F., and Wu, A. (2017). Inference of Immune Cell Composition on the Expression Profiles of Mouse Tissue. *Sci. Rep.* 7, 40508. doi:10.1038/srep40508

Clough, E., and Barrett, T. (2016). The Gene Expression Omnibus Database. *Methods Mol. Biol.* 1418, 93–110. doi:10.1007/978-1-4939-3578-9_5

de Freitas e Silva, R., and von Stebut, E. (2021). Unraveling the Role of Immune Checkpoints in Leishmaniasis. *Front. Immunol.* 12, 620144. doi:10.3389/fimmu.2021.620144

de Martino, M., Lodi, L., Galli, L., and Chiappini, E. (2019). Immune Response to *Mycobacterium tuberculosis*: A Narrative Review. *Front. Pediatr.* 7, 350. doi:10.3389/fped.2019.00350

Guo, J., Wang, H., Li, L., Yuan, Y., Shi, X., and Hou, S. (2018). Treatment with IL-19 Improves Locomotor Functional Recovery after Contusion Trauma to the Spinal Cord. *Br. J. Pharmacol.* 175 (13), 2611–2621. doi:10.1111/bph.14193

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics* 4 (2), 249–264. doi:10.1093/biostatistics/4.2.249

Kim, G.-D., Ng, H. P., Patel, N., and Mahabeleshwar, G. H. (2019). Kruppel-like Factor 6 and miR-223 Signaling axis Regulates Macrophage-mediated Inflammation. *FASEB J.* 33 (10), 10902–10915. doi:10.1096/fj.201900867RR

Kong, X., and Gao, J. (2017). Macrophage Polarization: a Key Event in the Secondary Phase of Acute Spinal Cord Injury. *J. Cell. Mol. Med.* 21 (5), 941–954. doi:10.1111/jcmm.13034

Kramer, A. A., Olson, G. M., Chakraborty, A., and Blackmore, M. G. (2021). Promotion of Corticospinal Tract Growth by KLF6 Requires an Injury Stimulus and Occurs within Four Weeks of Treatment. *Exp. Neurol.* 339, 113644. doi:10.1016/j.expneurol.2021.113644

Lago, N., Pannunzio, B., Amo-Aparicio, J., López-Vales, R., and Peluffo, H. (2018). CD200 Modulates Spinal Cord Injury Neuroinflammation and Outcome through CD200R1. *Brain, Behav. Immun.* 73, 416–426. doi:10.1016/j.bbi.2018.06.002

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034

Li, C., Qin, T., Liu, Y., Wen, H., Zhao, J., Luo, Z., et al. (2021). Microglia-Derived Exosomal microRNA-151-3p Enhances Functional Healing after Spinal Cord Injury by Attenuating Neuronal Apoptosis via Regulating the p53/p21/CDK1 Signaling Pathway. *Front. Cell Dev. Biol.* 9, 783017. doi:10.3389/fcell.2021.783017

Liu, J.-J., Guo, L., Lv, J., Huang, Y.-F., and Hao, D.-J. (2019). Bioinformatics Analyses of Differentially Expressed Genes Associated with Spinal Cord Injury: A Microarray-Based Analysis in a Mouse Model. *Neural Regen. Res.* 14 (7), 1262–1270. doi:10.4103/1673-5374.251335

Mamet, J., Klukinov, M., Harris, S., Manning, D. C., Xie, S., Pascual, C., et al. (2017). Intrathecal Administration of AYX2 DNA Decoy Produces a Long-Term Pain Treatment in Rat Models of Chronic Pain by Inhibiting the KLF6, KLF9, and KLF15 Transcription Factors. *Mol. Pain* 13, 174480691772791. doi:10.1177/1744806917727917

Mehta, S., Cronkite, D. A., Basavappa, M., Saunders, T. L., Adiliaghdam, F., Amatullah, H., et al. (2017). Maintenance of Macrophage Transcriptional Programs and Intestinal Homeostasis by Epigenetic Reader SP140. *Sci. Immunol.* 2 (9), eaag3160. doi:10.1126/sciimmunol.aag3160

Moroni, F., Ammirati, E., Norata, G. D., Magnoni, M., and Camici, P. G. (2019). The Role of Monocytes and Macrophages in Human Atherosclerosis, Plaque Neoangiogenesis, and Atherothrombosis. *Mediat. Inflamm.* 2019, 1–11. doi:10.1155/2019/7434376

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Park-Min, K.-H., Ji, J.-D., Antoniv, T., Reid, A. C., Silver, R. B., Humphrey, M. B., et al. (2009). IL-10 Suppresses Calcium-Mediated Costimulation of Receptor Activator NF-Kb Signaling during Human Osteoclast Differentiation by Inhibiting TREM-2 Expression. *J. Immunol.* 183 (4), 2444–2455. doi:10.4049/jimmunol.0804165

Pei, G., Repnik, U., Griffiths, G., and Gutierrez, M. G. (2014). Identification of an Immune Regulated Phagosomal Rab Cascade in Macrophages. *J. Cell Sci.* 127 (Pt 9), 2071–2082. doi:10.1242/jcs.144923

Pei, G., Schnettger, L., Bronietzki, M., Repnik, U., Griffiths, G., and Gutierrez, M. G. (2015). Interferon-γ-inducible Rab20 Regulates Endosomal Morphology and EGFR Degradation in Macrophages. *MBoC* 26 (17), 3061–3070. doi:10.1091/mbc.E14-11-1547

Peterson, S. L., Nguyen, H. X., Mendez, O. A., and Anderson, A. J. (2017). Complement Protein C3 Suppresses Axon Growth and Promotes Neuron Loss. *Sci. Rep.* 7 (1), 12904. doi:10.1038/s41598-017-11410-x

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Ropper, A. E., and Ropper, A. H. (2017). Acute Spinal Cord Compression. *N. Engl. J. Med.* 376 (14), 1358–1369. doi:10.1056/NEJMra1516539

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinforma.* 8, 25. doi:10.1186/1471-2105-8-25

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Suykens, J. A. K., and Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* 9 (3), 293–300. doi:10.1023/a:1018628609742

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58 (1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Torchinsky, M. B., and Blander, J. M. (2010). T Helper 17 Cells: Discovery, Function, and Physiological Trigger. *Cell. Mol. Life Sci.* 67 (9), 1407–1421. doi:10.1007/s00018-009-0248-3

Tseng, C. C., Lin, Y. Z., Lin, C. H., Li, R. N., Yen, C. Y., Chan, H. C., et al. (2019). Next-Generation Sequencing Profiles of the Methylome and Transcriptome in Peripheral Blood Mononuclear Cells of Rheumatoid Arthritis. *Jcm* 8 (9), 1284. doi:10.3390/jcm8091284

Wang, H., Yang, F., and Luo, Z. (2016). An Experimental Study of the Intrinsic Stability of Random Forest Variable Importance Measures. *BMC Bioinforma.* 17, 60. doi:10.1186/s12859-016-0900-5

Wang, N., Feng, Z., Zhao, W., Zhang, Z., and Zhang, L. (2018). Ultrashortwave Radiation Promotes the Recovery of Spinal Cord Injury by Inhibiting Inflammation via Suppression of the MK2/TNF-α P-athway. *Int. J. Mol. Med.* 42 (4), 1909–1916. doi:10.3892/ijmm.2018.3786

Wang, P., Qi, X., Xu, G., Liu, J., Guo, J., Li, X., et al. (2019). CCL28 Promotes Locomotor Recovery after Spinal Cord Injury via Recruiting Regulatory T Cells. *Aging* 11 (18), 7402–7415. doi:10.18632/aging.102239

Wang, Z., Mehra, V., Simpson, M. T., Maunze, B., Chakraborty, A., Holan, L., et al. (2018). KLF6 and STAT3 Co-occupy Regulatory DNA and Functionally Synergize to Promote Axon Growth in CNS Neurons. *Sci. Rep.* 8 (1), 12565. doi:10.1038/s41598-018-31101-5

Wei, L., He, F., Zhang, W., Chen, W., and Yu, B. (2019). Identification of Critical Genes Associated with Spinal Cord Injury Based on the Gene Expression Profile of Spinal Cord Tissues from trkB.T1 Knockout Mice. *Mol. Med. Rep.* 19 (3), 2013–2020. doi:10.3892/mmr.2019.9884

Wu, J., Mao, X., Cai, T., Luo, J., and Wei, L. (2006). KOBAS Server: a Web-Based Platform for Automated Annotation and Pathway Identification. *Nucleic Acids Res.* 34 (Web Server issue), W720–W724. doi:10.1093/nar/gkl167

Wu, J., Renn, C. L., Faden, A. I., and Dorsey, S. G. (2013). TrkB.T1 Contributes to Neuropathic Pain after Spinal Cord Injury through Regulation of Cell Cycle Pathways. *J. Neurosci.* 33 (30), 12447–12463. doi:10.1523/JNEUROSCI.0846-13.2013

Xu, L., Wang, J., Ding, Y., Wang, L., and Zhu, Y.-J. (2021). Current Knowledge of Microglia in Traumatic Spinal Cord Injury. *Front. Neurol.* 12, 796704. doi:10.3389/fneur.2021.796704

Xu, P., Zhang, F., Chang, M.-m., Zhong, C., Sun, C.-H., Zhu, H.-R., et al. (2021). Recruitment of γδ T Cells to the Lesion via the CCL2/CCR2 Signaling after Spinal Cord Injury. *J. Neuroinflammation* 18 (1), 64. doi:10.1186/s12974-021-02115-0

Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., et al. (2020). Study of Cardiovascular Disease Prediction Model Based on Random Forest in Eastern China. *Sci. Rep.* 10 (1), 5245. doi:10.1038/s41598-020-62133-5

Yap, H.-Y., Tee, S., Wong, M., Chow, S.-K., Peh, S.-C., and Teow, S.-Y. (2018). Pathogenic Role of Immune Cells in Rheumatoid Arthritis: Implications in Clinical Treatment and Biomarker Development. *Cells* 7 (10), 161. doi:10.3390/cells7100161

Zha, J., Smith, A., Andreansky, S., Bracchi-Ricard, V., and Bethea, J. R. (2014). Chronic Thoracic Spinal Cord Injury Impairs CD8+ T-Cell Function by Up-Regulating Programmed Cell Death-1 Expression. *J. Neuroinflammation* 11, 65. doi:10.1186/1742-2094-11-65

Zhang, D.-Y., Kou, Y.-H., Niu, S.-P., Zhang, Y.-J., Han, N., and Yin, X.-F. (2021). Identification of Four Differentially Expressed Genes Associated with Acute and Chronic Spinal Cord Injury Based on Bioinformatics Data. *Neural Regen. Res.* 16 (5), 865–870. doi:10.4103/1673-5374.297087

Zhao, S. J., Zhou, W., Chen, J., Luo, Y. J., and Yin, G. Y. (2018). Bioinformatics Analysis of the Molecular Mechanisms Underlying Traumatic Spinal Cord Injury. *Mol. Med. Rep.* 17 (6), 8484–8492. doi:10.3892/mmr.2018.8918

Zhou, X., Wahane, S., Friedl, M.-S., Kluge, M., Friedel, C. C., Avrampou, K., et al. (2020). Microglia and Macrophages Promote Corralling, Wound Compaction and Recovery after Spinal Cord Injury via Plexin-B2. *Nat. Neurosci.* 23 (3), 337–350. doi:10.1038/s41593-020-0597-7

Zong, S., Zeng, G., Fang, Y., Peng, J., Tao, Y., Li, K., et al. (2014). The Role of IL-17 Promotes Spinal Cord Neuroinflammation via Activation of the Transcription Factor STAT3 after Spinal Cord Injury in the Rat. *Mediat. Inflamm.* 2014, 1–10. doi:10.1155/2014/786947

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fgene.2022.890961

Identifying Antidepressant Effects of Brain-Derived Neurotrophic Factor and IDO1 in the Mouse Model Based on RNA-Seq Data

Jing Ren[1,2†], Chenyang Li[3†], Songren Wei[1], Yanjun He[4], Peng Huang[5] and Jiangping Xu[1]**

[1]Department of Neuropharmacology and Novel Drug Discovery, School of Pharmaceutical Sciences, Southern Medical University, Guangzhou, China, [2]Students Affairs Division, Zhujiang Hospital of Southern Medical University, Guangzhou, China, [3]The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China, [4]Emergency Department, Affiliated Foshan Maternity & Child Healthcare Hospital, Southern Medical University, Foshan, China, [5]Women and Children Medical Research Center, Affiliated Foshan Maternity & Child Healthcare Hospital, Southern Medical University, Foshan, China

OPEN ACCESS

Edited by:
Liang Cheng,
Harbin Medical University, China

Reviewed by:
Yanglan Gan,
Donghua University, China
Hui Ding,
University of Electronic Science and Technology of China, China

***Correspondence:**
Peng Huang
huangpengsmu@126.com
Jiangping Xu
jpx@smu.edu.cn

†These authors have contributed equally to this work and share first authorship.

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: 07 March 2022
Accepted: 15 April 2022
Published: 30 May 2022

Citation:
Ren J, Li C, Wei S, He Y, Huang P and Xu J (2022) Identifying Antidepressant Effects of Brain-Derived Neurotrophic Factor and IDO1 in the Mouse Model Based on RNA-Seq Data. Front. Genet. 13:890961. doi: 10.3389/fgene.2022.890961

Deletion of brain-derived neurotrophic factor (BDNF) and upregulation of indoleamine 2,3-dioxygenase 1 (IDO1) are associated with depression severity in animals. The neurotransmitter hypothesis of depression at the transcriptomic level can be tested using BDNF- and IDO1-knockout mouse models and RNA-seq. In this study, $BDNF^{+/-}$, $IDO1^{-/-}$, and chronic ultra-mild stress (CUMS)-induced depression mouse models and controls were developed, and the differentially expressed genes were analyzed. Furthermore, the ceRNA package was used to search the lncRNA2Target database for potential lncRNAs. Finally, a protein–protein interaction (PPI) network was constructed using STRINGdb. By comparing the control and CUMS model groups, it was found that pathway enrichment analysis and ceRNA network analysis revealed that most differentially expressed genes (DEGs) were associated with protection of vulnerable neuronal circuits. In addition, we found the enriched pathways were associated with nervous system development and synapse organization when comparing the control and $BDNF^{+/-}$model groups. When replicating the neurotransmitter disruption features of clinical patients, such comparisons revealed the considerable differences between CUMS and knockdown BDNF models, and the $BDNF^{+/-}$model may be superior to the classic CUMS model. The data obtained in the present study implicated the potential DEGs and their enriched pathway in three mouse models related to depression and the regulation of the ceRNA network-mediated gene in the progression of depression. Together, our findings may be crucial for uncovering the mechanisms underlying the neurotransmitter hypothesis of depression in animals.

Keywords: depression, RNA-seq, pathway enrichment analysis, protein–protein interaction (PPI) network, brain-derived neurotrophic factor (BDNF), indoleamine 2,3-dioxygenase 1 (IDO1)

INTRODUCTION

Depression is a common mental disorder characterized by high morbidity and suicidal risk (Auerbach et al., 2018; Devendorf et al., 2020). Previous studies have shown that depression is a complex disorder involving multiple genes (Fan et al., 2020; Kang et al., 2020). The brain-derived neurotrophic factor (BDNF) gene, which is widely involved in emotion and cognition, has neurotrophic effects and modulates neuron regeneration, synaptic plasticity, and dendritic growth (Kowianski et al., 2018; Lima Giacobbo et al., 2019). Several studies have shown that BDNF is involved in the pathogenesis of neuropsychiatric diseases (Lima Giacobbo et al., 2019; Colucci-D'amato et al., 2020). Chronic social defeat stress in a rat model of depression has revealed a significant reduction of BDNF levels in the hippocampus and prefrontal cortex (Amidfar et al., 2018).

Increasing studies have shown that rats that have been deprived of maternal care during their young stage exhibit reduced hippocampal BDNF levels, short- and long-term deficits in aversion, and recognition memory, as well as cognitive flexibility (Menezes et al., 2020). Environmental enrichment interventions restore the levels of hippocampal BDNF in rats and protect their memory and cognitive flexibility (Zhang et al., 2020). Furthermore, the reduced level of BDNF has been associated with anhedonia (Dong et al., 2018) which is the main symptom of depression. The deletion of brain-derived neurotrophic factor (BDNF) and upregulation of indoleamine 2,3-dioxygenase 1 (IDO1) are associated with depression severity in animals. The neurotransmitter hypothesis of depression at the transcriptomic level can be tested using BDNF- and IDO1-knockout mouse models and RNA-seq. In this study, $BDNF^{+/-}$, $IDO1^{-/-}$, and chronic ultra-mild stress (CUMS)-induced depression mouse models and controls were developed, and the differentially expressed genes were analyzed. Furthermore, the ceRNA package was used to search the lncRNA2Target database for potential lncRNAs. Finally, a protein–protein interaction (PPI) network was constructed using STRINGdb. By comparing the control and CUMS model groups, it was found that pathway enrichment analysis and ceRNA network analysis revealed that most differentially expressed genes (DEGs) were associated with the protection of vulnerable neuronal circuits. In addition, we found the enriched pathways were associated with nervous system development and synapse organization when comparing the control and $BDNF^{+/-}$model groups. When replicating the neurotransmitter disruption features of clinical patients, such comparisons revealed the considerable differences between CUMS and knockdown BDNF models, and the $BDNF^{+/-}$model may be superior to the classic CUMS model. The data obtained in the current study implicated the potential DEGs and their enriched pathway in three mouse models related to depression and the regulation of the ceRNA network-mediated gene in the progression of depression. Together, our findings may be crucial for uncovering the mechanisms.

Indoleamine 2,3-dioxygenase 1 (IDO1), which is the tryptophan catabolizing enzyme, affects the nervous system through two mechanisms. The first mechanism involves tryptophan depletion through over-activation of IDO1 which increases tryptophan catabolism and thereby reduces the levels of tryptophan, as well as suppressing the synthesis of 5-HT, hence resulting in depression (Chaves Filho et al., 2018). The second mechanism is the increase in kynurenine toxicity mediated by IDO1 (Jiang et al., 2020). It has been found that although kynurenine is neuroprotective, it is neurotoxic at excessive levels.

Therefore, it is evident that the reduction of BDNF can cause depression-like symptoms in mice (Jiang et al., 2019) whereas the knockout of IDO1 has antidepressant-like effects (Gao et al., 2021). Furthermore, there is no corresponding report on the mRNA sequencing of the comparison between BDNF and IDO1, but this study sequenced the mRNA expression in $BDNF^{+/-}$, $IDO1^{-/-}$, chronic ultra-mild stress (CUMS), and control mice.

MATERIALS AND METHODS

Animals and Experimental Groups

To avoid the effects of sex differences and hormones, only male mice were selected for the current study. Mice (10 per group) were randomly assigned to the control (untreated), CUMS-exposed (mimicking adult stress), $BDNF^{+/-}$ (strain $BDNF^{tm1Krj/J}$, C57BL6/J background, Jax Strain #006579), and $IDO1^{-/-}$ (strain $IDO1^{tm1Alm/J}$, Jax Strain #005867) groups. The detailed information about the mice is shown in **Supplementary Table S1**. They were housed in a pathogen-free, temperature-controlled environment (22 ± 1°C) and subjected to 12/12 h light/dark cycles, with ad libitum access to food and water except during model building. Animal experimental protocols in the current study were approved by the National Institutional Animal Care and Ethical Committee of Southern Medical University.

Chronic Ultra-Mild Stress Protocol

CUMS modeling was performed, as previously described (Huang et al., 2017; Gao et al., 2018). Briefly, the protocol involved the sequential application of various mild stressors: 1) 24 h of food and water deprivation, 2) 1 h of empty bottle, 3) 17 h of 45° cage tilt, 4) overnight illumination, 5) 24 h of wet cage, 6) 5 min swimming in water at 4°C, 7) 24 h of disrupting the squirrel cage, 8) 24 h of foreign body stimulation, and 9) 4 h of restriction in movement.

RNA Sequencing

TRIzol reagent was used to isolate RNA (Invitrogen, United States). The mRNA sequencing libraries were constructed using multiplex PCR amplification techniques. The sequencing of mRNA was carried out on the Illumina sequencing platform NextSeq 550, while the sequencing of microRNA was carried out on the Illumina sequencing platform Hiseq 4000.

Mapping

Adaptors were removed by FastQC and Trimmomatic. The alignment of mRNA was conducted by STAR software with the reference mm10, while miRNA was aligned with data from

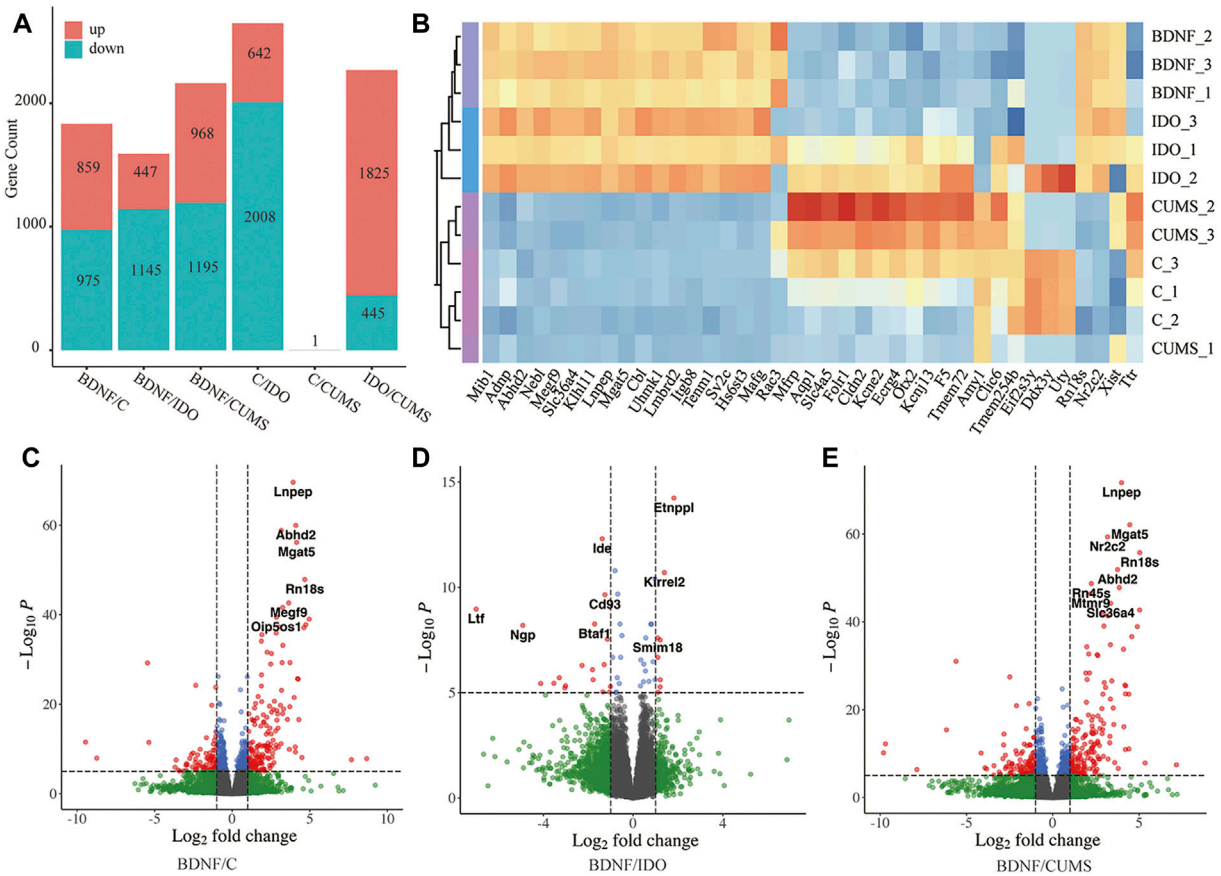

FIGURE 1 | Identified DEGs in each group. **(A)** Bar plot showing statistical data of DEGs. **(B)** Sample clustering based on the expression level of top DEGs. **(C)** Volcano plot of DEGs between BDNF$^{+/-}$ and control. **(D)** Volcano plot of DEGs between BDNF$^{+/-}$ and IDO1$^{-/-}$. **(E)** Volcano plot of DEGs between BDNF$^{+/-}$ and CUMS. In the Volcano plot, blue and green scatter points represent insignificant DEGs, red scatter points represent upregulated DEGs, and blue scatter points represent downregulated DEGs. The statistical method is the default cut-off for log2FC which is >|2|, and the default cut-off for *p*-value is 10e-6 to highlight the top genes with red color.

miRBase. Downstream statistical analyses were carried out in R software.

Differential Expression Analysis

The mRNA expression differential analysis was carried out using DESeq2. Volcano plots were plotted by the EnhancedVolcano package with a default cut-off for log2FC >|2|, and the default cut-off for *p*-value 10e-6 to highlight the top genes.

Differential miRNA: mRNA Interaction

miRNAs were searched on multiple miRNA-mRNA databases using multiMiR. The differential miRNA–mRNA interaction was calculated by using the binomial test. FDR was also used to adjust for multiple tests.

ceRNA Network Analysis

The potential lncRNAs targeting differentially expressed genes (DEGs) were searched on lncRNA2Target for the analysis of ceRNA. In addition, the ceRNA network of the collected miRNAs and lncRNAs was constructed and visualized by using the igraph package by querying interactions between them from multiple miRNA-lncRNA databases from multiMiR.

Protein–Protein Interaction Network Analysis

The analysis of the protein–protein interaction (PPI) network of the mRNA DEGs was performed using the R package STRINGdb to generate an interaction table, and the interaction network was visualized by using the igraph package.

RESULTS

Identification of Differentially Expressed Genes

It was found that the differences in expressed genes were highly significant between BDNF$^{+/-}$ and IDO1$^{-/-}$ mice, whereas there was a less evident difference in the gene expression between the CUMS and control groups.

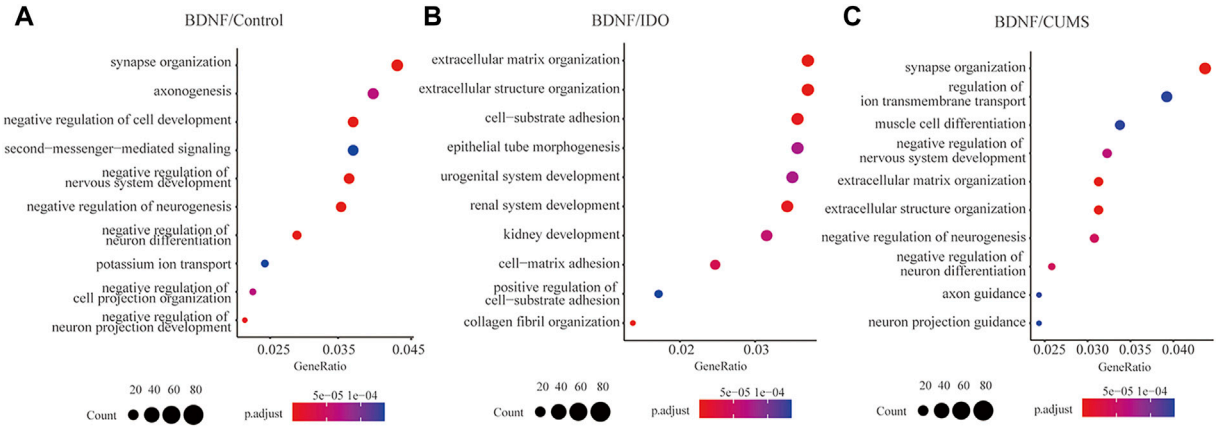

FIGURE 2 | Pathway enrichment of DEGs between BDNF$^{+/-}$ and controls. **(A)** Top 10 enriched pathways in GO terms for BDNF$^{+/-}$ and control groups. **(B)** Top 10 enriched pathways for GO terms for BDNF$^{+/-}$ and IDO1$^{-/-}$ groups. **(C)** Top 10 enriched pathways for GO terms in BDNF$^{+/-}$ and CUMS groups.

Mouse medial prefrontal cortex (mPFC) was obtained for sequencing from BDNF$^{+/-}$, IDO1$^{-/-}$, CUMS-exposed, and control mice. Results of the DEG analysis revealed gene expression differences between BDNF$^{+/-}$ and other groups, as well as modest gene expression differences in CUMS vs. control groups (**Figure 1A**). Consistently, the results of clustering analysis revealed close clustering between the control and CUMS samples (**Figure 1B**).

It was found that the analysis of gene expression identified 859 significantly upregulated and 975 significantly downregulated genes in BDNF$^{+/-}$ vs. control samples (**Figure 1A**, **Supplementary Table S2**). Furthermore, the results of volcano plot visualization revealed that the top DEGs included Lnpep, Adhd2, Nf2c2, Mgat2, and Rn18s (**Figure 1C**). A heatmap with sample clustering showed the most genes that were upregulated in the expression of the top 50 different genes in BDNF$^{+/-}$ (**Supplementary Figure S1A**). In addition, the results of analysis of the top five DEGs revealed that relative to BDNF$^{+/-}$, Lnpep, Abhd2, Mgat5, Nr2c2, and Rn18s expressions were significantly higher in controls (**Supplementary Figure 1B**). It was evidently noted that among the DEGs, Mgat5 influences behavior and physical outcomes in response to early life stress by remodeling N-glycans and cell surface glycoproteins.

Comparison BDNF$^{+/-}$ vs. IDO1$^{-/-}$ identified a total of 1,145 downregulated and 447 upregulated DEGs (**Figure 1A**, **Supplementary Table S2**), including Entppl, Idem, and Kirrel2 (**Figure 1D**). A heatmap showed an even regulated difference among the expressions of the top 50 DEGs, indicating that IDO1$^{-/-}$ may have a unique expression pattern under different biological mechanisms as compared with BDNF$^{+/-}$ (**Supplementary Figure S2A**). It was found that the top five DEGs exhibited an evenly matched relationship between these two groups (**Supplementary Figure S2B**). Etnppl was evaluated as an astrocyte-specific fasting-induced gene that induces the catabolization of phosphoethanolamine (PEtN), regulating brain lipid homeostasis (White et al., 2021). The altered Etnppl expression has also been associated with mood disorders (White et al., 2021). Both genes indicated a strong change in the neural level under these two groups of models.

A comparison of BDNF$^{+/-}$ vs. CUMS groups identified a total of 1,195 downregulated and 968 upregulated genes (**Supplementary Table S2**, **Figure 1A**). The DEGs included Lnpep, Mgat5, Rn18s, and Abdh2, which are quite similar to the results from BDNF$^{+/-}$ vs. control (**Figure 1E**). Similar to the control group, the heatmap also showed the most upregulated expression in BDNF$^{+/-}$ among the top 50 DEGs, and the top five DEGs, Abhd2, Lnpep, Mgat5, Nr2c2, and Rn18s also presented a higher expression in BDNF$^{+/-}$ (**Supplementary Figure S3**). This comparison illustrated a similar result of DEGs with previous groups of BDNF$^{+/-}$ and control, indicating that there was likely no significant difference in the gene expression between the CUMS and control groups. For the significantly different aforementioned genes , the significance threshold for statistical analysis was log2FC >|2|, and the default cut-off for *p*-value was 10e-6 to highlight the top genes with red dots.

Pathway Enrichment Among Models

It was found that there was little difference in neural activities between BDNF$^{+/-}$ that were involved in negative neuromodulatory pathways and IDO1$^{-/-}$ mice, but the CUMS model did not significantly differ from controls as compared with BDNF$^{+/-}$.

To assess pathway activation differences between the models, we subjected the DEGs to pathway enrichment analysis. Gene ontology (GO) term enrichment analysis of the BDNF$^{+/-}$ vs. control groups identified a total of 427 pathways (**Supplementary Table S3**), including the negative regulation of neurogenesis, negative regulation of nervous system development, synapse organization, and negative regulation of neuron differentiation (**Figure 2A**), which indicated a negative neural regulation in BDNF$^{+/-}$ mice.

The results of the heatmap and upset plot showed a common sharing gene enriched by different pathways (**Supplementary Figure S7A,B**). Furthermore, a comparison between the top pathways in the upset plot and their significant genes

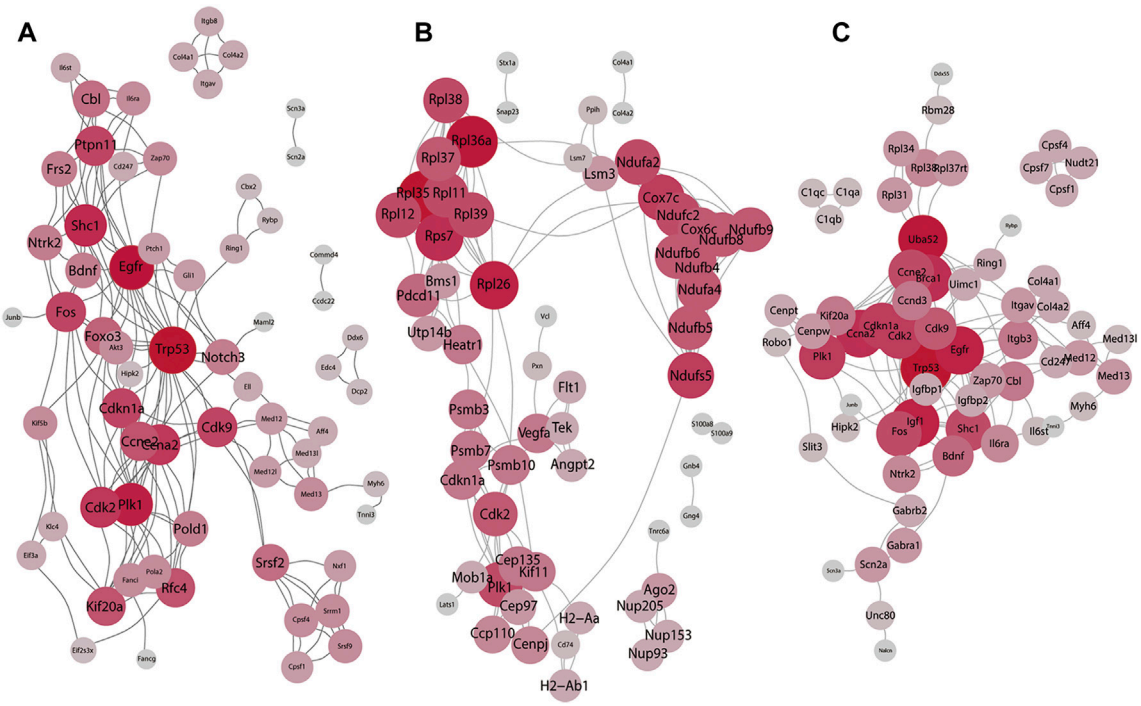

FIGURE 3 | Result of the PPI network analysis. **(A)** PPI network of the top 60 DEGs in BDNF$^{+/-}$ and control groups. **(B)** PPI network of the top 60 DEGs in BDNF$^{+/-}$ and IDO1$^{-/-}$ groups. **(C)** PPI network of the top 60 DEGs in BDNF$^{+/-}$ and CUMS. The degree of red color and the size of each vertex indicate the number of connections.

identified a high concentrated gene set that included Mib1, Foxo3, Ptbp1, Sema3c, and Sorl, enriched in a cluster of neural regulation pathways such as negative regulation of neuron differentiation, neuron projection guidance, and axonogenesis (**Supplementary Figure 7C**).

We identified a total of 237 significant GO terms and revealed the DEGs to be enriched for various pathways that are not related to neural regulation, including extracellular matrix organization, extracellular structure organization, collagen fibril organization, cell-substrate adhesion, and renal system development (**Figure 2B**, **Supplementary Table S3**). This indicated little difference in neural activities between BDNF$^{+/-}$ and IDO1$^{-/-}$ mice. The count of shared genes among top pathways was lower as compared to BDNF$^{+/-}$ vs. control, which indicates a discrete distribution of biological functions (**Supplementary Figure S8A,B**).

In the top five pathways, the high concentrated gene set, including Cxcr2, Tnxb, P4ha1, Adams1, and Col4a5 among others, was not highly related to neural function (**Supplementary Figure S8C**). The CXCL1 chemokine deletion can cause rat depression-like behaviors, and CXCL1/CXCL2 correlates with depression-like behavior in response to chronic stress (Chai et al., 2019; Song et al., 2020).

We identified a total of 625 significant GO terms and revealed that the DEGs were significantly enriched in synapse organization, negative regulation of neuron differentiation, and negative regulation of neurogenesis (**Figure 2C**, **Supplementary Table S3**). Enriched pathways were highly associated with neural activities but slightly differed from the results of the analysis of BDNF$^{+/-}$ vs. control which indicated that the main pathways in BDNF$^{+/-}$ vs. CUMS and BDNF$^{+/-}$ vs. control were the same. The current study found genes similar to those identified in BDNF$^{+/-}$ vs. control pathway enrichment, including Mib1, Sema3c, and Foxo3, which were still enriched in relevant negative regulation of neuron activities, indicating that the CUMS model did not differ significantly from the controls as compared with BDNF$^{+/-}$ (**Supplementary Figure S9**).

Network Analysis of the Protein–Protein Interaction

PPI differences between BDNF$^{+/-}$, a series of strong protein interactions, and IDO1$^{-/-}$ were not focused or related to neural activities, whereas internal consistency was similar between the control and CUMS groups.

To analyze the interactions with other molecules, we performed PPI based on the DEGs. Results in the BDNF$^{+/-}$ vs. control groups and the PPI network of DEGs revealed highly confident interactions which illustrated a series of strong interactions between proteins in BDNF$^{+/-}$ mice (**Figure 3A**). It was found that the whole network includes 171 links with the highest confidence among 60 nodes (score: >700). In addition, the whole network was mainly connected using several hub genes, including Trp53, Foxo3, EGFR, and CDK families. Furthermore, Trp53 responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, and senescence, as well as commonly interacts with CDKs which indicate cell cycle regulation changes in BDNF$^{+/-}$ mice (Rufini et al., 2013).

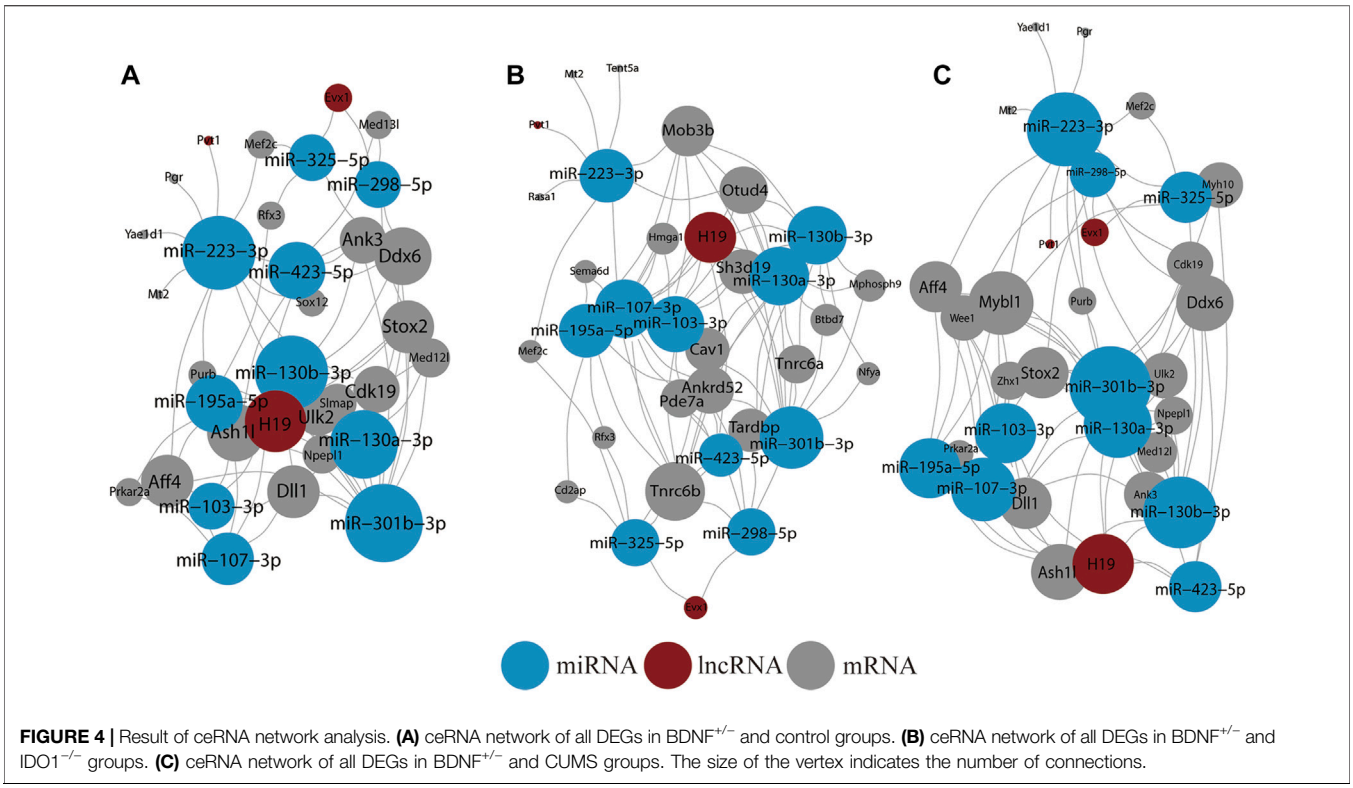

FIGURE 4 | Result of ceRNA network analysis. **(A)** ceRNA network of all DEGs in BDNF$^{+/-}$ and control groups. **(B)** ceRNA network of all DEGs in BDNF$^{+/-}$ and IDO1$^{-/-}$ groups. **(C)** ceRNA network of all DEGs in BDNF$^{+/-}$ and CUMS groups. The size of the vertex indicates the number of connections.

In BDNF$^{+/-}$ vs. IDO1$^{-/-}$, it was found that the PPI network contained 178 links and 60 nodes (**Figure 3B**). Notably, the network had three dense subnetworks of nearly equal size. The densest was mostly composed of Rpl family genes, including Rpl36a, Rpl38, and Rpl39. Furthermore, the Rpl family is composed of L ribosomal proteins. It was found that between the other two subnetworks one was led by Cdk2, P1k1, and Psmb10, and the other was led by Ndufb6, Ndufb4, Ndufb9, and the relevant gene of the NADH dehydrogenase subunit. The three subnetworks showed a dispersion in different biological functions, indicating that PPI differences between BDNF$^{+/-}$ vs. IDO1$^{-/-}$ are not focused or related to neural activities.

It was found that in BDNF$^{+/-}$ vs. CUMS, the network was composed of the top 60 DEGs with 163 interaction links (**Figure 3C**). The results of the PPI network revealed only one cluster of similar topology to the one in BDNF$^{+/-}$ vs. control, as well as similar hub genes, including Trp53, EGRF, Fox, Foxo3, and CDKs, reflecting consistent similarity between control and CUMS. However, it contained other hub genes, including Uba52, Bdnf, and Zap70.

Network Analysis of lncRNA–miRNA–mRNA ceRNA

In BDNF$^{+/-}$ vs. control, BDNF$^{+/-}$ vs. CUMS, and BDNF$^{+/-}$ vs. IDO1$^{-/-}$ mice, most differentially expressed genes were associated with the protection of vulnerable neuronal circuits. To investigate the potential interactions between DEGs and lncRNAs, we analyzed ceRNA based on DEGs among different models. For each comparison, lncRNAs and miRNAs that may interact with the DEGs were identified, and relevant interaction networks were built.

The BDNF$^{+/-}$ vs. control lncRNA-mRNA data were obtained from lncRNA2 targets. The lncRNA-mRNA network revealed 150 interactions between 40 DEGs and 46 lncRNAs (**Supplementary Figure S10A**). Lnpep, Slc36a4, and Amy1 interacted with most lncRNAs whereas AK040954, Linc-RAM, H19, and Linc1388 targeted most mRNAs. The hub genes in the miRNA-mRNA network included miR-124-3p, miR-132-3p, and miR-9-5p in miRNA and Dyrk2 as well as Nr2c2 and Nbeal1 in mRNA. miR-124-3p, which had the most connections in the current study, is a well-known biomarker of neural diseases (**Supplementary Figure S10B**).

A ceRNA network was further reconstructed (**Figure 4A**). In addition, it was noted that the network included lncRNAs H19, Evx1, and Pvt1, whereby H19 connected most miRNAs. The hub miRNAs included miR-130a-3p, miR-130b-3p, miR-223-3p, miR-423-5p, and miR-301b-3p whereas the hub mRNAs included Stox2, Ulk2, Npepl1, Aff4, and Ddx6.

In BDNF$^{+/-}$ vs. IDO1$^{-/-}$, the lncRNA-mRNA network was composed of 147 interactions between 39 DEGs and 44 lncRNAs (**Supplementary Figure S11A**). Myh9, Adam12, Iqgap1, and Tfrc interacted with most lncRNAs whereas linc1388, linc1382, linc1470, and linc1558 targeted most mRNAs. In the miRNA-mRNA network, miR-124-3p, miR-30e-5p, and miR-30a-5p connected with most mRNAs, whereas Ptpn13, Tfrc, Zfp36l1, and Myh9 connected with most miRNAs (**Supplementary Figure S11B**).

The mRNA–miRNA–lncRNA ceRNA network had three lncRNA nodes, 20 mRNA nodes, and 10 miRNA nodes (**Figure 4B**). The lncRNA nodes with the most connections were

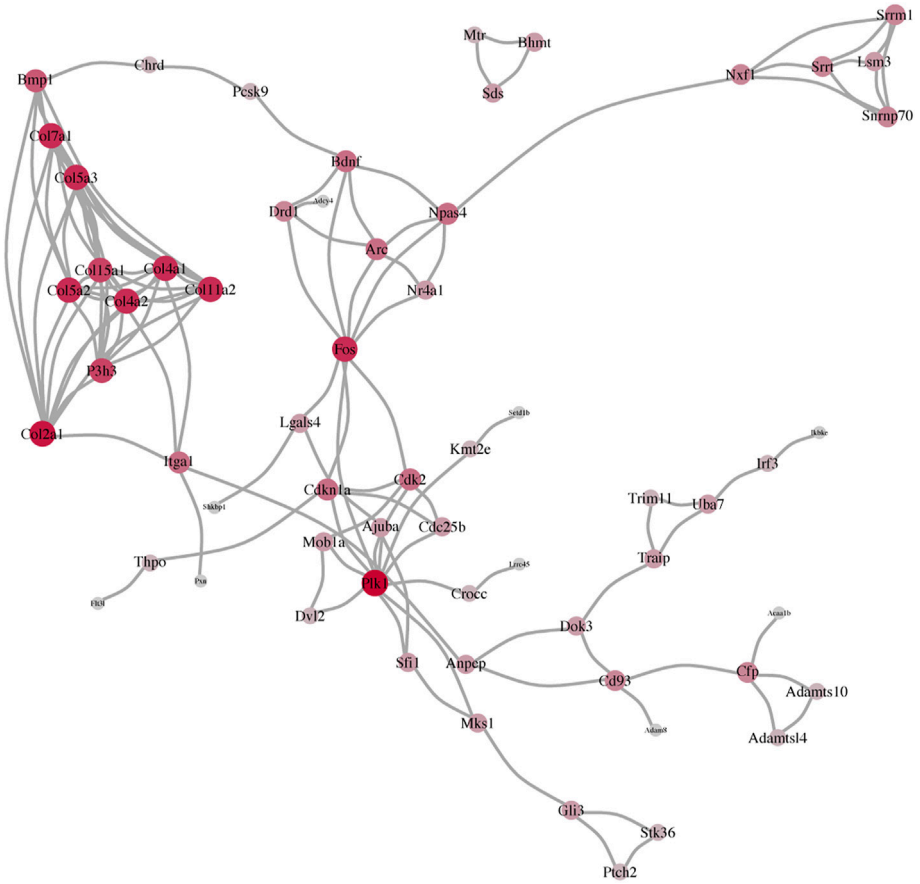

FIGURE 5 | Consistency of the three PPI networks, including BDNF$^{+/-}$ vs. control, BDNF$^{+/-}$ vs. IDO1$^{-/-}$, and BDNF$^{+/-}$ vs. CUMS.

H19, Evx1, and Pvt1 as compared with BDNF$^{+/-}$ vs. control. Furthermore, the miRNA nodes included miR-107-3p, miR-130b-3p, miR-130a-3p, miR-195a-5p, miR-301b-3p, and miR-103-3p. The average connection per miRNA was higher than in BDNF$^{+/-}$ vs. control. The hub mRNAs included Tnrc6b, Mob3b, Otud4, Ankrd52m, Tardbp, Sh3d19, and Cav1, and it was found that they had little overlap with results from BDNF$^{+/-}$ vs. control.

In BDNF$^{+/-}$ vs. CUMS, the lncRNA-mRNA network showed 139 interactions between 37 DEGs and 46 lncRNAs (**Supplementary Figure S12A**). Lnpep, Slc36a4, and Amy1 still interacted with most lncRNAs, whereas AK040954, Linc-RAM, H19, and Linc1388 targeted most of the mRNAs. In the miRNA-mRNA network, hub miRNAs included miR-124-3p, miR-106-5p, miR-132-3p, and miR-9-5p, whereas hub mRNAs included Dyrk2, Nr2c2, Nbeal1, and Ptbp1 (**Supplementary Figure S12B**). In the mRNA–miRNA–lncRNA network, lncRNA nodes still included H19, Evx1, and Pvt1, with H19 still having the most connections (**Figure 4C**). Furthermore, the hub miRNAs included miR-301b-3p, miR-223-3p, miR-130a-3p, miR-130b-3p, and miR-223-3p whereas the hub mRNA gene included Mybl1, Ddx6, Aff4, Stox2, and Ddx6, which was similar to BDNF$^{+/-}$ vs. control.

Following the consistency of the aforementioned three PPI networks, we determined the mRNA network of the BDNF$^{+/-}$ vs. control, BDNF$^{+/-}$ vs. IDO1$^{-/-}$, and BDNF$^{+/-}$ vs. CUMS mice. It showed that BDNF was a common difference between them, which was also in line with the differential expression of the prefrontal lobe after the knockdown of the Bdnf. Among them, we found that not only was the upstream Bmp1 of Bdnf different but also the downstream Fos of Bdnf and Fos was also an important indicator of activating neuronal activity (**Figure 5**).

DISCUSSION

The current study identified several differentially expressed genes in normal vs. depression-like mouse tissues from diverse genomic locations. These genes were collected in an mPFC manner. Pathway enrichment and ceRNA network analyses evidently revealed that most differentially expressed genes were associated with the protection of vulnerable neuronal circuits, and enriched pathways were associated with nervous system development and synapse organization.

Consistent with several previous studies, it was found that there were no significant gene expression differences in control vs. CUMS mice (Ma et al., 2016; Ma et al., 2019). It was evident that the possible differences are not reflected at the transcriptomic level but in protein

modification or neurotransmitter content. However, it was found that BDNF-knockdown mice exhibited depression-like features based on reduced levels of neurotransmitter content (Kojima et al., 2020). Furthermore, it was found that the $BDNF^{+/-}$ mice exhibited significant gene expression differences as compared with control or $IDO1^{-/-}$ mice.

It was also evident that various genes, including Ptbp1, were predominantly expressed in $BDNF^{+/-}$ as compared with other groups which suggested that they were purposefully produced. This study focused on mouse mPFC sequencing of gene modification, especially in $BDNF^{+/-}$ and $IDO1^{-/-}$ mice. Other previous studies have reported more differential mRNA expressions in the hippocampus, and there are possibilities of molecular lateralization in other subcortical areas (Hu et al., 2020; Chae et al., 2021). Furthermore, various abundant genes are specifically expressed in the gene-editing group and differentially expressed in the depression-like group as compared with the normal or depression-like antagonism groups, hence indicating that they serve specific functions in specific pathways (Le et al., 2018; Xu et al., 2019).

The current study had some limitations. The first limitation was the lack of sequencing comparison between other brain regions such as the hippocampus of the limbic system or the parahippocampal gyrus and cingulate gyrus. The lack of comparison of human samples was also a shortcoming of this study. Adding human-derived depression samples would have enriched the understanding of the degree of gene expression changes in depression-like lesions. Overall, the current study only performed RNA sequencing studies on a CUMS-based depression-like mouse model, BDNF knockdown mice (simulating depression-like), and IDO knock-out mice (antagonizing depression-like). The sample area was the prefrontal cortex, and because no human samples were analyzed for comparison in the current study, there was no experimental verification of whether the differential gene expression, including Ptbp1, is associated with depression pathogenesis.

Nevertheless, the results of the current study suggest that in a mouse model of depression ($BDNF^{+/-}$), CXCL1 deletion (Chai et al., 2019) and Slc17a7 reduction (Lindstrom et al., 2020) are related to the loss of excitatory neurons in the prefrontal lobe, whereas Ptbp1 downregulation (Qian et al., 2020) correlates with neuronal regeneration. However, there is a need for experimental validation of these findings in future research.

CONCLUSION

Depression mouse models and controls were studied for possible DEGs and enriched pathways. The findings show a function for ceRNA network-mediated genes in the development of depression. There is a difference in the expression between $BDNF^{+/-}$ and CUMS model depressed mice, showing that the BDNF knockout model can only assist in imitating neurotransmitter models. A neurotransmitter disruption was not seen in the $IDO1^{-/-}$ mouse model, in contrast to the CUMS and $BDNF^{+/-}$ models. Our findings may help unravel the neurotransmitter hypothesis of depression in animals.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI repository, accession number PRJNA825298.

ETHICS STATEMENT

The animal study was reviewed and approved by the National Institutional Animal Care and Ethical Committee of Southern Medical University.

AUTHOR CONTRIBUTIONS

JX and PH conceived the project, designed the experiments, and edited the manuscript; JR and CL performed animal experiments and contributed to the writing of the manuscript and figures; JR and SW analyzed and interpreted enrichment DEGs; and CL and YH constructed the PPI network. All authors read and approved the submitted version.

FUNDING

The work was supported by the National Natural Science Foundation of China (81901362), the Postdoctoral Research Foundation of China (2019M652969), the Traditional Chinese Medicine Bureau of Guangdong Province Scientific Research Project (20201340), the Bureau of Science and Technology of Foshan City Project (1920001000269), and the Innovation Project of Women and Children Medical Research Center affiliated to the Foshan Institute of Fetal Medicine (FEYJZX-2021-003 and FEYJZX-2021-004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.890961/full#supplementary-material

REFERENCES

Amidfar, M., Kim, Y.-K., and Wiborg, O. (2018). Effectiveness of Memantine on Depression-like Behavior, Memory Deficits and Brain mRNA Levels of BDNF and TrkB in Rats Subjected to Repeated Unpredictable Stress. *Pharmacol. Rep.* 70, 600–606. doi:10.1016/j.pharep.2017.12.007

Auerbach, R. P., Mortier, P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., et al. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and Distribution of Mental Disorders. *J. Abnormal Psychol.* 127, 623–638. doi:10.1037/abn0000362

Chae, S., Hong, J., Kang, K., Shin, A., Kim, D.-G., Lee, S., et al. (2021). Molecular Laterality Encodes Stress Susceptibility in the Medial Prefrontal Cortex. *Mol. Brain* 14, 92. doi:10.1186/s13041-021-00802-w

Chai, H. H., Fu, X. C., Ma, L., Sun, H. T., Chen, G. Z., Song, M. Y., et al. (2019). The Chemokine CXCL1 and its Receptor CXCR2 Contribute to Chronic Stress-induced Depression in Mice. *FASEB j.* 33, 8853–8864. doi:10.1096/fj.201802359RR

Chaves Filho, A. J. M., Lima, C. N. C., Vasconcelos, S. M. M., de Lucena, D. F., Maes, M., and Macedo, D. (2018). Ido Chronic Immune Activation and Tryptophan Metabolic Pathway: A Potential Pathophysiological Link between Depression and Obesity. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* 80, 234–249. doi:10.1016/j.pnpbp.2017.04.035

Colucci-D'Amato, L., Speranza, L., and Volpicelli, F. (2020). Neurotrophic Factor BDNF, Physiological Functions and Therapeutic Potential in Depression, Neurodegeneration and Brain Cancer. *Ijms* 21, 7777. doi:10.3390/ijms21207777

Devendorf, A., Bender, A., and Rottenberg, J. (2020). Depression Presentations, Stigma, and Mental Health Literacy: A Critical Review and YouTube Content Analysis. *Clin. Psychol. Rev.* 78, 101843. doi:10.1016/j.cpr.2020.101843

Dong, B. E., Xue, Y., and Sakata, K. (2018). The Effect of Enriched Environment across Ages: A Study of Anhedonia and BDNF Gene Induction. *Genes, Brain Behav.* 17, e12485. doi:10.1111/gbb.12485

Fan, T., Hu, Y., Xin, J., Zhao, M., and Wang, J. (2020). Analyzing the Genes and Pathways Related to Major Depressive Disorder via a Systems Biology Approach. *Brain Behav.* 10, e01502. doi:10.1002/brb3.1502

Gao, L., Gao, T., Zeng, T., Huang, P., Wong, N.-K., Dong, Z., et al. (2021). Blockade of Indoleamine 2, 3-dioxygenase 1 Ameliorates Hippocampal Neurogenesis and BOLD-fMRI Signals in Chronic Stress Precipitated Depression. *Aging* 13, 5875–5891. doi:10.18632/aging.202511

Gao, L., Huang, P., Dong, Z., Gao, T., Huang, S., Zhou, C., et al. (2018). Modified Xiaoyaosan (MXYS) Exerts Anti-depressive Effects by Rectifying the Brain Blood Oxygen Level-dependent fMRI Signals and Improving Hippocampal Neurogenesis in Mice. *Front. Pharmacol.* 9, 1098. doi:10.3389/fphar.2018.01098

Hu, J., Cao, S., Zhang, Z., Wang, L., Wang, D., Wu, Q., et al. (2020). Effects of Caffeic Acid on Epigenetics in the Brain of Rats with Chronic Unpredictable Mild Stress. *Mol. Med. Rep.* 22, 5358–5368. doi:10.3892/mmr.2020.11609

Huang, P., Dong, Z., Huang, W., Zhou, C., Zhong, W., Hu, P., et al. (2017). Voluntary Wheel Running Ameliorates Depression-like Behaviors and Brain Blood Oxygen Level-dependent Signals in Chronic Unpredictable Mild Stress Mice. *Behav. Brain Res.* 330, 17–24. doi:10.1016/j.bbr.2017.05.032

Jiang, N., Lv, J.-w., Wang, H.-x., Lu, C., Wang, Q., Xia, T.-j., et al. (2019). Dammarane Sapogenins Alleviates Depression-like Behaviours Induced by Chronic Social Defeat Stress in Mice through the Promotion of the BDNF Signalling Pathway and Neurogenesis in the hippocampus. *Brain Res. Bull.* 153, 239–249. doi:10.1016/j.brainresbull.2019.09.007

Jiang, X., Lin, Q., Xu, L., Chen, Z., Yan, Q., Chen, L., et al. (2020). Indoleamine-2,3-Dioxygenase Mediates Emotional Deficits by the Kynurenine/Tryptophan Pathway in the Ethanol Addiction/Withdrawal Mouse Model. *Front. Cel. Neurosci.* 14, 11. doi:10.3389/fncel.2020.00011

Kang, H.-J., Park, Y., Yoo, K.-H., Kim, K.-T., Kim, E.-S., Kim, J.-W., et al. (2020). Sex Differences in the Genetic Architecture of Depression. *Sci. Rep.* 10, 9927. doi:10.1038/s41598-020-66672-9

Kojima, M., Otabi, H., Kumanogoh, H., Toyoda, A., Ikawa, M., Okabe, M., et al. (2020). Reduction in BDNF from Inefficient Precursor Conversion Influences Nest Building and Promotes Depressive-like Behavior in Mice. *Ijms* 21, 3984. doi:10.3390/ijms21113984

Kowiański, P., Lietzau, G., Czuba, E., Waśkow, M., Steliga, A., and Moryś, J. (2018). BDNF: A Key Factor with Multipotent Impact on Brain Signaling and Synaptic Plasticity. *Cell Mol Neurobiol* 38, 579–593. doi:10.1007/s10571-017-0510-4

Le, T. T., Savitz, J., Suzuki, H., Misaki, M., Teague, T. K., White, B. C., et al. (2018). Identification and Replication of RNA-Seq Gene Network Modules Associated with Depression Severity. *Transl Psychiatry* 8, 180. doi:10.1038/s41398-018-0234-3

Lima Giacobbo, B., Doorduin, J., Klein, H. C., Dierckx, R. A. J. O., Bromberg, E., and de Vries, E. F. J. (2019). Brain-Derived Neurotrophic Factor in Brain Disorders: Focus on Neuroinflammation. *Mol. Neurobiol.* 56, 3295–3312. doi:10.1007/s12035-018-1283-6

Lindström, S. H., Sundberg, S. C., Larsson, M., Andersson, F. K., Broman, J., and Granseth, B. (2020). VGluT1 Deficiency Impairs Visual Attention and Reduces the Dynamic Range of Short-Term Plasticity at Corticothalamic Synapses. *Cereb. Cortex* 30, 1813–1829. doi:10.1093/cercor/bhz204

Ma, K., Guo, L., Xu, A., Cui, S., and Wang, J.-H. (2016). Molecular Mechanism for Stress-Induced Depression Assessed by Sequencing miRNA and mRNA in Medial Prefrontal Cortex. *PLoS One* 11, e0159093. doi:10.1371/journal.pone.0159093

Ma, K., Zhang, H., Wei, G., Dong, Z., Zhao, H., Han, X., et al. (2019). Identification of Key Genes, Pathways and miRNA/mRNA Regulatory Networks of CUMS-Induced Depression in Nucleus Accumbens by Integrated Bioinformatics Analysis. *Ndt* 15, 685–700. doi:10.2147/NDT.S200264

Menezes, J., Souto Das Neves, B.-H., Gonçalves, R., Benetti, F., and Mello-Carpes, P. B. (2020). Maternal Deprivation Impairs Memory and Cognitive Flexibility, Effect that Is Avoided by Environmental Enrichment. *Behav. Brain Res.* 381, 112468. doi:10.1016/j.bbr.2020.112468

Qian, H., Kang, X., Hu, J., Zhang, D., Liang, Z., Meng, F., et al. (2020). Reversing a Model of Parkinson's Disease with *In Situ* Converted Nigral Neurons. *Nature* 582, 550–556. doi:10.1038/s41586-020-2388-4

Rufini, A., Tucci, P., Celardo, I., and Melino, G. (2013). Senescence and Aging: the Critical Roles of P53. *Oncogene* 32, 5129–5143. doi:10.1038/onc.2012.640

Song, A.-Q., Gao, B., Fan, J.-J., Zhu, Y.-J., Zhou, J., Wang, Y.-L., et al. (2020). NLRP1 Inflammasome Contributes to Chronic Stress-Induced Depressive-like Behaviors in Mice. *J. Neuroinflammation* 17, 178. doi:10.1186/s12974-020-01848-8

White, C. J., Ellis, J. M., and Wolfgang, M. J. (2021). The Role of Ethanolamine Phosphate Phospholyase in Regulation of Astrocyte Lipid Homeostasis. *J. Biol. Chem.* 297, 100830. doi:10.1016/j.jbc.2021.100830

Xu, Y.-Y., Xia, Q., Xia, Q., Zhang, X., and Liang, J. (2019). MicroRNA-Based Biomarkers in the Diagnosis and Monitoring of Therapeutic Response in Patients with Depression. *Ndt* 15, 3583–3597. doi:10.2147/NDT.S237116

Zhang, X., Li, H., Sun, H., Jiang, Y., Wang, A., Kong, Y., et al. (2020). Effects of BDNF Signaling on Anxiety-Related Behavior and Spatial Memory of Adolescent Rats in Different Length of Maternal Separation. *Front. Psychiatry* 11, 709. doi:10.3389/fpsyt.2020.00709

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fgene.2022.905619

Association of Single-Nucleotide Polymorphisms of rs2383206, rs2383207, and rs10757278 With Stroke Risk in the Chinese Population: A Meta-analysis

Xuemei Hu[1,2‡], *Dongsen Wang*[1,2‡], *Chunying Cui*[2] *and Qingjian Wu*[2*†]

[1]*Clinical Medical College of Jining Medical University, Jining, China,* [2]*Department of Emergency, Jining No. 1 People's Hospital, Jining, China*

OPEN ACCESS

Edited by:
Guiyou Liu,
Tianjin Institute of Industrial Biotechnology (CAS), China

Reviewed by:
Yiyan Li,
University of Chinese Academy of Sciences, China
Xingli Xu,
Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, China

***Correspondence:**
Qingjian Wu
wqw110@163.com

†ORCID:
Qingjian Wu
orcid.org/0000-0002-8746-8743

‡*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: *27 March 2022*
Accepted: *13 May 2022*
Published: *28 June 2022*

Citation:
Hu X, Wang D, Cui C and Wu Q (2022) Association of Single-Nucleotide Polymorphisms of rs2383206, rs2383207, and rs10757278 With Stroke Risk in the Chinese Population: A Meta-analysis. Front. Genet. 13:905619. doi: 10.3389/fgene.2022.905619

Several studies have reported that chromosome 9p21 is significantly associated with ischemic stroke (IS) risk, with the G allele associated with increased risk. However, controversial results have been reported in the literature. We systematically assessed the relationship between stroke and three 9p21 loci (rs2383206, rs2383207, and rs10757278) in this meta-analysis. First, we searched the PubMed and Embase databases for relevant studies. We then calculated odds ratios using the chi-squared test. The evaluation of experimental data was performed using bias tests and sensitivity analyses. We analyzed data from 16 studies involving 18,584 individuals of Chinese ancestry, including 14,033 cases and 14,656 controls. Our results indicated that chromosome 9p21 is significantly associated with IS (odds ratio: 1.15, 95% confidence interval: 1.1–1.20, $p < 0.0001$). Because the three single-nucleotide polymorphisms (rs2383206, rs2383207, and 10757278) have a linkage disequilibrium relationship, all three may increase the risk of IS.

Keywords: ischemic stroke, chromosome 9p21, rs2383206, rs2383207, rs10757278, Chinese

INTRODUCTION

Stroke is a severe disease and is the leading cause of disability and death in China (Liu et al., 2011). It is an acute cerebrovascular disease that is characterized by focal loss of nerve function and high mortality and disability, and it currently poses a serious threat to human life and health (Li et al., 2021). Stroke is thought to be caused by environmental risk factors, multiple genes, and their interactions. To date, however, a large proportion of stroke risk remains unexplained (Ganesh et al., 2016). Genetic variation on chromosome 9p21 is widely believed to be linked to risk of coronary heart disease (McPherson et al., 2007; Samani et al., 2007), but it has a different role in stroke (Matarin et al., 2008; Gschwendtner et al., 2009). Previously, genome-wide association studies (GWAS) have analyzed genes associated with ischemic stroke (IS) (Söderholm et al., 2019). Single-nucleotide polymorphisms (SNPs) of rs2383206, rs2383207, and rs10757278 on chromosome 9p21 are linked to stroke. However, although several recent genetic studies have reported that chromosome 9p21 plays an important role in the mechanism of stroke, studies of different races and from different geographic locations have provided very different results. Therefore, an association between 9p21 polymorphisms and stroke has been established for individuals of European descent; the main aim of this meta-analysis was to study the relationship between three SNPs on chromosome 9p21 and stroke in the Chinese population.

TABLE 1 | Sixteen studies in 11 articles investigating the association between rs2383207, rs2383206, and rs10757278 and IS.

SNP	First author; year	Population	Case	Control	Case genotype			Control genotype		
					GG	GA	AA	GG	GA	AA
rs2383207	Lin-2011	Chinese	627	1,349	288	274	65	568	609	172
	Jin-2021	Chinese	1,640	1755	795	665	180	815	690	250
	Yang-2018	Chinese	550	548	236	237	77	244	251	53
	Li-2017	Chinese	1,429	1,191	633	642	154	492	525	174
	Li-2021[1]	Chinese	987	946	480	425	82	410	407	129
	Zhang-2012[2]	Chinese	1,657	1,664	700	743	214	652	796	216
rs2303206	Hua-2009	Chinese	352	423	67	188	97	78	191	154
	Ding-2009[4]	Chinese	440	498	113	213	114	94	264	140
	Li-2021[1]	Chinese	1,006	949	233	493	280	197	447	305
	Xiong-2018[3]	Chinese	200	205	48	96	56	46	98	61
	Zhang-2012[2]	Chinese	1,657	1,664	379	802	476	317	833	514
rs10757278	Bi-2015	Chinese	116	118	29	49	38	15	47	56
	Han-2020	Chinese	505	652	149	235	121	140	310	203
	Xiong-2018[3]	Chinese	200	205	52	95	53	47	99	59
	Ding-2009[4]	Chinese	441	501	40	181	220	45	236	220
	Zhang-2021[2]	Chinese	1,657	1,664	509	774	374	420	832	412

Note: The same numbers indicate the same article. SNP, single-nucleotide polymorphism.

METHODS

Search Strategy

We searched the PubMed and Embase databases and selected all possible studies using the keywords "Stroke Chinese" and "rs2383206," "rs2383207," "rs10757278," and "9p21." The relevant literature was updated on 31 January 2022.

Selection Criteria

The following selection criteria were used: (1) an association between the proposed SNPs and stroke was evaluated using a case–control design; (2) an accurate genotype number was provided or could be calculated (Liu et al., 2013); (3) the odds ratio (OR) and 95% confidence interval (CI) were provided to measure the risk of disease; (4) the OR value and 95% CI were calculated by providing enough data; (5) the same diagnostic criteria were used for stroke. The exclusion criteria were (1) the research was presented as a poster presentation, summary, meta-analysis, conference summary or article, or case series analysis; (2) the study was not performed in a Chinese population; (3) the three SNPs were not used; (4) the study was not consistent with the research topic; and (5) the exact number of genotypes was not provided and could not be calculated and/or the OR and 95% CI were not provided and could not be calculated. Two authors (DW and XH) independently screened all studies by their title or abstract and then evaluated the full text. Any differences in opinion were resolved through discussion.

Data Extraction

Trial data from each identified study were extracted separately by two investigators (DW and XH). Any differences were eliminated by discussing the data extraction for each study using standard data collection tables. The data and information that were extracted for inclusion in the analysis included the first author's name, publication year, language, population, study type, sample size, numbers, and frequencies of rs2383206, rs2383207, and rs10757278 polymorphism genotypes in the cases and controls, ORs, and 95% CIs. All extracted data are presented in **Tables 1**, **2**.

Statistical Analysis

We investigated the Hardy–Weinberg equilibrium of rs2383206, rs2383207, and rs10757278. We also investigated their association with stroke using the chi-squared test, which was performed using R (http://www.r-project.org/) (Liu et al., 2013). For the meta-analysis, we determined the heterogeneity among datasets using Cochran's Q test and $I^2 = {}^{(}Q - (k - 1^{))}/_{Q} \times 100\%$. The Q statistic approximately follows a $\chi 2$ distribution, with k-1 degrees of freedom (k is the number of studies in the analysis) (Liu et al., 2017). When I^2 was greater than 50% and the p-value was less than 0.1 (Higgins et al., 2021), the DerSimonian and Laird random-effects model was used as the pooling method; otherwise, the Mantel–Haenszel or inverse variance fixed-effects model was used as the pooling method, as appropriate. We also used funnel plots to assess potential publication bias. When there is no bias, funnel plots are symmetrical; conversely, when bias is present, funnel plots are asymmetrical (Liu et al., 2014).

RESULTS

Characteristics of Included Studies

In this meta-analysis, 18,584 participants were included: 14,033 in the IS group (7,235 cases with rs2383207, 3,762 cases with rs2383206, and 3,036 cases with rs10757278) and 14,656 cases in the control group (7,653 cases with rs2383207, 3,807 cases with rs2383206, and 3,196 cases with rs10757278). Eleven articles were selected, comprising 16 studies, of which six investigated rs2383207 (Lin et al., 2011; Zhang et al., 2012; Li et al., 2017; Yang et al., 2018; Jin et al., 2021; Li et al., 2021), five investigated rs2383206 (Ding et al., 2009; Hu et al., 2009; Zhang et al., 2012; Xiong et al., 2018; Li et al., 2021), and five investigated rs10757278

TABLE 2 | Correlation analysis between different genetic patterns of rs2383207, rs2383206, and rs10757278 at 9p21 locus and IS susceptibility.

SNP	First author; year	G (case/control)	A (case/control)	OR	95% CI	SE (ln (OR))
rs2383207	Lin-2011	850/1745	404/953	1.15	0.997 ~ 1.325	0.073
	Jin-2021	2255/2320	1,025/1,190	1.13	1.019 ~ 1.249	0.052
	Yang-2018	709/391	739/357	0.88	0.734 ~ 1.045	0.09
	Li-2017	1908/1,509	950/873	1.16	1.037 ~ 1.302	0.058
	Li-2021[1]	1,385/1,227	589/665	1.27	1.114 ~ 1.458	0.069
	Zhang-2012[2]	2143/2100	1,171/1,228	1.07	0.968 ~ 1.183	0.051
rs2303206	Hua-2009	322/347	382/499	1.21	0.988 ~ 1.479	0.103
	Ding-2009[4]	439/452	441/544	1.19	0.999 ~ 1.437	0.093
	Li-2021[1]	959/841	1,053/1,057	1.14	1.009 ~ 1.298	0.064
	Xiong-2018[3]	192/190	208/220	1.07	0.811 ~ 1.408	0.141
	Zhang-2012[2]	1,560/1,467	1754/1801	1.12	1.024 ~ 1.243	0.049
rs10757278	Bi-2015	107/77	125/159	1.75	1.199 ~ 2.541	0.192
	Han-2020	533/590	477/714	1.35	1.147 ~ 1.594	0.084
	Xiong-2018[3]	199/193	201/217	1.11	0.845 ~ 1.467	0.141
	Ding-2009[4]	261/326	621/676	0.87	0.716 ~ 1.060	0.100
	Zhang-2021[2]	1792/1,672	1,522/1,656	1.17	1.059 ~ 1.284	0.049

Note: The same numbers indicate the same article; SNP, single-nucleotide polymorphism; OR, odds ratio; CI, confidence interval; SE, standard error.

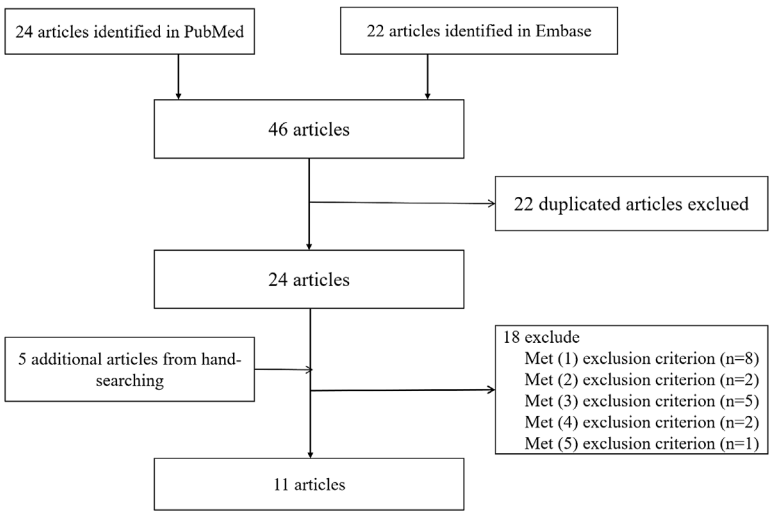

FIGURE 1 | Flow chart of study selection in this meta-analysis.

(Ding et al., 2009; Zhang et al., 2012; Bi et al., 2015; Xiong et al., 2018; Han et al., 2020). The study identification and selection process is shown in detail in **Figure 1**.

Linkage Disequilibrium

The three SNPs—rs10757278, rs2383206, and rs2383207—are located within 10 kb of one another on chromosome 9p21 (https://snipa.helmholtz-muenchen.de/snipa3/).

Meta-Analysis Results of 9p21

There is a linkage disequilibrium relationship among the three SNPS (rs2303206, rs2383207, and rs10757278). Thus, we performed an analysis of the OR values of all studies involving rs2383206, rs2383207, and rs10757278 in which the G allele was a minor allele. Because I^2 = 50%, a random-effects model was used to compare alleles (**Figure 2**). Chromosome 9p21 was significantly associated with IS risk, and the G allele was associated with increased IS risk (OR: 1.14, 95% CI: 1.08–1.19, $p < 0.0001$, **Figure 2**).

Publication Bias

The Harbord test was used to evaluate publication bias. The bias = 0.3228, $p = 0.7661$, indicating no publication bias in the studies of 9p21 (**Figure 3**).

Sensitivity Analysis

Because the I^2 is > 50% in this meta-analysis, a random-effects model was used. To assess the impact of each individual study on the pooled effect estimate, we performed a sensitivity analysis by removing one study at a time. The pooled estimate I^2 = 49.8%;

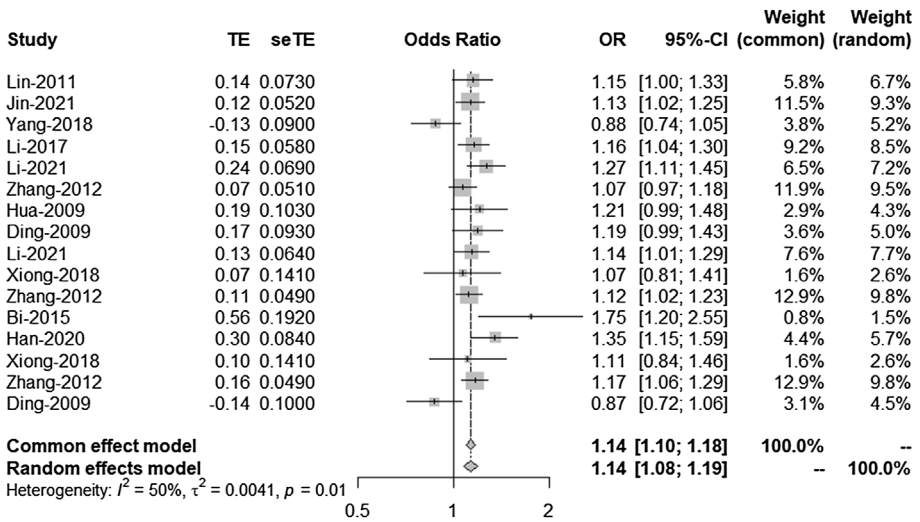

FIGURE 2 | Random-effects meta-analysis of the association between the three single-nucleotide polymorphisms (SNPs; rs2383207, rs2383206, and rs10757278) and ischemic stroke (IS). CI, confidence interval; OR, odds ratios.

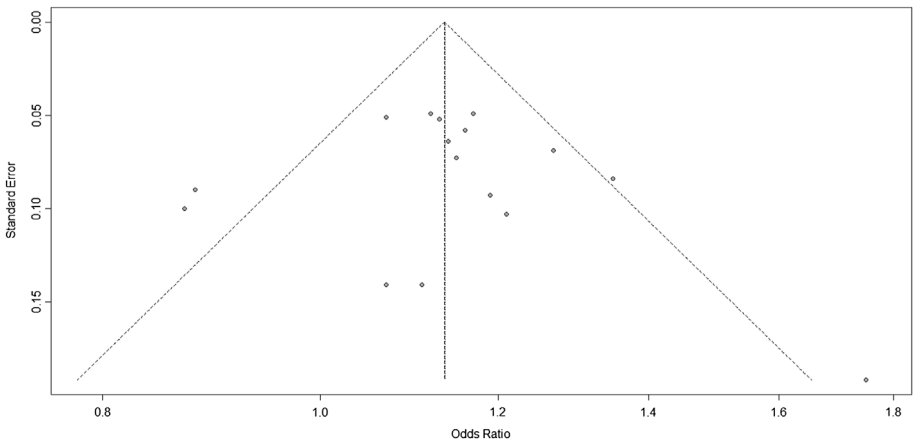

FIGURE 3 | Funnel plots corresponding to the random-effects meta-analysis of all studies of the three single-nucleotide polymorphisms (SNPs; rs2383207, rs2383206, and rs10757278) and ischemic stroke (IS).

thus, no single study significantly affected the results of each single-locus sensitivity analysis.

Second and Third Analyses

According to the results of the bias test and sensitivity analysis, it was found that the studies by Yang et al. (rs2383207) (Yang et al., 2018) and Ding et al. (rs10757278) (Ding et al., 2009) had roughly the same weight and were outside the funnel plot. We decided to remove the two studies and re-analyze the results. After removing two studies, we used R program to re-analyze the remaining studies. In the second analysis, chromosome 9p21 remained significantly associated with IS risk, and the G allele was associated with increased IS risk (OR: 1.16, 95% CI: 1.12–1.20, $p < 0.0001$, **Figure 4**). The results of this second analysis further confirmed that the two removed studies had little influence on the initial results. Moreover, there was homogeneity between the studies ($I^2 = 5\%$, $p = 0.40$), and the two experiments were outliers. A second bias test revealed that bias = 1.4217, $p = 0.0696$ (**Figure 5**). Sensitivity tests for the individual studies were again performed to ensure that no single study significantly affected the results of each single-locus sensitivity analysis.

We know from **Figure 4** that the included studies were homogeneous, and the sensitivity analysis of each study also indicated that no single experiment significantly affected the experimental results. Therefore, based on the forest map and funnel plot, we also removed the study by Bi et al. (Bi et al., 2015), located outside the funnel plot, in the third analysis. In this third analysis, an increased risk of IS was associated with the G allele (OR: 1.16, 95% CI: 1.11–1.20, $p < 0.0001$, **Figure 6**). Further analysis

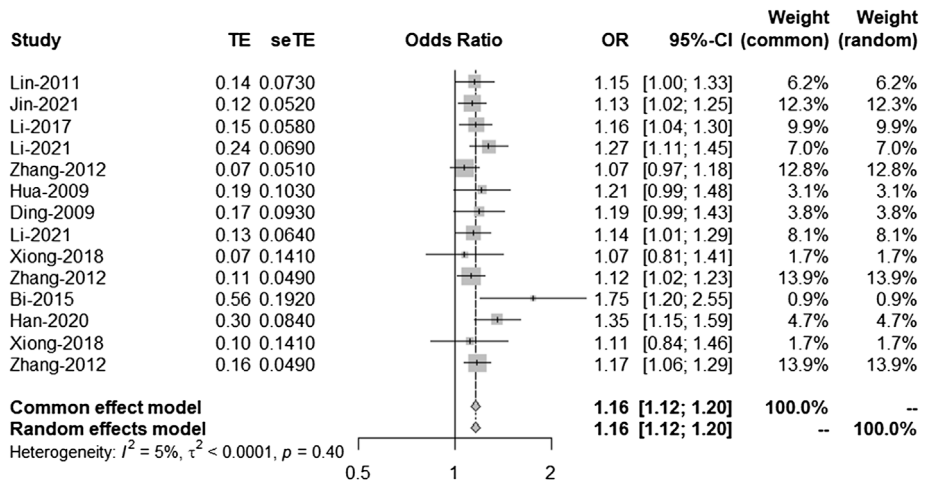

FIGURE 4 | Fixed effects meta-analysis of the association between the three single-nucleotide polymorphisms (SNPs; rs2383207, rs2383206, and rs10757278) and ischemic stroke (IS) in the second analysis. CI, confidence interval; OR, odds ratios.

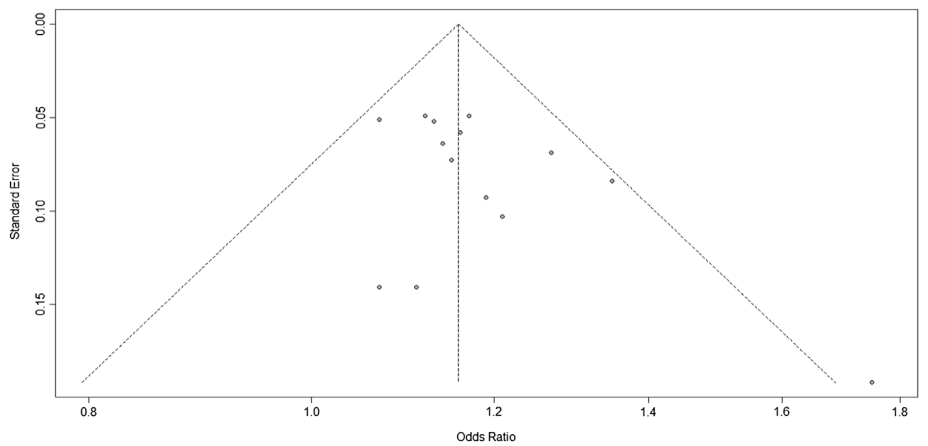

FIGURE 5 | Funnel plots corresponding to the fixed-effects meta-analysis of the three single-nucleotide polymorphisms (SNPs; rs2383207, rs2383206, and rs10757278) and ischemic stroke (IS) in the second analysis.

confirmed that the homogeneity between studies was more significant after removing the study by Bi-2015 (I^2 = 0%, p = 0.71, **Figure 7**), and there was a more significant correlation between chromosome 9p21 and IS risk. Thus, the removal of the study by Bi et al. (2015) further verified our original conclusions. The experimental results indicate that chromosome 9p21 is significantly associated with IS risk, and an increased risk of IS is associated with the G allele.

DISCUSSION

Stroke is currently the main cause of death in China; it has high morbidity, mortality, and disability rates (Kim et al., 2015). Stroke can be clinically divided into two types: IS and hemorrhagic stroke. Among the stroke subtypes, hemorrhagic stroke accounts for 20–40% of strokes in Chinese population, while in most Western populations, the majority of strokes (80–90%) are cerebral infarctions (Reed, 1990). Furthermore, IS accounts for approximately 87% of all stroke types, and IS a multifactorial disease that is influenced by both genetic and environmental factors (Wang et al., 2021). Chromosome 921 was originally reported to be associated with coronary heart disease (Matarin et al., 2008). There are some similarities between the etiologies and mechanisms of coronary heart disease and stroke, and 9p21 variants are associated with both diseases (Matarin et al., 2008). However, when investigating the association between 9p21 and IS, the conclusions drawn by researchers in China and in the rest of the world have been inconsistent. Stroke is influenced by many factors, including genetic, environmental, and vascular risk factors. The main method of studying susceptibility sites and genes in complex diseases is GWAS, based on SNPs (McPherson et al., 2007).

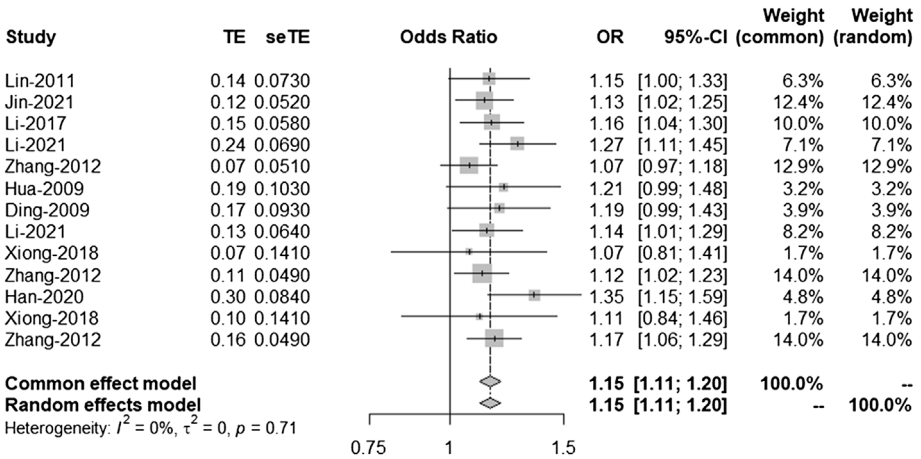

FIGURE 6 | Fixed-effects meta-analysis of the associated between the three single-nucleotide polymorphisms (SNPs; rs2383207, rs2383206, and rs10757278) and ischemic stroke (IS) in the third analysis. CI, confidence interval; OR, odds ratios.

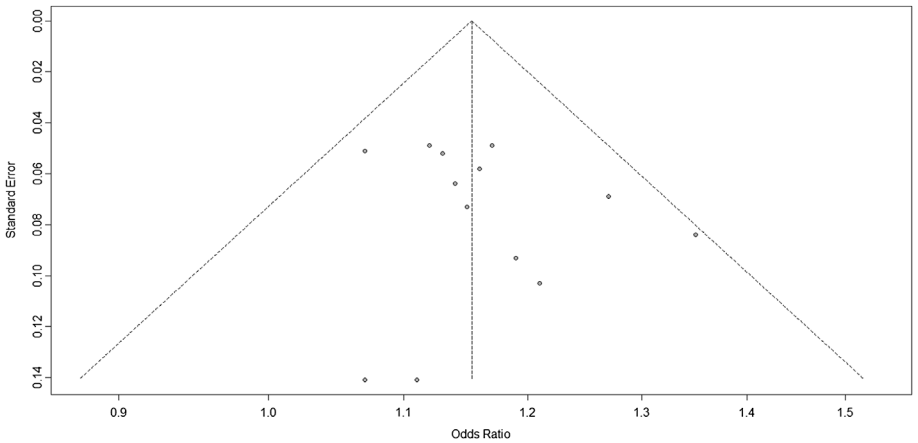

FIGURE 7 | Funnel plots corresponding to the fixed-effects meta-analysis of the three single-nucleotide polymorphisms (SNPs; rs2383207, rs2383206, and rs10757278) and ischemic stroke (IS) in the third analysis.

Matthew Traylor et al. found that chromosome 9p21 and histone deacetylase were associated with stroke in individuals of European ancestry (Traylor et al., 2012). Furthermore, Akinyemi et al. reported that rs2383207 increases IS incidence in indigenous West African men (Akinyemi et al., 2017). Previously, GWAS was also used to demonstrate that the antisense non-coding RNA in the *INK4* locus (ANRIL) variants rs2383207 and rs1333049 increases the risk of IS and coronary heart disease in Caucasian populations (Dichgans et al., 2014; Dehghan et al., 2016). Notably, studies investigating the genetic associations of chromosome 9p21 variants have mainly been performed among Caucasian populations, and relatively few studies have been carried out in Han Chinese populations. Although Chen et al. (2019) studied chromosome 9p21 variants in Chinese populations, they concluded that mutations in rs2383207 may reduce the risk of IS but reported no definite correlation between rs10757278 and IS (Chen et al., 2019). In the present study, we once again focused on the relationship between stroke and chromosome 9p21.

In this meta-analysis, 18,584 participants were included; the IS and control groups contained 14,033 and 14,656 individuals, respectively. The three investigated SNPs have a linkage disequilibrium relationship, and we arrived at the same conclusions through unified analysis. All three SNPs were associated with IS risk. However, there was heterogeneity between the experimental results and studies; thus, bias detection and sensitivity analyses were carried out. **Figure 3** suggested that the research may have been biased; therefore, to remove any possible bias, we performed another set of analyses. These further analyses had similar results that were more significant than those of the original analysis, further confirming that our analysis was correct.

In conclusion, our results indicate that rs2383206, rs2383207, and rs10757278 are significantly associated with IS risk and the G allele is associated with an increased risk of IS. Because the three SNPs in the present study have linkage disequilibrium and are in similar positions on chromosome 9p21, a unified analysis was

performed. Environmental factors such as smoking and alcohol use may also be associated with IS risk, but not all studies considered these risk factors. Therefore, the influence of genes and the environment on IS pathogenesis needs to be further studied.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

DW, XH, CC, and QW participated in the design of this study. XH and DW conducted the literature search. DW, XH, CC, and QW retrieved and selected the articles. DW and XH conducted the data extraction. DW and XH performed the statistical analysis of the data. XH, DW, and CC wrote the manuscript draft. QW supervised the study. All authors contributed to the article and read and approved the final manuscript.

FUNDING

This study was supported by the Shandong Medicine and Health Science Technology Development Program (Grant No. 2018WS470), the Shandong Traditional Chinese Medicine Science and Technology Development Program (Grant No. 2019-0746), the Natural Science Foundation of Shandong Province (Grant No. ZR2021MH133), and the Jining Key Research and Development Project (Grant No. 2020YXNS035). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We are deeply grateful to all the participants in this study. We also thank Bronwen Gardner, PhD, from Liwen Bianji (Edanz) (www.liwenbianji.cn/), for editing the English text of a draft of this manuscript.

REFERENCES

Akinyemi, R., Arnett, D. K., Tiwari, H. K., Ovbiagele, B., Sarfo, F., Srinivasasainagendra, V., et al. (2017). Interleukin-6 (IL-6) rs1800796 and Cyclin Dependent Kinase Inhibitor (CDKN2A/CDKN2B) rs2383207 are Associated with Ischemic Stroke in Indigenous West African Men. *J. Neurological Sci.* 379, 229–235. doi:10.1016/j.jns.2017.05.046

Bi, J., Yang, L., Liu, D., Wu, J., Tong, X., Cen, S., et al. (2015). Sequence Variants on Chromosome 9p21 are Associated with Ischemic Stroke and the Lipids Level in Chinese Han Population. *J. Stroke Cerebrovasc. Dis.* 24, 894–900. doi:10.1016/j.jstrokecerebrovasdis.2014.12.020

Chen, J.-X., Liu, J., Hu, F., Bi, Y., Li, M., and Zhao, L. (2019). Genetic Variants on Chromosome 9p21 Confer Risks of Cerebral Infarction in the Chinese Population: A Meta-Analysis. *Int. J. Immunopathol. Pharmacol.* 33, 2058738419847852. doi:10.1177/2058738419847852

Dehghan, A., Bis, J. C., White, C. C., Smith, A. V., Morrison, A. C., Cupples, L. A., et al. (2016). Genome-Wide Association Study for Incident Myocardial Infarction and Coronary Heart Disease in Prospective Cohort Studies: The CHARGE Consortium. *PLoS One* 11, e0144997. doi:10.1371/journal.pone.0144997

Dichgans, M., Malik, R., König, I. R., Rosand, J., Clarke, R., Gretarsdottir, S., et al. (2014). Shared Genetic Susceptibility to Ischemic Stroke and Coronary Artery Disease: a Genome-Wide Analysis of Common Variants. *Stroke* 45, 24–36. doi:10.1161/strokeaha.113.002707

Ding, H., Xu, Y., Wang, X., Wang, Q., Zhang, L., Tu, Y., et al. (2009). 9p21 is a Shared Susceptibility Locus Strongly for Coronary Artery Disease and Weakly for Ischemic Stroke in Chinese Han Population. *Circ. Cardiovasc Genet.* 2, 338–346. doi:10.1161/circgenetics.108.810226

Ganesh, C., Corey, A., Audrey, C., Myriam, F., Azadeh, R., Joshua, B., et al. (2016). Identification of Additional Risk Loci for Stroke and Small Vessel Disease: A Meta-Analysis of Genome-Wide Association Studies. *Lancet Neurol.* 15, 695–707. doi:10.1016/s1474-4422(16)00102-2

Gschwendtner, A., Bevan, S., Cole, J. W., Plourde, A., Matarin, M., Ross-Adams, H., et al. (2009). Sequence Variants on Chromosome 9p21.3 Confer Risk for Atherosclerotic Stroke. *Ann. Neurol.* 65, 531–539. doi:10.1002/ana.21590

Han, X., Wang, C., Tang, D., Shi, Y., and Gao, M. (2020). Association of Genetic Polymorphisms in Chromosome 9p21 with Risk of Ischemic Stroke. *Cytokine* 127, 154921. doi:10.1016/j.cyto.2019.154921

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., et al. (2021). *Cochran Handbook for Systematic Reviews of Interventions*. version 6.2. Available at: http://www.trainingcochraneorg/handbook (Accessed January 2, 2022).

Hu, W.-l., Li, S.-j., Liu, D.-t., Wang, Y., Niu, S.-q., Yang, X.-c., et al. (2009). Genetic Variants on Chromosome 9p21 and Ischemic Stroke in Chinese. *Brain Res. Bull.* 79, 431–435. doi:10.1016/j.brainresbull.2009.04.001

Jin, M., Zhang, H., Zhang, Q., Li, X., Wu, N., Hu, Y., et al. (2021). Association of Single-Nucleotide Polymorphism on Chromosome 9 and Ischemic Stroke in Heilongjiang Province in China. *Int. J. Clin. Exp. Pathol.* 14, 726–733.

Kim, A. S., Cahill, E., and Cheng, N. T. (2015). Global Stroke Belt: Geographic Variation in Stroke Burden Worldwide. *Stroke* 46, 3564–3570. doi:10.1161/strokeaha.115.008226

Li, R., Zhang, X., Yin, W., Wang, Y., and Liu, Y. (2021). Common Genetic Variants on Chromosome 9p21 Confers Risk of Ischemic Stroke: A Large-Scale Genetic Association Study. *Cell Mol. Biol. (Noisy-le-grand)* 67, 132–137. doi:10.14715/cmb/2021.67.2.20

Li, S., Xu, Y. M., Zheng, H., Randell, E., Wang, H. Z., Cui, J., et al. (2017). A Reduced Interval of Chromosome 9p21 Locus Is Associated with Ischemic Stroke in Chinese Northern Han Population. *Int. J. Genet. Genomics* 5, 14–18. doi:10.11648/j.ijgg.20170501.12

Lin, H.-F., Tsai, P.-C., Liao, Y.-C., Lin, T.-H., Tai, C.-T., Juo, S.-H. H., et al. (2011). Chromosome 9p21 Genetic Variants Are Associated with Myocardial Infarction but Not with Ischemic Stroke in a Taiwanese Population. *J. Investig. Med.* 59, 926–930. doi:10.2310/JIM.0b013e318214ea49

Liu, G., Wang, H., Liu, J., Li, J., Li, H., Ma, G., et al. (2014). The CLU Gene rs11136000 Variant Is Significantly Associated with Alzheimer's Disease in Caucasian and Asian Populations. *Neuromol Med.* 16, 52–60. doi:10.1007/s12017-013-8250-1

Liu, G., Xu, Y., Jiang, Y., Zhang, L., Feng, R., and Jiang, Q. (2017). PICALM rs3851179 Variant Confers Susceptibility to Alzheimer's Disease in Chinese Population. *Mol. Neurobiol.* 54, 3131–3136. doi:10.1007/s12035-016-9886-2

Liu, G., Zhang, S., Cai, Z., Ma, G., Zhang, L., Jiang, Y., et al. (2013). PICALM Gene rs3851179 Polymorphism Contributes to Alzheimer's Disease in an Asian Population. *Neuromol Med.* 15, 384–388. doi:10.1007/s12017-013-8225-2

Liu, L., Wang, D., Wong, K. S. L., and Wang, Y. (2011). Stroke and Stroke Care in China: Huge Burden, Significant Workload, and a National Priority. *Stroke* 42, 3651–3654. doi:10.1161/strokeaha.111.635755

Matarin, M., Brown, W. M., Singleton, A., Hardy, J. A., and Meschia, J. F. (2008). Whole Genome Analyses Suggest Ischemic Stroke and Heart Disease Share an Association with Polymorphisms on Chromosome 9p21. *Stroke* 39, 1586–1589. doi:10.1161/strokeaha.107.502963

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., et al. (2007). A Common Allele on Chromosome 9 Associated with Coronary Heart Disease. *Science* 316, 1488–1491. doi:10.1126/science.1142447

Reed, D. M. (1990). The Paradox of High Risk of Stroke in Population with Low Risk of Coronary Heart Disease. *Am. J. Epidemiol.* 131, 579–588. doi:10.1093/oxfordjournals.aje.a115542

Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., et al. (2007). Genomewide Association Analysis of Coronary Artery Disease. *N. Engl. J. Med.* 357, 443–453. doi:10.1056/NEJMoa072366

Söderholm, M., Pedersen, A., Lorentzen, E., Stanne, T. M., Bevan, S., Olsson, M., et al. (2019). Genome-Wide Association Meta-Analysis of Functional Outcome after Ischemic Stroke. *Neurology* 92, e1271–e1283. doi:10.1212/wnl.0000000000007138

Traylor, M., Farrall, M., Holliday, E. G., Sudlow, C., Hopewell, J. C., Cheng, Y.-C., et al. (2012). Genetic Risk Factors for Ischaemic Stroke and its Subtypes (The METASTROKE Collaboration): A Meta-Analysis of Genome-Wide Association Studies. *Lancet Neurol.* 11, 951–962. doi:10.1016/s1474-4422(12)70234-x

Wang, Q., Zhao, J., Chang, H., Liu, X., and Zhu, R. (2021). Association between lncRNA ANRIL Genetic Variants with the Susceptibility to Ischemic Stroke: From a Case-Control Study to Meta-Analysis. *Med. Baltim.* 100, e25113. doi:10.1097/md.0000000000025113

Xiong, L., Liu, W., Gao, L., Mu, Q., Liu, X., Feng, Y., et al. (2018). The ANRIL Genetic Variants and Their Interactions with Environmental Risk Factors on Atherothrombotic Stroke in a Han Chinese Population. *J. Stroke Cerebrovasc. Dis.* 27, 2336–2347. doi:10.1016/j.jstrokecerebrovasdis.2018.04.020

Yang, J., Gu, L., Guo, X., Huang, J., Chen, Z., Huang, G., et al. (2018). LncRNA ANRIL Expression and ANRIL Gene Polymorphisms Contribute to the Risk of Ischemic Stroke in the Chinese Han Population. *Cell Mol. Neurobiol.* 38, 1253–1269. doi:10.1007/s10571-018-0593-6

Zhang, W., Chen, Y., Liu, P., Chen, J., Song, L., Tang, Y., et al. (2012). Variants on Chromosome 9p21.3 Correlated with ANRIL Expression Contribute to Stroke Risk and Recurrence in a Large Prospective Stroke Population. *Stroke* 43, 14–21. doi:10.1161/strokeaha.111.625442

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 06 July 2022
doi: 10.3389/fgene.2022.860122

Identification of TLR2 as a Key Target in Neuroinflammation in Vascular Dementia

*Yuye Wang [1,2], Shuang Lv [1,3], Xiao Zhou [1,2], Xiaoqian Niu [1,3], Leian Chen [1,2], Ziyuan Yang [1,3] and Dantao Peng [1,2,3]**

[1]Department of Neurology, China-Japan Friendship Hospital, Beijing, China, [2]Graduate School of Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China, [3]Peking University China-Japan Friendship School of Clinical Medicine, Beijing, China

Vascular dementia (VaD) is the second most common cause of dementia. At present, precise molecular processes of VaD are unclear. We attempted to discover the VaD relevant candidate genes, enrichment biological processes and pathways, key targets, and the underlying mechanism by microarray bioinformatic analysis. We selected GSE122063 related to the autopsy samples of VaD for analysis. We first took use of Weighted Gene Co-expression Network Analysis (WGCNA) to achieve modules related to VaD and hub genes. Second, we filtered out significant differentially expressed genes (DEGs). Third, significant DEGs then went through Geno Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. Fourth, Gene Set Enrichment Analysis (GSEA) was performed. At last, we constructed the protein–protein interaction (PPI) network. The results showed that the yellow module had the strongest correlation with VaD, and we finally identified 21 hub genes. Toll-like receptor 2 (TLR2) was the top hub gene and was strongly correlated with other possible candidate genes. In total, 456 significant DEGs were filtered out and these genes were found to be enriched in the Toll receptor signaling pathway and several other immune-related pathways. In addition, Gene Set Enrichment Analysis results showed that similar pathways were significantly over-represented in TLR2-high samples. In the PPI network, TLR2 was still an important node with high weight and combined scores. We concluded that the TLR2 acts as a key target in neuroinflammation which may participate in the pathophysiological process of VaD.

Keywords: vascular dementia, TLR2, neuroinflammation, bioinformatic analysis, WGCNA

OPEN ACCESS

Edited by:
Rossen Donev,
MicroPharm Ltd., United Kingdom

Reviewed by:
Shirong Liu,
Department of Infectious Disease, Genentech Inc., United States
Wen Youliang,
Gannan Medical University, China

****Correspondence:***
Dantao Peng
pengdantao2000@163.com

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: *22 January 2022*
Accepted: *10 June 2022*
Published: *06 July 2022*

Citation:
Wang Y, Lv S, Zhou X, Niu X, Chen L, Yang Z and Peng D (2022) Identification of TLR2 as a Key Target in Neuroinflammation in Vascular Dementia. Front. Genet. 13:860122. doi: 10.3389/fgene.2022.860122

INTRODUCTION

Vascular dementia (VaD), following Alzheimer's disease (AD), is one of the most prevalent causes of dementia (O'Brien and Thomas, 2015). A study in 6,481 Korean older adults showed that in 2016 disability-adjusted life-years (DALYs) caused by VaD (316 per 100,000) comprised 20% of the total DALYs caused by mild cognitive impairment (MCI) and dementia. In 2065, DALYs due to VaD (3654 per 100,000) would comprise 38% of the total DALYs as mentioned before. In parallel, the years of life lived with disability (YLDs) attributed to VaD (85 per 100,000) accounted for 18% of the total YLDs caused by MCI and dementia in 2016, while in 2065 YLDs attributed to VaD (410 per 100,000) will account for 15% of total YLDs (Moon et al., 2021). As the data shows, DALYs and YLDs of VaD are estimated to increase. However, there are

fewer relative studies about VaD than those about AD, and there are no licensed treatments for VaD.

As a multifactorial disease, various risk factors participate in the development of VaD. Age and stroke are both major risk factors for the pathogenesis of VaD. VaD is also associated with vascular risk factors (O'Brien and Thomas, 2015; Iadecola et al., 2019). In addition, genetic linkage analyses investigated penetrant monogenic causes of VaD (Romay et al., 2019). Thus, a comprehensive understanding of key risk factors and genetic predispositions that lead to VaD needs to be clarified.

In nervous system, Toll-like receptors (TLRs) were reported to regulate the numbers of neurons and the size of brain, modulating structural plasticity in the adult brain (Li G et al., 2020). TLRs were an ancient family of pattern recognition receptors (PRRs). The role of TLRs in immunity control has been broadly discussed (Fitzgerald and Kagan, 2020). In neurological diseases, TLRs were reported to participate in AD (7), Parkinson's disease (PD) (Kouli et al., 2019), ischemic stroke (IS) (Wang et al., 2013; Tajalli-Nezhad et al., 2019), and multiple sclerosis (MS) (Racke and Drew, 2009). However, the role of TLRs in VaD remained unclear.

In the present study, we performed a bioinformatic analysis based on GSE122063 (McKay et al., 2019). We first tried to figure out hub genes and top hub gene. Then we conducted a basic analysis on DEGs. Last, we performed relative analyses centered on the top hub gene to further investigate the probable mechanism of that gene in VaD.

MATERIALS AND METHODS

Microarray Data Processing

In the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database, we chose GSE122063 which included the autopsy samples of VaD for analysis. GSE122063 was based on GPL16699 which used Agilent-039494 SurePrint G3 Human GE v2 8 × 60 K Microarray to detect the expression of genes. The microarray data includes eight VaD patients, 12 AD patients, and 11 controls postmortem frontal and temporal cortex samples. Each sample was run with at least two technical replicates. Data from AD patients were excluded from analysis and VaD sample 1063 was removed due to poor data quality according to the clustering result. The raw expression matrix was directly downloaded from the website, and the SOFT format file was downloaded and parsed by the GEOquery package (Davis and Meltzer, 2007). Then we used GPL1699 to transit ID into gene names and gene symbols using merge function in R. In addition, we checked if the data need log transformation or normalization. After pre-processing, a normalized expression matrix was constructed. The group matrix was constructed based on clinical information. All bioinformatic analyses and visualization were processed based on R.

Weighted Gene Co-Expression Network Analysis (WGCNA)

The WGCNA package (Langfelder and Horvath, 2008) was used to create a gene co-expression network. By median absolute deviation (MAD), the top 5,000 ranking genes were selected at first. Then a soft-thresholding power β was calculated by using the "pickSoftThreshold" function. A suitable power value was defined as the first number reaching which the degree of independence was at least 0.9. The gene expression matrix was then converted into a topological overlap matrix (TOM), and the genes were divided into several gene modules, each represented by a distinct color. Next, a hierarchical clustering analysis was performed by using the hclust function. Except for the WGCNA package, the gplots package (Warnes et al., 2020) was used for visualization. In addition, the top 100 networks sorted by weight were exported to Cytoscape software for visualization.

In WGCNA, gene significance (GS) was used to describe the relationship between gene and phenotype. Module membership (MM) was calculated to evaluate the importance of a gene in the module by using the cor function. In this study, genes with both GS > 0.3 and MM > 0.9 was defined as hub genes among the candidate gene modules (Jin et al., 2021). The correlation relationship of hub genes was explored by using the gpairs package (Emerson and Green, 2020).

Identification of DEGs

We first used lmFit and eBayes functions in the limma package (Ritchie et al., 2015) to identify the DEGs between VaD and control groups. The statistical method to calculate false discovery rate (FDR) was the Benjamini–Hochberg method. Then a threshold of adjust-p < 0.05 and the absolute value of $\log_2$ fold change ($\log_2$FC) > 1 were set, and the significant DEGs between the VaD and controls were filtered out. A volcano plot was presented by using EnhancedVolcano (Blighe et al., 2018). The distribution shape of TLR2 was shown in the violin plot by using the ggpubr package (Kassambara, 2020).

Geno Ontology and Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis

A GO enrichment analysis was run to annotate the functions of the significant DEGs with GO terms. The GO enrichment analysis could explain the features of changed genes from the following three structural networks of terms: biological processes (BP), cellular components (CC), and molecular functions (MF). The KEGG pathway analysis was performed to investigate the pathway that the significant DEGs might be involved in. The org. Hs.eg.db package (Carlson, 2021) was used for transition from gene symbols to Entrez ID. Then the clusterProfiler package (Yu et al., 2012; Wu et al., 2021) was used for the enrichment analysis. At last, the ggplot2 (Wickham, 2016) package was used for visualization. The aforementioned analysis results enabled us to discover the biological pathways of the altered genes in the VaD group.

Gene Set Enrichment Analysis (GSEA)

In the GSE122063 datasets, GSEA was used to explore distinct GO terms and KEGG pathways that may be associated with TLR2. All genes were included in the analysis. Gene sets were directly downloaded from the website (http://www.gsea-msigdb.org/gsea/downloads.jsp). Except for the VaD and control groups, we set the median expression level of TLR2 as the cutoff value to

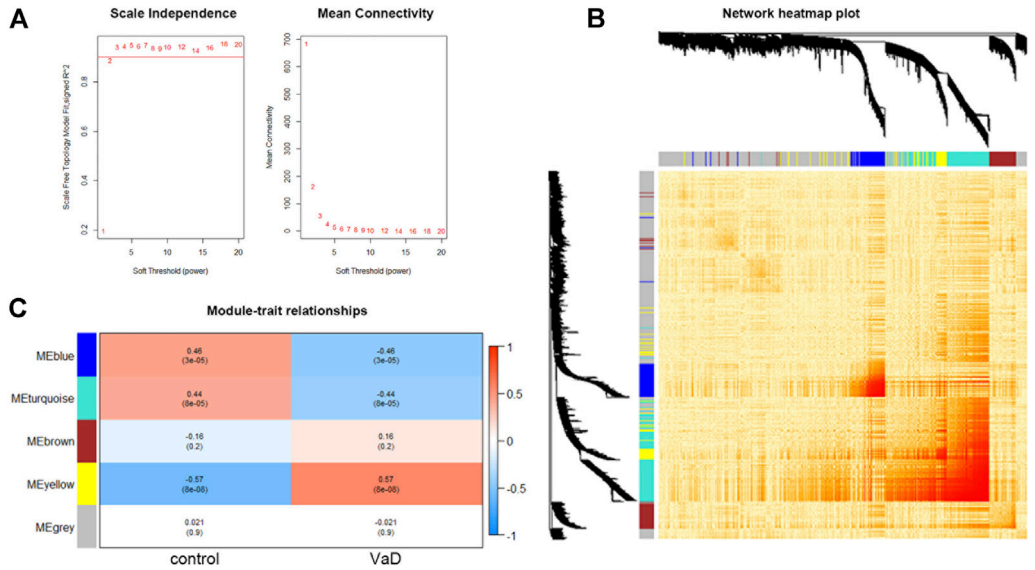

FIGURE 1 | Results and visualization of Weighted Gene Co-expression Network Analysis (WGCNA) analysis. **(A)** Determination of soft threshold β. Left: scale independence; right: mean connectivity. **(B)** Heat map showing the TOM among all 5,000 genes involved in the WGCNA with cluster dendrogram showing on the axis. Each color represents one specific co-expression module; the above branches represent genes. The genes with strong correlations are clustered into modules, which are represented as dark sections symmetrically distributed along the diagonal in the heatmap, corresponding to the cluster dendrogram. **(C)** Module–trait relationships among the five gene modules. The yellow module is the most correlated module (correlation coefficient = 0.57, $^{***}p < 0.001$).

divide patients into TLR2-high and TLR2-low expression groups. The org. Hs.eg.db package (Carlson, 2021) was used for Entrez ID transition, and the clusterProfiler package (Yu et al., 2012; Wu et al., 2021) was used for the enrichment analysis. Furthermore, the gseaplot2 function in the enrichplot package (Yu, 2021) was used for visualization of enrichment results.

Construction of a Protein–Protein Interaction Network

We used the STRING online database (https://string-db.org/) to construct a PPI network. Significant DEGs were uploaded to the STRING website. After being filtered by the "no more than 50 interactors" and "k-means clustering" options, the PPI network was exported into a TSV file. At last, the analysis and visualization of the interaction network were achieved by Cytoscape software. The function of network analysis function in the Cytoscape software calculated the degree which was utilized as the continuous mapping of nodes both in size and fill color (from blue to red). The combined score exported directly from the string database was used for the continuous mapping of edges both in width and stroke color (from blue to red). Larger size and bluer nodes indicated the higher degree, while wider and bluer lines indicated the higher combined scores.

RESULTS

WGCNA and Module Related With VaD

Using the expression matrix, WGCNA was used to determine the main module which was most linked with VaD. At first, we chose the top 5,000 genes sorted by MAD in the GSE122063 microarray assay for analysis. According to the calculation result, the soft-thresholding power β was 2 as the plot showed, with the scale-free topology R2-value achieving 0.9 (**Figure 1A**). To visualize the weighted network, a heat map was plotted. The gene co-expression network was created, and the genes were clarified into five modules represented by distinct colors including grey, turquoise, blue, brown, and yellow. This is called a cluster dendrogram, and it was presented along the axis. a network heat map of all 5,000 genes was shown by using the TOMplot function in **Figure 1B**. Each row and column in the heat map represented the same gene, and thus the network heat map is a symmetric plot. The genes with strong correlations were clustered into modules, which were represented as dark sections symmetrically distributed along the diagonal in the heat map, corresponding to the cluster dendrogram. The biggest grey module included 2,783 genes, and the smallest yellow module included 400 genes. As shown in the module–trait relationships plot, the yellow module was most positively associated with VaD (correlation coefficient = 0.57, $^{***}p < 0.001$; **Figure 1C**) and was chosen as the key module. The functional annotation of three significantly related modules (blue, turquoise, and yellow) are shown in **Supplementary Figure S1**. The yellow module was most related to immunity and inflammation.

Identification of Hub Genes and Top Hub Gene

Among the 400 genes in the yellow module, genes with MM > 0.9 and GS > 0.3 were sorted out as hub genes. The red dotted lines represent the thresholds value of MM > 0.9 and GS > 0.3 set for

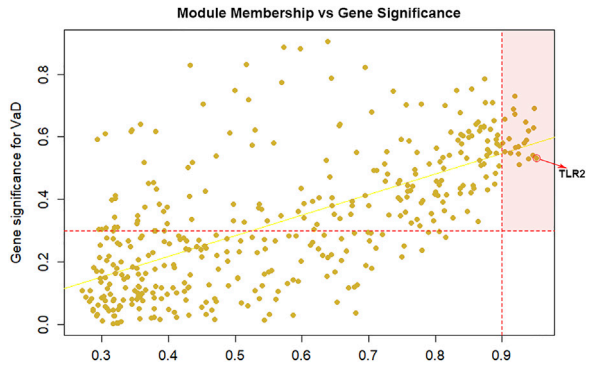

FIGURE 2 | Selection of hub genes. Module membership (MM) vs. gene significance (GS) in the yellow module (correlation coefficient = 0.65, ***p < 0.001). The red dotted lines represented the thresholds of MM > 0.9 and GS > 0.3 set for hub genes and separated an area in the upper right corner. Toll-like receptor 2 (TLR2) is selected as the top hub gene.

hub genes and separated an area in the upper right corner. The correlation analysis between yellow module memberships and gene significance showed statistical significance (correlation coefficient = 0.65, ***p < 0.001). In total, 21 hub genes were identified (TLR2, CD163, VSIG4, SLAMF8, C1QB, CD16a, CD32, ALOX5AP, integrinβ2, EBI3, HCLS1, CD14, LAIR-1, CD300a, IFI30, LCP1, C1orf162, γ-parvin, ALOX5, SLA, and CMTM7). According to MM or the chooseTopHubInEachModule function, TLR2 was the top hub gene in the yellow module (**Figure 2**). Furthermore, we found that TLR2 shows a strong positive correlation with other candidate genes, which indicated that changes in TLR2 expression might cause changes in these genes (**Supplementary Figure S2**).

Identification of DEGs in VaD

The gene expression levels of the samples were distributed at the same baseline after normalization. Compared to the control group, significant DEGs were identified in the VaD group by setting the threshold value as adjust-p < 0.05 and |log2FC| > 1. The expression of the genes was displayed as a volcano plot in which the size of the dot reflects |log2FC| of the gene (**Figure 3A**). There were 456 significant DEGs between the VaD and control groups among the 23,320 genes detected in microarray, including 198 upregulated ones and 258 downregulated ones. TLR2 was one of the significant DEGs and was marked out in the volcano plot. Specifically, the expression level of TLR2 in the VaD and control groups was shown in the violin plot (***p < 0.001, **Figure 3B**). TLR2 was significantly differentially expressed between the two groups.

Results of GO and KEGG Analysis

Significantly upregulated and downregulated DEGs were enriched in BP, CC, and MF terms and the KEGG pathway, respectively. The horizontal axis represents –log10 (p-value), while the color indicated the change direction. In detail, BP, Toll-like receptor signaling pathway was enriched, which was consistent with our previous result. Other BPs such as negative regulation of immune system process, antigen processing, and presentation and regulation of B, T, and NK cells were examples of significantly enriched upregulated GO terms (*p < 0.05, **Figure 4A**), while CCs, including azurophil granule, endocytic vesicle, and secretory granule membrane are shown (*p < 0.05, **Figure 4B**). Upregulated MFs, such as scavenger receptor activity and RAGE receptor activity, were significantly enriched. Neuropeptide hormone activity, neuropeptide receptor binding, and signaling receptor activation activity were downregulated (*p < 0.05, **Figure 4C**). Most enriched KEGG pathways did not reach statistical significance in which we

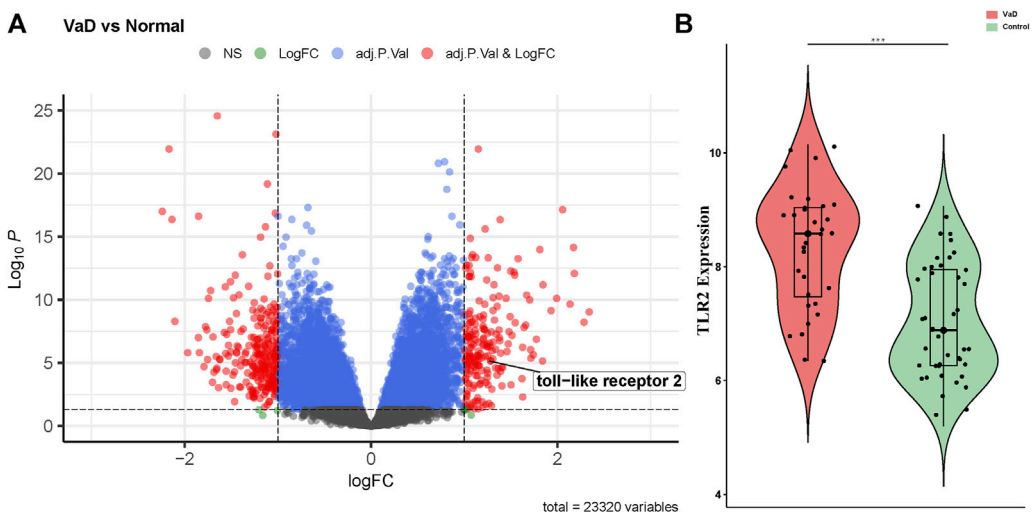

FIGURE 3 | Differentially expressed genes (DEGs) present in vascular dementia (VaD) and control groups in microarray from GSE122063 and the expression level of Toll-like receptor-2 (TLR2). **(A)** Volcano plot showed the distribution of the DEGs between two groups. The red dots correspond to the significantly regulated genes. **(B)** Violin plot of TLR2. TLR2 is upregulated in the VaD group (***p < 0.001).

FIGURE 4 | Results of the Geno Ontology (GO) terms enrichment analysis of significant DEGs. (*$p < 0.05$). Blue bars showed the results of upregulated genes while red bars showed the results of downregulated genes.

observed a trend in Toll-like receptor signaling pathway and neuroactive ligand–receptor interaction was significantly downregulated (**Supplementary Figure S3**).

GSEA Enrichment Results

GSEA was analyzed in the disease group versus control as well as groups divided by the expression level of TLR2. When comparing the VaD group with the control group, the Toll-like receptor pathway was enriched, which was the same as the results from DEGs. Other immunity and inflammation-related processes were also enriched which indicated the representativeness of the data and complemented evidence for the role of TLR2 in neuroinflammation. The results are shown in **Supplementary Figure S4**. When comparing the TLR2-high group with the low group, the results showed that BPs, such as cytokine-mediated signaling pathway and defense response to other organism, were significantly enriched in the TLR2-high samples (*$p < 0.05$, **Figure 5A**). CCs, such as synapse, vacuole, and cell surface granule, and MFs, such as immune receptor activity and molecular transducer activity, were significantly enriched in the TLR2-high samples, shown in **Figures 5B, C**, respectively (*$p < 0.05$, **Figures 5B,C**). When it comes to the KEGG enrichment analysis, pathways such as antigen processing and presentation, ribosome, and cytokine–cytokine receptor reaction were significantly over-represented in TLR2-high samples (*$p < 0.05$, **Figure 5D**). The similar enrichment results in VaD and

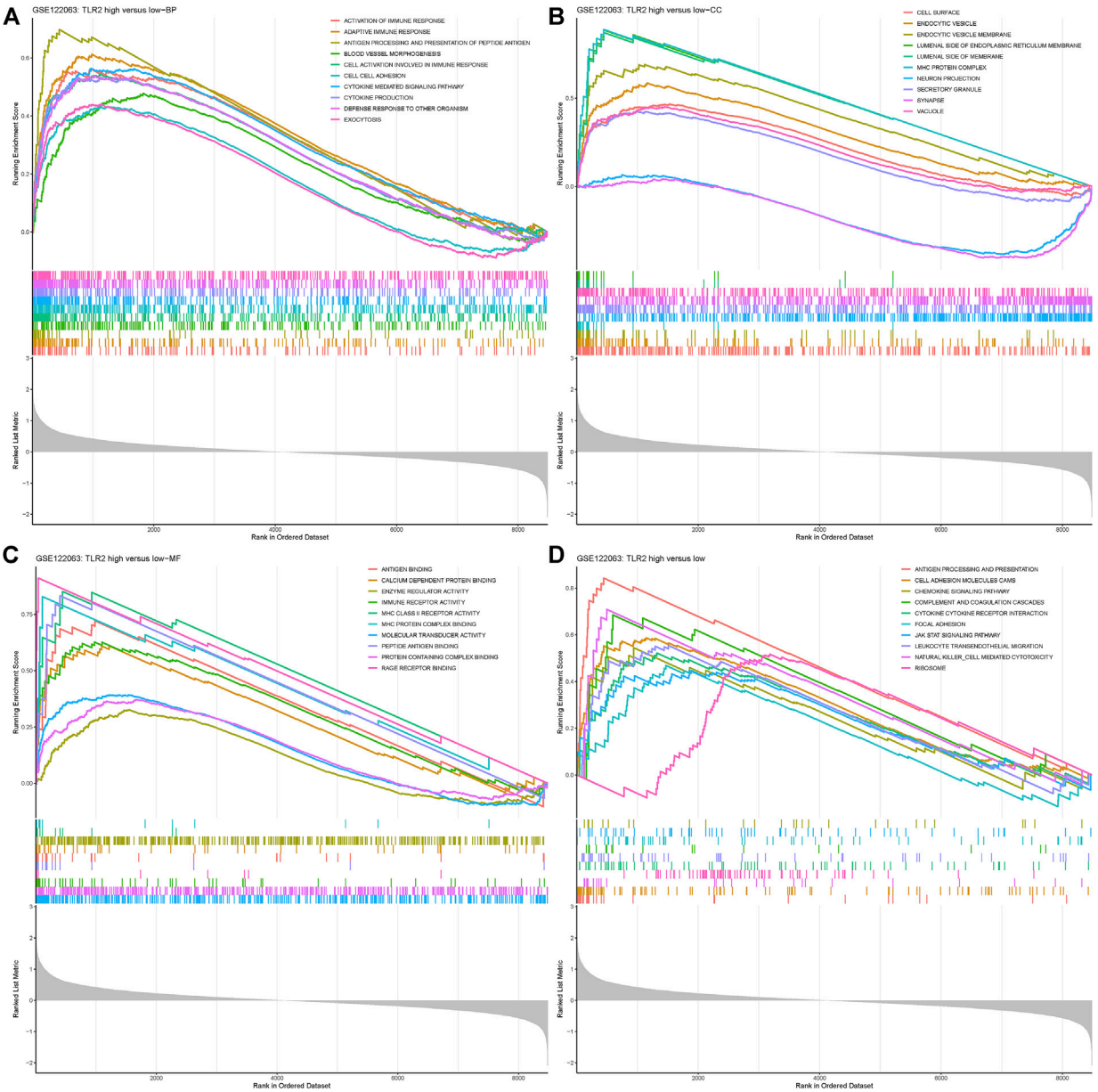

FIGURE 5 | Gene Set Enrichment Analysis (GSEA) results grouped by the expression level of TLR2. **(A)** BP enriched in TLR2-high group. **(B)** CC enriched in TLR2-high group. **(C)** MF enriched in TLR2-high group. **(D)** KEGG pathways enriched in TLR2-high group.

control groups, as well as in the TLR2-high and low groups further demonstrated the important role of TLR2 in VaD. Moreover, high expression level of TLR2 was related to many genes, including myeloid differentiation factor 88 (MyD88), nuclear factor kappa B (NF-κB), protein kinase B (AKT), glial fibrillary acidic protein (GFAP), ionized calcium-binding adapter molecule 1 (Iba1), and many cytokines according to the expression matrix and the KEGG pathway.

PPI Network Construction

With the combined use of STRING and Cytoscape, the PPI network of the significant DEGs was created. The size and color reflected the degree of nodes in which the more edges connected to this node, the greater its degree. The larger size and bluer node indicated the higher degree. The width and color reflected the combined score of edges in which the combined scores were positively related to the interaction relationships between the two proteins. The wider and bluer line indicated higher combined scores. The overall network of DEG-correlated proteins is shown in **Figure 6A**. TLR2 got a relatively high degree in this overall network which suggested that TLR2 played a crucial role in the network. Considering the complication of the network, a new network centered on TLR2 was further constructed and amplified. TLR2 was most associated with Complement C5a Receptor 1 (C5AR1), Heat Shock Protein Family A Member 1 A

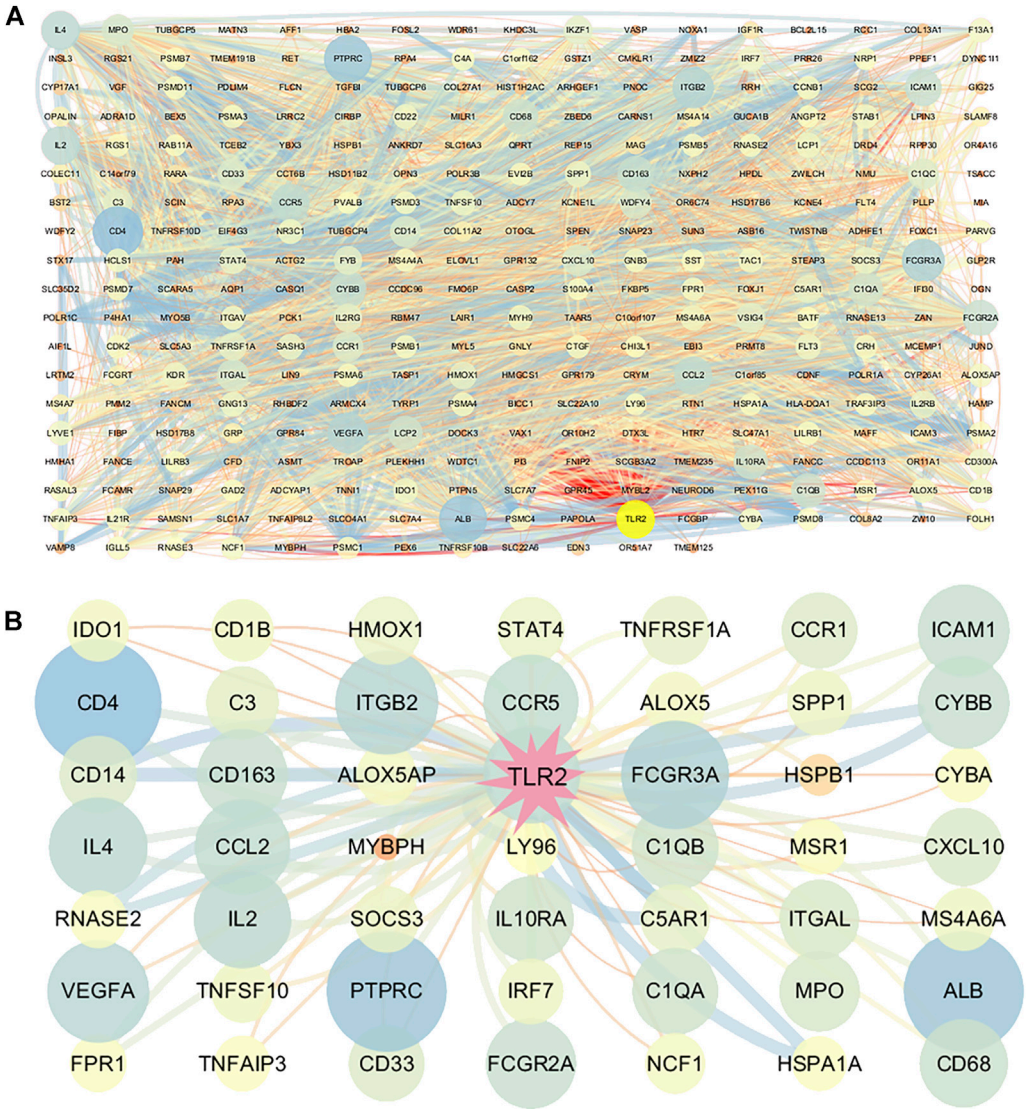

FIGURE 6 | Construction of the protein–protein interaction (PPI) network consisting of DEGs. **(A)** PPI of DEGs. **(B)** Partial network centered on TLR2. The size and color of the nodes reflect the degree and the width and color of the edges reflect the combined scores (color: from blue to red). Larger size and bluer nodes indicated the higher degree while wider and bluer lines indicated the higher combined scores.

(HSPA1A), cluster of differentiation (CD14), and cytochrome B-245 Beta Chain (CYBB) (**Figure 6B**).

DISCUSSION

Cognitive impairment related to aging has become one of the major public health burdens for us. Although Alzheimer's disease is the most prevalent cause of clinically diagnosed dementia in western nations, vascular etiology is the second most common cause. Also, vascular etiology is the most common cause in East Asia (Iadecola et al., 2019). Thus, it is worthwhile to investigate the underlying mechanism of VaD development. Much progress has been made during the past years; however, several controversies remain to be interpreted.

In the present study, we first took use of WGCNA to achieve modules related to VaD and hub genes. According to the correlation coefficient, a yellow module was chosen which was closely related to immunity and we finally identified 21 hub genes. TLR2 was the top hub gene which was strongly correlated with other possible candidate genes. Second, we filtered out 456 significant DEGs by adjust-$p < 0.05$ and $|\log2FC| > 1$. TLR2 was one of the DEGs and was significantly upregulated in the VaD group. Third, significantly upregulated and downregulated DEGs were gone through GO and KEGG analyses and the Toll-like receptor pathway, and other inflammation related processes were found to be upregulated in the VaD group. Fourth, GSEA results showed that cytokine-mediated signaling pathway, cell surface, immune receptor activity, and cytokine–cytokine receptor reaction

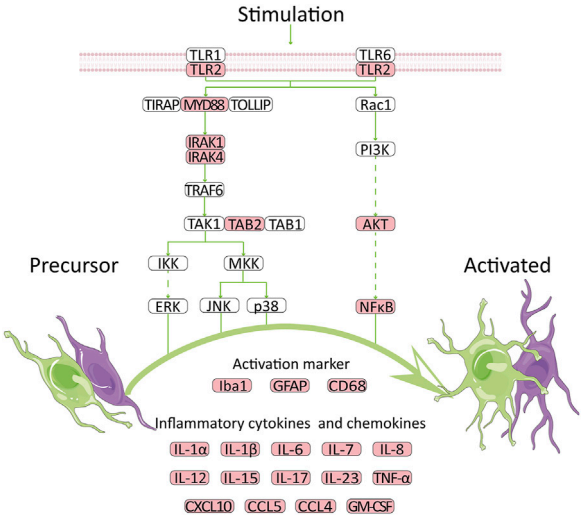

FIGURE 7 | Potential mechanism for high expression of TLR2 to promote VaD. The network is summarized according to GSE122063 database and public KEGG pathway. Red indicates the upregulated genes.

were significantly over-represented in TLR2-high samples. The results were similar to enrichment results achieved by samples being divided by disease status. Finally, in the PPI analysis, TLR2 was an important node with a higher degree and combined scores edges which indicated that TLR2 remained a key target at the protein level. In summary, with five approaches complementing each other, TLR2 might participate in the pathophysiological process of VaD *via* the neuroinflammation pathway.

TLRs were proved to be involved in the control of immunity and neurological diseases (Racke and Drew, 2009; Kouli et al., 2019; Lin et al., 2019; Tajalli-Nezhad et al., 2019; Fitzgerald and Kagan, 2020). TLR2, as a member of TLRs, also played a vital role in nervous system. Based on the KEGG Toll-like receptor signaling pathway, we summarized a mechanism chart. After comparing the pathway with our analysis results, we found that a high expression level of TLR2 was related to many genes, including MYD88, AKT, NF-κB, Iba1, GFAP, and many cytokines, suggesting that TLR2 might participate in the development of VaD *via* the neuroinflammation pathway. The genes that were upregulated in this microarray were marked in red. High expression of TLR2 induced activation of astrocytes and microglia, which further lead to the secretion of cytokines (**Figure 7**). Previous studies were consistent with our results and provided a foundation for this prediction. Knockdown of MyD88 attenuated the mRNA expression of TNF-α and inducible nitric oxide synthase (iNOS) (Jana et al., 2008) in AD, while reduced inflammatory response was observed in MYD88 knockdown mice with traumatic brain injury (TBI) (Krieg et al., 2017). These results revealed the role of MYD88 in neuroinflammation. Meanwhile, AKT and NF-κB were involved in the neuroinflammation pathway in experimental models of AD (Yang et al., 2020). In addition, GFAP is an activation marker of astrocytes, while Iba1 and CD68 are the activation markers of microglia. The anti-TLR2 antibody group had lower GFAP and CD68 immunoreactivity than the control group (McDonald et al., 2016). At last, the expression levels of inflammatory cytokines increased (Brea et al., 2011; Dzamko et al., 2017; Sun et al., 2017). At the protein level, TLR2 was proven to be strongly correlated with proteins such as C5AR1 (Mödinger et al., 2018), HSPA1A (Yang et al., 2013), and CD14 (Aguilar-Briseño et al., 2020), according to the previous study which was coincident with our results. These molecules, as well as CYBB, were all related to neuroinflammation which further proved our results (Tarassishin et al., 2014; Qu et al., 2017; Michailidou et al., 2018; Keller et al., 2021). All the results proved that TLR2 could be an efficient target to regulate the unwanted inflammatory responses in neurological conditions (Hayward and Lee, 2014). Thus, we suggested that TLR2 might participate in the development of VaD *via* the neuroinflammation pathway.

In parallel, there was other evidence that also supported the role of TLR2 in the development of VaD. First, TLR2 regulated the risk factors of vascular diseases which further affect VaD development, such as atherosclerosis (Li B et al., 2020) and diabetes. TLR2 was found to promote vascular smooth muscle cell chondrogenic differentiation and consequent calcification in atherosclerosis by activating p38 and extracellular regulated protein kinases (ERK) 1/2 signaling (Lee et al., 2019). Activation of TLR2 stimulated the pro-inflammatory cytokines and chemokines secretion, which would cause vascular injuries. Diabetes-induced changes in cerebral blood flow and cognitive deficits were prevented when TLR2 was knocked out (Hardigan et al., 2017).

Second, TLR2 participated in the pathophysiological process of stroke and other neurodegeneration diseases. In IS, TLR2 was associated with the outcome (Brea et al., 2011), and TLR2 inhibition improved neuronal survival (Ziegler et al., 2011), which indicated a future therapy. Repeated exposure to TLR2 agonists may exacerbate neurodegeneration in AD by their microglial-mediated toxicity (Lax et al., 2020) and inhibition of TLR2 in microglia (Liu et al., 2012) or mouse model could be beneficial in AD pathogenesis. Similarly, TLR2 was reported to exert a prominent role in the microglial-mediated responses which is vital for PD progression (Doorn et al., 2014).

Third, TLR2 exerted functions in biological processes or other neurological diseases *via* the neuroinflammation pathway. Neuraminidase-induced inflammatory reaction *in vivo* was partly dependent on TLR2 (Fernández-Arjona et al., 2019), while interferon-γ (IFN-γ) enhanced α-syn stimulation and inflammatory responses *via* TLR2, TLR3, and TNF-α *in vitro* (Wang et al., 2019). TLR2 and TLR4 could serve as important mediators of repeated social defeat stress (R-SDS)–induced microglial activation in the medial prefrontal cortex (mPFC), which caused neuronal and behavioral alternations *via* inflammatory-related cytokines (Nie et al., 2018). In addition, TLR2 and TLR4 were shown to potentially advance secondary brain injury after experimentally controlled cortical impact (CCI) *via* neuroinflammation (Krieg et al., 2017) while activation of microglia, *via* a TLR2-sphingosine kinase 1 (Sphk1)-pro-inflammatory cytokines (IL-1β, TNF-α, IL-17, and IL-23) pathway, may be involved in ischemia/reperfusion (I/R) injury (Sun et al., 2017). In IS, TLR2 activation was associated with a higher interleukin (IL)-1β, tumor necrosis factor-α (TNF-α) and IL-6 expression level (Brea et al., 2011). The expression of TLR2 was increased in affected regions, further inducing TNF-α expression and

increased phosphorylation of NF-κB p105 subunit in PD (32). In AD, TLR2 was proved to be a natural receptor for Aβ to trigger neuroinflammatory activation (Richard et al., 2008; Liu et al., 2012). TLR2 deficits in microglia shifted related inflammatory activation *in vivo,* while TLR2 insufficiency reduced Aβ42-triggered inflammatory activation and increased Aβ phagocytosis *in vitro*, which were both related to improved neuronal function (Jana et al., 2008; Liu et al., 2012; McDonald et al., 2016). TLR2 could enhance macrophage receptor with collagenous structure (Marco)–induced neuroinflammation by acting on the scavenger receptors cysteine-reach (SRCR) domain of Marco, which also suggested that TLR2 could serve as a novel target for reducing neuroinflammation in neurodegenerative diseases (Wang et al., 2021). Therefore, it is reasonable to speculate that TLR2 participates in the pathophysiological process of VaD through the neuroinflammation pathway and could serve as a key target.

Our research showed that using bioinformatics to investigate the molecular processes underlying VaD could provide valuable information. Bioinformatic techniques, however, were used to identify probable critical pathways and genes. Thus, molecular experiments based on clinical samples or animal models should be performed to further validate the results. It remained to be clarified whether TLR2 is involved in the pathophysiological process of VaD and inhibition of TLR2 would contribute to VaD treatment.

In conclusion, we identified TLR2 as a neuroinflammatory leading change during VaD.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/geo/, GSE122063.

AUTHOR CONTRIBUTIONS

Study design and paper revision: PT; experimental implementation, data analysis, and paper drafting: WY; provision of reagents/materials/analysis tools: LS and ZX; data analysis: NQ, CA, and YY; all authors approved the final version of the paper.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 81974220) and Central health research project (Grant No. 2020ZD10).

ACKNOWLEDGMENTS

We thank Dr. Jianming Zeng (University of Macau), and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.860122/full#supplementary-material

REFERENCES

Aguilar-Briseño, J. A., Upasani, V., Ellen, B. M. T., Moser, J., Pauzuolis, M., Ruiz-Silva, M., et al. (2020). TLR2 on Blood Monocytes Senses Dengue Virus Infection and its Expression Correlates with Disease Pathogenesis. *Nat. Commun.* 11 (1), 3177. doi:10.1038/s41467-020-16849-7

Blighe, K., Rana, S., and Lewis, M. (2018). EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced Colouring and Labeling. Available at: https://github.com/kevinblighe/EnhancedVolcano. (Accessed May 4, 2022).

Brea, D., Blanco, M., Ramos-Cabrer, P., Moldes, O., Arias, S., Pérez-Mato, M., et al. (2011). Toll-like Receptors 2 and 4 in Ischemic Stroke: Outcome and Therapeutic Values. *J. Cereb. Blood Flow. Metab.* 31 (6), 1424–1431. doi:10.1038/jcbfm.2010.231

Carlson, M. (2021). *org.Hs.eg.db: Genome Wide Annotation for Human*. R package version 3.14.0.

Davis, S., and Meltzer, P. S. (2007). GEOquery: a Bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23 (14), 1846–1847. doi:10.1093/bioinformatics/btm254

Doorn, K. J., Moors, T., Drukarch, B., van de Berg, W., Lucassen, P. J., and van Dam, A.-M. (2014). Microglial Phenotypes and Toll-like Receptor 2 in the Substantia Nigra and hippocampus of Incidental Lewy Body Disease Cases and Parkinson's Disease Patients. *Acta neuropathol. Commun.* 2, 90. doi:10.1186/preaccept-2025829283135633

Dzamko, N., Gysbers, A., Perera, G., Bahar, A., Shankar, A., Gao, J., et al. (2017). Toll-like Receptor 2 Is Increased in Neurons in Parkinson's Disease Brain and May Contribute to Alpha-Synuclein Pathology. *Acta Neuropathol.* 133 (2), 303–319. doi:10.1007/s00401-016-1648-8

Emerson, J. W., and Green, W. A. (2020). Gpairs: The Generalized Pairs Plot. R package version 1.3.3. Available at: https://CRAN.R-project.org/package=gpairs.

Fernández-Arjona, M. d. M., Grondona, J. M., Fernández-Llebrez, P., and López-Ávalos, M. D. (2019). Microglial Activation by Microbial Neuraminidase through TLR2 and TLR4 Receptors. *J. Neuroinflammation* 16 (1), 245. doi:10.1186/s12974-019-1643-9

Fitzgerald, K. A., and Kagan, J. C. (2020). Toll-like Receptors and the Control of Immunity. *Cell* 180 (6), 1044–1066. doi:10.1016/j.cell.2020.02.041

Hardigan, T., Hernandez, C., Ward, R., Hoda, M. N., and Ergul, A. (2017). TLR2 Knockout Protects against Diabetes-Mediated Changes in Cerebral Perfusion and Cognitive Deficits. *Am. J. Physiology-Regulatory, Integr. Comp. Physiology* 312 (6), R927–R937. doi:10.1152/ajpregu.00482.2016

Hayward, J. H., and Lee, S. J. (2014). A Decade of Research on TLR2 Discovering its Pivotal Role in Glial Activation and Neuroinflammation in Neurodegenerative Diseases. *Exp. Neurobiol.* 23 (2), 138–147. doi:10.5607/en.2014.23.2.138

Iadecola, C., Duering, M., Hachinski, V., Joutel, A., Pendlebury, S. T., Schneider, J. A., et al. (2019). Vascular Cognitive Impairment and Dementia: JACC Scientific Expert Panel. *J. Am. Coll. Cardiol.* 73 (25), 3326–3344. doi:10.1016/j.jacc.2019.04.034

Jana, M., Palencia, C. A., and Pahan, K. (2008). Fibrillar Amyloid-β Peptides Activate Microglia via TLR2: Implications for Alzheimer's Disease. *J. Immunol.* 181 (10), 7254–7262. doi:10.4049/jimmunol.181.10.7254

Jin, F., Li, L., Hao, Y., Tang, L., Wang, Y., and He, Z. (2021). Identification of Candidate Blood mRNA Biomarkers in Intracerebral Hemorrhage Using

Integrated Microarray and Weighted Gene Co-expression Network Analysis. *Front. Genet.* 12, 707713. doi:10.3389/fgene.2021.707713

Kassambara, A. (2020). Ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. Available at: https://CRAN.R-project.org/package=ggpubr.

Keller, C. W., Kotur, M. B., Mundt, S., Dokalis, N., Ligeon, L.-A., Shah, A. M., et al. (2021). CYBB/NOX2 in Conventional DCs Controls T Cell Encephalitogenicity during Neuroinflammation. *Autophagy* 17 (5), 1244–1258. doi:10.1080/15548627.2020.1756678

Kouli, A., Horne, C. B., and Williams-Gray, C. H. (2019). Toll-like Receptors and Their Therapeutic Potential in Parkinson's Disease and α-synucleinopathies. *Brain, Behav. Immun.* 81, 41–51. doi:10.1016/j.bbi.2019.06.042

Krieg, S. M., Voigt, F., Knuefermann, P., Kirschning, C. J., Plesnila, N., and Ringel, F. (2017). Decreased Secondary Lesion Growth and Attenuated Immune Response after Traumatic Brain Injury in Tlr2/4(−/−) Mice. *Front. Neurol.* 8, 455. doi:10.3389/fneur.2017.00455

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Lax, N., Fainstein, N., Nishri, Y., Ben-Zvi, A., and Ben-Hur, T. (2020). Systemic Microbial TLR2 Agonists Induce Neurodegeneration in Alzheimer's Disease Mice. *J. Neuroinflammation* 17 (1), 55. doi:10.1186/s12974-020-01738-z

Lee, G.-L., Yeh, C.-C., Wu, J.-Y., Lin, H.-C., Wang, Y.-F., Kuo, Y.-Y., et al. (2019). TLR2 Promotes Vascular Smooth Muscle Cell Chondrogenic Differentiation and Consequent Calcification via the Concerted Actions of Osteoprotegerin Suppression and IL-6-Mediated RANKL Induction. *Arterioscler. Thromb. Vasc. Biol.* 39 (3), 432–445. doi:10.1161/atvbaha.118.311874

Li G, G., Forero, M. G., Wentzell, J. S., Durmus, I., Wolf, R., Anthoney, N. C., et al. (2020). A Toll-Receptor Map Underlies Structural Brain Plasticity. *Elife* 9. e52743, doi:10.7554/eLife.52743

Li B, B., Xia, Y., and Hu, B. (2020). Infection and Atherosclerosis: TLR-dependent Pathways. *Cell. Mol. Life Sci.* 77 (14), 2751–2769. doi:10.1007/s00018-020-03453-7

Lin, C., Zhao, S., Zhu, Y., Fan, Z., Wang, J., Zhang, B., et al. (2019). Microbiota-gut-brain axis and Toll-like Receptors in Alzheimer's Disease. *Comput. Struct. Biotechnol. J.* 17, 1309–1317. doi:10.1016/j.csbj.2019.09.008

Liu, S., Liu, Y., Hao, W., Wolf, L., Kiliaan, A. J., Penke, B., et al. (2012). TLR2 Is a Primary Receptor for Alzheimer's Amyloid β Peptide to Trigger Neuroinflammatory Activation. *J. Immunol.* 188 (3), 1098–1107. doi:10.4049/jimmunol.1101121

McDonald, C. L., Hennessy, E., Rubio-Araiz, A., Keogh, B., McCormack, W., McGuirk, P., et al. (2016). Inhibiting TLR2 Activation Attenuates Amyloid Accumulation and Glial Activation in a Mouse Model of Alzheimer's Disease. *Brain, Behav. Immun.* 58, 191–200. doi:10.1016/j.bbi.2016.07.143

McKay, E. C., Beck, J. S., Khoo, S. K., Dykema, K. J., Cottingham, S. L., Winn, M. E., et al. (2019). Peri-Infarct Upregulation of the Oxytocin Receptor in Vascular Dementia. *J. Neuropathol. Exp. Neurol.* 78 (5), 436–452. doi:10.1093/jnen/nlz023

Michailidou, I., Jongejan, A., Vreijling, J. P., Georgakopoulou, T., de Wissel, M. B., Wolterman, R. A., et al. (2018). Systemic Inhibition of the Membrane Attack Complex Impedes Neuroinflammation in Chronic Relapsing Experimental Autoimmune Encephalomyelitis. *acta neuropathol. Commun.* 6 (1), 36. doi:10.1186/s40478-018-0536-y

Mödinger, Y., Rapp, A., Pazmandi, J., Vikman, A., Holzmann, K., Haffner-Luntzer, M., et al. (2018). C5aR1 Interacts with TLR2 in Osteoblasts and Stimulates the Osteoclast-Inducing Chemokine CXCL10. *J. Cell Mol. Med.* 22 (12), 6002–6014. doi:10.1111/jcmm.13873

Moon, W., Han, J. W., Bae, J. B., Suh, S. W., Kim, T. H., Kwak, K. P., et al. (2021). Disease Burdens of Alzheimer's Disease, Vascular Dementia, and Mild Cognitive Impairment. *J. Am. Med. Dir. Assoc.* 22 (10), 2093–2099. doi:10.1016/j.jamda.2021.05.040

Nie, X., Kitaoka, S., Tanaka, K., Segi-Nishida, E., Imoto, Y., Ogawa, A., et al. (2018). The Innate Immune Receptors TLR2/4 Mediate Repeated Social Defeat Stress-Induced Social Avoidance through Prefrontal Microglial Activation. *Neuron* 99 (3), 464–479. doi:10.1016/j.neuron.2018.06.035

O'Brien, J. T., and Thomas, A. (2015). Vascular Dementia. *Lancet* 386 (10004), 1698–1706. doi:10.1016/s0140-6736(15)00463-8

Qu, J., Tao, X.-Y., Teng, P., Zhang, Y., Guo, C.-L., Hu, L., et al. (2017). Blocking ATP-Sensitive Potassium Channel Alleviates Morphine Tolerance by Inhibiting HSP70-TLR4-NLRP3-Mediated Neuroinflammation. *J. Neuroinflammation* 14 (1), 228. doi:10.1186/s12974-017-0997-0

Racke, M. K., and Drew, P. D. (2009). Toll-like Receptors in Multiple Sclerosis. *Curr. Top. Microbiol. Immunol.* 336, 155–168. doi:10.1007/978-3-642-00549-7_9

Richard, K. L., Filali, M., Prefontaine, P., and Rivest, S. (2008). Toll-Like Receptor 2 Acts as a Natural Innate Immune Receptor to Clear Amyloid 1-42 and Delay the Cognitive Decline in a Mouse Model of Alzheimer's Disease. *J. Neurosci.* 28 (22), 5784–5793. doi:10.1523/jneurosci.1146-08.2008

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Romay, M. C., Toro, C., and Iruela-Arispe, M. L. (2019). Emerging Molecular Mechanisms of Vascular Dementia. *Curr. Opin. Hematol.* 26 (3), 199–206. doi:10.1097/moh.0000000000000502

Sun, W., Ding, Z., Xu, S., Su, Z., and Li, H. (2017). Crosstalk between TLR2 and Sphk1 in Microglia in the Cerebral Ischemia/reperfusion-Induced Inflammatory Response. *Int. J. Mol. Med.* 40 (6), 1750–1758. doi:10.3892/ijmm.2017.3165

Tajalli-Nezhad, S., Karimian, M., Beyer, C., Atlasi, M. A., and Azami Tameh, A. (2019). The Regulatory Role of Toll-like Receptors after Ischemic Stroke: Neurosteroids as TLR Modulators with the Focus on TLR2/4. *Cell. Mol. Life Sci.* 76 (3), 523–537. doi:10.1007/s00018-018-2953-2

Tarassishin, L., Suh, H.-S., and Lee, S. C. (2014). LPS and IL-1 Differentially Activate Mouse and Human Astrocytes: Role of CD14. *Glia* 62 (6), 999–1013. doi:10.1002/glia.22657

Wang, Y., Ge, P., and Zhu, Y. (2013). TLR2 and TLR4 in the Brain Injury Caused by Cerebral Ischemia and Reperfusion. *Mediat. Inflamm.* 2013, 124614. doi:10.1155/2013/124614

Wang, J., Chen, Z., Walston, J. D., Gao, P., Gao, M., and Leng, S. X. (2019). Interferon-γ Potentiates α-Synuclein-induced Neurotoxicity Linked to Toll-like Receptors 2 and 3 and Tumor Necrosis Factor-α in Murine Astrocytes. *Mol. Neurobiol.* 56 (11), 7664–7679. doi:10.1007/s12035-019-1567-5

Wang, L., Yang, H.-Y., Zang, C.-X., Shang, J.-M., Liu, H., Zhang, Z.-H., et al. (2021). TLR2 Potentiates SR-Marco-Mediated Neuroinflammation by Interacting with the SRCR Domain. *Mol. Neurobiol.* 58 (11), 5743–5755. doi:10.1007/s12035-021-02463-1

Warnes, G. R., Ben, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., et al. (2020). Gplots: Various R Programming Tools for Plotting Data. R package version 3.1.1. Available at: https://CRAN.R-project.org/package=gplots.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *Innovation* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Yang, Y., Takeda, A., Yoshimura, T., Oshima, Y., Sonoda, K.-H., and Ishibashi, T. (2013). IL-10 Is Significantly Involved in HSP70-Regulation of Experimental Subretinal Fibrosis. *PloS one* 8 (12), e80288. doi:10.1371/journal.pone.0080288

Yang, W., Liu, Y., Xu, Q. Q., Xian, Y. F., and Lin, Z. X. (2020). Sulforaphene Ameliorates Neuroinflammation and Hyperphosphorylated Tau Protein via Regulating the PI3K/Akt/GSK-3β Pathway in Experimental Models of Alzheimer's Disease. *Oxid. Med. Cell Longev.* 2020, 4754195. doi:10.1155/2020/4754195

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Yu, G. (2021). Enrichplot: Visualization of Functional Enrichment Result. R package version 1.13.2. Available at: https://yulab-smu.top/biomedical-knowledge-mining-book/.

Ziegler, G., Freyer, D., Harhausen, D., Khojasteh, U., Nietfeld, W., and Trendelenburg, G. (2011). Blocking TLR2 *In Vivo* Protects against Accumulation of Inflammatory Cells and Neuronal Injury in Experimental Stroke. *J. Cereb. Blood Flow. Metab.* 31 (2), 757–766. doi:10.1038/jcbfm.2010.161

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fnins.2022.880105

Assessing the Association Between Lead Pollution and Risk of Alzheimer's Disease by Integrating Multigenomics

*Chunying Li[1,2], Yuwei Zhang[1], Jiandong Liang[3], Changyan Wu[3] and Xiao Zou[1]**

[1] Institute of Fungus Resources, College of Life Sciences, Guizhou University, Guiyang, China, [2] Department of Environmental Engineering, Chongqing Vocational College of Resources and Environmental Protection, Chongqing, China, [3] Basic Medical School, Guizhou University of Traditional Chinese Medicine, Guiyang, China

Alzheimer's disease (AD) is a life-threatening neurodegenerative disease of the elderly. In recent observations, exposure to heavy metals environmental may increase the risk of AD. However, there are few studies on the causal relationship between heavy metal exposure and AD. In this study, we integrated two large-scale summaries of AD genome-wide association study (GWAS) datasets and a blood lead level GWAS dataset and performed the two-sample Mendelian randomization analysis to assess the causality of blood lead level and AD risk. The results showed that there is a significantly positive causality between blood lead level and AD risk both in the inverse-variance weighted (IVW) model and the weighted median estimator (WME) model. An independent additional verification also reached a consistent conclusion. These findings further confirm the conclusions of previous studies and improve the understanding of the relationship between AD pathogenesis and the toxicity of lead in environmental pollution.

Keywords: blood lead, heavy metal pollution, Alzheimer's disease, multigenomics, Mendelian randomization

OPEN ACCESS

Edited by:
Guiyou Liu,
Tianjin Institute of Industrial Biotechnology (CAS), China

Reviewed by:
Ziliang Zhu,
University of North Carolina at Chapel Hill, United States
Qingxia Yang,
Nanjing University of Posts and Telecommunications, China
Yang Zhang,
Hebei Medical University, China

****Correspondence:***
Xiao Zou
xzou@gzu.edu.cn

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

Received: *21 February 2022*
Accepted: *30 May 2022*
Published: *22 July 2022*

Citation:
Li C, Zhang Y, Liang J, Wu C and Zou X (2022) Assessing the Association Between Lead Pollution and Risk of Alzheimer's Disease by Integrating Multigenomics. Front. Neurosci. 16:880105. doi: 10.3389/fnins.2022.880105

INTRODUCTION

Heavy metals are non-biodegradable, and well-documented evidence supports that chronic exposure to heavy metals can cause neurodegenerative diseases (Bush, 2003; Mates et al., 2010). These pollutants arise from rapid urbanization and industrialization, such as municipal waste, traffic, aquaculture, agricultural chemicals, paint coatings, petrochemical industry, electronic industry, mining, and smelting (Tchounwou et al., 2012; Wang et al., 2013; Ojuederie and Babalola, 2017; Fan et al., 2020). Human exposure to heavy metals mainly *via* ingestion of metal-contaminated food, water, and employment in metal-contaminated workplaces (Tchounwou et al., 2012). Several epidemiological studies have shown a significant association between cumulative metals exposure and neurodegenerative diseases (Bjorklund et al., 2018; Bakulski et al., 2020). There is robust evidence that heavy metals can disturb neurotransmitter systems by multiple mechanisms, including the interaction with neurotransmitter receptors, the modification of certain gene and/or protein expression, and the collateral damage of their functions following Reactive Oxygen Species (ROS) production (Bertram and Tanzi, 2005; Carmona et al., 2021). A previous study found that some amyotrophic lateral sclerosis patients have a 2.3- to 3.6-fold increase both in the patellar and tibial lead, which is a dose-dependent increased risk of this disease (Kamel et al., 2002).

There are many kinds of neurodegenerative diseases, including Parkinson's disease, amyotrophic lateral sclerosis, Lewy body dementia, Alzheimer's disease (AD), and so on. Among them, AD is the most typical neurodegenerative disease (Bakulski et al., 2020). AD is a neurodegenerative disease that threatens the life of the elderly, and currently, there is no efficient treatment for AD (Bakulski et al., 2020). AD is caused by a variety of environmental, lifestyle, and genetic factors that influence the degeneration of neuronal cells over some time (Bakulski et al., 2020; Huang et al., 2022). The neuropathological features of AD are hyperphosphorylated tau (a microtubule-binding protein), neurofibrillary tangles (NFTs), and aging plaques consisting of accumulated amyloid protein (Aβ) and contained metal ions (Han et al., 2019; Bakulski et al., 2020).

Accumulating evidence suggests that heavy metal pollution may be an important contributor to AD, but there is no comprehensive understanding of the effects of heavy metal pollution on AD. This study attempts to analyze the correlation between heavy metal pollution and AD by the Mendelian randomization. Mendelian randomization analysis is an analytical method for evaluating the observed correlation between a changeable risk or exposure factor and a clinically relevant outcome (Sekula et al., 2016). The use of as many instrumental variables as possible can reduce the concern of weak instrumental bias (Burgess and Thompson, 2011). This research uses genetic variants to assess the causal relationship between heavy metal exposure and AD.

In this study, we first selected the genome-wide association study (GWAS) summary data of AD and environmental pollutants from multiple authoritative databases. Then, we filtered the GWAS summary data and selected independent and matched exposure risk factor-related SNPs as the instrumental variables. Next, based on the instrumental variables with their GWAS summary results, we used two models to assess the causality of environmental pollutants and AD risk by the two-samples Mendelian randomization analysis. Finally, we used three check methods to ensure the reliability of the results of the Mendelian randomization analysis.

MATERIALS AND METHODS

Data Sources

The common water quality pollutants were considered as the exposure risk factors in this study. The related genetic variations of these exposure risk factors were selected by searching the NHGRI-EBI GWAS Catalog[1] using the keywords: "Cadmium," "Chromium," "Mercury," "Manganese," "Lead," "Molybdenum," and "Nickel." The NHGRI-EBI GWAS Catalog is a curated collection for delivering the high-quality published (GWAS) summary results of various human traits (Buniello et al., 2019). Finally, we only identified 14 blood lead level-related SNPs from a 5,433-sample size European ancestry GWAS study. This study used the blood samples from the Queensland Institute of Medical Research in Australia and the Avon Longitudinal Study of Parents and Children in the United Kingdom to measure blood lead levels and genotype of the SNPs (Warrington et al., 2015). The details were shown in **Supplementary Table 1**. The summary of GWAS data on AD is derived from a consortium consisting of the Alzheimer's Disease Genetics Consortium (ADGC), European Alzheimer's Disease Initiative (EADI), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES). A total of 10,528,610 variants are genotyped and measured using 21,982 AD individuals and 41,944 controls (Kunkle et al., 2019). In addition, to ensure the reliability of the results, we further used an independent GWAS dataset EFO_0000249, which includes 5,918 AD individuals and 212,874 controls, to conduct a verification using the Mendelian randomization analysis.[2]

Selection and Filtration of Instrumental Variables

According to the threshold of significant association $P < 10^{-5}$, we first selected the 14 blood lead level-related SNPs as the instrumental variables and further discarded the non-biallelic SNPs. Then, we matched the remaining SNPs to the AD GWAS results and attempted to align strands of the palindromic SNPs for allele harmonization. Next, to ensure mutual independence between the instrumental variables, we performed a linkage disequilibrium (LD) analysis and filtered the non-independent SNPs according to the significance threshold, i.e., $r^2 < 0.001$ within the 10,000 kb window. The samples used to estimate the LD effect were derived from the 1,000 Genome Project European ancestry individuals (Consortium, 2012). Finally, if blood lead level-related SNP is not present in the AD GWAS results, we tried to use the proxy SNPs through LD tagging ($r^2 = 1$) instead of it and integrated the filtered SNPs with the GWAS results of blood lead level and AD as the instrumental variables.

Mendelian Randomization Analysis

We used the R package "TwoSampleMR" and its web server "MRBASE" to perform the two-sample Mendelian randomization analysis (Hemani et al., 2018). Particularly, we conducted the inverse-variance weighted (IVW) model and the weighted median estimator (WME) model to assess the causal effect of blood lead level on AD risk. The IVW model ignores the intercept in the regression analysis and uses the inverse of the variance as a weight for the fit. The WME model is a consistency estimator under the assumption that more than half of the instrumental variables are valid. For the IVW model, each inverse-variance was estimated by dividing SNP-AD associations by SNP-blood lead level associations (i.e., Wald ratios). Then, the mean effect of blood lead level on AD risk was estimated by a random effect meta-analysis of the Wald ratios. When the inverse-variance satisfies the primary assumptions of Mendelian randomization analysis [i.e., the inverse-variance: (1) is associated with the exposure, (2) is not associated with the confounders, and (3) does not influence the

[1]https://www.ebi.ac.uk/gwas

[2]https://r5.risteys.finngen.fi/phenocode/G6_AD_WIDE

TABLE 1 | The causality of blood lead level and Alzheimer's disease (AD) risk by two-sample Mendelian randomization (MR) analysis using the data of AD consortium (Kunkle et al., 2019).

SNP	Position	Effect allele	GWAS of blood lead level		GWAS of AD		Model	MR analysis		Heterogeneity test		Horizontal pleiotropy	
			Beta	*P*-value	Beta	*P*-value		Beta	*P*-value	Cochran's Q	*P*-value	Intercept	*P*-value
rs76153987	chr3:9214817	T	−0.195	4×10^{-6}	−0.073	0.0266	IVW	0.245	0.0103	2.161	0.34	−0.029	0.44
rs116864947	chr7:11705786	T	−0.431	3×10^{-7}	−0.123	0.0376	WME	0.262	0.0367	0.684	0.41		
rs6462018	chr7:27519118	G	−0.084	4×10^{-6}	−0.002	0.8994							

TABLE 2 | The causality of blood lead level and AD risk by two-sample MR analysis using the data of EFO_0000249.

SNP	Position	Effect allele	GWAS of blood lead level		GWAS of AD		Model	MR analysis		Heterogeneity test		Horizontal pleiotropy	
			Beta	*P*-value	Beta	*P*-value		Beta	*P*-value	Cochran's Q	*P*-value	Intercept	*P*-value
rs76153987	chr3:9214817	T	−0.195	4×10^{-6}	−0.036	0.4025	IVW	0.242	0.0046	3.297	0.51		
rs116864947	chr7:11705786	T	−0.431	3×10^{-7}	0.128	0.0333							
rs6462018	chr7:27519118	G	−0.084	4×10^{-6}	−0.005	0.8136						−0.026	0.385
rs798338	chr7:78287721	A	−0.111	4×10^{-6}	−0.015	0.4901	WME	0.220	0.0059	2.269	0.52		
rs10121150	chr9:113369415	C	−0.143	3×10^{-8}	0.069	0.0183							

outcome through some pathways other than the exposure], IVW model can provide accurate estimates (Burgess et al., 2013; Staley and Burgess, 2017). For the WME model, the intercept of the fitted curve was calculated to estimate the average pleiotropy effect across the genetic variants. The WME can also provide a consistent estimate when more than half of the inverse variance satisfies the primary assumptions of Mendelian randomization analysis (Verbanck et al., 2018). The threshold of significant causal effect was set as $P < 0.05$. Moreover, the causal effect was considered positive and negative when the beta value was greater and less than zero, respectively.

Reliability Check

To ensure the reliability of the results of Mendelian randomization analysis, we performed the horizontal pleiotropy test, heterogeneity test, and sensitivity analysis. Particularly, we used the Egger regression intercept to estimate the magnitude of horizontal pleiotropy. If the SNPs influence the AD risk through a pathway other than the blood lead level, the significant horizontal pleiotropic ($P < 0.05$) can bias the Mendelian randomization estimates (Burgess and Thompson, 2017). Then, we assessed the heterogeneity by a funnel plot. The asymmetry and large spread of the funnel plot indicate a high heterogeneity. The significant threshold was set as $P < 0.05$ (Van Kippersluis and Rietveld, 2018). Finally, we conducted the sensitivity analysis by removing each SNP from the original Mendelian randomization analysis. The leave-one-out sensitivity analysis was used to ascertain if an association is being disproportionately influenced by a single SNP, and the forest plot was used to show the results.

RESULTS AND DISCUSSION

The Selected Instrumental Variables for Mendelian Randomization Analysis

We collected the summary GWAS data of blood lead levels from the NHGRI-EBI GWAS Catalog, and AD from the EFO_0000249 and a consortium consisting of the ADGC, EADI, CHARGE, and GERAD/PERADES, respectively. All of the samples are derived from European ancestry individuals. The blood lead level GWAS dataset was intersected with two AD GWAS datasets, respectively. After the allele harmonization, LD filtering, and SNP proxy, we selected a total of three SNPs as the instrumental variables for Mendelian randomization analysis which are significantly associated with the blood lead level and independent of each other for the consortium's AD GWAS dataset. Particularly, SNP rs76153987 (chr3:9173133), rs116864947 (chr7:11666159), and rs6462018 (chr7:27479499) are located in genes SRGAP3, THSD7A, and EVX1, respectively, and all of them are negatively associated with the blood lead level (beta = −0.195, −0.431, and −0.084; $P = 4 \times 10^{-6}$, 3×10^{-7} and 4×10^{-6}, respectively) (Warrington et al., 2015). The AD GWAS results of them are beta = −0.073, −0.123, and −0.002 and $P = 0.033$, and 0.059 and 0.015, respectively (**Table 1**). For the EFO_0000249 dataset, we identified two additional SNPs, rs798338 (chr7:78287721 in MAGI2), and rs10121150 (chr9:113369415 in BSPRY), after the screening process. The AD GWAS results of them are beta = −0.015 and 0.069 and $P = 0.490$ and 0.018, respectively (**Table 2**). The human reference genome hg38 was used in

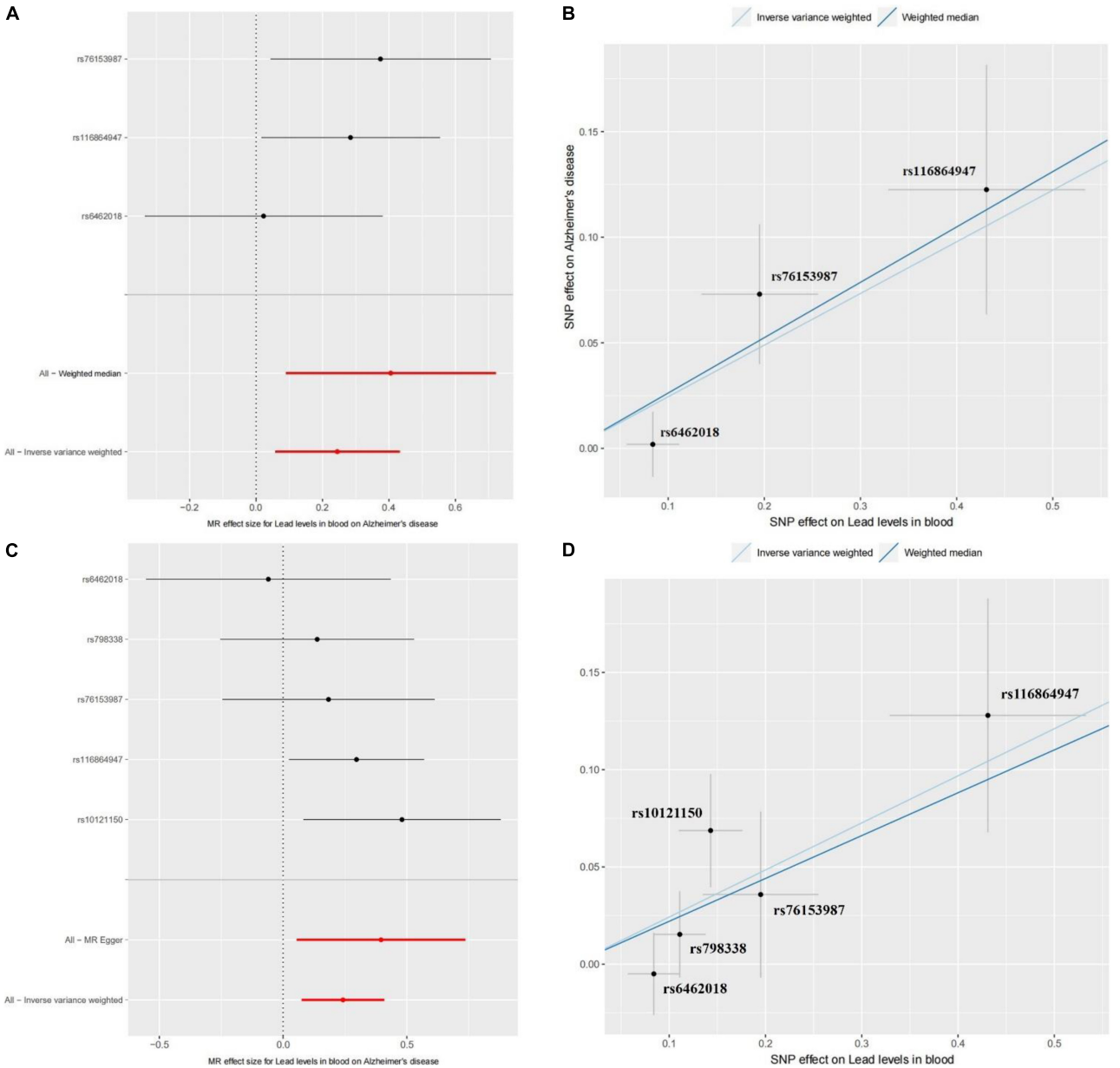

FIGURE 1 | The Mendelian randomization (MR) analysis for the causality of blood lead level and Alzheimer's disease (AD) risk. **(A)** The forest plot represents the causal effect of blood lead level on AD using the Wald ratio. The Mendelian randomization using singly each SNP and all SNPs by the WME and IVW models are shown in it. **(B)** The method comparison plot shows the SNP effects on AD against SNP effects on blood lead levels in the WME and IVW models. Each method has a different line, and the slope of the line represents the causal association. Panels **(C,D)** show the forest plot of causal effect and the method comparison plot of WME and IVW models for the EFO_0000249 dataset, respectively.

this study. The more detailed information was shown in **Supplementary Table 1**.

The Causality of Blood Lead Level and Alzheimer's Disease Risk

Using the three SNPs with their GWAS results about blood lead level and AD, we performed the two-sample Mendelian randomization analysis to assess the causal effect of blood lead level on AD risk. The results of the IVW model showed that there is a significant positive causality between blood lead level and AD risk (beta = 0.2445 and P = 0.0103). The whole confidence interval of Mendelian randomization effect size for blood lead level on AD is greater than zero (**Figure 1A**). The WME model showed similar results (beta = 0.2621 and P = 0.0382) (**Figure 1A**). As **Figure 1B** shows, the influence of the three SNPs on blood lead level and AD in the two models exhibits good consistency. To ensure the reliability of the results, we

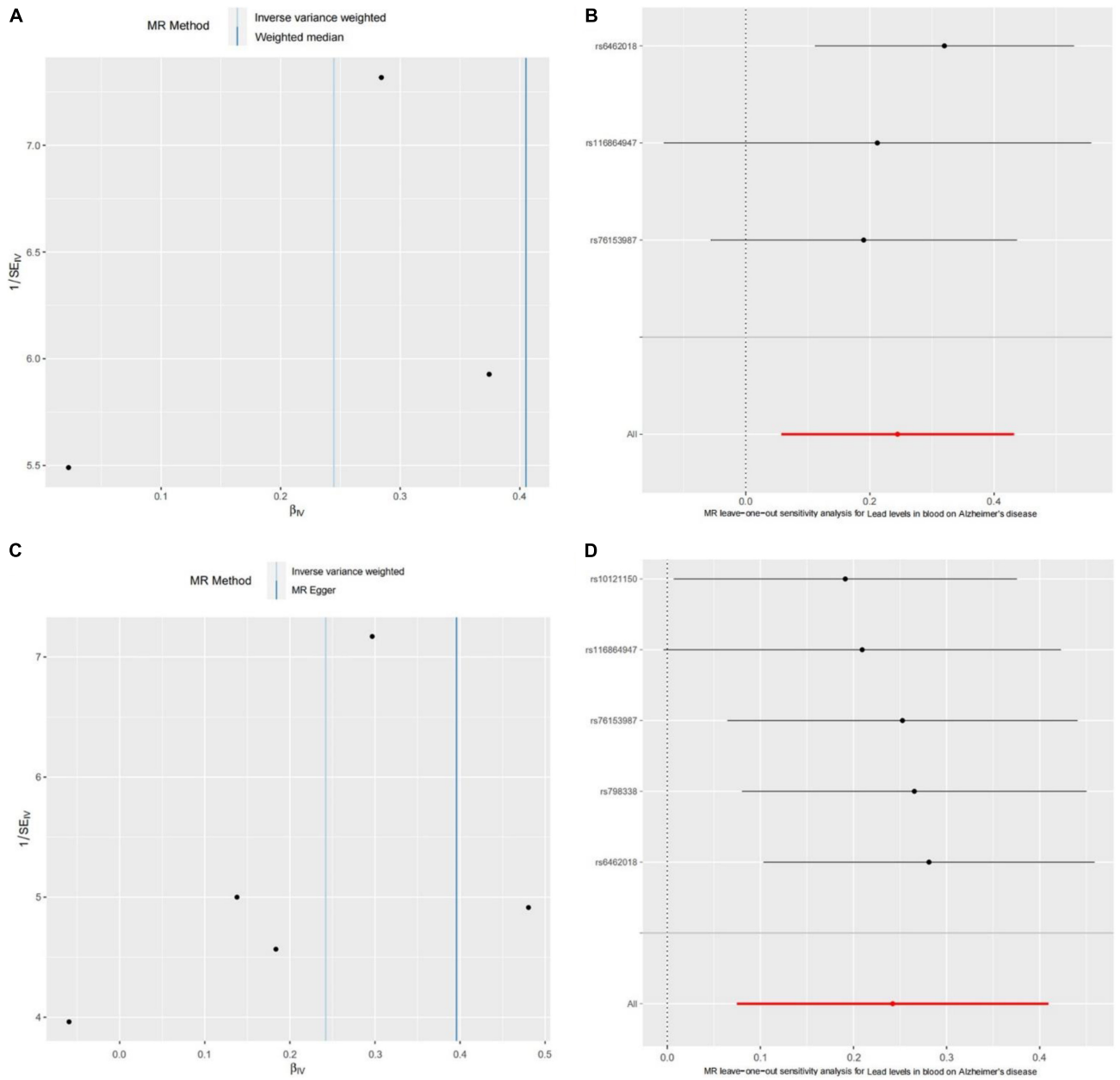

FIGURE 2 | The heterogeneity test and sensitivity analysis of the Mendelian randomization analysis. **(A)** Funnel plot to assess heterogeneity. Asymmetry and large spread suggest a high heterogeneity. **(B)** The forest plot of leave-one-out sensitivity analysis shows if an association is being disproportionately influenced by a single SNP. Each black point represents the Mendelian randomization analysis excluding that particular SNP. Panels **(C,D)** show the results of the heterogeneity test and leave-one-out sensitivity analysis for the EFO_0000249 dataset, respectively.

further performed a Mendelian randomization analysis using the five SNPs from the EFO_0000249 dataset. We found a similar result, i.e., beta = 0.2421 and 0.2203 and $P = 0.0046$ and 0.0059 in IVW and WME model, respectively (**Figure 1C**). The influence of the five SNPs on blood lead level and AD in the two models also exhibits a good consistency (**Figure 1D**). These results suggest that the elevated blood lead level increases the risk of AD. The previous studies reported that the toxicity of lead gives rise to severe environmental pollution with the use of petrol and its exposure results in cognitive decline in elderly men and women. Moreover, the blood lead level was found significantly higher in the patients with AD and is associated with an increase in AD mortality after adjusting for identified confounders (Laidlaw et al., 2017; Fathabadi et al., 2018; Horton et al., 2019). Our findings are consistent with these studies and further confirm previous conclusions, which suggest that the exposure of lead may damage the nervous system and increase risk of AD.

Reliability Check

We further performed the horizontal pleiotropy test, heterogeneity test, and sensitivity analysis to check reliability of the Mendelian randomization analysis. For the consortium's AD GWAS dataset, the results showed that there is no directional horizontal pleiotropy affecting the IVW and WME estimate (intercept = −0.029 and $P = 0.438$) (**Table 1**). Then, Cochran's Q test showed that there is also no significant heterogeneity in IVW (Cochran's Q-statistic = 2.161 and $P = 0.340$) and WME estimate (Cochran's Q-statistic = 0.684 and $P = 0.408$) (**Figure 2A**). Moreover, the leave-one-out sensitivity analysis showed that the results of the Mendelian randomization analysis do not extremely change when we removed each of the SNP orderly (**Figure 2B**). For the EFO_0000249 dataset, the similar results also showed a non-directional horizontal pleiotropy (intercept = −0.026 and $P = 0.385$) (**Table 2**), non-significant heterogeneity in IVW (Cochran's Q-statistic = 3.297 and $P = 0.510$), WME estimate (Cochran's Q-statistic = 2.269 and $P = 0.520$) (**Figure 2C**), and insignificant changes in sensitivity analysis (**Figure 2D**). These results demonstrate that the causality of blood lead level and AD is reliable, and further suggest that the elevated blood lead level increases the risk of AD.

CONCLUSION

The lead pollution is a serious environmental problem and damages to the human central nervous system. In this study, we integrated two large-scale summary AD GWAS datasets and a blood lead level GWAS dataset to assess the causality of blood lead level and AD risk by the two-sample Mendelian randomization analysis. After the reliability check, we found a significant positive causality between blood lead level and AD risk. Our findings suggest that the exposure of lead may increase risk of AD, which is further confirm the results of previous studies and benefit to understanding of AD pathogenesis and the toxicity of lead in environmental pollution.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XZ and CL designed and performed the research, analyzed the data, and wrote the manuscript. XZ, CL, YZ, JL, and CW collected the data. XZ reviewed and modified the manuscript. All authors discussed the results and contributed to the final manuscript.

FUNDING

This research was financially supported by the Technology Funds of Guizhou Province (2020-1Y391) and the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202113101).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.880105/full#supplementary-material

REFERENCES

Bakulski, K. M., Seo, Y. A., Hickman, R. C., Brandt, D., Vadari, H. S., Hu, H., et al. (2020). Heavy metals exposure and Alzheimer's disease and related dementias. *J. Alzheimers Dis.* 76, 1215–1242. doi: 10.3233/JAD-200282

Bertram, L., and Tanzi, R. E. (2005). The genetic epidemiology of neurodegenerative disease. *J. Clin. Invest.* 115, 1449–1457. doi: 10.1172/JCI24761

Bjorklund, G., Stejskal, V., Urbina, M. A., Dadar, M., Chirumbolo, S., Mutter, J., et al. (2018). Metals and Parkinson's disease: mechanisms and biochemical processes. *Curr. Med. Chem.* 25, 2198–2214. doi: 10.2174/0929867325666171129124616

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi: 10.1093/nar/gky1120

Burgess, S., and Thompson, S. G. (2011). Bias in causal estimates from Mendelian randomization studies with weak instruments. *Statist. Med.* 30, 1312–1323. doi: 10.1002/sim.4197

Burgess, S., and Thompson, S. G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* 32, 377–389.

Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* 37, 658–665. doi: 10.1002/gepi.21758

Bush, A. I. (2003). The metallobiology of Alzheimer's disease. *Trends Neurosci.* 26, 207–214.

Carmona, A., Roudeau, S., and Ortega, R. (2021). Molecular mechanisms of environmental metal neurotoxicity: a focus on the interactions of metals with synapse structure and function. *Toxics* 9:98. doi: 10.3390/toxics9090198

Consortium, G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632

Fan, X. Y., Lu, X. W., Liu, H. M., and Qin, Q. (2020). [Pollution and source analysis of heavy metal in surface dust from Xi'an university campuses]. *Huan Jing Ke Xue* 41, 3556–3562. doi: 10.13227/j.hjkx.201912041

Fathabadi, B., Dehghanifiroozabadi, M., Aaseth, J., Sharifzadeh, G., Nakhaee, S., Rajabpour-Sanati, A., et al. (2018). Comparison of blood lead levels in patients with Alzheimer's disease and healthy people. *Am. J. Alzheimers Dis. Dement.* 33, 541–547. doi: 10.1177/1533317518794032

Han, Z., Xue, W., Tao, L., and Zhu, F. (2019). Identification of key long non-coding RNAs in the pathology of Alzheimer's disease and their functions based on genome-wide associations study, microarray, and RNA-seq data. *J. Alzheimers Dis.* 68, 339–355. doi: 10.3233/JAD-181051

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 7:e34408. doi: 10.7554/eLife.34408

Horton, C. J., Weng, H.-Y., and Wells, E. M. (2019). Association between blood lead level and subsequent Alzheimer's disease mortality. *Environ. Epidemiol.* 3:e045. doi: 10.1097/EE9.0000000000000045

Huang, L., Wang, Y., Tang, Y., He, Y., and Han, Z. (2022). Lack of causal relationships between Chronic Hepatitis C virus infection and Alzheimer's disease. *Front. Genet.* 13:828827. doi: 10.3389/fgene.2022.828827

Kamel, F., Umbach, D. M., Munsat, T. L., Shefner, J. M., Hu, H., Sandler, D. P., et al. (2002). Lead exposure and amyotrophic lateral sclerosis. *Epidemiology* 13, 311–319.

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., et al. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* 51, 414–430. doi: 10.1038/s41588-019-0358-2

Laidlaw, M. A., Poropat, A. E., Ball, A., and Mielke, H. W. (2017). Exposure to lead in petrol and increased incidence of dementia. *Lancet* 389, 2371–2372.

Mates, J. M., Segura, J. A., Alonso, F. J., and Marquez, J. (2010). Roles of dioxins and heavy metals in cancer and neurological diseases using ROS-mediated mechanisms. *Free Radic. Biol. Med.* 49, 1328–1341. doi: 10.1016/j.freeradbiomed.2010.07.028

Ojuederie, O. B., and Babalola, O. O. (2017). Microbial and plant-assisted bioremediation of heavy metal polluted environments: a review. *Int. J. Environ. Res. Public Health* 14:1504. doi: 10.3390/ijerph14121504

Sekula, P., Del Greco, M. F., Pattaro, C., and Kottgen, A. (2016). Mendelian randomization as an approach to assess causality using observational data. *J. Am. Soc. Nephrol.* 27, 3253–3265. doi: 10.1681/ASN.2016010098

Staley, J. R., and Burgess, S. (2017). Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet. Epidemiol.* 41, 341–352. doi: 10.1002/gepi.22041

Tchounwou, P. B., Yedjou, C. G., Patlolla, A. K., and Sutton, D. J. (2012). Heavy metal toxicity and the environment. *Exp. Suppl.* 101, 133–164.

Van Kippersluis, H., and Rietveld, C. A. (2018). Pleiotropy-robust Mendelian randomization. *Int. J. Epidemiol.* 47, 1279–1288.

Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698.

Wang, S. L., Xu, X. R., Sun, Y. X., Liu, J. L., and Li, H. B. (2013). Heavy metal pollution in coastal areas of South China: a review. *Mar. Pollut. Bull.* 76, 7–15. doi: 10.1016/j.marpolbul.2013.08.025

Warrington, N. M., Zhu, G., Dy, V., Heath, A. C., Madden, P. A., Hemani, G., et al. (2015). Genome-wide association study of blood lead shows multiple associations near ALAD. *Hum. Mol. Genet.* 24, 3871–3879. doi: 10.1093/hmg/ddv112

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 10 August 2022
DOI 10.3389/fnins.2022.926830

OPEN ACCESS

EDITED BY
Zhijie Han,
Chongqing Medical University, China

REVIEWED BY
Sandra Rebelo,
University of Aveiro, Portugal
Steven Estus,
University of Kentucky, United States

*CORRESPONDENCE
Jin Ma
103118900@qq.com

SPECIALTY SECTION
This article was submitted to Neurogenomics, a section of the journal Frontiers in Neuroscience

RECEIVED 23 April 2022
ACCEPTED 11 July 2022
PUBLISHED 10 August 2022

CITATION
Ma J and Qiu S (2022) Genetic variant rs11136000 upregulates clusterin expression and reduces Alzheimer's disease risk.
Front. Neurosci. 16:926830.
doi: 10.3389/fnins.2022.926830

Genetic variant rs11136000 upregulates clusterin expression and reduces Alzheimer's disease risk

Jin Ma[1]* and Shizheng Qiu[2]

[1]Department of Emergency Medicine, Affiliated Kunshan Hospital of Jiangsu University, Kunshan, China, [2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

Clusterin (*CLU*) is an extracellular chaperone involved in reducing amyloid beta (Aβ) toxicity and aggregation. Although previous genome-wide association studies (GWAS) have reported a potential protective effect of *CLU* on Alzheimer's disease (AD) patients, how intron-located rs11136000 (*CLU*) affects AD risk by regulating *CLU* expression remains unknown. In this study, we integrated multiple omics data to construct the regulated pathway of rs11136000-*CLU*-AD. In step 1, we investigated the effects of variant rs11136000 on AD risk with different genders and diagnostic methods using GWAS summary statistics for AD from International Genomics of Alzheimer's Project (IGAP) and UK Biobank. In step 2, we assessed the regulation of rs11136000 on *CLU* expression in AD brain samples from Mayo clinic and controls from Genotype-Tissue Expression (GTEx). In step 3, we investigated the differential gene/protein expression of *CLU* in AD and controls from four large cohorts. The results showed that rs11136000 T allele reduced AD risk in either clinically diagnosed or proxy AD patients. By using expression quantitative trait loci (eQTL) analysis, rs11136000 variant downregulated *CLU* expression in 13 normal brain tissues, but upregulated *CLU* expression in cerebellum and temporal cortex of AD samples. Importantly, *CLU* was significantly differentially expressed in temporal cortex, dorsolateral prefrontal cortex and anterior prefrontal cortex of AD patients compared with normal controls. Together, rs11136000 may reduce AD risk by regulating *CLU* expression, which may provide important information about the biological mechanism of rs9848497 in AD progress.

KEYWORDS

Alzheimer's disease, genetic variant, *CLU*, genome-wide association study, rs11136000, eQTL

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease of the central nervous system characterized by progressive cognitive dysfunction and behavioral impairment (Scheltens et al., 2016; Van Cauwenberghe et al., 2016; Pimenova et al., 2018; Hu et al., 2021b). It is estimated that at least 40 million middle-aged and elderly people worldwide suffer from AD (Van Cauwenberghe et al., 2016). Among all the AD susceptibility genes, Apolipoprotein E (*APOE*), which mediates the binding, internalization, and catabolism of lipoprotein particles, is considered to be the major risk factor (Namba et al., 1991; Belloy et al., 2019). *APOE* not only co-deposits with beta-amyloid (Aβ) through protein-protein interaction, but also directly leads to secretion and impaired clearance of Aβ (Namba et al., 1991; Huynh et al., 2017; Belloy et al., 2019).

Other susceptibility genes, such as *CLU*, also affect the occurrence and progression of AD through the accumulation and clearance of Aβ, nerve inflammation, and lipid metabolism (Foster et al., 2019; Uddin et al., 2020a). Previous genome-wide association studies (GWAS) have shown that rs11136000 (*CLU*) is a protective locus for AD risk, and several case-control association studies replicate this result (Harold et al., 2009; Lancaster et al., 2015; Balcar et al., 2021). However, some of other studies report no statistically significant association of rs11136000 on AD or no association in non-European populations (Carrasquillo et al., 2010; Seshadri et al., 2010; Seripa et al., 2018; Zhu et al., 2018). The conflicting results of these studies made us interested in investigating the effect of rs11136000 on AD. Moreover, how rs11136000 regulates *CLU* expression and leads to AD needs further evaluation (Hu et al., 2020, 2021a).

In this study, we integrated multiple omics data, including genome-wide association study (GWAS), expression quantitative trait loci (eQTLs), transcriptome and proteome data, to investigate whether rs11136000 regulates *CLU* expression and thereby contribute to AD. In addition, we identified the different effects of rs11136000 on AD patients of different genders.

Materials and methods

Genome-wide association studies datasets

Genome-wide association studies uses single nucleotide polymorphisms (SNPs) in the human genome as molecular genetic markers to analyze the correlation between genotype and phenotype, aiming to discover genetic risk variants that affect phenotype (Tam et al., 2019). A total of five large-scale GWAS datasets for AD were included in the statistical analysis of this study. First, we obtained two GWAS datasets for AD patients with clinical or autopsy diagnosis from International Genomics of Alzheimer's Project (IGAP), including 17,008 AD cases and 37,154 controls, and 21,982 AD cases and 41,944 controls, respectively (**Table 1**; Lambert et al., 2013; Kunkle et al., 2019). In addition, we obtained three large GWAS cohorts for AD proxy from UK Biobank, including family history of maternal AD (27,696 cases and 260,980 controls), family history of patrilineal AD (14,338 cases and 245,941 controls), and family history of all AD patients (Marioni et al., 2018). All of the participants were of European descent.

Expression quantitative trait loci datasets

Expression quantitative trait loci are genetic variants that control the expression levels of quantitative trait genes. In particular, variants located in non-coding regions may cause disease by modulating gene expression. In this study, we obtained datasets that rs11136000 regulates gene expression in AD patients and controls, respectively. The eQTL data of AD and non-AD samples were obtained from Mayo clinic and Genotype-Tissue Expression (GTEx) project, respectively (**Table 2**; Allen et al., 2012; GTEx Consortium., 2017). The Mayo dataset contained gene expression data for temporal cortex (TCX) in 186 AD subjects and 170 normal subjects, and cerebellar tissue (CER) in 191 AD subjects and 181 normal subjects (Allen et al., 2012). In addition, eQTL data of 13 brain tissues, including amygdala, anterior cingulate cortex, caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord, and substantia nigra were obtained from GTEx (version 8) as controls (GTEx Consortium., 2017). The donors in GTEx were of multiple descents including European (85.3%), African (12.3%), Asian (1.4%), etc., (GTEx Consortium., 2017).

RNA expression datasets

RNA-seq data for AD versus controls was generated from over 2,100 samples from post-mortem brains of more than 1,100 individuals from seven distinct brain regions from three human cohort studies, including Religious Orders Study and Memory and Aging Project (ROSMAP), Mayo RNAseq (MAYO), and Mount Sinai Brain Bank (MSBB) (Bennett et al., 2012a,b; Zou et al., 2012; Ng et al., 2017). The seven brain regions contained dorsolateral prefrontal cortex (DLPFC), CER, TCX, frontal pole (FP), inferior frontal gyrus (IFG), parahippocampal gyrus (PHG), and superior temporal gyrus (STG).

TABLE 1 Data sources of GWAS.

Study	Traits	Diagnosis	Cases	Controls	Ethnicity
IGAP2013 (Lambert et al., 2013)	GWAS	Clinical or autopsy	25,580	48,466	European
IGAP2019 (Kunkle et al., 2019)	GWAS	Clinical or autopsy	35,274	59,163	European
UK Biobank (all) (Marioni et al., 2018)	GWAX	Clinical or autopsy	42,034	272,244	European
UK Biobank (maternal) (Marioni et al., 2018)	GWAX	Proxy	27,696	260,980	European
UK Biobank (paternal) (Marioni et al., 2018)	GWAX	Proxy	14,338	245,941	European

GWAX, genome-wide association studies by proxy. GWAS, genome-wide association studies.

Proteomics datasets

Proteomic data was generated from post-mortem brains of more than 500 individuals from four human cohort studies, including Banner Sun Health Research Institute (Banner), Baltimore Longitudinal Study on Aging (BLSA), MAYO and MSBB. Brain samples consisted of four different brain regions [DLPFC, Middle Frontal Gyrus (MFG), TCX and Anterior Prefrontal Cortex (AntPFC)]. Protein abundance was quantified using liquid-free quantification (LFQ). The proteomic data was adjusted for age, sex, and post mortem interval (PMI).

TABLE 2 The effect of genetic variant rs11136000 on *CLU* expression in AD and normal samples.

Data sources	Brain tissue	No. Samples	Beta	*P*-value
GTEx	Amygdala (non-AD)	88	−0.065	0.23
	Anterior cingulate cortex (non-AD)	109	−0.10	0.027
	Caudate (non-AD)	144	−0.041	0.16
	Cerebellar Hemisphere (non-AD)	125	−0.0012	0.98
	Cerebellum (non-AD)	154	−0.049	0.12
	Cortex (non-AD)	136	−0.058	0.065
	Frontal Cortex (non-AD)	118	−0.068	0.036
	Hippocampus (non-AD)	111	−0.069	0.076
	Hypothalamus (non-AD)	108	−0.0095	0.81
	Nucleus accumbens (non-AD)	130	−0.16	0.00023
	Putamen (non-AD)	111	−0.12	0.00082
	Spinal cord (non-AD)	83	−0.064	0.24
	Substantia nigra (non-AD)	80	−0.021	0.75
MAYO	Cerebellum (AD)	186	0.0635	0.23
	Cerebellum (non-AD)	170	−0.0905	0.048
	Temporal cortex (AD)	191	0.0588	0.031
	Temporal cortex (non-AD)	181	0.286	0.00029

Beta is the regression coefficient based on the effect allele. Beta > 0 and beta < 0 mean that this effect allele could increase and reduce gene expression, respectively. The statistically significant association is defined to be $P < 0.05/17 = 0.00294$.

The effect of genetic variant rs11136000 on Alzheimer's disease risk

We investigated the effect of rs11136000 T allele on AD risk in GWAS summary statistics for AD of clinically diagnosed or autopsy and first-degree relative proxies, respectively. In addition, we explored the effect of rs11136000 on AD patients with different genders using GWAS by proxy (GWAX) from UK Biobank. The statistically significant association is defined to be $P < 5E\text{-}08$ after adjusting for multiple testing.

The effect of rs11136000 on clusterin expression in Alzheimer's disease and controls

We investigated the potential differential *cis*-regulated effect of rs11136000 on *CLU* in AD versus controls using an additive model eQTL analysis (Hu et al., 2020, 2021b; Qiu et al., 2022). According to the additive model, each allele has an independent effect on the trait. Here, we coded the possible genotypes of rs11136000 (TT = 2, TC = 1, CC = 0), where T is an effect allele and C is a non-effect allele. Thus, the differential regulation of *CLU* expression in rs11136000 T allele carriers of AD and controls can be calculated using linear regression models. The statistically significant association is defined to be $P < 0.05/(\text{number of brain tissues}) = 0.05/17 = 0.00294$ after multiple testing.

Differential expression of clusterin between Alzheimer's disease and normal individuals

We evaluated the differential mRNA expression of *CLU* in seven brain regions between AD and controls from ROSMAP,

MAYO and MSBB. Meanwhile, we investigated the differential protein expression of *CLU* in four brain regions of AD versus controls from DLPFC, MFG, TCX, and AntPFC. The differential expression was determined via ANOVA. The significance level of differential expression was defined as $P < 0.05/7 = 0.00714$ and $P < 0.05/4 = 0.0125$ after multiple testing.

Results

rs11136000 T allele reduced Alzheimer's disease risk

rs11136000 T allele significantly reduced AD risk in both clinically diagnosed Alzheimer's cohorts from IGAP (OR: 0.92, 95%CI: 0.91-0.94, P = 1.38E-24; OR: 0.88, 95%CI: 0.86-0.91, P = 4.90E-16) (**Figure 1**). In the UK Biobank cohort (using participants whose parents suffered from AD as a proxy for cases), rs11136000 T allele was suggestively protective against AD (OR: 0.95, 95%CI: 0.93-0.97, P = 1.88E-07) (**Figure 1**). However, rs11136000 only potentially affected female individuals with AD (OR: 0.94, 95%CI: 0.92-0.96, P = 3.96E-07).

rs11136000 upregulated clusterin expression in Alzheimer's disease

rs11136000 downregulated *CLU* expression in all 13 normal brain tissues from GTEx, two of which passed multiple testing ($P_{nucleusaccumbens}$ = 0.00023 and $P_{putamen}$ = 0.00082) (**Table 2**). However, rs11136000 suggestively upregulated *CLU* expression in cerebellum (β = 0.0635, P = 0.23) and temporal cortex samples of AD (β = 0.0588, P = 0.031) (**Table 2**).

Clusterin differentially expressed in Alzheimer's disease versus controls

To further determine the effect of *CLU* in AD patients, we investigated the differential expression of *CLU* between AD and controls in various brain regions at the level of gene expression and protein expression, respectively. The *CLU* mRNA expression in temporal cortex region of AD patients significantly differed from controls regardless of gender ($\log_2^{FC}$ = 0.83, P_{TCX} = 2.96E-10) (**Table 3**). However, *CLU* was only significantly differentially expressed in parahippocampal gyrus region of female AD patients compared to controls ($\log_2^{FC}$ = 0.34, P_{PHG} = 0.00032). Furthermore, *CLU* protein was detected in DLPFC and AntPFC, and was significantly differentially expressed in both two brain tissues ($\log_2^{FC}$ = 0.29, P_{AntPFC} = 0.00022; $\log_2^{FC}$ = 0.23, P_{DLPFC} = 5.09E-06) (**Figure 2**).

Discussion

Large-scale GWAS in recent years have identified substantial genetic variants and genes associated with AD risk (Jansen et al., 2019; Kunkle et al., 2019; Schwartzentruber et al., 2021). Susceptibility loci including *APOE* have been confirmed by numerous studies (Al Mamun et al., 2020; Uddin et al., 2020b). *CLU*, also known as apolipoprotein J (APOJ) protein, is identified as the third-highest risk gene for late-onset AD (LOAD), contributing approximately 9% of AD risk (Bertram et al., 2007; Foster et al., 2019; Uddin et al., 2020a). Previous studies have shown that elevated *CLU* levels have been detected in the brain and plasma of AD individuals and are involved in neuroinflammation, lipid metabolism, and Aβ clearance in AD patients (Lidstrom et al., 1998; Bu, 2009; Thambisetty et al., 2010; Uddin et al., 2020a). However, some studies have also reported that *CLU* incorporation into amyloid aggregates is

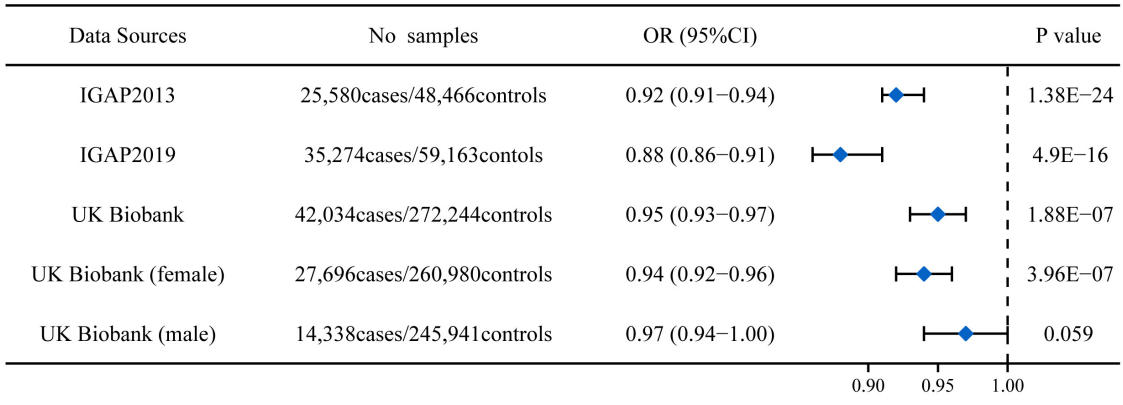

FIGURE 1
Association between rs11136000 variant T allele and AD. IGAP, International Genomics of Alzheimer's Project. The statistically significant association is defined to be P < 5E-08.

TABLE 3 Differential mRNA expression of *CLU* in AD and normal samples.

Phenotype	Brain tissue	log_2^{FC}	*P*-value
AD (all)	Cerebellum	0.12	0.43
	Dorsolateral Prefrontal Cortex	0.016	0.78
	Frontal Pole	0.15	0.033
	Inferior Frontal Gyrus	0.13	0.13
	Parahippocampal Gyrus	0.31	8.17E-07
	Superior Temporal Gyrus	0.17	0.27
	Temporal Cortex	0.83	2.96E-10
AD (female)	Cerebellum	0.20	0.29
	Dorsolateral Prefrontal Cortex	−0.013	0.87
	Frontal Pole	0.15	0.23
	Inferior Frontal Gyrus	0.19	0.14
	Parahippocampal Gyrus	0.34	0.00032
	Superior Temporal Gyrus	0.23	0.061
	Temporal Cortex	0.82	1.89E-06
AD (male)	Cerebellum	0.012	0.97
	Dorsolateral Prefrontal Cortex	0.028	0.78
	Frontal Pole	0.12	0.38
	Inferior Frontal Gyrus	0.64	0.79
	Parahippocampal Gyrus	0.24	0.084
	Superior Temporal Gyrus	0.076	0.67
	Temporal Cortex	0.84	0.00004

log_2^{FC}: log fold change value.

more harmful than Aβ42 aggregates alone (Uddin et al., 2020a). The ambiguous and complex role of *CLU* in AD prevents it from becoming a therapeutic target for AD.

In this study, we integrated GWAS, eQTL, gene expression and protein expression data to investigate whether rs11136000 (*CLU*) affects AD risk by regulating *CLU* expression. We successfully explained the pathway of rs11136000-CLU-AD. The results showed that rs11136000 significantly reduced AD risk in both clinically diagnosed AD and AD proxy. The effects of rs11136000 on AD risk with different genders and different diagnostic modalities were slightly different. In addition, previous meta-analyses and systematic reviews suggested that the heterogeneity of rs11136000 on AD risk was also reflected by race. Both Han et al. (2018) and Zhu et al. (2018) believed that rs11136000 only reduced the risk of AD in the European population, while the association was weak in the East Asian population. Subsequent eQTL analysis revealed heterogeneity of rs11136000 expression in various brain tissues. Significant difference of the regulation of rs11136000 on *CLU* expression was only showed in temporal cortex region between AD patients versus controls. Interestingly, *CLU*-immunopositive Aβ deposits were found in the temporal cortex of AD patients, and 29% of Aβ in brain tissue was associated with *CLU* protein (Martin-Rehrmann et al., 2005; Uddin et al., 2020a).

The study has some advantages. The multiple omics data used in this study were all from European populations, avoiding the bias associated with population stratification. Multiple omics data constructed a complete pathway that genetic variants regulate gene expression and then affect disease phenotype, which better explains the role of rs11136000 in the brain of AD patients than previous studies. However, this study also has certain limitations. It is difficult for us to obtain gender- and ethnic-specific multi-omics data, which limits the further disclosure of the specific regulatory role of rs11136000 on AD patients in different populations.

FIGURE 2
Differential protein expression of *CLU* between AD and normal samples. The gray boxplots represent the expression levels of *CLU* protein in the brain tissues of healthy participants. The orange dots represent the expression levels of *CLU* protein in the brain tissues of AD patients. AntPFC, anterior prefrontal cortex; DLPFC, dorsolateral prefrontal cortex.

In conclusion, this study highlights the potential role of the variant rs11136000 on AD risk by regulating *CLU* expression. These findings reveal the importance of a better understanding of *CLU* function and dysfunction in the context of normal and AD individuals.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SQ analyzed the data and drafted the manuscript. JM designed the study, revised the manuscript, and supervised this work. Both authors reviewed and approved the final version.

Acknowledgments

We acknowledge the consortiums that provided the data in this study. The results published here are in part based on data obtained from Agora, a platform initially developed by the NIA-funded AMP-AD consortium that shares evidence in support of AD target discovery. We thank the International Genomics of Alzheimer's Project (IGAP) and UK Biobank (https://www.ukbiobank.ac.uk) for providing GWAS summary data. The investigators within IGAP 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 was funded by the French National Foundation on AD 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 was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z), and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193, 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, U01 AG016976, and the Alzheimer's Association grant ADGC-10-196728. This research has been conducted using the UK Biobank resource.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Al Mamun, A., Uddin, M. S., Bin Bashar, M. F., Zaman, S., Begum, Y., Bulbul, I. J., et al. (2020). Molecular insight into the therapeutic promise of targeting APOE4 for Alzheimer's disease. *Oxid. Med. Cell. Longev.* 2020:5086250. doi: 10.1155/2020/5086250

Allen, M., Zou, F., Chai, H. S., Younkin, C. S., Crook, J., Pankratz, V. S., et al. (2012). Novel late-onset Alzheimer disease loci variants associate with brain gene expression. *Neurology* 79, 221–228. doi: 10.1212/WNL.0b013e3182605801

Alzheimer's Association (2021). 2021 Alzheimer's disease facts and figures. *Alzheimers Dement.* 17, 327–406. doi: 10.1002/alz.12328

Balcar, V. J., Zeman, T., Janout, V., Janoutova, J., Lochman, J., and Sery, O. (2021). Single nucleotide polymorphism rs11136000 of CLU gene (Clusterin, ApoJ) and the risk of late-onset Alzheimer's disease in a central European population. *Neurochem. Res.* 46, 411–422. doi: 10.1007/s11064-020-03176-y

Belloy, M. E., Napolioni, V., and Greicius, M. D. (2019). A quarter century of APOE and Alzheimer's disease: progress to date and the path forward. *Neuron* 101, 820–838. doi: 10.1016/j.neuron.2019.01.056

Bennett, D. A., Schneider, J. A., Arvanitakis, Z., and Wilson, R. S. (2012a). Overview and findings from the religious orders study. *Curr. Alzheimer Res.* 9, 628–645. doi: 10.2174/156720512801322573

Bennett, D. A., Schneider, J. A., Buchman, A. S., Barnes, L. L., Boyle, P. A., and Wilson, R. S. (2012b). Overview and findings from the rush memory and aging project. *Curr. Alzheimer Res.* 9, 646–663. doi: 10.2174/156720512801322663

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.* 39, 17–23. doi: 10.1038/ng1934

Bu, G. J. (2009). Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nat Rev Neurosci.* 10, 333–344. doi: 10.1038/nrn2620

Carrasquillo, M. M., Belbin, O., Hunter, T. A., Ma, L., Bisceglio, G. D., Zou, F., et al. (2010). Replication of CLU, CR1, and PICALM associations with alzheimer disease. *Arch Neurol.* 67, 961–964. doi: 10.1001/archneurol.2010.147

Foster, E. M., Dangla-Valls, A., Lovestone, S., Ribe, E. M., and Buckley, N. J. (2019). Clusterin in Alzheimer's disease: mechanisms, genetics, and lessons from other pathologies. *Front Neurosci.* 13:164. doi: 10.3389/fnins.2019.00164

GTEx Consortium. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213.

Han, Z. J., Qu, J. J., Zhao, J. H., and Zou, X. (2018). Analyzing 74,248 samples confirms the association between CLU rs11136000 polymorphism and Alzheimer's disease in caucasian but not Chinese population. *Sci. Rep.* 8:11062. doi: 10.1038/s41598-018-29450-2

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* 41, 1088–1093.

Hu, Y., Sun, J. Y., Zhang, Y., Zhang, H., Gao, S., Wang, T., et al. (2021b). rs1990622 variant associates with Alzheimer's disease and regulates TMEM106B expression in human brain tissues. *BMC Med.* 19:11. doi: 10.1186/s12916-020-01883-5

Hu, Y., Qiu, S., and Cheng, L. (2021a). Integration of multiple-omics data to analyze the population-specific differences for coronary artery disease. *Comput. Math. Methods Med.* 2021:7036592. doi: 10.1155/2021/7036592

Hu, Y., Zhang, H., Liu, B., Gao, S., Wang, T., Han, Z., et al. (2020). rs34331204 regulates TSPAN13 expression and contributes to Alzheimer's disease with sex differences. *Brain* 143:e95.

Huynh, T. V., Davis, A. A., Ulrich, J. D., and Holtzman, D. M. (2017). Apolipoprotein E and Alzheimer's disease: the influence of apolipoprotein E on amyloid-beta and other amyloidogenic proteins. *J. Lipid Res.* 58, 824–836.

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.

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., et al. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat. Genet.* 51, 414–430.

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, 1452–1458. doi: 10.1038/ng.2802

Lancaster, T. M., Brindley, L. M., Tansey, K. E., Sims, R. C., Mantripragada, K., Owen, M. J., et al. (2015). Alzheimer's disease risk variant in CLU is associated with neural inefficiency in healthy individuals. *Alzheimers Dement.* 11, 1144–1152. doi: 10.1016/j.jalz.2014.10.012

Lidstrom, A. M., Bogdanovic, N., Hesse, C., Volkman, I., Davidsson, P., and Blennow, K. (1998). Clusterin (apolipoprotein J) protein levels are increased in hippocampus and in frontal cortex in Alzheimer's disease. *Exp. Neurol.* 154, 511–521.

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* 8:99.

Martin-Rehrmann, M. D., Hoe, H. S., Capuani, E. M., and Rebeck, G. W. (2005). Association of apolipoprotein J-positive beta-amyloid plaques with dystrophic neurites in Alzheimer's disease brain. *Neurotox. Res.* 7, 231–241. doi: 10.1007/BF03036452

Namba, Y., Tomonaga, M., Kawasaki, H., Otomo, E., and Ikeda, K. (1991). Apolipoprotein E immunoreactivity in cerebral amyloid deposits and neurofibrillary tangles in Alzheimer's disease and kuru plaque amyloid in Creutzfeldt-Jakob disease. *Brain Res.* 541, 163–166.

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, 1418–1426. doi: 10.1038/nn.4632

Pimenova, A. A., Raj, T., and Goate, A. M. (2018). Untangling genetic risk for Alzheimer's disease. *Biol. Psychiatry* 83, 300–310.

Qiu, S., Hu, Y., and Cheng, L. (2022). BIN1 rs744373 located in enhancers of brain tissues upregulates BIN1 mRNA expression, thereby leading to Alzheimer's disease. *Alzheimers Dement.* [Epub ahead of print]. doi: 10.1002/alz.12548

Scheltens, P., Blennow, K., Breteler, M. M., De Strooper, B., Frisoni, G. B., Salloway, S., et al. (2016). Alzheimer's disease. *Lancet* 388, 505–517.

Schwartzentruber, J., Cooper, S., Liu, J. Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., et al. (2021). Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* 53, 392–402.

Seripa, D., Panza, F., Paroni, G., D'onofrio, G., Bisceglia, P., Gravina, C., et al. (2018). Role of CLU, PICALM, and TNK1 genotypes in aging with and without Alzheimer's disease. *Mol. Neurobiol.* 55, 4333–4344. doi: 10.1007/s12035-017-0547-x

Seshadri, S., Fitzpatrick, A. L., Ikram, M. A., Destefano, A. L., Gudnason, V., Boada, M., et al. (2010). Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 303, 1832–1840. doi: 10.1001/jama.2010.574

Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. doi: 10.1038/s41576-019-0127-1

Thambisetty, M., Simmons, A., Velayudhan, L., Hye, A., Campbell, J., Zhang, Y., et al. (2010). Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. *Arch Gen Psychiatry.* 67, 739–748. doi: 10.1001/archgenpsychiatry.2010.78

Uddin, M. S., Kabir, M. T., Begum, M. M., Islam, M. S., Behl, T., and Ashraf, G. M. (2020a). Exploring the role of CLU in the pathogenesis of Alzheimer's disease. *Neurotox. Res.* 39, 2108–2119. doi: 10.1007/s12640-020-00271-4

Uddin, M. S., Kabir, M. T., Rahman, M. S., Behl, T., Jeandet, P., Ashraf, G. M., et al. (2020b). Revisiting the amyloid cascade hypothesis: from anti-a beta therapeutics to auspicious new ways for Alzheimer's disease. *Int. J. Mol. Sci.* 21:5858. doi: 10.3390/ijms21165858

Van Cauwenberghe, C., Van Broeckhoven, C., and Sleegers, K. (2016). The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet. Med.* 18, 421–430. doi: 10.1038/gim.2015.117

Zhu, R. X., Liu, X., and He, Z. Y. (2018). Association between CLU gene rs11136000 polymorphism and Alzheimer's disease: an updated meta-analysis. *Neurol. Sci.* 39, 679–689. doi: 10.1007/s10072-018-3259-8

Zou, F. G., Chai, H. S., Younkin, C. S., Allen, M., Crook, J., Pankratz, V. S., et al. (2012). Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.* 8:e1002707. doi: 10.1371/journal.pgen.1002707

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 18 August 2022
doi: 10.3389/fgene.2022.905560

Variants rs2200733 and rs6843082 Show Different Associations in Asian and Non-Asian Populations With Ischemic Stroke

Dongsen Wang[1,2‡], *Xuemei Hu*[1,2‡], *Xue Yang*[2], *Mingfeng Yang*[3*†] *and Qingjian Wu*[2*†]

[1]Clinical Medical College of Jining Medical University, Jining, China, [2]Department of Emergency, Jining No. 1 People's Hospital, Jining, China, [3]Second Affiliated Hospital, Key Laboratory of Cerebral Microcirculation in Universities of Shandong, Brain Science Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Taian, China

OPEN ACCESS

Edited by:
Guiyou Liu,
Tianjin Institute of Industrial Biotechnology (CAS), China

Reviewed by:
Hui Liu,
Capital Medical University, China
Zhijie Han,
Chongqing Medical University, China

***Correspondence:**
Mingfeng Yang
mfyang@163.com
Qingjian Wu
wqw110@163.com

†ORCID:
Mingfeng Yang
orcid.org/0000-0002-6857-4743
Qingjian Wu
orcid.org/0000-0002-8746-8743

‡These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Neurogenomics, a section of the journal Frontiers in Genetics

Received: 27 March 2022
Accepted: 30 May 2022
Published: 18 August 2022

Citation:
Wang D, Hu X, Yang X, Yang M and Wu Q (2022) Variants rs2200733 and rs6843082 Show Different Associations in Asian and Non-Asian Populations With Ischemic Stroke. Front. Genet. 13:905560. doi: 10.3389/fgene.2022.905560

A previous genome-wide association study (GWAS) has reported that variants rs2200733 and rs6843082 in the paired-like homeodomain transcription factor 2 (*PITX2*) gene may be one of the risk factors for ischemic stroke (IS) in European populations. However, more recently, studies in Asia have reported that rs2200733 and rs6843082 are only weakly or not associated with increased risk of IS. This difference may be caused by the sample size and genetic heterogeneity of rs2200733 and rs6843082 among different races. For this study, we selected eight articles with nine studies from the PubMed and Embase databases, including five articles from Asian and three articles from non-Asian, to evaluate the risk of IS caused by rs2200733 and rs6843082. Then, we investigated rs2200733 and rs6843082 single-nucleotide polymorphisms (SNPs) by analysis using allele, recessive, dominant, and additive models. We identified that rs2200733 and rs6843082 are weakly significantly associated with IS for the allele model (p = 0.8), recessive model (p = 0.8), dominant model (p = 0.49), and additive model (p = 0.76) in a pooled population. Next, we performed a subgroup analysis of the population, the result of which showed that rs2200733 and rs6843082 covey genetic risk for IS in a non-Asian population, but not in an Asian population. In conclusion, our analysis shows that the effect of *PITX2* rs2200733 and rs6843082 SNPs on IS risk in Asia is inconsistent with the effect observed in European IS cohorts.

Keywords: ischemic stroke, genome-wide association study, rs2200733, rs6843082, population

INTRODUCTION

Ischemic stroke (IS) is the second leading cause of death worldwide and the main leading cause of intellectual disability in adults (Orellana-Urzua et al., 2020). The pathogenesis of IS has been studied using genome-wide association studies (GWAS), which provide a crucial direction for studying the genetic mechanism of IS (Chauhan and Debette, 2016; Liu et al., 2019b; Wei et al., 2019). In 2008, the *PITX2* rs2200733 and rs10033464 variants were identified as significant contributors to IS in a European population ($p = 2.18 \times 10^{-10}$) (Gretarsdottir et al., 2008). However, a series of subsequent studies failed to replicate those results.

In 2009, Shi et al. analyzed 383 patients with atrial fibrillation (AF) *versus* (*vs.*) 851 patients without AF and 811 patients with IS *vs.* 688 patients without IS, all of Chinese. After analysis,

rs2200733 was meaningfully correlated with AF ($p = 4.1 \times 10^{-12}$) but not IS in this Chinese population (Shi et al., 2009).

In 2012, Bertrand et al. analyzed 3548 patients with stroke *vs.* 5972 patients without stroke and then replicated their result in 5859 patients with stroke *vs.* 6281 patients without stroke, all of European ancestry. Their results showed that both rs2200733 and rs1906599 were associated with IS (OR = 1.32) (International Stroke Genetics Consortium et al., 2012). Their study again identified a significant association between rs2200733 and IS.

In 2022, Zhao et al. analyzed 476 patients with IS *vs.* 501 control individuals, all Chinese (Zhao et al., 2022). Their analysis found no meaningful association between rs6843082 and IS ($p = 0.448$).

In summary, previous studies have reported different results as to whether rs2200733 and rs6843082 increase susceptibility to IS. It is not clear whether the two SNPs (rs2200733 and rs6843082) are related to IS susceptibility. In this study, we further evaluate whether these two SNPs (rs2200733 and rs68430828) increase the risk of IS using nine studies from eight articles.

MATERIALS AND METHODS

Literature Search

The relevant literature was searched in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Embase (https://www.embase.com/) databases. We filtered all relevant studies based on the keywords "Stroke," "PITX2," "rs2200733," and "rs6843082." The literature search was completed by 10 March 2022. In the following paragraph, we describe the criteria for inclusion.

Inclusion Criteria

The inclusion criteria for our meta-analysis were as follows: (1) the study used a case–control design, (2) the study evaluated whether the two SNPs (rs2200733 and rs6843082) are risk factors for IS, (3) the study provided a clear and definite number of genotypes or alleles or enough data to calculate these numbers, and (4) the study provided an explicit odds ratio (OR) and 95% confidence interval (CI) or sufficient data to calculate the OR and 95% CI. All studies that did not meet the inclusion criteria were eliminated.

Data Extraction

For each study that met the inclusion criteria, we extracted the following information: (1) first author, (2) year of publication, (3) race of the study subjects, (4) number of cases and controls, and (5) quantity of rs2200733 and rs6843082 genotypes in cases and controls. The full results are shown in **Table 1**.

Genetic Model

We used four common genetic models for this meta-analysis, including the allele model (A *vs.* G), recessive model (AA *vs.* AG+GG), dominant model (AA+AG *vs.* GG), and additive model (AA *vs.* GG). These results are helpful to evaluate the susceptibility to IS with the two SNPs (rs2200733 and rs6843082): A allele *vs.* G allele.

Hardy–Weinberg Equilibrium

The HWE of the two SNPs (rs2200733 and rs6843082) in IS cases and the control group were analyzed using the Chi-square test. The relationship between the two SNPs (rs2200733 and rs6843082) and IS was analyzed using four gene models: allele model (A *vs.* G), recessive model (AA *vs.* AG+GG), dominant model (AA+AG *vs.* GG), and additive model (AA *vs.* GG). We performed all relevant Chi-square tests using the R program (http://www.r-project.org/).

Heterogeneity Test

First, we extracted the summary statistical information corresponding to the two SNPs (rs2200733 and rs6843082) in the above study. Then, Cochran's Q test and $I^2 = [Q—(k—1)]/Q \times 100\%$ (Liu et al., 2017) were used to analyze the heterogeneity of the two SNPs (rs2200733 and rs6843082) among these datasets. The Q statistic approximately follows a χ^2 distribution with k-1 degrees of freedom (k stands for the number of studies for analysis). When the P value from Cochran's Q statistic <0.1 and the I^2 value from Cochran's Q statistic >50%, the data showed considerable heterogeneity (Hu et al., 2017; Liu et al., 2017).

Meta-Analysis

In Cochran's Q statistic, if $p < 0.05$ or I^2 >50%, it indicated that there was heterogeneity between studies, and a random-effect model (DerSimonian–Laird) was used to calculate the pooled OR.

TABLE 1 | Characteristics of studies included in the meta-analysis.

SNPs	Study	Population	Case	Control	Case genotypes			Control genotypes		
					AA	AG	GG	AA	AG	GG
rs2200733	Gretarsdottir et al. (2008)	European	29474	6222	514	6754	22206	71	1189	4962
	Shi et al. (2009)	Chinese	811	688	200	405	206	180	344	164
	Bevan et al. (2012)	European	5859	6281	NR	NR	NR	NR	NR	NR
	Cao et al. (2013)	Chinese	1388	1629	311	692	385	342	809	478
	Su et al. (2015)	Chinese	816	816	194	417	205	191	408	217
rs6843082	Su et al. (2015)	Chinese	816	816	49	305	462	60	316	440
	Wu et al. (2015)	Chinese	167	176	12	66	89	20	78	78
	Ferreira et al. (2019)	Brazilian	240	285	128	95	17	140	120	25
	Zhao et al. (2022)	Chinese	476	501	34	187	255	40	203	258

SNP, single-nucleotide polymorphisms; NA, not publicly available; IS, ischemic stroke.

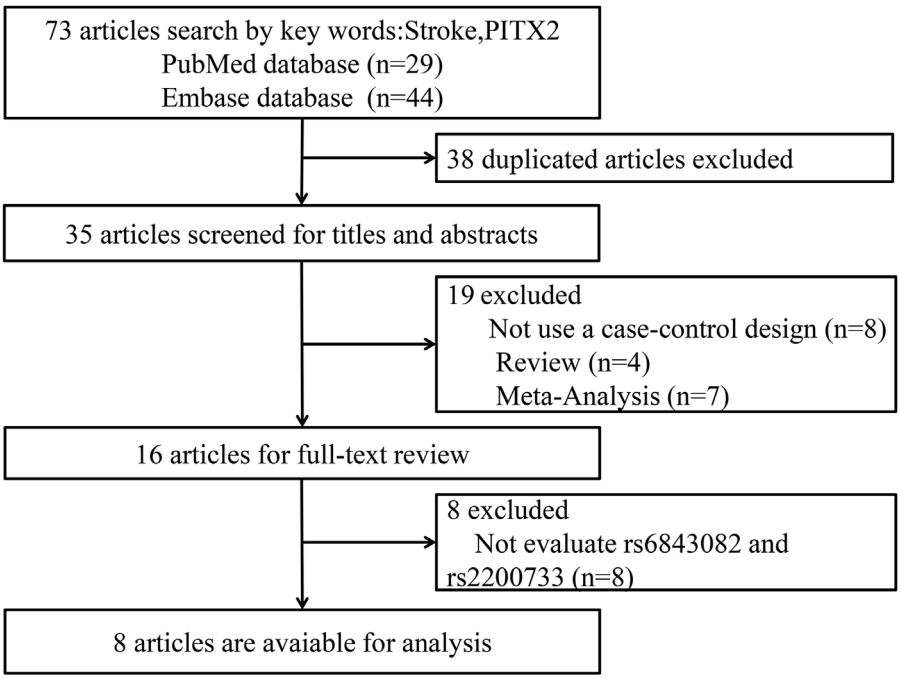

FIGURE 1 | Flowchart of the selection of studies included in this meta-analysis.

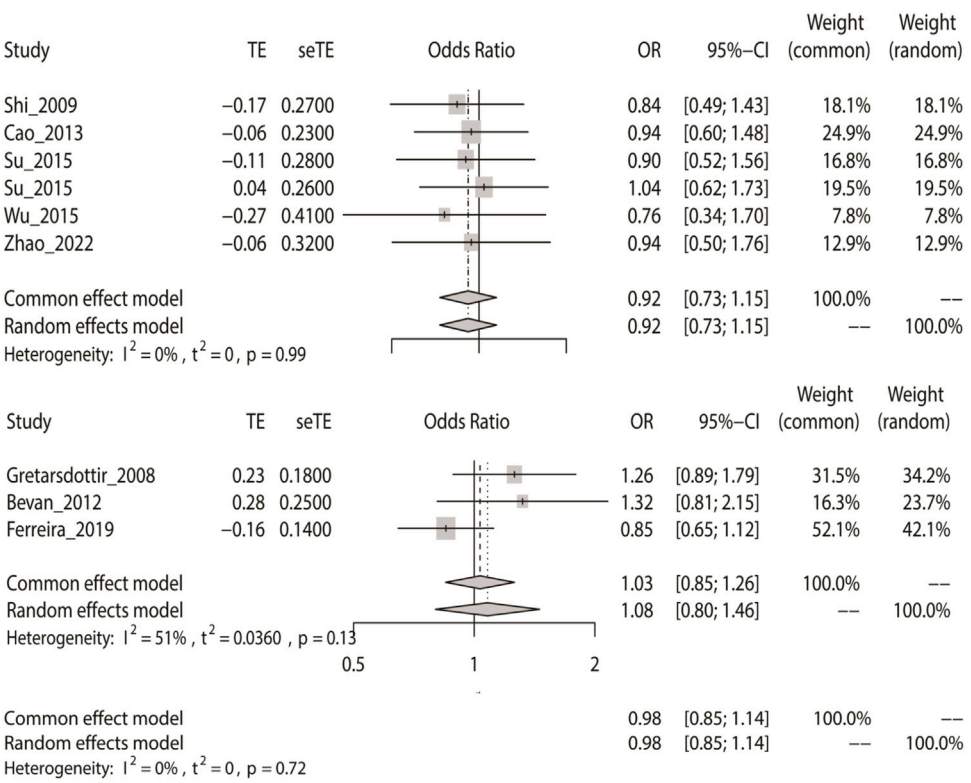

FIGURE 2 | Fixed-effect meta-analysis of the allele model for rs2200733 and rs6843082 in the Asian, non-Asian, and pooled populations.

TABLE 2 | Analysis of four genetic models' association of rs2200733 and rs6843082 with ischemic stroke.

Model	Asian		Non-Asian		Pooled population	
	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P
Allele (A *vs*. G)	0.92 (0.73–1.15)	0.45	1.03 (0.85–1.26)	0.74	0.98 (0.85–0.14)	0.80
Recessive (AA *vs*. AG+GG)	0.94 (0.69–1.28)	0.70	1.38 (0.81–2.35)	0.24	1.04 (0.79–1.36)	0.80
Dominant (AA+AG *vs*. GG)	0.95 (0.72–1.26)	0.73	1.30 (0.93–1.81)	0.13	1.08 (0.87–1.33)	0.49
Additive (AA *vs*. GG)	0.93 (0.67–1.30)	0.67	1.54 (0.84–2.82)	0.16	1.05 (0.78–1.40)	0.76

OR: odds ratio.

If not, we used a fixed-effect model (Mantel–Haenszel). All statistical methods in the meta-analysis were applied by program R (http://www.r-project.org/).

Publication Bias Analyses

In this analysis, we used funnel plots to assess the possible publication bias. When there was no publication bias, the plot of the funnel was symmetrically inverted. Otherwise, it was an asymmetric inverted funnel (Liu et al., 2014). The asymmetry of the funnel plot was evaluated by the Egger test. We performed all statistical tests using the R program (http://www.r-project.org/).

RESULTS

Comprehensive Literature Search

We retrieved 29 articles from PubMed and 44 articles from the Embase database. Finally, eight articles (Gretarsdottir et al., 2008; Shi et al., 2009; Bevan et al., 2012; Cao et al., 2013; Su et al., 2015; Wu et al., 2015; Ferreira et al., 2019; Zhao et al., 2022), including nine studies, were chosen for meta-analysis by excluding overlapping studies. A total of 55,829 participants were included in this meta-analysis: 39,231 cases in the case group (38,348 cases with rs2200733 and 1699 cases with rs6843082) and

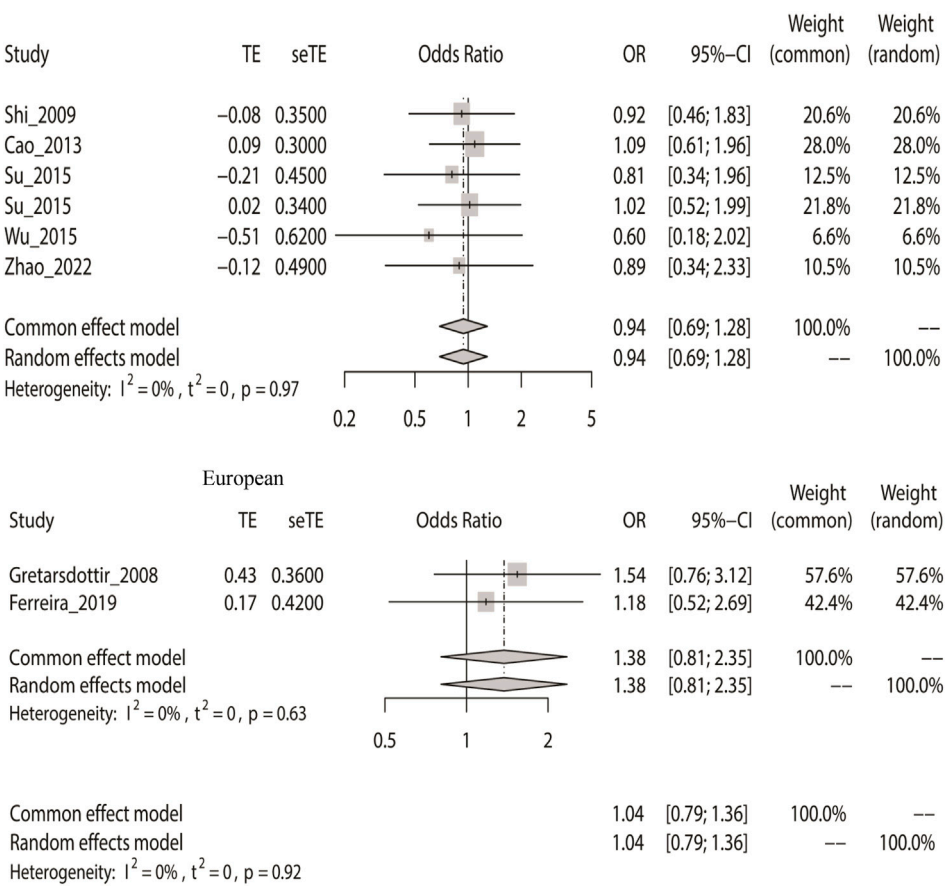

FIGURE 3 | Fixed-effect meta-analysis of the recessive model for rs2200733 and rs6843082 in the Asian, non-Asian, and pooled populations.

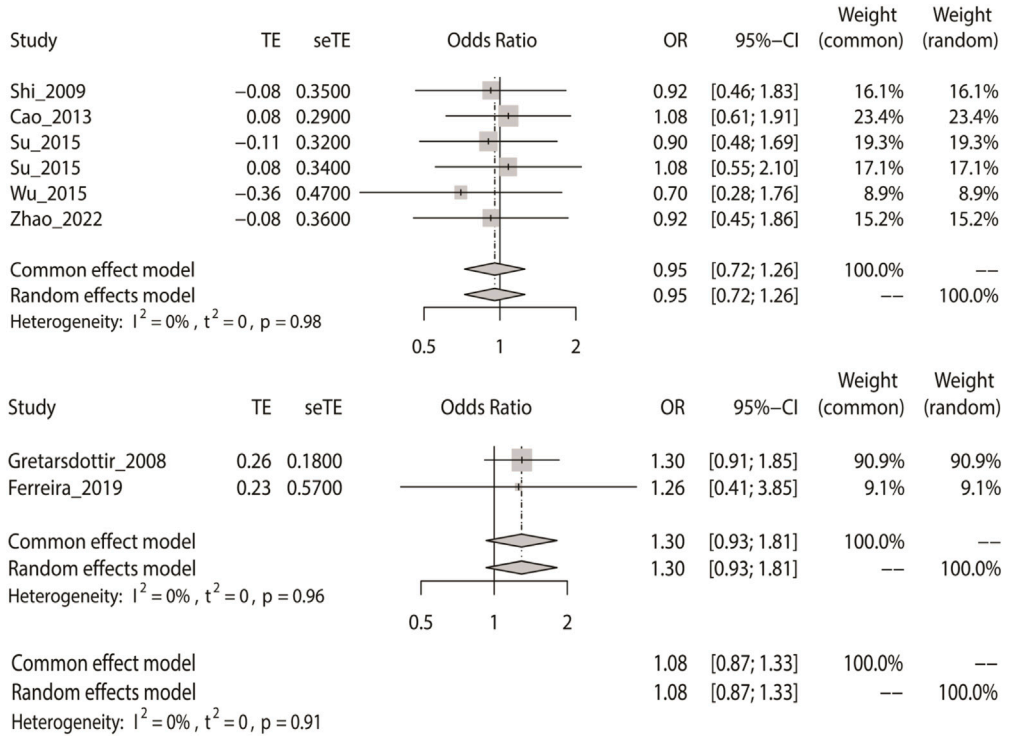

FIGURE 4 | Fixed-effect meta-analysis of the dominant model for rs2200733 and rs6843082 in the Asian, non-Asian, and pooled populations.

16,598 cases in the control group (15,636 cases with rs2200733 and 1778 cases with rs6843082). The study by Su et al. analyzed the two SNPs based on 1632 participants, and so the total number of case and controls overlapped. The specifics are shown in **Figure 1**. The primary features of the studies we included are presented in **Table 1.**

Linkage Disequilibrium

The rs2200733 and rs6843082 SNPs were located within 10 kb on *PITX2* gene (https://snipa.helmholtz-muenchen.de/snipa3/).

Heterogeneity Test

The study of Bevan et al. was excluded from the dominant, recessive, and additive models. For this analysis, we observed no remarkable heterogeneity in the pooled population when using the four genetic models (**Table 2**).

Meta-Analysis With the Allele Model

We computed the overall OR using a fixed-effect model in accordance with the outcomes of the heterogeneity test. The allele model tests showed that IS did not have a relationship with rs2200733 and rs6843082 in the Asian (p = 0.45), non-Asian (p = 0.74), and pooled populations (p = 0.80) (**Table 2**). Our results also showed that the two SNPs did not contribute to IS in Asian populations (OR = 0.92), but interestingly, the opposite results were seen in non-Asian populations, where both rs2200733 and rs6843082 were genetic risk factors for IS (OR = 1.03) (**Figure 2**).

Meta-Analysis With the Recessive Model

Similarly, we calculated the overall OR using a fixed-effect model based on the recessive model. The recessive model indicated that rs2200733 and rs6843082 and IS in the Asian (p = 0.70), non-Asian (p = 0.24), and pooled population (p = 0.80) (**Table 2**) were not closely related. The two SNPs were not associated with IS in Asian populations (OR = 0.94). Conversely, rs2200733 and rs6843082 could increase the incidence of IS disease in non-Asian populations (OR = 1.38) (**Figure 3**).

Meta-Analysis With the Dominant Model

Likewise, we calculated the overall OR using a fixed-effect model in accordance with the dominant model in the three groups. The dominant model showed that the two SNPs (rs2200733 and rs6843082) had no significant relationship with IS in Asian (p = 0.73), non-Asian (p = 0.13), and pooled (p = 0.49) populations (**Table 2**). The result of the subgroup analysis indicated that in the Asian population, the two SNPs were not genetic risk factors for IS (OR = 0.95); however, in the non-Asian population (OR = 1.30), the two SNPs were genetic risk factors for IS (OR = 1.08) (**Figure 4**).

Meta-Analysis With the Additive Model

Finally, we used the fixed-effect model to calculate the overall OR based on the additive model, where IS had no meaningful relationship with the two SNPs (rs2200733 and rs6843082) in Asian (p = 0.67), non-Asian (p = 0.16), and pooled populations (p = 0.76) (**Table 2**). The results were the same as the three

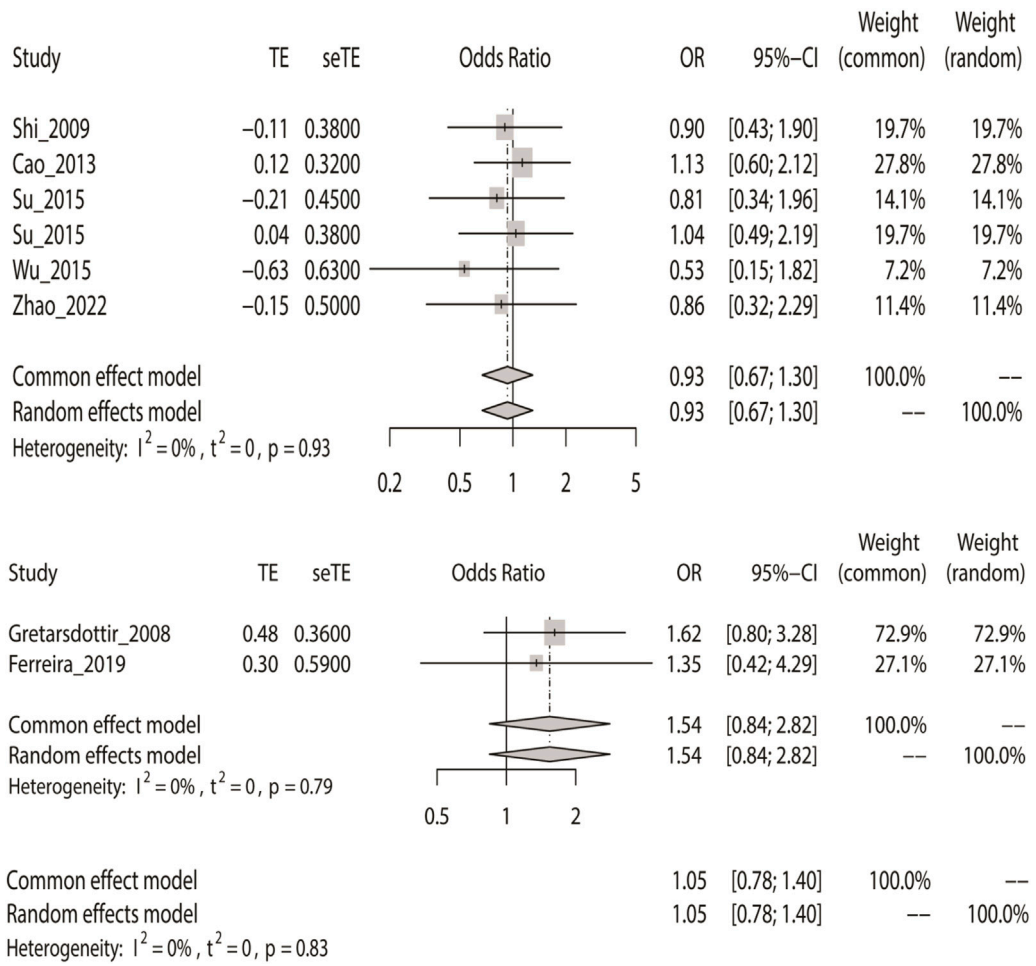

FIGURE 5 | Fixed-effect meta-analysis of the additive model for rs2200733 and rs6843082 in the Asian, non-Asian, and pooled populations.

previous genetic models; the two SNPs were not associated with IS in the Asian population (OR = 0.93); however, rs2200733 and rs6843082 were associated with an increased incidence of IS disease in the non-Asian population (OR = 1.54) (**Figure 5**).

Publication Bias Analysis

The funnel plot and Egger's test were applied to assess the existence of the potential publication bias in the four genetic models. There was no bias in the four plots, which were symmetrical inverted funnels. For the allele, recessive, dominant, and additive models, p = 0.943, 0.133, 0.053, and 0.204, respectively (**Figure 6**).

DISCUSSION

Previous GWAS studies have shown that rs2200733 and rs6843082 SNPs in *PITX2* are associated with genetic susceptibility to IS in European populations (Gretarsdottir et al., 2008). Subsequently, however, our results indicated that rs2200733 and rs6843082 conveyed no increased risk of IS. Overall, most studies have shown that the rs2200733 SNP in *PITX2* is associated with European IS, but five studies conducted in Chinese populations all concluded that the rs2200733 SNP was not associated with the IS risk. Meanwhile, three studies analyzed the association between rs2200733 and AF in the Chinese population, and the results suggested that the expression of rs2200733 had a potential genetic risk for AF, but not IS (Shi et al., 2009; Cao et al., 2013; Su et al., 2015).

In accordance with the analysis of the two SNPs (rs2200733 and rs6843082) in a pooled population, we can conclude that the G allele has low importance for the risk of IS. In the subgroup analysis, the results showed that the two SNPs had no correlation with the risk of IS in an Asian population, but the results in a non-Asian population showed a significant relevance with the risk of IS. These results suggest that the specific gene expression of that population and/or disease could be affected by genetic variation (Liu et al., 2019a). Therefore, two possibilities may lead to different associations between the two SNPs and human IS gene expression. The first factor is the racial difference, such as the genetic difference between Asian and non-Asian populations. For example, Gretarsdottir et al. indicated that

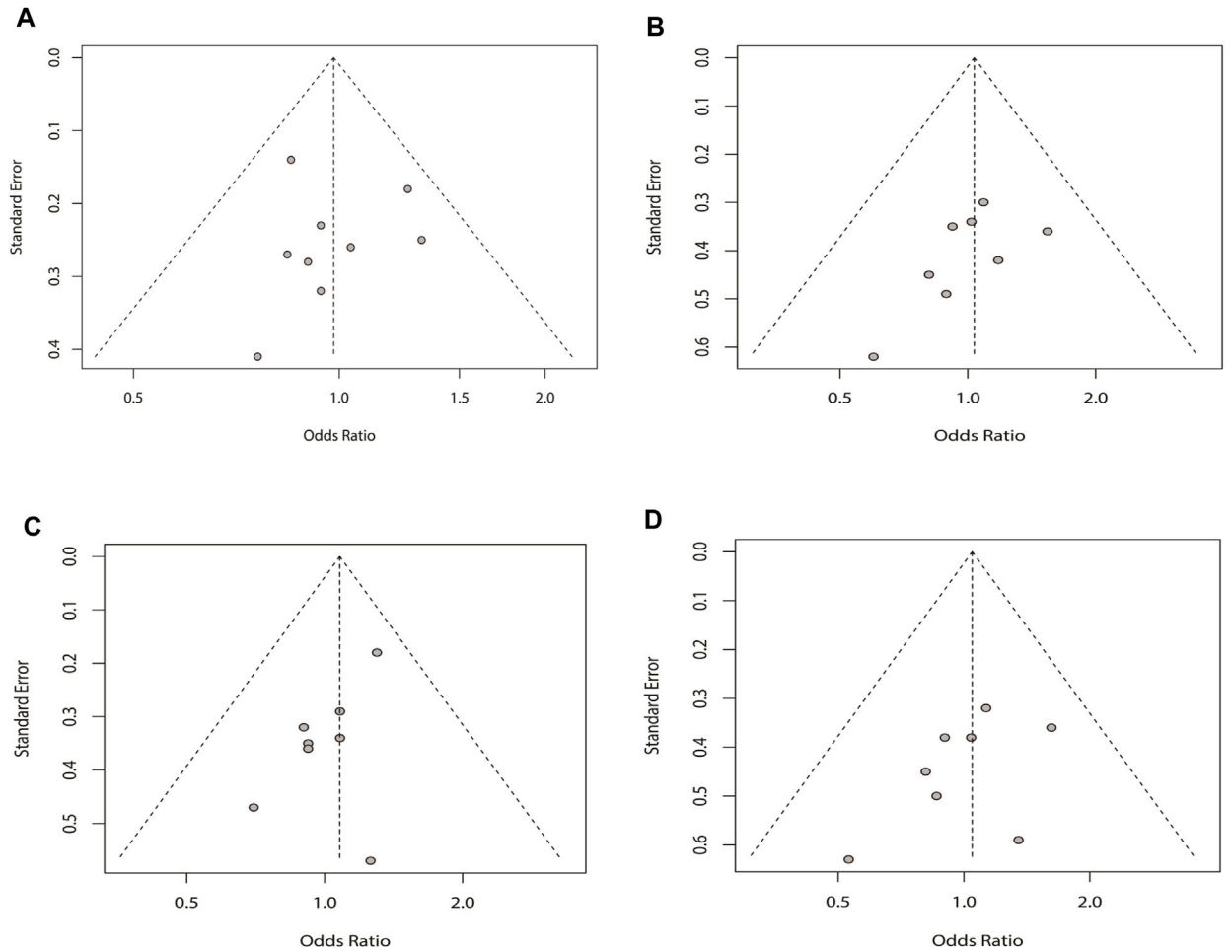

FIGURE 6 | Sensitivity analysis of the four genetic models for rs2200733 and rs6843082 in the pooled population: **(A)** allele model for the two SNPs in the pooled population, **(B)** recessive model for the two SNPs in the pooled population, **(C)** dominant model for the two SNPs in the pooled population, and **(D)** additive model for the two SNPs in the pooled population.

rs2200733 has a strong association with IS (Gretarsdottir et al., 2008), but Cao et al. showed that rs2200733 has no association with any type of stroke (Cao et al., 2013). The second probability is that the disease condition has an effect on gene expression (Tammen et al., 2013; Maruthai et al., 2022). Haplotype association analysis by Su et al. showed that rs6843082 was significantly correlated with serum total cholesterol (TC) in IS patients in the additive and dominant models (Su et al., 2015). Moreover, for female individuals, the result of the recessive model also showed that rs2200733 was associated with high levels of TC, increasing the risk of IS (Su et al., 2015). The two SNPs (rs2200733 and rs6843082) are closely related to the *PITX2* gene on chromosome 4q25, which is associated with cardiac morphogenesis, particularly the differential identity of the left atrium from the right atrium and the growth of myocardial sleeves of pulmonary veins (Logan et al., 1998; Mommersteeg et al., 2007; Tessari et al., 2008). Therefore, the *PITX2* gene may be expressed during the development of the circulatory system and play a role in IS-related risk factors. Further studies are required to determine whether these two SNPs are risk factors for IS and provide a new direction for the treatment of IS.

Numerous studies on the relationship between *PITX2* and IS have produced conflicting results. So far, there is still no final and unanimous conclusion. Our analysis includes samples from multiple researchers; therefore, the results of our study may be more reliable than those of single studies. However, some potential limitations in our meta-analysis should be acknowledged. First, there was a small sample of GWAS and candidate gene studies, which may influence the pooled estimated value. Second, environmental factors, such as smoking and alcohol use, may affect the risk of IS, but some studies did not consider these risk factors. Third, different populations have different genetic susceptibilities both to the two SNPs (rs2200733 and rs6843082) and to IS, which can make the primary cause difficult to distinguish. IS is a disease that is effected by the interactions of multiple environmental and genetic factors (Dichgans, 2007; Cai et al., 2020), and so the influence of genes and

environment on the pathogenesis of IS needs to be more deeply investigated.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

Written informed consent was not obtained from the individual(s) for the publication of no identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

MY and QW participated in the design of this study. XH, DW, and YX conducted the literature search. DW, XH, YX, MY, and QW retrieved and selected the articles. DW and XH conducted data extraction. DW and XH performed the statistical analysis of the data. DW and XH wrote the manuscript draft. MY and QW supervised the study. All authors contributed to the article. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Natural Science Foundation of Shandong Province (Grant No. ZR2021MH133), the Shandong Medicine and Health Science Technology Development Program (Grant No.2018WS470), the Shandong Traditional Chinese Medicine Science and Technology Development Program (Grant No.2019-0746), the Jining Key Research and Development Project (Grant No.2020YXNS035), and the National Natural Science Foundation of China (Grant No. 81871855).

ACKNOWLEDGMENTS

The authors thank Catherine Perfect, MA (Cantab), from Liwen Bianji (Edanz) (www.liwenbianji.cn), for editing the English text of a draft of this manuscript.

REFERENCES

International Stroke Genetics ConsortiumBellenguez, C., Bevan, S., Gschwendtner, A., Spencer, C. C. A., Burgess, A. I., Pirinen, M., et al. (2012). Genome-Wide Association Study Identifies a Variant in HDAC9 Associated with Large Vessel Ischemic Stroke. *Nat. Genet.* 44, 328–333. doi:10.1038/ng.1081

Bevan, S., Bertrand, S., Bellenguez, C., Bevan, S., Gschwendtner, A., Spencer, C. C., et al. (2012). Genome-Wide Association Study Identifies a Variant in HDAC9 Associated with Large Vessel Ischemic Stroke. *Nat. Genet.* 44, 328–333. doi:10.1038/ng.1081

Cai, H., Cai, B., Liu, Z., Wu, W., Chen, D., Fang, L., et al. (2020). Genetic Correlations and Causal Inferences in Ischemic Stroke. *J. Neurol. Sci.* 267, 1980–1990. doi:10.1007/s00415-020-09786-4

Cao, Y.-y., Ma, F., Wang, Y., Wang, D. W., and Ding, H. (2013). Rs2200733 and Rs10033464 on Chromosome 4q25 Confer Risk of Cardioembolic Stroke: An Updated Meta-Analysis. *Mol. Biol. Rep.* 40, 5977–5985. doi:10.1007/s11033-013-2707-z

Chauhan, G., and Debette, S. (2016). Genetic Risk Factors for Ischemic and Hemorrhagic Stroke. *Curr. Cardiol. Rep.* 18, 124. doi:10.1007/s11886-016-0804-z

Dichgans, M. (2007). Genetics of Ischaemic Stroke. *Lancet Neurol.* 6, 149–161. doi:10.1016/s1474-4422(07)70028-5

Ferreira, L. E., Secolin, R., Lopes-Cendes, I., Cabral, N. L., and França, P. H. C. D. (2019). Association and Interaction of Genetic Variants with Occurrence of Ischemic Stroke Among Brazilian Patients. *Gene* 695, 84–91. doi:10.1016/j.gene.2019.01.041

Gretarsdottir, S., Thorleifsson, G., Manolescu, A., Styrkarsdottir, U., Helgadottir, A., Gschwendtner, A., et al. (2008). Risk Variants for Atrial Fibrillation on Chromosome 4q25 Associate with Ischemic Stroke. *Ann. Neurol.* 64, 402–409. doi:10.1002/ana.21480

Hu, Y., Zheng, L., Cheng, L., Zhang, Y., Bai, W., Zhou, W., et al. (2017). GAB2 Rs2373115 Variant Contributes to Alzheimer's Disease Risk Specifically in European Population. *J. Neurol. Sci.* 375, 18–22. doi:10.1016/j.jns.2017.01.030

Liu, G., Hu, Y., and Jiang, Q. (2019a). Population Difference and Disease Status Affect the Association between Genetic Variants and Gene Expression. *Gastroenterology* 157, 894–896. doi:10.1053/j.gastro.2019.01.278

Liu, G., Wang, H., Liu, J., Li, J., Li, H., Ma, G., et al. (2014). The CLU Gene Rs11136000 Variant is Significantly Associated with Alzheimer's Disease in Caucasian and Asian Populations. *Neuromol Med.* 16, 52–60. doi:10.1007/s12017-013-8250-1

Liu, G., Xu, Y., Jiang, Y., Zhang, L., Feng, R., and Jiang, Q. (2017). PICALM Rs3851179 Variant Confers Susceptibility to Alzheimer's Disease in Chinese Population. *Mol. Neurobiol.* 54, 3131–3136. doi:10.1007/s12035-016-9886-2

Liu, G., Zhang, H., Liu, B., and Ji, X. (2019b). Rs2293871 Regulates HTRA1 Expression and Affects Cerebral Small Vessel Stroke and Alzheimer's Disease. *Brain* 142, e61. doi:10.1093/brain/awz305

Logan, M., Pagán-Westphal, S. M., Smith, D. M., Paganessi, L., and Tabin, C. J. (1998). The Transcription Factor Pitx2 Mediates Situs-Specific Morphogenesis in Response to Left-Right Asymmetric Signals. *Cell* 94, 307–317. doi:10.1016/s0092-8674(00)81474-9

Maruthai, K., Sankar, S., and Subramanian, M. (2022). Methylation Status of VDR Gene and its Association with Vitamin D Status and VDR Gene Expression in Pediatric Tuberculosis Disease. *Immunol. Investig.* 51, 73–87. doi:10.1080/08820139.2020.1810702

Mommersteeg, M. T. M., Brown, N. A., Prall, O. W. J., de Gier-de Vries, C., Harvey, R. P., Moorman, A. F. M., et al. (2007). Pitx2c and Nkx2-5 are Required for the Formation and Identity of the Pulmonary Myocardium. *Circulation Res.* 101, 902–909. doi:10.1161/CIRCRESAHA.107.161182

Orellana-Urzúa, S., Rojas, I., Líbano, L., and Rodrigo, R. (2020). Pathophysiology of Ischemic Stroke: Role of Oxidative Stress. *Curr. Pharm. Des.* 26, 4246–4260. doi:10.2174/1381612826666200708133912

Shi, L., Li, C., Wang, C., Xia, Y., Wu, G., Wang, F., et al. (2009). Assessment of Association of Rs2200733 on Chromosome 4q25 with Atrial Fibrillation and Ischemic Stroke in a Chinese Han Population. *Hum. Genet.* 126, 843–849. doi:10.1007/s00439-009-0737-3

Su, L., Shen, T., Xie, J., Yan, Y., Chen, Z., Wu, Y., et al. (2015). Association of GWAS-Supported Variants Rs2200733 and Rs6843082 on Chromosome 4q25

with Ischemic Stroke in the Southern Chinese Han Population. *J. Mol. Neurosci.* 56, 585–592. doi:10.1007/s12031-015-0520-y

Tammen, S. A., Friso, S., and Choi, S.-W. (2013). Epigenetics: The Link between Nature and Nurture. *Mol. Aspects Med.* 34, 753–764. doi:10.1016/j.mam.2012.07.018

Tessari, A., Pietrobon, M., Notte, A., Cifelli, G., Gage, P. J., Schneider, M. D., et al. (2008). Myocardial Pitx2 Differentially Regulates the Left Atrial Identity and Ventricular Asymmetric Remodeling Programs. *Circulation Res.* 102, 813–822. doi:10.1161/CIRCRESAHA.107.163188

Wei, C.-J., Cui, P., Li, H., Lang, W.-J., Liu, G.-Y., and Ma, X.-F. (2019). Shared Genes between Alzheimer's Disease and Ischemic Stroke. *CNS Neurosci. Ther.* 25, 855–864. doi:10.1111/cns.13117

Wu, Q., Wu, H., Gegentana, G., Huo, W., Suyalatu, S., Wu, N., et al. (2015). "Identification of the Susceptibility Gene Loci Associated with Ischemic Stroke in a Mongolian Population in China," in Computational Molecular Biology MCCMB 2015 16, 50.

Zhao, W., Hu, X., Hao, J., Guo, L., Zhang, W., Liu, J., et al. (2022). Effect of PITX2 Genetic Variants on the Susceptibility to Stroke in the Chinese Han Population. *Infect. Genet. Evol.* 98, 105201. doi:10.1016/j.meegid.2021.105201

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership