MULTIMODAL AND INTEGRATIVE ANALYSIS OF SINGLE-CELL OR BULK SEQUENCING DATA

EDITED BY: Geng Chen, Zhichao Liu and Cheng Peng

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-668-3
DOI 10.3389/978-2-88966-668-3

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

MULTIMODAL AND INTEGRATIVE ANALYSIS OF SINGLE-CELL OR BULK SEQUENCING DATA

Topic Editors:
Geng Chen, East China Normal University, China
Zhichao Liu, National Center for Toxicological Research (FDA), United States
Cheng Peng, Yunnan University, China

Citation: Chen, G., Liu, Z., Peng, C., eds. (20121). Multimodal and Integrative Analysis of Single-Cell or Bulk Sequencing Data. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-668-3

Table of Contents

frontiers in Genetics

EDITORIAL
published: 26 February 2021
doi: 10.3389/fgene.2021.658185

Editorial: Multimodal and Integrative Analysis of Single-Cell or Bulk Sequencing Data

Geng Chen [1,2]*

[1] Center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, [2] Genecast Biotechnology Co., Ltd., Wuxi, China

Keywords: multi-omics, integrative analysis, single-cell omics, multi-omics analysis, next-generation sequencing

Editorial on the Research Topic

Multimodal and Integrative Analysis of Single-Cell or Bulk Sequencing Data

Biological systems often involve the complex interactions among the molecules from different omics layers, including genome, methylome, transcriptome, proteome, metabolome, and even microbiome. At the genome level, diverse types of variants (e.g., single nucleotide variations, small insertions or deletions, and structural variations) that could be associated with a wide range of phenotypes or diseases may occur on the genome. For epigenome, it includes a variety of epigenetic modifications, such as covalent modifications on DNA and histones, chromatin accessibility and compaction, as well as the higher-order conformation of chromosome domains, which form an intricate regulatory network that can influence the chromatin structure and gene expression (Weinhold, 2006; Allis and Jenuwein, 2016). Exploration of the transcriptome was greatly revolutionized by RNA-seq technologies, which have gradually replaced traditional microarrays and provided unprecedented insights into the dynamics and complexity of gene expression (Costa et al., 2010; Stark et al., 2019). Specifically, many long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) were found to have critical regulatory functions in diverse biological processes (Marchese et al., 2017; Xiao et al., 2020). Proteins encoded by mRNAs are generally organized into higher-order structures and networks to perform catalytic, synthetic, and regulatory functions at specific times and locations (Aebersold and Mann, 2016). Mass spectrometry (MS)-based methods [such as liquid chromatography-MS/MS (LC-MS/MS)] greatly revolutionized proteome profiling and largely facilitated the dissection of complex biological processes and phenotypes (Angel et al., 2012). Furthermore, metabolome can theoretically link the genome, transcriptome, and proteome to phenotype (Misra et al., 2018). The levels and relative ratios of metabolites could generally reflect the metabolic functions, thus abnormal perturbations that beyond the normal range may indicate certain diseases (Hasin et al., 2017). Additionally, microbiomes may also significantly contribute to the biology and diverse phenotypes of their partner organisms, which can reveal the interactions between the genome and environment of the host organism (Knight et al., 2017; Lynch and Hsiao, 2019). Therefore, multi-omics analysis can promote the development of systems biology, which is essential for comprehensively investigating the dynamic changes and interactions of cellular molecules as well as understanding the underlying mechanisms (**Figure 1**).

Edited and reviewed by:
Richard D. Emes, University of Nottingham, United Kingdom

***Correspondence:**
Geng Chen
gchen@bio.ecnu.edu.cn

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

Received: *25 January 2021*
Accepted: *08 February 2021*
Published: *26 February 2021*

Citation:
Chen G (2021) Editorial: Multimodal and Integrative Analysis of Single-Cell or Bulk Sequencing Data. Front. Genet. 12:658185. doi: 10.3389/fgene.2021.658185

In this research topic, a number of research teams conducted integrative analyses to explore the molecular mechanisms or identify potential biomarkers for certain diseases/disorders. Chen et al. systematically investigated the expression profiles of placenta accreta spectrum (PAS) at both transcriptomic and proteomic levels, which provided novel insights into the underlying molecular mechanism of PAS. Through joint analysis of the interaction networks among miRNAs, mRNAs, and lncRNAs, Wang et al. identified novel potential prognostic markers for luminal breast cancer patients. Sun et al. revealed that HIF-1α pathway-related lncRNA-HEIPP (high expression in preeclampsia placenta) could play an important role in the pathogenesis of preeclampsia based on the multi-omics exploration. Wang et al. performed an integrative analysis of the underlying mechanisms of noise-induced hearing loss (NIHL) and suggested that the inflammatory pathways are closely associated with the auditory organ changes of NIHL. Zhang et al. found that G-quadruplexes could be potential targets for the drug discovery of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by systematically analyzing the non-canonical secondary structures of G-quadruplexes in both positive and negative-sense strands of SARS-CoV-2. Wang et al. revealed that the expression profile of IGFBP7 could be a potential biomarker for vasculature in response to traumatic brain injury and TGFβ signaling might be closely correlated to the upregulation of IGFBP7. Yuan et al. systematically explored the expression patterns of major depressive disorder (MDD) and uncovered that Ephrin signaling and Ras protein signal transduction could be associated with the MDD pathogenesis. Wang et al. identified potential diagnostic and prognostic biomarkers for colorectal cancer based on an integrative analysis of the datasets from different public databases. Moreover, Xiong et al. developed an integrative computational approach of ASDmiR for identifying the potential pathogenic genes, networks, and modules correlated with autism spectrum disorder. Overall, these studies performed joint analyses on the data from distinct omics layers, which gained novel insights into different diseases.

Although the development and innovation of single-cell approaches will gradually lead to a shift from bulk integrative analysis to a detailed exploration of individual cells, bulk strategies are complementary with single-cell approaches to get whole-system and cell-based perspectives and mechanisms for health and disease. One main limitation of bulk profiling methods is that they cannot accurately disentangle the cellular heterogeneity, thus single-cell exploration is essential for better elucidating the cellular behaviors and cell-to-cell variations for both basic and clinical research. However, compared to conventional bulk approaches, the single-cell technologies currently available for distinct omics are still in the early stages of development, which are often with relatively lower capture efficiency and higher technical noise (Chen et al., 2019). The improvement of the experimental procedures for single-cell protocols will reduce the technical noise and sparsity of multi-omics data, and increase the sensitivity and specificity of multimodal dissection. Moreover, conducting the single-cell study on a large number of samples is still expensive and time-consuming, the decreasing cost and simplified operation of single-cell profiling will make the multi-omics analysis more affordable and practicable. Since bulk strategies are feasible to study large-scale samples, combing bulk and single-cell data with deconvolution methods could be a good solution to investigate a multitude of individuals in a cell-type-specific manner (Li et al., 2020). Besides, the computational

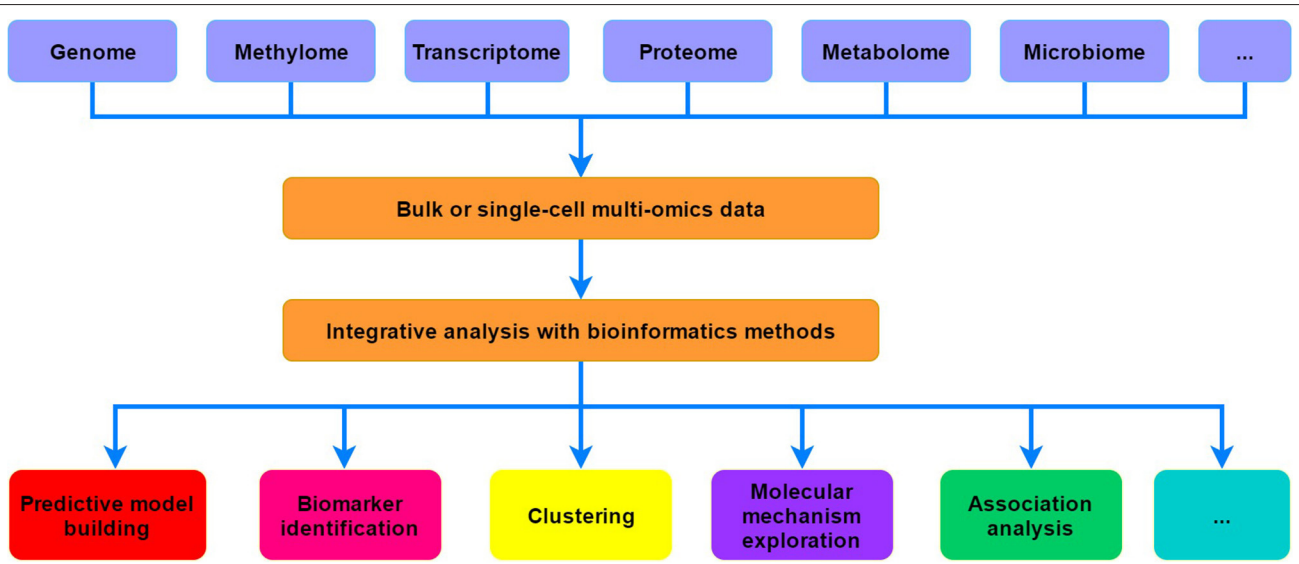


FIGURE 1 | A simple schematic view of systems biology researches based on multi-omics data.

methods for joint analysis of single-cell multimodal data are just emerging in recent years, novel bioinformatics tools are required to more efficiently integrate single-cell multi-omics data. On the other hand, existing bulk and single-cell sequencing protocols are mainly based on next-generation sequencing technologies. We envision that the utilization of third-generation sequencing approaches [e.g., Nanopore (Garalde et al., 2018)] to produce super long reads in the bulk or single-cell omics studies will largely benefit the downstream data analysis and facilitate the development of systems biology.

AUTHOR CONTRIBUTIONS

GC conceived and wrote the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31771460, 32070680, and 91629103) and the National Key Research and Development Program of China (2016YFC0902100).

REFERENCES

Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355. doi: 10.1038/nature19949

Allis, C. D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* 17, 487–500. doi: 10.1038/nrg.2016.59

Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., et al. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem. Soc. Rev.* 41, 3912–3928. doi: 10.1039/c2cs15331a

Chen, G., Ning, B., and Shi, T. (2019). Single-cell RNA-Seq technologies and related computational data analysis. *Front. Genet.* 10:317. doi: 10.3389/fgene.2019.00317

Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. *J. Biomed. Biotechnol.* 2010:853916. doi: 10.1155/2010/853916

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15, 201–206. doi: 10.1038/nmeth.4577

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

Knight, R., Callewaert, C., Marotz, C., Hyde, E. R., Debelius, J. W., McDonald, D., et al. (2017). The microbiome and human biology. *Annu. Rev. Genomics Hum. Genet.* 18, 65–86. doi: 10.1146/annurev-genom-083115-022438

Li, Y., Xu, Q., Wu, D., and Chen, G. (2020). Exploring additional valuable information from single-cell RNA-Seq data. *Front. Cell Dev. Biol.* 8:593007. doi: 10.3389/fcell.2020.593007

Lynch, J. B., and Hsiao, E. Y. (2019). Microbiomes as sources of emergent host phenotypes. *Science* 365, 1405–1409. doi: 10.1126/science.aay0240

Marchese, F. P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18:206. doi: 10.1186/s13059-017-1348-2

Misra, B. B., Langefeld, C. D., Olivier, M., and Cox, L. A. (2018). Integrated omics: tools, advances, and future approaches. *J. Mol. Endocrinol.* 62, R21–R45. doi: 10.1530/JME-18-0055

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20, 631–656. doi: 10.1038/s41576-019-0150-2

Weinhold, B. (2006). Epigenetics: the science of change. *Environ. Health Perspect.* 114, A160–A167. doi: 10.1289/ehp.114-a160

Xiao, M. S., Ai, Y., and Wilusz, J. E. (2020). Biogenesis and functions of circular RNAs come into focus. *Trends Cell Biol.* 30, 226–240. doi: 10.1016/j.tcb.2019.12.004

Conflict of Interest: GC was employed by company Genecast Biotechnology Co., Ltd.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 August 2020
doi: 10.3389/fgene.2020.00968

Systematic Transcriptome Analysis of Noise-Induced Hearing Loss Pathogenesis Suggests Inflammatory Activities and Multiple Susceptible Molecules and Pathways

Quan Wang[1,2], Yilin Shen[1,2], Haixia Hu[1,2], Cui Fan[1,2], Andi Zhang[1,2], Rui Ding[1,2], Bin Ye[1,2] and Mingliang Xiang[1,2]**

[1] Department of Otolaryngology & Head and Neck Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] Ear Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China

OPEN ACCESS

Edited by:
Zhichao Liu,
National Center for Toxicological Research (FDA), United States

Reviewed by:
Yifan Zhang,
University of Arkansas at Little Rock, United States
Ping Gong,
Engineer Research and Development Center (ERDC), United States

***Correspondence:**
Bin Ye
aydyebin@126.com
Mingliang Xiang
mingliangxiang@163.com

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

Received: *19 May 2020*
Accepted: *31 July 2020*
Published: *28 August 2020*

Citation:
Wang Q, Shen Y, Hu H, Fan C, Zhang A, Ding R, Ye B and Xiang M (2020) Systematic Transcriptome Analysis of Noise-Induced Hearing Loss Pathogenesis Suggests Inflammatory Activities and Multiple Susceptible Molecules and Pathways. Front. Genet. 11:968. doi: 10.3389/fgene.2020.00968

Noise-induced hearing loss (NIHL) is characterized by damage to cochlear neurons and associated hair cells; however, a systematic evaluation of NIHL pathogenesis is still lacking. Here, we systematically evaluated differentially expressed genes of 22 cochlear samples in an NIHL mouse model. We performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and weighted gene co-expression network analysis (WGCNA). Core modules were detected using protein–protein interactions and WGCNA with functional annotation, diagnostic value evaluation, and experimental validation. Pooled functional annotation suggested the involvement of multiple inflammatory pathways, including the TNF signaling pathway, IL-17 signaling pathway, NF-kappa B signaling pathway, rheumatoid arthritis, and p53 signaling pathway. The core modules suggested that responses to cytokines, heat, cAMP, ATP, mechanical stimuli, and immune responses were important in NIHL pathogenesis. These activities primarily occurred on the external side of the plasma membrane, the extracellular region, and the nucleus. Binding activities, including CCR2 receptor binding, protein binding, and transcription factor binding, may be important. Additionally, the hub molecules with diagnostic value included Relb, Hspa1b, Ccl2, Ptgs2, Ldlr, Plat, and Ccl17. An evaluation of Relb and Hspa1b protein levels showed that Relb was upregulated in spiral ganglion neurons, which might have diagnostic value. In conclusion, this study indicates that the inflammatory response is involved in auditory organ changes in NIHL pathogenesis; moreover, several molecules and activities have essential and subtle influences that have translational potential for pharmacological intervention.

Keywords: systematically evaluation, diagnostic value, NIHL pathogenesis, spiral ganglion neurons, functional annotation

Abbreviations: AE, ArrayExpress; AUC, area under curve; DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; NIHL, noise-induced hearing loss; OR, odds ratio; PTS, permanent threshold shift; ROS/RNS, reactive oxygen/nitrogen species; TTS, temporary threshold shift; TOM, topological overlap matrix; WGCNA, weighted gene co-expression network analysis.

INTRODUCTION

Noise-induced hearing loss (NIHL) threatens workers in many occupations, and can also affect those exposed to loud music and other recreational activities, such as concerts, clubs, and bars (Clark and Bohne, 1999; Basner et al., 2014). In 2015, the World Health Organization estimated that more than a billion teenagers and young adults suffer from NIHL, which is caused by excess exposure to loud sounds without adequate hearing protection (World Health Organization, 2015).

The biology and epidemiology of NIHL are still under extensive investigation, and data from animal models have shown that several distinct structures within the cochlea are involved (Lynch and Kil, 2005). Intense noise exposure can result in reversible or irreversible hearing loss, including temporary threshold shift (TTS) or permanent threshold shift (PTS). Both TTS and PTS involve damage to or loss of multiple cellular structures within the cochlea, including hair cells, supporting cells, spiral ganglion neurons, and cells within the stria vascularis and spiral ligament. Functional and structural damage is thought to comprise a cascade triggered by reactive oxygen/nitrogen species (ROS/RNS) and anti-ROS/RNS (Yamane et al., 1995; Yamasoba et al., 1998). Glutathione peroxidase 1 (GPx1) is an enzyme that catalyzes glutathione (GSH) to balance ROS levels, and may be essential to maintain normal cochlear function. Glutathione peroxidase 1 is highly expressed in the cochlea of mammals (Kil et al., 2007) and is deleted after noise exposure (Ohlemiller et al., 1999). In support of this hypothesis, mice with GPx1 deletion have shown increased NIHL vulnerability relative to their wild-type littermates (Ohlemiller et al., 2000). Notably, the safety and efficacy of a novel GPx1 mimic was observed in a randomized, double-blind, placebo-controlled phase 2 clinical trial for the prevention of NIHL (Kil et al., 2017). Similarly, the proliferation of peroxisomes and auditory dysfunction was rescued when the gene was restored in pejvakin-deficient mice (Delmaghani et al., 2015).

Although NIHL is characterized by damage to cochlear neurons and associated hair cells (Kurabi et al., 2017), we have yet to obtain a systematic understanding of the molecules and pathways involved. Here, we aimed to analyze differences in the gene profiles of cochlea from wild-type and noise-exposed mice, to detect changes in important molecules and pathways. An understanding of the specific molecules and pathways would significantly facilitate the advancement of future studies, health care, and clinical work.

MATERIALS AND METHODS

Data Preparation, Extraction, and Processing

Two authors performed an independent study retrieval from the Gene Expression Omnibus (GEO)[1] and ArrayExpress (AE)[2] databases for NIHL mRNA expression profiling studies. The medical subject terms included ("noise" [MeSH Terms] OR noise [All Fields]) AND ("hearing loss" [MeSH Terms] OR hearing loss [All Fields]). We limited the series study type and a 5+ sample size to increase credibility, and the two databases were screened for title, summary, general design, and description. We systematically compared the values of cochlear mRNA between noise-exposed mice and healthy controls following full-text dataset examinations and consensus meetings. Our results were validated prior to formal data extraction and processing according to previous studies (Zhang et al., 2018). Quantile normalization, log2 transformation, official gene symbol translation, and grouping annotation were saved as series matrix files for qualified studies.

After the batch effect was removed, susceptible mRNAs or differentially expressed genes (DEGs) were determined by the area under curve (AUC) and p value criteria. The *p*-value cutoff was set to 0.05, and the best threshold was calculated using Youden's J statistic (Youden, 1950; Robin et al., 2011). The maximum distance to the diagonal line was considered as the optimal cutoff point value.

Establishment of Gene Co-expression Network

As an efficient system biology method, weighted gene co-expression network analysis (WGCNA) can construct a scale-free network using gene expression data (Horvath and Dong, 2008). Experiments were conducted in accordance with the criteria of outlier samples detection and appropriate soft threshold power selection. Module identification procedures included topological overlap matrix (TOM) formation representing adjacency, hierarchical gene clustering with a deep-split value of 2, a minimum size cutoff of 30, and similar modules merging with a height cutoff of 0.4. All procedures above were computed with the module preservation function implemented in the WGCNA package (Langfelder and Horvath, 2008; Miao et al., 2018).

Module Detection and Functional Annotation

Weighted gene co-expression network analysis module detection was performed following conventional procedures. Briefly, candidates were modules with high membership scores (module-trait correlation) and genes with high significance scores (gene-trait correlation). The gene significance cutoff was set at 0.2, and the module membership score at 0.5, with a threshold *p*-value <0.05 (Horvath and Dong, 2008).

Protein-protein interaction (PPI) modules were performed in accordance with a previous study (Wang et al., 2018). Briefly, DEGs were processed for identification of the most significant interactive associations (a combined score >0.4) using the Search Tool for the Retrieval of Interacting Genes (STRING)[3] and significant cluster calculation using the Cytoscape (version 3.6.1) software (Shannon et al., 2003). Molecular Complex Detection application. Cutoff of false degree, node score, haircut, false K-score, and max depth from seed were set to 2, 0.2, true, 2, and 100, respectively.

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

[2] https://www.ebi.ac.uk/arrayexpress

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

Noise Exposure

Five-week-old mice were exposed to 10 kHz octave band noise at 108 dB sound pressure for 2 h, as previously described (Kujawa and Liberman, 2009). Briefly, mice were placed in a circular cage with four fan-shaped compartments and could move within the compartment. Setup included a soundproof chamber for cage placement, soundproofing acoustical foam for reflection minimization, a Fostex FT17H Tweeter Speaker for noise recordings at the top of the sound chamber, and a B&K sound level meter for calibration of a variation of 1.5 dB across the cage.

Histopathological Analysis

After temporal bone division, the cochlea was excised, immersed in a fixative containing 4% paraformaldehyde in phosphate buffered saline solution for 1 day, and decalcified in 10% EDTA for 7 days. Specimens were sliced into 3 μm sections and mounted onto silane-coated slides, stained, and observed under a light microscope. The evaluation of cochlear histology included three regions (apical, middle, and basal) in the Rosenthal canal. Our sample size included five per group, and every fifth modiolar section (a total of five) was subjected to histopathological assessment. The same animals were used for immunohistochemistry (IHC) staining.

Primary antibodies against the following antigens were applied: Relb (Abcam, ab180127), Hspa1b (Abcam, ab5442), and Caspase 3 (Abcam, ab44976).

Statistical Analysis

Computations and data visualization were performed using Python 3.7 or R 3.5.1. R packages included clusterProfiler (version 3.8.1), GGally (version 1.4.0), limma (version 3.36.5), network (version 1.13.0.1), pheatmap (version 1.0.10), pROC (version 1.12.1), Rtsne (version 0.13), rms (version 5.1-2), sna (version 2.4), and WGCNA (version 1.66). Significance was defined as $p < 0.05$.

RESULTS

Systematic Evaluation of DEGs in NIHL

We retrieved suitable studies, which included nine relevant mRNA expression datasets in GEO and one in AE. After duplicates were removed, the title description, summary information, and overall design of the remaining nine studies (E-GEOD-8342, GSE100365, GSE81667, GSE85290, GSE72722, GSE59416, GSE59415, GSE65249, and GSE12810) were reviewed. According to the inclusion principle, three studies (EGEOD8342, GSE100365, and GSE12810) were enrolled for the subsequent assessment. In total, 22 samples (clinical traits indicated in **Supplementary Table S1A**) were included in our study. Among them, 11 (50%) were from experimental groups, and 11 (50%) were from normal controls. The number of mRNAs among the included platforms ranged from 9,131 to 21,602 (**Supplementary Figure S1A**). Assessment of these samples was conducted using the tSNE algorithm. Among the 22 samples, notable outliers were not detected (**Supplementary Figure S1B**).

Integrated analysis of the three studies indicated 512 upregulated and 166 downregulated genes, among which 105 were transcription factors (**Supplementary Table S1B**), and the following PPI network construction and hub genes detection were based on these DEGs. As shown in **Figure 1A**, the Manhattan plot indicated the parameter of p values and affiliated chromosomes of all genes including the DEGs. The top 15 upregulated genes included cAMP responsive element modulator (Crem), heat shock protein 1B (Hspa1b), GTP binding protein (Gem), oncostatin M receptor (Osmr), immediate early response 3 (Ier3), growth differentiation factor 15 (Gdf15), jun proto-oncogene (Jun), FBJ osteosarcoma oncogene (Fos), activating transcription factor 3 (Atf3), tropomyosin 4 (Tpm4), basic helix-loop-helix family, member e40 (Bhlhe40), reticuloendotheliosis viral oncogene related B (Relb), growth arrest and DNA-damage-inducible 45 alpha (Gadd45a), squalene epoxidase (Sqle), and iodothyronine deiodinase 2 (Dio2). The top 15 downregulated genes included aldehyde oxidase 1 (Aox1), sulfatase modifying factor 1 (Sumf1), immunoglobulin binding protein 1b (Igbp1b), glutathione S-transferase alpha 4 (Gsta4), receptor activity modifying protein 1 (Ramp1), fumarylacetoacetate hydrolase (Fah), PR domain containing 5 (Prdm5), S100P binding protein (S100pbp), interferon regulatory factor 4 (Irf4), refilin B (Rflnb), cytochrome c oxidase subunit 8B (Cox8b), amine oxidase, copper containing 3 (Aoc3), NAD(P)H quinone dehydrogenase 1 (Nqo1), DNA segment Chr 8 ERATO Doi 67 expressed (D8ertd67e), and potassium voltage-gated channel subfamily J member 2 (Kcnj2) (**Figure 1B**).

These top genes could construct a network that has significantly more interactions than expected ($p < 0.0001$) (**Supplementary Figure S1D**). The mainly involved biological processes include response to stimulus, cellular process, response to abiotic stimulus, circadian rhythm, and response to external stimulus. The most enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways are Tyrosine metabolism, MAPK signaling pathway, HTLV-I infection, Fluid shear stress and atherosclerosis, and Colorectal cancer (**Supplementary Table S1F**). These results indicated that noise-induced cochlear pathology might have similar mechanisms with other diseases, such as cancer and atherosclerosis.

Analysis of Series-Based DEGs

As frequently measured DEGs are more reliable for diagnosing NIHL, we analyzed series-based DEGs. Individual DEG analysis indicated that 34 overlapped in all three studies and 870 in two studies, with overlapping counts shown in **Figure 1C**.

Based on Youden's J statistic and the AUC calculation (AUC ranged from 0.84 to 1.00), we ranked these 34 DEGs (**Supplementary Table S2**). Among the 34 DEGs, 13 displayed significant diagnostic value in the same direction (upregulated or downregulated), including Crem, casein kinase 2 alpha 2 (Csnk2a2), cleavage stimulation factor subunit 3 (Cstf3), cysteine rich angiogenic inducer 61 (Cyr61), Hspa1b, interleukin 13 receptor subunit alpha 1 (Il13ra1), Jun, pleckstrin homology domain containing family H member 1 (Plekhh1), purine rich element binding protein A (Pura), Ras association domain family member 1 (Rassf1), Relb, S100pbp, and tropomodulin 3 (Tmod3)

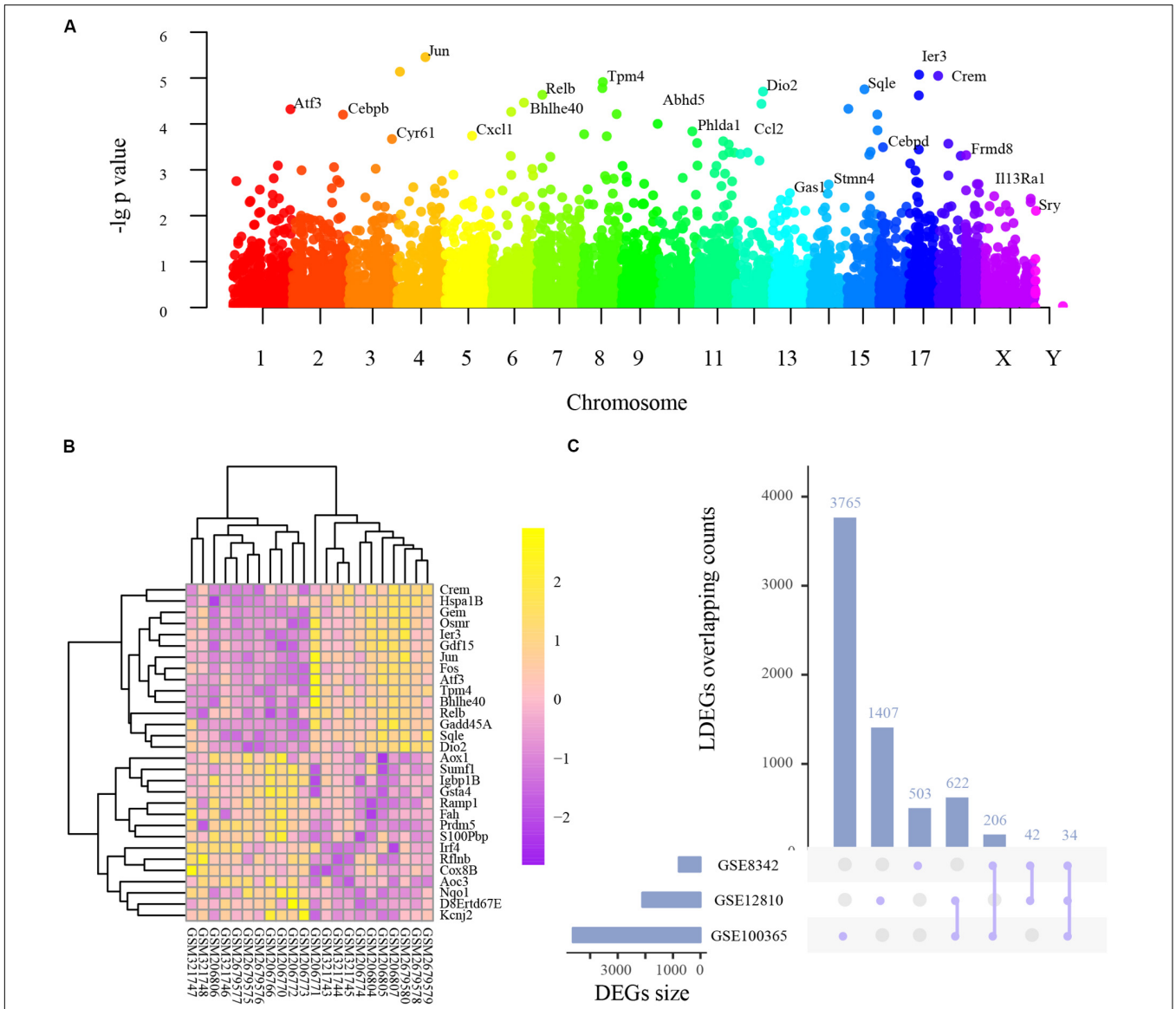


FIGURE 1 | The basic information of all detected genes and representative differentially expressed genes (DEGs). **(A)** Manhattan plot of all genes including the DEGs. The abscissa indicates the chromosomes, and the ordinate represents the p values of those genes. The most significant gene of each chromosome is labeled respectively. **(B)** Heatmap of the top 30 upregulated and downregulated genes. Purple indicates relatively lower expression, and yellow indicates relatively higher expression. **(C)** The overlapping counts of DEGs in the three studies. The abscissa indicates the seven possible sets, and the ordinate suggests to which set the number of DEGs affiliate.

(**Table 1**). We present the expression levels of these genes in **Supplementary Figure S1C**, and the statistical results in **Table 1**.

These DEGs could also construct a network that has significantly more interactions than expected (p = 0.0228) (**Supplementary Figure S1E**). The mainly involved biological processes include regulation of cell cycle, circadian rhythm, negative regulation of the cellular process, regulation of cell cycle process, and myeloid cell differentiation. The most enriched KEGG pathways are Epstein-Barr virus infection, MAPK signaling pathway, Mitophagy – animal, HTLV-I infection, and NF-kappa B signaling pathway (**Supplementary Table S1G**). These results indicated that noise-induced cochlear pathology might be an innate immune response and might have similarities with the infection of Epstein-Barr virus or HTLV-I.

Pooled Functional Annotation

To avoid type two error in further analyses, we subjected the DEGs to functional annotation using the clusterProfiler R package (Yu et al., 2012). Enriched KEGG pathways included the TNF signaling pathway, the IL-17 signaling pathway, the NF-kappa B signaling pathway, rheumatoid arthritis, legionellosis, Salmonella infection, osteoclast differentiation, pertussis, Chagas disease, the AGE-RAGE signaling pathway, the prolactin signaling pathway, the adipocytokine signaling pathway, the p53

TABLE 1 | Thirteen differentially expressed genes displayed significant diagnostic value in the same direction.

Gene	J1	J2	J3	AUC1	AUC2	AUC3	TOP	TF	HG
Crem	7.220904	5.386261	6.545	0.88	0.8889	1	Y	Y	Y
Csnk2a2	6.553653	7.273568	3.46	0.96	1	1	N	N	N
Cstf3	7.503475	7.361176	8.55	0.92	1	1	N	N	Y
Cyr61	7.979865	3.887051	7.28	1	1	0.8889	N	N	N
Hspa1b	9.530462	3.392444	6.44	0.92	0.8889	1	Y	N	Y
Il13ra1	6.634921	4.304466	5.07	1	1	1	N	N	N
Jun	8.993056	7.038203	7.25	0.96	1	1	Y	Y	Y
Plekhh1	5.264403	4.338483	6.34	0.84	1	1	N	N	N
Pura	10.53739	5.376438	10.055	0.96	1	1	N	Y	N
Rassf1	7.538773	6.913554	5.84	0.96	1	0.8889	N	N	N
Relb	4.052692	3.586543	5.885	0.96	0.8889	1	Y	Y	Y
S100pbp	8.250666	4.456705	7.555	0.88	1	1	Y	N	N
Tmod3	12.055	3.754847	7.535	0.96	1	1	N	N	N

J1, J2, J3 mean Youden's J statistics in three enrolled studies, and AUC1, AUC2, AUC3 mean area under curve, respectively. Y(yes) or N(no) indicates whether respective gene is in TOP (TOP30 DEGs), TF (Transcription Factors), HG (Hub Genes) or not.

signaling pathway, the C-type lectin receptor signaling pathway, insulin resistance, transcriptional misregulation in cancer, the FoxO signaling pathway, prostate cancer, small cell lung cancer, the Toll-like receptor signaling pathway, herpes simplex infection, colorectal cancer, apoptosis, the NOD-like receptor signaling pathway, hepatitis B, Kaposi sarcoma-associated herpesvirus infection, the chemokine signaling pathway, cytokine-cytokine receptor interaction, the MAPK signaling pathway, and human cytomegalovirus infection (**Figure 2A** and **Supplementary Table S1C**).

We noticed that inflammation-associated pathways occurred frequently, including TNF, IL-17, NF-kappa B, and Toll-like receptor signaling pathways (Cai et al., 2014). These results coincide with previous studies (Tornabene et al., 2006; Moon et al., 2007; Vethanayagam et al., 2016), indicating the enormous impact of inflammation in auditory organs. We also observed that apoptosis was upregulated (**Figure 2B**). The Caspase 3 IF staining showed that the caspase 3 positive cells were significantly increased in noise group, especially in neurons and stria vascularis areas (**Figures 2C,D**).

Gene Co-expression Network and PPI Network

For the gene co-expression network, a soft threshold of 10 was selected with the correlation coefficient set at 0.85 (**Supplementary Figure S2A**) and seven modules were detected by the dynamic tree cut (**Supplementary Figure S2B**). A topological overlap profile of these modules was shown in **Figure 3A**, darker squares along with the diagonal corresponding to modules. The relationship between the detected modules and mouse group (control or treated) was also calculated and modules with insufficient relationship values were abandoned (**Figure 3B**). The PPIs of the most enriched WGCNA module (green, **Figure 3B**) was shown in **Figure 3C**. A total of 92 nodes and 117 protein pairs were obtained with a combined weight score >0.4.

For the PPI network, a total of 2,945 pairs were detected (**Supplementary Table S1E**) and improved network visualization of them was presented in **Supplementary Figure S3** with a combined score set at more than 0.9. The most enriched PPI module (score = 13.45, nodes = 41, edge = 269) indicated 41 susceptible molecules (**Figure 3D**) including Relb, Hspa1b, C-C motif chemokine ligand 2 (Ccl2), prostaglandin-endoperoxide synthase 2 (Ptgs2), low density lipoprotein receptor (Ldlr), plasminogen activator tissue type (Plat), and chemokine C-C motif ligand 17 (Ccl17). These DEGs could also construct a network that has significantly more interactions than expected ($p < 0.0001$) (**Supplementary Figure S1F**).

Functional Annotation of WGCNA and PPI Enriched Modules

To explore the biological relevance of these modules, the genes were subjected to Gene Ontology (GO) functional and KEGG pathway enrichment analyses. Thirteen overlapping pathways between the two modules included the TNF signaling pathway, rheumatoid arthritis, the NOD-like receptor signaling pathway, the prolactin signaling pathway, Salmonella infection, cytokine-cytokine receptor interaction, the MAPK signaling pathway, HTLV-I infection, malaria, herpes simplex infection, osteoclast differentiation, legionellosis, and the FoxO signaling pathway (**Figure 4** and **Supplementary Table S1D**).

Thirty-seven biological processes were consistently observed, indicating several potential responses, including the response to cytokines, response to heat, response to cAMP, response to ATP, immune response, and the response to mechanical stimulus (**Figure 4** and **Supplementary Table S1D**). Six cellular components were consistently observed, suggesting that the main molecular and biological changes occur in the external side of the plasma membrane, extracellular region, and nucleus (**Figure 4** and **Supplementary Table S1D**). Nine overlapping molecular function terms indicated that binding activities might be important. Those activities included CCR2 receptor binding,

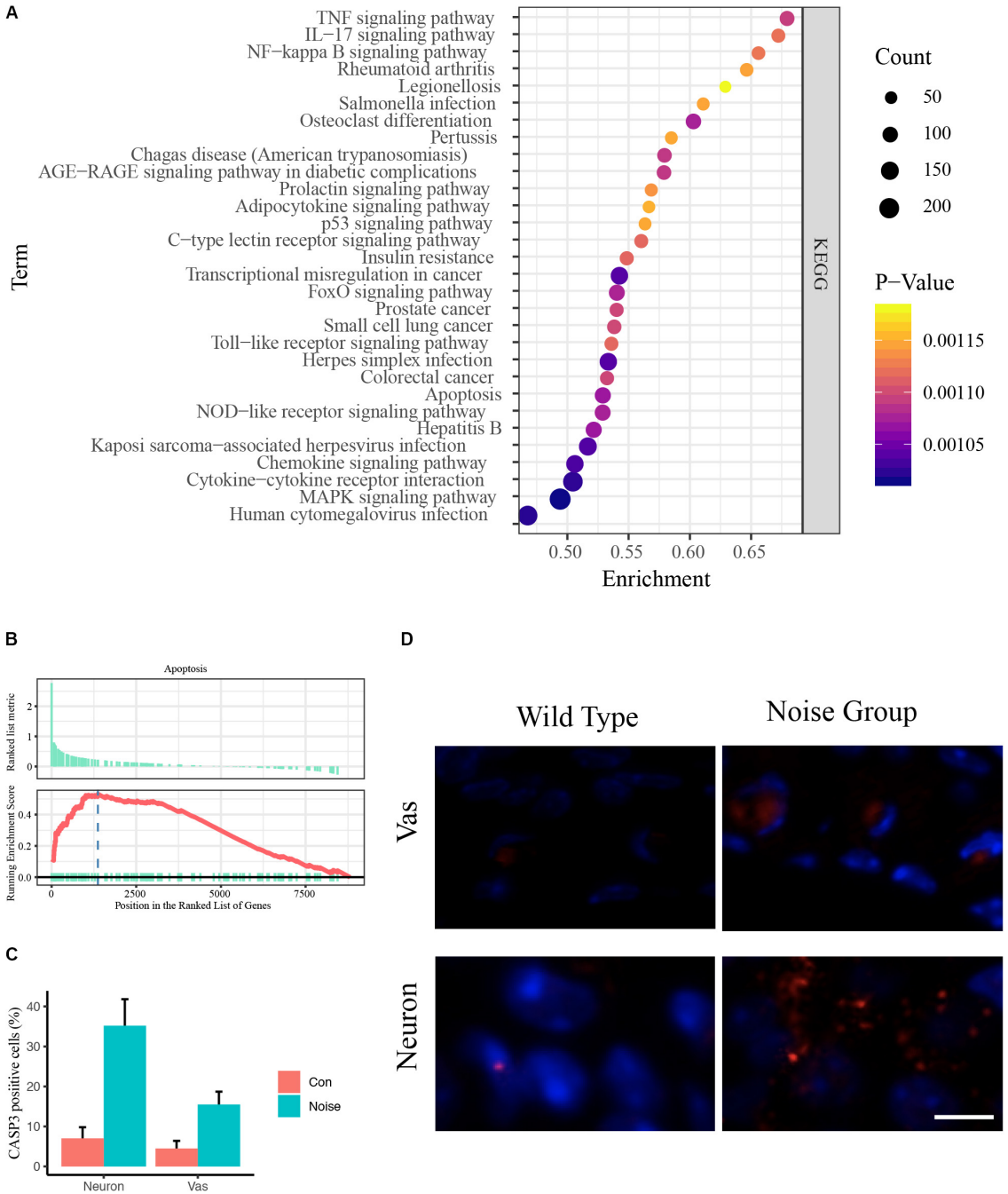


FIGURE 2 | Pooled functional annotation and validation. **(A)** KEGG pathway enrichment analysis of all differentially expressed genes. The abscissa represents the enrichment scores, and the ordinate suggests KEGG terms. **(B)** Gene set enrichment analysis (GSEA) plot of apoptosis. The abscissa represents the ranks of all detected genes, and the ordinate suggests the running score. Most apoptosis molecules lie in the upregulated position, indicating an upregulation of apoptosis in NIHL. **(C)** Quantitative analysis of CASP3 positive cells in wild-type and noise-exposed cochlea. Data are shown as Mean ± SEM. Vas: stria vascularis. **(D)** Representative image of CASP3 IF staining in spiral ganglion area and stria vascularis area of wild-type **(left)** and noise-exposed **(right)** cochlea. Vas: stria vascularis. Scale bar: 10 μm.

protein binding, and transcription factor binding (**Figure 4** and **Supplementary Table S1D**).

Interestingly, our data were consistent with pooled annotations (**Supplementary Table S1C**). Notably, IL6, NFKBIA, NFKB1, NOS3, PTPN1, PIK3R1, and STAT3 (members in the enriched WGCNA module) have been shown to alter insulin resistance (Seo et al., 2016). Researchers have shown that insulin resistance can decrease the number of ribbon synapses and

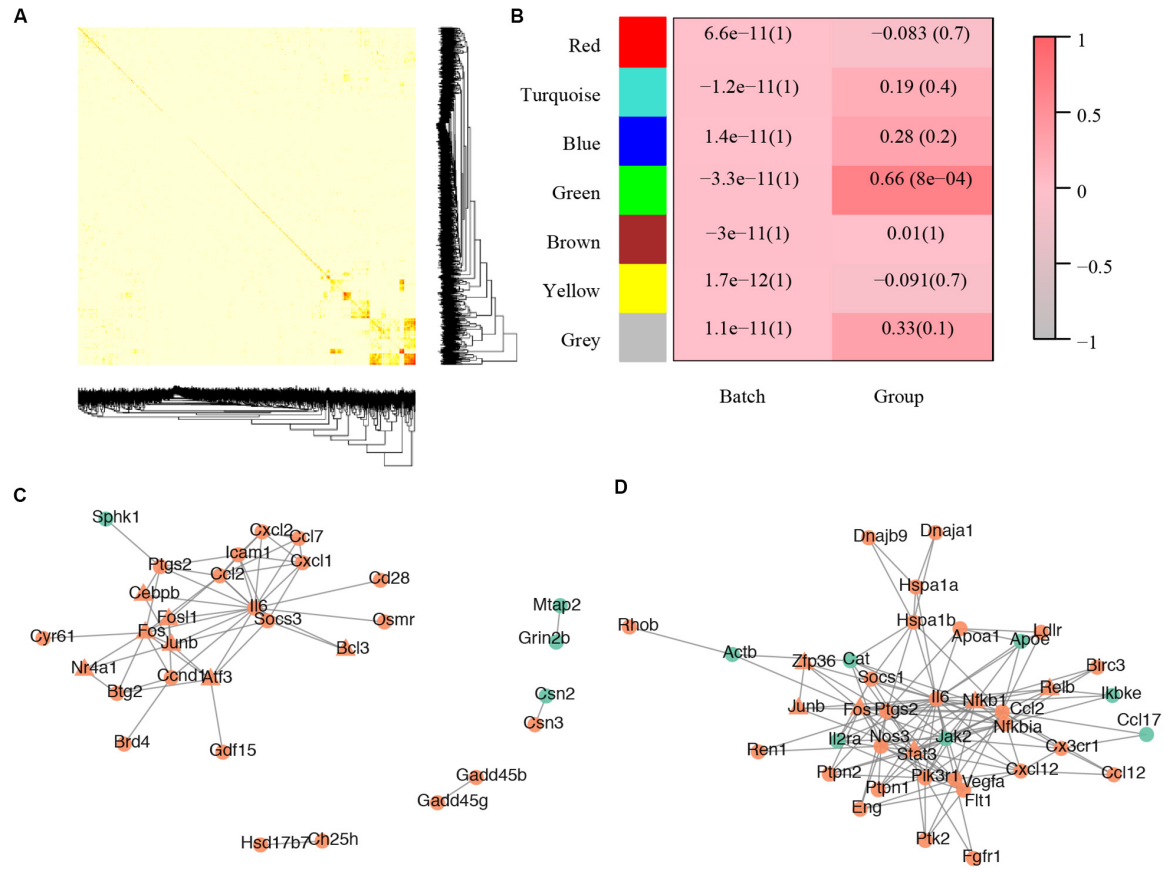


FIGURE 3 | Core modules of noise-induced hearing loss pathogenesis. **(A)** Heatmap plot of network topology. In the right and bottom are gene dendrogram and module assignment. A light color denotes low topological overlap, and progressively darker red denotes higher topological overlap. Darker squares along the diagonal correspond to modules. Seven modules were detected. **(B)** Module-trait correlation. Each cell contains the respective correlation and *p*-value. Red represents high correlation. **(C)** Protein–protein interaction of molecules in the most enriched WGCNA module. **(D)** Protein–protein interaction of the most enriched cluster of all differentially expressed genes.

elevate the ABR threshold in an age-related hearing loss model without affecting OHCs, IHCs, and SGNs (Yu et al., 2015).

Experimental Validation of Susceptible Hub Genes

All PPI susceptible molecules were utilized in logistic regression (stepwise) to predict NIHL as shown in **Figure 5A**. Logistic regression predicted that Ln OR = −19.078 + Relb × 1.608 + Hspa1b × 0.308 + Ccl2 × −0.227 + Ptgs2 × 0.155 + Ldlr × 0.814 + Plat × 0.813 + Ccl17 × −0.364 (Multiple R-squared: 0.9822). The AUCs for Relb, Hspa1b, Ccl2, Ptgs2, Ldlr, Plat, and Ccl17 were 0.9669, 0.9587, 0.8926, 0.8843, 0.876, 0.8678, and 0.8347, respectively, with the receiver operating characteristic curve presented in **Figure 5B**. We hypothesized that Relb, Hspa1b, Ccl2, Ptgs2, Ldlr, Plat, and Ccl17 were hub genes with diagnostic value at the transcription level, and Relb and Hspa1b IHC were assessed to validate our hypothesis (**Figures 5C,D**) at the translation level. While Hspa1b may have post-transcription modifications, the IHC results of Relb were quantitatively consistent with the microarray results.

DISCUSSION

Prevention and treatment are two strategies used against NIHL. However, prevention strategies, including hearing conservation programs, education, and ear protection devices are ineffective because of their lack of efficiency and accessibility (Chen and Tsai, 2003; Rajguru, 2013; Sekhar et al., 2014).

High doses of corticoids (a widely used anti-inflammatory hormone) are often prescribed by physicians after an acoustic trauma to mitigate the inflammatory response. In 2006, researchers revealed that stressors (sound, heat, or stress), as a pre-emptive medication, could increase the receptiveness of glucocorticoids by increasing the number of receptors of the anti-inflammatory hormone (Tahera et al., 2007). It reduces the inflammation caused by the acoustic trauma, which causes subsequent damage to hair cells.

For NIHL treatment, a series of clinical trials have been conducted. Intratympanical injection with AM-111 may have a therapeutic effect in cases of acute trauma after firecracker exposure (Suckfuell et al., 2007). Immediate treatment with a combination of prednisolone and piracetam appeared to

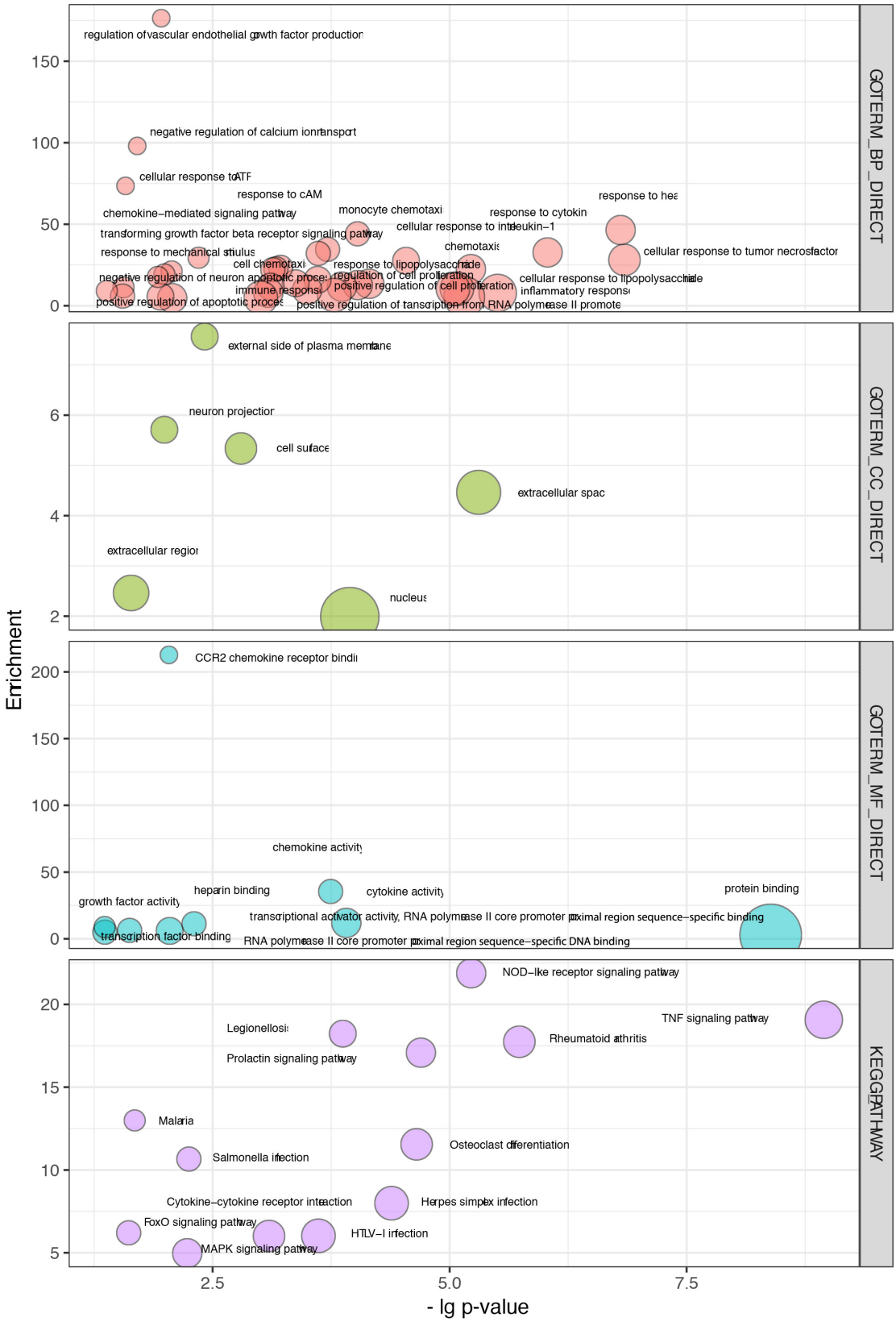


FIGURE 4 | Overlapping terms of GO and KEGG pathway enrichment in WGCNA and PPI modules. The abscissa indicates the *p* values, and the ordinate represents the enrichment scores of those terms.

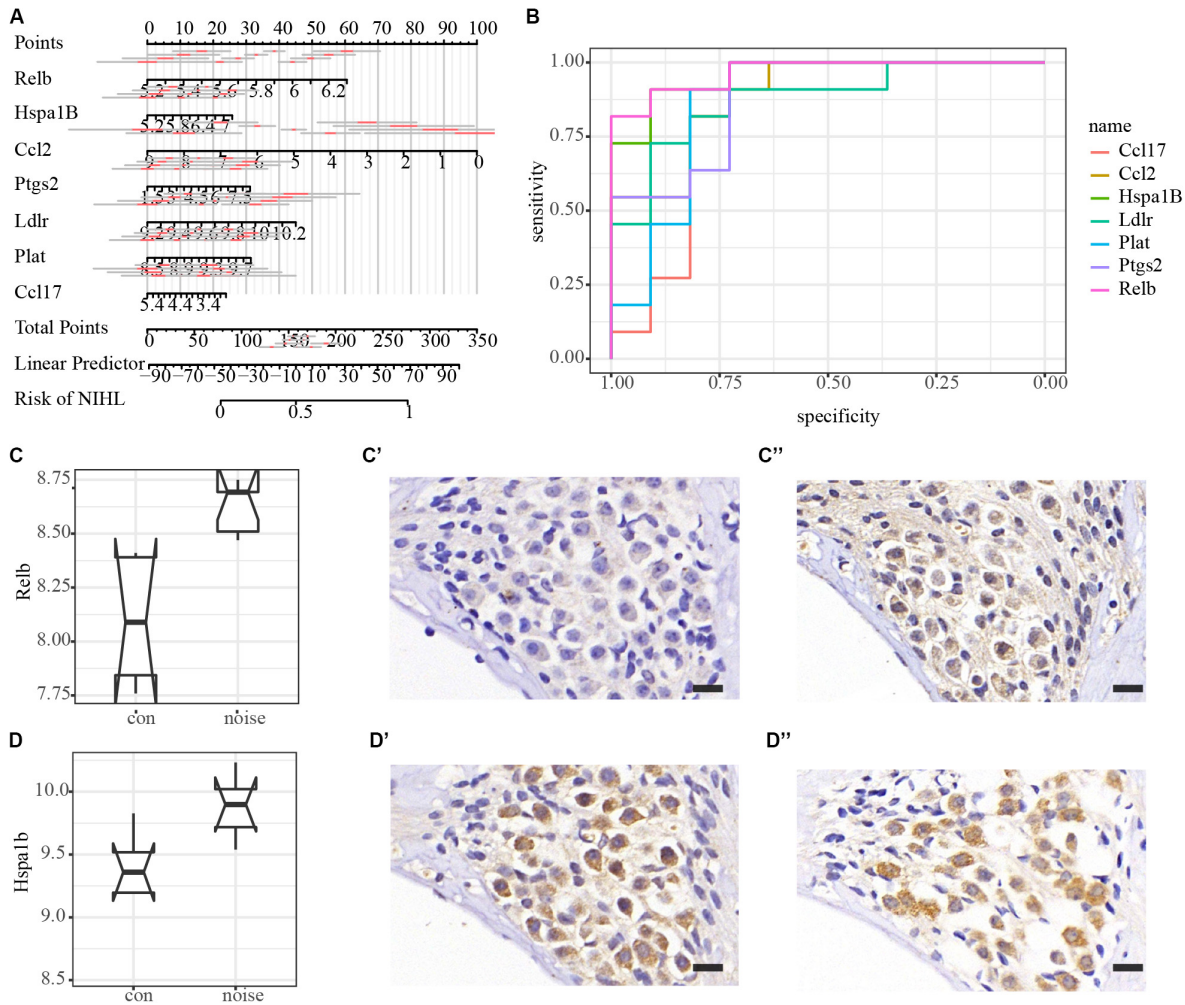


FIGURE 5 | Logistic regression, diagnostic value, and immunohistochemistry validation of hub genes. **(A)** Nomogram plot of the most susceptible molecules ($p < 0.001$) to visualize logistic regression in the PPI module. **(B)** Receiver operating characteristic (ROC) curves of hub genes. AUCs of Relb, Hspa1b, Ccl2, Ptgs2, Ldlr, Plat, and Ccl17 are 0.9669, 0.9587, 0.8926, 0.8843, 0.876, 0.8678, and 0.8347, respectively. **(C–C")** Immunohistochemical validation of Relb in the spiral ganglion area. Scale bar: 15 μm. **(D–D")** Immunohistochemical validation of Hspa1b in the spiral ganglion area. Scale bar: 15 μm.

rescue patients with acute trauma after exposure to gunshots, as significantly lower threshold shifts were observed (Psillas et al., 2008). In recent clinical trials, antioxidants that reduce reactive oxygen after traumatic noise events appear promising (Prasher, 1998). Ebselen has been shown to have promising results for both TTS and PTS by reducing the threshold shift (Lynch and Kil, 2005; Satheeshkumar and Mugesh, 2011; Kil et al., 2017).

Research has indicated that RelB, a member of the NF-kappaB family, is upregulated at the transcription level and results in increases in the lateral wall and the rest area after acoustic overstimulation (Yamamoto et al., 2009). Notably, the expression of adhesion molecules (Icam1, Icam5, and Nrcam) and iNOS was observed in tissues around the capillaries in the stria vascularis, indicating hemodynamics and changes in the cellular integrity (**Supplementary Table S1B**).

After acoustic overstimulation, heat shock protein (HSP) induction was observed in the cochlea (Lim et al., 1993). Unsurprisingly, HSP induction protected against noise trauma in guinea pig cochlea (Mikuriya et al., 2005). In contrast, mice deficient in heat shock factor 1 (the major transcription factor triggering HSP expression) exhibited a decreased TTS recovery ability following noise overstimulation (Fairfield et al., 2005). Notably, HSP70-2 (Hspa1B) polymorphism (rs1061581) was the only HSP70 polymorphism that had a significant association with NIHL in both Chinese and Swedish sample sets (Yang et al., 2006; Konings et al., 2009), which is thought to be evidence for NIHL susceptibility.

The mitochondrial Ptgs2 (COX2) mutation is associated with aminoglycoside antibiotic-induced deafness in a Han Chinese pedigree (Chen et al., 2013). Inhibition of cyclooxygenase-2 (COX2) by NS398 attenuates NIHL in mice (Sun et al., 2016), whereas celecoxib does not (Pourbakht, 2013; Li et al., 2015).

Ccl family members (such as Ccl2 and Ccl17)-associated inflammatory responses may be detrimental to hearing recovery, suggesting their role as potential post-insult therapeutic targets for treatment (Fujioka et al., 2014).

Currently, no clinical treatments exist to reverse the effects of permanent NIHL (Oishi and Schacht, 2011). Limitations of our present study include the following: first, validations for the detected molecules and pathways were still needed, and technologies with higher resolution such as single cell RNA-seq and spatial transcriptome would facilitate our study; second, even if the pathways influence NIHL pathogenesis, it is unclear how to make treatments safer, more convenient, and accessible. However, given that molecular imaging is utilized as a powerful tool in the investigation of parkinsonian disorders, the identification of biomarkers of early changes remains a challenge to predict the clinical trajectory of these disorders (Strafella et al., 2017). Exciting new tracer developments are aiding the investigation of *in vivo* markers in its treatment (Géléoc and Holt, 2014). Similarly, NIHL and even all sensorineural hearing loss will benefit from these technologies. The first step is to identify potential markers, and our current research does so in a comprehensive and systematic way.

In conclusion, our results suggest that genes involved in multiple inflammatory pathways are influenced most significantly after noise exposure. Relb, Hspa1b, Ccl2, Ptgs2, Ldlr, Plat, and Ccl17 are crucial for NIHL pathogenesis in mouse models. While further validation studies are required, these genes could potentially be used as novel diagnostic and therapeutic targets for NIHL.

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 XinHua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine.

AUTHOR CONTRIBUTIONS

MX and QW contributed for initiation and conduction. QW contributed for the writing of the manuscript. QW, YS, BY, HH, and CF performed the experiments. BY, QW, and YS processed and analyzed the data. All authors contributed to the article and approved the submitted version.

FUNDING

The present study was supported by grants from the National Natural Science Foundation of China (grant no. 81670926), the Cultivation Project of the Major Research Plan of the National Natural Science Foundation of China (grant no. 91949119), and Ruijin Youth NSFC Cultivation Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00968/full#supplementary-material

REFERENCES

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., et al. (2014). Auditory and non-auditory effects of noise on health. *Lancet* 383, 1325–1332. doi: 10.1016/S0140-6736(13)61613-X

Cai, Q., Vethanayagam, R. R., Yang, S., Bard, J., Jamison, J., Cartwright, D., et al. (2014). Molecular profile of cochlear immunity in the resident cells of the organ of Corti. *J. Neuroinflammation* 11:173. doi: 10.1186/s12974-014-0173-8

Chen, J.-D., and Tsai, J.-Y. (2003). Hearing loss among workers at an oil refinery in Taiwan. *Arch. Environ. Health* 58, 55–58. doi: 10.3200/AEOH.58.1.55-58

Chen, T., Liu, Q., Jiang, L., Liu, C., and Ou, Q. (2013). Mitochondrial COX2 G7598A mutation may have a modifying role in the phenotypic manifestation of aminoglycoside antibiotic-induced deafness associated with 12S rRNA A1555G mutation in a Han Chinese pedigree. *Genet. Test Mol. Biomark.* 17, 122–130. doi: 10.1089/gtmb.2012.0251

Clark, W. W., and Bohne, B. A. (1999). Effects of noise on hearing. *JAMA* 281, 1658–1659.

Delmaghani, S., Defourny, J., Aghaie, A., Beurg, M., Dulon, D., Thelen, N., et al. (2015). Hypervulnerability to sound exposure through impaired adaptive proliferation of peroxisomes. *Cell* 163, 894–906. doi: 10.1016/j.cell.2015.10.023

Fairfield, D. A., Lomax, M. I., Dootz, G. A., Chen, S., Galecki, A. T., Benjamin, I. J., et al. (2005). Heat shock factor 1-deficient mice exhibit decreased recovery of hearing following noise overstimulation. *J. Neurosci. Res.* 81, 589–596. doi: 10.1002/jnr.20417

Fujioka, M., Okamoto, Y., Shinden, S., Okano, H. J., Okano, H., Ogawa, K., et al. (2014). Pharmacological inhibition of cochlear mitochondrial respiratory chain induces secondary inflammation in the lateral wall: a potential therapeutic target for sensorineural hearing loss. *PLoS One* 9:e90089. doi: 10.1371/journal.pone.0090089

Géléoc, G. S. G., and Holt, J. R. (2014). Sound strategies for hearing restoration. *Science* 344:1241062. doi: 10.1126/science.1241062

Horvath, S., and Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* 4:e1000117. doi: 10.1371/journal.pcbi.1000117

Kil, J., Lobarinas, E., Spankovich, C., Griffiths, S. K., Antonelli, P. J., Lynch, E. D., et al. (2017). Safety and efficacy of ebselen for the prevention of noise-induced hearing loss: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet* 390, 969–979. doi: 10.1016/S0140-6736(17)31791-9

Kil, J., Pierce, C., Tran, H., Gu, R., and Lynch, E. D. (2007). Ebselen treatment reduces noise induced hearing loss via the mimicry and induction of glutathione peroxidase. *Hear. Res.* 226, 44–51. doi: 10.1016/j.heares.2006.08.006

Konings, A., Van Laer, L., Michel, S., Pawelczyk, M., Carlsson, P.-I., Bondeson, M.-L., et al. (2009). Variations in HSP70 genes associated with noise-induced hearing loss in two independent populations. *Eur. J. Hum. Genet.* 17, 329–335. doi: 10.1038/ejhg.2008.172

Kujawa, S. G., and Liberman, M. C. (2009). Adding insult to injury: cochlear nerve degeneration after "Temporary" noise-induced hearing loss. *J. Neurosci.* 29, 14077–14085. doi: 10.1523/JNEUROSCI.2845-09.2009

Kurabi, A., Keithley, E. M., Housley, G. D., Ryan, A. F., and Wong, A. C.-Y. (2017). Cellular mechanisms of noise-induced hearing loss. *Hear. Res.* 349, 129–137. doi: 10.1016/j.heares.2016.11.013

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

Li, B., Su, K., Yang, G., Feng, Y., Xia, L., and Yin, S. (2015). Assessment of the potential ototoxicity of high-dose celecoxib, a selective cyclooxygenase-2 inhibitor, in rats. *Otolaryngol. Head Neck Surg.* 152, 1108–1112. doi: 10.1177/0194599815573702

Lim, H. H., Jenkins, O. H., Myers, M. W., Miller, J. M., and Altschuler, R. A. (1993). Detection of HSP 72 synthesis after acoustic overstimulation in rat cochlea. *Hear. Res.* 69, 146–150.

Lynch, E. D., and Kil, J. (2005). Compounds for the prevention and treatment of noise-induced hearing loss. *Drug Discov. Today* 10, 1291–1298. doi: 10.1016/S1359-6446(05)03561-0

Miao, L., Yin, R.-X., Pan, S.-L., Yang, S., Yang, D.-Z., and Lin, W.-X. (2018). Weighted gene co-expression network analysis identifies specific modules and hub genes related to hyperlipidemia. *Cell. Physiol. Biochem.* 48, 1151–1163. doi: 10.1159/000491982

Mikuriya, T., Sugahara, K., Takemoto, T., Tanaka, K., Takeno, K., Shimogori, H., et al. (2005). Geranylgeranylacetone, a heat shock protein inducer, prevents acoustic injury in the guinea pig. *Brain Res.* 1065, 107–114. doi: 10.1016/j.brainres.2005.10.045

Moon, S. K., Woo, J.-I., Lee, H.-Y., Park, R., Shimada, J., Pan, H., et al. (2007). Toll-like receptor 2-dependent NF-kappaB activation is involved in nontypeable *Haemophilus influenzae*-induced monocyte chemotactic protein 1 up-regulation in the spiral ligament fibrocytes of the inner ear. *Infect. Immun.* 75, 3361–3372. doi: 10.1128/IAI.01886-06

Ohlemiller, K. K., McFadden, S. L., Ding, D. L., Lear, P. M., and Ho, Y. S. (2000). Targeted mutation of the gene for cellular glutathione peroxidase (Gpx1) increases noise-induced hearing loss in mice. *J. Assoc. Res. Otolaryngol.* 1, 243–254.

Ohlemiller, K. K., Wright, J. S., and Dugan, L. L. (1999). Early elevation of cochlear reactive oxygen species following noise exposure. *Audiol. Neurootol.* 4, 229–236. doi: 10.1159/000013846

Oishi, N., and Schacht, J. (2011). Emerging treatments for noise-induced hearing loss. *Expert Opin. Emerg. Drugs* 16, 235–245. doi: 10.1517/14728214.2011.552427

Pourbakht, A. (2013). The effect of celecoxib, a cyclooxygenase-2 inhibitor on noise- induced hearing loss. *Iran J. Basic Med. Sci.* 16, 726–730.

Prasher, D. (1998). New strategies for prevention and treatment of noise-induced hearing loss. *Lancet* 352, 1240–1242. doi: 10.1016/S0140-6736(05)70483-9

Psillas, G., Pavlidis, P., Karvelis, I., Kekes, G., Vital, V., and Constantinidis, J. (2008). Potential efficacy of early treatment of acute acoustic trauma with steroids and piracetam after gunshot noise. *Eur. Arch. Otorhinolaryngol.* 265, 1465–1469. doi: 10.1007/s00405-008-0689-6

Rajguru, R. (2013). Military aircrew and noise-induced hearing loss: prevention and management. *Aviat. Space Environ. Med.* 84, 1268–1276.

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

Satheeshkumar, K., and Mugesh, G. (2011). Synthesis and antioxidant activity of peptide-based ebselen analogues. *Chemistry* 17, 4849–4857. doi: 10.1002/chem.201003417

Sekhar, D. L., Clark, S. J., Davis, M. M., Singer, D. C., and Paul, I. M. (2014). Parental perspectives on adolescent hearing loss risk and prevention. *JAMA Otolaryngol. Head Neck Surg.* 140, 22–28. doi: 10.1001/jamaoto.2013.5760

Seo, M., Lee, Y.-S., and Moon, S.-S. (2016). Association of hearing impairment with insulin resistance, β-cell dysfunction and impaired fasting glucose before onset of diabetes. *Diabet. Med.* 33, 1275–1282. doi: 10.1111/dme.13096

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

Strafella, A. P., Bohnen, N. I., Perlmutter, J. S., Eidelberg, D., Pavese, N., Van Eimeren, T., et al. (2017). Molecular imaging to track Parkinson's disease and atypical parkinsonisms: New imaging frontiers. *Mov. Disord.* 32, 181–192. doi: 10.1002/mds.26907

Suckfuell, M., Canis, M., Strieth, S., Scherer, H., and Haisch, A. (2007). Intratympanic treatment of acute acoustic trauma with a cell-permeable JNK ligand: a prospective randomized phase I/II study. *Acta Otolaryngol.* 127, 938–942. doi: 10.1080/00016480601110212

Sun, Y., Yu, J., Lin, X., and Tang, W. (2016). Inhibition of cyclooxygenase-2 by NS398 attenuates noise-induced hearing loss in mice. *Sci. Rep.* 6:22573. doi: 10.1038/srep22573

Tahera, Y., Meltser, I., Johansson, P., Salman, H., and Canlon, B. (2007). Sound conditioning protects hearing by activating the hypothalamic-pituitary-adrenal axis. *Neurobiol. Dis.* 25, 189–197. doi: 10.1016/j.nbd.2006.09.004

Tornabene, S. V., Sato, K., Pham, L., Billings, P., and Keithley, E. M. (2006). Immune cell recruitment following acoustic trauma. *Hear. Res.* 222, 115–124. doi: 10.1016/j.heares.2006.09.004

Vethanayagam, R. R., Yang, W., Dong, Y., and Hu, B. H. (2016). Toll-like receptor 4 modulates the cochlear immune response to acoustic injury. *Cell Death Dis.* 7:e2245. doi: 10.1038/cddis.2016.156

Wang, Q., Shen, Y., Ye, B., Hu, H., Fan, C., Wang, T., et al. (2018). Gene expression differences between thyroid carcinoma, thyroid adenoma and normal thyroid tissue. *Oncol. Rep.* 40, 3359–3369. doi: 10.3892/or.2018.6717

World Health Organization (2015). *Hearing Loss Due to Recreational Exposure to Loud Sounds: A Review*. Geneva: World Health Organization.

Yamamoto, H., Omelchenko, I., Shi, X., and Nuttall, A. L. (2009). The influence of NF-kappaB signal-transduction pathways on the murine inner ear by acoustic overstimulation. *J. Neurosci. Res.* 87, 1832–1840. doi: 10.1002/jnr.22018

Yamane, H., Nakai, Y., Takayama, M., Iguchi, H., Nakagawa, T., and Kojima, A. (1995). Appearance of free radicals in the guinea pig inner ear after noise-induced acoustic trauma. *Eur. Arch. Otorhinolaryngol.* 252, 504–508.

Yamasoba, T., Harris, C., Shoji, F., Lee, R. J., Nuttall, A. L., and Miller, J. M. (1998). Influence of intense sound exposure on glutathione synthesis in the cochlea. *Brain Res.* 804, 72–78.

Yang, M., Tan, H., Yang, Q., Wang, F., Yao, H., Wei, Q., et al. (2006). Association of hsp70 polymorphisms with risk of noise-induced hearing loss in Chinese automobile workers. *Cell Stress Chaperones* 11, 233–239.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* 3, 32–35.

Yu, F., Hao, S., Yang, B., Zhao, Y., Zhang, R., Zhang, W., et al. (2015). Insulin resistance due to dietary iron overload disrupts inner hair cell ribbon synapse plasticity in male mice. *Neurosci. Lett.* 597, 183–188. doi: 10.1016/j.neulet.2015.04.049

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, Z., Pan, B., Lv, S., Ji, Z., Wu, Q., Lang, R., et al. (2018). Integrating MicroRNA expression profiling studies to systematically evaluate the diagnostic value of MicroRNAs in pancreatic cancer and validate their prognostic significance with the cancer genome atlas data. *Cell. Physiol. Biochem.* 49, 678–695. doi: 10.1159/000493033

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: 15 September 2020
doi: 10.3389/fgene.2020.551495

Systematic Identification of Hub Genes in Placenta Accreta Spectrum Based on Integrated Transcriptomic and Proteomic Analysis

*Bingnan Chen[1,2,3,4], Di Wang[5], Yue Bian[1,2,3,4], Jiapo Li[1,2,3,4], Tian Yang[1,2,3,4], Na Li[1,2,3,4] and Chong Qiao[1,2,3,4]**

[1] Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China, [2] Key Laboratory of Maternal-Fetal Medicine of Liaoning Province, China Medical University, Shenyang, China, [3] Key Laboratory of Obstetrics and Gynecology of Higher Education of Liaoning Province, Shenyang, China, [4] Research Center of China Medical University Birth Cohort, Shenyang, China, [5] Department of Internal Medicine, Shengjing Hospital of China Medical University, Shenyang, China

Placenta accreta spectrum (PAS) is a pathological condition of the placenta with abnormal adhesion or invasion of the placental villi to the uterine wall, which is associated with a variety of adverse maternal and fetal outcomes. Although some PAS-related molecules have been reported, the underlying regulatory mechanism is still unclear. Compared with the study of single gene or pathway, omics study, using advanced sequencing technology and bioinformatics methods, can increase our systematic understanding of diseases. In this study, placenta tissues from 5 patients with PAS and 5 healthy pregnant women were collected for transcriptomic and proteomic sequencing and integrated analysis. A total of 728 messenger RNAs and 439 proteins were found to be significantly different between PAS group and non-PAS group, in which 23 hub genes were differentially expressed in both transcriptome and proteome. Functional enrichment analysis showed that the differentially expressed genes were mainly related to cell proliferation, migration and vascular development. Totally 18 long non-coding RNA were found that might regulate the expression of hub genes. Many kinds of single nucleotide polymorphism, alternative splicing and gene fusion of hub genes were detected. This is the first time to systematically explore the hub genes and gene structure variations of PAS through integrated omics analysis, which provided a genetic basis for further in-depth study on the underlying regulatory mechanism of PAS.

Keywords: placenta accreta spectrum, biomarker, transcriptomics, proteomics, integrated analysis, bioinformatics

OPEN ACCESS

Edited by:
Geng Chen,
East China Normal University, China

Reviewed by:
Yuriy L. Orlov,
First Moscow State Medical University, Russia
Adi Laurentiu Tarca,
Wayne State University, United States

***Correspondence:**
Chong Qiao
qiaochong2002@hotmail.com

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

Received: *13 April 2020*
Accepted: *25 August 2020*
Published: *15 September 2020*

Citation:
Chen B, Wang D, Bian Y, Li J, Yang T, Li N and Qiao C (2020) Systematic Identification of Hub Genes in Placenta Accreta Spectrum Based on Integrated Transcriptomic and Proteomic Analysis. Front. Genet. 11:551495. doi: 10.3389/fgene.2020.551495

INTRODUCTION

Placenta accreta spectrum (PAS) is a pathological condition of the placenta with abnormal adhesion or invasion of the placental villi to the uterine wall (Society of Gynecologic Oncology et al., 2018). There are three types of PAS, including placenta accreta, placenta increta and placenta percreta (Jauniaux et al., 2018b). In recent years, the prevalence of PAS has increased, which may be directly related to the increase in cesarean section rates in most high-income and middle-income countries

(Thurn et al., 2016; Jauniaux et al., 2018a). The prevalence of PAS has increased about eight times since 1970s (Silver and Branch, 2018). A study showed that the overall proportion of PAS in recent years had even reached 0.91% (El Gelany et al., 2019).

Placenta accreta spectrum is associated with a variety of adverse maternal and fetal outcomes, including preterm birth, low birth weight infants, increased perinatal mortality, maternal rupture of the uterus and postpartum hemorrhage, etc. (Kabiri et al., 2014; Vinograd et al., 2015; Farquhar et al., 2017). This makes PAS become one of the important factors that affect the safety and prognosis of mothers and fetuses in perinatal period. The pathogenesis of PAS is mainly attributed to the absence of decidual or basal layer, abnormal maternal vascular remodeling and excessive invasion of extravillous trophoblasts (EVTs) (Tantbirojn et al., 2008). However, the hub genes and underlying mechanism involved in PAS are still poorly understood.

With the development of the next-generation sequencing technology, the bioinformatics analysis has enabled us to understand the full picture of the biological sample under a disease on a multi-omics level (Bani Baker and Nuser, 2019). Because of the presence of post-translational control, the true state of cells or tissues can not be reflected accurately only through transcriptomics research. With the development of proteomics technology, the integrated transcriptomic and proteomic analysis has become a powerful tool to discover the regulation of gene expression (Vogel and Marcotte, 2012). The application of bioinformatics methods allowed many hub genes to be found as potential therapeutic targets in cancer research (Blum et al., 2018; Danaher et al., 2018; Poma et al., 2018). However, there are still few bioinformatics studies on PAS. In this study, transcriptome and proteome sequencing was performed in the placenta tissues of 5 women with PAS and 5 healthy pregnant women. Through the integrated analysis, the hub genes closely related to PAS were screened out, and the gene structure was preliminarily explored. The results provided evidence for the further study on the underlying regulatory mechanism of PAS.

MATERIALS AND METHODS

Study Population

Patients diagnosed with PAS and healthy pregnant women of maternal age and gestational age at delivery matched in a large teaching hospital of north China were included in this study. Studies have shown that most patients with PAS have placenta previa and/or prior cesarean section (Thurn et al., 2016; Jauniaux et al., 2018a). Considering the interference of these factors to the study, the inclusion criteria for this study were as follows: placenta previa, prior cesarean section, cesarean delivery this time, singleton. And the exclusion criteria were: other uterine cavity operation or uterus-related diseases (such as uterus bicornis and adenomyosis), obstetric complications (such as hypertensive disorder complicating pregnancy, gestational diabetes), systemic disease. Patients with preterm birth were not excluded from this study because patients with PAS are at significantly increased risk of preterm birth (Vinograd et al., 2015). The diagnosis of PAS was based on the International Federation of Gynecology and Obstetrics (FIGO). According to the clinical and histologic criteria in FIGO, PAS was divided into placenta creta, increta and precreta (Jauniaux et al., 2019). Placenta previa was diagnosed by ultrasound and further divided into complete placenta previa (placenta completely covered the cervical internal os) and incomplete placenta previa (placenta margin reached or partially covered the cervical internal os) (Orbach-Zinger et al., 2018). Diagnosis of other diseases was based on the International Classification of Diseases, 11th edition (ICD-11)[1]. The study was approved by the Medical Ethics Committee of China Medical University. All participants signed written informed consent.

Collection of Placenta Tissues

A piece of placenta tissue on maternal side for each patient was taken immediately after cesarean section. The placental tissues of patients with PAS were taken at the place of accreta, containing villi, decidua and myometrial fibers. And the placental tissues of patients without PAS were taken at the place which was 1/2 length of the placenta radius from the umbilical cord. Since no PAS cases required hysterectomy in this study, the placenta tissue that deeply invaded the uterus was not obtained. The collected placenta tissues were rinsed with sterile saline. Tissues were then frozen in liquid nitrogen and stored at −80°C until processing transcriptomic and proteomic analysis.

RNA-Seq

Total RNA was extracted from placenta samples using Trizol (Invitrogen, Carlsbad, CA, United States) according to the manual instructions. For placenta samples, grind about 60 mg with liquid nitrogen into powder and transfer the powder samples into the 2 ml tube contains 1.5 ml Trizol reagent. The mix was centrifuged at 12000 × *g* for 5 min at 4°C. The supernatant was transferred to a new 2.0 ml tube which was added 0.3 ml of Chloroform/isoamyl alcohol (24:1) per 1.5 ml of Trizol reagent. After the mix was centrifuged at 12000 × *g* for 10 min at 4°C, the aqueous phase was transferred to a new 1.5 ml tube in which an equal volume of supernatant of isopropyl alcohol was added. The mix was centrifuged at 12000 × *g* for 20 min at 4°C and then removed the supernatant. After washed with 1 ml 75% ethanol, the RNA pellet was air-dried in the biosafety cabinet and then dissolved by add 25∼100 μL of DEPC-treated water. Subsequently, total RNA was qualified and quantified using a NanoDrop and Agilent 2100 Bioanalyzer (Thermo Fisher Scientific, MA, United States).

Approximately 1 μg total RNA per sample was treated with Ribo-Zero™ Magnetic Kit (Epicentre) to deplete rRNA. The retrieved RNA was fragmented by adding First Strand Master Mix (Invitrogen). First-strand cDNA was generated using random primers reverse transcription, followed by a second-strand cDNA synthesis. The synthesized cDNA was subjected to end-repair and then was 3′ adenylated. Adapters were ligated to the ends of these 3′ adenylated cDNA fragments. Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix are performed to enrich the cDNA fragments. Then the

[1]http://www.who.int/classifications/icd

PCR products are purified with Ampure XP Beads. The final library was quality and quantitated in two methods: check the distribution of the fragments size using the Agilent 2100 Bioanalyzer, and quantify the library using real-time quantitative PCR (QPCR) (TaqMan Probe). The Qualified libraries were sequenced pair end on the Hiseq 4000 or Hiseq X-ten platform (BGI-Shenzhen, China).

Proteomic Analysis

A 5 mm magnetic bead and an appropriate amount of Lysis Buffer 3 were added into a 1.5 ml tube containing the placenta sample. PMSF (final concentration: 1 mM) and EDTA (final concentration: 2 mM) were also added into the tube. The tube was then vortexed. After let stand for 5 min, DTT (final concentration: 10 mM) was added into the tube. The mix was shaken in a tissue grinder for 2 min (power = 50 HZ, Time = 120 s). After centrifuged at 25000 × *g* for 20 min at 4°C, the supernatant was transferred to a new tube which was added DTT (final concentration: 10 mM) again at 56°C water bath for 1 h. After returning to room temperature, IAM (final concentration: 55 mM) was added in a dark room and let stand for 45 min. Then cold acetone was added and let stand at −20°C for 2 h. This step was repeated until the supernatant was colorless. After centrifuged at 25000 × *g* for 20 min, a 5 mm magnetic bead and an appropriate amount of Lysis Buffer 3 were added into the precipitate. The mix was shaken in a tissue grinder for 2 min (power = 50 HZ, Time = 120 s). Finally the mix was centrifuged at 25000 × *g* for 20 min at 4°C, and the supernatant was taken for quantification. Bradford quantitative and SDS-PAGE were used for quality control of protein extraction.

Totally 2.5 μg trypsin enzyme was added into 100 μg protein solution for each sample (protein: enzyme = 40: 1) at 37°C for 4 h. Then the trypsin was added again according to the above ratio at 37°C for 8 h. Peptides after enzymolysis were desalted using Strata X column and vacuum dried. High pH reversed-phase separation was carried out by LC-20AB liquid system (Shimadzu). Peptides were measured by Liquid chromatography-MS/MS (LC-MS/MS) on an UltiMate 3000 UHPLC (Thermo Fisher Scientific). Q-Exactive HF (Thermo Fisher Scientific) was used for data-dependent acquisition (DDA) mode detection and data-independent acquisition (DIA) mode detection. The obtained DDA data were identified using the integrated Andromeda engine of MaxQuant (Cox and Mann, 2008) 1.5.3.30 (false discovery rate, FDR ≤ 1%). According to the results, Spectronaut (Bruderer et al., 2015) was used to construct a spectral library, of which the information was used to complete the deconvolution and extraction of DIA data. Quality control was carried out by mProphet algorithm. Finally significant quantitative results were obtained based on the Target-decoy model applicable to SWATH-MS (FDR ≤ 1%).

Bioinformatics Analysis

The R package "edgeR" was used to identify differentially expressed messenger RNAs (mRNAs) and proteins in the samples. String (version 11.0) (Szklarczyk et al., 2019) and Metascape (Zhou et al., 2019) were used to assess the function of the differentially expressed mRNAs and proteins according to the gene ontology (GO) (Ashburner et al., 2000), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), and the Reactome pathway database (Jassal et al., 2020). Gene-sets enrichment analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) was used to analyze the KEGG pathways significantly correlated with concordant and discordant mRNA-protein expressions. String (version 11.0), Metascape and Cytoscape (Shannon et al., 2003) were used to construct visual interaction networks of differentially expressed mRNAs and proteins. Soapfuse (version 1.18) (Jia et al., 2013) was used to detect the fusion gene in each sample. The Genome Analysis Toolkit (GATK, version 3.4-0) (McKenna et al., 2010) was used to detect single nucleotide polymorphisms (SNP) information in each sample, which was further annotated by Snpeff (Cingolani et al., 2012). ASprofile[2] was used to quantitatively detect splicing events of each sample. The data were analyzed in SPSS (version 25.0, IBM) and R (version 3.6.2).

RESULTS

A total of 10 placenta tissues (5 PAS patients: PAS1, PAS2, PAS3, PAS4, PAS5; 5 non-PAS patients: NPAS1, NPAS2, NPAS3, NPAS4, NPAS5) were sequenced for transcriptomic and proteomic analysis. There were no significant differences between PAS group and non-PAS group in maternal age and gestational age at delivery ($p > 0.05$, see **Table 1** for detailed information of pregnant women).

Transcriptomic Profiling of Placenta Tissue

A total of 17,860 known mRNAs in placenta tissue were detected quantitatively. And 728 differentially expressed mRNAs (false discovery rate, FDR < 0.05, fold change >1.5) were obtained, including 481 up-regulated genes and 247 down-regulated genes (**Figures 1A,B**). Intra-group correlation was good in both PAS group and non-PAS group (**Figure 1C**). GO enrichment analysis in String showed that the differentially expressed mRNAs were mainly related to response to stimulus, vascular development and protein binding (**Figure 2A**). In addition, the biological processes related to cell proliferation and differentiation were also enriched (see **Supplementary Table S1** for detailed results of GO analysis). KEGG enrichment analysis showed that differently expressed mRNAs were related to cancer and cell adhesion. PI3K/AKT signaling pathway, HIF-1 signaling pathway, notch signaling pathway and other classical pathways were enriched (**Figure 2B**, see **Supplementary Table S2** for detailed results of KEGG analysis). In order to further explore the functional interaction of differentially expressed mRNAs, we constructed the interaction network with String, and isolated the tightly interacting gene clusters through MCODE in Cytoscape. Totally 26 gene clusters were obtained and three main clusters that were related to cell proliferation, differentiation and vascular development were shown in **Figure 3**.

[2]http://ccb.jhu.edu/software/ASprofile

TABLE 1 | Detailed information of participants.

Patient ID	Age	Height (centimeter, cm)	Weight before delivery (kilogram, kg)	Previous caesarian section times	Type of placenta previa	Intraoperative blood transfusion	Gestational age at delivery (week)	Type of placenta accreta spectrum
PAS1	36	167	81	1	Complete	No	35.00	Increta
PAS2	28	160	64	1	Complete	Yes	36.56	Increta
PAS3	24	160	80	1	Complete	Yes	36.00	Increta
PAS4	37	160	80	1	Complete	Yes	36.00	Increta
PAS5	28	162	71.5	1	Complete	No	36.42	Increta
NPAS1	40	168	55	1	Complete	No	37.56	No
NPAS2	36	168	90	1	Complete	No	37.28	No
NPAS3	28	165	77	1	Complete	No	35.42	No
NPAS4	36	160	80	1	Complete	No	37.14	No
NPAS5	33	168	70	1	Complete	No	36.28	No

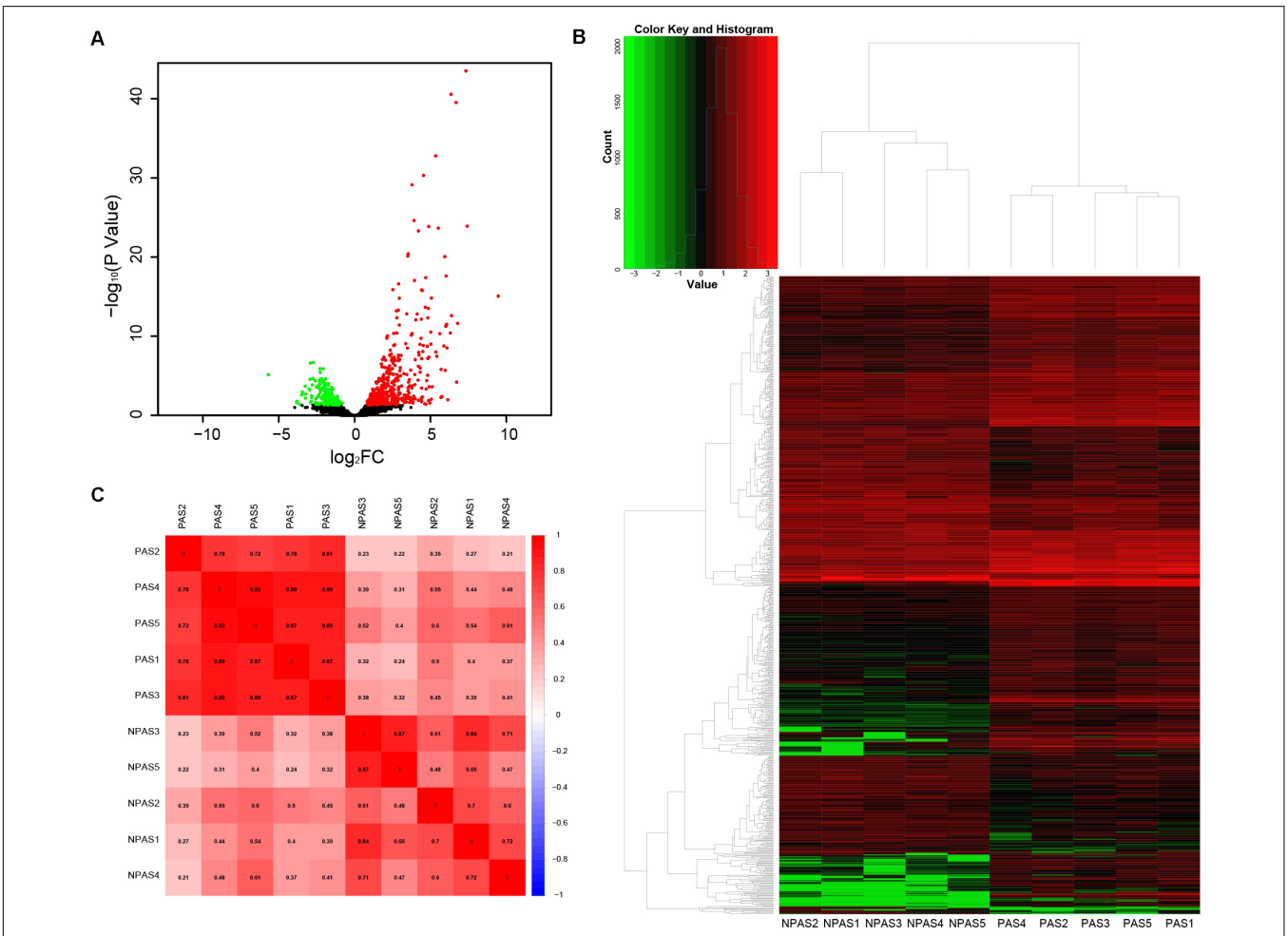


FIGURE 1 | Analysis of differentially expressed mRNAs. **(A)** Volcano plot, **(B)** heat map, and **(C)** correlation between PAS group and non-PAS group.

Proteomic Profiling of Placenta Tissue

A total of 4,800 known proteins were detected quantitatively. In the same way, we obtained 439 differentially expressed proteins (DEPs) (FDR < 0.05, fold change >1.5). The GO enrichment analysis showed that DEPs were mainly associated with cellular component organization or biogenesis and intracellular organelle (**Figure 2C**). The biological processes related to cell migration, differentiation and tube development were also enriched (see **Supplementary Table S3** for detailed results of GO analysis). In KEGG enrichment analysis, DEPs were only associated with metabolic pathways (**Figure 2D**, see **Supplementary Table S4** for detailed results of KEGG analysis). Similar to the differentially

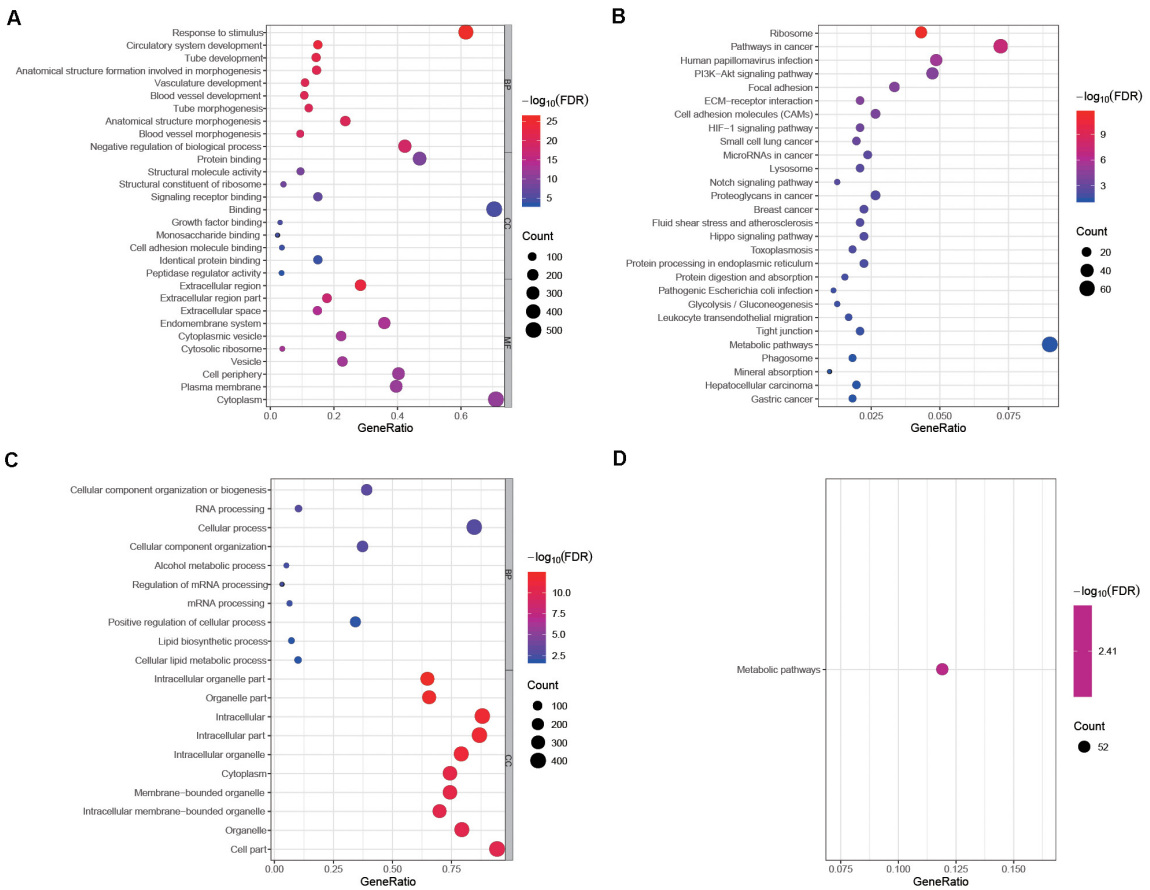


FIGURE 2 | Enrichment analyses of differentially expressed mRNAs and DEPs. The top 10 biological process GO terms, cellular component GO terms, molecular function GO terms of **(A)** differentially expressed mRNAs and **(B)** DEPs sorted by false discovery rate (FDR) value. Significantly enriched KEGG pathways of **(C)** differentially expressed mRNAs and **(D)** DEPs. Some functions and pathways related to cell proliferation, adhesion and angiogenesis are not shown in this figure because the corresponding FDR values are not in the top 10. No molecular function GO term was enriched in the GO analysis of DEPs, so this part is not shown in **(B)**.

expressed genes, all the DEPs were constructed into a visual interaction network, and 20 gene clusters were obtained by MCODE analysis. The three main clusters related to cell proliferation, differentiation and vascular development were shown in **Figure 4**.

Integrated Analysis of Transcriptome and Proteome

Overall Correlation and Functional Enrichment

First, we evaluated the correlation between mRNA expression and protein abundance in 10 samples. Specifically, we aggregated genes that were quantitatively detected in both the transcriptome and proteome and calculated the Pearson correlation coefficient between mRNA and protein expression levels of each gene in 10 samples. And 62.3% of the genes showed positive correlations (**Supplementary Figure S1**). Furthermore, GSEA was performed with the Pearson correlation coefficient as the rank of each gene in order to investigated whether the trend of mRNA-protein correlation was associated with specific KEGG pathways. As a result, we identified several pathways involved in cell proliferation, cell adhesion, vascular development, immunity, and metabolism (**Figure 5A**), which were all related to PAS.

Identification of Hub Genes and Related Functions

Subsequently, in order to explore the hub genes in PAS, we took the intersection of differentially expressed mRNAs and DEPs, and obtained a total of 31 co-differentially expressed genes (**Figure 5B**), among which the expression trend of 23 genes was consistent in transcriptome and proteome (**Table 2**). Pearson correlation analysis showed that the expression levels of these 31 genes in the transcriptome and proteome had a strong positive correlation ($R = 0.63$, $p < 0.001$, **Figure 5C**). Given the small sample size, we also ran a stochastic simulation to verify the reliability of these genes. In the simulation, 728 fake differentially expressed mRNAs and 439 fake DEPs were randomly selected from 17,860 mRNAs and 4,800 proteins, respectively, and the number of genes shared by both selected mRNAs and proteins was counted. Based on R (version 3.6.2), we performed the simulation 10,000 times and found that there was only a probability of 0.0019 that the number of shared genes

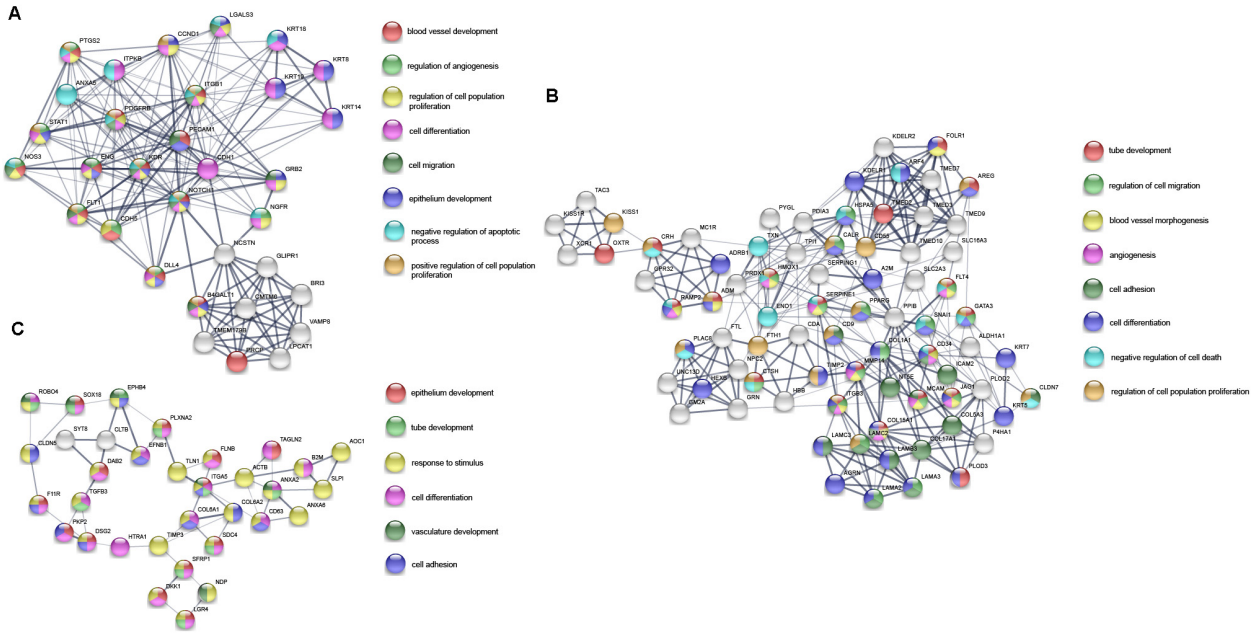


FIGURE 3 | Three main gene clusters **(A–C)** related to cell proliferation, differentiation and vascular development in PPI network of differentially expressed mRNAs. The color of the nodes represents the participating function. The thickness of the edges indicates the strength of interaction evidence.

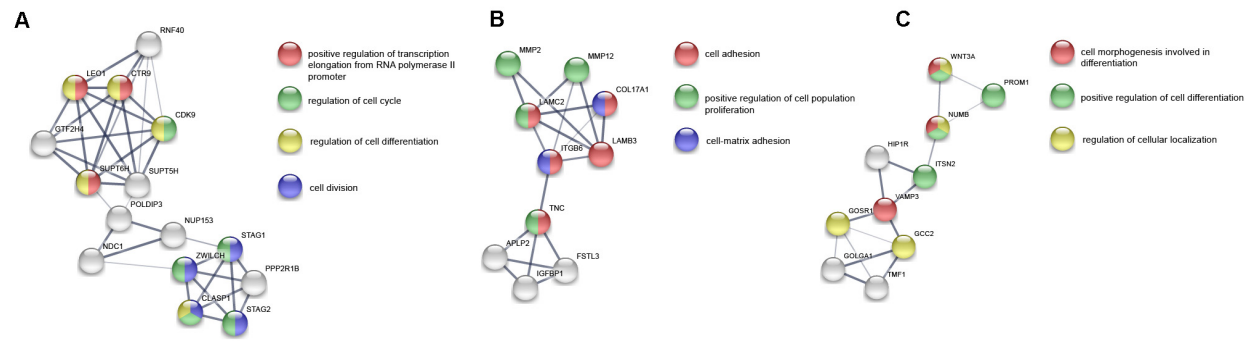


FIGURE 4 | Three main gene clusters **(A–C)** related to cell proliferation, differentiation and vascular development in PPI network of DEPs. The color of the nodes represents the participating function. The thickness of the edges indicates the strength of interaction evidence.

was greater than or equal to 31, which further illustrated the reliability of the results.

The functional enrichment analysis of the 23 hub genes with consistent expression trends was then carried out by Metascape, in which 9 function clusters including functions and pathways related to cell proliferation, adhesion and vascular development were enriched (**Supplementary Table S5**). The connection between each hub gene and 9 clusters is shown in **Figures 5D,E**. This was highly consistent with the above enrichment results in String and GSEA, which suggested that these 23 hub genes were likely to play an important role in the pathogenesis of PAS.

Regulation and Structure Variation of Hub Genes

In order to explore which long non-coding RNA (lncRNA) might regulate the expression of these 23 hub genes, we extracted the sequence results of lncRNA from placenta tissue and predicted the target genes. Totally 18 lncRNAs (including known and novel lncRNAs) might be involved in the regulation of some hub genes (**Figure 6**). The sequence of novel lncRNAs were shown in **Supplementary Table S6**. Pearson and Spearman correlation coefficients between lncRNA and mRNA were shown in **Supplementary Table S7**. In addition, we performed a series of analyses of hub genes in gene structure level, including SNP, alternative splicing and gene fusion. In terms of SNP, the presence of one or more types of SNP was found in most hub genes (**Figure 7**). The intron variation of *ADGRA2* (4 of 5 cases in NPAS group, $p = 0.048$, Fisher's exact test, same for the following), the intergenic variation of *FSTL3* (4 of 5 cases in PAS group, $p = 0.048$) and the synonymous variation of *LAMB3* (4 of 5 cases in PAS group, $p = 0.048$) were significantly correlated with PAS.

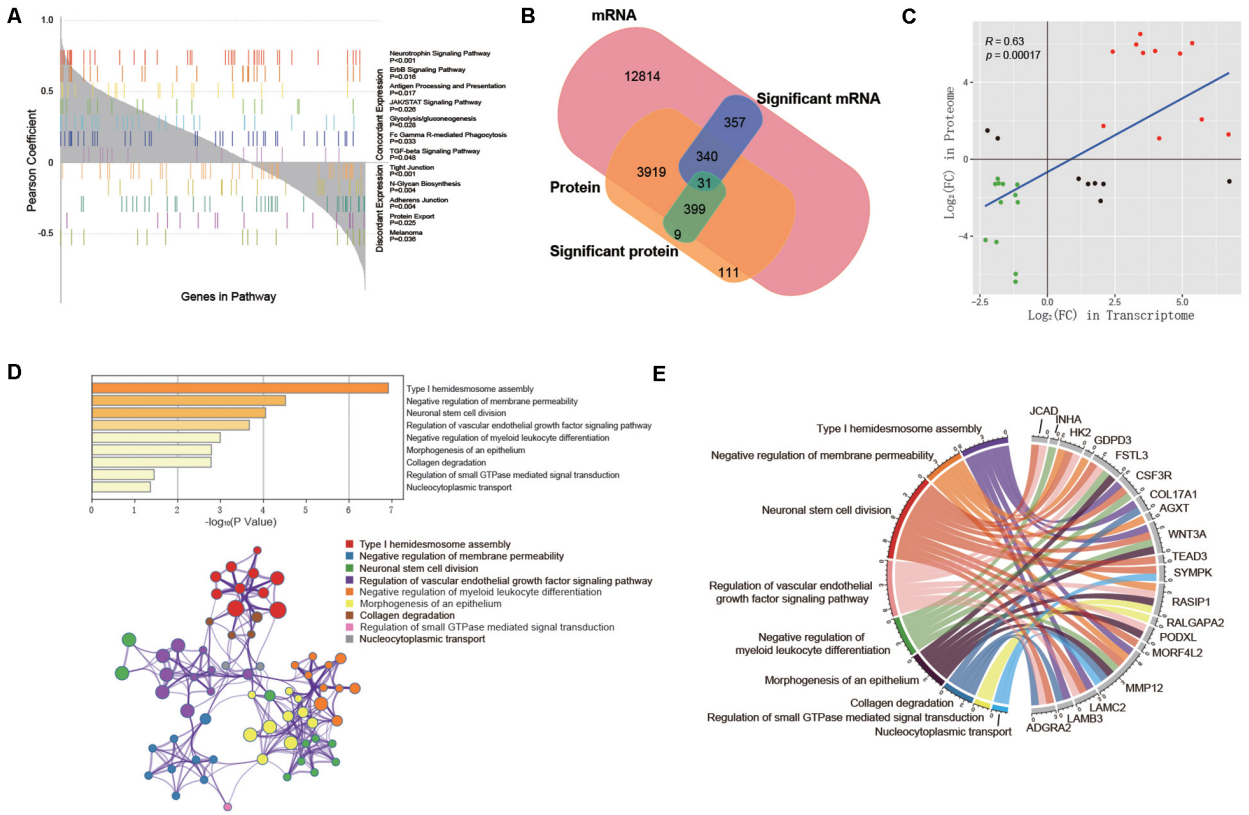


FIGURE 5 | Integrative analysis between transcriptome and proteome. **(A)** KEGG pathways significantly correlated with concordant or discordant mRNA-protein expressions. Genes are aligned along the x axis by the rank of their Pearson correlation coefficient between mRNA and protein expression levels. Each color represents one significantly associated pathway, and each bar represents one gene in the pathway. **(B)** Venn diagram showing all identified mRNAs and proteins and their overlap. **(C)** The expression levels of 31 co-differentially expressed genes in the transcriptome and proteome. **(D)** Functional enrichment analysis based on Metascape. Bar graph demonstrates *P*-value of enriched clusters. Each node in network represents an enriched term of corresponding clusters. For large clusters, the network only displays the top 10 terms with *P*-value. The edge reflects the connection of functions and pathways. The size of the nodes reflects the number of genes, and the thickness of the edges reflects correlation level of terms **(E)** Chord diagram shows the hub genes with their representative enriched clusters.

In the detection of alternative splicing events for all genes, 7 types of alternative splicing events, mainly alternative 5′ first exon (TSS) and alternative 3′ last exon (TTS), were found (**Figure 8A**). Four types of alternative splicing events were detected in the hub genes, among which the alternative splicing of *AGXT*, *MMP12*, and *TEAD3* were only detected in the PAS group (**Figure 8B**). The expression of the hub genes in alternative splicing events was analyzed by using the Mann-Whitney *U* test. The results showed that TSS of *LAMB3* at chromosome 1 sites 209652369 to 209652475 ($p = 0.032$) and TTS of *HK2* at chromosome 2 sites 74890797 to 74893354 ($p = 0.032$) were significantly associated with PAS. Finally, there was a gene fusion between *INHA* and *STK11IP* on chromosome 2, which was detected in PAS1, PAS2, and PAS5 (**Figure 9**).

DISCUSSION

In this study, we sequenced the placenta tissues of 5 women with PAS and 5 healthy pregnant women. And the results were analyzed systematically. Totally 23 hub genes which are closely related to PAS were screened out. As expected, most of these genes are directly related to cell proliferation and migration, as well as vascular development. There are also some hub genes related to trophoblast fusion, energy metabolism and DNA repair. Further, we found lncRNAs that might regulate the expression of hub genes and analyzed the hub genes at the gene structure level.

Totally 17,860 mRNAs and 4,800 proteins were detected quantitatively in this study. Although there was a strong positive correlation between the transcriptome and the proteome, only 23 genes were differentially expressed at both mRNA and protein levels and the expression trends were the same. This suggested that these genes might play a key role in PAS. And on the other hand, it also reflected the importance of post-transcriptional regulation in gene expression. The asynchronous expression of mRNAs and corresponding proteins has been reported in many studies (Dai et al., 2018; Wang D.et al., 2019). Therefore integrated analysis of transcriptomics and proteomics is of great benefit to the discovery of hub genes and underlying mechanisms (Maier et al., 2009).

TABLE 2 | Detailed expression information of hub genes in transcriptome and proteome.

Hub genes	log_2(FC) in transcriptome	FDR in transcriptome	log_2(FC) in proteome	FDR in proteome
COL17A1	6.693	< 0.001	1.288	0.003
MMP12	5.701	< 0.001	2.064	< 0.001
FSTL3	5.342	< 0.001	6.037	< 0.001
AGXT	4.899	< 0.001	5.501	< 0.001
INHA	4.123	< 0.001	1.094	0.035
LAMC2	3.975	< 0.001	5.614	< 0.001
HK2	3.525	< 0.001	5.534	< 0.001
LAMB3	3.423	0.005	6.519	< 0.001
APOBR	3.275	0.003	5.973	< 0.001
MORF4L2	2.403	< 0.001	5.581	< 0.001
GDPD3	2.055	0.001	1.739	0.006
HID1	−2.308	< 0.001	−4.205	0.012
CSF3R	−1.949	0.003	−1.287	0.001
WNT3A	−1.904	0.021	−4.3	0.007
RASIP1	−1.856	0.017	−1.034	0.030
PODXL	−1.819	0.049	−1.273	0.001
ADGRA2	−1.741	0.001	−2.249	< 0.001
RALGAPA2	−1.662	0.031	−1.303	0.002
ERVW-1	−1.206	0.033	−1.87	< 0.001
JCAD	−1.196	0.028	−6.379	< 0.001
SYMPK	−1.19	0.011	−5.956	< 0.001
TEAD3	−1.131	0.008	−1.326	0.001
INTS3	−1.116	0.023	−2.236	< 0.001

FC, fold change; FDR, false discovery rate.

Excessive invasion of extravillous trophoblasts (EVTs) is essential in the development of PAS. At this point, EVTs have a strong ability of proliferation and migration (Matsukawa et al., 2019). The functions and pathways related to cell proliferation and migration (cell adhesion) were enriched in all functional enrichment analyses in this study. Among the up-regulated hub genes, *MMP12* has been reported to be an important molecule in the process of trophoblast invasion and vascular remodeling under hypoxia (Chakraborty et al., 2016), and *FSTL3* can promote trophoblast proliferation, invasion, and lipid storage and inhibit trophoblast apoptosis (Xie et al., 2018). The functional enrichment analysis indicated that *GDPD3*, *LAMB3*, and *LAMC2* were involved in the PI3K/AKT signaling pathway to promote cell proliferation and migration, which was confirmed in the related reports (Rao et al., 2017; Wang Q.et al., 2019; Zhang H. et al., 2019). *MORF4L2* and *COL17A1* could also promote cell proliferation and migration (Shadeo et al., 2008; Thangavelu et al., 2016). In addition, up-regulated expression of *HK2* can inhibit apoptosis (Xu et al., 2017). Among the down-regulated hub genes, *TEAD3* was reported to be down-regulated during trophoblast implantation into the endometrium (Bai et al., 2018); *HID1*, *WNT3A*, and *SYMPK* were also associated with cell proliferation, invasion and adhesion (Sonderegger et al., 2010; Chang et al., 2012; Aydin and Arga, 2019). In general, these genes played a role in reducing cell adhesion, promoting cell proliferation and migration, and inhibiting cell apoptosis, which is very important in the EVTs invasion into the myometrium. It should also be noted that both *PODXL* and *CSF3R* have been reported to promote cell proliferation or migration (Yeo et al., 2018; Zhang J. et al., 2019), but their expression in the PAS group was down-regulated in this study. Co-expression of enhancement signals and suppression signals in maternal regulation of placental cells were found in order to maintain the homeostasis (Pavlicev et al., 2017). Increased proliferation and migration of EVTs might trigger some negative feedback mechanism that led to low expression of *PODXL* and *CSF3R*.

Another pathological change accompanied by EVTs invasion in PAS is the remodeling of maternal blood vessels (Tantbirojn et al., 2008). The functional enrichment analysis in various parts of this study also enriched multiple functions and pathways related to vascular development, including the PI3K/AKT signaling pathway that was reported to be related to PAS (Long et al., 2020). The hub genes related to vascular development included *HK2*, *RASIP1*, *ADGRA2*, and *JCAD*. It has been reported that HK2 promoted aerobic glycolysis and activated p38-MAPK signal conduction in angiogenesis of melanoma (Lu et al., 2019). *HK2* was highly expressed in the PAS group and might be involved in maternal vascular remodeling. The remaining three genes, including *RASIP1*, *ADGRA2,* and *JCAD* were low expressed in the PAS group. *RASIP1* increased the stability of endothelium connection (Koo et al., 2016), which was not conducive to vascular remodeling; *ADGRA2* and *JCAD* were also found to be related to the inhibition of vascular development (Vallon et al., 2010; Hara et al., 2017). These molecules may act synergistically to promote the formation of blood vessels.

It is worth mentioning that some other hub genes and functions might also be involved in the pathogenesis of PAS. *ERVW-1* and *INHA* were the hub genes in this study. It has been reported that *ERVW-1* induced the fusion of trophoblast cells and reduced the proliferation and migration ability of trophoblast cells (Dunk et al., 2018), while another study showed that the expression level of *INHA* in cytotrophoblast cells was higher than that in syncytiotrophoblast cells (Azar et al., 2018). We found that *INHA* was up-regulated and *ERVW-1* was down-regulated in the PAS group. Therefore, we thought that the fusion of trophoblast cells was probably inhibited by enhanced expression of *INHA* and inhibited expression of *ERVW-1*, thereby ensuring the strong proliferation and migration ability of trophoblast cells. In the KEGG pathways significantly correlated with concordant and discordant mRNA-protein expressions, the glycolysis/gluconeogenesis pathway was found to be significantly associated with PAS. *HK2*, one of the hub genes involved in this pathway, was highly expressed in the PAS group. *HK2* depletion was shown to inhibit glycolysis in hepatocellular carcinoma and increased cell death (Xu et al., 2017). Similarly, the high expression of *HK2* in the PAS group might promote glycolysis and maintain energy supply during EVTs invasion and vascular remodeling. Among the hub genes, *APOBR* and *AGXT* are, respectively, related to lipid metabolism and alanine metabolism (Brown et al., 2000;

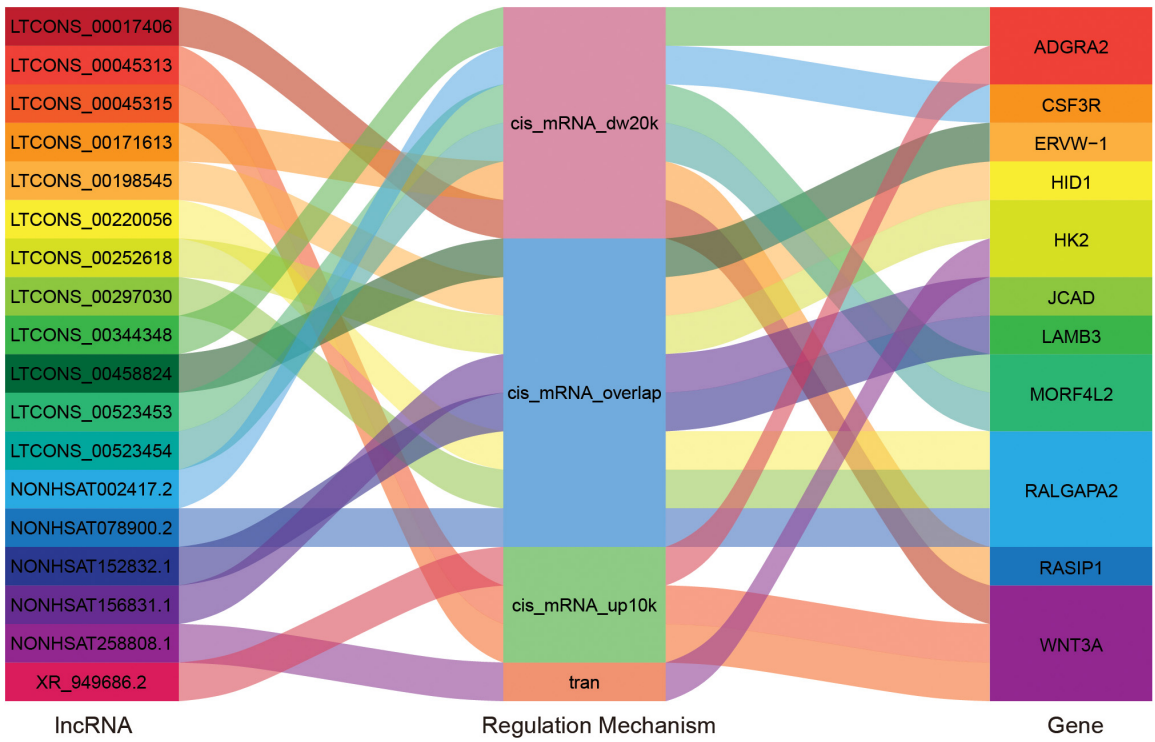


FIGURE 6 | "Sankey" diagram for the correlation between lncRNA and hub genes. The middle column reflects regulation mechanism. cis_mRNA_up10k: Cis action, in which lncRNA is within 10 k upstream of mRNA; cis_mRNA_dw20k: Cis action, in which lncRNA is within 20 k downstream of mRNA; cis_mRNA_overlap: Cis action, in which lncRNA overlaps with mRNA; tran: Tran action.

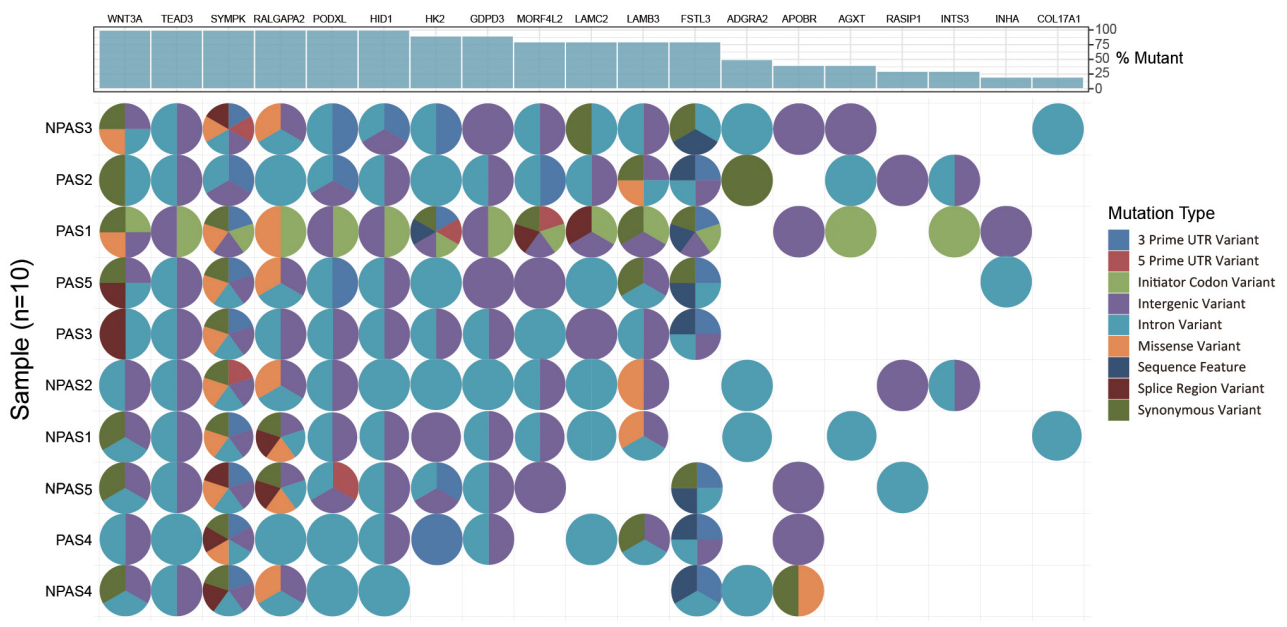


FIGURE 7 | Waterfall plot demonstrates the types of SNP in hub genes.

Montioli et al., 2012), both of which were up-regulated in the PAS group and might also participate in the energy metabolism process in PAS. In terms of immunity, dNK cells was reported to regulate the invasion ability of EVTs by secreting colony-stimulating factors (CSFs) (Vento-Tormo et al., 2018). *CSF3R* expression was low in the PAS group, which might be related

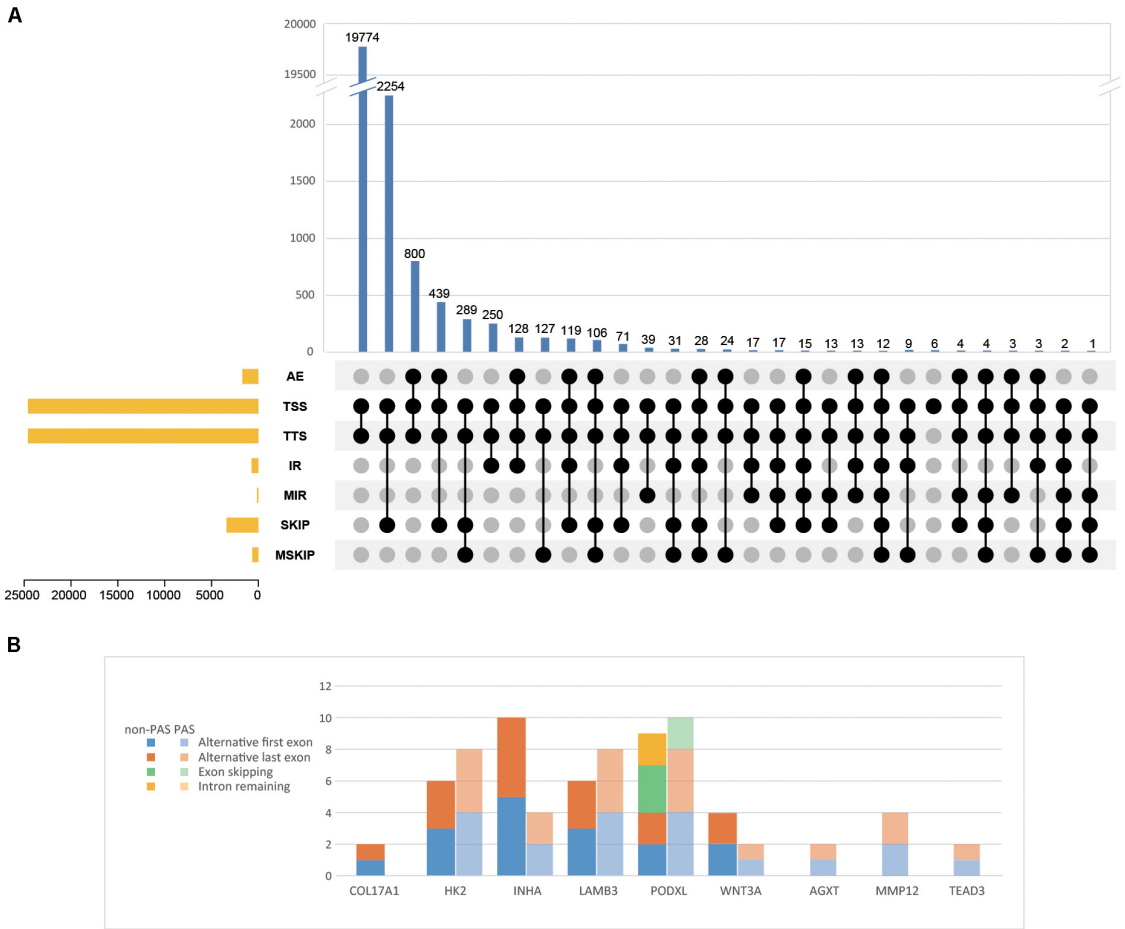


FIGURE 8 | Alternative splicing events of **(A)** all identified genes and **(B)** hub genes. AE, alternative exon ends; TSS, alternative 5′ first exon; TTS, alternative 3′ last exon; IR, intron retention; MIR, multi-IR; SKIP, skipped exon; MSKIP, multi-exon SKIP.

to the regulation of EVTs by dNK cells. In addition, *INTS3*, a gene widely involved in DNA repair (Yang et al., 2015), was found to be downregulated in the PAS group, which might cause mutations in some genes and RNA splicing (Yoshimi et al., 2019). We speculated that the SNP and alternative splicing events, which were found in this study and significantly related to PAS, might be related to the low expression of INTS3.

In this study, multiple SNP and alternative splicing events of hub genes were found. Intron variation of *ADGRA2*, intergene variation of *FSTL3* and synonymous variation of *LAMB3* were significantly related to PAS ($p < 0.05$). SNP of *ADGRA2* was reported to be related to angiogenesis of embryonic central nervous system (Weinsheimer et al., 2012), while SNP of *FSTL3* and *LAMB3* have not been reported. However, both SNP and alternative splicing events of *LAMB3* were significantly related to PAS in this study. In some diseases, synonymous mutations can affect splicing events and cause phenotypic changes (Liu et al., 2018; Sharma et al., 2019). It is worth exploring whether there is a similar mechanism for *LAMB3* in PAS. In addition, the alternative splicing of *HK2* was also significantly related to PAS. It has been reported that *HK2* can be amplified into two different fragments through alternative splicing and influence the invasion and metastasis of prostate cancer (Slawin et al., 2000). Besides, we also found the gene fusion between *INHA* and *STK11IP*. Gene fusion is one of the main causes of tumor formation, which is often used as a tumor diagnosis and prognostic marker (Mertens et al., 2015). The gene fusion between *INHA* and *STK11IP* only occurred in 3 of 5 PAS placenta tissues, so it might not be used as an accurate genetic marker of PAS, but whether it has the potential to assist diagnosis remains to be studied. Similarly, although various SNP and alternative splicing events were found to be significantly related to PAS, considering the small sample size in this study, more samples with sequencing data are needed to verify these variations.

Although some PAS-related molecules and mechanisms have been reported in other studies, this is the first time that the integrated analysis of transcriptomics and proteomics has been performed to systematically identify hub genes and gene structure variations involved in the pathological process of PAS. In addition, lncRNAs that may regulate hub genes were identified

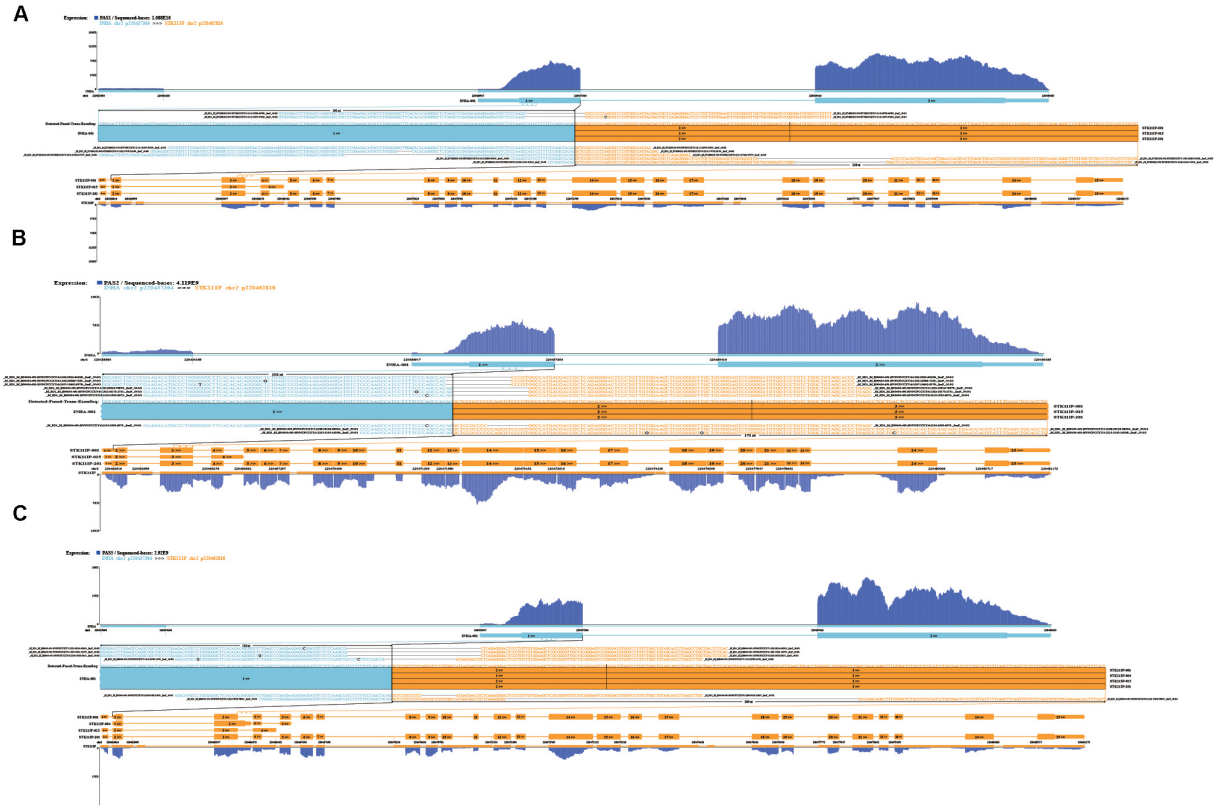


FIGURE 9 | Gene fusion between INHA and STK11IP in **(A)** PAS1, **(B)** PAS2, and **(C)** PAS5.

and analysis at the gene structure level was carried out, which is also one of the advantages of this study. Meanwhile, this study also has some shortcomings. First, the sample size of the study is small, and it is susceptible to external factors. In order to make up for this deficiency, the patients enrolled in this study were limited to having a history of cesarean section and no other history of uterine cavity operation or uterine-related diseases, which excluded the interference of different PAS etiologies and some other potential confounding factors. However, the hub genes and related functions discovered in this study still need to be verified by a large number of further experiments. Second, there were few omics studies related to PAS. Only one relevant sequencing data set can be found on Gene Expression Omnibus (GSE126552), in which PAS and normal placenta tissue was sequenced to find potential biomarkers and molecular mechanism of PAS. We downloaded and analyzed the data set, but unfortunately, there were no differentially expressed lncRNAs and mRNAs in this data set. Therefore, the hub genes found in this study cannot be effectively verified externally. Nevertheless, several hub genes, including *MMP12*, *FSTL3*, *TEAD3*, *ERVW-1*, and *INHA*, have been confirmed in other studies to regulate the invasion of EVTs or vascular development, which has been mentioned in the corresponding part of the discussion.

Omics analysis provides a new method for systematically exploring the hub genes of PAS. However, except for tumors, omics study of most other diseases is still in its infancy. This systematic study provided a genetic basis for further in-depth study on the underlying regulatory mechanism of PAS. With the progress of omics studies in PAS and the increasing sample size, the genetic features of PAS will be more clear and the structure variation of hub genes will be more statistically significant, which will contribute to the gene-level intervention in the development of PAS.

DATA AVAILABILITY STATEMENT

The RNA sequencing data was uploaded to NCBI Sequence Read Archive. The BioProject Number is PRJNA627183 and you can learn more about the data at http://www.ncbi.nlm.nih.gov/bioproject/627183.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of China Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CQ and BC designed the study. CQ supervised the whole study. BC and DW analyzed the data and wrote the manuscript. YB, JL, TY, and NL collected and sequenced the samples. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key R&D Program of China (2016YFC1000404), the National Natural Science Foundation of China (General Program; 81370735 and General Program; 81771610), the Outstanding Scientific Fund of Shengjing Hospital (201706), and the Distinguished professor of Liaoning Province (2017).

ACKNOWLEDGMENTS

We are very grateful to the patients who agreed to participate in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.551495/full#supplementary-material

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Aydin, B., and Arga, K. Y. (2019). Co-expression network analysis elucidated a core module in association with prognosis of non-functioning non-invasive human pituitary adenoma. *Front. Endocrinol.* 10:361. doi: 10.3389/fendo.2019.00361

Azar, C., Valentine, M., Trausch-Azar, J., Druley, T., Nelson, D. M., and Schwartz, A. L. (2018). RNA-Seq identifies genes whose proteins are transformative in the differentiation of cytotrophoblast to syncytiotrophoblast, in human primary villous and BeWo trophoblasts. *Sci. Rep.* 8:5142. doi: 10.1038/s41598-018-23379-2

Bai, R., Kusama, K., Nakamura, K., Sakurai, T., Kimura, K., Ideta, A., et al. (2018). Down-regulation of transcription factor OVOL2 contributes to epithelial-mesenchymal transition in a noninvasive type of trophoblast implantation to the maternal endometrium. *FASEB J.* 32, 3371–3384. doi: 10.1096/fj.201701131RR

Bani Baker, Q., and Nuser, M. S. (2019). Bioinformatics in Jordan: status, challenges, and future directions. *PLoS Comput. Biol.* 15:e1007202. doi: 10.1371/journal.pcbi.1007202

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-analyzed tumors. *Cell* 173:530. doi: 10.1016/j.cell.2018.03.059

Brown, M. L., Ramprasad, M. P., Umeda, P. K., Tanaka, A., Kobayashi, Y., Watanabe, T., et al. (2000). A macrophage receptor for apolipoprotein B48: cloning, expression, and atherosclerosis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7488–7493. doi: 10.1073/pnas.120184097

Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinovic, S. M., Cheng, L. Y., Messner, S., et al. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteom.* 14, 1400–1410. doi: 10.1074/mcp.M114.044305

Chakraborty, D., Cui, W., Rosario, G. X., Scott, R. L., Dhakal, P., Renaud, S. J., et al. (2016). HIF-KDM3A-MMP12 regulatory circuit ensures trophoblast plasticity and placental adaptations to hypoxia. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7212–E7221. doi: 10.1073/pnas.1612626113

Chang, H., Zhang, C., and Cao, Y. (2012). Expression and distribution of symplekin regulates the assembly and function of the epithelial tight junction. *Histochem. Cell Biol.* 137, 319–327. doi: 10.1007/s00418-011-0906-z

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511

Dai, W., Wang, Q., Zhao, F., Liu, J., and Liu, H. (2018). Understanding the regulatory mechanisms of milk production using integrative transcriptomic and proteomic analyses: improving inefficient utilization of crop by-products as forage in dairy industry. *BMC Genom.* 19:403. doi: 10.1186/s12864-018-4808-5

Danaher, P., Warren, S., Lu, R., Samayoa, J., Sullivan, A., Pekker, I., et al. (2018). Pan-cancer adaptive immune resistance as defined by the tumor inflammation signature (TIS): results from the cancer genome atlas (TCGA). *J. Immunother. Cancer* 6:63. doi: 10.1186/s40425-018-0367-1

Dunk, C. E., Pappas, J. J., Lye, P., Kibschull, M., Javam, M., Bloise, E., et al. (2018). P-Glycoprotein (P-gp)/ABCB1 plays a functional role in extravillous trophoblast (EVT) invasion and is decreased in the pre-eclamptic placenta. *J. Cell Mol. Med.* 22, 5378–5393. doi: 10.1111/jcmm.13810

El Gelany, S., Mosbeh, M. H., Ibrahim, E. M., Mohammed, M., Khalifa, E. M., Abdelhakium, A. K., et al. (2019). Placenta Accreta Spectrum (PAS) disorders: incidence, risk factors and outcomes of different management strategies in a tertiary referral hospital in Minia, Egypt: a prospective study. *BMC Pregn. Childbirth* 19:313. doi: 10.1186/s12884-019-2466-5

Farquhar, C. M., Li, Z., Lensen, S., Mclintock, C., Pollock, W., Peek, M. J., et al. (2017). Incidence, risk factors and perinatal outcomes for placenta accreta in Australia and New Zealand: a case-control study. *BMJ Open* 7:e017713. doi: 10.1136/bmjopen-2017-017713

Hara, T., Monguchi, T., Iwamoto, N., Akashi, M., Mori, K., Oshita, T., et al. (2017). Targeted Disruption of JCAD (junctional protein associated with coronary artery disease)/KIAA1462, a coronary artery disease-associated gene product, inhibits angiogenic processes in vitro and in vivo. *Arterioscler. Thromb. Vasc. Biol.* 37, 1667–1673. doi: 10.1161/ATVBAHA.117.309721

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503. doi: 10.1093/nar/gkz1031

Jauniaux, E., Ayres-De-Campos, D., Langhoff-Roos, J., Fox, K. A., Collins, S., Diagnosis, F. P. A., et al. (2019). FIGO classification for the clinical diagnosis of placenta accreta spectrum disorders. *Int. J. Gynaecol. Obstet.* 146, 20–24. doi: 10.1002/ijgo.12761

Jauniaux, E., Chantraine, F., Silver, R. M., Langhoff-Roos, J., and for the Figo Placenta Accreta Diagnosis and Management Expert Consensus Panel (2018a). FIGO consensus guidelines on placenta accreta spectrum disorders: epidemiology. *Int. J. Gynaecol. Obstet.* 140, 265–273. doi: 10.1002/ijgo.12407

Jauniaux, E., Collins, S., and Burton, G. J. (2018b). Placenta accreta spectrum: pathophysiology and evidence-based anatomy for prenatal ultrasound imaging. *Am. J. Obstet. Gynecol.* 218, 75–87. doi: 10.1016/j.ajog.2017.05.067

Jia, W., Qiu, K., He, M., Song, P., Zhou, Q., Zhou, F., et al. (2013). SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol.* 14:R12. doi: 10.1186/gb-2013-14-2-r12

Kabiri, D., Hants, Y., Shanwetter, N., Simons, M., Weiniger, C. F., Gielchinsky, Y., et al. (2014). Outcomes of subsequent pregnancies after conservative treatment for placenta accreta. *Int. J. Gynaecol. Obstet.* 127, 206–210. doi: 10.1016/j.ijgo.2014.05.013

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Koo, Y., Barry, D. M., Xu, K., Tanigaki, K., Davis, G. E., Mineo, C., et al. (2016). Rasip1 is essential to blood vessel stability and angiogenic blood vessel growth. *Angiogenesis* 19, 173–190. doi: 10.1007/s10456-016-9498-5

Liu, W. Q., Dong, J., Peng, Y. X., Li, W. L., and Yang, J. (2018). Synonymous mutation adenomatous polyposis coliDelta486s affects exon splicing and may predispose patients to adenomatous polyposis coli/mutY DNA glycosylase mutationnegative familial *Adenomatous polyposis*. *Mol. Med. Rep.* 18, 4931–4939. doi: 10.3892/mmr.2018.9495

Long, Y., Jiang, Y., Zeng, J., Dang, Y., Chen, Y., Lin, J., et al. (2020). The expression and biological function of chemokine CXCL12 and receptor CXCR4/CXCR7 in placenta accreta spectrum disorders. *J. Cell Mol. Med.* 24, 3167–3182. doi: 10.1111/jcmm.14990

Lu, J., Liang, X., Gao, Y., Fu, G., and Shen, Q. (2019). Hexokinase2 controls angiogenesis in melanoma by promoting aerobic glycolysis and activating the p38-MAPK signaling. *J. Cell. Biochem.* 120, 19721–19729. doi: 10.1002/jcb.29278

Maier, T., Guell, M., and Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* 583, 3966–3973. doi: 10.1016/j.febslet.2009.10.036

Matsukawa, S., Sumigama, S., Kotani, T., Wang, J., Miki, R., Moriyama, Y., et al. (2019). Possible association between cathepsin V and the development of placenta accreta spectrum disorders. *Gynecol. Obstet. Invest.* 84, 396–406. doi: 10.1159/000496609

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA SEQUENCING data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Mertens, F., Johansson, B., Fioretos, T., and Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer* 15, 371–381. doi: 10.1038/nrc3947

Montioli, R., Fargue, S., Lewin, J., Zamparelli, C., Danpure, C. J., Borri Voltattorni, C., et al. (2012). The N-terminal extension is essential for the formation of the active dimeric structure of liver peroxisomal alanine:glyoxylate aminotransferase. *Int. J. Biochem. Cell Biol.* 44, 536–546. doi: 10.1016/j.biocel.2011.12.007

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34, 267–273. doi: 10.1038/ng1180

Orbach-Zinger, S., Weiniger, C. F., Aviram, A., Balla, A., Fein, S., Eidelman, L. A., et al. (2018). Anesthesia management of complete versus incomplete placenta previa: a retrospective cohort study. *J. Matern. Fetal Neonatal Med.* 31, 1171–1176. doi: 10.1080/14767058.2017.1311315

Pavlicev, M., Wagner, G. P., Chavan, A. R., Owens, K., Maziarz, J., Dunn-Fletcher, C., et al. (2017). Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res.* 27, 349–361. doi: 10.1101/gr.207597.116

Poma, A. M., Torregrossa, L., Bruno, R., Basolo, F., and Fontanini, G. (2018). Hippo pathway affects survival of cancer patients: extensive analysis of TCGA data and review of literature. *Sci. Rep.* 8:10623. doi: 10.1038/s41598-018-28928-3

Rao, G., Pierobon, M., Kim, I. K., Hsu, W. H., Deng, J., Moon, Y. W., et al. (2017). Inhibition of AKT1 signaling promotes invasion and metastasis of non-small cell lung cancer cells with K-RAS or EGFR mutations. *Sci. Rep.* 7:7066. doi: 10.1038/s41598-017-06128-9

Shadeo, A., Chari, R., Lonergan, K. M., Pusic, A., Miller, D., Ehlen, T., et al. (2008). Up regulation in gene expression of chromatin remodelling factors in cervical intraepithelial neoplasia. *BMC Genomics* 9:64. doi: 10.1186/1471-2164-9-64

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, Y., Miladi, M., Dukare, S., Boulay, K., Caudron-Herger, M., Gross, M., et al. (2019). A pan-cancer analysis of synonymous mutations. *Nat. Commun.* 10:2569. doi: 10.1038/s41467-019-10489-2

Silver, R. M., and Branch, D. W. (2018). Placenta accreta spectrum. *N. Engl. J. Med.* 378, 1529–1536. doi: 10.1056/NEJMcp1709324

Slawin, K. M., Shariat, S. F., Nguyen, C., Leventis, A. K., Song, W., Kattan, M. W., et al. (2000). Detection of metastatic prostate cancer using a splice variant-specific reverse transcriptase-polymerase chain reaction assay for human glandular kallikrein. *Cancer Res.* 60, 7142–7148.

Society of Gynecologic Oncology, American College of Obstetricians and Gynecologists and the Society for Maternal-Fetal Medicine, Cahill, A. G., Beigi, R., Heine, R. P., Silver, R. M., et al. (2018). Placenta accreta spectrum. *Am. J. Obstet. Gynecol.* 219, B2–B16. doi: 10.1016/j.ajog.2018.09.042

Sonderegger, S., Haslinger, P., Sabri, A., Leisser, C., Otten, J. V., Fiala, C., et al. (2010). Wingless (Wnt)-3A induces trophoblast migration and matrix metalloproteinase-2 secretion through canonical Wnt signaling and protein kinase B/AKT activation. *Endocrinology* 151, 211–220. doi: 10.1210/en.2009-0557

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, 15545–15550. doi: 10.1073/pnas.0506580102

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Tantbirojn, P., Crum, C. P., and Parast, M. M. (2008). Pathophysiology of placenta creta: the role of decidua and extravillous trophoblast. *Placenta* 29, 639–645. doi: 10.1016/j.placenta.2008.04.008

Thangavelu, P. U., Krenacs, T., Dray, E., and Duijf, P. H. (2016). In epithelial cancers, aberrant COL17A1 promoter methylation predicts its misexpression and increased invasion. *Clin. Epigenet.* 8:120. doi: 10.1186/s13148-016-0290-6

Thurn, L., Lindqvist, P. G., Jakobsson, M., Colmorn, L. B., Klungsoyr, K., Bjarnadottir, R. I., et al. (2016). Abnormally invasive placenta-prevalence, risk factors and antenatal suspicion: results from a large population-based pregnancy cohort study in the Nordic countries. *BJOG* 123, 1348–1355. doi: 10.1111/1471-0528.13547

Vallon, M., Rohde, F., Janssen, K. P., and Essler, M. (2010). Tumor endothelial marker 5 expression in endothelial cells during capillary morphogenesis is induced by the small GTPase Rac and mediates contact inhibition of cell proliferation. *Exp. Cell Res.* 316, 412–421. doi: 10.1016/j.yexcr.2009.10.013

Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., et al. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* 563, 347–353. doi: 10.1038/s41586-018-0698-6

Vinograd, A., Wainstock, T., Mazor, M., Beer-Weisel, R., Klaitman, V., Dukler, D., et al. (2015). Placenta accreta is an independent risk factor for late pre-term birth and perinatal mortality. *J. Matern. Fetal Neonatal Med.* 28, 1381–1387. doi: 10.3109/14767058.2014.955004

Vogel, C., and Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. doi: 10.1038/nrg3185

Wang, D., Eraslan, B., Wieland, T., Hallstrom, B., Hopf, T., Zolg, D. P., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* 15:e8503. doi: 10.15252/msb.20188503

Wang, Q., Lu, X., Li, C., Zhang, W., Lv, Y., Wang, L., et al. (2019). Down-regulated long non-coding RNA PVT1 contributes to gestational diabetes mellitus and preeclampsia via regulation of human trophoblast cells. *Biomed. Pharmacother.* 120:109501. doi: 10.1016/j.biopha.2019.109501

Weinsheimer, S., Brettman, A. D., Pawlikowska, L., Wu, D. C., Mancuso, M. R., Kuhnert, F., et al. (2012). G protein-coupled receptor 124 (GPR124) gene polymorphisms and risk of brain arteriovenous malformation. *Transl. Stroke Res.* 3, 418–427. doi: 10.1007/s12975-012-0202-9

Xie, J., Xu, Y., Wan, L., Wang, P., Wang, M., and Dong, M. (2018). Involvement of follistatin-like 3 in preeclampsia. *Biochem. Biophys. Res. Commun.* 506, 692–697. doi: 10.1016/j.bbrc.2018.10.139

Xu, D., Jin, J., Yu, H., Zhao, Z., Ma, D., Zhang, C., et al. (2017). Chrysin inhibited tumor glycolysis and induced apoptosis in hepatocellular carcinoma by targeting hexokinase-2. *J. Exp. Clin. Cancer Res.* 36:44. doi: 10.1186/s13046-017-0514-4

Yang, S., Quaresma, A. J., Nickerson, J. A., Green, K. M., Shaffer, S. A., Imbalzano, A. N., et al. (2015). Subnuclear domain proteins in cancer cells support the functions of RUNX2 in the DNA damage response. *J. Cell Sci.* 128, 728–740. doi: 10.1242/jcs.160051

Yeo, B., Redfern, A. D., Mouchemore, K. A., Hamilton, J. A., and Anderson, R. L. (2018). The dark side of granulocyte-colony stimulating factor: a supportive therapy with potential to promote tumour progression. *Clin. Exp. Metast.* 35, 255–267. doi: 10.1007/s10585-018-9917-7

Yoshimi, A., Lin, K. T., Wiseman, D. H., Rahman, M. A., Pastore, A., Wang, B., et al. (2019). Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature* 574, 273–277. doi: 10.1038/s41586-019-1618-0

Zhang, H., Pan, Y. Z., Cheung, M., Cao, M., Yu, C., Chen, L., et al. (2019). LAMB3 mediates apoptotic, proliferative, invasive, and metastatic behaviors in pancreatic cancer by regulating the PI3K/Akt signaling pathway. *Cell Death Dis.* 10:230. doi: 10.1038/s41419-019-1320-z

Zhang, J., Zhu, Z., Wu, H., Yu, Z., Rong, Z., Luo, Z., et al. (2019). PODXL, negatively regulated by KLF4, promotes the EMT and metastasis and serves as a novel prognostic indicator of gastric cancer. *Gastr. Cancer* 22, 48–59. doi: 10.1007/s10120-018-0833-y

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523. doi: 10.1038/s41467-019-09234-6

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 Genetics

METHODS
published: 14 October 2020
doi: 10.3389/fgene.2020.562971

ASDmiR: A Stepwise Method to Uncover miRNA Regulation Related to Autism Spectrum Disorder

Chenchen Xiong[1†], Shaoping Sun[2†], Weili Jiang[1], Lei Ma[1] and Junpeng Zhang[3*]*

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China, [2] Department of Medical Engineering, People's Hospital of Yuxi City, Yuxi, China, [3] School of Engineering, Dali University, Dali, China

Autism spectrum disorder (ASD) is a class of neurodevelopmental disorders characterized by genetic and environmental risk factors. The pathogenesis of ASD has a strong genetic basis, consisting of rare *de novo* or inherited variants among a variety of multiple molecules. Previous studies have shown that microRNAs (miRNAs) are involved in neurogenesis and brain development and are closely associated with the pathogenesis of ASD. However, the regulatory mechanisms of miRNAs in ASD are largely unclear. In this work, we present a stepwise method, ASDmiR, for the identification of underlying pathogenic genes, networks, and modules associated with ASD. First, we conduct a comparison study on 12 miRNA target prediction methods by using the matched miRNA, lncRNA, and mRNA expression data in ASD. In terms of the number of experimentally confirmed miRNA–target interactions predicted by each method, we choose the best method for identifying miRNA–target regulatory network. Based on the miRNA–target interaction network identified by the best method, we further infer miRNA–target regulatory bicliques or modules. In addition, by integrating high-confidence miRNA–target interactions and gene expression data, we identify three types of networks, including lncRNA–lncRNA, lncRNA–mRNA, and mRNA–mRNA related miRNA sponge interaction networks. To reveal the community of miRNA sponges, we further infer miRNA sponge modules from the identified miRNA sponge interaction network. Functional analysis results show that the identified hub genes, as well as miRNA-associated networks and modules, are closely linked with ASD. ASDmiR is freely available at https://github.com/chenchenxiong/ASDmiR.

Keywords: miRNA, lncRNA, mRNA, miRNA regulation, autism spectrum disorder

OPEN ACCESS

Edited by:
Cheng Peng,
Yunnan University, China

Reviewed by:
Guoli Ji,
Xiamen University, China
Yuriy L. Orlov,
First Moscow State Medical University, Russia

****Correspondence:***
Lei Ma
roy_murray@qq.com
Junpeng Zhang
zhangjunpeng_411@yahoo.com

† These authors have contributed equally to this work

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

Received: *17 May 2020*
Accepted: *31 August 2020*
Published: *14 October 2020*

Citation:
Xiong C, Sun S, Jiang W, Ma L and Zhang J (2020) ASDmiR: A Stepwise Method to Uncover miRNA Regulation Related to Autism Spectrum Disorder. Front. Genet. 11:562971. doi: 10.3389/fgene.2020.562971

INTRODUCTION

Autism spectrum disorder (ASD) encompasses a variety of complex inheritable neurodevelopment disorders that usually occur before 3 years old and last throughout a person's life (Fregeac et al., 2016; Quesnel-Vallieres et al., 2019). ASD patients are characterized by controlled social interactions, restricted activities, and repetitive behavior (Chen et al., 2015). The current diagnosis of ASD is mainly based on behavioral characteristics (Gillian et al., 2003), which may cause misdiagnosis or delay treatment. Previous transcriptomic studies (Voineagu et al., 2011;

Gupta et al., 2014; Ansel et al., 2016; Quesnel-Vallieres et al., 2019) have reported that ASD has strong genetic complexity, and many genes are involved in the ASD-related biological processes, including neuronal activity (Voineagu et al., 2011), immune response (Gupta et al., 2014; Ansel et al., 2016), and signaling pathways (Quesnel-Vallieres et al., 2019). Although great progress has been made to study the pathogenesis of ASD, the gene regulation in ASD is largely unknown because of the heterogeneity and complexity of ASD. Therefore, it is necessary to investigate the pathogenesis and molecular mechanisms underlying ASD for improving the diagnosis and therapeutic strategies of patients.

At the genetic level, microRNAs (miRNAs) are important regulators of brain function and neuronal development (Rajman and Schratt, 2017; Shen et al., 2019). By binding with messenger RNAs (mRNAs) at the post-transcriptional level, miRNAs as tiny non-coding RNA molecules (∼22 nucleotides) can induce repression or translational inhabitation of mRNAs (Ambros, 2004). Previous studies (Ander et al., 2015; Hu et al., 2017; Shen et al., 2016) have elucidated that miRNAs participate in several biological processes that are closely associated with ASD, including synaptic plasticity and neuronal development (Hu et al., 2017), immune response (Ander et al., 2015), and signaling pathways (Shen et al., 2016). These studies have also indicated that miRNAs and their corresponding targets could help to uncover ASD pathogenesis.

Long non-coding RNAs (LncRNAs) are transcripts with a length of more than 200 nucleotides, and they play critical roles in the progression of neuropsychiatric disorders including ASD (Hosseini et al., 2019). In the developmental processes of ASD, lncRNAs take part in several important biological processes, including neuronal architecture and immune response (Kerin et al., 2012), synaptic and neuronal excitatory dysfunction (Noor et al., 2010), neurite elaboration (Wang et al., 2015), and alternative splicing (Parikshak et al., 2016). These studies have demonstrated the potential contribution of lncRNAs on revealing the molecular mechanisms of ASD.

According to competing endogenous RNA hypothesis (Salmena et al., 2011), coding and non-coding RNA transcripts compete with each other by base pairing with miRNA-recognition elements (MREs). These transcripts are also known as miRNA sponges, including mRNAs (Tay et al., 2011), lncRNAs (Cesana et al., 2011), pseudogenes (Poliseno et al., 2010), and circular RNAs (circRNAs) (Hansen et al., 2013). All types of miRNA sponges crosstalk with other through MREs and form a large-scale miRNA sponge interaction network (Salmena et al., 2011). Although accumulating miRNA sponges have been experimentally identified and are closely relevant to various cancers (Le et al., 2017), the roles of miRNA sponges in ASD are largely unknown. To uncover potential roles of miRNA sponges in ASD, we focus on investigating lncRNA and mRNA related miRNA sponge interaction networks in ASD in this work.

There have been growing computational methods to effectively explore miRNA functions based on gene expression data. However, current bioinformatics research on miRNA regulatory mechanisms related to ASD is still in its infancy. In this work, we propose a novel stepwise method, ASDmiR, to uncover miRNA regulation in ASD. ASDmiR has two main contributions as follows. First, ASDmiR can be used to study ASD-related miRNA regulation at both the network and module level. Secondly, ASDmiR can help to explore both direct and indirect miRNA regulation in ASD. At the network level, we identify two types of ASD-related networks: miRNA–target regulatory network and miRNA sponge interaction network. Meanwhile, at the module level, we infer two types of ASD-related modules: miRNA–target regulatory modules and miRNA sponge modules. Topological analysis and functional analysis have shown that the identified miRNA-associated networks and modules are highly implicated in ASD.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

Differential Expression Analysis

Previous studies (Mohr and Liew, 2007; Segura et al., 2015; Ansel et al., 2016) have discovered that peripheral blood samples are more accessible than brain tissue samples in the transcriptomic study of ASD. In this work, we obtained the matched miRNA, lncRNA, and mRNA expression profiles of ASD from Kong et al. (2012). The samples of gene expression profiles are from peripheral blood samples and are categorized as ASD (104 samples) and normal (82 samples). We apply the *miRBaseConverter* (Xu et al., 2018) R package to convert miRNA names into the latest version of miRBase. To discover the differentially expressed miRNAs, lncRNAs, and mRNAs between ASD samples and normal samples, we conduct differential expression analysis using the *limma* R package (Ritchie et al., 2015). In the ASD dataset, the changes in mRNA expression level between ASD samples and normal samples are large, while the changes in the case of miRNAs and lncRNAs are small. To cover important ASD-related miRNAs and lncRNAs and to have a moderate number of mRNAs for ASDmiR, we select top 100 miRNAs, 300 lncRNAs, and 4,000 mRNAs ranked by adjusted p-values (adjusted by the Benjamini and Hochberg method) in the differential gene expression analysis for subsequent analysis. The detailed results of differentially expressed miRNAs, lncRNAs, and mRNAs can be seen in **Supplementary Table 1**.

MiRNA-Target Interactions

For putative miRNA–mRNA interactions, we have obtained 762,540 unique interactions between 2,600 miRNAs and 21,538 mRNAs from miRTarBase v8.0 (Chou et al., 2018) and TarBase v8.0 (Karagkouni et al., 2018) databases. By combining the interactions from LncBase v2.0 (Paraskevopoulou et al., 2016) and NPInter v4.0 (Teng et al., 2020) databases, we have collected 138,951 unique miRNA–lncRNA interactions between 1,044 miRNAs and 13,243 lncRNAs. The obtained miRNA–target interactions could be seen in **Supplementary Table 2**.

ASD-Related Genes

In this work, we collect a list of miRNAs, lncRNAs, and mRNAs associated with ASD to investigate ASD-related miRNA regulation. In total, we have obtained a list of 141 ASD-related

miRNAs from HMDD v3.2 (Huang et al., 2019) and MNDR v2.0 (Cui et al., 2018), a list of 117 ASD-related lncRNAs from LncRNADisease v2.0 (Bao et al., 2019) and MNDR v2.0 (Cui et al., 2018), and a list of 1,658 ASD-related mRNAs from the Simons Foundation Autism Research Initiative (SFARI)[1] and DisGeNET v7.0 (Pinero et al., 2020). The obtained ASD-related miRNAs, lncRNAs, and mRNAs could be seen in **Supplementary Table 2**.

Methods

Overview of ASDmiR

In **Figure 1**, the workflow of ASDmiR includes three major steps for identifying miRNA-associated networks and modules related to ASD. In the first step, by using the matched miRNA, lncRNA, and mRNA expression profiles, we conduct a comparison study of 12 commonly used miRNA target prediction methods from Le et al. (2015). In terms of the number of experimentally validated miRNA–mRNA interactions, we select the best performing method to identify miRNA–target regulatory network in ASD dataset. Furthermore, we infer miRNA–target regulatory modules based on the identified miRNA–target regulatory network. In the second step, we use the well-cited sensitivity partial Pearson correlation (SPPC) method (Paci et al., 2014) to identify miRNA sponge interaction network by integrating putative miRNA–target interactions and gene expression data. Moreover, the Markov cluster (MCL) algorithm (Enright et al., 2002) is used to discover miRNA sponge modules for investigating the community of miRNA sponges. In the final step, we conduct functional analysis of the identified miRNA-associated networks and modules. In the following, we will describe the details of these steps.

Identification of miRNA-Target Regulatory Network and Modules

To identify miRNA–target interactions, we use 12 existing computational methods implemented in the *miRLAB* R package (Le et al., 2015). These miRNA target prediction methods could be categorized into four types: correlation methods, regression methods, causal inference methods, and other methods. The first type of computational methods, including Pearson (Pearson, 1920), Spearman (Spearman, 1904), Kendall (Kendall, 1938), Distance correlation (Székely et al., 2007), and Hoeffding's D measure (Hoeffding, 1948), could calculate linear correlation relationships between miRNAs and targets. To capture non-linear relationships between miRNAs and targets, the randomized dependence coefficient (Lopez-Paz et al., 2013) and Mutual Information (MI) (Moon et al., 1995) methods are utilized. For the second type of computational methods, Lasso (Tibshirani, 1996) and Elastic-net (Zou and Hastie, 2005) are used to identify the associations between miRNAs and targets. As for the third type of computational methods, the IDA (Intervention calculus when the Directed acyclic graph is Absent) method (Maathuis et al., 2009) is selected to estimate the causal effects that miRNAs have on mRNAs. For the fourth type of computational methods, *Z* score (Prill et al., 2010) and probabilistic MiRNA–mRNA Interaction Signature (ProMISe) (Li et al., 2014) are used. The *Z* score method is commonly used in gene-knockdown experiments to estimate the effect of knocking out a miRNA on mRNAs, and the ProMISe method estimates the probability of a miRNA targeting each mRNA by considering the competition among mRNAs and the competition among miRNAs. In this work, miRNAs are upstream variables, and targets (lncRNAs and mRNAs) are downstream variables. For each computational method, we use experimentally validated miRNA–target interactions as the ground truth to validate top 50, 100, 150, 200 predicted targets of each miRNA. The more the number of miRNA–target interactions validated by the ground truth is, the better the computational method performs.

It is known that genes tend to implement a specific biological process in the form of a community or module (Choobdar et al., 2019). Therefore, we further identify miRNA–target regulatory modules based on the identified miRNA–target regulatory network. Different from other biological networks (i.e., protein-protein interaction network), the miRNA–target regulatory network is a bipartite network. Consequently, the generated miRNA–target modules are actually bicliques where every miRNA of the miRNA set is connected to each target gene of the target gene set (Yoon et al., 2019). In this work, we utilize the R package *biclique* (Zhang et al., 2014) to enumerate all bicliques from the identified miRNA–target bipartite network. Here, a biclique corresponds to a miRNA–target regulatory module, and we only consider the bicliques with at least 3 miRNAs and 3 targets.

Identification of MiRNA Sponge Interaction Network and Modules

In this section, we apply the SPPC method (Paci et al., 2014) implemented in the *miRspongeR* R package (Zhang et al., 2019) to infer miRNA sponge interactions. The SPPC method takes miRNA, lncRNA, and mRNA expression data into account for identifying miRNA sponge interactions, and quantitatively evaluates the effect of sharing miRNAs on each miRNA sponge interaction pair at the expression level. This method uses three constraints (significant sharing of common miRNAs, significant positive correlation, and adequate sensitivity correlation) to evaluate whether a candidate RNA–RNA pair (lncRNA–lncRNA, lncRNA–mRNA, and mRNA–mRNA pair) is a miRNA sponge interaction or not. Given two competing RNAs (RNA_i and RNA_j), the significance *p*-value of sharing miRNAs and positive correlation is usually set to be 0.05. The Sensitive Correlation (*SC*) between the RNA_i–RNA_j pair is calculated as follows:

$$SC = \rho_{ij} - \rho_{ij|n} \quad (1)$$

Where ρ_{ij} denotes Pearson correlation (Pearson, 1920) between RNA_i and RNA_j, and $\rho_{ij|n}$ is partial Pearson correlation between RNA_i and RNA_j on the condition of *n* sharing miRNAs. In this work, the cutoff of *SC* is set to be 0.25 (see "The Identified MiRNA-Associated Modules Are Functional" for details). After assembling the identified miRNA sponge interactions, we could gain three types of networks, including lncRNA–lncRNA, lncRNA–mRNA, and mRNA–mRNA related

[1] https://www.sfari.org/

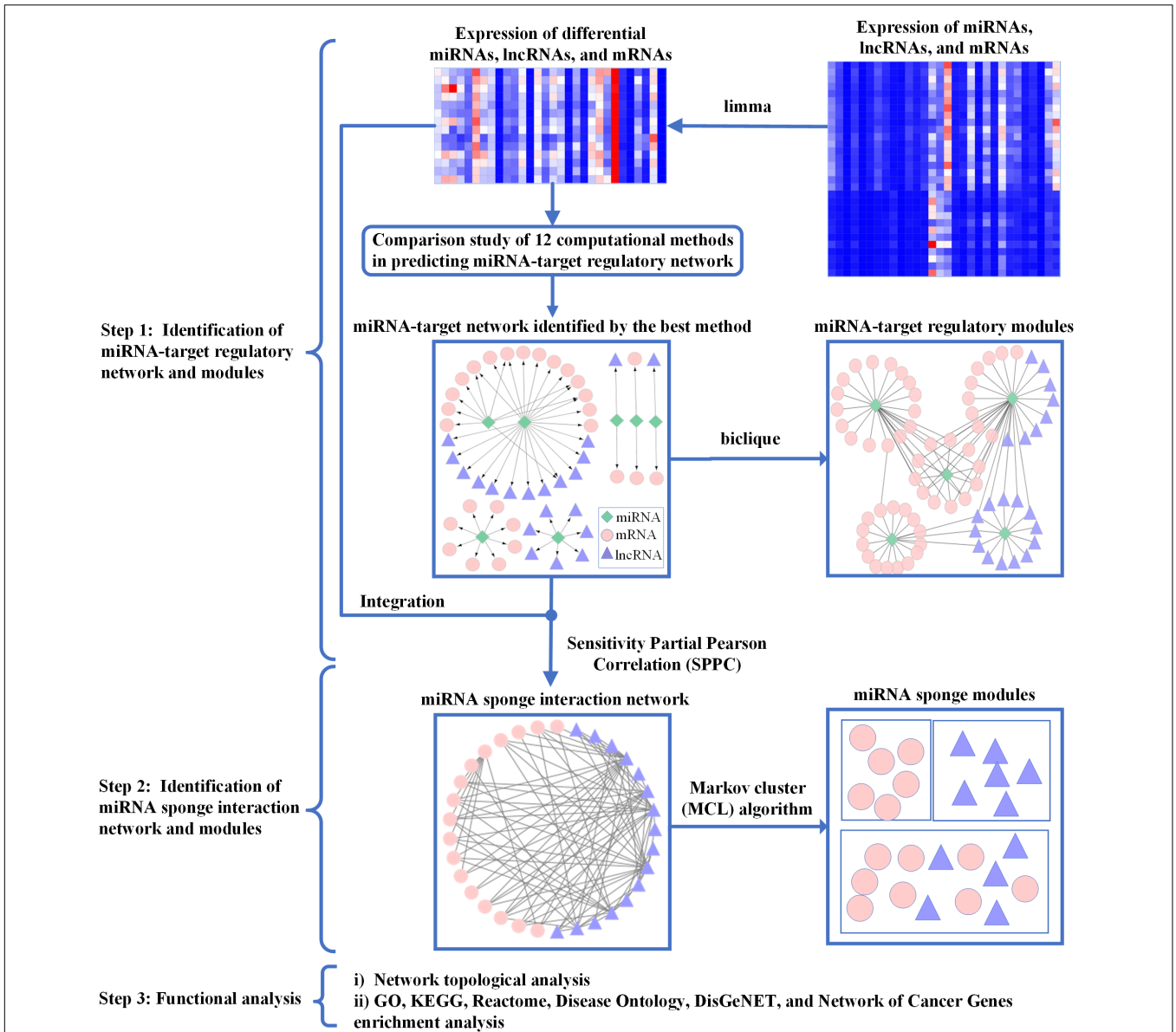


FIGURE 1 | The workflow of ASDmiR. In Step 1, by integrating expression data of differential miRNAs, lncRNAs, and mRNAs and putative miRNA–target interactions, we identify miRNA–target regulatory networks using 12 existing computational methods. The miRNA–target regulatory network predicted by the best method is used for identifying miRNA–target regulatory modules. In Step 2, based on putative miRNA–mRNA interactions and gene expression data, we infer miRNA sponge interaction network using the sensitivity partial Pearson correlation method. Furthermore, we identify miRNA sponge modules from the identified miRNA sponge interaction network by using the Markov cluster algorithm. In Step 3, we conduct functional analysis of the identified miRNA-associated networks and modules. Green rhombic, pink circle, and purple triangle nodes represent miRNAs, mRNAs, and lncRNAs, respectively.

miRNA sponge interaction networks. At the module level, we further infer miRNA sponge modules by using the Markov cluster (MCL) algorithm (Enright et al., 2002). For each module, the number of miRNA sponges (lncRNAs or mRNAs) is at least 3.

Functional Analysis

The hub genes may play key roles in the characteristics and development of complex diseases (Zhang et al., 2018). Consequently, at the network level, we focus on identifying hub genes from both the identified miRNA–target regulatory network and miRNA sponge interaction network. Empirically, we choose top 20% miRNAs or miRNA sponges with the largest degree as hub miRNAs or hub miRNA sponges. Furthermore, we use the miEAA (Kern et al., 2020) online tool to conduct functional enrichment analysis of hub miRNAs, and the *miRspongeR* (Zhang et al., 2019) R package for functional enrichment analysis of hub miRNA sponges.

At the module level, to know the potential diseases, biological processes, and pathways associated with the identified miRNA-associated modules, we conduct functional enrichment analysis

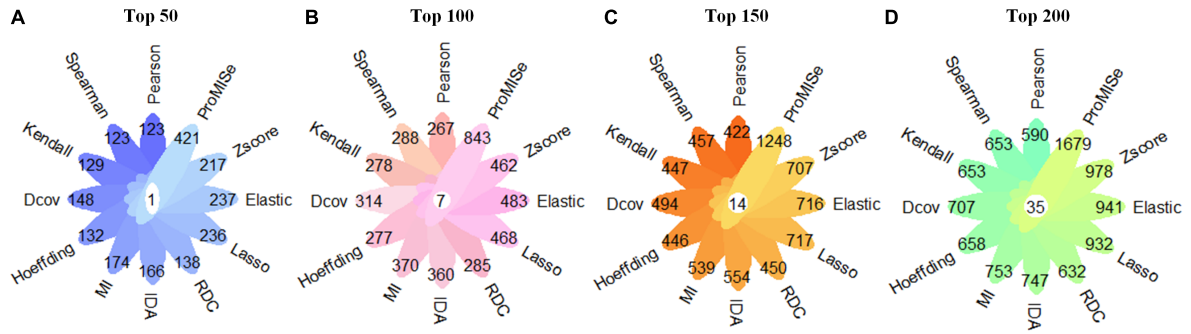
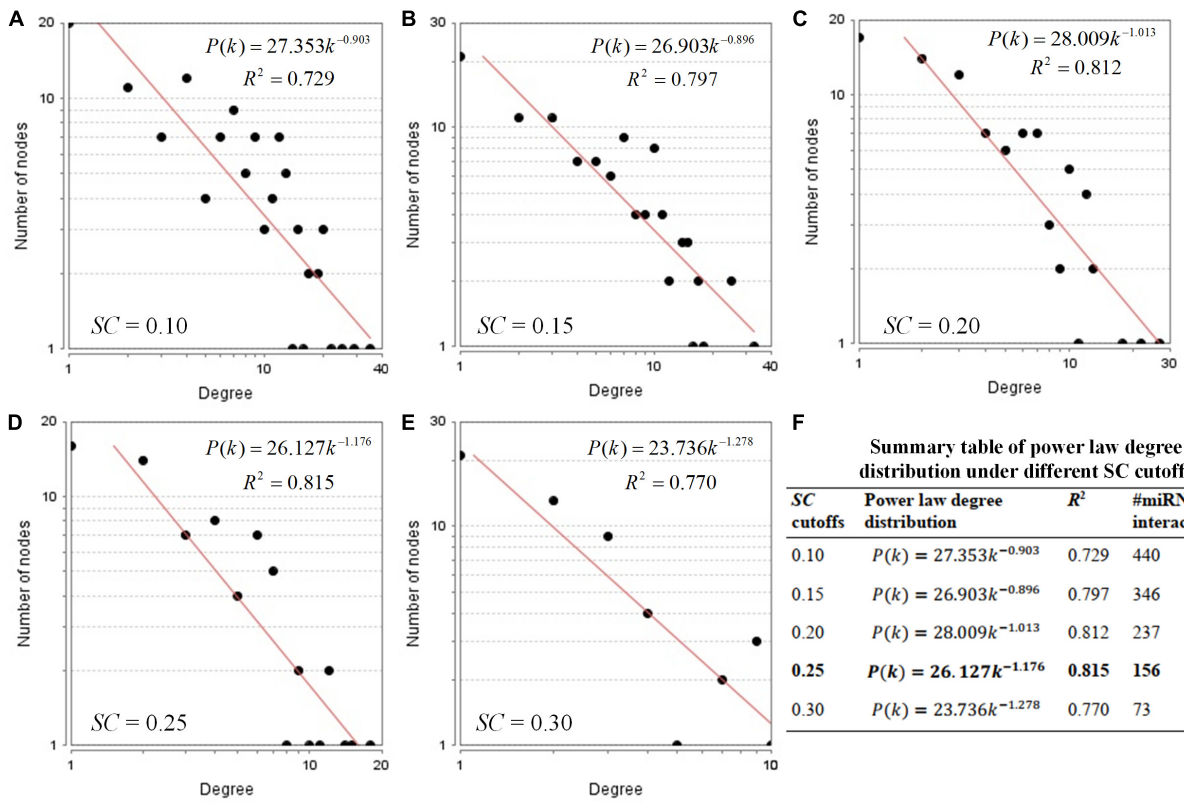


FIGURE 2 | Comparison in terms of the number of confirmed miRNA–target interactions using 12 miRNA–target prediction methods. **(A)** The number of validated miRNA–target interactions in the case of top 50 miRNA–target interactions of each miRNA. **(B)** The number of validated miRNA–target interactions in the case of top 100 miRNA–target interactions of each miRNA. **(C)** The number of validated miRNA–target interactions in the case of top 150 miRNA–target interactions of each miRNA. **(D)** The number of validated miRNA–target interactions in the case of top 200 miRNA–target interactions of each miRNA. The numbers in the white circle denote the overlap of validated miRNA–target interactions by 12 computational methods.



F

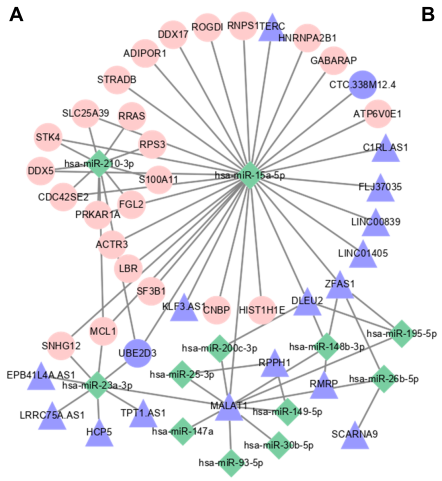
Summary table of power law degree distribution under different SC cutoffs

SC cutoffs	Power law degree distribution	R^2	#miRNA sponge interactions
0.10	$P(k) = 27.353k^{-0.903}$	0.729	440
0.15	$P(k) = 26.903k^{-0.896}$	0.797	346
0.20	$P(k) = 28.009k^{-1.013}$	0.812	237
0.25	$\boldsymbol{P(k) = 26.127k^{-1.176}}$	**0.815**	**156**
0.30	$P(k) = 23.736k^{-1.278}$	0.770	73

FIGURE 3 | Power law degree distribution of the identified miRNA sponge interaction networks. **(A–E)** Node degree distribution of the identified miRNA sponge interaction networks using different SC cutoffs from 0.1 to 0.3 with a step of 0.05. **(F)** Summary table of power law degree distribution under different SC cutoffs.

using the well-cited *clusterProfiler* (Yu et al., 2012) R package. The third-party databases for functional enrichment analysis include Gene Ontology database (GO)[2], Kyoto Encyclopedia of Genes and Genomes Pathway database (KEGG)[3], Reactome Pathway database (Reactome)[4], Disease Ontology database[5], DisGeNET database[6], and Network of Cancer Genes database[7]. The enriched term (GO, KEGG, Reactome, Disease Ontology,

[2]http://www.geneontology.org/
[3]http://www.genome.jp/kegg
[4]http://reactome.org/
[5]http://disease-ontology.org/
[6]http://www.disgenet.org/
[7]http://ncg.kcl.ac.uk/



B **ASD-related enriched terms related to the target genes of hub miRNAs**

Category	Subcategory	Enrichment	*p*-adjusted	#miRNAs
Gene Ontology	GO0007612: learning	enriched	7.00E-04	10
Gene Ontology	GO0007420: brain development	enriched	1.75E-03	12
Gene Ontology	GO0007623: circadian rhythm	enriched	1.81E-03	11
Gene Ontology	GO0007409: axonogenesis	enriched	3.75E-03	10
Gene Ontology	GO0007050: cell cycle arrest	enriched	1.39E-02	12
Gene Ontology	GO0000165: MAPK cascade	enriched	3.45E-02	12
Gene Ontology	GO0007268: chemical synaptic transmission	enriched	3.55E-02	11
Gene Ontology	GO0050776: regulation of immune response	enriched	3.85E-02	10
Gene Ontology	GO0007399: nervous system development	enriched	4.06E-02	12
KEGG	hsa04064: NF-kappa B signaling pathway	enriched	3.87E-03	12
KEGG	hsa04730: Long-term depression	enriched	3.87E-03	11
KEGG	hsa04727: GABAergic synapse	enriched	4.12E-03	11
KEGG	hsa04370: VEGF signaling pathway	enriched	4.12E-03	11
KEGG	hsa04330: Notch signaling pathway	enriched	4.88E-03	10
KEGG	hsa04630: JAK-STAT signaling pathway	enriched	1.87E-02	12
KEGG	hsa04310: Wnt signaling pathway	enriched	1.95E-02	12
KEGG	hsa05226: Gastric cancer	enriched	3.32E-02	12

FIGURE 4 | Visualization and functional enrichment analysis of hub miRNA regulatory network. **(A)** Hub miRNA regulatory network. Green rhombic, pink circle, and purple triangle nodes denote miRNAs, mRNAs, and lncRNAs, respectively. **(B)** ASD-related enriched terms related to the target genes of hub miRNAs.

TABLE 1 | Disease and functional enriched terms of top 20 largest miRNA–target regulatory modules related to ASD.

Items	Descriptions	Module ID	Evidence
umls:C1378703	Renal carcinoma	1, 8, 9, 19	Lam et al., 2018
umls:C0003469	Anxiety disorders	19	White et al., 2009
GO:0000380	Alternative mRNA splicing, via spliceosome	1, 2, 4, 6	Quesnel-Vallieres et al., 2019
GO:0007623	Circadian rhythm	1, 8	Hu et al., 2017
GO:0120111	Neuron projection cytoplasm	1, 2	Guo et al., 2019
GO:0099640	Axodendritic protein transport	1, 2	Mandal and Drerup, 2019
hsa03040	Spliceosome	5, 11, 12, 15, 16, 17, 19	Kong et al., 2013
R-HSA-210500	Glutamate neurotransmitter release cycle	5, 14	Horder et al., 2018
R-HSA-165159	mTOR signaling	5	Khlebodarova et al., 2018

DisGeNET, or Network of Cancer Genes term) with adjusted $p < 0.05$ (adjusted by the Benjamini and Hochberg method) is regarded as a significantly enriched term.

RESULTS

MiRNA-Associated Networks Are Scale-Free Networks

We first follow Step 1 to obtain miRNA–target interactions predicted by each of the 12 computational methods (details in the section "Methods"). The aim of comparing the performance of these methods is to select the best prediction method to identify miRNA–target regulatory network in ASD. For each method, we select top 50, 100, 150, and 200 targets of each miRNA for the comparison. The method of predicting the largest number of experimentally validated miRNA–target interactions is used to identify miRNA–target regulatory network in ASD. As displayed in **Figure 2**, the ProMISe method performs the best in terms of the number of experimentally validated miRNA–target interactions. Thus, we merge top 200 targets of each miRNA identified by the ProMISe method as our final predicted miRNA–target regulatory network (consisting of 20,000 miRNA–lncRNA interactions and 20,000 miRNA–mRNA interactions). In total, we obtain a list of 1,679 validated miRNA–target interactions, consisting of 241 validated miRNA–lncRNA interactions and 1,438 validated miRNA–mRNA interactions. We further analyze the node degree distribution of the identified miRNA–target regulatory network using the Network Analyzer plugin (Assenov et al., 2008) in Cytoscape (Shannon, 2003), and discover that our identified miRNA–target regulatory network follows power law distribution well in the form of $P(k) = 67.593k^{-1.022}$ with $R^2 = 0.783$, where $P(k)$ represents the number of nodes with the node degree k. A higher R^2 (range from 0 to 1) indicates that the identified miRNA–target regulatory network is more likely to be a scale-free network that occurs in the real world. The detailed results of the identified miRNA–target regulatory network can be found in **Supplementary Table 3**.

By following Step 2, we use different *SC* cutoffs from 0.1 to 0.3 with a step of 0.05, to infer the miRNA sponge interaction network with better power law distribution. Under different *SC* cutoffs, we use R^2 value to evaluate the goodness of power law degree distribution for the identified miRNA sponge interaction network. If a miRNA sponge interaction network with higher R^2 value, the network is more likely to be a real biological network. As shown in **Figure 3**, according to the principle of

TABLE 2 | Disease and functional enriched terms of miRNA sponge modules related to ASD.

Module ID	Items	Description	Adjusted *p*-value
1	GO:0061437	Renal system vasculature development	3.06E-02
	GO:0071364	Cellular response to epidermal growth factor stimulus	3.06E-02
	GO:1903844	Regulation of cellular response to transforming growth factor β stimulus	4.09E-02
	R-HSA-380972	Energy dependent regulation of mTOR by LKB1-AMPK	1.36E-02
	R-HSA-165159	mTOR signaling	1.68E-02
	R-HSA-198933	Immunoregulatory interactions between a Lymphoid and a non-lymphoid cell	4.90E-02
2	DOID:10155	Intestinal cancer	3.75E-02
	umls:C1845055	α-Thalassemia/mental retardation syndrome, non-deletion type, X-linked	2.47E-02
	GO:0000380	Alternative mRNA splicing, via spliceosome	4.23E-02
3	GO:0034134	Toll-like receptor 2 signaling pathway	3.33E-04
	GO:0002758	Innate immune response-activating signal transduction	2.82E-02
	GO:0007616	Long-term memory	2.97E-02
	GO:0000380	Alternative mRNA splicing, via spliceosome	3.71E-02
	hsa03040	Spliceosome	2.61E-02
	R-HSA-5260271	Diseases of immune system	3.03E-02
	R-HSA-1236974	ER-phagosome pathway	4.66E-02
	R-HSA-168179	Toll-like receptor TLR1:TLR2 cascade	4.66E-02
4	GO:0000784	Nuclear chromosome, telomeric region	8.97E-03
	GO:0005912	Adhere junction	1.96E-02
5	GO:0099640	Axodendritic protein transport	4.86E-02
	GO:0042754	Negative regulation of circadian rhythm	4.86E-02
	GO:1900016	Negative regulation of cytokine production involved in inflammatory response	4.86E-02
	GO:0006658	Phosphatidylserine metabolic process	4.86E-02
	GO:0002534	Cytokine production involved in inflammatory response	4.86E-02
	R-HSA-8950505	Gene and protein expression by JAK-STAT signaling after Interleukin-12 stimulation	3.64E-02
6	GO:0000783	Nuclear telomere cap complex	4.43E-02
	R-HSA-1980145	Signaling by NOTCH2	3.66E-02
	R-HSA-177929	Signaling by EGFR	3.66E-02
	R-HSA-2644603	Signaling by NOTCH1 in Cancer	3.66E-02
	umls:C0267446	Acute gastroenteritis	4.81E-02
	umls:C0588008	Severe depression	4.81E-02
7	GO:0005930	Axoneme	3.05E-02
8	umls:C0027889	Hereditary sensory and autonomic neuropathies	1.19E-02
	umls:C0235025	Peripheral motor neuropathy	1.68E-02
	umls:C0151313	Sensory neuropathy	1.93E-02
	umls:C1270972	Mild cognitive disorder	3.32E-02
	GO:0007173	Epidermal growth factor receptor signaling pathway	4.00E-02
	GO:0038127	ERBB signaling pathway	4.25E-02
	GO:0002433	Immune response-regulating cell surface receptor signaling pathway involved in phagocytosis	4.25E-02
	GO:0038094	Fc-gamma receptor signaling pathway	4.25E-02
	hsa04144	Endocytosis	8.81E-03
9	DOID:0060116	Sensory system cancer	4.63E-02
	GO:0000380	Alternative mRNA splicing, via spliceosome	8.07E-03
	GO:0007050	Cell cycle arrest	3.34E-02
	GO:0099640	Axodendritic protein transport	3.44E-02
	GO:1904357	Negative regulation of telomere maintenance via telomere lengthening	4.29E-02
	GO:0032839	Dendrite cytoplasm	4.62E-02
	GO:0005925	Focal adhesion	4.62E-02
	hsa04218	Cellular senescence	1.90E-02
	hsa03040	Spliceosome	1.37E-02
	R-HSA-9617828	FOXO-mediated transcription of cell cycle genes	3.36E-02
10	DOID:0050735	X-linked disease	1.31E-02
	GO:0005160	Transforming growth factor β receptor binding	3.18E-02
	R-HSA-2173789	TGF-β receptor signaling activates SMADs	4.82E-02
	R-HSA-2029480	Fc-gamma receptor–dependent phagocytosis	4.82E-02

the largest R^2 value, we select the *SC* cutoff as 0.25 to infer miRNA sponge interaction network (containing 156 miRNA sponge interactions) that fits power law distribution well in the form of $P(k) = 26.127k^{-1.176}$ with $R^2 = 0.815$. The detailed results of the identified miRNA sponge interaction network can be found in **Supplementary Table 3**.

Hub Genes Are Closely Associated With ASD

In this work, we have identified 12 hub miRNAs (*hsa-miR-195-5p*, *hsa-miR-15a-5p*, *hsa-miR-26b-5p*, *hsa-miR-23a-3p*, *hsa-miR-93-5p*, *hsa-miR-210-3p*, *hsa-miR-25-3p*, *hsa-miR-30b-5p*, *hsa-miR-148b-3p*, *hsa-miR-149-5p*, *hsa-miR-200c-3p*, and *hsa-miR-147a*) and 15 hub miRNA sponges (*SLC38A2*, *SHOC2*, *DDX6*, *WSB1*, *PURB*, *DDX5*, *DLEU2*, *USP15*, *C6orf62*, *ADAM10*, *STK4*, *LBR*, *PNISR*, *ANKRD44*, and *SERINC1*). It is noted that four hub miRNAs (*hsa-miR-148b-3p*, *hsa-miR-15a-5p*, *hsa-miR-23a-3p*, and *hsa-miR-93-5p*) and two hub miRNA sponges (*DLEU2* and *USP15*) are experimentally validated ASD-related hub genes.

In **Figure 4A**, we discover that 12 hub miRNAs are highly connected with their target genes, and several hub miRNAs synergistically regulate their target genes. To investigate the underlying biological implications of these hub miRNAs, we conduct functional 1enrichment analysis of the target genes of these hub miRNAs. Functional enrichment analysis results show that 439 GO terms and 128 KEGG pathways are significantly associated with the target genes of the hub miRNAs. Moreover, several significantly enriched GO biological processes and KEGG pathways, including Cell cycle arrest (GO: 0007050), Regulation of immune response (GO: 0050776) (Gupta et al., 2014) (Ander et al., 2015), Nervous system development (GO:0007399) (Hu et al., 2017), NF-kappa B signaling pathway (hsa04064) (Malik et al., 2011), Long-term depression (hsa04730) (Monday et al., 2018), Wnt signaling pathway (hsa04310) (Shen et al., 2016), and gastric cancer (hsa05226) (Wasilewska and Klukowski, 2015) are closely associated with the progression and development of ASD (**Figure 4B**). As for the identified hub miRNA sponges, functional enrichment analysis results indicate that they are significantly enriched in SMAD binding (GO: 0046332). A previous study (Avazzadeh et al., 2019) has demonstrated that SMAD binding is closely related to ASD. Altogether, the above functional enrichment analysis results imply that the identified hub genes are closely associated with the occurrence and development of ASD. The functional enrichment analysis results of hub genes can be found in **Supplementary Table 4**.

The Identified MiRNA-Associated Modules Are Functional

Based on the identified miRNA–target network, we have identified 9,625 miRNA–target regulatory modules. In this work, we are only interested in studying the potential biological functions of top 20 largest miRNA–target regulatory modules. Moreover, we have obtained 10 miRNA sponge modules from the identified miRNA sponge interaction network. Disease and functional enrichment analysis indicate that the top 20 largest miRNA–target regulatory modules are significantly enriched in 397 GO terms, 3 KEGG pathways, 12 Reactome pathways, and 69 DisGeNET terms. Specifically, several biological processes, pathways, and diseases, including anxiety disorder (umls: C0003469) (White et al., 2009), Alternative mRNA splicing (GO: 0000380) (Quesnel-Vallieres et al., 2019), circadian rhythm (GO: 0007623) (Hu et al., 2017), and mTOR signaling pathway (R-HSA-165159) (Khlebodarova et al., 2018) are closely related to ASD (**Table 1**).

Furthermore, the identified 10 miRNA sponge modules are significantly enriched in 711 GO terms, 23 KEGG pathways, 117 Reactome pathways, 22 Disease Ontology terms, and 157 DisGeNET terms. In **Table 2**, several GO, KEGG, Reactome, Disease Ontology, DisGeNET, and Network of Cancer Genes terms are closely associated with ASD. For instance, severe depression (umls: C0588008) (Hedley et al., 2018), Mild cognitive disorder (umls: C1270972) (Leekam, 2016), acute gastroenteritis (umls: C0267446) (Wasilewska and Klukowski, 2015), Focal adhesion (GO: 0005925) (Ander et al., 2015), long-term memory (GO: 0007616) (Toichi and Kamio, 2002), and spliceosome (hsa03040) (Kong et al., 2013) are experimentally confirmed to be ASD-related terms. Taken together, the above enrichment analysis results indicate that the identified miRNA-associated modules are functional. The disease and functional enrichment analysis results of miRNA-associated modules can be seen in **Supplementary Table 5**.

DISCUSSION AND CONCLUSION

Given the high prevalence rate of ASD, it becomes more and more urgent to reveal the underlying molecular mechanisms associated with ASD. Growing evidence (Fregeac et al., 2016; Shen et al., 2016; Hu et al., 2017) has revealed that miRNA dysregulation has made a great contribution to the pathology of ASD. However, there is still a lack of computational methods to uncover miRNA regulation in ASD at both the network and module level.

In this work, we propose a stepwise method, ASDmiR, to reveal miRNA regulation in ASD. The comparison study suggests that the ProMISe method is the best miRNA target prediction method for identifying miRNA–target regulatory network in ASD dataset. Network topological analysis indicates that the identified miRNA–target network and miRNA sponge interaction network are all scale-free networks. Moreover, functional enrichment analysis shows that hub miRNAs and hub miRNA sponges are closely associated with ASD. As functional units, the identified miRNA-associated modules are found to be significantly enriched in several important ASD-related terms.

ASDmiR can be improved in the following aspects. First, ASD-related samples can be obtained from peripheral blood, post-mortem brain, gastrointestinal tissue, adult olfactory stem cells, and scalp hair follicles (Ansel et al., 2016). In future, we will apply ASDmiR into other types of ASD-related datasets. Second, we will conduct a more comprehensive comparison to identify miRNA–target regulatory network by considering more miRNA target prediction methods. Third, we will cover other types of miRNA sponges (e.g., circRNAs, pseudogenes) to further uncover the potential roles of miRNA sponges in ASD. Finally, a previous

study (Iakoucheva et al., 2019) has shown that *de novo* mutations (e.g., structure variants, protein-altering point mutations) and genetic variants (e.g., copy number variations, single nucleotide polymorphisms) could also contribute to the occurrence of ASD. Therefore, to further understand the molecular mechanisms of ASD, it is necessary to integrate these heterogeneous data to explore miRNA regulation.

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**. The source codes for this study can be found at: https://github.com/chenchenxiong/ASD.

AUTHOR CONTRIBUTIONS

CX and JZ designed the methods. CX wrote the codes. CX, SS, and WJ analyzed the data with supervision from LM and JZ. CX, SS, and JZ wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61702069 and 61963001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.562971/full#supplementary-material

Supplementary Table 1 | Differentially expressed miRNAs, lncRNAs, and mRNAs.

Supplementary Table 2 | miRNA-target interactions and ASD-related miRNAs, lncRNAs, and mRNAs collected in relevant databases.

Supplementary Table 3 | The Identified miRNA-target interactions and miRNA sponge interactions.

Supplementary Table 4 | The enriched GO terms and KEGG pathways associated with hub genes.

Supplementary Table 5 | The enriched disease and functional terms associated with miRNA-associated modules.

REFERENCES

Ambros, V. (2004). The functions of animal microRNAs. *Nature* 421, 350–355. doi: 10.1038/nature02871

Ander, B. P., Barger, N., Stamova, B., Sharp, F. R., and Schumann, C. M. (2015). Atypical miRNA expression in temporal cortex associated with dysregulation of immune, cell cycle, and other pathways in autism spectrum disorders. *Mol. Aut.* 6:37. doi: 10.1186/s13229-015-0029-9

Ansel, A., Rosenzweig, J. P., Zisman, P. D., Melamed, M., and Gesundheit, B. (2016). Variation in gene expression in autism spectrum disorders: an extensive review of transcriptomic studies. *Front Neurosci.* 10:601. doi: 10.3389/fnins.2016.00601

Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics* 24, 282–284. doi: 10.1093/bioinformatics/btm554

Avazzadeh, S., McDonagh, K., Reilly, J., Wang, Y., Boomkamp, S. D., McInerney, V., et al. (2019). Increased Ca2+ signaling in NRXN1α+/− neurons derived from ASD induced pluripotent stem cells. *Mol. Aut.* 10:52. doi: 10.1186/s13229-019-0303-3

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* 47, D1034–D1037. doi: 10.1093/nar/gky905

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., et al. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147, 358–369. doi: 10.1016/j.cell.2011.09.028

Chen, J. A., Penagarikano, O., Belgard, T. G., Swarup, V., and Geschwind, D. H. (2015). The emerging picture of autism spectrum disorder: genetics and pathology. *Annu. Rev. Pathol.* 10, 111–144. doi: 10.1146/annurev-pathol-012414-040405

Choobdar, S., Ahsen, M. E., Crawford, J., Tomasoni, M., Fang, T., Lamparter, D., et al. (2019). Assessment of network module identification across complex diseases. *Nat. Methods* 16, 843–852. doi: 10.1038/s41592-019-0509-5

Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., et al. (2018). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46, D296–D302. doi: 10.1093/nar/gkx1067

Cui, T., Zhang, L., Huang, Y., Yi, Y., Tan, P., Zhao, Y., et al. (2018). MNDR v2.0: an updated resource of ncRNA-disease associations in mammals. *Nucleic Acids Res.* 46, D371–D374. doi: 10.1093/nar/gkx1025

Enright, J., Van Donges, S., and Bouzoukis, C. A. (2002). An efficient algorithm for largescale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575

Fregeac, J., Colleaux, L., and Nguyen, L. S. (2016). The emerging roles of MicroRNAs in autism spectrum disorders. *Neurosci. Biobehav. Rev.* 71, 729–738. doi: 10.1016/j.neubiorev.2016.10.018

Gillian, B., Hilary, C. S., and Slonims, V. (2003). Diagnosis of autism. *BMJ* 327, 488–493. doi: 10.1136/bmj.327.7413.488

Guo, H., Li, Y., Shen, L., Wang, T., Jia, X., Liu, L., et al. (2019). Disruptive variants of CSDE1 associate with autism and interfere with neuronal development and synaptic transmission. *Sci. Adv.* 5:eaax2166. doi: 10.1126/sciadv.aax2166

Gupta, S., Ellis, S. E., Ashar, F. N., Moes, A., Bader, J. S., Zhan, J., et al. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* 5:5748. doi: 10.1038/ncomms6748

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993

Hedley, D., Uljarevic, M., Foley, K. R., Richdale, A., and Trollor, J. (2018). Risk and protective factors underlying depression and suicidal ideation in Autism Spectrum Disorder. *Depress Anxiety* 35, 648–657. doi: 10.1002/da.22759

Hoeffding, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* 19, 546–557. doi: 10.2307/2236021

Horder, J., Petrinovic, M. M., Mendez, M. A., Bruns, A., Takumi, T., Spooren, W., et al. (2018). Glutamate and GABA in autism spectrum disorder—a translational magnetic resonance spectroscopy study in man and rodent models. *Transl. Psychiatry* 8:106. doi: 10.1038/s41398-018-0155-1

Hosseini, E., Bagheri-Hosseinabadi, Z., De Toma, I., Jafarisani, M., and Sadeghi, I. (2019). The importance of long non-coding RNAs in neuropsychiatric disorders. *Mol. Aspects Med.* 70, 127–140. doi: 10.1016/j.mam.2019.07.004

Hu, Y., Ehli, E. A., and Boomsma, D. I. (2017). MicroRNAs as biomarkers for psychiatric disorders with a focus on autism spectrum disorder: current

progress in genetic association studies, expression profiling, and translational research. *Autism Res.* 10, 1184–1203. doi: 10.1002/aur.1789

Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017. doi: 10.1093/nar/gky1010

Iakoucheva, L. M., Muotri, A. R., and Sebat, J. (2019). Getting to the cores of autism. *Cell* 178, 1287–1298. doi: 10.1016/j.cell.2019.07.037

Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., et al. (2018). DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.* 46, D239–D245. doi: 10.1093/nar/gkx1141

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* 30, 81–93. doi: 10.1093/biomet/30.1-2.81

Kerin, T., Ramanathan, A., Rivas, K., Grepo, N., and Campbell, D. B. (2012). A noncoding RNA antisense to moesin at 5p14.1 in autism. *Sci. Transl. Med.* 4:128ra40. doi: 10.1126/scitranslmed.3003479

Kern, F., Fehlmann, T., Solomon, J., Schwed, L., Backes, C., Meese, E., et al. (2020). miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. *Nucleic Acids Res.* 48, W521–W528. doi: 10.1101/2020.03.05.978890

Khlebodarova, T. M., Kogai, V. V., Trifonova, E. A., and Likhoshvai, V. A. (2018). Dynamic landscape of the local translation at activated synapses. *Mol. Psychiatry* 23, 107–114. doi: 10.1038/mp.2017.245

Kong, S. W., Collins, C. D., Shimizu-Motohashi, Y., Holm, I. A., Campbell, M. G., Lee, I. H., et al. (2012). Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS One* 7:e49475. doi: 10.1371/journal.pone.0049475

Kong, S. W., Shimizu-Motohashi, Y., and Campbell, M. G. (2013). Peripheral blood gene expression signature differentiateschildren with autism from unaffected siblings. *Neurogenetics* 14, 143–152. doi: 10.1007/s10048-013-0363-z

Lam, H. C., Siroky, B. J., and Henske, E. P. (2018). Renal disease in tuberous sclerosis complex: pathogenesis and therapy. *Nat. Rev. Nephrol.* 14, 704–716. doi: 10.1038/s41581-018-0059-6

Le, T. D., Zhang, J., Liu, L., and Li, J. (2017). Computational methods for identifying miRNA sponge interactions. *Brief Bioinform.* 18, 577–590. doi: 10.1093/bib/bbw042

Le, T. D., Zhang, J., Liu, L., Liu, H., and Li, J. (2015). miRLAB: an R based dry lab for exploring miRNA-mRNA regulatory relationships. *PLoS One* 10:e0145386. doi: 10.1371/journal.pone.0145386

Leekam, S. (2016). Social cognitive impairment and autism: what are we trying to explain? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371:20150082. doi: 10.1098/rstb.2015.0082

Li, Y., Liang, C., Wong, K. C., Jin, K., and Zhang, Z. (2014). Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. *Nucleic Acids Res.* 42:e76. doi: 10.1093/nar/gku182

Lopez-Paz, D., Hennig, P., and Schölkopf, B. (2013). The randomized dependence coefficient. *Adv. Neural Inform. Process. Syst.* 1, 1–9.

Maathuis, M. H., Kalisch, M., and Bühlmann, P. (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Stat.* 37, 3133–3164. doi: 10.1214/09-aos685

Malik, M., Tauqeer, Z., Sheikh, A. M., Wen, G., Nagori, A., Yang, K., et al. (2011). NF-kappaB signaling in the brain of autistic subjects. *Mediat. Inflamm.* 2011:785265. doi: 10.1155/2011/785265

Mandal, A., and Drerup, C. M. (2019). Axonal transport and mitochondrial function in neurons. *Front. Cell Neurosci.* 13:373. doi: 10.3389/fncel.2019.00373

Mohr, S., and Liew, C. C. (2007). The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends Mol. Med.* 13, 422–432. doi: 10.1016/j.molmed.2007.08.003

Monday, H. R., Younts, T. J., and Castillo, P. E. (2018). Long-term plasticity of neurotransmitter release: emerging mechanisms and contributions to brain function and disease. *Annu. Rev. Neurosci.* 41, 299–322. doi: 10.1146/annurev-neuro-080317-062155

Moon, Y. I., Rajagopalan, B., and Lall, U. (1995). Estimation of mutual information using kernel density estimators. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* 52, 2318–2321. doi: 10.1103/physreve.52.2318

Noor, A., Whibley, A., Marshall, C. R., Gianakopoulos, P. J., Piton, A., Carson, A. R., et al. (2010). Disruption at the PTCHD1 locus on Xp22.11 in autism spectrum disorder and intellectual disability. *Sci. Transl. Med.* 2:49ra68. doi: 10.1126/scitranslmed.3001267

Paci, P., Colombo, T., and Farina, L. (2014). Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst. Biol.* 8:83. doi: 10.1186/1752-0509-8-83

Paraskevopoulou, M. D., Vlachos, I. S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., et al. (2016). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.* 44, D231–D238. doi: 10.1093/nar/gkv1270

Parikshak, N. N., Swarup, V., Belgard, T. G., Irimia, M., Ramaswami, G., Gandal, M. J., et al. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540, 423–427. doi: 10.1038/nature20612

Pearson, K. (1920). Notes on the history of correlation. *Biometrika* 13, 25–45. doi: 10.1093/biomet/13.1.25

Pinero, J., Ramirez-Anguita, J. M., Sauch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855. doi: 10.1093/nar/gkz1021

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038. doi: 10.1038/nature09144

Prill, R. J., Marbach, D., Saez-Rodriguez, J., Sorger, P. K., Alexopoulos, L. G., Xue, X., et al. (2010). Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PLoS One* 5:e9202. doi: 10.1371/journal.pone.0009202

Quesnel-Vallieres, M., Weatheritt, R. J., Cordes, S. P., and Blencowe, B. J. (2019). Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. *Nat. Rev. Genet.* 20, 51–63. doi: 10.1038/s41576-018-0066-2

Rajman, M., and Schratt, G. (2017). MicroRNAs in neural development: from master regulators to fine-tuners. *Development* 144, 2310–2322. doi: 10.1242/dev.144337

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:e47. doi: 10.1093/nar/gkv007

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Segura, M., Pedreno, C., Obiols, J., Taurines, R., Pamias, M., Grunblatt, E., et al. (2015). Neurotrophin blood-based gene expression and social cognition analysis in patients with autism spectrum disorder. *Neurogenetics* 16, 123–131. doi: 10.1007/s10048-014-0434-9

Shannon, P. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shen, L., Lin, Y., Sun, Z., Yuan, X., Chen, L., and Shen, B. (2016). Knowledge-guided bioinformatics model for identifying autism spectrum disorder diagnostic MicroRNA biomarkers. *Sci. Rep.* 6:39663. doi: 10.1038/srep39663

Shen, L., Zhao, Y., Zhang, H., Feng, C., Gao, Y., Zhao, D., et al. (2019). Advances in biomarker studies in autism spectrum disorders. *Adv. Exp. Med. Biol.* 1118, 207–233. doi: 10.1007/978-3-030-05542-4_11

Spearman, C. (1904). general intelligence, objectively determined and measured. *Am. J. Psychol.* 15, 201–292. doi: 10.2307/1412107

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Stat.* 35, 2769–2794. doi: 10.1214/009053607000000505

Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S. M., Ala, U., et al. (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 147, 344–357. doi: 10.1016/j.cell.2011.09.029

Teng, X., Chen, X., Xue, H., Tang, Y., Zhang, P., Kang, Q., et al. (2020). NPInter v4.0: an integrated database of ncRNA interactions. *Nucleic Acids Res.* 48, D160–D165. doi: 10.1093/nar/gkz969

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Toichi, M., and Kamio, Y. (2002). Long-term memory and levels-of-processing in autism. *Neurogenetics* 40, 964–969. doi: 10.1016/s0028-3932(01)00163-4

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384. doi: 10.1038/nature10110

Wang, Y., Zhao, X., Ju, W., Flory, M., Zhong, J., Jiang, S., et al. (2015). Genome-wide differential expression of synaptic long noncoding RNAs in autism spectrum disorder. *Transl. Psychiatry.* 5:e660. doi: 10.1038/tp.2015.144

Wasilewska, J., and Klukowski, M. (2015). Gastrointestinal symptoms and autism spectrum disorder: links and risks - a possible new overlap syndrome. *Pediatr. Health Med. Ther.* 6, 153–166. doi: 10.2147/PHMT.S85717

White, S. W., Oswald, D., Ollendick, T., and Scahill, L. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Clin. Psychol. Rev.* 29, 216–229. doi: 10.1016/j.cpr.2009.01.003

Xu, T., Su, N., Liu, L., Zhang, J., Wang, H., Zhang, W., et al. (2018). miRBaseConverter: an R/Bioconductor Package for Converting and Retrieving miRNA Name, Accession, Sequence and Family Information in Different Versions of miRBase. *BMC Bioinform.* 19(Suppl. 19):514. doi: 10.1101/407148

Yoon, S., Nguyen, H. C. T., Jo, W., Kim, J., Chi, S. M., Park, J., et al. (2019). Biclustering analysis of transcriptome big data identifies condition-specific microRNA targets. *Nucleic Acids Res.* 47:e53. doi: 10.1093/nar/gkz139

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, J., Liu, L., Li, J., and Le, T. D. (2018). LncmiRSRN: identification and analysis of long non-coding RNA related miRNA sponge regulatory network in human cancer. *Bioinformatics* 34, 4232–4240. doi: 10.1093/bioinformatics/bty525

Zhang, J., Liu, L., Xu, T., Xie, Y., Zhao, C., Li, J., et al. (2019). miRspongeR: an R/Bioconductor package for the identification and analysis of miRNA sponge interaction networks and modules. *BMC Bioinform.* 20:235. doi: 10.1186/s12859-019-2861-y

Zhang, Y., Phillips, C. A., Rogers, G. L., Baker, E. J., Chesler, E. J., and Langston, M. A. (2014). On finding bicliques in bipartite graphs: a novel algorithm and its application to the integration of diverse biological data types. *BMC Bioinform.* 15:110. doi: 10.1186/1471-2105-15-110

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

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: 10 November 2020
doi: 10.3389/fgene.2020.555865

Identification of Tumor Microenvironment-Related Prognostic Biomarkers in Luminal Breast Cancer

*Yanyan Wang[1], Mingzhi Zhu[1], Feng Guo[2], Yi Song[2], Xunjie Fan[2] and Guijun Qin[2]**

[1]*Department of Breast Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China,* [2]*Department of Endocrinology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China*

Background: The tumor microenvironment (TME) has been reported to have significant value in the diagnosis and prognosis of cancers. This study aimed to identify key biomarkers in the TME of luminal breast cancer (BC).

Methods: We obtained immune scores (ISs) and stromal scores (SSs) for The Cancer Genome Atlas (TCGA) luminal BC cohort from the online ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) portal. The relationships between ISs and SSs and the overall survival of luminal BC patients were assessed by the Kaplan-Meier method. The differentially expressed messenger RNAs (DEmRNAs) related to the ISs and SSs were subjected to functional enrichment analysis. Additionally, a competing endogenous RNA (ceRNA) network was constructed with differentially expressed microRNAs (DEmiRNAs) and long noncoding RNAs (DElncRNAs). Furthermore, a protein–protein interaction (PPI) network was established to analyze the DEmRNAs in the ceRNA network. Then, survival analysis of biomarkers involved in the ceRNA network was carried out to explore their prognostic value. Finally, these biomarkers were validated using the luminal BC dataset from the Gene Expression Omnibus (GEO) database.

Results: The results showed that ISs were significantly associated with longer survival times of luminal BC patients. Functional enrichment analysis showed that the DEmRNAs were mainly associated with immune response, antigen binding, and the extracellular region. In the PPI network, the top 10 DEmRNAs were identified as hub genes that affected the TME of luminal BC. Finally, two DEmiRNAs, two DElncRNAs, and 17 DEmRNAs of the ceRNA network associated with the TME were shown to have prognostic value. Subsequently, the expression of 15 prognostic biomarkers was validated in one additional dataset (GSE81002). In particular, one lncRNA (GVINP1) and five mRNAs (CCDC69, DOCK2, IKZF1, JCHAIN, and NCKAP1L) were novel biomarkers.

Conclusions: Our studies demonstrated that ISs were associated with the survival of luminal BC patients, and a set of novel biomarkers that might play a prognostic role in the TME of luminal BC was identified.

Keywords: luminal breast cancer, tumor microenvironment, ESTIMATE algorithm, messenger RNA, microRNA, long noncoding RNA

OPEN ACCESS

Edited by:
Cheng Peng,
Yunnan University, China

Reviewed by:
Manal Said Fawzy,
Suez Canal University, Egypt
Meng Zhou,
Wenzhou Medical University, China

****Correspondence:***
Guijun Qin
hyqingj@zzu.edu.cn

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

Received: *26 April 2020*
Accepted: *23 September 2020*
Published: *10 November 2020*

Citation:
Wang Y, Zhu M, Guo F, Song Y, Fan X and Qin G (2020) Identification of Tumor Microenvironment-Related Prognostic Biomarkers in Luminal Breast Cancer. Front. Genet. 11:555865. doi: 10.3389/fgene.2020.555865

INTRODUCTION

Breast cancer (BC) is one of the most common malignant tumors and the main cause of cancer-associated death in women worldwide. Treatment measures for BC patients, such as surgical methods and drug regimens, have been constantly improved, while the clinical outcomes of individual patients remain difficult to predict (Clough et al., 2010; Graham et al., 2015). BC is a heterogeneous disease in terms of cellular composition, molecular alterations, and clinical outcomes within different tumor subtypes and within a single tumor. BC is commonly categorized by gene expression profiling into four main subtypes: luminal A, luminal B, human epidermal growth factor receptor 2 (HER2)-enriched, and triple-negative (TN; Lorona et al., 2019). More than 70% of the diagnosed BC cases are the luminal subtype and positive for estrogen receptor and/or progesterone receptor (ER+ and PR+, respectively; Polyak and Metzger Filho, 2012). Although the recurrent risk of luminal BC patients can be reduced by endocrine therapy, some patients still relapse after 5 years.

In recent years, an increasing number of studies have focused on the tumor microenvironment, which consists of immune cells, stromal cells, endothelial cells, mesenchymal cells, inflammatory mediators, extracellular matrix molecules, and numerous cytokines and chemokines (Hanahan and Coussens, 2012; De Nola and Menga, 2019). Among them, immune cells and stromal cells function as two essential components of the tumor microenvironment (TME) and have been reported to have significant value in the diagnosis and prognosis of cancers. ESTIMATE (Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data) is an algorithm that can be used to predict the infiltration of stromal and immune cells and their prognostic value by accessing the stromal scores (SSs) and immune scores (ISs) in tumors based on RNA sequencing (RNA-Seq) data extracted through The Cancer Genome Atlas (TCGA) database (Yoshihara et al., 2013). To date, several studies have investigated the microenvironment of many cancers, including BC and TNBC (Deng and Lu, 2019; Wang et al., 2019). For luminal BC, Li and Sang (2020) mainly investigated DNA methylation and key genes that regulate immune cell infiltration in luminal BC. Furthermore, the roles of competing endogenous RNAs (ceRNAs) [long noncoding RNAs (lncRNAs) sharing microRNA (miRNA) response elements (MREs) with messenger RNAs (mRNAs)] associated with the ISs and/or SSs of luminal BC have not been investigated.

LncRNAs are defined as RNA transcripts of more than 200 nucleotides and were once considered transcriptional "noise" without protein-coding capacity (Jathar et al., 2017). They play a fundamental role in the interaction between cancer cells and the surrounding environment during tumor progression (Sun et al., 2018). Recent studies have suggested that lncRNAs, as regulators, play important roles in cancer immune processes, such as infiltration into cancer tissues, immune activation, and immune cell migration (Carpenter and Fitzgerald, 2018; Denaro et al., 2019). For example, hepatoma cell lines overexpressing the lncRNA HOTAIR secreted more C-C motif chemokine ligand 2 (CCL2) than control cell lines, which promoted the proliferation of tumor-associated macrophages and myeloid-derived suppressor cells (Botti et al., 2019). In addition, studies have also shown that an immune-related lncRNA signature has important value for survival prediction in several cancers, including diffuse large B cell lymphoma, hepatocellular carcinoma, non-small-cell lung cancer (NSCLC), and BC (Zhou et al., 2017; Shen et al., 2020; Sun et al., 2020; Zhang et al., 2020).

In the current study, by taking advantage of the RNA-Seq and clinical data of 566 luminal BC tumor samples from TCGA database and the ESTIMATE algorithm, we aimed to explore the association between overall survival and ISs and SSs of luminal BC and construct a ceRNA network of differentially expressed mRNAs (DEmRNAs), lncRNAs (DElncRNAs), and miRNAs (DEmiRNAs) related to the ISs and SSs. Then, we compiled a list of TME-related biomarkers that were associated with the survival of luminal BC patients. Finally, these biomarkers were validated using the luminal BC dataset from the Gene Expression Omnibus (GEO) database. Thus, our study might help better understand the molecular mechanisms in the TME of luminal BC and identify potential immune biomarkers that predict the survival outcomes of patients.

MATERIALS AND METHODS

Data Processing and Estimation of Immune Scores and Stromal Scores

In order to provide an overall estimation of the molecular mechanisms in the TME of luminal BC, we combined RNA-Seq datasets and clinical information of luminal A and luminal B patients from TCGA database.[1] The inclusion criteria were as follows: (1) patients who had no other tumors, (2) samples with mRNA, miRNA, and lncRNA sequencing data, and (3) luminal A and luminal B subtypes classified by PAM50 gene expression profile. Finally, 556 luminal BC patients were finally included in this study. The ESTIMATE algorithm was used to estimate the ISs and SSs of each luminal BC sample. The gene expression profiles of the GSE81002 cohort were downloaded from the GEO database for validation.[2]

Association Between Immune Scores/ Stromal Scores and the Prognoses of Luminal Breast Cancer Patients

According to each sample IS and SS, all luminal BC patients were classified into high- and low-score groups. The overall survival of these two groups was estimated with the Kaplan-Meier survival estimator, and the corresponding survival outcomes of the two groups were compared by log-rank tests.

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

Identification of Differentially Expressed mRNAs, miRNAs, and lncRNAs Based on Immune Scores and Stromal Scores

By comparing the two groups above, DEmRNAs, DEmiRNAs, and DElncRNAs were filtered with the cutoff criteria of $p < 0.05$, false discovery rate (FDR) <0.05, and fold change (FC) >1.2 by using the R package limma. Venn diagrams were generated to show the intersection between the DEmRNAs, DEmiRNAs, and DElncRNAs of both the IS and SS groups.

Functional Enrichment Analysis

We used the Database for Annotation, Visualization, and Integrated Discovery (DAVID) online tool to perform Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to explore the functional roles of intersecting DEmRNAs (Ashburner et al., 2000; Kanehisa, 2002).[3] The GO terms are grouped by biological processes (BPs), molecular functions (MFs), and cellular components (CCs). $p < 0.05$ was considered statistically significant.

Construction of the Competing Endogenous RNA and Protein: Protein Interaction Networks

The target genes and lncRNAs of the overlapping DEmiRNAs were predicted. The target genes were predicted using the MiRanda and TargetScan algorithms, and the target lncRNAs were predicted through the MiRanda and PITA algorithms. Out of all the predictions, only the target mRNAs and lncRNAs predicted in both algorithms were considered candidates, and the common mRNAs and lncRNAs with the IS/SS-associated overlapping DEmRNAs and DElncRNAs mentioned above were chosen to construct the ceRNA network.

The DEmRNAs involved in the ceRNA network mentioned above were selected to construct the protein–protein interaction (PPI) network by using the Search Tool for the Retrieval of Interacting Genes (STRING) database.[4] The resulting network was visualized with Cytoscape software (version 3.6.1).

Prognostic Value of Biomarkers Involved in the Competing Endogenous RNA Network

Based on their median survival time, luminal BC patients were divided into high-risk and low-risk groups. Kaplan-Meier survival analysis and the log-rank test were used to assess the relationship between patient overall survival and the expression levels of DEmiRNAs, DEmRNAs, and DElncRNAs that were involved in the ceRNA network. $p < 0.05$ was considered statistically significant.

Statistical Analysis

R version 3.5.0 was used to conduct statistical analyses. The figures were generated using related packages including limma, heatmap, survival, and so on. $p < 0.05$ was considered statistically significant.

RESULTS

The Prognostic Value of Immune Scores and Stromal Scores in Luminal Breast Cancer Patients

The RNA sequencing datasets downloaded from TCGA database consisted of 556 luminal BC samples, of which 381 were luminal A and 175 were luminal B samples. By using the ESTIMATE algorithm, we found that the ISs of the patients ranged from −1,276.49 to 3,801.82, and the SSs ranged from −2,204.39 to 2,216.92.

Additionally, to determine the association between ISs and SSs and the overall survival time of luminal BC patients, the patients were divided into high- and low-score groups. Then, the patient data were analyzed by Kaplan-Meier survival analysis. The results indicated that patients with high ISs had longer survival times than patients with low ISs ($p = 0.015$; **Figure 1A**). However, there was no difference between the high-SS group and the low-SS group (**Figure 1B**).

Identification of Differentially Expressed mRNAs, miRNAs, and lncRNAs Based on Immune Scores and Stromal Scores

In total, 882 and 1,063 DEmRNAs, 18 and 38 DEmiRNAs, and 214 and 342 DElncRNAs were identified based on the comparison of the high- vs. low-IS and high- vs. low-SS groups, respectively (**Figures 2A–F**). Moreover, through Venn diagram constructions, 593 common DEmRNAs (577 upregulated and 16 downregulated), 11 common DEmiRNAs (eight upregulated and three downregulated), and 152 common DElncRNAs (93 upregulated and 59 downregulated) were obtained (**Figures 3A–C**). Subsequently, these common DEmRNAs, DEmiRNAs, and DElncRNAs were used for further analysis.

Functional Enrichment Analysis

Among the common DEmRNAs, only 16 DEmRNAs were downregulated; thus, we only chose 577 co-upregulated DEmRNAs for biological function exploration, and the data are presented in three subontologies: BP, CC, and MF. With regard to the BP category, the upregulated DEmRNAs were mainly enriched in immune response, complement activation, and leukocyte migration. For the MF category, the upregulated DEmRNAs were mainly involved in antigen binding, serine-type endopeptidase activity, and protein binding. The upregulated DEmRNAs were mainly associated with the extracellular region, extracellular space, and plasma membrane terms in the CC category. In addition, according to the pathway analysis, the upregulated DEmRNAs were mainly associated with *Staphylococcus aureus* infection, phagosomes, and cell adhesion molecules (CAMs). The top 10 GO and KEGG pathway terms of the co-upregulated DEmRNAs are shown in **Figure 3D**.

[3]https://david.ncifcrf.gov/
[4]https://string-db.org/

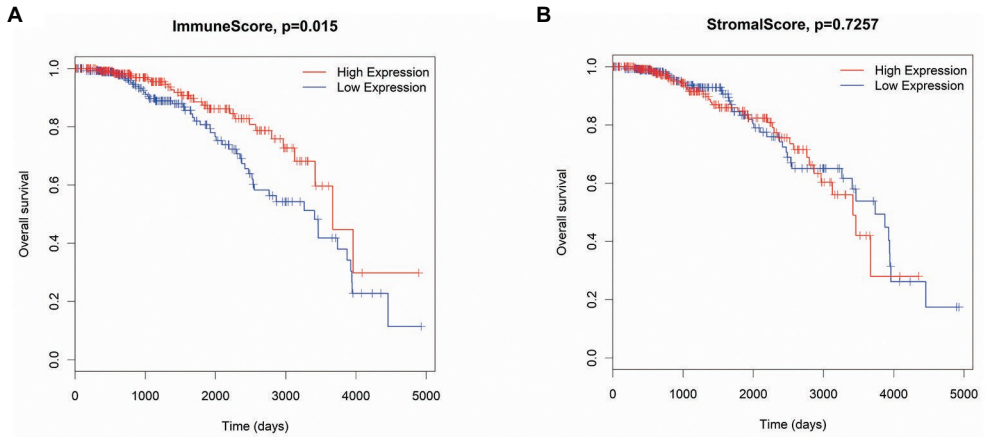


FIGURE 1 | The association between immune scores (ISs; **A**) and stromal scores (SSs; **B**) and overall survival in luminal breast cancer (BC) patients.

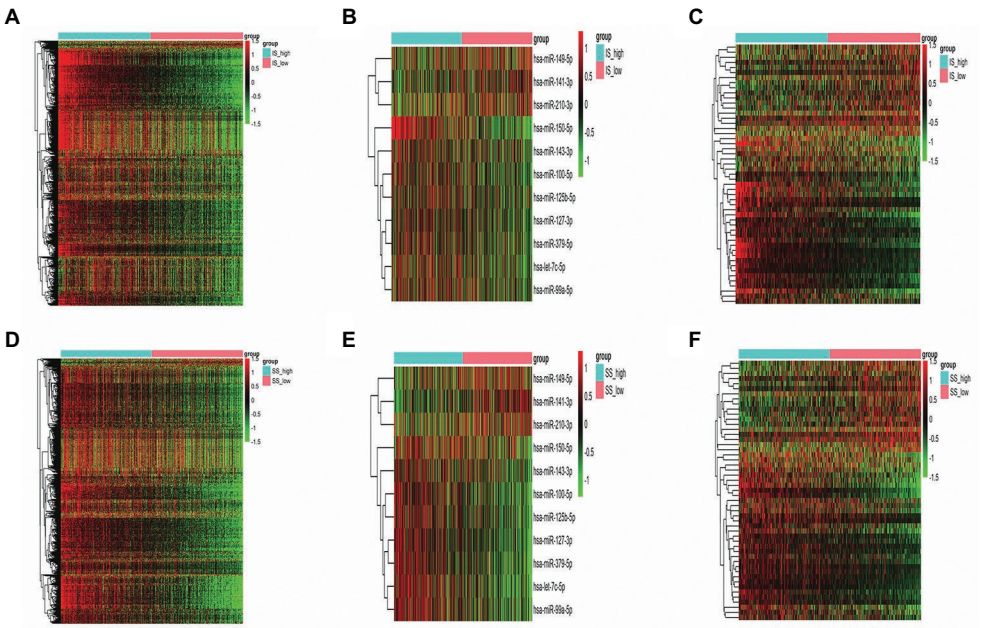


FIGURE 2 | (A–C) Heatmaps of differentially expressed messenger RNAs (DEmRNAs), microRNAs (DEmiRNAs), and long noncoding RNAs (DElncRNAs) in the high- vs. low-immune score (IS) groups. **(D–F)** Heatmaps of DEmRNAs, DEmiRNAs, and DElncRNAs in the high- vs. low-stromal score (SS) groups. Green represents high expression, and red represents low expression.

Construction of the Competing Endogenous RNA and Protein: Protein Interaction Networks

Based on the negatively regulated miRNA-mRNA and miRNA-lncRNA pairs, a ceRNA network was constructed with 99 DEmRNAs, nine DEmiRNAs, and 49 DElncRNAs (**Figure 4**). The nine DEmiRNAs (hsa-miR-149-5p, hsa-miR-141-3p, hsa-let-7c-5p, hsa-miR-210-3p, hsa-miR-125b-5p, hsa-miR-150-5p, hsa-miR-127-3p, hsa-miR-143-3p, and hsa-miR-379-5p) were highly correlated with other DEmRNAs and DElncRNAs. For instance, hsa-miR-149-5p, which acted as a ceRNA, was correlated with 65 DEmRNAs and 25 DElncRNAs. This means that the expression of these hub miRNAs might be regulated by a number of lncRNAs and mRNAs to affect the TME of luminal BC.

Next, to better understand the interactions among DEmRNAs of the ceRNA network, a PPI network was constructed with 75 nodes and 215 edges. The top 10 DEmRNAs (CD86, LCP2, CXCL12, MMP9, CD48, CCR1, IKZF1, PLEK, LOX, and ESR1) with the highest degree of connectivity were identified as hub genes in the PPI network (**Figure 5**).

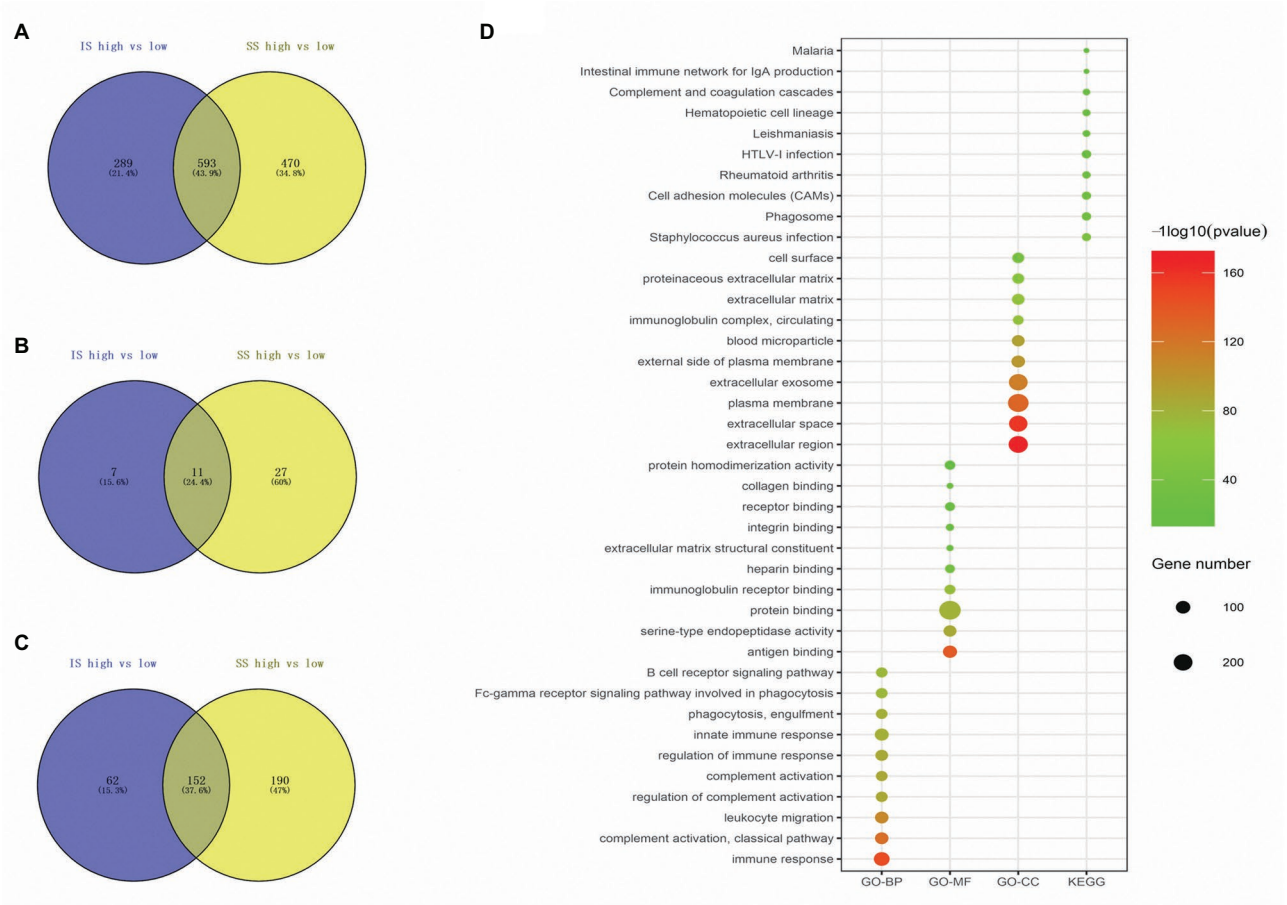


FIGURE 3 | (A–C) Venn diagrams showing the number of commonly regulated differentially expressed messenger RNAs (DEmRNAs), microRNAs (DEmiRNAs), and long noncoding RNAs (DElncRNAs) in the immune score (IS) and stromal score (SS) groups. **(D)** Biological process (BP), molecular function (MF), cellular component (CC), and pathway terms of the top 10 upregulated DEmRNAs.

Prognostic Values of Biomarkers Involved in the Competing Endogenous RNA Network

Ninety-nine DEmRNAs, nine DEmiRNAs, and 49 DElncRNAs in the ceRNA network were included in the survival analysis. Two DElncRNAs (GVINP1 and PCED1B-AS1), two DEmiRNAs (hsa-let-7c-5p and hsa-miR-150-5p), and 17 DEmRNAs (BIN2, CCDC69, CCR4, CD3E, CD5, CD48, DOCK2, F2RL2, HLA-E, IKZF1, JCHAIN, LRRC15, NCKAP1L, PIK3CD, SFRP1, SPN, and TNFAIP3) were found to be closely related to the overall survival of luminal BC patients ($p < 0.05$; **Figure 6**). Of these survival-associated biomarkers, only the high expression of LRRC15 was related to unfavorable survival outcomes of luminal BC patients. For the rest of the biomarkers, high expression levels were associated with favorable survival outcomes.

Validation in the Gene Expression Omnibus Database

To determine whether these survival-associated biomarkers were also of prognostic significance in an independent database, a cohort of 247 luminal BC cases was downloaded from the GEO database (GSE81002). However, due to the lack of follow-up information, we only validated their differential expression levels between the high- and low-IS/SS groups. The results showed that the expression levels of the lncRNA GVINP1 and 14 mRNAs (BIN2, CCDC69, CCR4, CD5, CD48, DOCK2, HLA-E, IKZF1, JCHAIN, LRRC15, NCKAP1L, PIK3CD, SPN, and TNFAIP3) were consistent in the two comparisons (all $p < 0.05$), which was consistent with the results for TCGA cohort (**Supplementary Figures 1**, **2**). In particular, the lncRNA GVINP1 and five mRNAs (CCDC69, DOCK2, IKZF1, JCHAIN, and NCKAP1L) were identified as novel biomarkers.

DISCUSSION

An increasing amount of evidence has elucidated the clinicopathological significance of the TME in the prediction of treatment effects (Binnewies et al., 2018; Zeng et al., 2019). Evaluating the proportions of microenvironment-associated cell types may help explore the role of TME and provide

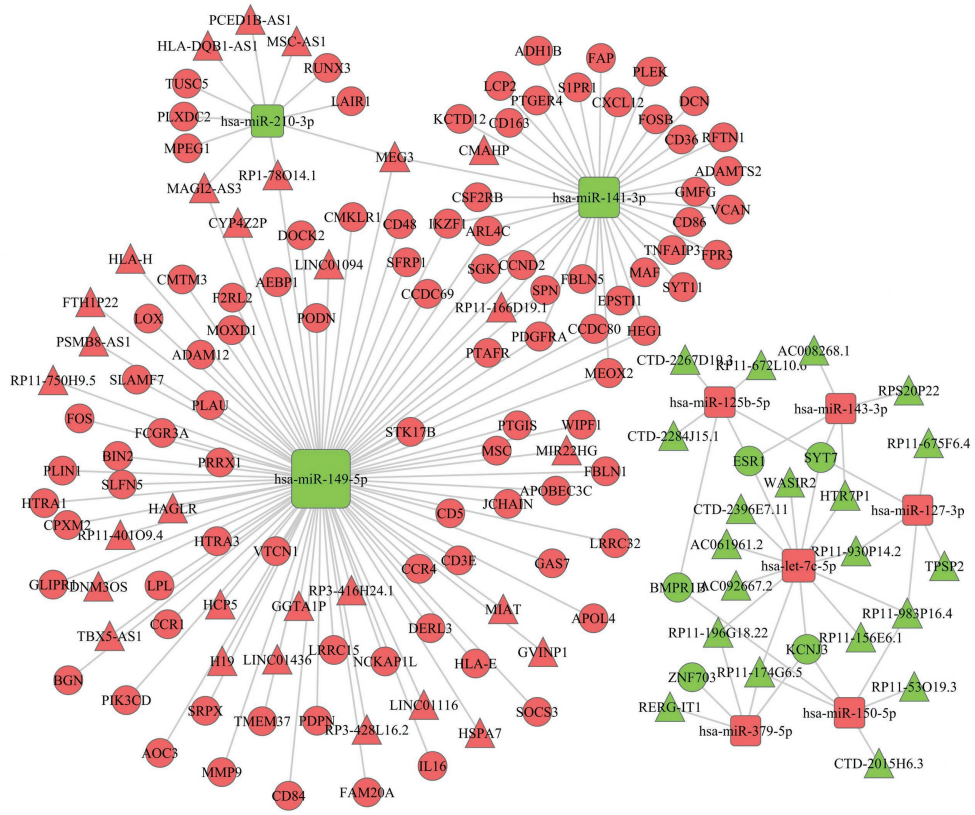


FIGURE 4 | The competing endogenous RNA (ceRNA) network. Red represents upregulation, and green represents downregulation. The circle nodes represent differentially expressed messenger RNAs (DEmRNAs), the rectangle nodes denote differentially expressed microRNAs (DEmiRNAs), and the triangle nodes represent differentially expressed long noncoding RNAs (DElncRNAs).

perspectives in cancer research. Recently, researchers have suggested the clinical importance of stromal cells and immune cells in the microenvironment of BC tissues (Wang et al., 2019; Li and Sang, 2020). However, the roles of ceRNAs associated with the ISs and/or SSs of luminal BC have not been investigated. The ceRNA hypothesis suggests a novel regulatory mechanism that can be mediated by lncRNAs, and lncRNAs with sequences similar to those of their target miRNAs are able to regulate the expression of mRNAs by acting as sponges for miRNAs (Salmena et al., 2011).

In this study, we evaluated the ISs and SSs of luminal BC patients. Our results showed that high ISs were significantly associated with longer survival time of patients than low ISs, which was consistent with a previous study that reported on IS evaluation in BC patients (Wang et al., 2019). This might be because higher ISs suggest enhanced immune system and function, which can be stimulated to increase the antitumor immunity of the TME to then control and eliminate the tumor (Blattman and Greenberg, 2004; Chen and Mellman, 2017). Furthermore, although no correlation between SSs and luminal BC patient survival was found, this did not necessarily mean that stroma-specific biomarkers are not associated with prognosis. Thus, we used the intersecting DEmRNAs, DEmiRNAs, and DElncRNAs obtained from the two groups (high vs. low ISs and high vs. low SSs) to explore their prognostic value in luminal BC patients.

A total of 593 DEmRNAs, 11 DEmiRNAs, and 152 DElncRNAs common between the three comparisons were selected for further analysis. Of the 593 shared DEmRNAs, 577 were upregulated, accounting for 97.3%; thus, we mainly focused on their biological functions. The results indicated that many of them were associated with TME-related processes, such as the immune response, leukocyte migration, antigen binding, and the CAM pathway (Harjunpaa et al., 2019; Cao et al., 2020; Messex et al., 2020; Sulea et al., 2020). Next, ceRNA and PPI networks were constructed. In the ceRNA network, nine DEmiRNAs (hsa-miR-149-5p, hsa-miR-141-3p, hsa-let-7c-5p, hsa-miR-210-3p, hsa-miR-125b-5p, hsa-miR-150-5p, hsa-miR-127-3p, hsa-miR-143-3p, and hsa-miR-379-5p) acted as ceRNAs that had important roles in regulating the network along with many other DEmRNAs and DElncRNAs. Among these miRNAs, hsa-let-7c-5p and hsa-miR-150-5p were associated with increased survival times of luminal BC patients. Additionally, 10 DEmRNAs (CD86, LCP2, CXCL12, MMP9, CD48, CCR1, IKZF1, PLEK, LOX, and ESR1) with the highest degree of connectivity were identified as hub genes in the PPI network. Only CD48 and IKZF1 were related to favorable survival outcomes. Apart from the abovementioned

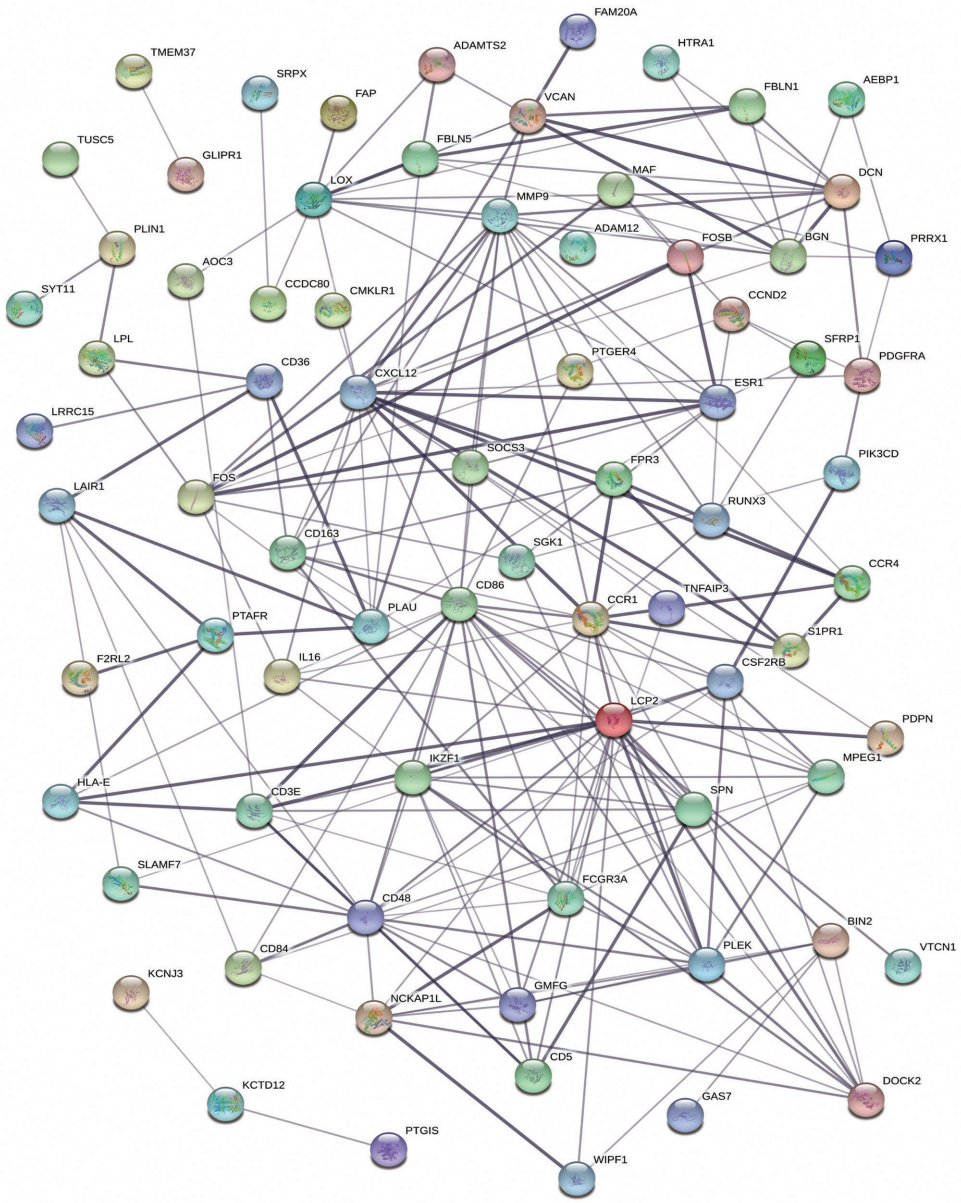


FIGURE 5 | The protein-protein interaction (PPI) network of the differentially expressed mRNAs (DEmRNAs) in the competing endogenous RNA (ceRNA) network.

prognostic biomarkers, there were another two DElncRNAs (GVINP1 and PCED1B-AS1) and 15 DEmRNAs (BIN2, CCDC69, CCR4, CD3E, CD5, DOCK2, F2RL2, HLA-E, JCHAIN, LRRC15, NCKAP1L, PIK3CD, SFRP1, SPN, and TNFAIP3) that contributed to the overall survival of patients with luminal BC. Finally, we downloaded data from a luminal BC cohort to verify these biomarkers. We only successfully confirmed the differential expression of some of the RNAs and failed to detect whether they were associated with prognosis due to a lack of survival information. Of the 15 validated biomarkers, the lncRNA GVINP1 and five mRNAs (CCDC69, DOCK2, IKZF1, JCHAIN, and NCKAP1L) had not been previously reported in BC, but these independent factors might serve as prognostic biomarkers.

The role of lncRNAs in BC has been widely studied. The lncRNAs HIF1A-AS2 and AK124454 have been shown to promote cell proliferation and invasion in TNBC cells and contribute to paclitaxel resistance (Jiang et al., 2016). Four lncRNAs (ADAMTS9-AS1, LINC00536, AL391421.1, and LINC00491) have been shown to have a significant prognostic value, and an lncRNA signature containing these four lncRNAs independently predicted overall survival in BC patients (Fan et al., 2018). LncRNAs might also act as novel candidate biomarkers to identify BC patients at high risk of tumor

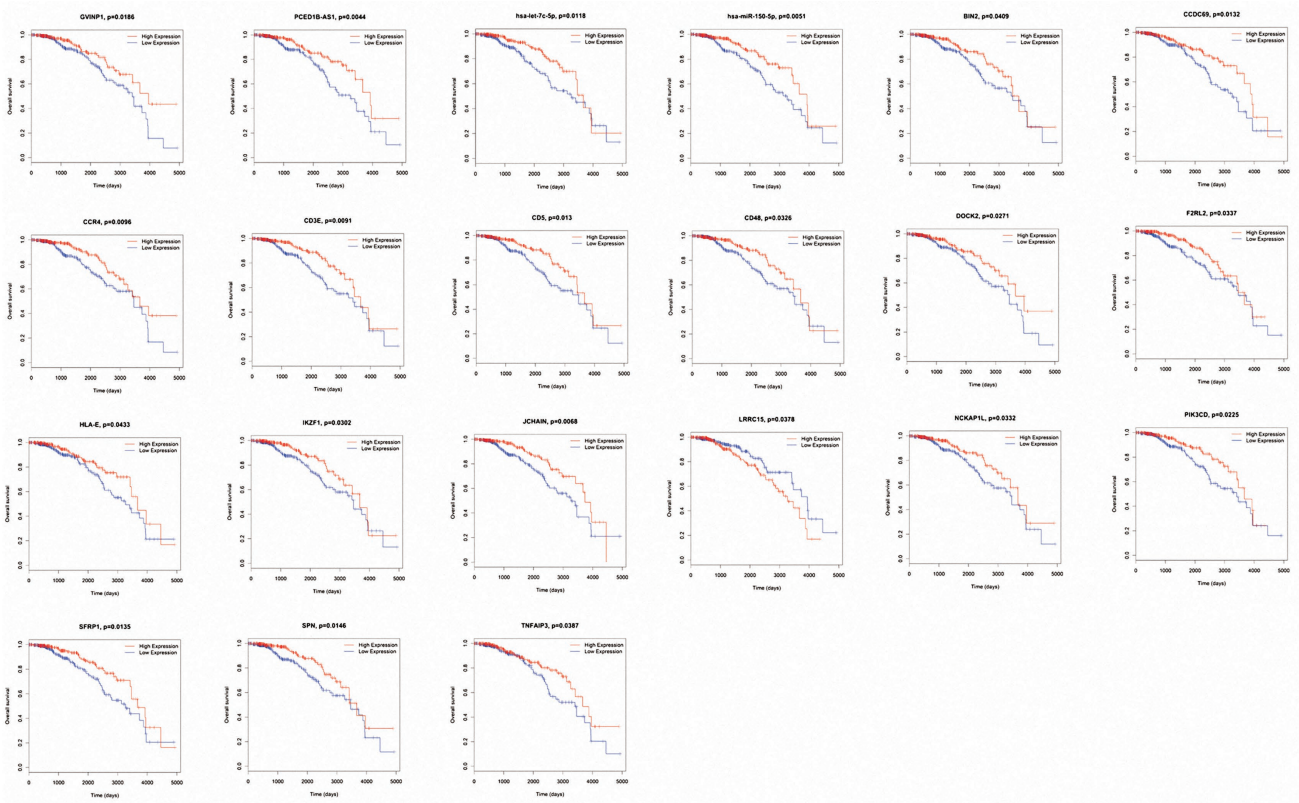


FIGURE 6 | Kaplan-Meier survival curves of the differentially expressed microRNAs (DEmiRNAs), long noncoding RNAs (DElncRNAs), and messenger RNAs (DEmRNAs) significantly associated with overall survival in luminal breast cancer (BC) patients.

recurrence (Zhou et al., 2016). In addition, Shen et al. (2020) observed that an 11-lncRNA prognostic signature for BC was associated with the infiltration of immune cell subtypes. In our study, only one lncRNA, GVINP1, was found to be associated with the overall survival in luminal BC. The lncRNA GVINP1 can bind with guanosine triphosphate selectively and noncovalently, and it has been found to be an independent prognostic marker for lung adenocarcinoma (LUAD) and NSCLC patients (Sui and Yang, 2019; Zhou et al., 2019). Our study also indicated that high expression of GVINP1 could compete with the downregulated miRNA hsa-miR-149-5p to regulate the expression of target genes, including several novel survival-associated genes such as CCDC69, DOCK2, IKZF1, JCHAIN, and NCKAP1L involved in the ceRNA network.

CCDC69 can act as a regulator of the formation of mitotic spindles and DNA replication in eukaryotic cells (Fu et al., 2006). A previous study found that women with high CCDC69 expression had longer survival times than those with low CCDC69 expression, and during cisplatin exposure, CCDC69 promoted the accumulation of p53 by activating p14ARF while inactivating MDM2 signaling to maintain p53 and p14ARF expression, suggesting that CCDC69 might be a potential therapeutic target in cancer (Cui et al., 2019). DOCK2 is a cytoplasmic protein and member of the DOCK-A subfamily of guanine exchange factors (GEFs) specific for Rac1 and Rac2 (Rac1/2) that is expressed primarily in leukocytes (Nishihara et al., 1999; Cote and Vuori, 2002). It can also regulate the functions of innate immune cells such as the migration and interferon (IFN) secretion of plasmacytoid dendritic cells, the cytotoxicity of natural killer (NK) cells, and the reactive oxygen species (ROS) production of neutrophils (Chen et al., 2018b). DOCK2 mutations have been observed in many cancers, including chronic lymphocytic leukemia (CLL), prostate cancer, and acute myeloid leukemia (AML; Hasan et al., 2018; Bjerre et al., 2019; Hu et al., 2019). In particular, DOCK2 might be a potential druggable target for AML with FLT3/internal tandem duplication (ITD) mutations (Wu et al., 2017). IKZF1 is encoded by Ikaros, which acts as a transcription factor that can control the specification and differentiation of lymphocytes (Chen et al., 2019). Overexpression of IKZF1 can activate autoimmune susceptibility through infiltrating NKG2D+ and CD8+ T cells (Chen et al., 2018a). Its prognostic value has been noted in osteosarcoma (Zhang and Yang, 2018). JCHAIN encodes the immunoglobulin J chain and links monomer units of IgA and IgM, and the upregulation of JCHAIN was likely related to tumor aggression, as studied in a cohort of patients with acute lymphoblastic leukemia (ALL) who had died (Tomar et al., 2019). NCKAP1L encodes a member

of the HEM family of tissue-specific transmembrane proteins and is only expressed in hematopoietic cells. However, the biological roles of JCHAIN and NCKAP1L in human cancer have rarely been investigated.

The main limitation of this study is that only one GEO dataset was used for verification. In particular, this cohort lacked useful follow-up information, which may lead to potential bias in the data analysis. Thus, further investigations on these prognostic biomarkers and more samples with outcome data are needed to validate our findings further.

CONCLUSION

In conclusion, our results confirm that ISs are associated with the survival of luminal BC patients and provide a set of novel biomarkers that may play prognostic roles in the TME of luminal BC.

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 and GQ contributed to the conception and design of the research. MZ and FG contributed to the acquisition of data and analysis and interpretation of data. XF contributed to statistical analysis. YW and GQ contributed to drafting the manuscript. FG, YS, and XF contributed to revision of the manuscript. GQ contributed to obtaining funding. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants from the Chinese Society of Clinical Oncology Foundation (Y-JS2019-016).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.555865/full#supplementary-material

Supplementary Figure 1 | The expression of survival-associated biomarkers in the high- vs. low-IS groups in the GEO samples.

Supplementary Figure 2 | The expression of survival-associated biomarkers in the high- vs. low-SS groups in the GEO samples.

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Binnewies, M., Roberts, E. W., Kersten, K., and Chan, V. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24, 541–550. doi: 10.1038/s41591-018-0014-x

Bjerre, M. T., Strand, S. H., Norgaard, M., Kristensen, H., Rasmussen, A. K., Mortensen, M. M., et al. (2019). Aberrant DOCK2, GRASP, HIF3A and PKFP hypermethylation has potential as a prognostic biomarker for prostate cancer. *Int. J. Mol. Sci.* 20:1173. doi: 10.3390/ijms20051173

Blattman, J. N., and Greenberg, P. D. (2004). Cancer immunotherapy: a treatment for the masses. *Science* 305, 200–205. doi: 10.1126/science.1100369

Botti, G., Scognamiglio, G., Aquino, G., Liguori, G., and Cantile, M. (2019). LncRNA HOTAIR in tumor microenvironment: what role? *Int. J. Mol. Sci.* 20:2279. doi: 10.3390/ijms20092279

Cao, J. Y., Guo, Q., Guan, G. F., Zhu, C., Zou, C. Y., Zhang, L. Y., et al. (2020). Elevated lymphocyte specific protein 1 expression is involved in the regulation of leukocyte migration and immunosuppressive microenvironment in glioblastoma. *Aging* 12, 1656–1684. doi: 10.18632/aging.102706

Carpenter, S., and Fitzgerald, K. A. (2018). Cytokines and long noncoding RNAs. *Cold Spring Harb. Perspect. Biol.* 10:a028589. doi: 10.1101/cshperspect.a028589

Chen, D. S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. *Nature* 541, 321–330. doi: 10.1038/nature21349

Chen, Y., Meng, F., Wang, B., He, L., Liu, Y., and Liu, Z. (2018b). Dock2 in the development of inflammation and cancer. *Eur. J. Immunol.* 48, 915–922. doi: 10.1002/eji.201747157

Chen, J. C., Perez-Lorenzo, R., Saenger, Y. M., Drake, C. G., and Christiano, A. M. (2018a). IKZF1 enhances immune infiltrate recruitment in solid tumors and susceptibility to immunotherapy. *Cell Syst.* 7, 92.e104–103.e104. doi: 10.1016/j.cels.2018.05.020

Chen, Q., Shi, Y., Chen, Y., Ji, T., Li, Y., and Yu, L. (2019). Multiple functions of Ikaros in hematological malignancies, solid tumor and autoimmune diseases. *Gene* 684, 47–52. doi: 10.1016/j.gene.2018.10.045

Clough, K. B., Kaufman, G. J., Nos, C., Buccimazza, I., and Sarfati, I. M. (2010). Improving breast cancer surgery: a classification and quadrant per quadrant atlas for oncoplastic surgery. *Ann. Surg. Oncol.* 17, 1375–1391. doi: 10.1245/s10434-009-0792-y

Cote, J. F., and Vuori, K. (2002). Identification of an evolutionarily conserved superfamily of DOCK180-related proteins with guanine nucleotide exchange activity. *J. Cell Sci.* 115, 4901–4913. doi: 10.1242/jcs.00219

Cui, L., Zhou, F., Chen, C., and Wang, C. C. (2019). Overexpression of CCDC69 activates p14(ARF)/MDM2/p53 pathway and confers cisplatin sensitivity. *J. Ovarian Res.* 12:4. doi: 10.1186/s13048-019-0479-3

Denaro, N., Merlano, M. C., and Lo Nigro, C. (2019). Long noncoding RNAs as regulators of cancer immunity. *Mol. Oncol.* 13, 61–73. doi: 10.1002/1878-0261.12413

Deng, L., and Lu, D. (2019). Immune profiles of tumor microenvironment and clinical prognosis among women with triple-negative breast cancer. *Cancer Epidemiol. Biomark. Prev.* 28, 1977–1985. doi: 10.1158/1055-9965.epi-19-0469

De Nola, R., and Menga, A. (2019). The crowded crosstalk between cancer cells and stromal microenvironment in gynecological malignancies: biological pathways and therapeutic implication. *Int. J. Mol. Sci.* 20:2401. doi: 10.3390/ijms20102401

Fan, C. N., Ma, L., and Liu, N. (2018). Systematic analysis of lncRNA-miRNA-mRNA competing endogenous RNA network identifies four-lncRNA signature as a prognostic biomarker for breast cancer. *J. Transl. Med.* 16:264. doi: 10.1186/s12967-018-1640-2

Fu, S., Hu, W., Kavanagh, J. J., and Bast, R. C. Jr. (2006). Targeting Aurora kinases in ovarian cancer. *Expert Opin. Ther. Targets* 10, 77–85. doi: 10.1517/14728222.10.1.77

Graham, P. J., Brar, M. S., Foster, T., McCall, M., Bouchard-Fortier, A., Temple, W., et al. (2015). Neoadjuvant chemotherapy for breast cancer, is practice changing? A population-based review of current surgical trends. *Ann. Surg. Oncol.* 22, 3376–3382. doi: 10.1245/s10434-015-4714-x

Hanahan, D., and Coussens, L. M. (2012). Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 21, 309–322. doi: 10.1016/j.ccr.2012.02.022

Harjunpaa, H., Llort Asens, M., Guenther, C., and Fagerholm, S. C. (2019). Cell adhesion molecules and their roles and regulation in the immune and tumor microenvironment. *Front. Immunol.* 10:1078. doi: 10.3389/fimmu.2019.01078

Hasan, M. K., Yu, J., Widhopf, G. F. 2nd, Rassenti, L. Z., Chen, L., Shen, Z., et al. (2018). Wnt5a induces ROR1 to recruit DOCK2 to activate Rac1/2 in chronic lymphocytic leukemia. *Blood* 132, 170–178. doi: 10.1182/blood-2017-12-819383

Hu, N., Pang, Y., Zhao, H., Si, C., Ding, H., Chen, L., et al. (2019). High expression of DOCK2 indicates good prognosis in acute myeloid leukemia. *J. Cancer* 10, 6088–6094. doi: 10.7150/jca.33244

Jathar, S., Kumar, V., Srivastava, J., and Tripathi, V. (2017). Technological developments in lncRNA biology. *Adv. Exp. Med. Biol.* 1008, 283–323. doi: 10.1007/978-981-10-5203-3_10

Jiang, Y. Z., Liu, Y. R., Xu, X. E., Jin, X., Hu, X., Yu, K. D., et al. (2016). Transcriptome analysis of triple-negative breast cancer reveals an integrated mRNA-lncRNA signature with predictive and prognostic value. *Cancer Res.* 76, 2105–2114. doi: 10.1158/0008-5472.can-15-3284

Kanehisa, M. (2002). The KEGG database. *Novartis Found. Symp.* 247, 91–101.

Li, W., and Sang, M. (2020). Gene expression and DNA methylation analyses suggest that immune process-related ADCY6 is a prognostic factor of luminal-like breast cancer. *J. Cell. Biochem.* 121, 3537–3546. doi: 10.1002/jcb.29633

Lorona, N. C., Cook, L. S., Tang, M. C., Hill, D. A., Wiggins, C. L., and Li, C. I. (2019). Recent use of oral contraceptives and risk of Luminal B, triple-negative, and HER2-overexpressing breast cancer. *Horm Cancer* 10, 71–76. doi: 10.1007/s12672-019-00362-5

Messex, J. K., Byrd, C. J., and Liou, G. Y. (2020). Signaling of macrophages that contours the tumor microenvironment for promoting cancer development. *Cells* 9:919. doi: 10.3390/cells9040919

Nishihara, H., Kobayashi, S., Hashimoto, Y., Ohba, F., Mochizuki, N., Kurata, T., et al. (1999). Non-adherent cell-specific expression of DOCK2, a member of the human CDM-family proteins. *Biochim. Biophys. Acta* 1452, 179–187. doi: 10.1016/s0167-4889(99)00133-0

Polyak, K., and Metzger Filho, O. (2012). SnapShot: breast cancer. *Cancer Cell* 22, 562–562.e561. doi: 10.1016/j.ccr.2012.06.021

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Shen, Y., Peng, X., and Shen, C. (2020). Identification and validation of immune-related lncRNA prognostic signature for breast cancer. *Genomics* 112, 2640–2646. doi: 10.1016/j.ygeno.2020.02.015

Sui, J., and Yang, S. (2019). Molecular characterization of lung adenocarcinoma: a potential four-long noncoding RNA prognostic signature. *J. Cell. Biochem.* 120, 705–714. doi: 10.1002/jcb.27428

Sulea, T., Rohani, N., and Baardsnes, J. (2020). Structure-based engineering of pH-dependent antibody binding for selective targeting of solid-tumor microenvironment. *MAbs* 12:1682866. doi: 10.1080/19420862.2019.1682866

Sun, Z., Yang, S., Zhou, Q., Wang, G., Song, J., Li, Z., et al. (2018). Emerging role of exosome-derived long non-coding RNAs in tumor microenvironment. *Mol. Cancer* 17:82. doi: 10.1186/s12943-018-0831-z

Sun, J., Zhang, Z., Bao, S., Yan, C., Hou, P., Wu, N., et al. (2020). Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. *J. Immunother. Cancer* 8:e000110. doi: 10.1136/jitc-2019-000110

Tomar, A. K., Agarwal, R., and Kundu, B. (2019). Most variable genes and transcription factors in acute lymphoblastic leukemia patients. *Interdiscip. Sci.* 11, 668–678. doi: 10.1007/s12539-019-00325-y

Wang, J., Li, Y., Fu, W., Zhang, Y., Jiang, J., Zhang, Y., et al. (2019). Prognostic nomogram based on immune scores for breast cancer patients. *Cancer Med.* 8, 5214–5222. doi: 10.1002/cam4.2428

Wu, M., Small, D., and Duffield, A. S. (2017). DOCK2: a novel FLT3/ITD leukemia drug target. *Oncotarget* 8, 88253–88254. doi: 10.18632/oncotarget.21390

Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4:2612. doi: 10.1038/ncomms3612

Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H., Shi, M., et al. (2019). Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol. Res.* 7, 737–750. doi: 10.1158/2326-6066.cir-18-0436

Zhang, Y., and Yang, F. (2018). Analyzing the disease module associated with osteosarcoma via a network- and pathway-based approach. *Exp. Ther. Med.* 16, 2584–2592. doi: 10.3892/etm.2018.6506

Zhang, Y., Zhang, L., Xu, Y., Wu, X., Zhou, Y., and Mo, J. (2020). Immune-related long noncoding RNA signature for predicting survival and immune checkpoint blockade in hepatocellular carcinoma. *J. Cell. Physiol.* 235, 9304–9316. doi: 10.1002/jcp.29730

Zhou, W., Liu, T., Saren, G., Liao, L., Fang, W., and Zhao, H. (2019). Comprehensive analysis of differentially expressed long non-coding RNAs in non-small cell lung cancer. *Oncol. Lett.* 18, 1145–1156. doi: 10.3892/ol.2019.10414

Zhou, M., Zhao, H., Xu, W., Bao, S., Cheng, L., and Sun, J. (2017). Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol. Cancer* 16:16. doi: 10.1186/s12943-017-0580-4

Zhou, M., Zhong, L., Xu, W., Sun, Y., Zhang, Z., Zhao, H., et al. (2016). Discovery of potential prognostic long non-coding RNA biomarkers for predicting the risk of tumor recurrence of breast cancer patients. *Sci. Rep.* 6:31038. doi: 10.1038/srep31038

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: 27 November 2020
doi: 10.3389/fgene.2020.587829

Whole Genome Identification of Potential G-Quadruplexes and Analysis of the G-Quadruplex Binding Domain for SARS-CoV-2

*Rongxin Zhang, Ke Xiao, Yu Gu, Hongde Liu and Xiao Sun**

State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, China

The coronavirus disease 2019 (COVID-19) pandemic caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) has become a global public health emergency. G-quadruplex, one of the non-canonical secondary structures, has shown potential antiviral values. However, little is known about the G-quadruplexes of the emerging SARS-CoV-2. Herein, we characterized the potential G-quadruplexes in both positive and negative-sense viral strands. The identified potential G-quadruplexes exhibited similar features to the G-quadruplexes detected in the human transcriptome. Within some bat- and pangolin-related betacoronaviruses, the G-tracts rather than the loops were under heightened selective constraints. We also found that the amino acid sequence similar to SUD (SARS-unique domain) was retained in SARS-CoV-2 but depleted in some other coronaviruses that can infect humans. Further analysis revealed that the amino acid residues related to the binding affinity of G-quadruplexes were conserved among 16,466 SARS-CoV-2 samples. Moreover, the dimer of the SUD-homology structure in SARS-CoV-2 displayed similar electrostatic potential patterns to the SUD dimer from SARS. Considering the potential value of G-quadruplexes to serve as targets in antiviral strategy, our fundamental research could provide new insights for the SARS-CoV-2 drug discovery.

Keywords: G-quadruplex, SARS-CoV-2, COVID-19, G4, coronavirus, G-quadruplex binding domain, SUD-homology structure

OPEN ACCESS

Edited by:
Geng Chen,
East China Normal University, China

Reviewed by:
Bernard Fongang,
The University of Texas Health Science Center at San Antonio, United States
Carson Andorf,
United States Department of Agriculture (USDA), United States

****Correspondence:***
Xiao Sun
xsun@seu.edu.cn

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

Received: *27 July 2020*
Accepted: *22 October 2020*
Published: *27 November 2020*

Citation:
Zhang R, Xiao K, Gu Y, Liu H and Sun X (2020) Whole Genome Identification of Potential G-Quadruplexes and Analysis of the G-Quadruplex Binding Domain for SARS-CoV-2. Front. Genet. 11:587829. doi: 10.3389/fgene.2020.587829

INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic, which first broke out in China, has rapidly become a global public health emergency within a few months (Lai et al., 2020). According to the statistics from the Johns Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/map.html), as of July 22, 2020, 15 million cases have been confirmed, with a death toll rising to 600,000. Since 2000, humans have suffered at least three coronavirus outbreaks, and they were severe acute respiratory syndrome (SARS) in 2003 (Zhong et al., 2003; Peiris et al., 2004; Zumla et al., 2016; Cui et al., 2019), Middle East respiratory syndrome (MERS) in 2012 (Zumla et al., 2016; Cui et al., 2019), and COVID-19. Scientists identified and sequenced the virus early in this outbreak, and named it SARS-CoV-2 (Gorbalenya et al., 2020). The symptoms of the patients infected with the novel coronavirus vary from person to person, and fever, cough, and fatigue are the most common ones

(Guan et al., 2020; Jin et al., 2020; Rothan and Byrareddy, 2020; Zu et al., 2020). The clinical chest CT (computed tomography) and nucleic acid testing are the most typical methods of diagnosing COVID-19 (Jin et al., 2020; Zu et al., 2020). It is worth noting that the recent achievements in AI (artificial intelligence) aid diagnosis technology (Li et al., 2020) and CRISPR-Cas12-based detection methods (Broughton et al., 2020) are expected to expand the diagnosis of COVID-19. Despite the great efforts of the researchers, no specific clinical drugs or vaccines had developed to cope with COVID-19 by the end of September 2020.

SARS-CoV-2 is a betacoronavirus within the Coronaviridae family that is the culprit responsible for the COVID-19 pandemic (**Figure 1A**) (Guo et al., 2020; Zheng, 2020). Studies have confirmed that SARS-CoV-2 is a positive-sense single-stranded RNA [(+)ssRNA] virus with a total length of approximately 30k. The positive-sense RNA strand of SARS-CoV-2 can serve as a template to produce viral proteins related to replication, structure composition, and other functions or events (Chen et al., 2020; Kim et al., 2020). One of the hotspots is how SARS-CoV-2 entry the host cells. SARS-CoV-2 has shown a great affinity to the angiotensin-converting enzyme 2 (ACE2), which has been proved to be the binding receptor for SARS-CoV-2 (Hoffmann et al., 2020; Walls et al., 2020). After entering the host cells, the viral genomic RNA will be released to the cytoplasm, and the ORF1a/ORF1ab is subsequently translated into replicase polyproteins of pp1a/pp1ab, which will be cleaved into some non-structural proteins (nsps). These non-structure proteins ultimately form the replicase–transcriptase complex for replication and transcription. Along with the full-length positive- and negative-sense RNAs, a nested set of subgenomic RNAs (sgRNAs) are also synthesized, and mainly translated into some structural proteins and accessory proteins. When the assembly is finished, the mature SARS-CoV-2 particles are released from the infected host cells via exocytosis (Shereen et al., 2020). Mounting evidence suggests that bats and pangolins are the suspected natural host and intermediate host of SARS-CoV-2 (Andersen et al., 2020; Lam et al., 2020; Zhang T. et al., 2020; Zhou et al., 2020). Intriguingly, a report from Yongyi Shen et al. showed that SARS-CoV-2 might be the recombination product of Bat-CoV-RaTG13-like virus and Pangolin-CoV-like virus (Xiao et al., 2020).

G-quadruplexes are the non-canonical nucleic acid structures usually formed in G-rich regions both in DNA and RNA strands (Bochman et al., 2012; Kwok and Merrick, 2017; Varshney et al., 2020). The G-quadruplex is formed by stacking G-quartets (**Figure 1B**) on top of each other, in which the four guanines making up a G-quartet are connected via Hoogsteen pairs (**Figure 1C**) (Bochman et al., 2012; Kwok and Merrick, 2017; Spiegel et al., 2020; Varshney et al., 2020). Extensive research indicated that G-quadruplexes were involved in many critical biological processes, including DNA replication (Hoshina et al., 2013; Valton et al., 2014; Valton and Prioleau, 2016; Prorok et al., 2019), telomere regulation (Tang et al., 2007; Wang et al., 2011; Takahama et al., 2013; Moye et al., 2015; Jansson et al., 2019), and RNA translation (Kumari et al., 2007; Gomez et al., 2010; Murat et al., 2018; Jodoin et al., 2019). It has been proven that G-quadruplexes existed in the viral genome and can regulate the viral biological processes, which made it possible to function as potential drug targets for antiviral strategy (Métifiot et al., 2014; Ruggiero and Richter, 2018; Saranathan and Vivekanandan, 2019). A study made by Jinzhi Tan et al. demonstrated that the SARS-unique domain within the nsp3 (nonstructural protein 3) of SARS coronavirus (SARS-CoV) exhibits the binding preference to the G-quadruplex structure in the human transcript and potentially interfere with host cell antiviral response (Tan et al., 2009). They also identified several amino acid residues that were tightly associated with its binding capacity. Yet, whether the SARS-CoV-2 contains a G-quadruplex binding domain and whether the amino acid residues related to the G-quadruplex binding affinity are conserved among SARS-CoV-2 samples need further interpretation. Besides, the G-quadruplexes in some well-known virus, such as HIV-1 (human immunodeficiency viruses

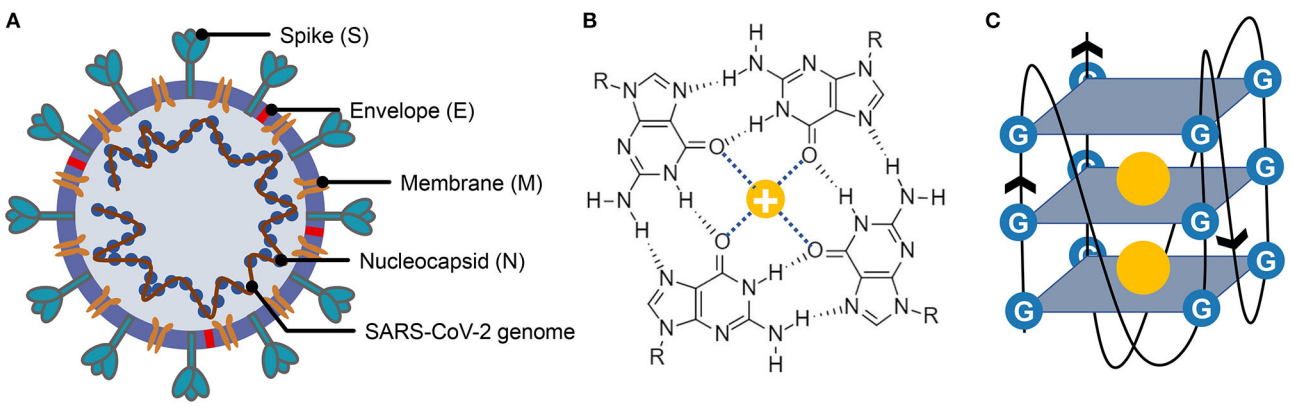


FIGURE 1 | Structure of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), G-quartet, and G-quadruplex. **(A)** The SARS-CoV-2 particle structure is composed of four structural proteins, which are the spike protein, the envelope protein, the membrane protein, and the nucleocapsid protein. The nucleocapsid proteins are bound to the SARS-CoV-2 genome. **(B)** Structure of G-quartet; the neighboring guanines are connected via Hoogsteen hydrogen. The cation is indicated by a yellow circle. **(C)** The G-quadruplex is formed by stacking multiple G-quartets, and the stabilization of the structure is partially determined by the central cations.

type 1) (Perrone et al., 2014; Piekna-Przybylska et al., 2014; Butovskaya et al., 2018, 2019), ZIKV (ZIKA virus) (Fleming et al., 2016), HPV (human papillomavirus) (Tlučková et al., 2013; Marušič et al., 2017) and EBOV (Ebola virus) (Wang et al., 2016a) have been studied. However, in our understanding of the G-quadruplexes, their potential roles in the emerging SARS-CoV-2 are lacking.

In this study, we depicted the potential G-quadruplexes (PG4s) in SARS-CoV-2 by merging several G-quadruplex prediction tools. The PG4s in SARS-CoV-2 presented similar features to the two-quartet G-quadruplexes in the human transcriptome, which potentially supported the formation and existence of the G-quadruplexes in SARS-CoV-2. Additionally, we investigated the difference in selective constraints between the G-tracts and other nucleotides in the SARS-CoV-2 genome. To further elucidate the possible pathogenic mechanism of SARS-CoV-2, we examined the sequence and structure of the SUD-homology in SARS-CoV-2 that are critical in binding the G-quadruplex structures in host transcripts.

MATERIALS AND METHODS

Data Collection

We obtained a total of 77 full-length bat-associated betacoronaviruses from the DBatVir (http://www.mgc.ac.cn/DBatVir/) database (**Supplementary Table 1**) (Chen et al., 2014). We also downloaded the bat coronavirus RaTG13 (MN996532.1) genome from the NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/), which has shown a high sequence similarity to the SARS-CoV-2 reference genome in previous reports. We acquired the SARS-CoV-2 reference genome from the NCBI virus database under the accession number of NC_045512. In addition to those sequences, nine pangolin coronaviruses were derived from GISAID (https://www.gisaid.org/) database (Shu and McCauley, 2017). To calculate the mutation frequency of each nucleotide in SARS-CoV-2, we downloaded the SARS-CoV-2 alignment sequence file from the GISAID database spanning from December 24, 2019, to April 24, 2020, which contains 16,466 SARS-CoV-2 samples. The ORF1ab amino acid sequences of 24 viruses were retrieved from the NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/). The detailed accession numbers of the sequence data used in this study are described in **Supplementary Tables 1, 2**.

Pairwise and Multiple Sequence Alignment, Phylogenetic, and Conservation Analysis

The EMBOSS Needle software, which is based on the Needleman–Wunsch algorithm and is a part of the EMBL-EBI web tools (Madeira et al., 2019), was employed for the pairwise sequence alignment. Clustal Omega (Sievers et al., 2011; Sievers and Higgins, 2018) is a reliable and accurate multiple sequence alignment (MSA) tool that can be performed on large data sets. We utilized this MSA tool to align the viral genomes and the protein sequences under the default paraments, respectively. UGENE (Okonechnikov et al., 2012) is a powerful and user-friendly bioinformatics software, and we choose UGENE to visualize the pairwise and multiple sequence alignment results. We used the MEGA X software (Kumar et al., 2018) to construct the neighbor-joining phylogenetic tree with 1,000 bootstrap replications. To depict the conservation state for each nucleotide site, the GERP++ software (Davydov et al., 2010) was applied to calculate the "Rejected Substitutions" score column by column, which can reflect the strength of constraints for each nucleotide site. The key software or databases are listed in the **Supplementary Table 3**.

Potential G-Quadruplex Detection

Several open-source G-quadruplex detection software were used to search the PG4s both in the SARS-CoV-2 positive- and negative-sense strands. G4CatchAll (Doluca, 2019), pqsfinder (Hon et al., 2017), and QGRS Mapper (Kikin et al., 2006) were employed to predict the putative G-quadruplexes, respectively; please see ref (Puig Lombardi and Londoño-Vallejo, 2019) for more information about the comparison of those tools mentioned above. The minimum G-tract length was set to two in the three pieces of software, while the max length of the predicted G-quadruplexes was limited to 30. Specifically, the minimum score of the predicted G-quadruplex was set to 10 when using pqsfinder. We utilized BEDTools (Quinlan and Hall, 2010) to sort the PG4s according to their coordinates. Apart from this, we adopted the cG/cC scoring system (Beaudoin et al., 2014) proposed by Jean-Pierre Perreault et al. to delineate the influence of sequence context on PG4s. The PG4s along with 15-nt (nucleotide) upstream and downstream sequence contexts were used to calculate the cG/cC score, and 2.05 was taken as the threshold for the preliminary inference of the G-quadruplex folding capability (**Supplementary Table 4**) (Beaudoin et al., 2014). Using a customized python script, we implemented the cG/cC scoring system, and the source code of the python script could be found at GitHub.

Homo-Dimer Homology Modeling and Electrostatic Potential Calculation

The homo-dimer of the SUD-homology in SARS-CoV-2 was modeled based on the template of the SARS-CoV SUD structure (PDB ID: 2W2G) through homology modeling. All the modeling process was performed in the Swiss Model (Waterhouse et al., 2018) website (https://swissmodel.expasy.org/) according to the default options. The electrostatic potential was calculated and visualized in the PyMOL software by using the APBS (adaptive Poisson–Boltzmann solver) plugin under the default parameters.

ΔG° Z-score Analysis

The ΔG° z-score for the SARS-CoV-2 genome was retrieved from RNAStructuromeDB (https://structurome.bb.iastate.edu/sars-cov-2) (Andrews et al., 2017). The ΔG° z-score is described as follows (Andrews et al., 2020).

$$\Delta G^{\circ} z - score = \frac{\left(MFE_{native} - \overline{MFE}_{random}\right)}{\sigma} \quad (1)$$

where the MFE_{native} means the MFE (minimum free energy) ΔG° value predicted by the RNAfold software with a window of 120 nt and step of 1 nt. In addition, the $\overline{MFE}_{random}$ represents the MFE

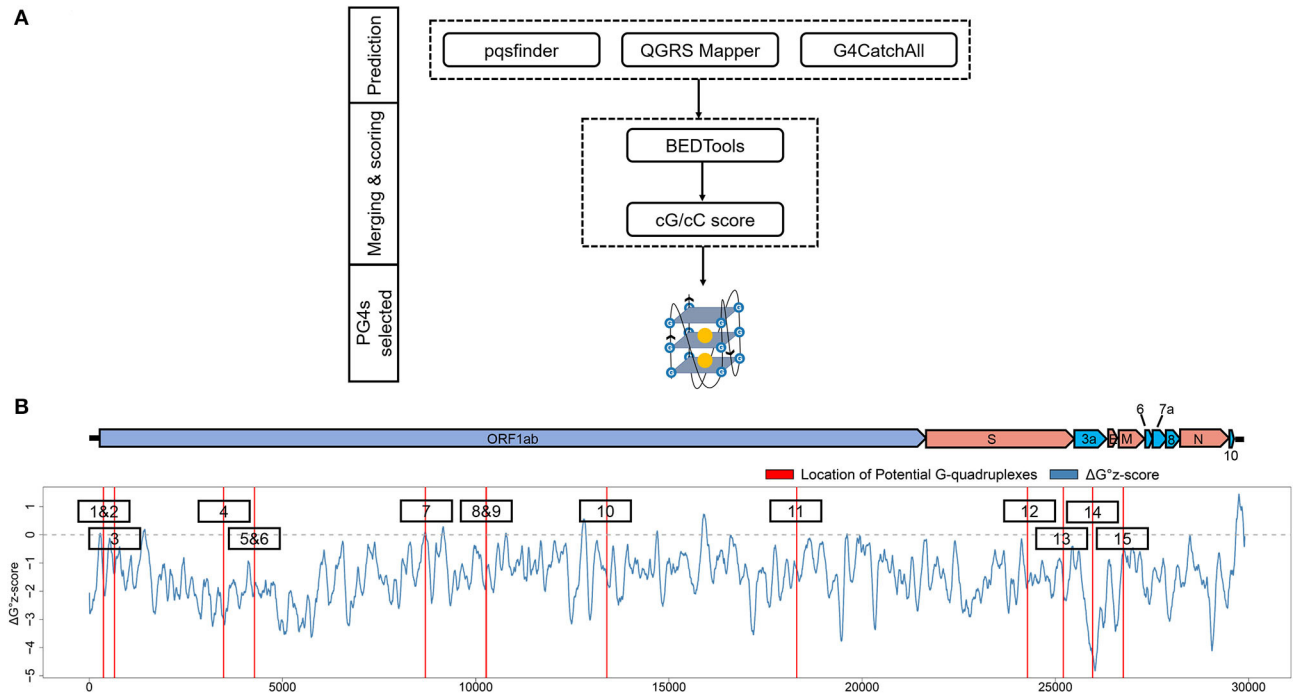


FIGURE 2 | Detection and annotation of the potential G-quadruplexes (PG4s). **(A)** The schematic flow of PG4s detection. The G-quadruplex prediction tools, pqsfinder, QGRS Mapper, and G4CatchAll were utilized for the prediction of PG4s. BEDTools and the cG/cC scoring system were applied to merge and score the PG4s. After screening, the PG4s that were used in this study were generated. **(B)** Visualization of the PG4s in the SARS-CoV-2 genome. Top: the genome organization of SARS-CoV-2. Bottom: the average ΔG° z-score for each nucleotide (blue curve) in the SARS-CoV-2 genome; the location of PG4s are plotted with red vertical lines. The order of PG4s is marked with a black box. Please note that only the PG4s in the positive-sense strand are visualized.

ΔG° value generated by the randomly shuffled sequence with the identical nucleotide composition. The σ is the standard deviation across all the MFE values.

To depict the ΔG° z-score for each nucleotide in the SARS-CoV-2 genome, we utilized the following formula.

$$z_i = \frac{\sum_{m=1}^{w} \Delta G^\circ z - score_m}{w} \quad (2)$$

where z_i is the average ΔG° z-score for nucleotide i, and w denotes the total number of the sliding windows that cover the nucleotide i. $\Delta Gz^\circ - score_m$ indicates the ΔG° z-score for the m-th window. For example, when considering the nucleotide 1,000 under the setting of 120 nt window length and 1-nt step, there are 120 sliding windows covering the nucleotide 1,000. So, the z_{200}, which means the average ΔG° z-score for nucleotide 200, is calculated as the sum of the ΔG° z-score of 120 sliding windows divided by the total number of the sliding windows.

RESULTS

Whole Genome Identification and Annotation of Potential G-Quadruplexes

To get the potential G-quadruplexes in SARS-CoV-2, we took the strategy described as follows (**Figure 2A**): (i) Predicting the PG4s with three software independently. (ii) Merging the prediction results of the PG4s and evaluating the G-quadruplex folding capabilities by the cG/cC scores. (iii) The PG4s with cG/cC scores higher than the threshold were selected as candidates for further analysis. Here, the threshold for determining whether PG4s can be folded was set to 2.05, as described in the study of Beaudoin et al. (2014). In total, we obtained 24 PG4s (**Table 1**) in the positive- and negative-sense strands for further analysis.

To annotate the PG4s, the reference annotation data (in gff3 format) of SARS-CoV-2 were downloaded from the NCBI database with the accession number of NC_045512. First, we focused on the PG4s in the positive-sense strand. Fifteen of the 24 PG4s (67.5%) were located in the positive-sense strand (**Table 1**); most of them were harbored in non-structural proteins including nsp1, nsp3, nsp4, nsp5, nsp10, and nsp14, with the remaining ones located in the spike protein, orf3a, and the membrane protein. Second, we examined the PG4s in the negative-sense strand, which is an intermediate product of replication. Nine PG4s were scattered in the negative-sense strand (**Table 1**).

To further characterize the potential canonical secondary structures competitive with G-quadruplexes, the landscape of thermodynamic stability of the SARS-CoV-2 genome was depicted by using ΔG° z-score (Andrews et al., 2020). In general, a positive ΔG° z-score implies that the secondary structure of this region tends to be less stable than the randomly shuffled sequence with the identical nucleotide composition, while a negative ΔG° z-score signifies higher stability than the randomly shuffled

TABLE 1 | The potential G-quadruplexes (PG4s) found in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

No.	Start	End	Strand	Sequence (5′→ 3′)	Annotation
1	15	37	–	GGTTGGTTTGTTACCTGGGAAGG	–
2	353	377	+	GGCTTTGGAGACTCCGTGG AGGAGG	nsp1
3	359	377	+	GGAGACTCCGTGGAGGAGG	nsp1
4	644	663	+	GGTAATAAAGGAGCTGGTGG	nsp1
5	2,449	2,472	–	GGGGCTTTTAGAGGCATGAGTAGG	–
6	3,467	3,483	+	GGAGGAGGTGTTGCAGG	nsp3
7	4,261	4,289	+	GGGTTTAAATGGTTACACTGTAG AGGAGG	nsp3
8	4,262	4,289	+	GGTTTAAATGGTTACACTGTAG AGGAGG	nsp3
9	4,886	4,901	–	GGTGGAATGTGGTAGG	–
10	6,011	6,027	–	GGATATGGTTGGTTTGG	–
11	8,687	8,709	+	GGATACAAGGCTATTGATGGTGG	nsp4
12	10,015	10,030	–	GGTTTGTGGTGGTTGG	–
13	10,015	10,039	–	GGTGATAGAGGTTTGTGGTG GTTGG	–
14	10,019	10,039	–	GGTGATAGAGGTTTGTGGTGG	–
15	10,255	10,282	+	GGTACAGGCTGGTAATGTTCAA CTCAGG	nsp5
16	10,261	10,290	+	GGCTGGTAATGTTCAACTC AGGGTTATTGG	nsp5
17	13,385	13,404	+	GGTATGTGGAAAGGTTATGG	nsp10
18	15,924	15,941	–	GGATCTGGGTAAGGAAGG	–
19	18,296	18,318	+	GGATTGGCTTCGATGTCGAGGGG	nsp14
20	24,268	24,291	+	GGCTTATAGGTTTAATGGTATTGG	S-S2
21	25,197	25,218	+	GGCCATGGTACATTTGGCTAGG	S-S2
22	25,951	25,979	+	GGTGGTTATACTGAAAAATGGGA ATCTGG	ORF3a
23	26,746	26,775	+	GGATCACCGGTGGAATTGC TATCGCAATGG	M
24	26,889	26,917	–	GGTCTGGTCAGAATAGTGC CATGGAGTGG	–

sequence. For each nucleotide in the SARS-CoV-2 genome, the ΔG° z-score was calculated for all the 120 nt windows covering the nucleotide, and an average ΔG° z-score was deduced then. Several PG4s are located in positions with a locally higher average ΔG° z-scores (**Figure 2B**) which implied the relative instability of a canonical secondary structure and the lower possibility to adopt such a competitive structure against the G-quadruplex structure, which may ultimately favor the formation of G-quadruplex structure.

Potential G-Quadruplexes in SARS-CoV-2 Show Analogical Features With the rG4s in the Human Transcriptome

In 2016, Chun Kit Kwok and coworkers profiled the RNA G-quadruplexes in the HeLa transcriptome by using the RNA G-quadruplex sequencing (rG4-seq) technology, and quantified the diversity of these RNA G-quadruplexes (Kwok et al., 2016). We set out to address the question of whether the potential G-quadruplexes in SARS-CoV-2 showed analogical features with the G-quadruplexes found in the human transcriptome and if these PG4s have the ability to form G-quadruplex structures. We noticed that the PG4s in SARS-CoV-2 are all in the two-quartet style. Therefore, we retrieved the two-quartet RNA G-quadruplex sequence data generated in the rG4-seq experiment under the condition of K+ and pyridostatin (PDS). However, for some RTS (reverse transcriptase stalling) sites labeled as two-quartet, there may exist overlapping G-quadruplexes with different loops (e.g.,GGCACAGCAGGCATCGGAGGTGAGGCGGGG), and it is difficult to determine which one was formed in the experiment. In order to eliminate the ambiguity, only the RTS sites containing non-overlapping two-quartet G-quadruplex (e.g., GTCATTTTTTGTGTTTGGTTTGGTGGTGGC) were considered.

First, we investigated the loop length distribution pattern of the two-quartet PG4s in both SARS-CoV-2 and the human transcriptome (**Figure 3A**). As a whole, the two-quartet PG4s in SARS-CoV-2 and the human transcriptome displayed similar loop length distribution patterns, and the loop length of the PG4s in SARS-CoV-2 falls into the scope of the ones from the human transcriptome. The distributions of loop length between the SARS-CoV-2 PG4s and the human two-quartet G-quadruplexes did not show discrepancies (**Supplementary Figure 1**, Wilcoxon test, $p = 0.4552$).

Considering the fact that the presence of multiple cytosine tracks may hinder the formation of G-quadruplex structures (Beaudoin, 2010; Beaudoin et al., 2014), we examined the cytosine ratio in G-quadruplex loops (**Figure 3B**). No significant difference in loop cytosine ratios was observed between the SARS-CoV-2 PG4s and the human two-quartet G-quadruplexes (Wilcoxon test, $p = 0.9911$), which suggested that the loop cytosine ratios between the two types of G-quadruplex were similar.

Taken together, our results suggested that the PG4s in SARS-CoV-2 displayed similar features to the rG4s in the human transcriptome.

Potential G-Quadruplexes Are Under Heightened Selective Constraints in Bat- and Pangolin-Related Betacoronaviruses

Recent research revealed that the G-quadruplexes in human UTRs (untranslated regions) are under selective pressures (Lee et al., 2020), and some coronaviruses on bats and pangolins are closely related to SARS-CoV-2 (Lam et al., 2020; Zhang T. et al., 2020). Consequently, we wondered whether the potential G-quadruplexes in the SARS-CoV-2 genome are under heightened selective constraints. We collected some betacoronavirus genomic sequences of bats and pangolins from several public databases and used the NJ (neighbor-joining) method to construct the phylogenetic tree with 1,000 bootstrap replications (**Supplementary Figure 2**). The RS (rejected substitutions) score for each site in the SARS-CoV-2 reference genome was evaluated by using the GERP++ software.

We checked the RS score difference between the G-tract (continuous runs of G) nucleotides and other nucleotides. A

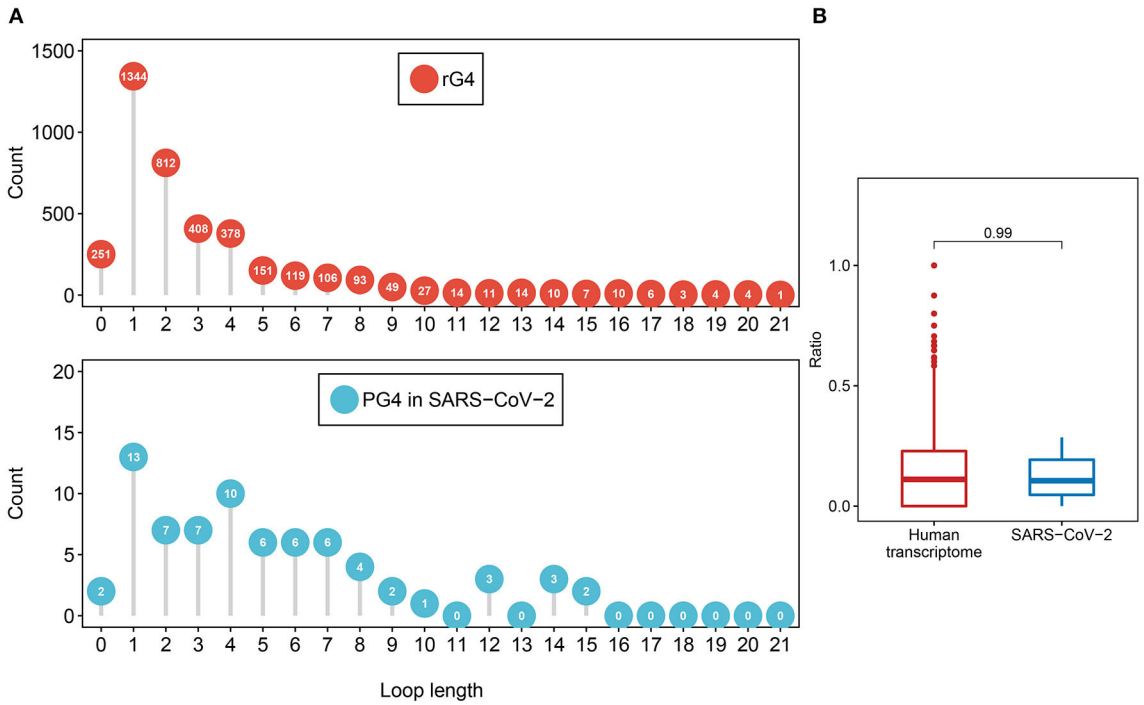


FIGURE 3 | Feature comparison of potential G-quadruplexes found in rG4-seq and SARS-CoV-2. **(A)** The dot chart represents the number of two-quartet G-quadruplex loops with different lengths in the human transcriptome and SARS-CoV-2, respectively. Top: G-quadruplex in the human transcriptome. Bottom: G-quadruple in SARS-CoV-2. **(B)** The red boxplot shows the ratio of cytosine in the human transcriptome two-quartet G-quadruplex loops, and the blue boxplot displays the ratio of cytosine in the SARS-CoV-2 PG4 loops.

significant discrepancy was observed, which means that the G-tract nucleotides exhibit heightened selective constraints than other nucleotides in the SARS-CoV-2 genome (**Figure 4A**, Wilcoxon test, $p = 9.254 \times 10^{-8}$). Considering that the G-tracts are composed of guanines, the conservation of guanines in and outside the G-tracts in the SARS-CoV-2 genome were also compared. We discovered that the guanines in G-tracts are under heightened selective constraints (**Figure 4B**, Wilcoxon test, $p = 3.363 \times 10^{-3}$). The nucleotides within G-tracts are more relevant to the G-quadruplex structural maintenance than the loops. Then we compared the G-tract and loop RS scores. As a result, the G-tract RS scores were significantly higher than that of the loops (**Figure 4C**, Wilcoxon test, $p = 3.962 \times 10^{-7}$), which suggests that the G-tracts experienced stronger selective constraints.

We also checked if the PG4s that are under heightened selective constraints is relevant to its inherent properties rather than the sequence contexts. A random test was performed to check whether the fragments containing PG4s manifested different average RS scores compared with random fragments in the SARS-CoV-2 genome. The fragments containing PG4s were designated as the sequence 100 nt upstream and downstream of the PG4 centers. We conducted 1,000 rounds of tests. In each test, we randomly selected 50 fragments from the SARS-CoV-2 genome with a length of 200 nt and carried out the Wilcoxon test to assess the average RS score difference among the randomly selected fragments and the fragments containing PG4s. The p-value for each round was retained. As a result, no evident difference was observed as few p-values (13/1,000) were < 0.05 (**Figure 4D**), suggesting that PG4s that are under heightened selective constraints are more likely to be related to its inherent properties rather than sequence contexts.

SARS-CoV-2 Contains Similar SUD to SARS-CoV

SARS-CoV and SARS-CoV-2 share similar nucleic acid sequence compositions, and both can cause acute disease symptoms. It has been confirmed that the SUD in the SARS-CoV could bind to the G-quadruplex structures in human transcripts (Tan et al., 2009), but whether SARS-CoV-2 possesses a similar SUD structure remains unclear. Thus, we explored whether the SARS-CoV-2 genome contains the protein-coding sequence potentially encoding SUD-homology and whether SARS-CoV-2 retains the ability to bind RNA G-quadruplex structures. We collected the ORF1ab amino acid sequences of some coronaviruses belonging to different genera, including seven known coronaviruses that can infect humans. The SUD-homology amino acid sequence was absent in some coronaviruses, especially in alpha, gamma, and delta coronaviruses (**Supplementary Figure 3**). In contrast, the SUD-homology sequence was retained in several betacoronavirus, particularly in bat- and pangolin-associated ones. Moreover, among the seven coronaviruses that can infect humans, only SARS-CoV and SARS-CoV-2 kept the SUD or SUD-homology sequence, while the sequence in MERS-CoV, HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1

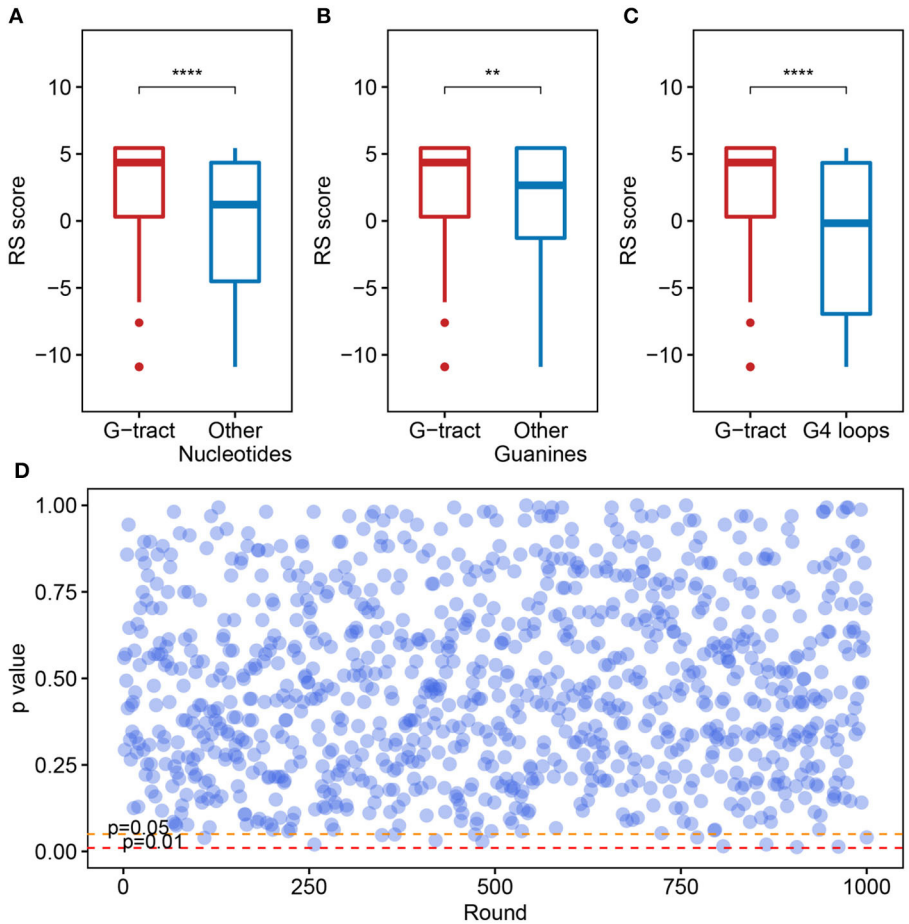


FIGURE 4 | Potential G-quadruplexes exhibit heightened selective constraints in bat and pangolin related betacoronavirus. **(A–C)** Boxplot showing the difference of nucleotide RS scores in G-tract, other nucleotides, other guanines, and PG4 loops (**$p \leq 0.01$, ****$p \leq 0.0001$). **(D)** Test of the RS score difference between the fragments containing PG4s and the randomly selected fragments. The abscissa indicates the round of the test, while the ordinate represents the *p*-value for each round.

were depleted. Next, we examined the eight key amino acid residues in SUD that are related to G-quadruplex binding affinity according to the previous reports (**Figure 5A**). Almost all the key amino acid residues are reserved in SARS-CoV-2, with one exception of a conservative replacement of K (Lysine) with R (Arginine). The conservation of the eight amino acid residues within SARS-CoV-2 samples was then investigated, to see if these residues were conserved and the G-quadruplex binding ability of SUD-homology was essential for SARS-CoV-2. We retrieved the nucleotide sequence alignment file of 16,466 SARS-CoV-2 samples from the GISAID database and calculated the mutation frequency for each nucleotide. As a result, limited mutation frequencies were found in the eight amino acid residues compared to the whole genome average mutation frequency (**Figure 5B**, frequency = 3.96). Although eight nucleotide mutations were detected of glutamate (2,432 E), seven of them were synonymous mutations. Next, we checked the electrostatic potential pattern of the SUD-homology dimer from SARS-CoV-2. The positively charged patches were observed in the core of the SUD-homology dimer, which was surrounded by negatively charged patches (**Figure 5C**). In contrast, when the dimer was rotated 180°, a slightly inclined narrow cleft with negative potential accompanied by the positively charged patches was discovered (**Figure 5D**). The above results indicated that the SUD-homology dimer of SARS-CoV-2 and the SUD_{core} (a shortened version of SUD) dimer of SARS presented analogical electrostatic potential patterns [see ref (Tan et al., 2009) for more details about the electrostatic potential surface of the SUD_{core} homodimer in SARS]. Similar to the SUD dimer of SARS, we identified the positively charged patches located in the center and back of the SARS-CoV-2 SUD-homology dimer that can potentially bind the G-quadruplex structures (**Figures 5C,D**).

DISCUSSION

The COVID-19 pandemic has caused huge losses to humans and made people pay more attention to public health. A large number of scientists around the world have participated in the fight against the epidemic. The SARS-CoV-2 coronavirus

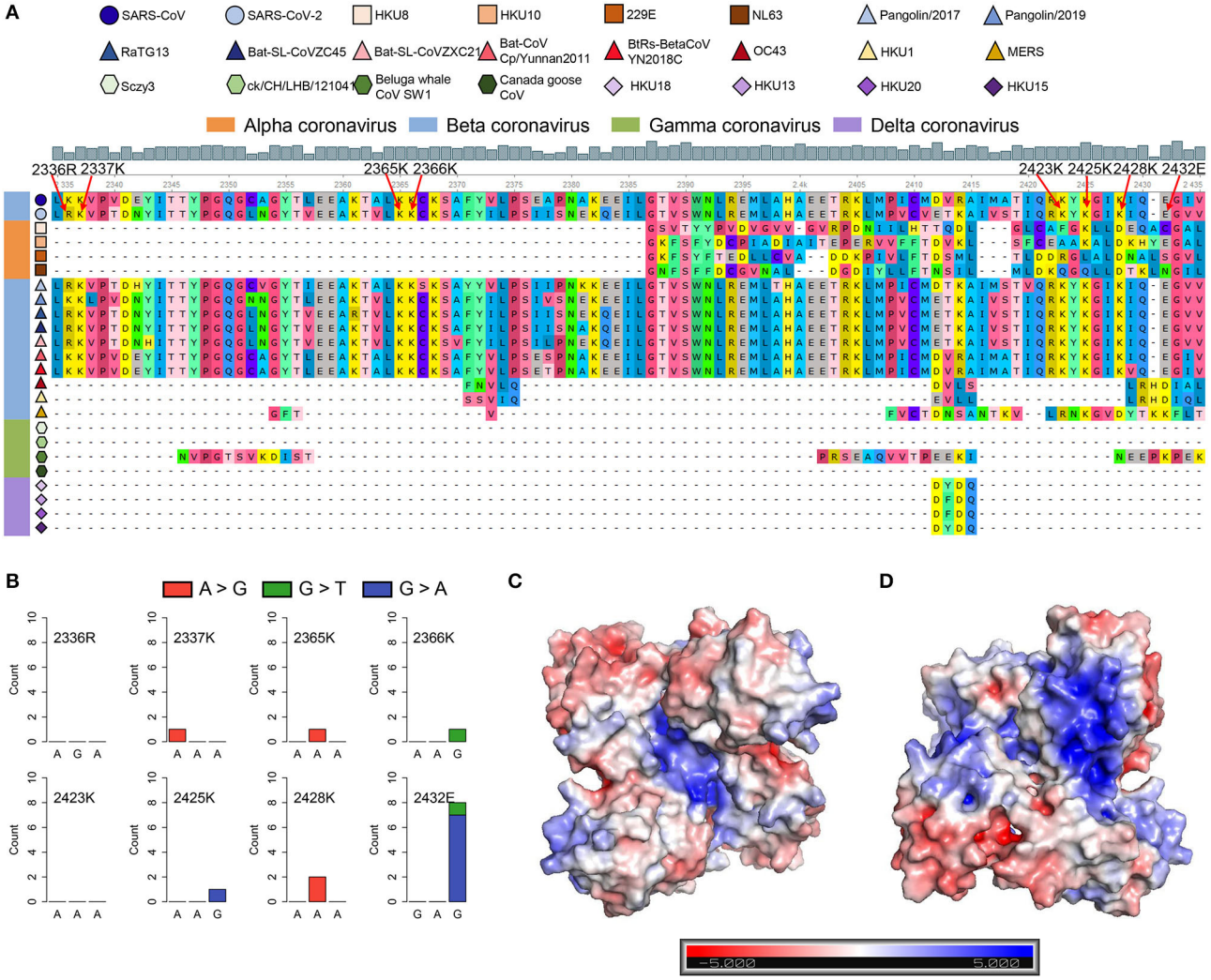


FIGURE 5 | SARS-CoV-2 contains the SARS-unique domain (SUD)-homology sequence and dimer structure. **(A)** Sequence alignment of four different genera coronaviruses. The shapes in various colors mark different kinds of coronaviruses. The color bar represents different genera of coronaviruses (orange, alphacoronavirus; blue, betacoronavirus; green, gammacoronavirus; purple, deltacoronavirus). The gray histogram shows the consensus of the alignment sites. The eight amino acid residues related to the G-quadruplex binding affinity are labeled by red arrows. **(B)** Nucleotide mutation count of eight amino acid residues among 16,466 SARS-CoV-2 samples. **(C,D)** The electrostatic potential surface of the SARS-CoV-2 SUD-homology dimer with different orientations [the front **(C)** and rotate 180° **(D)**]. The blue and red showed positive and negative potentials, respectively. The positively charged patches in the center **(C)** and back **(D)** of the SUD-homology dimer might be the binding domain for G-quadruplexes.

is the key culprit responsible for the outbreak, and no specific inhibitor drugs have been developed yet. G-quadruplexes have shown tremendous potential for the development of anticancer (Han and Hurley, 2000; Balasubramanian et al., 2011; Miller and Rodriguez, 2011; Neidle, 2017) and antiviral drugs (Perrone et al., 2015; Ruggiero and Richter, 2018, 2020), as G-quadruplexes can interfere with many biological processes that are critical to cancer cells and viruses. Therefore, it is necessary to quantify and characterize the PG4s in the SARS-CoV-2 genome to provide a possible novel method for the treatment of COVID-19.

In this study, besides three popular G-quadruplexes prediction tools, the cG/cC scoring system, which is specially designed for the identification of RNA G-quadruplexes, was adopted to determine the PG4s. Indeed, we did not find the G-quadruplexes with three or more G-quartets, which are generally considered to be more stable than the two-quartet G-quadruplexes. One of the controversial issues lies on the stability of the two-quartet G-quadruplexes, especially the folding capability of those G-quadruplexes *in vivo*. However, it is well-acknowledged that the RNA G-quadruplexes is more stable than their DNA counterparts (Joachimi et al., 2009; Zaccaria and Fonseca Guerra, 2018) and SARS-CoV-2 is a single-strand RNA virus, which may be conducive to its structure formation. Several emerging studies have demonstrated the formation of two-quartet G-quadruplexes in viral sequences (Perrone et al., 2013; Murat et al., 2014; Fleming et al., 2016; Wang et al., 2016a; Zahin et al.,

2018; Majee et al., 2020; Zhang Y. et al., 2020), which could serve as antiviral elements under the presence of G-quadruplex ligands. We also did a comprehensive search for viral two-quartet G-quadruplexes, whose formation has been confirmed by experimental methods *in vitro*, and a total of 16 two-quartet G-quadruplexes were found (**Supplementary Table 5**). The potential G-quadruplexes in SARS-CoV-2 showed a lower cytosine ratio in loops than that in experimental validated viral G-quadruplexes (**Supplementary Figure 4A**, Wilcoxon test, $p = 0.0018$), which is a positive signal for the G-quadruplex formation in SARS-CoV-2. When looking at the loop lengths, the experimental supported G-quadruplexes in other viruses displayed a short and concentrated distribution mode, while the potential G-quadruplexes in SARS-CoV-2 exhibited a relatively uniform distribution mode in <8 nt, which is a hint that maybe part of the sequence could form G-quadruplex structures (**Supplementary Figure 4B**). The cG/cC scores of the experimental supported G-quadruplexes in other viruses and their 15 nt flanking sequences were calculated as well (**Supplementary Table 6**). The cG/cC scores of some potential G-quadruplexes in SARS-CoV-2 were relatively high (**Supplementary Figure 5**); however, the overall cG/cC scores for the potential G-quadruplexes in SARS-CoV-2 were low. When only considering the G-quadruplex sequences itself, the potential G-quadruplexes in SARS-CoV-2 presented higher cG/cC scores (mean = 52.51, standard deviation = 77.29), while the experimental supported G-quadruplexes displayed relatively lower cG/cC scores (mean = 24.39, standard deviation = 37.27). These observations suggest that some of the potential G-quadruplexes in SARS-CoV-2 could be folded into secondary structures. Moreover, the K^+ (potassium ion), one of the primary positive ions inside human cells, can strongly support the formation of G-quadruplexes. Nevertheless, whether the SARS-CoV-2 G-quadruplexes could form *in vivo* requires overwhelming proofs.

Most of the PG4s we detected were located in the positive-sense strand. The G-quadruplex forming sequences in the SARS-CoV genome were presumed to function as the chaperones of SUD, and their interaction was essential for the SARS-CoV genome replication (Kusov et al., 2015). ORF1ab that encodes the replicase proteins is required for the viral replication and transcription. Some PG4s were found to harbor in ORF1ab, and whether these PG4s were related to the replication of the viral genome and interacted with the G-quadruplex binding domain in SARS-CoV-2 is worthy of further investigation. In addition to ORF1ab, there exist several PG4s in the structural and accessory protein-coding sequences as well as the sgRNAs that contain the above protein sequences. Some studies have

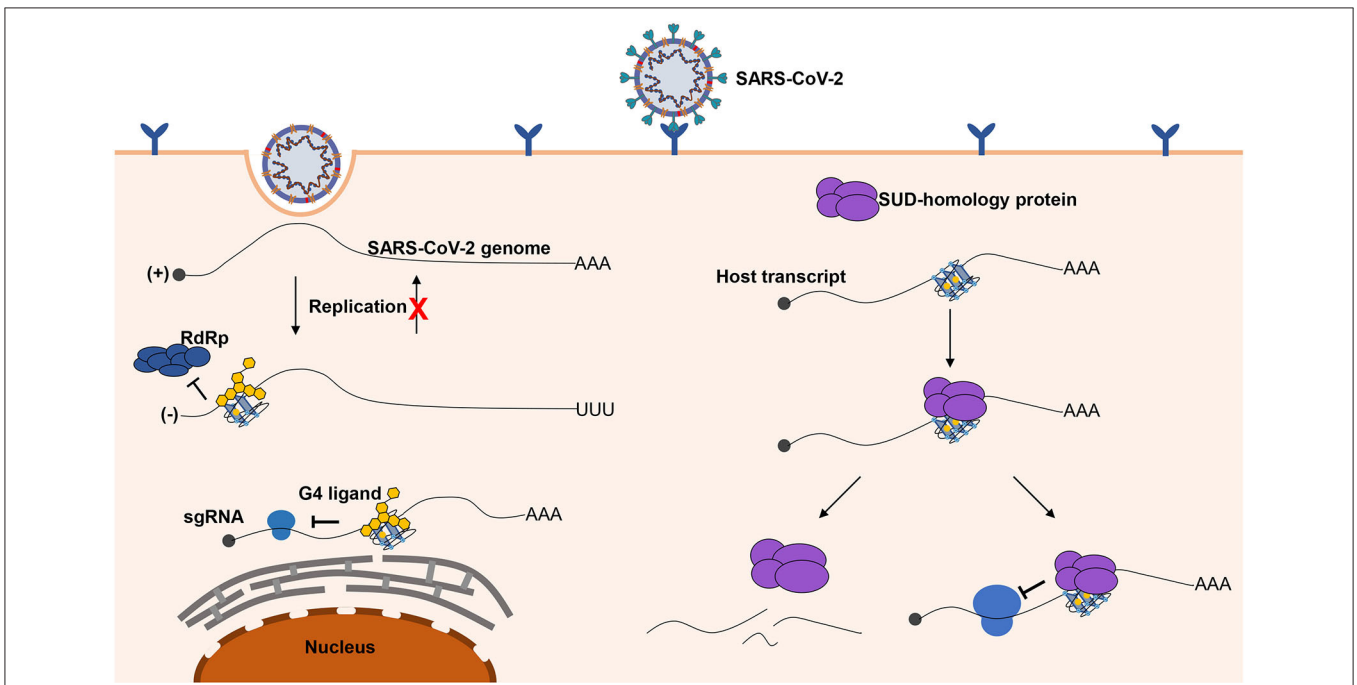


FIGURE 6 | Possible role of G-quadruplexes in the antiviral mechanism and pathogenicity. Left part, G-quadruplexes can function as inhibition elements in the SARS-CoV-2 life cycle. Both the replication and translation could be affected by the G-quadruplex structures. The stable G-quadruplex structures in the 3′ end of the negative-sense strand may interfere with the activity of RdRp; hence, the replication of the negative-sense strands to the positive-sense strands is repressed, so that the SARS-CoV-2 genomes cannot be produced in large quantities. The G-quadruplex structures can suppress the translation process by impairing the elongating of ribosomes, which can hinder the production of proteins required for the virus. The G-quadruplex structures could be stabilized by the specific ligands to enhance the inhibitory effects, which is a promising antiviral strategy. Right part, a possible mechanism for SARS-CoV-2 to impede the expression of human genes. G-quadruplex structures, particularly with longer G-stretches, are the potential binding targets for the G-quadruplex binding domain in SARS-CoV-2, and the interaction of the G-quadruplex binding domain of SARS-CoV-2 with G-quadruplex structures possibly leads to the transcript instability or obstructing of the translation efficiency.

characterized the impact of G-quadruplex structures on the translation of human transcripts, and an apparent inhibitory effect was observed (Kumari et al., 2007; Beaudoin, 2010; Shahid et al., 2010). The translation of some SARS-CoV-2 proteins requires the involvement of human ribosomes; thus, it is possible to repress the translation of SARS-CoV-2 proteins via stabilizing the G-quadruplex structures. In fact, this inhibition effect has been reported in some other viral studies (Wang et al., 2016b; Majee et al., 2020). The negative-sense strand serves as templates for the synthesis of the positive-sense strand and the subgenomic RNAs. The identified potential G-quadruplexes were broadly distributed in the negative-sense strand of SARS-CoV-2. Notably, we observed one PG4 located at the 3′ end of the negative-sense strand. A previous study confirmed that the stable G-quadruplex structures located at the 3′ end of the hepatitis C virus negative-sense strand could inhibit the RNA synthesis by reducing the activity of the RdRp (RNA-dependent RNA polymerase) (Jaubert et al., 2018). Therefore, it is necessary to further investigate whether the PG4 at the 3′ end of the negative-sense strand of SARS-CoV-2 could inhibit RNA synthesis. In addition, recent studies have detected high-frequency trinucleotide mutations (G28881A, G2882A, and G28883C) in the SARS-CoV-2 genome (Yao et al., 2020; Yin, 2020). G28881A and G28882A always co-occur within the same codon, which means a positive selection of amino acid (Mishra et al., 2020), but the consequence of the trinucleotide mutations was still elusive. We noticed that the trinucleotide mutations were in the G-rich sequence from 28,881 to 28,917 nt (5′ GGGGAACTTCTCCTGCTAGAAT**GGCTGGCAATGGCGGG** 3′). The potential G-quadruplex downstream of the trinucleotide mutations was filtered by the cG/cC score system as the presence of cytosine tracks within and flanking of the potential G-quadruplex reduce the cG/cC score; however, in fact, this potential G-quadruplex showed a relative lower MFE (Minimum Free Energy) among all the potential G-quadruplexes we detected. Whether the mutations have an internal causality with the G-rich sequence still needs to be elucidated.

The SUD in SARS, which is thought to be related to its terrible pathogenicity, has displayed binding preference to the G-quadruplexes in human transcripts (Tan et al., 2009). Our analysis revealed that the novel coronavirus SARS-CoV-2 contained a similar domain to SUD as well. Furthermore, several amino acid residues previously reported to be an indispensable part of the G-quadruplexes binding capability are almost retained in SARS-CoV-2. Further exploration indicated that the eight key amino acid residues were conserved in numerous SARS-CoV-2 samples across countries all over the world, suggesting the essentiality of the above residues. It is supposed that the binding of SUD to G-quadruplexes could affect transcript stability and translation, hence, impairing the immune response of host cells (Tan et al., 2009). The expression of host genes in SARS-CoV-2-infected cells is extremely inhibited (Kim et al., 2020); therefore, we speculate that the SARS-CoV-2 may possess a similar mechanism to SARS-CoV that can inhibit the expression of some important genes.

Herein, we briefly depict the possible role of G-quadruplexes in the antiviral mechanism and pathogenicity, and the development of certain G-quadruplex-specific ligands might be a promising antiviral strategy (**Figure 6**). We call for more researchers to shed light on the relationship between G-quadruplexes and coronaviruses. Only if we have a deeper understanding of coronaviruses can we better cope with the possible novel coronavirus pandemics in the future.

DATA AVAILABILITY STATEMENT

The datasets generated for 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

This project was under the supervision of XS. RZ, KX, and YG took part in the project. The article was written by RZ and revised by XS and KX. HL participated in the discussion of this project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 61972084 and 81830053).

ACKNOWLEDGMENTS

We would like to thank Miss Jiaqing Xue for language polishing of this manuscript. This manuscript has been released as a preprint at bioRxiv, https://www.biorxiv.org/content/10.1101/2020.06.05.135749v1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.587829/full#supplementary-material

REFERENCES

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452. doi: 10.1038/s41591-020-0820-9

Andrews, R. J., Baber, L., and Moss, W. N. (2017). RNAStructuromeDB: a genome-wide database for RNA structural inference. *Sci. Rep.* 7:17269. doi: 10.1038/s41598-017-17510-y

Andrews, R. J., Peterson, J. M., Haniff, H. S., Chen, J., Williams, C., Grefe, M., et al. (2020). An *in silico* map of the SARS-CoV-2 RNA structurome. *bioRxiv [preprint]*. doi: 10.1101/2020.04.17.045161

Balasubramanian, S., Hurley, L. H., and Neidle, S. (2011). Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov*. 10, 261–275. doi: 10.1038/nrd3428

Beaudoin, J. D. (2010). Perreault J-P, 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res*. 38, 7022–7036. doi: 10.1093/nar/gkq557

Beaudoin, J. D., Jodoin, R., and Perreault, J. P. (2014). New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res*. 42, 1209–1223. doi: 10.1093/nar/gkt904

Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet*. 13, 770–780. doi: 10.1038/nrg3296

Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., et al. (2020). CRISPR–Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol*. 38, 870–874. doi: 10.1038/s41587-020-0513-4

Butovskaya, E., Heddi, B., Bakalar, B., Richter, S. N., and Phan, A. T. (2018). Major G-quadruplex form of HIV-1 LTR reveals a (3 + 1) folding topology containing a stem-loop. *J. Am. Chem. Soc*. 140, 13654–13662. doi: 10.1021/jacs.8b05332

Butovskaya, E., Soldà, P., Scalabrin, M., Nadai, M., and Richter, S. N. (2019). HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands. *ACS Infect. Dis*. 5, 2127–2135. doi: 10.1021/acsinfecdis.9b00272

Chen, L., Liu, B., Yang, J., and Jin, Q. (2014). DBatVir: the database of bat-associated viruses. *Database* 2014:bau021. doi: 10.1093/database/bau021

Chen, Y., Liu, Q., and Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol*. 92, 418–423. doi: 10.1002/jmv.25681

Cui, J., Li, F., and Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol*. 17, 181–192. doi: 10.1038/s41579-018-0118-9

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Comp. Biol*. 6:e1001025. doi: 10.1371/journal.pcbi.1001025

Doluca, O. (2019). G4Catchall: a G-quadruplex prediction approach considering atypical features. *J. Theoretic. Biol*. 463, 92–98. doi: 10.1016/j.jtbi.2018.12.007

Fleming, A. M., Ding, Y., Alenko, A., and Burrows, C. J. (2016). Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the flaviviridae family. *ACS Infect. Dis*. 2, 674–681. doi: 10.1021/acsinfecdis.6b00109

Gomez, D., Guédin, A., Mergny, J. L., Salles, B., Riou, J. F., Teulade-Fichou, M. P., et al. (2010). A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res*. 38, 7187–7198. doi: 10.1093/nar/gkq563

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., et al. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol*. 5, 536–544. doi: 10.1038/s41564-020-0695-z

Guan, W. J., Ni, Z., Y, Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med*. 382, 1708–1720. doi: 10.1056/NEJMoa2002032

Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., et al. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Mil. Med. Res*. 7:11. doi: 10.1186/s40779-020-00240-0

Han, H., and Hurley, L. H. (2000). G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol. Sci*. 21, 136–142. doi: 10.1016/S0165-6147(00)01457-7

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052

Hon, J., Martínek, T., Zendulka, J., and Lexa, M. (2017). pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* 33, 3373–3379. doi: 10.1093/bioinformatics/btx413

Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., et al. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded *DNA J. Biol. Chem*. 288, 30161–30171. doi: 10.1074/jbc.M113.492504

Jansson, L. I., Hentschel, J., Parks, J. W., Chang, T. R., Lu, C., Baral, R., et al. (2019). Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proc. Nat. Acad. Sci. U.S.A*. 116, 9350–9359. doi: 10.1073/pnas.1814777116

Jaubert, C., Bedrat, A., Bartolucci, L., Di Primo, C., Ventura, M., Mergny, J. L., et al. (2018). RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. *Sci. Rep*. 8:8120. doi: 10.1038/s41598-018-26582-3

Jin, Y., Yang, H., Ji, W., Wu, W., Chen, S., Zhang, W., et al. (2020). Virology, epidemiology, pathogenesis, and control of COVID-19. *Viruses* 12:372. doi: 10.3390/v12040372

Joachimi, A., Benz, A., and Hartig, J. S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorgan. Med. Chem*. 17, 6811–6815. doi: 10.1016/j.bmc.2009.08.043

Jodoin, R., Carrier, J. C., Rivard, N., Bisaillon, M., and Perreault, J. P. (2019). G-quadruplex located in the 5′UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance. *Nucleic Acids Res*. 47, 10247–10266. doi: 10.1093/nar/gkz777

Kikin, O., D'Antonio, L., and Bagga, P. S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res*. 34(Suppl. 2), W676–W682. doi: 10.1093/nar/gkl253

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell* 181, 914–921.e10. doi: 10.1016/j.cell.2020.04.011

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol*. 35, 1547–1549. doi: 10.1093/molbev/msy096

Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol*. 3, 218–221. doi: 10.1038/nchembio864

Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., and Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the "SARS-unique domain" is essential for the activity of the SARS-coronavirus replication-transcription complex. *Virology* 484, 313–322. doi: 10.1016/j.virol.2015.06.016

Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S., and Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* 13, 841–844. doi: 10.1038/nmeth.3965

Kwok, C. K., and Merrick, C. J. (2017). G-quadruplexes: prediction, characterization, biological application. *Trends Biotechnol*. 35, 997–1013. doi: 10.1016/j.tibtech.2017.06.012

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., and Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents*. 55:105924. doi: 10.1016/j.ijantimicag.2020.105924

Lam, T. T. Y., Shum, M. H-H., Zhu, H. C., Tong, Y. G., Ni, X. B., Liao, Y-S., et al. (2020). Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* 583, 282–285. doi: 10.1038/s41586-020-2169-0

Lee, D. S. M., Ghanem, L. R., and Barash, Y. (2020). Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat. Commun*. 11:527. doi: 10.1038/s41467-020-14404-y

Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., et al. (2020). Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology* 296, E65–E71. doi: 10.1148/radiol.2020200905

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 47, W636–W641. doi: 10.1093/nar/gkz268

Majee, P., Kumar Mishra, S., Pandya, N., Shankar, U., Pasadi, S., Muniyappa, K., et al. (2020). Identification and characterization of two conserved G-quadruplex forming motifs in the Nipah virus genome and their interaction with G-quadruplex specific ligands. *Sci. Rep*. 10:1477. doi: 10.1038/s41598-020-58406-8

Marušič, M., Hošnjak, L., Krafčikova, P., Poljak, M., Viglasky, V., and Plavec, J. (2017). The effect of single nucleotide polymorphisms in G-rich regions of high-risk human papillomaviruses on structural diversity of DNA. *Biochim.*

Biophys. Acta Gen. Subj. 1861, 1229–1236. doi: 10.1016/j.bbagen.2016.11.007

Métifiot, M., Amrane, S., Litvak, S., and Andreola, M. L. (2014). G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.* 42, 12352–12366. doi: 10.1093/nar/gku999

Miller, K. M., and Rodriguez, R. (2011). G-quadruplexes: selective DNA targeting for cancer therapeutics? *Expert Rev. Clin. Pharmacol.* 4, 139–142. doi: 10.1586/ecp.11.4

Mishra, A., Pandey, A. K., Gupta, P., Pradhan, P., Dhamija, S., Gomes, J., et al. (2020). Mutation landscape of SARS-CoV-2 reveals three mutually exclusive clusters of leading and trailing single nucleotide substitutions. *bioRxiv [preprint]*. doi: 10.1101/2020.05.07.082768

Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., et al. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* 6:7643. doi: 10.1038/ncomms8643

Murat, P., Marsico, G., Herdy, B., Ghanbarian, A., Portella, G., and Balasubramanian, S. (2018). RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* 19:229. doi: 10.1186/s13059-018-1602-2

Murat, P., Zhong, J., Lekieffre, L., Cowieson, N. P., Clancy, J. L., Preiss, T., et al. (2014). G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* 10, 358–364. doi: 10.1038/nchembio.1479

Neidle, S. (2017). Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat. Rev. Chem.* 1:0041. doi: 10.1038/s41570-017-0041

Okonechnikov, K., Golosova, O., and Fursov, M. (2012). the Ut, unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091

Peiris, J. S. M., Guan, Y., and Yuen, K. Y. (2004). Severe acute respiratory syndrome. *Nat. Med.* 10, S88–S97. doi: 10.1038/nm1143

Perrone, R., Artusi, S., Butovskaya, E., Nadai, M., Pannecouque, C., and Richter, S. N. (2015). "G-quadruplexes in the human immunodeficiency virus-1 and herpes simplex virus-1: new targets for antiviral activity by small molecules," in *5th International Conference on Biomedical Engineering in Vietnam*, eds V.V. Toi and T.H. Lien Phuong (Cham: Springer International Publishing), 207–210.

Perrone, R., Butovskaya, E., Daelemans, D., Palù, G., Pannecouque, C., and Richter, S. N. (2014). Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.* 69, 3248–3258. doi: 10.1093/jac/dku280

Perrone, R., Nadai, M., Poe, J. A., Frasson, I., Palumbo, M., Pal,ù G., et al. (2013). Formation of a unique cluster of G-quadruplex structures in the HIV-1 nef coding region: implications for antiviral activity. *PLoS ONE* 8:e73121. doi: 10.1371/journal.pone.0073121

Piekna-Przybylska, D., Sullivan, M. A., Sharma, G., and Bambara, R. A. (2014). U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry* 53, 2581–2593. doi: 10.1021/bi4016692

Prorok, P., Artufel, M., Aze, A., Coulombe, P., Peiffer, I., Lacroix, L., et al. (2019). Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat. Commun.* 10:3274. doi: 10.1038/s41467-019-11104-0

Puig Lombardi, E., and Londoño-Vallejo, A. (2019). A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* 48, 1–15. doi: 10.1093/nar/gkz1097

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rothan, H. A., and Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmunity* 109:102433. doi: 10.1016/j.jaut.2020.102433

Ruggiero, E., and Richter, S. N. (2018). G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.* 46, 3270–3283. doi: 10.1093/nar/gky187

Ruggiero, E., and Richter, S. N. (2020). Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.* 54, 101–131. doi: 10.1016/bs.armc.2020.04.001

Saranathan, N., and Vivekanandan, P. (2019). G-quadruplexes: more than just a kink in microbial genomes. *Trends Microbiol.* 27, 148–163. doi: 10.1016/j.tim.2018.08.011

Shahid, R., Bugaut, A., and Balasubramanian, S. (2010). The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* 49, 8300–8306. doi: 10.1021/bi100957h

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., and Siddique, R. (2020). COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. *J. Adv. Res.* 24, 91–98. doi: 10.1016/j.jare.2020.03.005

Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to *reality*. *Euro Surveill* 22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494

Sievers, F., and Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 27, 135–145. doi: 10.1002/pro.3290

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. *Trends Chem.* 2, 123–136. doi: 10.1016/j.trechm.2019.07.002

Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., et al. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.* 20, 341–350. doi: 10.1016/j.chembiol.2013.02.013

Tan, J., Vonrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., et al. (2009). The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* 5:e1000428. doi: 10.1371/journal.ppat.1000428

Tang, J., Kan, Z. Y., Yao, Y., Wang, Q., Hao, Y. H., and Tan, Z. (2007). G-quadruplex preferentially forms at the very 3′ end of vertebrate telomeric DNA. *Nucleic Acids Res.* 36, 1200–1208. doi: 10.1093/nar/gkm1137

Tlučková, K., Marušič, M., Tóthová, P., Bauer, L., Šket, P., Plavec, J., et al. (2013). Human papillomavirus G-quadruplexes. *Biochemistry* 52, 7207–7216. doi: 10.1021/bi400897g

Valton, A. L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintomé, C., et al. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* 33, 732–746. doi: 10.1002/embj.201387506

Valton, A. L., and Prioleau, M. N. (2016). G-quadruplexes in DNA replication: a problem or a necessity? *Trends Genet.* 32, 697–706. doi: 10.1016/j.tig.2016.09.004

Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., and Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* 21, 459–474. doi: 10.1038/s41580-020-0236-x

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292.e6. doi: 10.1016/j.cell.2020.02.058

Wang, Q., Liu, J. Q., Chen, Z., Zheng, K. W., Chen, C. Y., Hao, Y. H., et al. (2011). G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.* 39, 6229–6237. doi: 10.1093/nar/gkr164

Wang, S. R., Min, Y. Q., Wang, J. Q., Liu, C. X., Fu, B. S., Wu, F., et al. (2016b). A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci. Adv.* 2:e1501535. doi: 10.1126/sciadv.1501535

Wang, S. R., Zhang, Q. Y., Wang, J. Q., Ge, X. Y., Song, Y. Y., Wang, Y. F., et al. (2016a). Chemical targeting of a G-quadruplex RNA in the *Ebola Virus L* gene. *Cell Chem. Biol.* 23, 1113–1122. doi: 10.1016/j.chembiol.2016.07.019

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi: 10.1093/nar/gky427

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J. J., et al. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* 583, 286–289. doi: 10.1038/s41586-020-2313-x

Yao, H., Lu, X., Chen, Q., Xu, K., Chen, Y., Cheng, L., et al. (2020). Patient-derived mutations impact pathogenicity of SARS-CoV-2. *medRxiv [preprint]*. doi: 10.1101/2020.04.14.20060160

Yin, C. (2020). Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics* 112, 3588–3596. doi: 10.1016/j.ygeno.2020.04.016

Zaccaria, F., and Fonseca Guerra, C. (2018). RNA versus DNA G-quadruplex: the origin of increased stability. *Chemistry* 24, 16315–16322. doi: 10.1002/chem.201803530

Zahin, M., Dean, W. L., Ghim, S. J., Joh, J., Gray, R. D., Khanal, S., Bossart, G. D., et al. (2018). Identification of G-quadruplex forming sequences in three manatee papillomaviruses. *PLoS ONE* 13:e0195625. doi: 10.1371/journal.pone.0195625

Zhang, T., Wu, Q., and Zhang, Z. (2020). Probable pangolin origin of sars-cov-2 associated with the COVID-19 outbreak. *Curr. Biol.* 30, 1346–1351.e2. doi: 10.1016/j.cub.2020.03.022

Zhang, Y., Liu, S., Jiang, H., Deng, H., Dong, C., Shen, W., et al. (2020). G2-quadruplex in the 3'UTR of IE180 regulates pseudorabies virus replication by enhancing gene expression. *RNA Biol.* 17, 816–827. doi: 10.1080/15476286.2020.1731664

Zheng, J. (2020). SARS-CoV-2: an emerging coronavirus that causes a global threat. *Int. J. Biol. Sci*. 16, 1678–1685. doi: 10.7150/ijbs.45053

Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, L. L. M., Xie, Z. H., Chan, K. H., et al. (2003). Epidemiology and cause of severe acute respiratory syndrome (SARS) in guangdong, people's Republic of China, in February, 2003. *Lancet* 362, 1353–1358. doi: 10.1016/S0140-6736(03)14630-2

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

Zu, Z. Y., Jiang, M. D., Xu, P. P., Chen, W., Ni, Q. Q., Lu, G. M., et al. (2020). Coronavirus disease 2019 (COVID-19): a perspective from China. *Radiology* 2, E15–E25. doi: 10.1148/radiol.2020200490

Zumla, A., Chan, J. F. W., Azhar, E. I., Hui, D. S. C., and Yuen, K. Y. (2016). Coronaviruses — drug discovery and therapeutic options. *Nat. Rev. Drug Discov*. 15, 327–347. doi: 10.1038/nrd.2015.37

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: 11 December 2020
doi: 10.3389/fgene.2020.559478

H3K4me3-Mediated Upregulation of LncRNA-HEIPP in Preeclampsia Placenta Affects Invasion of Trophoblast Cells

Ningxia Sun[1†], Huaiyan Chen[2†], Yan Ma[1†], Wenjuan Pang[1], Xiang Wang[1], Qing Zhang[1], Lu Gao[2,3,4] and Wen Li[1*]*

[1] Department of Reproductive Medicine, Changzheng Hospital, Second Military Medical University, Shanghai, China, [2] Department of Physiology, Second Military Medical University, Shanghai, China, [3] School of Medicine, International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University, Shanghai, China, [4] Shanghai Key Laboratory for Assisted Reproduction and Reproductive Genetics, Shanghai, China

Preeclampsia (PE) is a pregnancy-related disease defined as onset of hypertension and proteinuria after the 20th week of pregnancy, which causes most maternal and perinatal morbidity and mortality. Although placental dysfunction is considered as the main cause of PE, the exact pathogenesis of PE is not yet fully understood. Long non-coding RNAs (lncRNAs) are implicated in a broad range of physiological and pathological processes, including the occurrence of PE. In this study, we investigated the expression and functions of HIF-1α pathway–related lncRNA-HEIPP (high expression in PE placenta) in the pathogenesis of PE. The expression of lncRNA-HEIPP in the placenta from women who underwent PE was screened by lncRNA microarray and then verified using real-time polymerase chain reaction. Then, the methylation profile of the *lncRNA-HEIPP* promoter and the enrichment of H3K4me3 binding were assessed by bisulfite pyrosequencing and chromatin immunoprecipitation (ChIP)–quantitative polymerase chain reaction (qPCR) assay, respectively. It was found that the level of lncRNA-HEIPP in the PE placenta was significantly higher than that in normal placenta and was increased in HTR-8/SVneo human trophoblast cells upon hypoxia treatment. Moreover, we reported that H3K4me3 manifested significantly higher promoter occupancy on *lncRNA-HEIPP* promoter in HTR-8/SVneo cells upon hypoxia treatment and found that the downregulation of lncRNA-HEIPP promoted trophoblast invasion. Our findings suggested that the hypoxia-induced expression of lncRNA-HEIPP mediated by H3K4me3 modification in trophoblast may contribute to the pathogenesis of PE.

Keywords: lncRNA, placenta, preeclampsia, hypoxia, H3K4me3

OPEN ACCESS

Edited by:
Geng Chen,
East China Normal University, China

Reviewed by:
Guodong Fu,
University of Toronto, Canada
Tian Fuju,
Shanghai Jiao Tong University School of Medicine, China

***Correspondence:**
Lu Gao
lu.gao@smmu.edu.cn
Wen Li
liwen@smmu.edu.cn

† These authors have contributed equally to this work

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

Received: *06 May 2020*
Accepted: *16 November 2020*
Published: *11 December 2020*

Citation:
Sun N, Chen H, Ma Y, Pang W, Wang X, Zhang Q, Gao L and Li W (2020) H3K4me3-Mediated Upregulation of LncRNA-HEIPP in Preeclampsia Placenta Affects Invasion of Trophoblast Cells. Front. Genet. 11:559478. doi: 10.3389/fgene.2020.559478

INTRODUCTION

Preeclampsia (PE) is a common complication of human pregnancy defined as the occurrence of hypertension and significant proteinuria after the 20th week of gestation, occurring in approximately 3 to 5% of all pregnancies and resulting in about 50,000 maternal deaths annually worldwide (Gormley et al., 2017). Preeclampsia is also the leading cause of fetal morbidity and

mortality (Mayrink et al., 2018). It is characterized by maternal arterial pressure >140/90 mmHg and proteinuria >0.3 g/24 h after 20 weeks' gestation (Armaly et al., 2018). Current studies suggested that the pathogenesis of PE is closely related to immune system and genetic dysfunctions (Hosseini et al., 2018; Yong et al., 2018). However, the dysfunction of placenta plays a central role in the pathogenesis of PE (Aouache et al., 2018; Gutierrez et al., 2019).

Recent studies have shown that abnormal trophoblast differentiation, migration, and apoptosis often cause decreased invasion of trophoblast cells and remodeling of uterine spiral artery, resulting in reduced uteroplacental perfusion (Cotechini et al., 2014; Chen and Khalil, 2017; Brosens et al., 2019). This further leads to hypoxia and inflammatory changes in the placenta, which will eventually lead to PE (Harati-Sadegh et al., 2018; Tong and Giussani, 2019). Recent studies have shown that the important phenotypes of PE, such as hypertension and proteinuria, are closely related with hypoxia inducible factor-1α (HIF-1α) (Wang et al., 2017; Alici Davutoglu et al., 2018). HIF-1α is a transcription factor that controls the expression of hypoxia-induced factors such as vascular endothelial growth factor (VEGF), platelet-derived growth factor, and so on (Kuzmanov et al., 2012; Zhang et al., 2019), which are considered to play important roles during the pathogenesis of PE. Recently, it has been found that long non-coding RNA (lncRNA) could regulate the invasion and apoptosis of trophoblast cell lines (Song et al., 2017; Xu et al., 2018), suggesting their potential roles in PE occurrence. However, whether hypoxia and HIF-1α regulate the expression of specific lncRNAs during PE has not been investigated yet. So, the objective of the current study is to investigate the differentially expressed lncRNAs in human placenta between normal pregnancy and PE and dissect the mechanisms underlying the regulation of lncRNAs by hypoxia as well as their functions in trophoblast cells migration.

MATERIALS AND METHODS

Patient Samples

All patients provided written informed consent, and this study was approved by the Human Research Ethics Committee of the Shanghai Changzheng Hospital at the Second Military Medical University. Thirty PE patients (mean age of 33.20 ± 0.46 years) and 30 control patients (mean age of 32.60 ± 0.47 years) who received a cesarean section were included in this study. Preeclampsia was clinically diagnosed before cesarean. The main criteria for PE diagnosis were as follows: systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg on two occasions at least 4 h apar, and proteinuria >0.3 g/24 h. The control pregnancy was defined as no medical complications and proteinuria, and maternal blood pressure <140/90 mm Hg. Clinical characteristics and detailed information of all patients enrolled in this study are summarized in **Table 1**. The original demographic data of enrolled patients can be found in **Supplementary Table 1**. The placenta specimen was resected from the middle of villous lobule, avoiding any visible blood clot or calcification. To minimize blood contamination, each piece of tissue was intensively washed with ice-cold phosphate-buffered saline and then immediately snap-frozen in liquid nitrogen after resection. Of the 30 pairs of placenta samples, four pairs were randomly selected for lncRNA and mRNA microarray analysis, and the other samples were used for the following verification.

TABLE 1 | Clinical characteristics of the pregnant women enrolled in this study[a].

	Normal (*n* = 30)	Preeclampsia (*n* = 30)	*P*-value
Maternal age (y)	32.60 ± 0.47	33.20 ± 0.46	0.368
Gestational age (wk)	38.35 ± 0.34	34.44 ± 0.15[b]	<0.001
BP (mm Hg)			
Systolic	108.13 ± 0.81	163.87 ± 1.30[b]	<0.001
Diastolic	70.17 ± 0.85	104.43 ± 1.23[b]	<0.001
Proteinuria (g/24 h)	N/A	2.62 ± 0.20	N/A
Fetal birth weight (g)	3,278.00 ± 58.01	2,577.33 ± 97.73[b]	<0.001

BP, blood pressure.
[a]*Data are shown as mean ± SEM.*
[b]*Compared with normal pregnancy: P < 0.001.*

RNA Extraction

Total RNA was extracted from snap-frozen placenta tissues obtained from PE placenta patients and control patients using TRIzol Reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's instruction. The RNA integrity and concentration were evaluated using the NanoDrop ND-1000 spectrophotometer and 2100 RNA Nano 6000 Assay Kit (Agilent Technologies, Santa Clara, CA, United States).

RNA Microarray and Computational Analysis

RNA with RNA integrity number greater than 6.5 purified from total RNA following the removal of rRNA was amplified and transcribed into fluorescent cDNA along the entire length of the transcripts without 3′ bias utilizing random priming method, and cDNA was labeled and hybridized to the Human LncRNA Array V2.0 (8 × 60, Arraystar). In addition, 30,215 coding transcripts were detected using the microarray as well. The microarray was performed by KangCheng Bio-tech (Shanghai, China). The arrays were scanned using the Agilent Scanner G2505B (Agilent Technologies), and the acquired array images were analyzed using Agilent Feature Extraction software (version 11.5.1; Agilent Technologies). Quantile normalization and subsequent data processing were performed using GeneSpring GX v11.5.1 software package (Agilent Technologies).

Gene Ontology Pathway Analysis

To investigate the roles and associated pathways of differentially expressed lncRNAs and mRNAs, Gene Ontology (GO) pathway analysis was performed to annotate the transcripts with terms under the biological process, cellular component, and molecular function ontologies. GO annotations of microarray genes were

downloaded from NCBI[1], UniProt[2], and the GO[3]. A Fisher exact test was performed in order to locate the significant enrichment pathway. The resulting *P*-values were adjusted using the Benjamini–Hochberg false discovery rate (BH FDR) algorithm. Pathway categories with a FDR < 0.05 were reported.

Quantitative Reverse Transcription Real-Time Polymerase Chain Reaction

Total RNA (2 μg) was reverse transcribed with random hexamers using the Reverse Transcription System Kit (Promega Corporation, Madison, WI). LncRNA expression in placenta tissues was quantified using a standard quantitative reverse transcription real-time PCR (qRT-PCR) on the StepOne Plus system (Applied Biosystems, Foster City, CA, United States), in a total reaction volume of 20 μL, including 10 μL SYBR Premix Ex Taq (2x), 1 μL of PCR Forward Primer (10 uM), 1 μL of PCR Reverse Primer (10 uM), 2 μL of cDNA, and 6 μL of double-distilled water. Primer sets are listed in **Supplementary Table 2**. The qRT-PCR was performed with an initial denaturation step of 10 min at 95°C; 95°C (15 s), 60°C (30 s), 72°C (30 s) for a total 40 cycles; and a final extension step at 72°C for 5 min. All experiments were performed in triplicate. All samples were normalized to GAPDH. The geometric mean in each triplicate was used to calculate the relative lncRNAs concentrations (ΔCt = Ct of lncRNAs − Ct of GAPDH). The fold change of expression was calculated using the $2^{-\Delta\Delta Ct}$ method.

Cell Line and Hypoxia Treatment

The immortalized EVT cell line HTR-8/SVneo was used. The cells were maintained in RPMI-1640 supplemented with 5% fetal calf serum, 1% L-glutamine 200 mM, and 1% penicillin–streptomycin (all from Invitrogen) under an atmosphere of 5% CO_2 at 37°C. To establish the hypoxia model, the cells were exposed to 94%N_2/1% O_2/5% CO_2 for 2, 24, 48, and 72 h. The control group was cultured under normal air condition for the corresponding period.

HIF-1α Overexpression in HTR-8/SVneo Cells

The coding sequence of human HIF-1α was amplified by PCR from human genomic DNA libraries (GENECHEM Company, Shanghai, China) and was subcloned into the pEZ-MO2 plasmid via homologous recombination according to manufacturer's protocol (Hieff Clone® Plus One Step Cloning Kit, YEASEN, Shanghai, China) to construct pEZ-MO2–HIF-1α overexpression. The sequence of the construct was confirmed by DNA sequencing at Sangon Biotech Company (Shanghai, China). The HTR-8/SVneo cells at 70% confluence were transiently transfected with pEZ-MO2-HIF-1α plasmid for 48 h using Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. The empty pEZ-MO2 plasmid was also transfected as negative control (NC).

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

[2]http://www.uniprot.org/

[3]http://www.geneontology.org/

Small Interfering RNA Transfection

lncRNA-HEIPP in HTR-8/SVneo cells were knocked down by transfection of preannealed stealth small interfering RNA (siRNA) duplexes designed online by Thermo Fisher's BLOCK-iT™ RNAi Designer[4], using Lipofectamine RNAiMAX transfection reagent (Invitrogen). *Silencer*® Negative Control #1 siRNA (si-NC, AM4611, Ambion) was transfected as an NC. The efficiency of knockdown was determined by RT-PCR.

Cell Counting Kit-8 Cell Proliferation Assay

Cell viability was assessed using the Cell Counting Kit-8 (Dojindo Laboratories, Kumamoto, Japan) according to the manufacturer's protocol. Cell proliferation curves were plotted using the absorbance at each time point (0, 12, 24, and 48 h). All experiments were performed in triplicate.

Matrigel Invasion Assay

The invasion of HTR-8/SVneo cells across Matrigel was objectively evaluated in an invasion chamber according to the manufacturer's protocol (Millipore). The cells transfected with si-NC or si-HEIPP (10 and 30 μL) were plated in the upper chambers of 24-well plates at a density of 1.0×10^4 cells/well in RPMI-1640 medium containing 5% fetal bovine serum; the membrane pore of the Transwell chamber was 8 μm in diameter. The membranes were coated with Matrigel (BD, United States) to form matrix barriers at the concentration of 200 μg/mL. A 700-μL aliquot of RPMI-1640 medium containing 10% FBS was immediately placed in the lower well of the chamber as a chemoattractant. After incubation for 48 h at 37°C, cells were completely removed from the upper surface of the filter using a sterile cotton swab, and cells that had migrated to the lower surface were fixed and stained with 0.1% crystal violet. Finally, the cell migration ability was determined by counting the number of stained cells on the membranes in 10 randomly selected, non-overlapping fields at 20× magnification under microscope. The migration index was calculated as the ratio of the percentage of cell migration in the various treatments to that of the vehicle. Each experiment was performed in triplicate and repeated three times.

Bisulfite Pyrosequencing

Bisulfite pyrosequencing was used to quantity DNA methylation at two sites in the DMR of the lncRNA-HEIPP gene. Two micrograms of DNA was subjected to bisulfate conversion using the EpiTect Bisulfite Kit (59104) (Qiagen Ltd., GmbH, Hilden, Germany). Pyrosequencing primers were designed using PyroMark Assay Design 2.0 (Qiagen Ltd., GmbH, Hilden, Germany) and listed in **Supplementary Table 2**. One hundred nanograms of bisulfite-converted DNA was PCR amplified in a 50-μL reaction volume containing 5 × PCR buffer (KAPA Ltd., United States), 10 mM of each dNTP (KAPA Ltd., United States), 50 mM forward and reverse primers (BGI Inc., Shanghai, China), and 1 U DNA polymerase (HotStart Taq, KAPA Ltd). The PCR

[4]http://rnaidesigner.thermofisher.com/rnaiexpress/rnaiDesign.jsp

cycling parameters were 95°C for 3 min; 40 cycles of 95°C for 30 s, a variable annealing temperature (Ta) for 30 s, and 72°C for 1 min; and extension at 72°C for 7 min. A 25-μL aliquot of each PCR product was subjected to pyrosequencing using the PyroMark Q96 ID Pyrosequencer (Qiagen Ltd.) following the manufacturer's recommended protocols. The degree of methylation at each CpG site was determined using Pyro Q-CpG Quantification software. All samples were analyzed in triplicate.

ChIP Assay

The HTR-8 trophoblast cell line was used for ChIP-based analysis for enrichment of H3K4me3 on *lncRNA-HEIPP* promoter. ChIP assays were conducted according to the manufacturer's instruction (Millipore, 17-10086). The chromatin extracted from cells was sheared, subjected to immunoprecipitation with H3K4me3 (Abcam) or immunoglobulin G (IgG) (Millipore) antibodies, reverse cross linked, and subjected to quantitative ChIP- PCR (qChIP). The qChIP was performed on sheared DNA with following primers: SP1: F: 5′ ATTTTGCTTCC TATCCCT 3′, R:5′ ACCGACAATCTCCTCAGT 3′, SP2: F: 5′ GA CTGAGGAGATTGTCGGT 3, R: 5′ CTTATGATGGTCTGGGT GA 3′, SP3: F: 5′ TGGAGGAGGGAGATGACA 3′, R: 5′ TGGA GGAGGGAGATGACA 3′, SP4: F: 5′ CATGCCCTCTCAGC CTAAC 3′, R: 5′ AGTCCCCCCAAGAAAAACA 3′. The four pairs of primers amplified −324 to −1,288 bp region upstream from the transcriptional starting site of lncRNA-HEIPP. The results were normalized to input and expressed as a percentage of the fold difference. IgG was used as an NC. Data were obtained from at least three independent experiments.

Statistical Analysis

All data are expressed as mean ± SD (standard deviation), and all statistical analyses were performed using the SPSS statistical software package (SPSS, Inc., Chicago, IL, United States). All the data were tested for homogeneity of variance by Bartlett test before analyzing the significance. Comparisons were made using Student *t*-test and analysis of variance. $P < 0.05$ was considered statistically significant.

RESULTS

Differential Expression of lncRNAs in Placenta Between PE and Normal Patients

The demographic data of PE women and normal pregnant women are summarized in **Table 1**. There was no significant difference in maternal age between normal pregnancies and pregnancies with PE ($P > 0.05$). The systolic and diastolic blood pressure was significantly higher in pregnant women with PE ($P < 0.001$). The average proteinuria is 2.62 ± 0.20 g/24 h, which means that the cases enrolled this study were severe PE patients (proteinuria >2.0 g/24 h). Women from the PE group delivered earlier than normal pregnant women mostly due to the medical termination of the pregnancy, which is the major cause of the lower infant birth weight in the PE group, as there was no intrauterine growth restriction patient in the PE group enrolled in this study.

As the transcriptome of the placenta is dynamically changed during advancing gestation (Deyssenroth et al., 2017), to ensure the identical or similar baselines between the two disease groups, we selected four pairs of PE patients and controls with matched gestational age within 2 weeks (**Supplementary Table 1**). We aligned lncRNA array data into RefSeq_NR, UCSC, and Ensembl database and compared the lncRNA expression levels between four human PE placenta (P) and four normal samples (N). The expression profiles of lncRNAs were shown by calculating the log-fold change (P/N). Among these lncRNAs, 405 were consistently upregulated, and 344 lncRNAs were consistently downregulated (**Figure 1A**). The clustering analysis also showed the differential expression patterns of mRNAs between PE placentas and normal samples (**Figure 1B**). The complete dataset regarding the significantly upregulated and downregulated lncRNA or mRNA was deposit on the public repository[5].

GO Pathway Analysis

GO pathway analysis was performed to determine the lncRNA and mRNA enrichment in biological processes, cellular components, and molecular functions. HIF-1α pathway is known to play an important role in the pathogenesis of PE and identified as one of the most significant enrichment signaling pathways under our experimental settings (**Figure 2A** and **Supplementary Table 3**). We found that a number of lncRNAs and molecules are closely related with HIF-1α pathway (**Figure 2B**), and the top 10 lncRNAs (AF064860.7, AC027612.3, RP11-244N9.4, RP11-574M7.1, RP11-80I15.4, RP11-939C17.2, XLOC_013276, AC023085.1, AC009236.1, and RP11-184D12.1) showing mostly dramatic different expression between PE and normal placentas were selected for further validation (**Supplementary Table 4**).

Validation of Differentially Expressed lncRNAs

To further validate the preliminary data from microarray, we performed qRT-PCR to determine the expression of the ten candidate lncRNAs related with HIF-1α pathway in 30 PE placenta tissues and 30 normal placenta tissues. We found that 6 of the 10 lncRNAs manifested significant upregulation in PE samples compared with normal samples (**Figure 3A**), which are consistent with the array results. Among the six lncRNAs, lncRNA-AF064860.7 showed the most significant differences between PE samples and normal samples, so that we named it lncRNA-HEIPP (high expression in PE placenta).

In the HTR-8/SVneo human trophoblast cell lines, we found that hypoxia could induce the expression of lncRNA-HEIPP, which reached the highest level at 24 h upon hypoxia treatment (**Figure 3B**). Although the other five candidate lncRNAs also showed time-dependent changes of expression upon hypoxia treatment, the increasing trend and fold change were less significant than that in lncRNA-HEIPP. Moreover, the

[5]http://doi.org/10.5281/zenodo.4039476

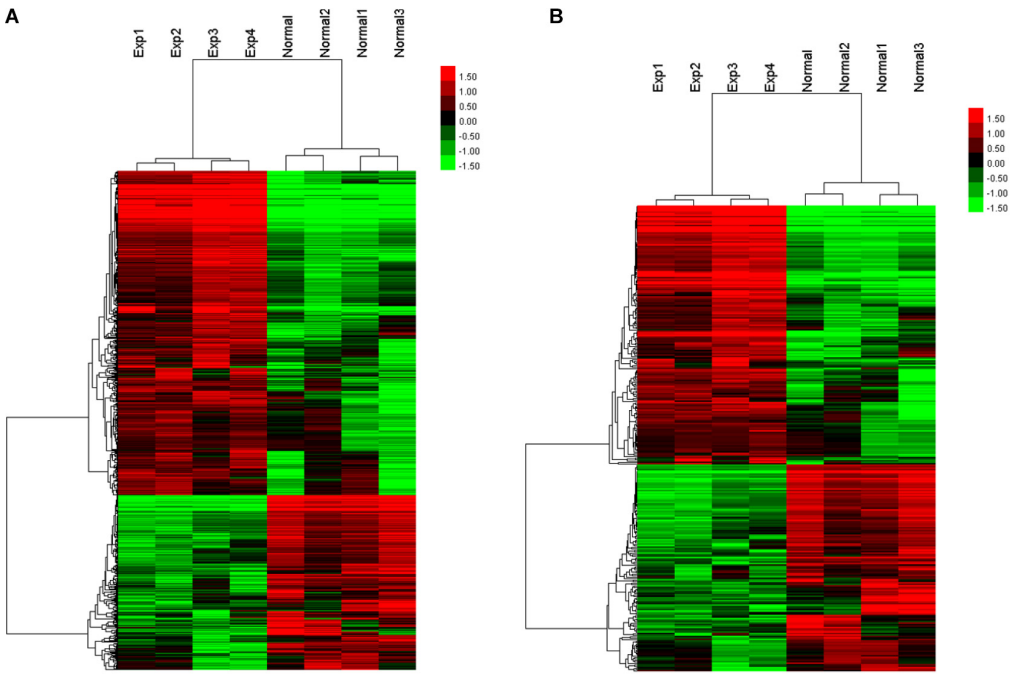


FIGURE 1 | Heatmap and hierarchical clustering of lncRNA expression **(A)** and mRNA expression **(B)** between the preeclampsia placenta (Exp) and normal samples (Normal).

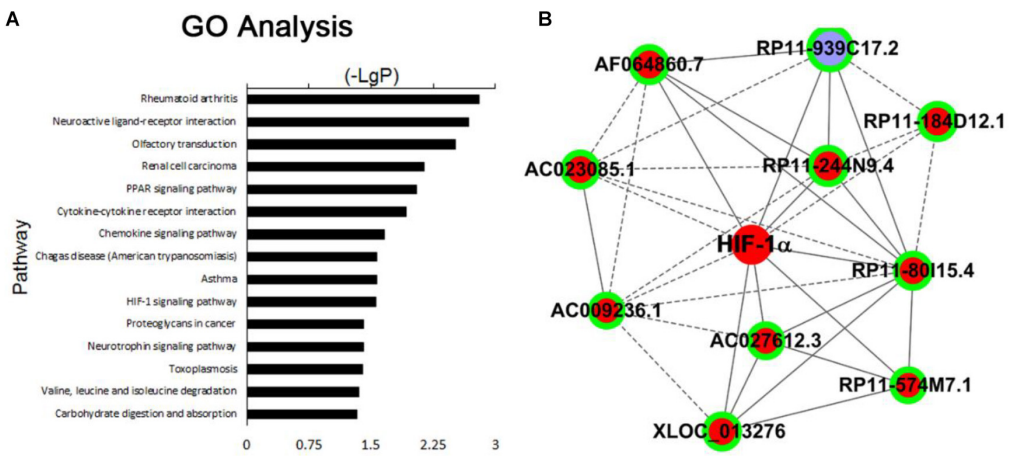


FIGURE 2 | (A) The top 15 signaling pathways that were most significantly enriched in PE placenta compared with normal placenta. **(B)** Top 10 lncRNAs related with HIF-1α pathway: AF064860.7, AC027612.3, RP11-244N9.4, RP11-574M7.1, RP11-80I15.4, RP11-939C17.2, XLOC_013276, AC023085.1, AC009236.1, and RP11-184D12.1.

overexpression of HIF-1α increased the expression of lncRNA-HEIPP, further suggesting the interaction between HIF-1α pathway and lncRNA-HEIPP (**Figure 3C**).

Hypoxia Did Not Change the Methylation Profile of lncRNA-HEIPP Promoter in Placental Trophoblast Cells

As the epigenetic regulatory elements such as DNA methylation and H3K4me3 occupancy were critical for the gene expression regulation in placental development and trophoblast invasion (Apicella et al., 2019; Kwak et al., 2019), we analyzed the epigenetic regulatory elements of lncRNA-HEIPP in the placenta in PE versus controls. Specifically, compared to the normal oxygen condition, the DNA methylation levels of lncRNA-HEIPP promoter in HTR-8/SVneo human trophoblast cells increased along with the extension of hypoxia time, which reached the highest level at 24 h upon hypoxia treatment (**Figure 4A**). However, the increase of methylation level was not significant at each time point (**Figure 4B**).

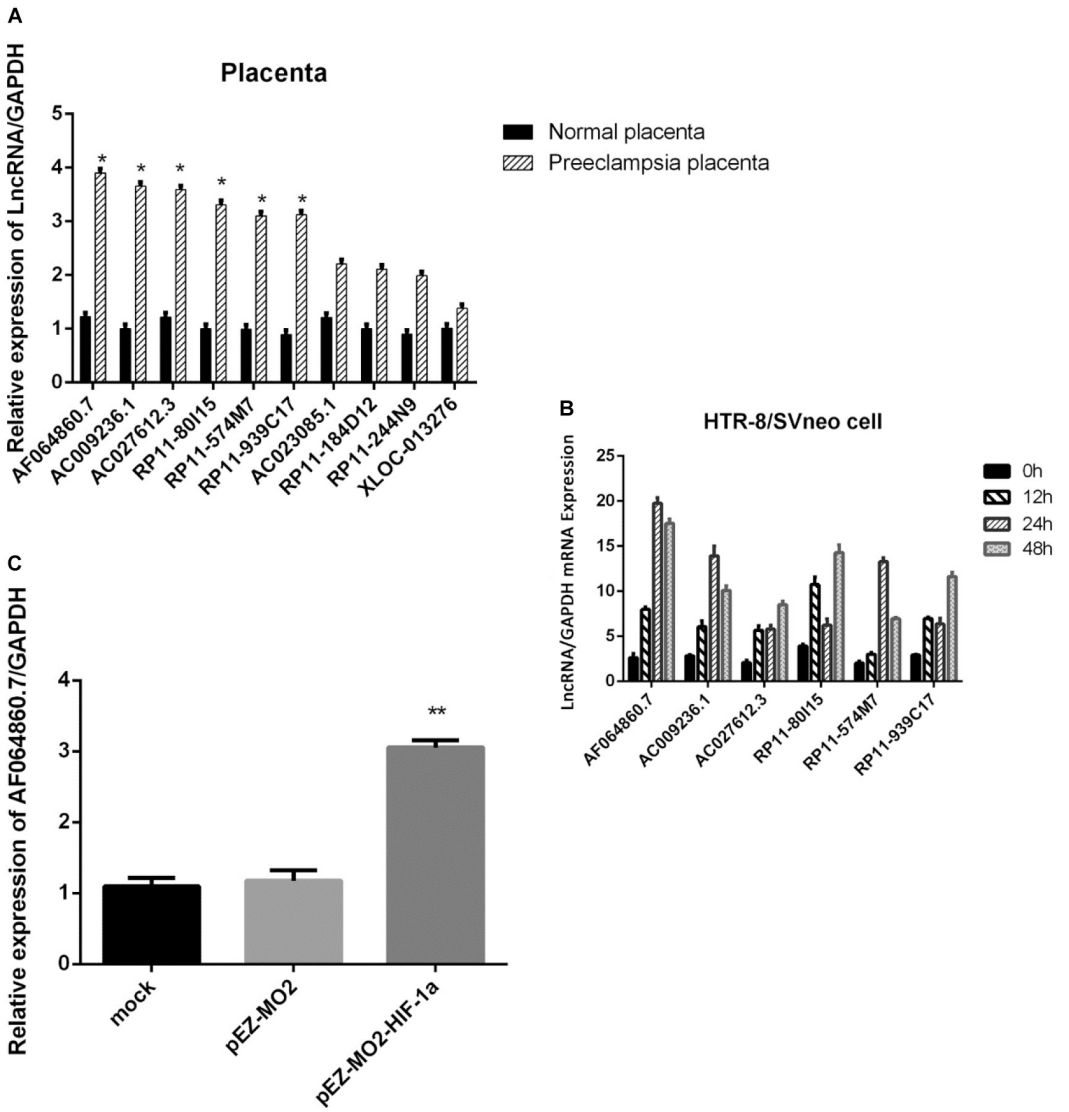


FIGURE 3 | (A) Verification of 10 lncRNAs related with HIF-1a pathway in 30 preeclampsia placenta tissues and 30 normal placenta tissues using qPCR. **(B)** Time-dependent changes of the six candidate lncRNA expression upon hypoxia treatment in HTR-8/SVneo cells. **(C)** The overexpression of HIF-1α increased the expression of lncRNA-AF064860.7 (lncRNA-HEIPP) in HTR-8/SVneo cells. $^{*}p < 0.05$, $^{**}p < 0.01$.

Binding of H3K4me3 in *lncRNA-HEIPP* Promoter Was Increased by Hypoxia Treatment

Enrichment of histone trimethylation at lysine 4 (H3K4me3) at *lncRNA-HEIPP* DMRs/promoters in HTR-8/SVneo human trophoblast cells that underwent hypoxia or normoxia was assessed using ChIP-qPCR. It was shown that the binding of H3K4me3 on *lncRNA-HEIPP* promoter was specifically enriched compared with the binding of IgG (**Figure 5A**). As shown in **Figure 5B**, the protein expression of H3K4me3 was not affected by hypoxia treatment. However, H3K4me3 manifested significantly higher promoter occupancy on *lncRNA-HEIPP* promoter in HTR-8/SVneo cells that underwent hypoxia for 24 h compared to those that underwent normoxia (**Figure 5C**), suggesting the histone modification, instead of the DNA methylation might be the epigenetic pathway that causes the increased expression of lncRNA-HEIPP in the placenta during PE undergone hypoxia.

Downregulation of lncRNA-HEIPP Promoted Trophoblast Invasion

Next, we assessed the biological functions of lncRNA-HEIPP in trophoblast cells. We knocked down the expression of lncRNA-HEIPP in HTR-8/SVneo human trophoblast cell line using HEIPP siRNA (si-HEIPP) and found that the proliferation of HTR8/SVneo cells was not affected (data not shown). However, the invasion of HTR8/SVneo cells was significantly promoted, compared to the cells treated with NC siRNA (**Figure 6**).

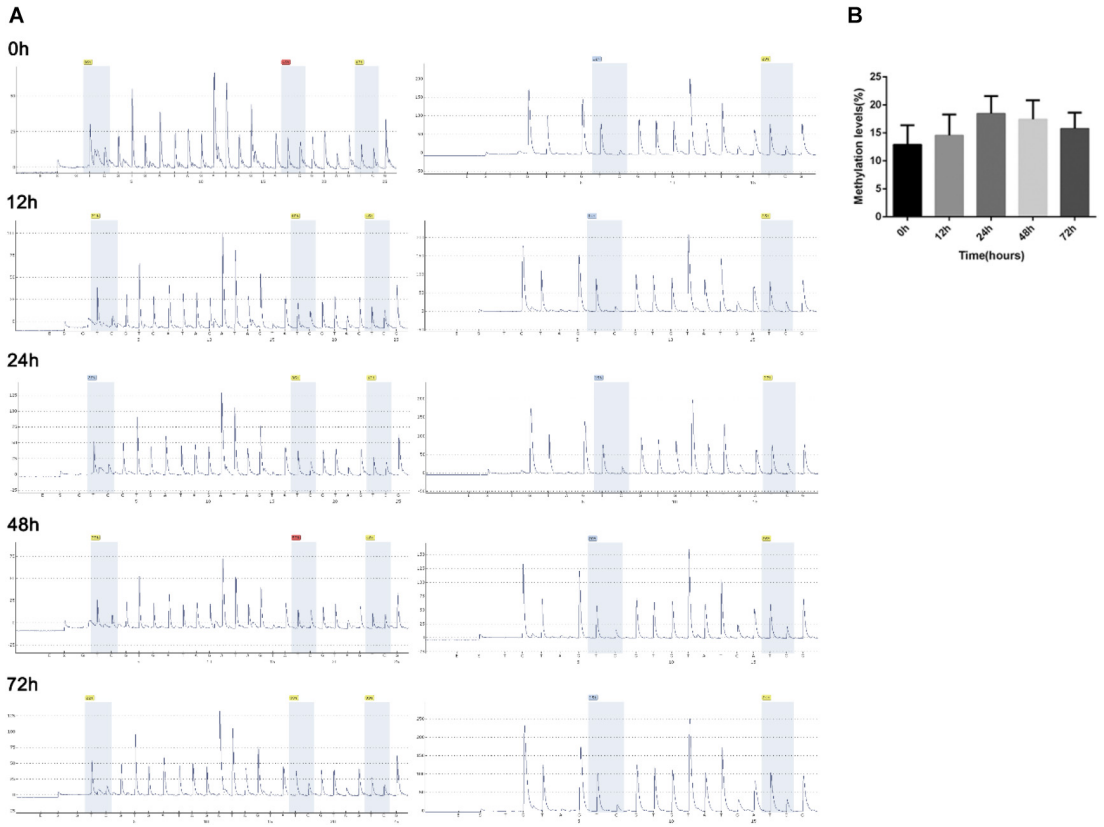


FIGURE 4 | Methylation levels of *lncRNA-HEIPP* promoter in HTR-8/SVneo human trophoblast cells that underwent hypoxia. **(A)** The degree of methylation at each CpG site of *lncRNA-HEIPP* promoter between normoxia (left panel) and hypoxia (right panel) treatment at different time point. **(B)** The statistical graph showing the changes of methylation level of *lncRNA-HEIPP* promoter in HTR-8/SVneo human trophoblast cells that underwent hypoxia at different time point. $n = 3$, and all samples were analyzed in triplicate.

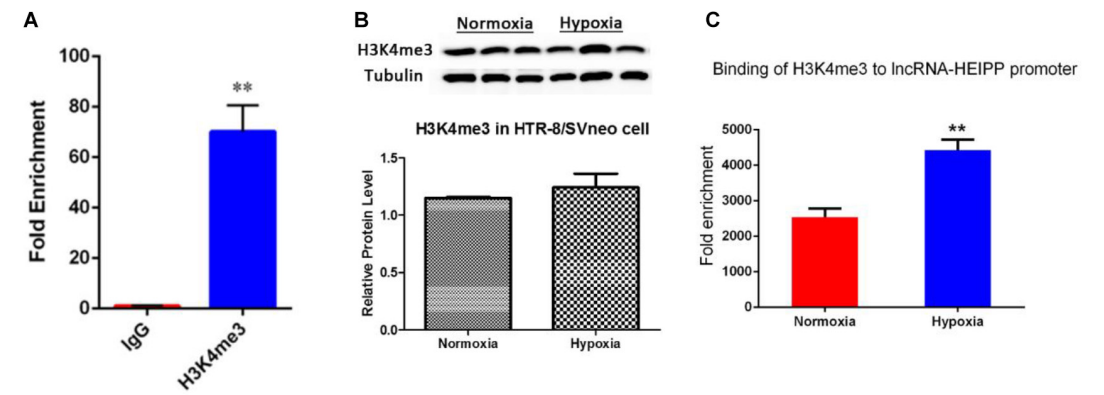


FIGURE 5 | (A) ChIP analysis detected the enrichment of H3K4me3 binding on *lncRNA-HEIPP* promoter. **(B)** The protein expression of H3K4me3 in HTR-8/SVneo cells that underwent hypoxia and normoxia. **(C)** The changes of H3K4me3 occupancy on *lncRNA-HEIPP* promoter in HTR-8/SVneo cells that underwent hypoxia compared to those that underwent normoxia, $^{**}p < 0.01$, $n = 3$.

DISCUSSION

Preeclampsia is a unique disease during pregnancy. It is a placental syndrome characterized by high blood pressure, edema, and proteinuria after 20 weeks of pregnancy (Gormley et al., 2017; Mayrink et al., 2018). Preeclampsia is the leading cause of perinatal and maternal death. At present, the PE incidence rate is 5–8% throughout the world and is 9.4% in China (Boriboonhirunsarn et al., 2017). Current studies strongly suggested that placenta play central roles in the pathogenesis

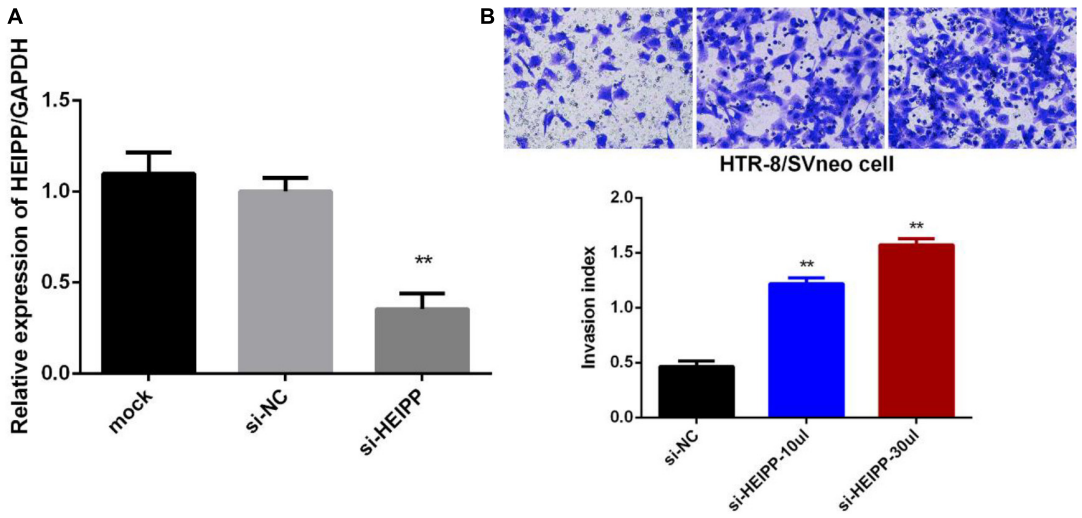


FIGURE 6 | (A) Transfection of si-HEIPP significantly knocked down the expression of lncRNA-HEIPP. **(B)** Downregulation of lncRNA-HEIPP promoted HTR8/SVneo trophoblast cells invasion, $^{**}p < 0.01$.

of PE (Kawasaki et al., 2019). The proliferation and invasion of placental trophoblast cells are the key to successful pregnancy and embryonic development (Knofler and Pollheimer, 2013). The earlier study found that abnormal trophoblast differentiation, migration, and apoptosis often caused the decrease in the invasion of trophoblast cells and the failure of uterine spiral artery recasting, resulting in reduced uteroplacental perfusion and subsequently leading to PE (Kadyrov et al., 2006). However, the factors that affect placental trophoblast cells and the pathogenesis of PE have not been fully elucidated.

In recent years, epigenetic regulation has been revealed to play important roles in the embryonic development and the pathogenesis in many diseases. LncRNAs are a class of transcripts whose lengths exceed 200 nt. Initially, lncRNAs were thought to be transcribed noise. However, an increasing number of studies have reported that these lncRNAs have a series of important functions and participate in the development of many diseases, such as cancer (Liu et al., 2016), cardiovascular illnesses (Leung et al., 2013; Wang et al., 2016), and neurological illnesses (Jiang et al., 2016). Recently, some lncRNAs have been shown to be associated with PE (He et al., 2013; Chen et al., 2015; Zhang et al., 2015). However, their related pathways and regulation were not fully revealed yet. In our current study, we performed lncRNA microarray and found that 1,051 lncRNAs were upregulated, and 344 lncRNAs were downregulated in PE placenta compared with normal placenta, which is comparable to the microarray results from previous study that 738 differently expressed lncRNAs in the placentas of PE in comparison with normal pregnancies (He et al., 2013), but is much less than another study in which 15,646 upregulated and 13,178 downregulated lncRNAs were identified in the placenta tissues of early-onset PE patients (Long et al., 2016). However, in the latter study, the tissues were collected from early-onset PE patients, and the controls were preterm birth, and the biased factor, i.e., preterm labor, may partially explain the huge differences in the numbers of differentially expressed lncRNAs between studies.

The decreased trophoblast proliferation, migration, invasion, and stimulated apoptosis constitute the pivotal reasons leading to PE (Fu et al., 2013). Previous studies revealed different lncRNAs that affect cellular functions of trophoblast cells (Cao et al., 2017; Song et al., 2017; Jiao et al., 2018; Li et al., 2018). However, the hypoxia and HIF-1α–induced pathways play central roles in the pathogenesis of PE (Korkes et al., 2017; Harati-Sadegh et al., 2018; Tong and Giussani, 2019), whereas no HIF-1α–related lncRNAs have yet been revealed in the previous studies. In the current study, we first showed that lncRNAs related with HIF-1α pathway were significantly increased in PE placenta compared to normal placenta, among which lncRNA-HEIPP manifested most significant and time-dependent upregulation upon hypoxia treatment. In tumors, it was found that microRNAs and lncRNAs are important in enabling the key hypoxia-regulated processes (Ferdin et al., 2013; Choudhry et al., 2016). In placenta, some microRNAs, such as microRNA-218 and microRNA-18b, were shown to be induced by hypoxia and HIF-1a in PE patients (Fang et al., 2017; Wang et al., 2017). To our knowledge, lncRNA-HEIPP is the first lncRNA that was found to be upregulated by HIF-1a in placental tissues of PE patients. Moreover, Wang et al. (2018) also found that some lncRNAs, such as NONHSAT116812 and NONHSAT145880, were also confirmed in plasma specimens, making them to be potentially utilized as indicators for PE. Whether lncRNA-HEIPP can also be detected in plasma and used as novel indicators for PE warranted further confirmation.

In recent years, epigenetic regulation of placental gene expression has been considered to play an important role in human trophoblast differentiation (Gamage et al., 2018) and contribute to the onset of PE (Chelbi and Vaiman, 2008; Leavey et al., 2018). The studies showed higher expression levels of death domain-associated protein 6 (DAXX), MMP9,

and BCL2 in placenta from PE tissue, and this increased expression was well correlated to promoter demethylation (Wang et al., 2010; Novakovic et al., 2017; Mohammadpour-Gharehbagh et al., 2019). However, a significant number of the observed methylation changes were not associated with corresponding changes in gene expression, and vice versa (Gamage et al., 2018; Leavey et al., 2018), indicating that alternate methods of epigenetic regulation will need to be explored in PE. Hypoxia was shown to alter the epigenetic profile in cultured human placental trophoblasts (Yuen et al., 2013). In this study, the DNA methylation levels of lncRNA-HEIPP promoter manifested increase upon hypoxia treatment, but the increases showed no significance at each time point after hypoxia. However, it was shown that the H3K4me3 occupancy of *lncRNA-HEIPP* promoter was significantly enriched in the PE placenta compared to the normal placenta. H3K4me3 is a histone modification marker that usually activates the gene transcription upon its binding to the promoter regions (Sati et al., 2012; Zhang et al., 2018; Fanucchi et al., 2019). These results suggested that the histone modification, instead of the DNA methylation, might be the epigenetic pathway that causes the increased expression of lncRNA-HEIPP in the placenta during PE.

Furthermore, we demonstrated that lncRNA-HEIPP knockdown promoted the invasion of trophoblasts. LncRNAs seem to contribute to PE via different pathways. LncRNA PRNCR1 promoted the progression of eclampsia by regulating the MAPK signal pathway (Jiao et al., 2018), while LncRNA CCAT1 promotes the progression of PE by regulating CDK4 (Li et al., 2018). As hypoxia usually induced the activation of JAK/STAT3 signaling pathway to promote trophoblast cell viability and angiogenesis in PE (Xu et al., 2017), whether the cellular functions of lncRNA-HEIPP were also mediated by JAK/STAT3 signaling pathway needs to be further investigated.

CONCLUSION

Our findings suggest that HIF-1a–related lncRNA-HEIPP is upregulated in the placenta from women who underwent PE mediated by the enriched binding of H3K4me3 in lncRNA-HEIPP promoter. LncRNA-HEIPP hampers the invasion of trophoblast cells and contributes to the pathogenesis of PE. Our results demonstrated that HIF-1a pathway–related lncRNA-HEIPP expression profile is a potentially new biomarker, and the manipulation of lncRNA-HEIPP may provide theoretical basis for prevention and treatment of PE.

DATA AVAILABILITY STATEMENT

The microarray data has been uploaded to the GEO dataset and the accession number is GSE160888.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Research Ethics Committee of the Shanghai Changzheng Hospital at the Second Military Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NS, HC, and YM: validation, formal analysis, and investigation. YM: writing – original draft. WP: methodology and investigation. XW and QZ: validation. LG: conceptualization, writing – review and editing, project administration, and funding acquisition. WL: conceptualization, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Project Nos. 2018YFC1002802 (WL) and 2017YFC1001404 (LG and WL), National Natural Science Foundation of China Nos. 81901482 (NS), 81622020 and 81771608 (LG), and 81873821 (WL), Program of Shanghai Academic/Technology Research Leader No. 18XD1405100 (LG), "Dawn" Program of Shanghai Education Commission No. 17SG36 (LG), Program of Shanghai Subject Chief Scientist No. 2017BR017 (LG), and the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (LG).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.559478/full#supplementary-material

REFERENCES

Alici Davutoglu, E., Akkaya Firat, A., Ozel, A., Yilmaz, N., Uzun, I., Temel Yuksel, I., et al. (2018). Evaluation of maternal serum hypoxia inducible factor-1alpha, progranulin and syndecan-1 levels in pregnancies with early- and late-onset preeclampsia. *J. Maternal-fetal Neonatal Med.* 31, 1976–1982. doi: 10.1080/14767058.2017.1333098

Aouache, R., Biquard, L., Vaiman, D., and Miralles, F. (2018). Oxidative stress in preeclampsia and placental diseases. *Int. J. Mol. Sci.* 19:1496. doi: 10.3390/ijms19051496

Apicella, C., Ruano, C. S. M., Mehats, C., Miralles, F., and Vaiman, D. (2019). The role of epigenetics in placental development and the etiology of preeclampsia. *Int. J. Mol. Sci.* 20:2837. doi: 10.3390/ijms20112837

Armaly, Z., Jadaon, J. E., Jabbour, A., and Abassi, Z. A. (2018). Preeclampsia: novel mechanisms and potential therapeutic approaches. *Front. Physiol.* 9:973. doi: 10.3389/fphys.2018.00973

Boriboonhirunsarn, D., Pradyachaipimol, A., and Viriyapak, B. (2017). Incidence of superimposed preeclampsia among pregnant Asian women with chronic hypertension. *Hypertens. Pregnancy* 36, 226–231. doi: 10.1080/10641955.2017.1311340

Brosens, I., Puttemans, P., and Benagiano, G. (2019). Placental bed research: i. the placental bed: from spiral arteries remodeling to the great obstetrical syndromes. *Am. J. Obstetr. Gynecol.* 221, 437–456. doi: 10.1016/j.ajog.2019.05.044

Cao, C., Li, J., Li, J., Liu, L., Cheng, X., and Jia, R. (2017). Long non-coding RNA Uc.187 is upregulated in preeclampsia and modulates proliferation, apoptosis, and invasion of HTR-8/SVneo trophoblast cells. *J. Cell. Biochem.* 118, 1462–1470. doi: 10.1002/jcb.25805

Chelbi, S. T., and Vaiman, D. (2008). Genetic and epigenetic factors contribute to the onset of preeclampsia. *Mol. Cell. Endocrinol.* 282, 120–129. doi: 10.1016/j.mce.2007.11.022

Chen, H., Meng, T., Liu, X., Sun, M., Tong, C., Liu, J., et al. (2015). Long non-coding RNA MALAT-1 is downregulated in preeclampsia and regulates proliferation, apoptosis, migration and invasion of JEG-3 trophoblast cells. *Int. J. Clin. Exp. Pathol.* 8, 12718–12727.

Chen, J., and Khalil, R. A. (2017). Matrix metalloproteinases in normal pregnancy and preeclampsia. *Prog. Mol. Biol. Transl. Sci.* 148, 87–165. doi: 10.1016/bs.pmbts.2017.04.001

Choudhry, H., Harris, A. L., and McIntyre, A. (2016). The tumour hypoxia induced non-coding transcriptome. *Mol. Aspects Med.* 4, 35–53. doi: 10.1016/j.mam.2016.01.003

Cotechini, T., Komisarenko, M., Sperou, A., Macdonald-Goodfellow, S., Adams, M. A., and Graham, C. H. (2014). Inflammation in rat pregnancy inhibits spiral artery remodeling leading to fetal growth restriction and features of preeclampsia. *J. Exp. Med.* 211, 165–179. doi: 10.1084/jem.20130295

Deyssenroth, M. A., Peng, S., Hao, K., Lambertini, L., Marsit, C. J., and Chen, J. (2017). Whole-transcriptome analysis delineates the human placenta gene network and its associations with fetal growth. *BMC Genom.* 18:520. doi: 10.1186/s12864-017-3878-0

Fang, M., Du, H., Han, B., Xia, G., Shi, X., Zhang, F., et al. (2017). Hypoxia-inducible microRNA-218 inhibits trophoblast invasion by targeting LASP1: implications for preeclampsia development. *Int. J. Biochem. Cell Biol.* 87, 95–103. doi: 10.1016/j.biocel.2017.04.005

Fanucchi, S., Fok, E. T., Dalla, E., Shibayama, Y., Borner, K., Chang, E. Y., et al. (2019). Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat. Genet.* 51, 138–150. doi: 10.1038/s41588-018-0298-2

Ferdin, J., Nishida, N., Wu, X., Nicoloso, M. S., Shah, M. Y., Devlin, C., et al. (2013). HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts. *Cell Death. Differ.* 20, 1675–1687. doi: 10.1038/cdd.2013.119

Fu, G., Ye, G., Nadeem, L., Ji, L., Manchanda, T., Wang, Y., et al. (2013). MicroRNA-376c impairs transforming growth factor-beta and nodal signaling to promote trophoblast cell proliferation and invasion. *Hypertension* 61, 864–872. doi: 10.1161/hypertensionaha.111.203489

Gamage, T., Schierding, W., Hurley, D., Tsai, P., Ludgate, J. L., Bhoothpur, C., et al. (2018). The role of DNA methylation in human trophoblast differentiation. *Epigenetics* 13, 1154–1173. doi: 10.1080/15592294.2018.1549462

Gormley, M., Ona, K., Kapidzic, M., Garrido-Gomez, T., Zdravkovic, T., and Fisher, S. J. (2017). Preeclampsia: novel insights from global RNA profiling of trophoblast subpopulations. *Am. J. Obstet. Gynecol.* 217, 200.e1–200.e17.

Gutierrez, J. A., Gomez, I., Chiarello, D. I., Salsoso, R., Klein, A. D., Guzman-Gutierrez, E., et al. (2019). Role of proteases in dysfunctional placental vascular remodelling in preeclampsia. *Biochimica et Biophysica Acta. Mol. Basis Dis.* 1866:165448. doi: 10.1016/j.bbadis.2019.04.004

Harati-Sadegh, M., Kohan, L., Teimoori, B., Mehrabani, M., and Salimi, S. (2018). The association of the placental Hypoxia-inducible factor1-alpha polymorphisms and HIF1-alpha mRNA expression with preeclampsia. *Placenta* 67, 31–37. doi: 10.1016/j.placenta.2018.05.005

He, X., He, Y., Xi, B., Zheng, J., Zeng, X., Cai, Q., et al. (2013). LncRNAs expression in preeclampsia placenta reveals the potential role of LncRNAs contributing to preeclampsia pathogenesis. *PLoS One* 8:e81437. doi: 10.1371/journal.pone.0081437

Hosseini, A., Dolati, S., Hashemi, V., Abdollahpour-Alitappeh, M., and Yousefi, M. (2018). Regulatory T and T helper 17 cells: their roles in preeclampsia. *J. Cell. Physiol.* 233, 6561–6573. doi: 10.1002/jcp.26604

Jiang, Q., Shan, K., Qun-Wang, X., Zhou, R. M., Yang, H., Liu, C., et al. (2016). Long non-coding RNA-MIAT promotes neurovascular remodeling in the eye and brain. *Oncotarget* 7, 49688–49698. doi: 10.18632/oncotarget.10434

Jiao, S., Wang, S. Y., and Huang, Y. (2018). LncRNA PRNCR1 promoted the progression of eclampsia by regulating the MAPK signal pathway. *Eur. Rev. Med. Pharmacol. Sci.* 22, 3635–3642.

Kadyrov, M., Kingdom, J. C., and Huppertz, B. (2006). Divergent trophoblast invasion and apoptosis in placental bed spiral arteries from pregnancies complicated by maternal anemia and early-onset preeclampsia/intrauterine growth restriction. *Am. J. Obstet. Gynecol.* 194, 557–563. doi: 10.1016/j.ajog.2005.07.035

Kawasaki, K., Kondoh, E., Chigusa, Y., Kawamura, Y., Mogami, H., Takeda, S., et al. (2019). Metabolomic profiles of placenta in preeclampsia. *Hypertension* 73, 671–679. doi: 10.1161/hypertensionaha.118.12389

Knofler, M., and Pollheimer, J. (2013). Human placental trophoblast invasion and differentiation: a particular focus on Wnt signaling. *Front. Genet.* 4:190. doi: 10.3389/fgene.2013.00190

Korkes, H. A., De Oliveira, L., Sass, N., Salahuddin, S., Karumanchi, S. A., and Rajakumar, A. (2017). Relationship between hypoxia and downstream pathogenic pathways in preeclampsia. *Hypertens. Pregnancy* 36, 145–150. doi: 10.1080/10641955.2016.1259627

Kuzmanov, A., Wielockx, B., Rezaei, M., Kettelhake, A., and Breier, G. (2012). Overexpression of factor inhibiting HIF-1 enhances vessel maturation and tumor growth via platelet-derived growth factor-C. *Int. J. Cancer* 131, E603–E613.

Kwak, Y. T., Muralimanoharan, S., Gogate, A. A., and Mendelson, C. R. (2019). Human trophoblast differentiation is associated with profound gene regulatory and epigenetic changes. *Endocrinology* 160, 2189–2203. doi: 10.1210/en.2019-00144

Leavey, K., Wilson, S. L., Bainbridge, S. A., Robinson, W. P., and Cox, B. J. (2018). Epigenetic regulation of placental gene expression in transcriptional subtypes of preeclampsia. *Clin. Epigenet.* 10:28.

Leung, A., Trac, C., Jin, W., Lanting, L., Akbany, A., Saetrom, P., et al. (2013). Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. *Circ. Res.* 113, 266–278. doi: 10.1161/circresaha.112.300849

Li, J. L., Li, R., Gao, Y., Guo, W. C., Shi, P. X., and Li, M. (2018). LncRNA CCAT1 promotes the progression of preeclampsia by regulating CDK4. *Eur. Rev. Med. Pharmacol. Sci.* 22, 1216–1223.

Liu, D., Yu, X., Wang, S., Dai, E., Jiang, L., Wang, J., et al. (2016). The gain and loss of long noncoding RNA associated-competing endogenous RNAs in prostate cancer. *Oncotarget* 7, 57228–57238. doi: 10.18632/oncotarget.11128

Long, W., Rui, C., Song, X., Dai, X., Xue, X., Lu, Y., et al. (2016). Distinct expression profiles of lncRNAs between early-onset preeclampsia and preterm controls. *Clin. Chimica Acta; Int. J. Clin. Chem.* 463, 193–199. doi: 10.1016/j.cca.2016.10.036

Mayrink, J., Costa, M. L., and Cecatti, J. G. (2018). Preeclampsia in 2018: revisiting concepts, physiopathology, and prediction. *Sci. World J.* 2018:6268276.

Mohammadpour-Gharehbagh, A., Jahantigh, D., Eskandari, M., Sadegh, M. H., Nematollahi, M. H., Rezaei, M., et al. (2019). Genetic and epigenetic analysis of the BAX and BCL2 in the placenta of pregnant women complicated by preeclampsia. *Apoptosis : Int. J. Program. Cell Death* 24, 301–311. doi: 10.1007/s10495-018-1501-8

Novakovic, B., Evain-Brion, D., Murthi, P., Fournier, T., and Saffery, R. (2017). Variable DAXX gene methylation is a common feature of placental trophoblast differentiation, preeclampsia, and response to hypoxia. *FASEB J.* 31, 2380–2392. doi: 10.1096/fj.201601189rr

Sati, S., Ghosh, S., Jain, V., Scaria, V., and Sengupta, S. (2012). Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* 40, 10018–10031. doi: 10.1093/nar/gks776

Song, X., Rui, C., Meng, L., Zhang, R., Shen, R., Ding, H., et al. (2017). Long non-coding RNA RPAIN regulates the invasion and apoptosis of trophoblast cell lines via complement protein C1q. *Oncotarget* 8, 7637–7646. doi: 10.18632/oncotarget.13826

Tong, W., and Giussani, D. A. (2019). Preeclampsia link to gestational hypoxia. *J. Dev. Origins Health Dis.* 10, 322–333. doi: 10.1017/s204017441900014x

Wang, S., Wang, X., Weng, Z., Zhang, S., Ning, H., and Li, B. (2017). Expression and role of microRNA 18b and hypoxia inducible factor-1alpha in placental tissues of preeclampsia patients. *Exp. Therapeutic Med.* 14, 4554–4560.

Wang, X., Chen, Y., Du, L., Li, X., Li, X., and Chen, D. (2018). Evaluation of circulating placenta-related long noncoding RNAs as potential biomarkers for preeclampsia. *Exp. Therapeutic Med.* 15, 4309–4317.

Wang, Y. N., Shan, K., Yao, M. D., Yao, J., Wang, J. J., Li, X., et al. (2016). Long Noncoding RNA-GAS5: a novel regulator of hypertension-induced vascular remodeling. *Hypertension* 68, 736–748. doi: 10.1161/hypertensionaha.116.07259

Wang, Z., Lu, S., Liu, C., Zhao, B., Pei, K., Tian, L., et al. (2010). Expressional and epigenetic alterations of placental matrix metalloproteinase 9 in preeclampsia. *Gynecol. Endocrinol.* 26, 96–102. doi: 10.3109/09513590903184100

Xu, C., Li, X., Guo, P., and Wang, J. (2017). Hypoxia-Induced activation of JAK/STAT3 signaling pathway promotes trophoblast cell viability and angiogenesis in preeclampsia. *Med. Sci. Monitor : Int. Med. J. Exp. Clin. Res.* 23, 4909–4917. doi: 10.12659/msm.905418

Xu, Y., Lian, Y., Zhang, Y., Huang, S., Zuo, Q., Yang, N., et al. (2018). The long non-coding RNA PVT1 represses ANGPTL4 transcription through binding with EZH2 in trophoblast cell. *J. Cell Mol. Med.* 22, 1272–1282.

Yong, H. E. J., Murthi, P., Brennecke, S. P., and Moses, E. K. (2018). Genetic approaches in preeclampsia. *Methods Mol. Biol.* 1710, 53–72. doi: 10.1007/978-1-4939-7498-6_5

Yuen, R. K., Chen, B., Blair, J. D., Robinson, W. P., and Nelson, D. M. (2013). Hypoxia alters the epigenetic profile in cultured human placental trophoblasts. *Epigenetics* 8, 192–202. doi: 10.4161/epi.23400

Zhang, E., He, X., Zhang, C., Su, J., Lu, X., Si, X., et al. (2018). A novel long noncoding RNA HOXC-AS3 mediates tumorigenesis of gastric cancer by binding to YBX1. *Genome Biol.* 19:154.

Zhang, Y., Zhao, H. J., Xia, X. R., Diao, F. Y., Ma, X., Wang, J., et al. (2019). Hypoxia-induced and HIF1alpha-VEGF-mediated tight junction dysfunction in choriocarcinoma cells: implications for preeclampsia. *Clinica Chimica Acta; Int. J. Clin. Chem.* 489, 203–211. doi: 10.1016/j.cca.2017.12.010

Zhang, Y., Zou, Y., Wang, W., Zuo, Q., Jiang, Z., Sun, M., et al. (2015). Down-regulated long non-coding RNA MEG3 and its effect on promoting apoptosis and suppressing migration of trophoblast cells. *J. Cell. Biochem.* 116, 542–550. doi: 10.1002/jcb.25004

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: 11 January 2021
doi: 10.3389/fgene.2020.599834

An Integrated Transcriptome Analysis Reveals IGFBP7 Upregulation in Vasculature in Traumatic Brain Injury

Jianhao Wang[1†], Xiangyi Deng[1†], Yuan Xie[2†], Jiefu Tang[3], Ziwei Zhou[1], Fan Yang[1], Qiyuan He[2], Qingze Cao[2], Lei Zhang[2,4] and Liqun He[1,5*]*

[1] Key Laboratory of Post-Neuroinjury Neuro-Repair and Regeneration in Central Nervous System, Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin Neurological Institute, Ministry of Education and Tianjin City, Tianjin, China, [2] Key Laboratory of Ministry of Education for Medicinal Plant Resource and Natural Pharmaceutical Chemistry, National Engineering Laboratory for Resource Developing of Endangered Chinese Crude Drugs in Northwest of China, College of Life Sciences, Shaanxi Normal University, Xi'an, China, [3] Trauma Center, First Affiliated Hospital of Hunan University of Medicine, Huaihua, China, [4] Precision Medicine Center, The Second People's Hospital of Huaihua, Huaihua, China, [5] Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

OPEN ACCESS

Edited by:
Cheng Peng,
Yunnan University, China

Reviewed by:
Andre Obenaus,
University of California, Irvine, United States
Hadijat M. Makinde,
Northwestern University, United States

***Correspondence:**
Lei Zhang
zlsnnu@gmail.com
Liqun He
liqun.he@igp.uu.se

[†] These authors have contributed equally to this work

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

Received: 29 August 2020
Accepted: 30 November 2020
Published: 11 January 2021

Citation:
Wang J, Deng X, Xie Y, Tang J, Zhou Z, Yang F, He Q, Cao Q, Zhang L and He L (2021) An Integrated Transcriptome Analysis Reveals IGFBP7 Upregulation in Vasculature in Traumatic Brain Injury. Front. Genet. 11:599834. doi: 10.3389/fgene.2020.599834

Vasculature plays critical roles in the pathogenesis and neurological repair of traumatic brain injury (TBI). However, how vascular endothelial cells respond to TBI at the molecular level has not been systematically reviewed. Here, by integrating three transcriptome datasets including whole cortex of mouse brain, FACS-sorted mouse brain endothelial cells, and single cell sequencing of mouse brain hippocampus, we revealed the key molecular alteration of endothelial cells characterized by increased Myc targets and Epithelial-Mesenchymal Transition signatures. In addition, immunofluorescence staining of patients' samples confirmed that IGFBP7 was up-regulated in vasculature in response to TBI. TGFβ1, mainly derived from microglia and endothelial cells, sufficiently induces IGFBP7 expression in cultured endothelial cells, and is significantly upregulated in response to TBI. Our results identified IGFBP7 as a potential biomarker of vasculature in response to TBI, and indicate that TGFβ signaling may contribute to the upregulation of IGFBP7 in the vasculature.

Keywords: traumatic brain injury, vasculature, endothelial cell, IGFBP7, TGFβ

INTRODUCTION

Traumatic brain injury (TBI), one of the leading cause of morbidity and disability, accounts for 30% of all injury-related deaths (Maas et al., 2008). It has been estimated that in the United States, an estimated 1.7 million people experience TBI annually, and 5.3 million people suffer TBI-related complications and sequela including long-term neurological and psychiatric disorders, chronic inflammation, and chronic traumatic encephalopathy (Nguyen et al., 2016; Jassam et al., 2017). TBI is a complicated pathophysiological process that can be divided into primary and secondary brain injuries (Sun et al., 2017). Primary injury is the direct damage of neural tissue caused by mechanical effect occurring at the moment of trauma (Silverberg et al., 2019). By contrast, secondary injury is the indirect injurious biochemical cascade initiated by primary insult, significantly contributing to the aggravation and high mortality of TBI

(Bramlett and Dietrich, 2015; Silverberg et al., 2019). Secondary brain injury is mediated by several cellular and molecular pathways including excitotoxicity, inflammation, oxidative stress, and energy failure, and cerebral vasculature is a critical player to regulate these pathological processes (Jassam et al., 2017; Salehi et al., 2017; Simon et al., 2017). There is an increasing interest in the role that brain vasculature plays in the pathogenesis of TBI. Direct disruption of cerebral vasculature at the time of head impact leads to hemorrhage and blood flow abnormalities immediately after trauma, and dysfunction of vasculature leads to additional insults such as hypoxemia, hypoxia, hypo-perfusion, ischemia, and blood brain barrier (BBB) breakdown (Jullienne et al., 2016).

BBB is a specialized vascular structure in brain that maintains homeostasis and regulates the movement of molecular and cells across the brain vasculature (Yeoh et al., 2013). BBB disruption, as assessed by cerebrospinal fluid/serum albumin quotient, is found in 44% of non-penetrating TBI patients (Ho et al., 2014). Alteration of the vasculature in TBI causes vasogenic edema at both lesion and surrounding tissues resulting in tissue swelling and elevated intracranial pressure, contributing to about 50% of death in severe head injury (Marmarou, 2003). Moreover, increased vascular permeability observed in TBI patient allows harmful molecular and blood toxins into brain, which may lead to neuronal damage and long-lasting functional deficits (Badaut et al., 2011). In addition, up-regulation of cytokines/chemokines and leukocyte adhesion molecules in vascular endothelial cells (EC) augment inflammation and further increase the risk of edema formation and neuronal dysfunction (Ziebell and Morganti-Kossmann, 2010; Simon et al., 2017). Therefore, vasculature plays a critical role in the pathogenesis of TBI, and therapeutical manipulation of vasculature may represent a potential way for TBI treatment. However, how vascular EC respond to TBI at the molecular level has not been systematically reviewed.

Here, we integrated three transcriptome datasets in which the responses of whole cortex, FACS-sorted ECs, and isolated single ECs to brain injury were documented. Unbiased comparison by aligning three data sets to the same mouse genome revealed the key molecular alteration of ECs in response to brain injury. In addition, IGFBP7 was identified as a potential biomarker of vasculature in TBI pathogenesis.

MATERIALS AND METHODS

Sequencing Data Collection

The original sequencing data from three independent studies were obtained from the NCBI Sequence Read Archive (SRA) database[1], including the mouse whole brain cortex RNAseq data (SRP072117, six samples), the FACS-sorted mouse brain endothelial RNAseq data (SRP100777, six samples), and the mouse brain hippocampus single cell RNAseq data (SRP113600, six samples). Each sample contains about 40–60 million RNA sequence reads.

[1]www.ncbi.nlm.nih.gov/sra

Bulk RNAseq Data Analysis

The RNAseq sequencing data from the brain cortex and FACS-sorted EC were aligned to the mouse genome assembly (GRCm38) obtained from the Ensembl database using the TopHat2 software (version 2.1.1, with the following parameters: tophat2 -p 8 –keep-fasta-order –GTF < reference.gtf > < reference.genome > < read1.fastq > < read2.fastq >) (Kim et al., 2013). To quantify the differential expression between TBI and control groups, the default cuffdiff tests were performed using the Cufflinks tool (version 2.2.1, with the following parameters: cuffdiff -p 4 < reference.gtf > < control1.bam,control2.bam,control3.bam > < TBI1.bam,TBI2.bam,TBI3.bam >) (Trapnell et al., 2012). The genes with statistical multiple-test corrected (Benjamini-Hochberg) *p*-value smaller than 0.05 and more than 2-fold difference were selected as significant differentially expressed genes.

To compare the gene expression levels among different samples, heatmap visualization was performed using the pheatmap packages (version 1.0.12) with ward.D2 method. The genes were clustered using the Pearson correlation distance and the samples were clustered using the Euclidean distance.

Single Cell RNAseq Data Analysis

The single cell RNAseq data were processed using the Drop-seq tools (version 2.3.0)[2] with default parameters as described in the paper (Arneson et al., 2018). The fastq files were aligned to the mouse genome assembly (GRCm38) and the digital expression counts for each cell were quantified using the default parameters. The single cell expression counts data were then imported into the Seurat package (version: 3.1.1) for cell type clustering and cluster marker identifications (Butler et al., 2018).

Gene Set Variation Analysis (GSVA)

The GSVA was performed using the R GSVA package (version 1.32.0). To identify the gene sets with significant changes, the 50 hallmark gene sets from the MSigDB collections were tested[3] using the R limma package (version 3.40.6). The gene sets with statistical multiple-test (Benjamini-Hochberg) corrected *p*-value smaller than 0.05 were identified as significant.

Human TBI Sample Collection

The brain tissue was collected from the Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin 300052, PR China. The patient was an adult male. The preoperative Glasgow Coma Scale (GCS) was 11 points without underlying disease. The written informed consent from the patient was made before the surgery.

During the operation, the contusion tissue was extracted from the contusion area by bipolar electrocoagulation, and the brain contusion tissue was marked as TBI group. The control specimens were taken from the non-functional area about 1.5 cm from the tumor boundary of the glioblastoma patient. After removing the specimens, they were stored at -80°C. All

[2]http://mccarrolllab.org/dropseq/

[3]www.gsea-msigdb.org

human sample experiments were approved by the local hospital ethics committee.

Immunofluorescence Staining

After the specimens were taken out from -80°C, they were quickly embedded in optimal cutting temperature (O.C.T., Sakura, Oakland, CA, United States). Then, the specimens were cut into a thick coronal section of 6 μm, and immunofluorescence stained to determine the IGFBP7 and CD31 expression in the coronal brain section. The slices were first fixed in ice-cold methanol for 10 min and washed with phosphate buffered saline (PBS) for 10 min. Blocked with 3% BSA (Invitrogen) for 1 h at room temperature (RT). Then, the sections were incubated with the primary antibody against IGFBP7 (1:200, ab74169, Abcam) and CD31 (1:500, MA3105, Invitrogen) overnight at 4℃. The slides were washed three times with PBS plus 0.2% Tween (Sigma) and incubated with the secondary antibody Goat anti-Rabbit Alexa 555 (1:400, cat A-21428, Invitrogen) and Goat anti-Armenian hamster Alexa 488 (1:400, ab173003, Abcam) for 1 h at RT. After washing with PBS plus 0.2% Tween three times, the sections were mounted in Fluoromount (F4680, Sigma) containing 0.1% hoechst 33258 (14530, Sigma). Images were acquired by fluorescence microscope (Zeiss, Germany).

In vitro Stimulation of EC

Murine brain EC (bEND.3) were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (Biological Industries) and 1% 100 mM Sodium Pyruvate (Gibco). Human umbilical vein EC (HUVECs) were maintained up to passage 14 in the same culture medium as described above. The bEND.3 cells and HUVECs were starved in DMEM with 1% FBS or 0.5% FBS, respectively, overnight and stimulated with TGFβ1 (2 ng/mL; Peprotech) for 72 h, VEGFA (50 ng/mL; Peprotech) for 72 h, or cobalt chloride (400 μM $CoCl_2$; Sigma Aldrich) for 24 h, respectively. The experiment was performed duplicates and repeated three times for both bEND.3 cells and HUVECs.

Quantitative Real-Time Q-PCR

Total cell RNA was extracted with the RNeasy kit (Qiagen) and reverse-transcribed with the PrimeScript RT Master Mix (Takara). qPCR was performed on a Thermal Cycler iQ5 Multicolor Real-Time PCR Detection system (Bio-Rad) using TB Green Premix Ex taqTM II (Takara) and intron-spanning, gene-specific primers as listed below: mouse *Hprt* (forward: CAGTCCCAGCGTCGTGATTA, reverse: TGGCCTCCCATCTCCTTCAT); mouse *Igfbp7* (forward: CTG GTGCCAAGGTGTTCTTGA, reverse: CTCCAGAGTGATCC CTTTTTACC); human *HPRT* (forward: CTTTGCTGACCT GCTGGATT, reverse: TCCCGTGTTGACTGGTCATT); human *IGFBP7* (forward: GCGAGCAAGGTCCTTCCATA, reverse: TCTGAATGGCCAGGTTGTCC). Gene expression was normalized to the house keeping gene *HPRT*(*Hprt*).

RESULTS

Identification of Differentially Expressed Genes in TBI

In order to identify the genes that are differentially expressed in brain in TBI condition, we reanalyzed and integrated the data from three transcriptome sequencing profiling approaches (**Table 1**). These include one mouse whole cortex RNAseq study (Zhong et al., 2016), one mouse brain FACS-sorted endothelial RNAseq study (Munji et al., 2019), and one mouse brain single cell RNAseq study (Arneson et al., 2018). All three studies were performed with mouse TBI models and screened the gene expression changes at 24 h after TBI. The three datasets were processed differently in their corresponding studies. To achieve an unbiased comparison, we obtained the original sequence data for each sample and aligned them to the same mouse genome reference (**Supplementary Table 1**). Differentially expressed genes were identified by comparing the TBI and control in the respective studies. The complete gene analysis results from each of the three studies are listed in the supplements (whole cortex RNAseq study in **Supplementary Table 2**; FACS-sorted EC study in **Supplementary Table 3**; single cell RNAseq study in **Supplementary Table 4**).

In the whole cortex RNAseq study, 1,096 genes were tested as significantly differentially expressed (**Supplementary Table 5**). In the FACS-sorted EC study, 214 genes were identified (**Supplementary Table 6**).

For the single cell RNAseq study, all the qualified cells (N = 6,351, with 2,549 cells from the TBI samples and 3,370 cells from the control samples) were clustered and a pure endothelial cluster (n = 371) was identified (**Supplementary Figure 1**). Within the endothelial clusters, the comparison between the 185 cells in TBI samples and the 186 cells in the control samples yields 14 significantly differentially expressed genes (**Supplementary Table 7**).

Comparison and Integration of the Three Studies

We focused on the brain EC to evaluate and compare the results from three studies. First, we compared the two RNAseq studies (whole cortex and FACS-sorted EC). For a list of 10 well-known canonical EC markers, they showed a wild range of expression in RNAseq data, however, they all display significant enrichment in the FACS sorted EC samples (around 20–100-folds' enrichment) (**Figure 1A**). In the results of differentially expressed genes, there were 184 genes differentially expressed in both studies (**Figure 1B** and **Supplementary Table 8**). Among these 184 genes, 170 genes were regulated in the same direction (both were up-regulated or down-regulated), and only 14 (8%) showed different regulation directions (**Figure 1C**).

To identify the pathways that were affected during the TBI condition, Gene Set Variation Analysis (GSVA) was applied to each of the three datasets. The regulated hallmark gene sets in the two RNAseq studies as well as the major cell types in the single cell RNAseq studies were identified (**Figure 1D**). Myc targets and Epithelial-Mesenchymal Transition signature are enriched

TABLE 1 | Summary of the characteristics of the three studies.

Study	Species and age	Sample	Comparing samples	TBI types; Severity; and time points	Parameters	Methods	Total reads
Zhong et al., 2016	C57BL/6 mouse; 12 weeks	Brain whole cortex	Three TBI samples vs. three controls	CCi; Moderate; 24 h	Diameter: 3 mm, velocity: 5.0 m/s, depth: 2.0 mm, dwelling time:100 ms	RNAseq	319 million
Munji et al., 2019	Rosa-tdTomato mouse; 21 days	FACS purified mouse brain endothelial cells	Three TBI samples vs. three controls	CCi; Focal; 24 h	Diameter: 3 mm, velocity: 4.5 m/s, depth: 1.73 mm, dwelling time:150 ms	RNAseq	344 million
Arneson et al., 2018	C57BL/6 mouse; 10 weeks	Single-cell suspension from brain hippocampus	Three TBI samples vs. three controls	FPI; Mild; 24 h	Diameter: 1.5 mm, velocity: 4.5 m/s, fluid percussion pulse: 1.5–1.7 atm	scRNAseq	384 million

The table lists the three transcriptome sequencing analyses we reintegrated. The similarities and differences of type of TBI, severity, age, time points, and type of mouse are shown in detail. CCi, controlled cortical impact; FPI, fluid percussion injury.

in samples from TBI in all three studies. Interestingly, both signatures are also enriched in tumor EC and associate with tumor angiogenesis (Lambrechts et al., 2018).

A comparison among the three different studies to identify differentially expressed genes (DEGs) in whole brain or brain EC is illustrated in **Figure 2A**. The DEGs of the three different studies are classified according to the up-regulated and down-regulated genes (**Figure 2A**). Totally, 1,135 different genes were identified by at least one study, and only three genes (*Igfbp7*, *Fxyd5*, and *Itm2a*) were consistently regulated in samples from TBI compared to control in all three studies. Among them, *Igfbp7* and *Fxyd5* are up-regulated, and *Itm2a* is down-regulated. Their detailed expression profile in the FACS-sorted EC samples is visualized in **Figure 2B**. All three genes showed EC specific expression in the single cell RNAseq study (**Figure 2C**). Expression of *Igfbp7* and *Itm2a* in the vasculature can further be confirmed in the Allen Brain Atlas database (**Figure 2D**).

To uncover whether alteration of the three genes is TBI specific or reflect a broader response to brain injury, we analyzed the expression of *Igfbp7/IGFBP7*, *Fxyd5/FXYD5*, and *Itm2a/ITM2A* in EC in response to stoke, seizure, and EAE in Munji's study (Munji et al., 2019), as well as in the human glioblastoma vasculature in our previous study (Dieterich et al., 2012). Upregulation of *Fxyd5* in endothelial cells was observed in TBI and seizure. Down-regulation of *Itm2a* in endothelial cells was detected in TBI, seizure, and EAE (**Supplementary Table 9**). *Igfbp7/IGFBP7* expression was up-regulated in all these disease models as well as glioblastoma vasculature (**Supplementary Table 9**), suggesting that its up-regulation is a universal response in EC to pathological alterations rather than TBI-specific.

IGFBP7 Is Upregulated in the Vasculature in Response TBI

Among the three commonly regulated genes in all three studies, *Igfbp7* were up-regulated in all datasets (**Figures 3A–C**), and also showed endothelial enriched expression from single cell data (**Figure 3D**). IGFBP7 has been suggested as a critical regulator for angiogenesis, vessel integrity, and endothelial adhesion molecule (Hooper et al., 2009; Komiya et al., 2014; Rai et al., 2015), which all associate with the pathogenesis of TBI. Thus, we focused on IGFBP7, a matrix bound secreted protein, which belongs to insulin-like growth factor binding protein (IGFBP) family. Immunofluorescence co-staining of CD31 and IGFBP7 in human samples clearly confirms up-regulation of IGFBP7 in vasculature in response to TBI (**Figure 3E**).

TGFβ Induces *Igfbp7/IGFBP7* Expression in EC

To uncover the signal pathway mediating *IGFBP7* upregulation, we analyzed the expression of *Igfbp7/IGFBP7* in EC (bEND.3 cells and HUVECs) upon the stimulation of VEGFA and TGFβ, which can increase *IGFBP7* expression in breast cancer and glioblastoma vasculature respectively (Pen et al., 2008; Komiya et al., 2014). In addition, the effect of hypoxia, the most pronounced characteristic of brain injury causing pathogenesis, on *Igfbp7/IGFBP7* expression in EC was determined by

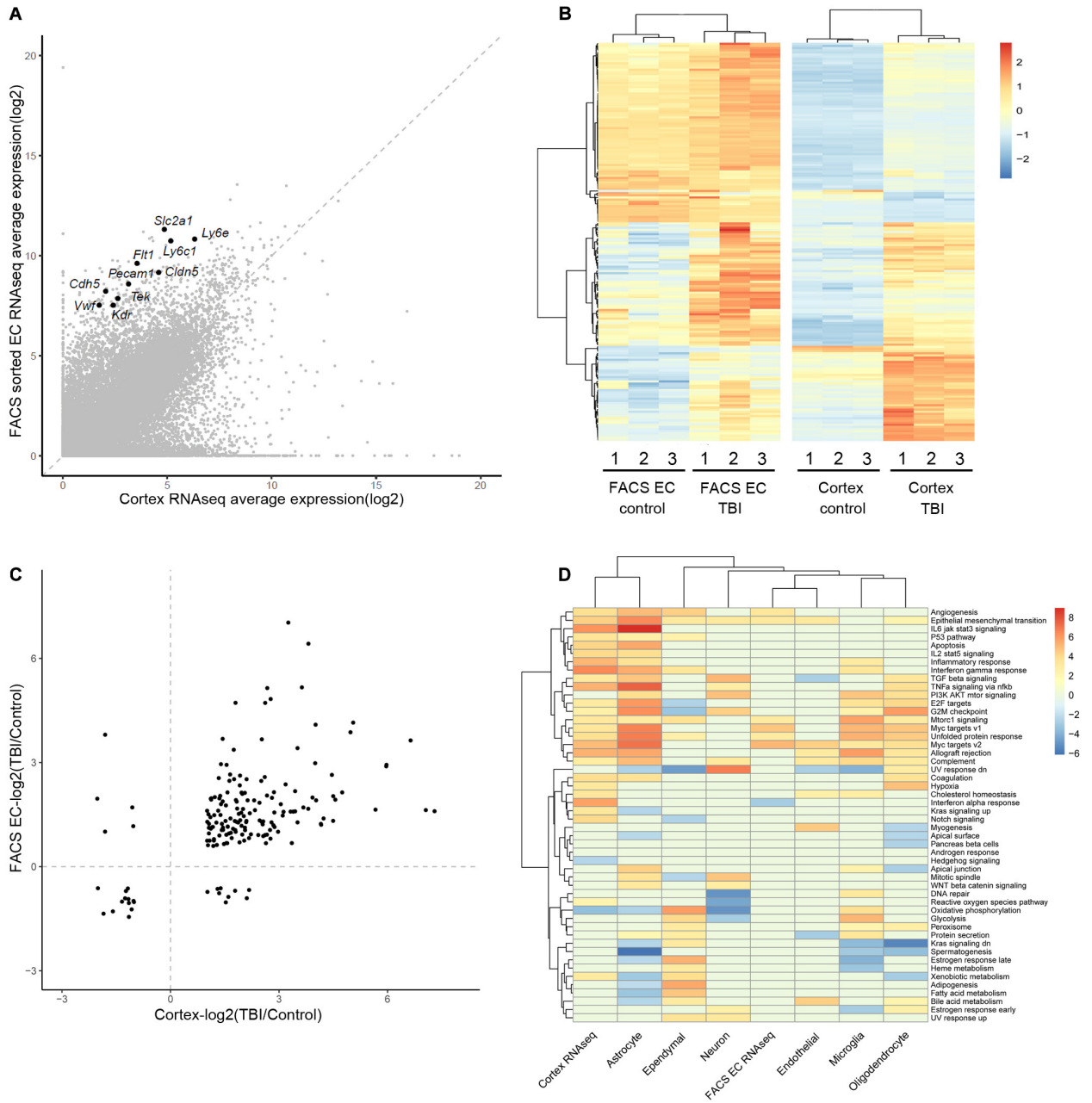


FIGURE 1 | Comparison of the gene expression analysis results from the three studies. **(A)** Comparison of the average expression of whole cortex and FACS sorted EC RNAseq data. Ten well-known canonical EC markers are labeled. **(B)** Heatmap of the different expression of 184 DEGs in the two studies, using the pheatmap packages (version 1.0.12) with ward.D2 method. The color legend shows the relative expression. **(C)** Comparison of the log2 scaled fold change (TBI/Control) of the 184 DEGs in the two RNAseq studies. **(D)** Heatmap of the GSVA analysis of the three studies. The two RNAseq studies and the major cell types in the scRNAseq study are listed in the columns. The rows represent the 50 hallmark gene sets. The color legend shows the GSVA t statistics, and red color indicates up-regulation and blue color indicates down-regulation.

stimulating cells with $CoCl_2$, which induces hypoxia stimulation through HIF1α stabilization (Xu et al., 2015). VEGFA stimulation failed to upregulate *Igfbp7/IGFBP7* expression in either bEND.3 or HUVEC. Exposure of HUVECs to $CoCl_2$ lead to a subtle upregulation of *IGFBP7* expression, while this effect could not be observed in bEND.3 cells (**Figures 4A,B**). Notably, *Igfbp7/IGFBP7* expression was significantly up-regulated in bEND.3 cells (7.2-folds) and HUVECs upon (1.6-folds) upon TGFβ stimulation (**Figures 4A,B**). In line with these findings, we observed a higher level of *TGFβ1* in cortex from TBI compared to control in whole cortex RNAseq study (Zhong et al., 2016; **Figure 4C**), suggesting TGFβ signaling may potentiate vascular

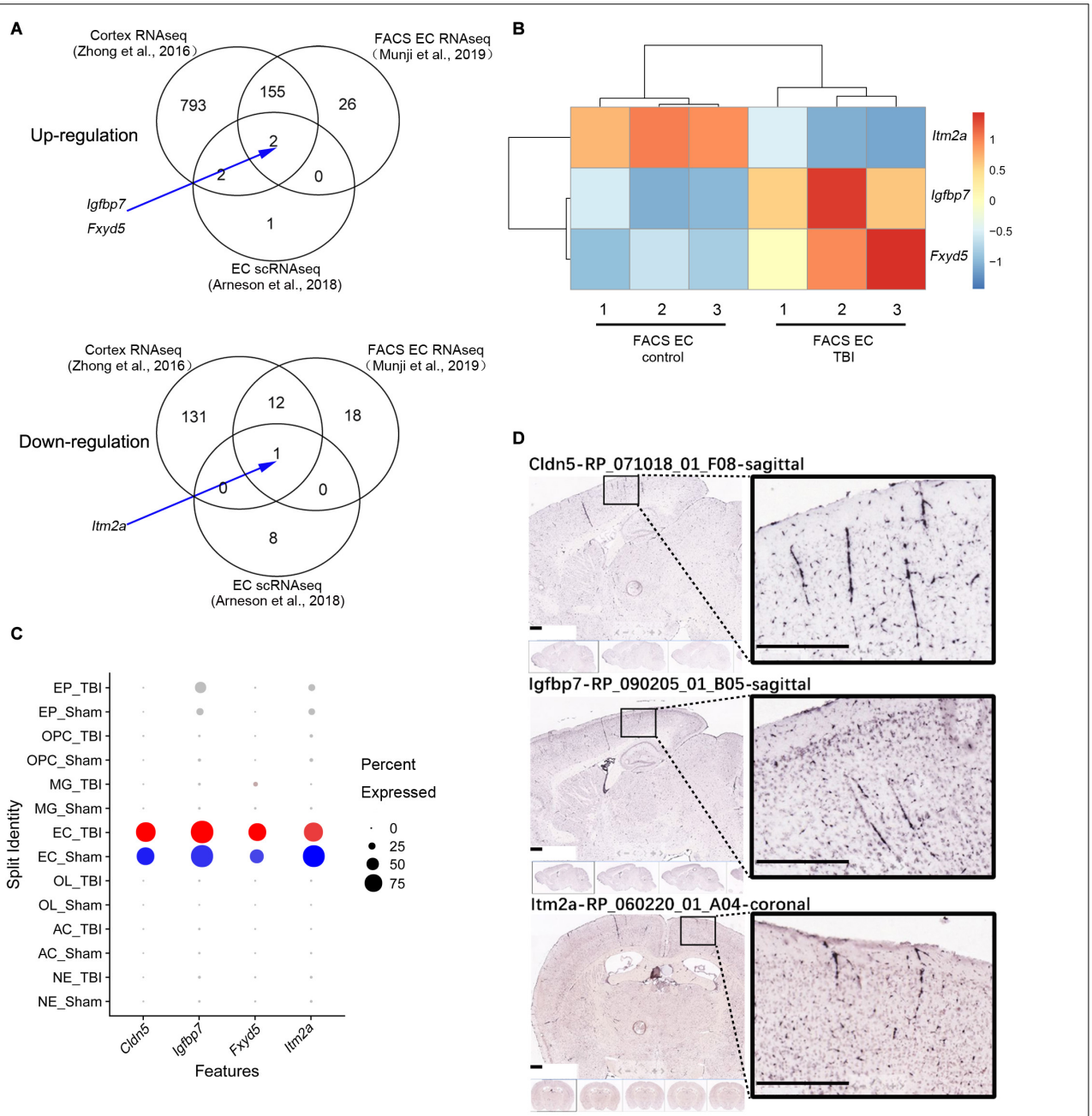


FIGURE 2 | Analysis of the three shared DEGs. **(A)** Venn diagram illustrating the overlaps of the DEGs identified in the three studies, and the arrow indicate the three shared DEGs. **(B)** The details of the gene expression regulation of *Itm2a*, *Igfbp7*, and *Fxyd5* in the TBI study from the FACS sorted EC samples. **(C)** The dot plot illustrating the expression level and the expression percentage of *Cldn5*, *Igfbp7*, *Fxyd5*, and *Itm2a* in the major cell types identified from the scRNAseq study. *Cldn5* is a commonly used EC maker and used as a reference. The size of the dot represents the expression percentage in the cell type and the color intensity represents the expression level. The well-known EC marker *Cldn5* is used as a reference. EP, Ependymocyte; OPC, Oligodendrocyte progenitor cell; MG, Microglia; EC, endothelial cell; OL, Oligodendrocyte; AC, Astrocyte; NE, Neuron. **(D)** The vascular expression of *Igfbp7* and *Itm2a* genes expression using the Allen Brain Atlas database. Scale bar, 400 μm.

IGFBP7 upregulation in TBI. Interestingly, there is no significant difference in *TGFβ2* level in cortex between TBI and control (**Figure 4D**). To identify the main source of *TGFβ1*, we have analyzed the *TGFβ1* expression in single cell RNAseq dataset. *TGFβ1* is mainly expressed in microglia and EC (**Figure 4E**). Based on these experiments *in vitro*, we proposed a potential mechanism that IGFBP7 up-regulation in endothelial cells in response to brain injury may be through TGFβ signaling. Experiments *in vivo* need to be further performed in the future to confirm the mechanism.

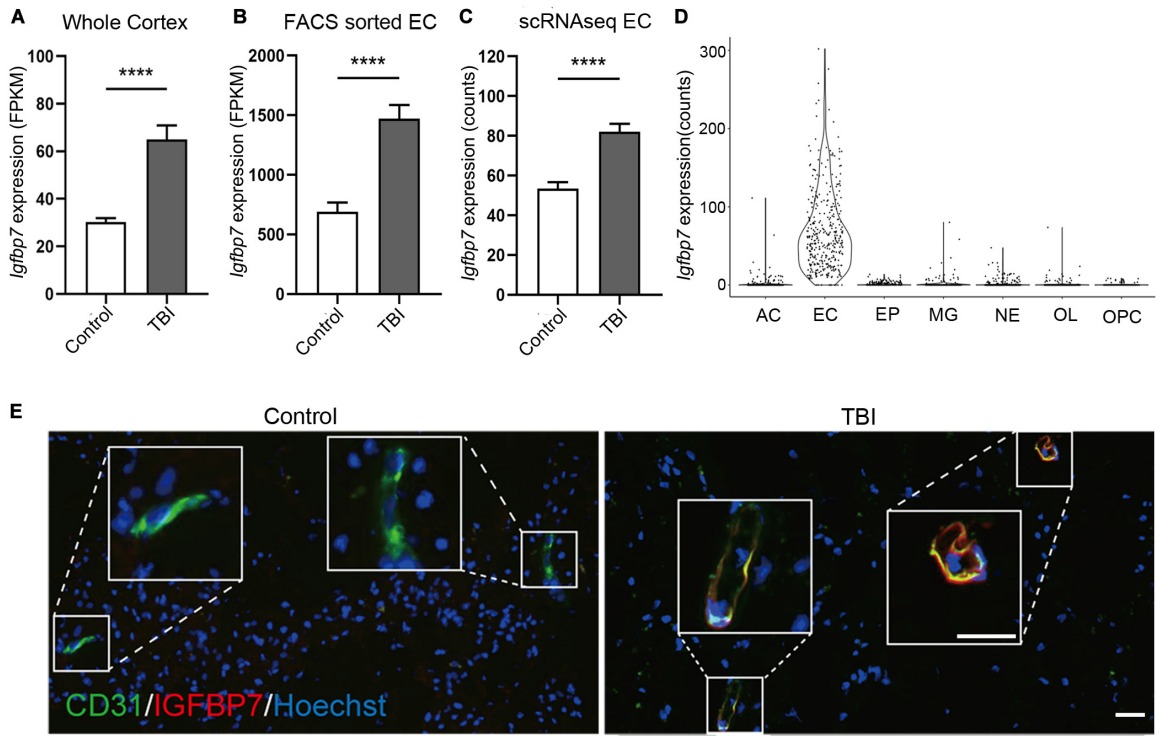


FIGURE 3 | *Igfbp7/IGFBP7* is upregulated in the vasculature in response TBI. **(A–C)** The expression level of *Igfbp7* in the TBI and control groups from the three data sets. From left to right, whole cortex, FACS sorted EC RNAseq, and scRNAseq data sets. **** indicates $p < 0.0001$. **(D)** The expression level of *Igfbp7* in the main cell types from the scRNAseq study. AC, Astrocyte; EC, endothelial cell; EP, Ependymocyte; MG, Microglia; NE, Neuron; OL, Oligodendrocyte; OPC, Oligodendrocyte progenitor cell. **(E)** Immunofluorescence co-staining of CD31 (green) and IGFBP7 (red) in human samples. The TBI and control group were both operated from the right frontal. Scale bar, 50 μm.

DISCUSSION

Brain vasculature plays critical roles in brain physiology and pathology. In our previous studies, we have illustrated transcriptome profiles in the normal vasculature (He et al., 2016, 2018; Vanlandewijck et al., 2018). In this study, we focus on the gene expression changes in the brain vascular EC under TBI condition. To cover common EC alteration in TBI pathogenesis, we integrated the results from two bulk RANseq studies (Zhong et al., 2016; Munji et al., 2019) and one single cell RNAseq study (Arneson et al., 2018). The controlled cortical impact (CCI) model was employed in Zhang's and Munji's studies to generate focal injury, while the fluid percussion injury (FPI) model was used in Arneson's study to induce mixed injury. All studies applied the mouse model and brain RNA expression analyses were performed at 24 h after injury. In the comparison of regulated genes from the three studies, despite the significant difference due to sample compositions and techniques, we identified three genes that were consistently regulated in all three analyses (*Igfbp7*, *Fxyd5*, and *Itm2a*).

IGFBP7 is highly expressed in the vasculature during development in the central nervous system. But, in adults, the IGFBP7 expression was reported to be only restricted to smooth muscle cells covering large vessels and choroid plexus vasculature, which is characterized by limited BBB property and high permeability (Hooper et al., 2009; Bar et al., 2020). Increased IGFBP7 expression in vasculature was observed in brain pathological conditions including glioblastoma and stroke, as well as other types of tumors (Hooper et al., 2009; Dieterich et al., 2012; Buga et al., 2014; Komiya et al., 2014). In this study, we showed that *Igfbp7/IGFBP7* is up-regulated in the vasculature in both TBI mice models and surgical samples from patients with brain injury. Our results together with previous findings support the notion that *Igfbp7/IGFBP7* may be a general marker of vasculature in response to pathological conditions in the brain. A higher level of TGFβ and hypoxia was observed in the brain tissue after TBI (Park et al., 2009), and they can sufficiently upregulate *Igfbp7/IGFBP7* expression in EC, which may explain the molecular mechanism for *Igfbp7/IGFBP7* up-regulation in vasculature in response to TBI.

The role of IGFBP7 on pathogenesis of brain injury remains largely unknown. Exposure of EC to IGFBP7 leads to stress fiber formation and disorganization of VE-cadherin mediated junctions, resulting in increased vascular permeability (Komiya et al., 2014), which indicates a role of IGFBP7 on BBB breakdown. In addition, stimulation of brain EC with IGFBP7 upregulated E-selectin, a crucial molecule in immune cell recruitment (Rai et al., 2015), indicating that IGFBP7 may regulate neuroinflammation in response to brain injury.

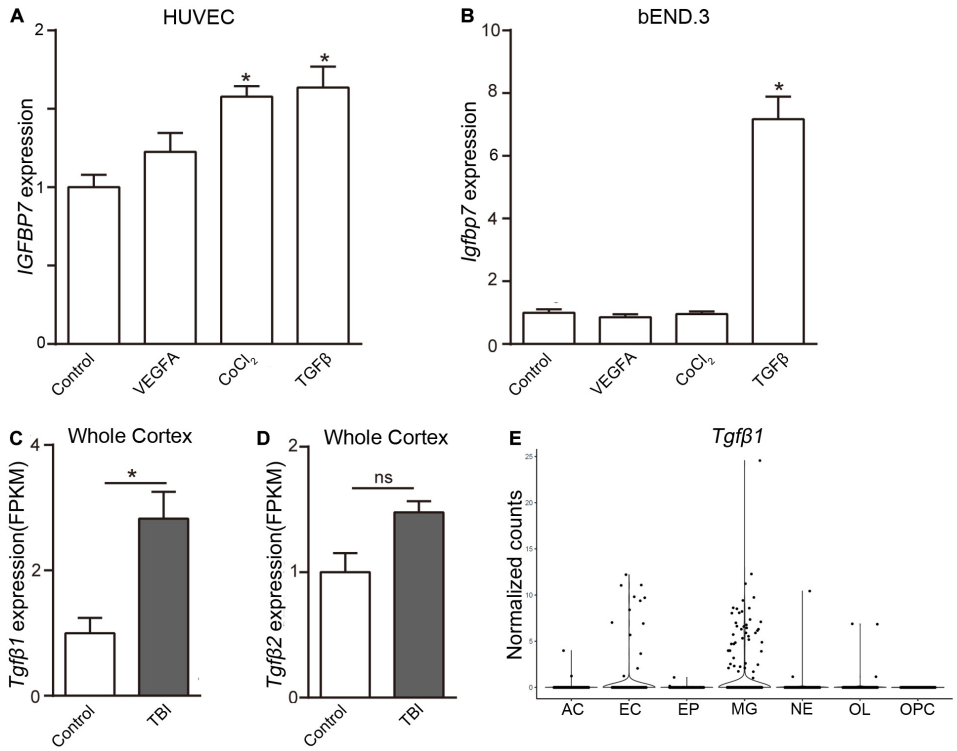


FIGURE 4 | TGFβ induces the expression of *Igfbp7/IGFBP7* in endothelial cells. The normalized *Igfbp7/IGFBP7* expression levels in Control, VEGFA, $CoCl_2$, and TGFβ stimulated HUVEC **(A)** and bEND.3 **(B)** cells are listed in the histogram, respectively. The expression levels of *TGFβ1* **(C)** and *TGFβ2* **(D)** from the whole cortex RNAseq data set. **(E)** The *TGFβ1* expression in the main cell types from the scRNAseq study. AC, Astrocyte; EC, endothelial cell; EP, Ependymocyte; MG, Microglia; NE, Neuron; OL, Oligodendrocyte; OPC, Oligodendrocyte progenitor cell. * indicates $p < 0.05$.

Emerging studies convincingly showed that the injury vasculature attempts to undergo repair by inducing angiogenesis (Park et al., 2009). IGFBP7 can act as both pro- and anti-angiogenic factors (Hooper et al., 2009; Pen et al., 2011; Tamura et al., 2009, 2014; Komiya et al., 2014). IGFBP7 could block VEGFA-induced tube formation, EC migration, proliferation, and vascular permeability (Tamura et al., 2009, 2014). In contrast, other studies suggested that IGFBP7 promotes angiogenesis by increasing EC adhesion and VEGFA bioavailability (Hooper et al., 2009; Komiya et al., 2014). Thus, the contradictory effect of IGFBP7 on angiogenesis in different systems may depend on cues in the microenvironment, such as distinct regional composition of ECM (Hooper et al., 2009). Whether IGFBP7 contributes to vascular repair by regulating angiogenesis after brain injury deserves further investigation.

Taken together, our study reveals the key molecular alteration of EC and identifies IGFBP7 as a potential biomarker of vasculature in response to brain injury.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the online repositories. 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 ethics committee, Tianjin Medical University General Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LH and LZ conceived the project. JW, YX, LZ, and ZZ performed the experiments. JW, XD, JT, FY, QH, QC, LZ, and LH analyzed the data. LZ and LH wrote the manuscript with significant input from JW and YX. All authors reviewed and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 81870978, 81702489, 81911530166, and 82002659), the National Key R&D Program of China (No. 2018YFC1313003), the Tianjin Natural Science Foundation (No. 18JCYBJC94000), the Natural Science Foundation of Shaanxi Province (No. 2020JQ-429), the Fundamental Research Funds for the Central University (Nos. GK202003050 and

GK202003048), and the Natural Science Foundation of Huaihua City (Nos. 2020R3118 and 2020R3116).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.599834/full#supplementary-material

Supplementary Figure 1 | Clustering and annotation of the 6,351 cells from the single cell RNAseq study. **(A)** A Uniform Manifold Approximation and Projection (UMAP) overview of the whole dataset with main cell type annotated. OPC, Oligodendrocyte progenitor cell; EC, endothelial cell. **(B)** The detailed expression levels of typical maker genes to illustrate the main cell types. The size of the dot represents the expression percentage in the cell type and the color represents the expression level. Cluster 5 displays unique expression of EC markers (*Pecam1*, *Cldn5*, and *Flt1*) and is selected as EC cluster for downstream analysis. Cluster 8 shows lower expression of EC markers, but with specific expression of mural cells markers (*Pdgfrb*, *Acta2*, and *Kcnj8*), suggesting that it is likely an EC-contaminated mural cell cluster.

REFERENCES

Arneson, D., Zhang, G., Ying, Z., Zhuang, Y., Byun, H. R., Ahn, I. S., et al. (2018). Single cell molecular alterations reveal target cells and pathways of concussive brain injury. *Nat. Commun.* 9:3894. doi: 10.1038/s41467-018-06222-0

Badaut, J., Ashwal, S., and Obenaus, A. (2011). Aquaporins in cerebrovascular disease: a target for treatment of brain edema? *Cerebrovasc. Dis.* 31, 521–531. doi: 10.1159/000324328

Bar, O., Gelb, S., Atamny, K., Anzi, S., and Ben-Zvi, A. (2020). Angiomodulin (IGFBP7) is a cerebral specific angiocrine factor, but is probably not a blood-brain barrier inducer. *Fluids Barriers CNS* 17:27. doi: 10.1186/s12987-020-00188-2

Bramlett, H. M., and Dietrich, W. D. (2015). Long-term consequences of traumatic brain injury: current status of potential mechanisms of injury and neurological outcomes. *J. Neurotrauma* 32, 1834–1848. doi: 10.1089/neu.2014.3352

Buga, A. M., Margaritescu, C., Scholz, C. J., Radu, E., Zelenak, C., and Popa-Wagner, A. (2014). Transcriptomics of post-stroke angiogenesis in the aged brain. *Front. Aging Neurosci.* 6:44. doi: 10.3389/fnagi.2014.00044

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. doi: 10.1038/nbt.4096

Dieterich, L. C., Mellberg, S., Langenkamp, E., Zhang, L., Zieba, A., Salomaki, H., et al. (2012). Transcriptional profiling of human glioblastoma vessels indicates a key role of VEGF-A and TGFbeta2 in vascular abnormalization. *J. Pathol.* 228, 378–390. doi: 10.1002/path.4072

He, L., Vanlandewijck, M., Mae, M. A., Andrae, J., Ando, K., Del Gaudio, F., et al. (2018). Single-cell RNA sequencing of mouse brain and lung vascular and vessel-associated cell types. *Sci. Data* 5:180160. doi: 10.1038/sdata.2018.160

He, L., Vanlandewijck, M., Raschperger, E., Andaloussi Mae, M., Jung, B., Lebouvier, T., et al. (2016). Analysis of the brain mural cell transcriptome. *Sci. Rep.* 6:35108. doi: 10.1038/srep35108

Ho, K. M., Honeybul, S., Yip, C. B., and Silbert, B. I. (2014). Prognostic significance of blood-brain barrier disruption in patients with severe nonpenetrating traumatic brain injury requiring decompressive craniectomy. *J. Neurosurg.* 121, 674–679. doi: 10.3171/2014.6.JNS132838

Hooper, A. T., Shmelkov, S. V., Gupta, S., Milde, T., Bambino, K., Gillen, K., et al. (2009). Angiomodulin is a specific marker of vasculature and regulates vascular endothelial growth factor-A-dependent neoangiogenesis. *Circ. Res.* 105, 201–208. doi: 10.1161/CIRCRESAHA.109.196790

Jassam, Y. N., Izzy, S., Whalen, M., McGavern, D. B., and El Khoury, J. (2017). Neuroimmunology of traumatic brain injury: time for a paradigm shift. *Neuron* 95, 1246–1265. doi: 10.1016/j.neuron.2017.07.010

Jullienne, A., Obenaus, A., Ichkova, A., Savona-Baron, C., Pearce, W. J., and Badaut, J. (2016). Chronic cerebrovascular dysfunction after traumatic brain injury. *J. Neurosci. Res.* 94, 609–622. doi: 10.1002/jnr.23732

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

Komiya, E., Sato, H., Watanabe, N., Ise, M., Higashi, S., Miyagi, Y., et al. (2014). Angiomodulin, a marker of cancer vasculature, is upregulated by vascular endothelial growth factor and increases vascular permeability as a ligand of integrin alphavbeta3. *Cancer Med.* 3, 537–549. doi: 10.1002/cam4.216

Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* 24, 1277–1289. doi: 10.1038/s41591-018-0096-5

Maas, A. I., Stocchetti, N., and Bullock, R. (2008). Moderate and severe traumatic brain injury in adults. *Lancet Neurol.* 7, 728–741. doi: 10.1016/S1474-4422(08)70164-9

Marmarou, A. (2003). Pathophysiology of traumatic brain edema: current concepts. *Acta Neurochir. Suppl.* 86, 7–10. doi: 10.1007/978-3-7091-0651-8_2

Munji, R. N., Soung, A. L., Weiner, G. A., Sohet, F., Semple, B. D., Trivedi, A., et al. (2019). Profiling the mouse brain endothelial transcriptome in health and disease models reveals a core blood-brain barrier dysfunction module. *Nat. Neurosci.* 22, 1892–1902. doi: 10.1038/s41593-019-0497-x

Nguyen, R., Fiest, K. M., McChesney, J., Kwon, C. S., Jette, N., Frolkis, A. D., et al. (2016). The international incidence of traumatic brain injury: a systematic review and meta-analysis. *Can. J. Neurol. Sci.* 43, 774–785. doi: 10.1017/cjn.2016.290

Park, E., Bell, J. D., Siddiq, I. P., and Baker, A. J. (2009). An analysis of regional microvascular loss and recovery following two grades of fluid percussion trauma: a role for hypoxia-inducible factors in traumatic brain injury. *J. Cereb. Blood Flow Metab.* 29, 575–584. doi: 10.1038/jcbfm.2008.151

Pen, A., Durocher, Y., Slinn, J., Rukhlova, M., Charlebois, C., Stanimirovic, D. B., et al. (2011). Insulin-like growth factor binding protein 7 exhibits tumor suppressive and vessel stabilization properties in U87MG and T98G glioblastoma cell lines. *Cancer Biol. Ther.* 12, 634–646. doi: 10.4161/cbt.12.7.17171

Pen, A., Moreno, M. J., Durocher, Y., Deb-Rinker, P., and Stanimirovic, D. B. (2008). Glioblastoma-secreted factors induce IGFBP7 and angiogenesis by modulating Smad-2-dependent TGF-beta signaling. *Oncogene* 27, 6834–6844. doi: 10.1038/onc.2008.287

Rai, S., Nejadhamzeeigilani, Z., Gutowski, N. J., and Whatmore, J. L. (2015). Loss of the endothelial glycocalyx is associated with increased E-selectin mediated adhesion of lung tumour cells to the brain microvascular endothelium. *J. Exp. Clin. Cancer Res.* 34:105. doi: 10.1186/s13046-015-0223-9

Salehi, A., Zhang, J. H., and Obenaus, A. (2017). Response of the cerebral vasculature following traumatic brain injury. *J. Cereb. Blood Flow Metab.* 37, 2320–2339. doi: 10.1177/0271678X17701460

Silverberg, N. D., Duhaime, A. C., and Iaccarino, M. A. (2019). Mild traumatic brain injury in 2019-2020. *JAMA* 323, 177–178. doi: 10.1001/jama.2019.18134

Simon, D. W., McGeachy, M. J., Bayir, H., Clark, R. S., Loane, D. J., and Kochanek, P. M. (2017). The far-reaching scope of neuroinflammation after traumatic brain injury. *Nat. Rev. Neurol.* 13, 171–191. doi: 10.1038/nrneurol.2017.13

Sun, D., Gu, G., Wang, J., Chai, Y., Fan, Y., Yang, M., et al. (2017). Administration of tauroursodeoxycholic acid attenuates early brain injury via Akt pathway activation. *Front. Cell. Neurosci.* 11:193. doi: 10.3389/fncel.2017.00193

Tamura, K., Hashimoto, K., Suzuki, K., Yoshie, M., Kutsukake, M., and Sakurai, T. (2009). Insulin-like growth factor binding protein-7 (IGFBP7) blocks vascular endothelial cell growth factor (VEGF)-induced angiogenesis in human vascular endothelial cells. *Eur. J. Pharmacol.* 610, 61–67. doi: 10.1016/j.ejphar.2009.01.045

Tamura, K., Yoshie, M., Hashimoto, K., and Tachikawa, E. (2014). Inhibitory effect of insulin-like growth factor-binding protein-7 (IGFBP7) on in vitro angiogenesis of vascular endothelial cells in the rat corpus luteum. *J. Reprod. Dev.* 60, 447–453. doi: 10.1262/jrd.2014-069

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments

with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Vanlandewijck, M., He, L., Mae, M. A., Andrae, J., Ando, K., Del Gaudio, F., et al. (2018). A molecular atlas of cell types and zonation in the brain vasculature. *Nature* 554, 475–480. doi: 10.1038/nature25739

Xu, X., Tan, X., Tampe, B., Sanchez, E., Zeisberg, M., and Zeisberg, E. M. (2015). Snail is a direct target of hypoxia-inducible factor 1alpha (HIF1alpha) in hypoxia-induced endothelial to mesenchymal transition of human coronary endothelial cells. *J. Biol. Chem.* 290, 16653–16664. doi: 10.1074/jbc.M115.636944

Yeoh, S., Bell, E. D., and Monson, K. L. (2013). Distribution of blood-brain barrier disruption in primary blast injury. *Ann. Biomed. Eng.* 41, 2206–2214. doi: 10.1007/s10439-013-0805-7

Zhong, J., Jiang, L., Cheng, C., Huang, Z., Zhang, H., Liu, H., et al. (2016). Altered expression of long non-coding RNA and mRNA in mouse cortex after traumatic brain injury. *Brain Res.* 1646, 589–600. doi: 10.1016/j.brainres.2016.07.002

Ziebell, J. M., and Morganti-Kossmann, M. C. (2010). Involvement of pro- and anti-inflammatory cytokines and chemokines in the pathophysiology of traumatic brain injury. *Neurotherapeutics* 7, 22–30. doi: 10.1016/j.nurt.2009.10.016

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 in Genetics

ORIGINAL RESEARCH
published: 14 January 2021
doi: 10.3389/fgene.2020.602922

Identification of Potential Diagnostic and Prognostic Biomarkers for Colorectal Cancer Based on GEO and TCGA Databases

Zhenjiang Wang[1†], *Mingyi Guo*[1†], *Xinbo Ai*[1], *Jianbin Cheng*[1], *Zaiwei Huang*[1], *Xiaobin Li*[2*] *and Yuping Chen*[1*]

[1] Department of Gastroenterology, Zhuhai People's Hospital (Zhuhai Hospital Affiliated With Jinan University), Zhuhai, China, [2] Zhuhai Precision Medical Center, Zhuhai People's Hospital (Zhuhai Hospital Affiliated With Jinan University), Zhuhai, China

OPEN ACCESS

Edited by:
Zhichao Liu,
National Center for Toxicological Research (FDA), United States

Reviewed by:
Dan Li,
University of Arkansas at Little Rock, United States
Dongying Li,
United States Food and Drug Administration, United States

***Correspondence:**
Yuping Chen
chenypmd@163.com
Xiaobin Li
li.xiaobin2009@163.com

[†] These authors have contributed equally to this work

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

Received: 04 September 2020
Accepted: 30 November 2020
Published: 14 January 2021

Citation:
Wang Z, Guo M, Ai X, Cheng J, Huang Z, Li X and Chen Y (2021) Identification of Potential Diagnostic and Prognostic Biomarkers for Colorectal Cancer Based on GEO and TCGA Databases. Front. Genet. 11:602922. doi: 10.3389/fgene.2020.602922

Colorectal cancer (CRC) is one of the most common neoplastic diseases worldwide. With a high recurrence rate among all cancers, treatment of CRC only improved a little over the last two decades. The mortality and morbidity rates can be significantly lessened by earlier diagnosis and prompt treatment. Available biomarkers are not sensitive enough for the diagnosis of CRC, whereas the standard diagnostic method, endoscopy, is an invasive test and expensive. Hence, seeking the diagnostic and prognostic biomarkers of CRC is urgent and challenging. With that order, we screened the overlapped differentially expressed genes (DEGs) of GEO (GSE110223, GSE110224, GSE113513) and TCGA datasets. Subsequent protein–protein interaction network analysis recognized the hub genes among these DEGs. Further functional analyses including Gene Ontology and KEGG pathway analysis and gene set enrichment analysis were processed to investigate the role of these genes and potential underlying mechanisms in CRC. Kaplan–Meier analysis and Cox hazard ratio analysis were carried out to clarify the diagnostic and prognostic role of these genes. In conclusion, our present study demonstrated that CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 are all potential diagnostic biomarkers for CRC and may also be potential treatment targets for clinical implication in the future.

Keywords: colorectal cancer, diagnostic biomarker, prognostic biomarker, GEO, TCGA

INTRODUCTION

Global Cancer Statistics 2018 indicates that colorectal cancer (CRC) accounts for ∼10% of all diagnosed cancers and cancer-related deaths in the world each year (Bray et al., 2018). According to the data from China Cancer Registry Annual Report, the incidence and mortality of CRCs have been increasing in the past 10 years (Zheng et al., 2019). With the improvement of surgical methods and the launch of early tumor diagnosis and treatment, the current levels of diagnosis and treatment of CRC have been greatly improved. However, the prognosis of clinical CRC is still not optimistic. Many researches have shown that the occurrence and development of CRCs may be related to genetic, lifestyle, obesity, and environmental factors, while the exact etiology and the mechanism are still unclear (Bray et al., 2018). To further clarify the pathogenesis of CRCs and to improve the precision of treatment of CRCs, genetic research, study of tumor signaling pathways, and biological target therapy are continuing to deepen, which are gradually being applied in clinic. Meanwhile,

molecular stratification therapy and application of biomarkers to guide prognosis and treatment decisions are also increasing (Bogaert and Prenen, 2014).

As we all know, the occurrence, development, overall survival time, and recurrence and non-recurrence of tumors are not only related to the pathological type and clinical stage of the tumor but also closely related to the expression and pathway of tumor genes (Bogaert and Prenen, 2014). More and more studies suggested that there are many abnormally expressed genes in the gene expression of CRCs, relative to normal tissues, which are closely related to the proliferation, differentiation, apoptosis, metastasis, recurrence, and survival time of CRC (Lu et al., 2012; Liu et al., 2013; Gan et al., 2018; Branchi et al., 2019). The analysis of abnormally expressed genes has very important clinical significance for the targeted therapy, prognosis analysis, and recurrence risk prediction of CRC. Currently, there have been a lot of clinical researches on tumor recurrence genes and signaling pathways, and the gene recurrent model (GRM) has been established to make up for the traditional tumor classification and staging recurrence prediction, providing more genetic information and more accurate prediction data (Chen et al., 2019; Yang et al., 2020). For example, Sun et al. (2019) found that exosomal CPNE3 showed potential implications in CRC diagnosis and prognosis. Carcinoembryonic antigen (CEA) was a recommended prognostic marker in CRC for tumor diagnosis and monitoring response to therapy (Campos-da-Paz et al., 2018). Ahluwalia et al. (2019) identified a novel 4-gene prognostic signature that had clinical utility in colorectal cancer. However, there are few clinical studies about biomarkers and gene pathways which have no risk of recurrence and tumor survival time.

In this study, we obtained advanced colorectal cancer gene profiles (GSE110223, GSE110224, and GSE113513) from the Gene Expression Omnibus. Differentially expressed genes (DEG) were identified by comparing CRC tissues with non-cancerous gastric tissues using the GEO2R online analysis. Subsequently, the DEG were analyzed according to Gene Ontology (GO), KEGG pathway enrichment analysis, coexpression, and protein–protein interaction (PPI) analyses. We then performed the overall survival analysis for candidate genes. Finally, GEPIA and UALCAN online tools were performed to associate candidate genes with colorectal cancer overall survival (OS), disease-free survival (RFS), and pathological staging analysis through the Cancer Genome Atlas [TCGA]) dataset and found that CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 could play an important role in colorectal cancer overall survival and disease-free survival, and these may be potential treatment targets for clinical implication in the future.

MATERIALS AND METHODS

Microarray Data

Studies from the GEO database were considered eligible satisfying the following criteria: (1) studies with CRC tissue samples, (2) studies containing information on technology and platform utilized for studies, and (3) studies including adjacent normal tissues as the control. Based on the above criteria, three datasets [GSE110223 (Vlachavas et al., 2019), GSE110224 (Vlachavas et al., 2019), GSE113513 (2018, unpublished)] are all downloaded from the GEO (Campos-da-Paz et al., 2018) database. The platform used by GSE110223 is [HG-U133A] Affymetrix Human Genome U133A Array, which includes 13 CRC tumor tissue samples and 13 normal tissue samples. The platform used by GSE110224 is [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, which includes 17 CRC tumor tissue samples and 17 normal tissue samples. The platform used by GSE113513 is [PrimeView] Affymetrix Human Gene Expression Array, which includes 14 CRC tumor tissue samples and 14 normal tissue samples. A total of 44 CRC tumor tissue samples and 44 normal tissue samples were included in this study (**Figure 1**).

Identification of DEGs

Using the GEOquery and limma R software packages in the Bioconductor project, the GEO2R web tool was used for identifying differential genes in selected datasets. We analyzed three GEO datasets online through GEO2R and selected the genes with $P < 0.05$ and fold change (FC) > 1.5 in the dataset as differentially expressed genes (DEGs). Then, we used FunRich_3.1.3 (Pathan et al., 2017) to make a Venn diagram and extract the differentially expressed genes common to the three datasets.

GO and KEGG Pathway Analysis

In this study, the GO term (www.geneontology.org) and Kyoto Encyclopedia of Genes and Genomes (KEGG, www.genome.jp) approach were identified and analyzed by using DAVID v6.8 (https://david-d.ncifcrf.gov/summary.jsp) (Huang da et al., 2009a,b). The identifier and species were selected as "official_gene_symbol" and "Homo Sapiens," respectively. The enrichment of $P < 0.05$ was set as the critical standard for significant enrichment and by using the ggplot2 package (version 3.3.1) (Wickham, 2009) and the R language (version 3.6.3, http://www.r-project.org/) to visualize the analysis results of the DAVID.

Protein–Protein Interaction Network Analysis

STRING (version 11.0), covering 24,584,628 proteins from 5,090 organisms, and integrating known and predicted interactions between more than 932,000,000 proteins from multiple organisms including Homo sapiens (Szklarczyk et al., 2019), was used to conduct PPI network analysis on DEGs. The "Multiple proteins" button was selected, and the species were selected as "Homo Sapiens." When the P-value was <0.05, the network interaction relationship is considered to be statistically significant, with interaction score >0.7 considered to be a high-confidence interaction relationship.

The Cytoscape software (version 3.6.0) (Shannon et al., 2003) was used to visualize the PPI network, and the plugin CytoHubba (Chin et al., 2014) was applied to identify the central node genes in the PPI network, and then the central node gene as a candidate DEG for the following analysis.

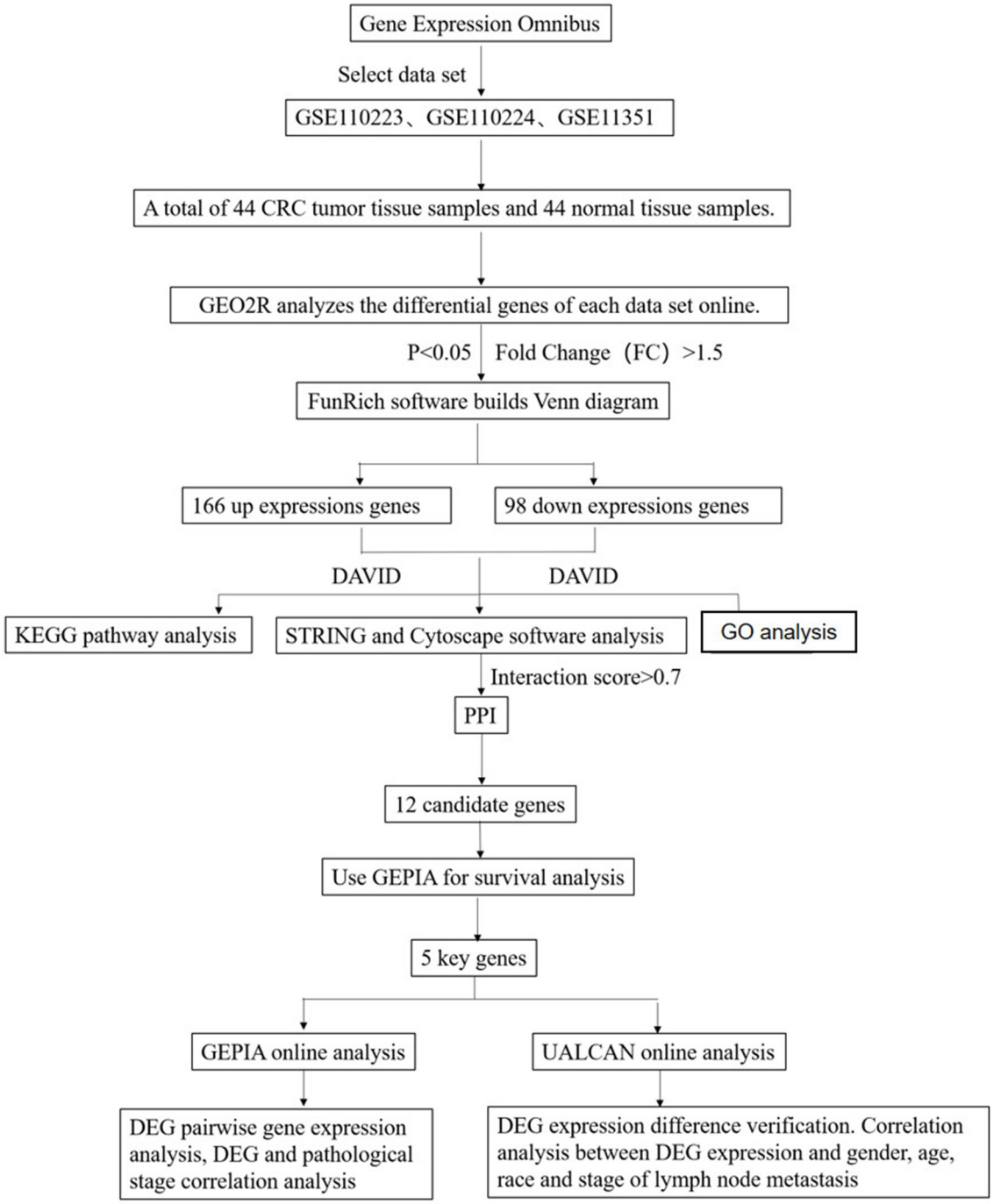


FIGURE 1 | Flowchart diagram for bioinformatics analysis of publicly available datasets from both GEO and TCGA databases.

The Expressions and Survival Analysis of Candidate DEGs

GEPIA (http://gepia.cancer-pku.cn/) is a database that uses standard processing methods to analyze the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA and GTEx projects (Tang et al., 2017). GEPIA provides various functions such as tumor/normal differential expression analysis, analysis according to cancer type or pathological stage, survival analysis, correlation analysis, etc. By using the GEPIA, we performed Kaplan–Meier survival analysis on the relative expression of candidate DEGs in CRC patients and the overall survival time and disease-free survival time, with hazard ratio

(HR) and corresponding 95% confidence interval. DEG related to the overall survival time and disease-free survival time was used as the study purpose DEG and conduct data analysis. Multiple-gene comparison and principal component analysis for the biomarker candidates were also conducted using the GEPIA.

UALCAN (http://ualcan.path.uab.edu/analysis.html) is an online database that uses TCGA transcriptome and clinical patient data to enable researchers to analyze the differential expression of tumor tissue and normal tissue, tumor stage, lymph node metastasis, and other related clinical parameters (Chandrashekar et al., 2017). We validated DEGs by using the UALCAN database, reanalyzed their expression differences in CRC tissue samples and normal tissue samples, and performed correlation analysis between DEG and gender, age, race, and stage of lymph node metastasis.

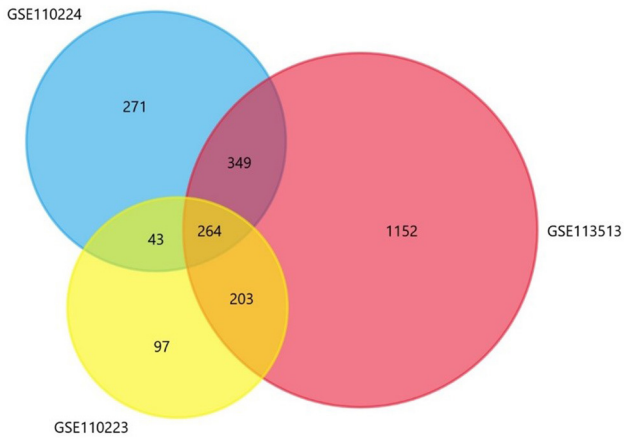


FIGURE 2 | The Venn diagram shows a total of 264 co-expressed differential genes.

RESULTS

Identification of DEGs

This study included three gene sets (GSE110223, GSE110224, GSE113513), of which GSE110223 included 13 tumor samples and 13 normal samples; GSE110224 included 17 tumor samples and 17 normal samples; and GSE113513 included 14 tumor samples and 14 normal samples. In all included datasets, compared with normal samples, there are 264 significantly different genes in all datasets (**Figure 2**), including 166 up-expression genes and 98 down-expression genes.

Enrichment Analysis of DEGs

The obtained DEGs were analyzed for functional enrichment by using DAVID. GO enrichment analysis mainly predicts the function of target genes through three aspects: biological process (BP), cell composition (CC), and molecular function (MF). By using DAVID, we found that DEG is mainly enriched in BP, including response to steroid hormone stimulus, response to nutrient, response to nutrient levels, response to hormone stimulus, response to endogenous stimulus,

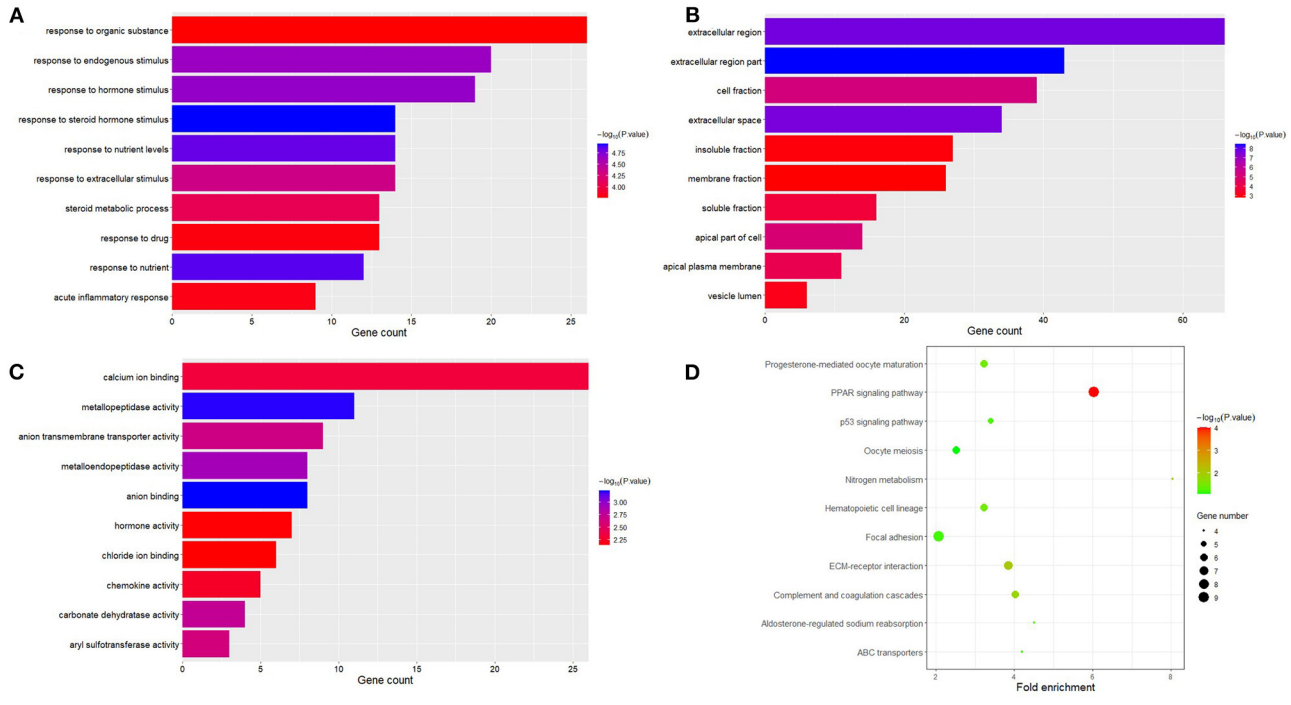


FIGURE 3 | Functional analysis of differential expression genes. Biological process (BP), cell composition (CC), and molecular function (MF) are the components of GO enrichment analysis results, and each part displays 10 GO terms **(A–C)**. The result shown by KEGG is a pathway of enrichment of DGEs **(D)**.

response to extracellular stimulus, steroid metabolic process, acute inflammatory response, response to drug, response to organic substance, etc. (**Figure 3A**). CC analysis includes extracellular region part, extracellular region, extracellular space, cell fraction, apical part of the cell, apical plasma membrane, soluble fraction, vesicle lumen, insoluble fraction, and membrane fraction (**Figure 3B**). MF analysis includes anion binding, metallopeptidase activity, metalloendopeptidase

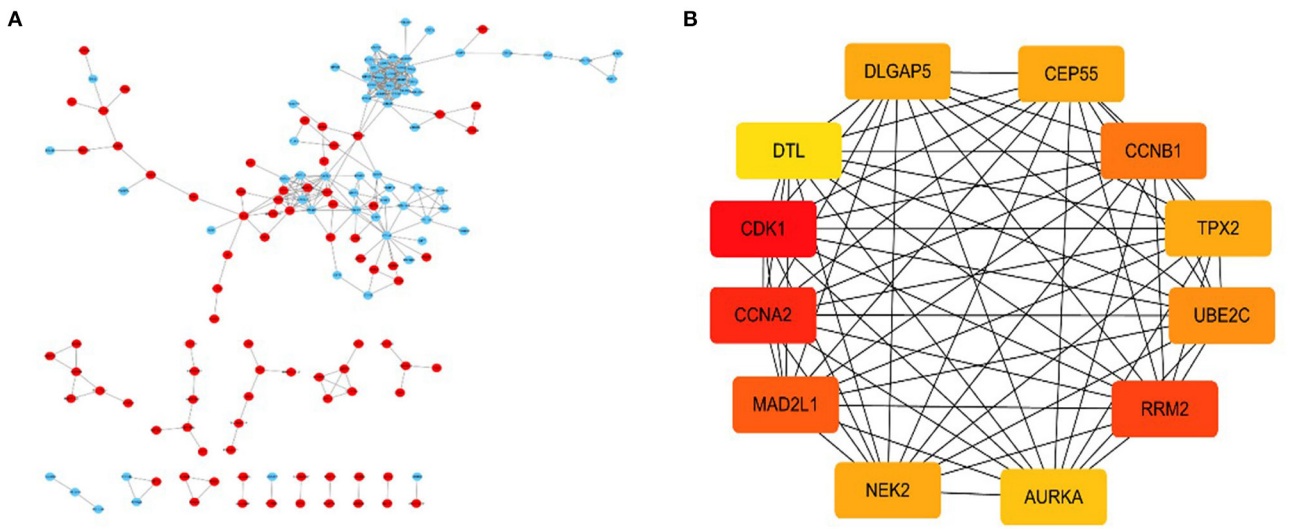


FIGURE 4 | A PPI network composed of 264 DGEs, in which red is an up-expression gene, and blue is a down-expression gene **(A)**. For the first 12 central genes calculated by Cytoscape software, the red represents the degree of connectivity. The deeper the red, the higher the degree of connectivity **(B)**.

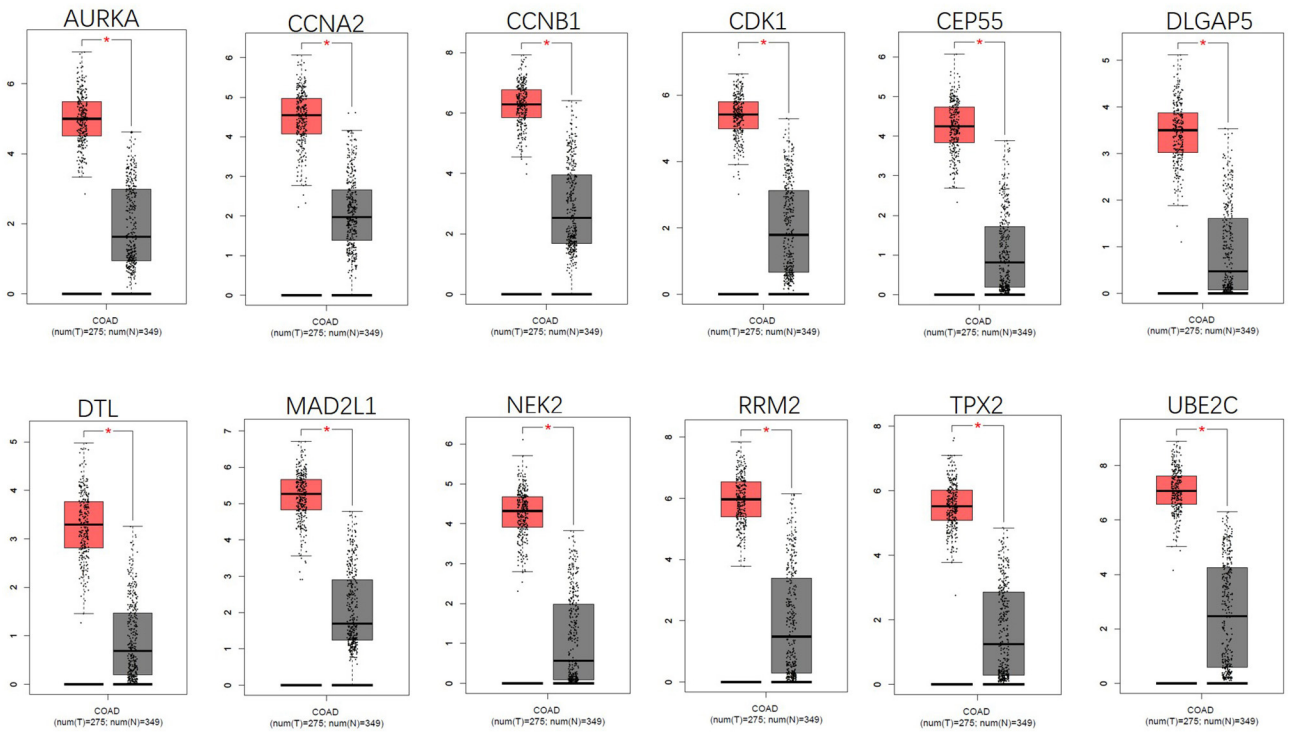


FIGURE 5 | Gene expression of 12 central genes (CDK1, CCNA2, RRM2, MAD2L1, CCNB1, UBE2C, CEP55, DLGAP5, NEK2, TPX2, AURKA, and DTL) based on GEPIA.

activity, carbonate dehydratase activity, anion transmembrane transporter activity, aryl sulfotransferase activity, calcium ion binding, chemokine activity, hormone activity, and chloride ion binding (**Figure 3C**).

At the same time, the analysis of the KEGG pathway shows that 264 DEGs are mainly enriched in eight pathways, namely, PPAR signaling pathway, ECM–receptor interaction, nitrogen metabolism, complement and coagulation cascades, hematopoietic cell lineage, progesterone-mediated oocyte maturation, aldosterone-regulated sodium reabsorption, p53 signaling pathway, focal adhesion, ABC transporters, and oocyte meiosis (**Figure 3D**).

PPI Network to Identify Central Genes

By using the STRING database and Cytoscape 3.6.0 software, a PPI network was constructed for the 264 DGEs obtained and the central genes were determined. The PPI network constructed by STRING (**Figure 4A**) has a total of 262 nodes and 802 edges, and an interaction score >0.7 is considered a high-confidence interaction relationship. Using Cytoscape 3.6.0 software, we identified the top 12 genes with the most connectedness (**Figure 4B**). The most connected gene is CDK1, followed by CCNA2, RRM2, MAD2L1, CCNB1, UBE2C, CEP55, DLGAP5, NEK2, TPX2, AURKA, and DTL. These 12 genes can form a module. Using GEPIA, gene expression profiles of the 12 central genes between CRC tumor samples and normal samples were displayed in **Figure 5**.

Overall Survival Analysis and Disease-Free Survival Analysis

Since CRC is mainly adenocarcinoma, which can account for 90%, we used the GEPIA database to analyze the overall survival and disease-free survival of colorectal adenocarcinoma on 12 central genes. We found that among the 12 central genes, CCNA2, MAD2L1, DLGAP5, and AURKA were associated with the overall survival of colorectal adenocarcinoma ($P < 0.05$) (**Figure 6**), and RRM2 and AURKA were associated with disease-free survival of colorectal adenocarcinoma ($P < 0.05$) (**Figure 7**). Therefore, this study will focus on the five genes CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2.

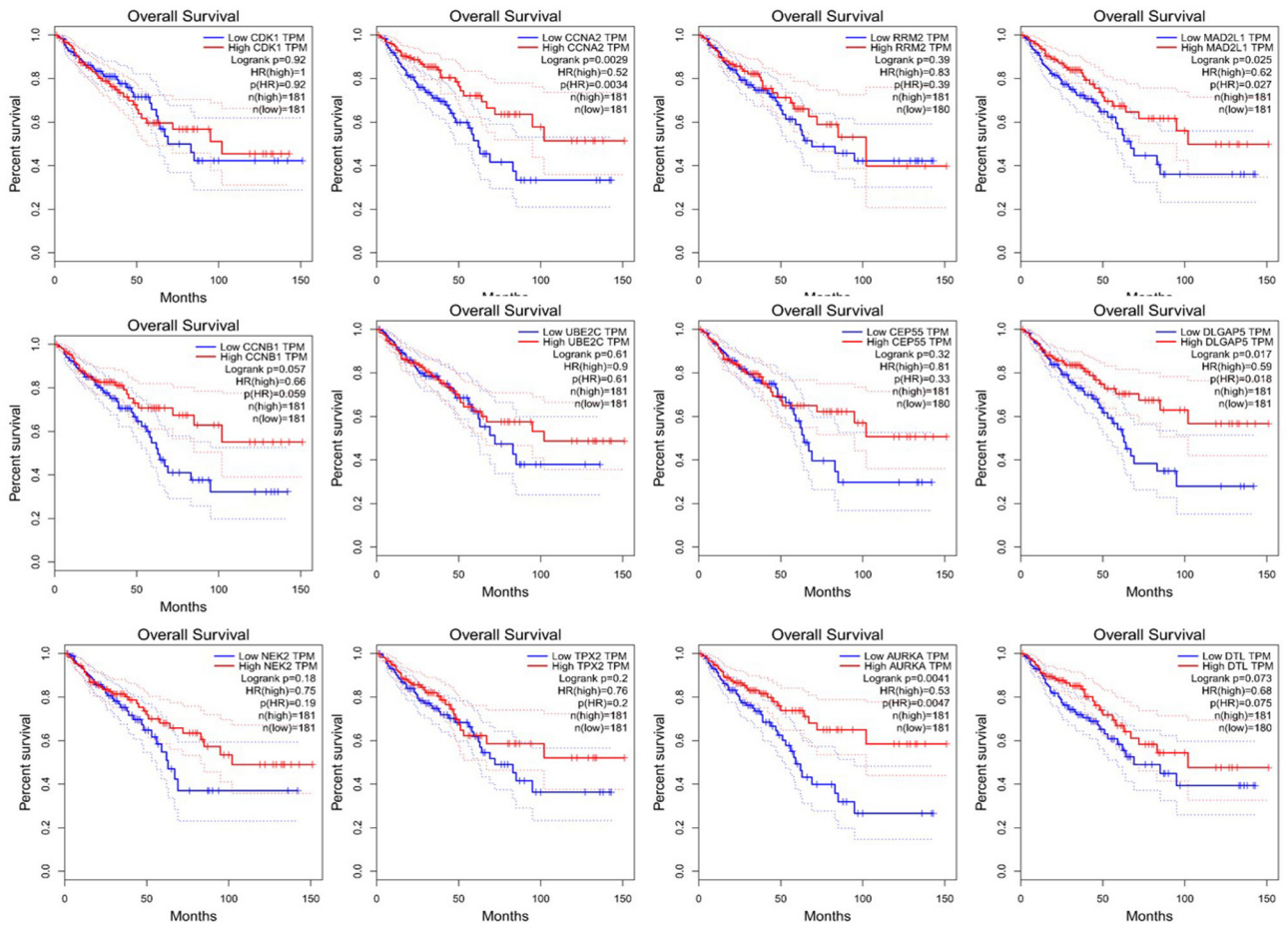


FIGURE 6 | Overall survival analysis of 12 central genes (CDK1, CCNA2, RRM2, MAD2L1, CCNB1, UBE2C, CEP55, DLGAP5, NEK2, TPX2, AURKA, and DTL) based on GEPIA.

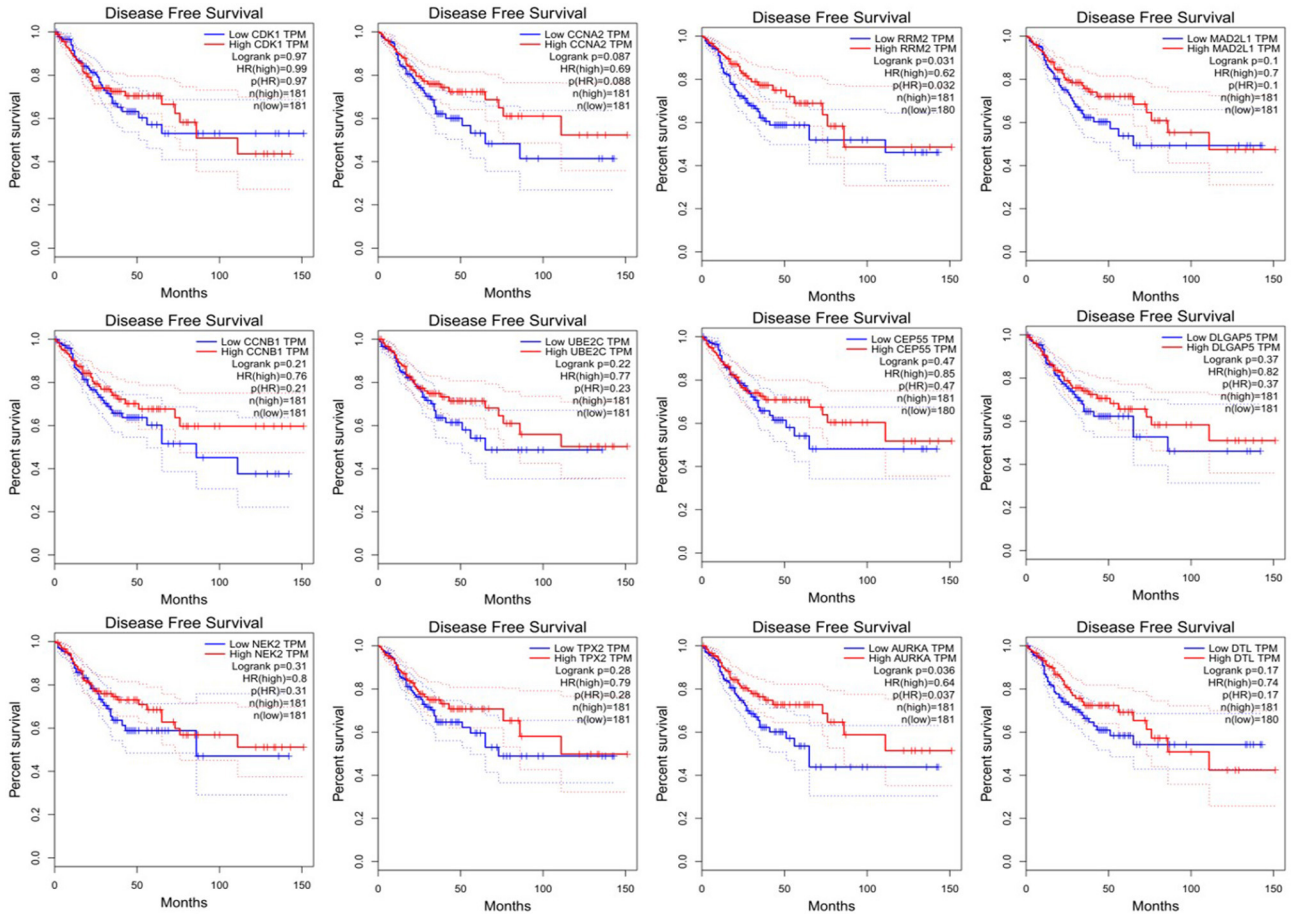


FIGURE 7 | Disease-free survival analysis 12 central genes (CDK1, CCNA2, RRM2, MAD2L1, CCNB1, UBE2C, DLGAP55, DLGAP5, NEK2, TPX2, AURKA, and DTL).

Correlation Analysis Based on the GEPIA

Through the analysis of the correlation between these five genes and the pathological staging of colorectal adenocarcinoma based on GEPIA, we found that CCNA2, MAD2L1, DLGAP5, and RRM2 are all significantly related to the pathological stage of COAD and READ ($P < 0.05$), while AURKA is associated with the colorectal gland, with no significant correlation between cancer pathological staging ($P > 0.05$) (**Figure 8A**). The correlation analysis of the expression of these five genes in colorectal adenocarcinoma showed that CCNA2 was highly correlated with MAD2L1 ($P < 0.001$, $R = 0.88$), and it was also correlated with DLGAP5 ($P < 0.001$, $R = 0.78$), AURKA ($P < 0.001$, $R = 0.53$), and RRM2 ($P < 0.001$, $R = 0.68$). Moderate positive correlations between MAD2L1 and DLGAP5 ($P < 0.001$, $R = 0.62$), MAD2L1 and AURKA ($P < 0.001$, $R = 0.54$), and MAD2L1 and RRM2 ($P < 0.001$, $R = 0.57$) were observed. DLGAP5 and AURKA ($P < 0.001$, $R = 0.42$) had low expression correlation, and there was a moderate positive correlation between DLGAP5 and RRM2 ($P < 0.001$, $R = 0.65$). AURKA and RRM2 ($P < 0.001$, $R = 0.48$) have a low expression correlation (**Figure 8B**).

Verification of the Differential Expression of CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 and the Analysis of Related Clinical Parameters

Analysis for gene differential expression through the UALCAN software (http://ualcan.path.uab.edu/analysis.html) indicated that there were significant differences in expression of the five genes in the normal group and the tumor group, both in colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) (**Figure 9**).

For COAD, taking the normal samples as the reference group, the expression of these five genes all significantly increased, whether in males or in females (**Figure S1A**), various age groups (**Figure S2A**), various races (**Figure S3A**), and various lymph node metastasis stages (**Figure S4A**).

For READ, taking the normal samples as the reference group, in terms of gender (**Figure S1B**), the expression of these five genes was significantly increased in males or females. In terms of age (**Figure S2B**), CCNA2, MAD2L1, and AURKA were 21–40. The expression of DLGAP5 was not obvious in the age groups

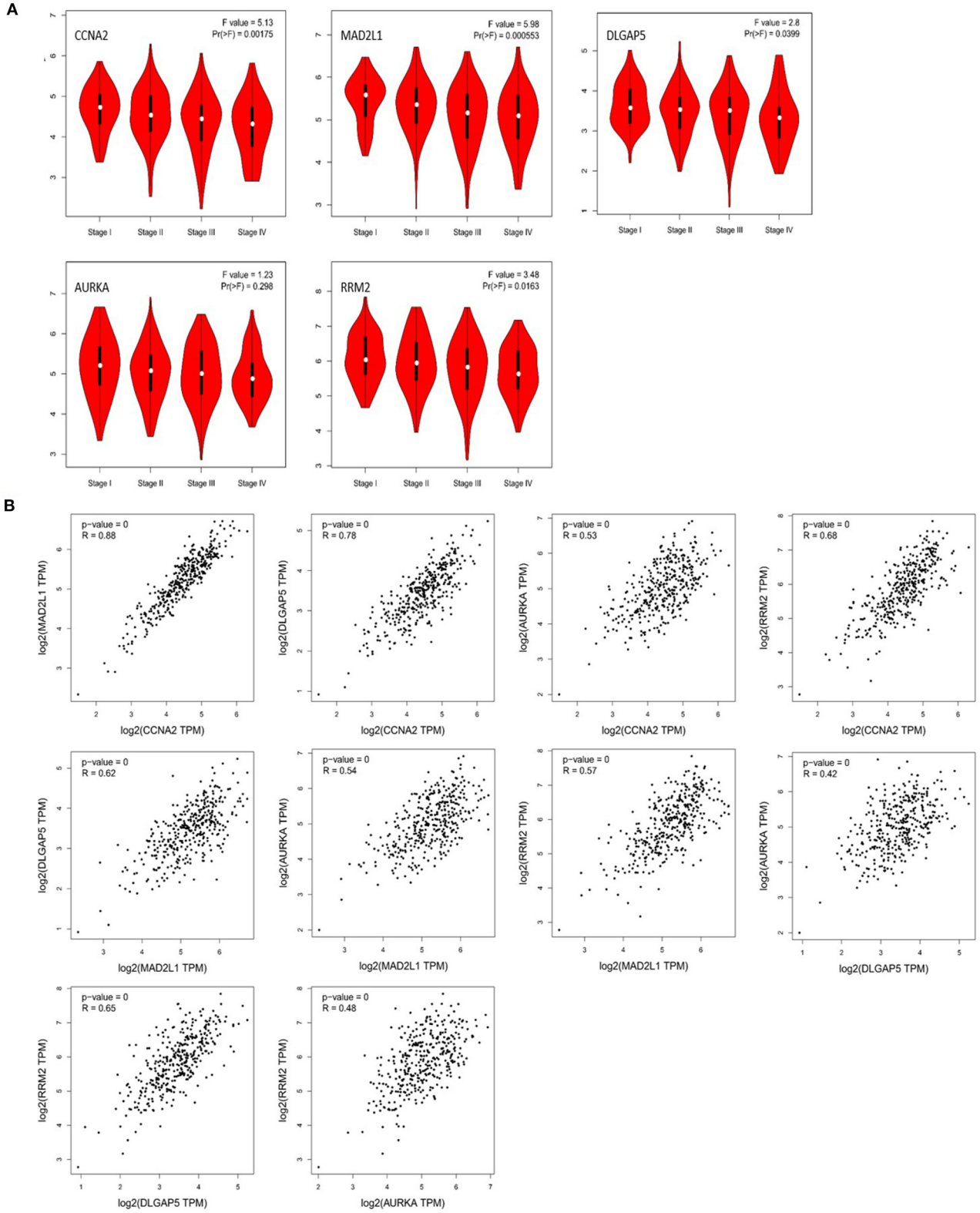


FIGURE 8 | Correlation analysis between CCNA2, MAD2L1, DLGAP5, and RRM2 and the pathological stage of colorectal adenocarcinoma **(A)**. Correlation analysis of the expression of CCNA2, MAD2L1, DLGAP5, and RRM2 in colorectal adenocarcinoma **(B)**.

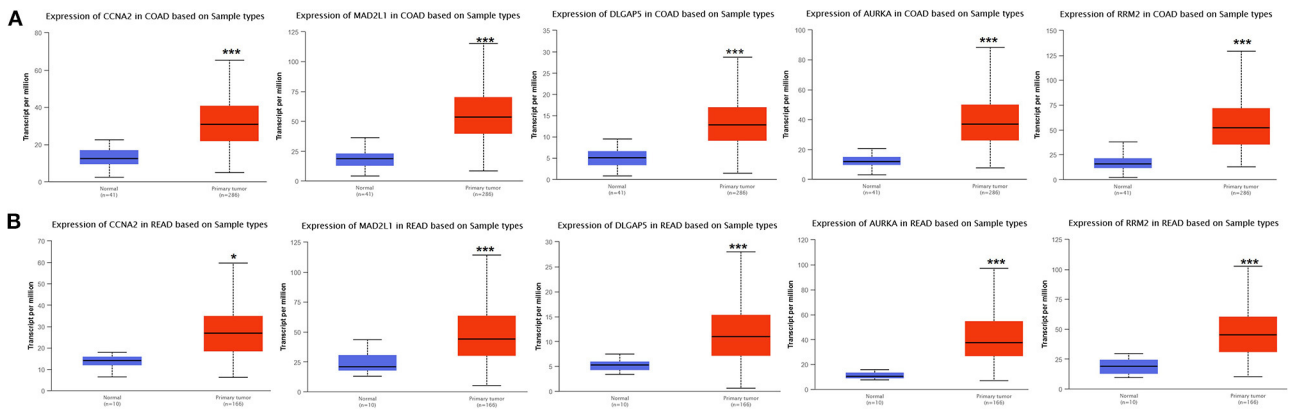


FIGURE 9 | Verification of the differential expression of CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 in CRC: COAD **(A)**; READ **(B)** (*$P < 0.05$, ***$P < 0.001$).

FIGURE 10 | Multiple-gene comparison analysis for the five CRC biomarker candidates.

of 41–60, 61–80, and 81–100. DLGAP5 was not significantly expressed in the groups of 21–40 and 81–100. Both the 60-year-old group and the 61–80-year-old group were significantly expressed, and RRM2 was significantly expressed in all age groups. In terms of race (**Figure S3B**), as the sample number of Asian patients was only one case, the Asian patient sample group was not tested. By comparison, the expression of these five genes was significantly increased in both the African-American patient group and the Caucasian patient group. In terms of lymph node metastasis (**Figure S4B**), the expression of CCNA2 was significantly increased in N1 and N2, but not in N3. MAD2L1, DLGAP5, AURKA, and RRM2 were significantly increased in the metastasis stage of each lymph node.

Possibility of a Five-Gene Biomarker in Diagnosis of CRC

Multiple-gene comparison analysis for the five CRC biomarker candidates was conducted using the GEPIA, with the only tumor data (COAD). Among the five genes, RRM2 had the highest expression level, followed by MAD2L1, AURKA, CCNA2, and DLGAP5 (**Figure 10**). Principal component analysis of the five genes was performed with TCGA tumor data, TCGA normal data, and GTEx data (both colon-sigmoid and colon-transverse); we found that the five genes could effectively distinguish between CRC samples and normal samples (**Figure 11**), indicating the possibility of a five-gene biomarker in diagnosis of CRC.

DISCUSSION

CRC is one of the most common cancers and carries a major global health burden. Globally, among all cancers, CRC ranked third in incidence and second in mortality in 2018 (Bray et al., 2018). There is no specific clinical symptom of colorectal cancer at the early stage, but when it was found, it had become in the middle and late stages. Surgery and radiotherapy and chemotherapy at the perioperative period or adjuvant chemotherapy is still the first choice for colorectal cancers (Schmoll et al., 2012). In those with metastatic disease, the treatment repertoire has been extended to include biologically targeted agents, including monoclonal antibodies targeting EGFR, such as cetuximab or panitumumab (Tabernero et al., 2015). As a result of improved treatment options, the overall

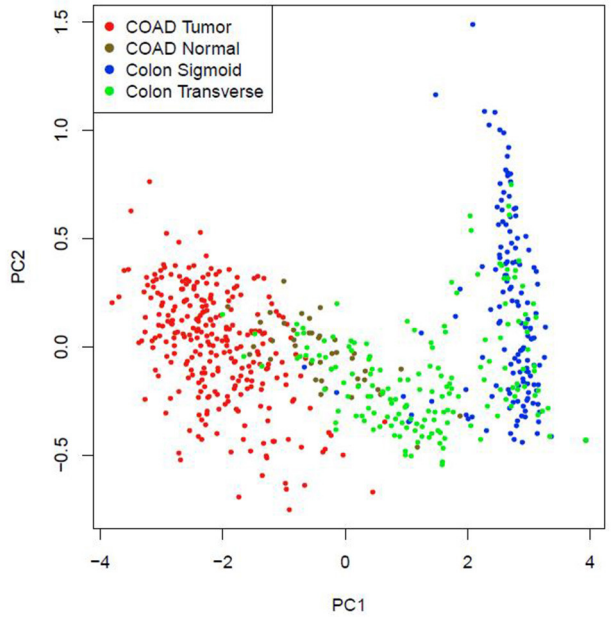


FIGURE 11 | Principal component analysis of the five CRC biomarker candidates.

survival (OS) of patients with metastatic CRC has increased from ∼1 year in the era of 5-fluorouracil (5-FU) therapy alone, to ∼3 years with currently available therapies (Cremolini et al., 2015). However, recurrence and overall survival are still the challenge for the treatment of colorectal cancer in clinic. Therefore, it is crucial to identify new markers that can predict CRC recurrence, overall survival (OS), and disease-free survival (RFS), subsequently separating patients into high- or low-risk groups for enhanced efficacy in further treatment. There are many clinical studies related to tumor recurrence. In this study, we focus on the genes about overall survival (OS) and disease-free survival (DFS) in CRC. Three datasets (GSE110223, GSE110224, GSE113513) from the GEO database were introduced into the analysis; 264 DEGs were overlapped in all the three datasets including 166 upregulated DEGs and 98 downregulated DEGs. Even though the three datasets had the similar number of samples, the number of DEGs of each dataset varied greatly. GSE113513 had the largest number of DEGs, largely more than those of other two datasets.

We found that CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 were closely related to the prognosis of CRC. Among them, we found that CCNA2, MAD2L1, DLGAP5, and AURKA were related to the overall survival (OS) of colorectal tumors, while RRM2 and AURKA had the relation between disease-free survival (DFS) and colorectal cancer. CCNA2, MAD2L1, DLGAP5, and RRM2 were all significantly related to the pathological stages of colorectal cancer and were also closely related to the stage of lymph node metastasis. There was a study that revealed that the expression of CCNA2 in CRC tissues is higher than that in normal tissues. The knockdown of CCNA2 could significantly suppress CRC cell growth by impairing cell cycle progression and inducing cell apoptosis (Gan et al., 2018). Our study showed that the abnormal expression of CCNA2 was also significantly related to the overall survival time, pathological stage of the tumor, and lymph node metastasis. There were articles that showed that CCNA2 is an important sign to judge the poor prognosis of the tumor, as it also highly expressed in pancreatic cancer, breast cancer, lung cancer, and other tumors (Gao et al., 2014; Peng et al., 2018; Brcic et al., 2019). MAD2L1, DLGAP5, and AURKA are the key genes for spindle assembly. When these genes are abnormally expressed, they will cause chromosome mismatch and other genetic problems during mitosis (Ooi et al., 2012). What is worse is that the unstable gene expression will eventually lead to cancer (Wassmann and Benezra, 2001; Weaver and Cleveland, 2005). Clinical studies had shown that DLGAP5 was related to the invasion and migration of CRC (Branchi et al., 2019); besides, DLGAP5 expression was also related to overall survival and lymph node metastasis but had no correlation with disease-free survival. It is an important measure of poor prognosis. Previous studies had verified that MAD2L1, DLGAP5, and AURKA were highly expressed in CRC (Chuang et al., 2016; Branchi et al., 2019; Ding et al., 2020). In our study, we found that these abnormally expressed genes not only induce the occurrence and development of tumors but also are significantly related to the overall survival, pathological staging of tumors, and tumor lymph node metastasis. When it comes to the expression of RRM2, studies showed that it related to the depth of invasion, degree of differentiation, disease-free survival (RFS), and metastasis of CRC (Lu et al., 2012; Liu et al., 2013). Our study showed that RRM2 was associated with disease-free survival of colorectal adenocarcinoma and was an important target gene predicted after tumor treatment.

In summary, it can be seen that CCNA2, MAD2L1, DLGAP5, RRM2, and AURKA are significantly related to the overall survival prognosis, disease-free survival, pathological stage, and lymph node metastasis stage of CRC, which are also important indicators for the evaluation of the prognosis of CRC and the evaluation of further treatment. Principal component analysis of the five genes indicated that they could effectively distinguish between CRC samples and normal samples. In order to increase the accuracy of diagnosing CRC, we suggested that the five biomarker candidates as a five-gene biomarker for diagnosis of CRC.

Once we understand the site where abnormal gene expression induces tumors, more targeted drugs will be applied. Most importantly, it provides a theoretical basis for future gene-level treatment of tumors and can achieve more precise targeted therapy. It is also important to guide the development of genetic kits and the non-invasive diagnosis of colorectal tumors.

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

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YC and XL designed and directed the research. ZW and MG conducted data collection, data statistics, and article writing for the research. JC, ZH, and XA assisted in related literature search. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by cultivation project of Zhuhai People's Hospital (2019PY-30) and cultivation project of Zhuhai People's Hospital (2019PY-19).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.602922/full#supplementary-material

REFERENCES

Ahluwalia, P., Mondal, A. K., Bloomer, C., Fulzele, S., Jones, K., Ananth, S., et al. (2019). Identification and clinical validation of a novel 4 gene-signature with prognostic utility in colorectal Cancer. *Int. J. Mol. Sci.* 20:3818. doi: 10.3390/ijms20153818

Bogaert, J., and Prenen, H. (2014). Molecular genetics of colorectal cancer. *Ann. Gastroenterol.* 27, 9–14. doi: 10.1111/j.1749-6632.1995.tb12114.x

Branchi, V., García, S. A., Radhakrishnan, P., Gyorffy, B., Hissa, B., Schneider, M., et al. (2019). Prognostic value of DLGAP5 in colorectal cancer. *Int. J. Colorectal Dis.* 34, 1455–1465. doi: 10.1007/s00384-019-03339-6

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 Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Brcic, L., Heidinger, M., Sever, A. Z., Zacharias, M., Jakopovic, M., Fediuk, M., et al. (2019). Prognostic value of cyclin A2 and B1 expression in lung carcinoids. *Pathology* 51, 481–486. doi: 10.1016/j.pathol.2019.03.011

Campos-da-Paz, M., Dórea, J. G., Galdino, A. S., Lacava, Z. G. M., and de Fatima Menezes Almeida Santos, M. (2018). Carcinoembryonic Antigen (CEA) and hepatic metastasis in colorectal cancer: update on biomarker for clinical and biotechnological approaches. *Recent Pat. Biotechnol.* 12, 269–279. doi: 10.2174/1872208312666180731104244

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., et al. (2017). UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 19, 649–658. doi: 10.1016/j.neo.2017.05.002

Chen, L., Lu, D., Sun, K., Xu, Y., Hu, P., Li, X., et al. (2019). Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* 692, 119–125. doi: 10.1016/j.gene.2019.01.001

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8(Suppl. 4):S11. doi: 10.1186/1752-0509-8-S4-S11

Chuang, T. P., Wang, J. Y., Jao, S. W., Wu, C. C., Chen, J. H., Hsiao, K. H., et al. (2016). Over-expression of AURKA, SKA3, and DSN1 contributes to colorectal adenoma to carcinoma progression. *Oncotarget* 7, 45803–45818. doi: 10.18632/oncotarget.9960

Cremolini, C., Schirripa, M., Antoniotti, C., Moretto, R., Salvatore, L., Masi, G., et al. (2015). First-line chemotherapy for mCRC—a review and evidence-based algorithm. *Nat. Rev. Clin. Oncol.* 12, 607–619. doi: 10.1038/nrclinonc.2015.129

Ding, X., Duan, H., and Luo, H. (2020). Identification of Core gene expression signature and key pathways in colorectal cancer. *Front. Genet.* 11:45. doi: 10.3389/fgene.2020.00045

Gan, Y., Li, Y., Li, T., Shu, G., and Yin, G. (2018). CCNA2 acts as a novel biomarker in regulating the growth and apoptosis of colorectal cancer. *Cancer Manag.Res.* 10, 5113–5124. doi: 10.2147/CMAR.S176833

Gao, T., Han, Y., Yu, L., Ao, S., Li, Z., and Ji, J. (2014). CCNA2 is a prognostic biomarker for ER+breast cancer and tamoxifen resistance. *PLoS One* 9:e91771. doi: 10.1371/journal.pone.0091771

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Liu, X., Zhang, H., Lai, L., Wang, X., Loera, S., Xue, L., et al. (2013). Ribonucleotide reductase small subunit M2 serves as a prognostic biomarker and predicts poor survival of colorectal cancers. *Clin. Sci. (Lond.)* 124, 567–578. doi: 10.1042/CS20120240

Lu, A. G., Feng, H., Wang, P. X., Han, D. P., Chen, X. H., and Zheng, M. H. (2012). Emerging roles of the ribonucleotide reductase M2 in colorectal cancer and ultraviolet-induced DNA damage repair. *World J. Gastroenterol.* 18, 4704–4713. doi: 10.3748/wjg.v18.i34.4704

Ooi, W. F., Re, A., Sidarovich, V., Canella, V., Arseni, N., Adami, V., et al. (2012). Segmental chromosome aberrations converge on overexpression of mitotic spindle regulatory genes in high-risk neuroblastoma. *Genes Chromosomes Cancer* 51, 545–556. doi: 10.1002/gcc.21940

Pathan, M., Keerthikumar, S., Chisanga, D., Alessandro, R., Ang, C. S., Askenase, P., et al. (2017). A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J. Extracell. Vesicles* 6:1321455. doi: 10.1080/20013078.2017.1321455

Peng, X., Pan, K., Zhao, W., Zhang, J., Yuan, S., Wen, X., et al. (2018). NPTX1 inhibits colon cancer cell proliferation through down-regulating cyclin A2 and CDK2 expression. *Cell Biol. Int.* 42, 589–597. doi: 10.1002/cbin.10935

Schmoll, H. J., Van Cutsem, E., Stein, A., Valentini, V., Glimelius, B., Haustermans, K., et al. (2012). ESMO consensus guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 23, 2479–2516. doi: 10.1093/annonc/mds236

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

Sun, B., Li, Y., Zhou, Y., Ng, T. K., Zhao, C., Gan, Q., et al. (2019). Circulating exosomal CPNE3 as a diagnostic and prognostic biomarker for colorectal cancer. *J. Cell. Physiol.* 234, 1416–1425. doi: 10.1002/jcp.26936

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Tabernero, J., Yoshino, T., Cohn, A. L., Obermannova, R., Bodoky, G., Garcia-Carbonero, R., et al. (2015). Ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that

progressed during or after first-line therapy with bevacizumab, oxaliplatin, and a fluoropyrimidine (RAISE): a randomised, double-blind, multicentre, phase 3 study. *Lancet Oncol.* 16, 499–508. doi: 10.1016/S1470-2045(15)70127-0

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45, W98–W102. doi: 10.1093/nar/gkx247

Vlachavas, E. I., Pilalis, E., Papadodima, O., Koczan, D., Willis, S., Klippel, S., et al. (2019). Radiogenomic analysis of F-18-fluorodeoxyglucose positron emission tomography and gene expression data elucidates the epidemiological complexity of colorectal cancer landscape. *Comput. Struct. Biotechnol. J.* 17, 177–185. doi: 10.1016/j.csbj.2019.01.007

Wassmann, K., and Benezra, R. (2001). Mitotic checkpoints: from yeast to cancer. *Curr. Opin. Genet. Dev.* 11, 83–90. doi: 10.1016/S0959-437X(00)00161-1

Weaver, B. A., and Cleveland, D. W. (2005). Decoding the links between mitosis, cancer, and chemotherapy: the mitotic checkpoint, adaptation, and cell death. *Cancer Cell* 8, 7–12. doi: 10.1016/j.ccr.2005.06.011

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer Publishing Company, Incorporated.

Yang, W. J., Wang, H. B., Wang, W. D., Bai, P. Y., Lu, H. X., Sun, C. H., et al. (2020). A network-based predictive gene expression signature for recurrence risks in stage II colorectal cancer. *Cancer Med.* 9, 179–193. doi: 10.1002/cam4.2642

Zheng, R. S., Sun, K. X., Zhang, S. W., Zeng, H. M., Zou, X. N., Chen, R., et al. (2019). [Report of cancer epidemiology in China, 2015]. *Zhonghua Zhong Liu Za Zhi* 41, 19–28. doi: 10.3760/cma.j.issn.0253-3766.2019.01.005

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 February 2021
doi: 10.3389/fgene.2020.565749

Integrating Clinical and Genomic Analyses of Hippocampal-Prefrontal Circuit Disorder in Depression

Naijun Yuan[1†], Kairui Tang[1†], Xiaoli Da[1], Hua Gan[1], Liangliang He[1,2], Xiaojuan Li[1], Qingyu Ma[1] and Jiaxu Chen[1,3]**

[1] Formula-Pattern Research Center, School of Traditional Chinese Medicine, Jinan University, Guangzhou, China, [2] College of Pharmacy, Jinan University, Guangzhou, China, [3] School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China

OPEN ACCESS

Edited by:
Zhichao Liu,
National Center for Toxicological Research (FDA), United States

Reviewed by:
Dan Li,
University of Arkansas at Little Rock, United States
Mary Yang,
University of Arkansas at Little Rock, United States

***Correspondence:**
Qingyu Ma
tmaqingyu@jnu.edu.cn
Jiaxu Chen
chenjiaxu@hotmail.com

[†] These authors have contributed equally to this work

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

Received: 26 May 2020
Accepted: 24 December 2020
Published: 05 February 2021

Citation:
Yuan NJ, Tang KR, Da XL, Gan H, He LL, Li XJ, Ma QY and Chen JX (2021) Integrating Clinical and Genomic Analyses of Hippocampal-Prefrontal Circuit Disorder in Depression. Front. Genet. 11:565749. doi: 10.3389/fgene.2020.565749

Major depressive disorder (MDD) is a prevalent, devastating and recurrent mental disease. Hippocampus (HIP)-prefrontal cortex (PFC) neural circuit abnormalities have been confirmed to exist in MDD; however, the gene-related molecular features of this circuit in the context of depression remain unclear. To clarify this issue, we performed gene set enrichment analysis (GSEA) to comprehensively analyze the genetic characteristics of the two brain regions and used weighted gene correlation network analysis (WGCNA) to determine the main depression-related gene modules in the HIP-PFC network. To clarify the regional differences and consistency for MDD, we also compared the expression patterns and molecular functions of the key modules from the two brain regions. The results showed that candidate modules related to clinical MDD of HIP and PFC, which contained with 363 genes and 225 genes, respectively. Ninety-five differentially expressed genes (DEGs) were identified in the HIP candidate module, and 51 DEGs were identified in the PFC candidate module, with only 11 overlapping DEGs in these two regional modules. Combined with the enrichment results, although there is heterogeneity in the molecular functions in the HIP-PFC network of depression, the regulation of the MAPK cascade, Ras protein signal transduction and Ephrin signaling were significantly enriched in both brain regions, indicating that these biological pathways play important roles in MDD pathogenesis. Additionally, the high coefficient protein–protein interaction (PPI) network was constructed via STRING, and the top-10 coefficient genes were identified as hub genes via the *cytoHubba* algorithm. In summary, the present study reveals the gene expression characteristics of MDD and identifies common and unique molecular features and patterns in the HIP-PFC network. Our results may provide novel clues from the gene function perspective to explain the pathogenic mechanism of depression and to aid drug development. Further research is needed to confirm these findings and to investigate the genetic regulation mechanisms of different neural networks in depression.

Keywords: depression, hippocampus-prefrontal circuit, integrative approaches, gene expression, clinical state

INTRODUCTION

Major depressive disorder (MDD) is a highly prevalent and debilitating psychiatric illness that can severely impair quality of life (Belzung et al., 2015). Estimates indicate that depression-related suicide is responsible for the loss of 1 million lives per year (Perlis et al., 2006; Kuo et al., 2015). The physical symptoms of MDD, which are chronic and interfere with daily tasks, behavior, and quality of life, are considered to result from detrimental emotional and cognitive processes.

Since the development of next-generation sequencing, numerous studies have performed transcriptome profiling to explore the mechanisms underlying neural plasticity and brain pathology (Galfalvy et al., 2013; Riglin et al., 2019). Chronic stress can cause wide-ranging changes in gene expression in the human brain, some of which contribute to functional deficits in brain cells (Kajiyama et al., 2010). Many previous expression profiling studies on human samples have identified individual genes as candidate contributors to the depression phenotype, but a comprehensive systematic analysis of the molecular and cellular changes in MDD is still lacking (Iwamoto et al., 2004a; Kroes et al., 2006; Zhang G. et al., 2020). Further analysis of only the representative molecules with the highest differential expression levels will cause the underlying connections among genes to be ignored.

Fortunately, a very well-known method, gene set enrichment analysis (GSEA), can be applied to analyze overall gene expression data among different groups (Subramanian et al., 2005; Reimand et al., 2019). GSEA can be used to identify functional categories or pathways in which genes exhibit coordinated alterations in expression under different kinds of conditions and is not limited to analyses of sets of differentially expressed genes (DEGs). One advantage of GSEA is its ability to highlight and analyze genes with significant phenotypes but relatively minimal changes in expression that may be difficult to detect with classical univariate statistics. Additionally, weighted gene correlation network analysis (WGCNA) is an efficient method that can be used determine the interactions between genes and disease-related phenotypes (Langfelder and Horvath, 2008). This approach involves the analysis and calculation of connectivity weights and topological overlap. The correlation matrix of co-expressed genes and the adjacency function formed by the gene network are defined, and the coefficient of dissimilarity of different nodes is calculated. Then, genes with similar expression profiles are clustered in gene modules. If certain genes always exhibit similar expression changes in the context of a specific physiological process or tissue, the genes are likely functionally related and can be defined as a module. Unlike the clustering criteria of conventional clustering methods (such as geometric distances), the clustering criteria of WGCNA have important biological meaning; thus, the results obtained by WGCNA have high reliability and can be used for numerous further analyses (Rangaraju et al., 2018; Zhang et al., 2019).

Previous autopsy and imaging studies on patients with MDD have indicated the existence of abnormalities in several brain regions, including the prefrontal cortex (PFC), cingulated cortex, hippocampus (HIP), and other brain areas (Li et al., 2015; Mamdani et al., 2015). The HIP-PFC circuit, the critical neural circuit in MDD research, has important functions in cognitive and emotional regulation (Carreno et al., 2016); however, little is known about the gene-related signatures and their correlations in the HIP-PFC neural circuit.

Therefore, we used WGCNA and GSEA with clinical information to explore the phenotype-centric gene networks of patients with MDD. This study used public data sets to ensure a comprehensive analysis, and the included data sets met the following fundamental criteria: (1) all patients had confirmed MDD; (2) complete public transcription data were available; and (3) annotation platform files were available. The available transcriptome datasets on CNS diseases, especially on mental disorders, are limited, and most prior studies focused on a single brain area. Previously, Lanz et al. (2019) collected multiple brain regions from patients with different psychiatric disorders, and systematically evaluated the genes shared between mental diseases. In this analysis, different brain region tissues were collected from the same patient, and all samples are based on the same platform for annotation. Therefore, to explore the transcriptional insights into HPC and PFC in major depression, we performed WGCNA and GSEA of the clinical information in GSE53987, which is the main dataset analyzed in the current study, and other single brain area datasets were incorporated for verification.

MATERIALS AND METHODS

Data Collection and Processing

Gene microarray datasets (GSE53987, GSE42546, and GSE12654) were retrieved from the Gene Expression Omnibus (GEO) database, and the basic information is listed in **Table 1**. GSE53987 contains the gene profiles of both the HIP (CTRL/MDD: 18/17) and PFC (CTRL/MDD: 19/17), and no demographic differences were found across the groups (**Supplementary Table S1**) (Lanz et al., 2019). Under the corresponding platform annotation, 20174 probes can be converted into genes. Then in the verification section, the GSE42546 is a high-throughput sequencing dataset for the HIP, which contains 29 healthy controls and 17 MDD patients with a total of 6297 genes (Kohen et al., 2014). GSE12654 is a gene set for PFC, which includes 15 healthy controls and 11 MDD, and a total of 8622 genes were annotated (Iwamoto et al., 2004b). We used R programming language software to analyze the datasets after converting the corresponding IDs of the probes to the official gene symbols. The subsequent analyses were based on the official gene symbols. Before further analysis, the raw profile data of the CEL files were processed for background correction, and the robust multiarray average (RMA) algorithm of the affy Bioconductor/R oligo package was used for noise reduction and normalized by applying quantile normalization (Gautier et al., 2004). The expression data of each dataset before and after adjustment are displayed in the **Supplementary Figures S1, S2**. The differential expression analysis of GSE53987 was performed using the R package limma, and the adjusted *P*-value was calculated using Benjamini and Hochberg (BH) correction. Genes with an adjusted *P*-value of

TABLE 1 | Fundamental information of included gene chips.

Data sets	Tissues	Samples (CTRL/MDD)	Platform	Type	Purposes
GSE53987	Hippocampus	18/17	GPL570	Affymetrix U133_Plus2 chips	Analysis
	Prefrontal cortex	19/17			
GSE42546	Hippocampus	29/17	GPL13393	RNA-seq	Validate
GSE12654	Prefrontal cortex	15/11	GPL8300	Affymetrix U95 Version 2	Validate

<0.05 were statistically significant, log (fold change) > 0 was considered to be upregulated, and <0 was considered to be downregulated compared with the control. For the RNA-seq data (GSE42546), the calcNormFactors function in the edgeR R package was used for correction and normalization (Ritchie et al., 2015). The flow chart of the overall study is shown in **Figure 1** (created with BioRender).

Gene Set Enrichment Analysis

GSEA version 3.0 software was used to perform the overall gene analysis on two groups of different samples (control versus depression). To examine whether significant regulatory differences were caused by depression at the functional and pathway levels, background data from the Molecular Signatures Database (MsigDB[1]) were analyzed. The normalized enrichment score (NES) and normalized *P*-value were used to determine statistical significance, as previously described (Subramanian et al., 2005). By calculating the NES, GSEA can resolve differences in gene set size and correlations between gene sets and the expression dataset. The NES value calculation is performed as follow:

$$NES = \frac{actual\ ES}{mean\ (ESs\ against\ all\ permutations\ of\ the\ dataset)}$$

According to the NES calculation method in GSEA, NES can be a positive or negative value. Positive and negative NES values indicate enrichment genes mainly at the top and bottom of the list, respectively. According to GSEA's instruction and recommendations[2], the nominal *P*-value estimates the statistical significance of the enrichment scores, thus a conventional 1000 permutation was used in our study to obtain a more accurate *P*-value.

WGCNA Analysis

To determine the interaction between genes and the disease-related phenotype, an expression matrix was calibrated though the system biology method using the R package WGCNA (Langfelder and Horvath, 2008). This method used gene expression data from two sample groups (the control and MDD groups in this study) to build co-expression pairwise correlation matrices. Because low-expressing or unvarying genes usually means noise, the expression matrix was pre-processed by calculating the variance and screening the genes in the top 50% of the variance. The algorithm assumes that the network is an adjacency matrix a_{ij}, and the co-expression similarity S_{ij} is defined as the absolute value of the correlation coefficient between the profiles of nodes i and j based on the default method, $S_{ij} = |\ cor(x_i, x_j)|$. Hereafter, the pickSoftThreshold function was used to calculate the soft threshold β when constructing each module. Then, by defining the gene co-expression correlation matrices, the algorithm constructs a hierarchical clustering tree and different gene expression modules. For each module, the quantification of module membership (MM) is defined as the correlation between the module eigengene (ME) and the gene expression profile. The modules are defined as having different degrees of clinical relevance based on the calculation of the association between ME and external features. The gene significance (GS) is represented as the correlation (absolute value) between the gene and the clinical traits. For the intramodular analysis of each module, a significant correlation between GS and MM indicates that the genes within that module have consistency with clinical traits. A module was considered important and analyzed further if it correlated with clinical traits and its internal genes significantly correlated with clinical properties. The important modules related to MDD were screened for further analysis.

Enrichment Analyses for Key Modules

Gene Ontology (GO) enrichment analysis was carried out to determine the key gene modules in the biological process (BP) category. Pathway enrichment analysis was carried out for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Reactome pathways. All genes in the genome were used as the enrichment background. The enrichment analysis was based on the Metascape platform (Saldanha, 2004; Zhou et al., 2019). The *P*-value was calculated for each enriched term (Hochberg and Benjamini, 1990). Terms with *P*-values < 0.05, and enrichment factors (ratios between the observed counts and the counts expected by chance) > 1.0 were clustered based on their membership similarities. More specifically, the *P*-values were calculated based on the cumulative hypergeometric distribution. Kappa scores were used as the similarity metric when performing hierarchical clustering on the enriched terms, and sub-trees with a similarity of >0.3 were considered to form a cluster. The most statistically significant term within a cluster was chosen to represent the cluster.

Identification and Validation of Hub Genes

The hub genes were identified based on protein interaction evidence from the STRING database (Altaf-Ul-Amin et al., 2006). Evidence for protein interaction of significantly dysregulated genes in key modules were retrieved from the STRING database,

[1]http://software.broadinstitute.org/gsea/msigdb

[2]https://www.gsea-msigdb.org/gsea/doc/GSEAUserGuideFrame.html

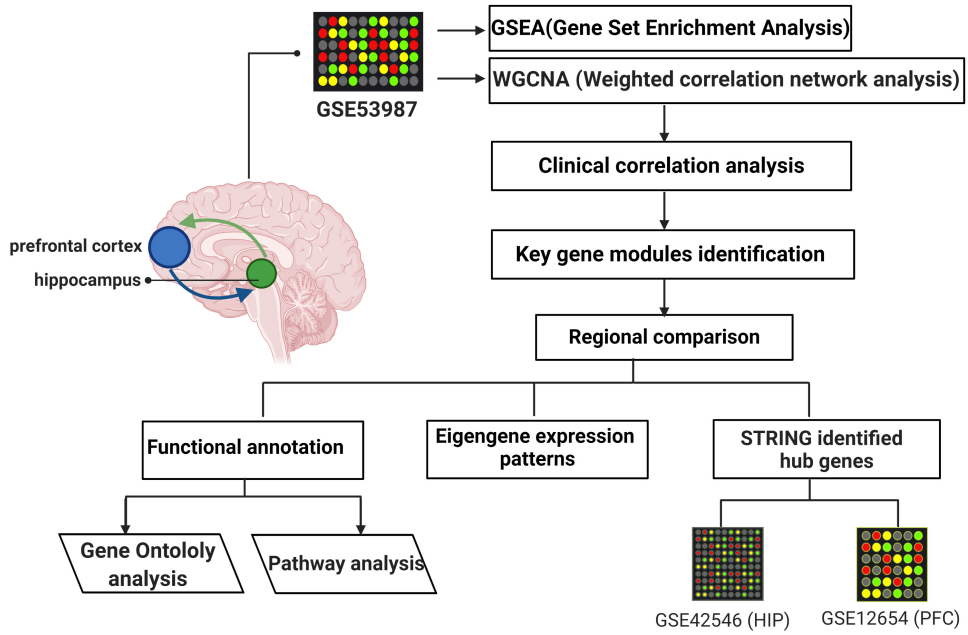


FIGURE 1 | The flow chart of overall study.

which required an interaction score with high confidence (0.7). The Cytoscape plugin *cytoHubba* was used to rank nodes based on the Maxial clique centrality (MCC) topological method, which predicts performance (Chin et al., 2014), and the top-10 genes were selected as hub genes for verification. MCC assumes that the node network is an undirected network; given a node v, S(v) is the set of the maximal cliques containing v, and (| C| −1)! is the product of all positive integers less than | C |. The calculation is as follows:

$$MCC(V) = \sum_{C \in S(v)} (|C| - 1)!$$

The separate datasets GSE42546 (HIP) and GSE12654 (PFC) were analyzed to verify gene expression with unpaired two-samples Wilcoxon tests.

RESULTS

GSEA for Different Brain Regions in MDD

GSEA was utilized to explore the functions and signaling pathways of gene sets that differed in the MDD groups compared with controls in different brain regions and thus to determine the potential biological significance of these genes sets in MDD. As listed in **Supplementary Table S2**, the HIP genes in the MDD group were primarily enriched for the GO processes of "dorsal/ventral neural tube patterning," "negative regulation of protein exit from endoplasmic reticulum" and "regulation of GTP binding," and the suppressed HIP genes were enriched for the "neuron remodeling," "serotonin secretion," and "synapse pruning" processes. The GSEA-based pathway results for the HIP indicated that several critical upstream pathways were suppressed in MDD, such as the "NF-kappa B signaling pathway" and the "TGF-beta signaling pathway"; downstream neuron-related pathways were also inactivated, such as the "GABAergic synapse" and "neuroactive ligand-receptor interaction" pathways (**Supplementary Table S2**).

For the PFC region, the GSEA results based on GO terms suggested that "stress-induced intrinsic apoptotic signaling pathway," "negative regulation of fibrinolysis," and "regulation of neuron projection arborization" were significantly activated in MDD (**Supplementary Table S2**). Moreover, pathways and GO terms related to synaptic functions and metabolic processes, such as "negative regulation of long-term synaptic potentiation," "synapse pruning," and the glutathione and glycerolipid metabolism pathways, were the most inhibited pathways/processes in MDD.

Key Gene Module Identification

The input for WGCNA contained all the gene profiles in the GSE53987 dataset; the corresponding clinical characteristics for different brain regions were analyzed separately. The R package WGCNA was applied to classify genes with similar expression patterns into different modules. We first selected a suitable soft threshold β value that met the scale-free conditions (R^2 = 0.9; HIP: 14; PFC: 12; **Figures 2A,B**). As shown in **Figures 2C,D**, we used a tree-cutting algorithm to calculate the average linkage clustering, obtain gene co-expression modules, merge similar modules, and gradually build a co-expression network. Ultimately, the gene modules were identified in the dataset for each brain area (HIP: 10; PFC: 12).

Then, we performed a relevance test to relate each module to clinical diagnostic parameters. The heatmap shown in **Figure 3**

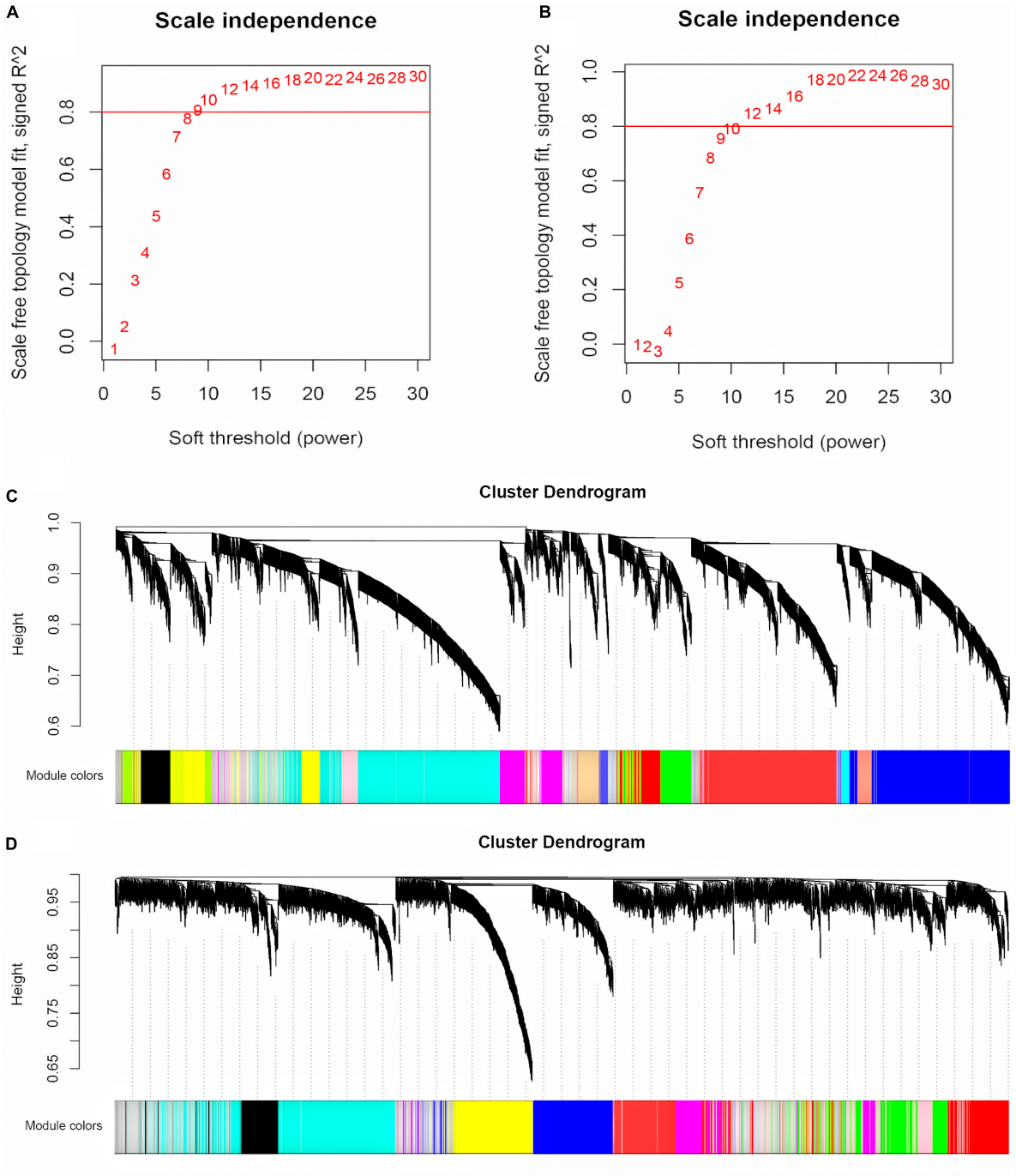


FIGURE 2 | Selecting the appropriate soft threshold and network construction in WGCNA. **(A)** Soft threshold filtering for HIP. **(B)** Soft threshold filtering for PFC. The horizontal axis is Soft threshold (power), and the vertical axis is the evaluation parameter of the scale-free network. The red horizontal line can clearly select the soft threshold when R^2 = 0.8. **(C)** Cluster dendrogram for HIP. **(D)** Cluster dendrogram for PFC. Hierarchical clustering tree showing each module, different colors represent different gene modules.

clearly indicates the clinical relevance of each module with an MDD diagnostic status. The clinical correlation analysis results suggested that the yellow module in the HIP ranked higher than any other module (**Figure 3A**) and was significantly related to clinical MDD (r = 0.37, P = 0.03). In the PFC (**Figure 3B**), the red module (r = 0.32, P = 0.07), magenta module (r = 0.30, P = 0.09), and turquoise module (r = 0.33, P = 0.06) ranked higher than other modules correlated with MDD. We then calculated the relationships between the nodes and modules (**Figure 4**). The correlation results indicated

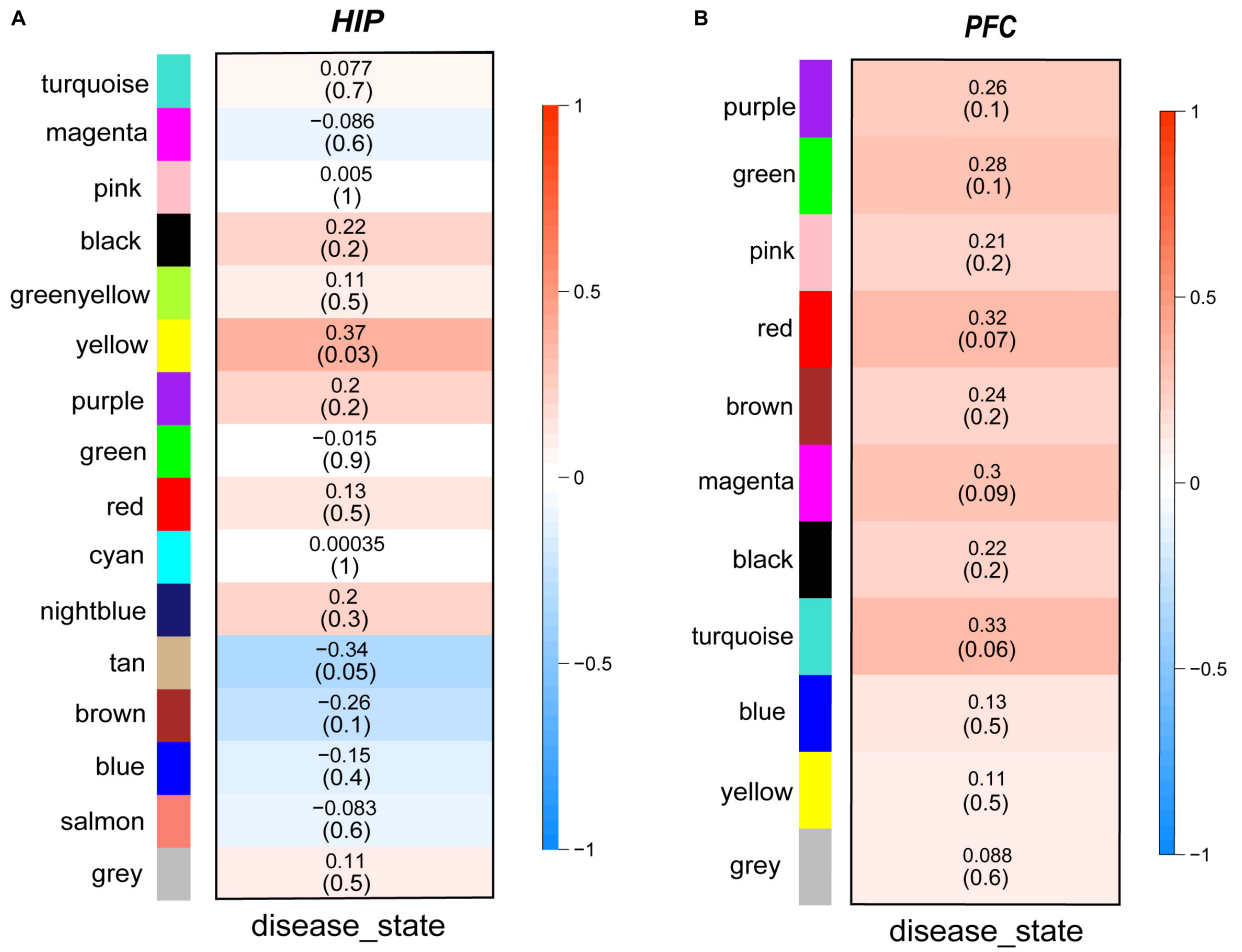


FIGURE 3 | (A,B) Module-trait relationships heatmaps. The heatmaps show the clinical relevance of different modules when the diagnosis status is MDD. The intensity of the color represents the strength of the correlation of modules. Red represents positive correlation and blue represents negative correlation. The specific correlation value and *P*-value are displayed in the corresponding module.

that the genes within the yellow module in the HIP were significantly correlated (cor = 0.44, P = 1.3e−18), and the genes within the red module in the PFC were significantly correlated (cor = 0.15, P = 0.024). These modules were consistent with the module analysis results, so they were identified as key modules for subsequent analyses (**Figure 4**). Additionally, the different modules contained different numbers of genes: the blue module in the HIP contained 363 genes, and the red module in the PFC contained 225 genes. Other genes were not classified into these modules because they lacked clinical relevance or co-expression characteristics; therefore, the modules containing these genes were not further analyzed.

Regional Comparison of Eigengene Expression via MDD Modules

To analyze the expression patterns of the MDD co-expression modules of the two brain regions, heatmaps were generated (**Figures 5A,C**). In the yellow module of HIP, the MDD co-expression genes have a different expression pattern than the normal control, with most showing upregulated gene expression; however, in the red module of PFC, the expression patterns of the module genes are not as clear as in HIP, and a wide range of differential expression is shown compared with the control group. To clearly determine the significance compared with the control, we combined the results of the differential analysis of the expression profiles (adjusted P-value < 0.05, **Figures 5B,D** and **Supplementary Table S2**). Among the 343 genes in the yellow module of HIP, 45 genes were significantly upregulated and 50 genes were significantly downregulated (**Figure 5B**). Compared with HIP, relatively few DEGs were identified in the PFC red module. Of the 225 genes in the PFC module, 22 genes were significantly upregulated and 29 were significantly downregulated (**Figure 5D**).

Interestingly, for these two gene sets based on clinical co-expression, few overlapping genes were found (32, **Figure 5E**). Eleven genes were differentially expressed in both HIP and PFC with similar expression trends. Among these intersecting dysregulated genes, 10 genes were significantly downregulated in HPC and PFC compared to the control group, and

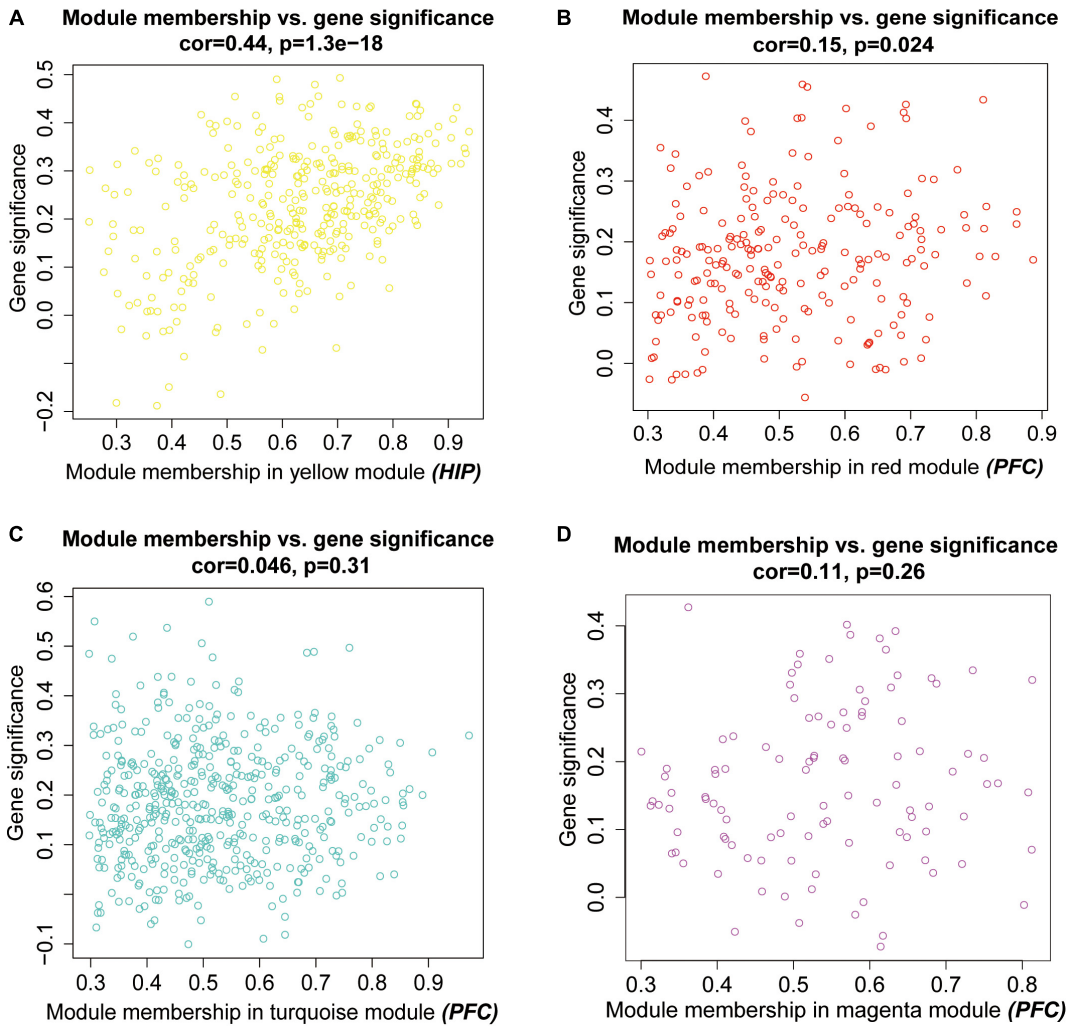


FIGURE 4 | Scatter plots for intramodular analysis. Each dot represents a gene, and different colors represent its module. **(A)** the intramodular analysis for the yellow model in hippocampus, **(B–D)** the intramodular analysis for the modules in prefrontal cortex.

BMP6 was significantly upregulated in both brain regions (**Supplementary Table S3**).

Regional Comparison of Functions via MDD Modules

To analyze the functional differences between the two key co-expression modules, we identified the BP and pathway characteristics with the Metascape database, which can provide GO clusters to more clearly define the enrichment results. The top significant BP enrichment clusters for each brain region are shown in **Figure 6A** and **Table 2**. The findings indicate significant differences between the HIP and PFC regions; the depression-related genes of the HIP are most significantly involved in processes associated with neuronal interaction and synaptic structure, such as signal release and cerebral cortex GABAergic interneuron migration, while the depression-related genes of the PFC are significantly involved in processes related to cellular component organization and development, such as telencephalon development, membrane depolarization and MAPKKK activity. The enriched pathways of HIP and PFC have distinguishable attributes (**Table 3** and **Figure 6B**); the HIP yellow module dysregulated genes were significantly enriched in signaling pathways associated with signal transmission between neurons, and synaptic transmission. The top enriched pathways included "Amine ligand-binding receptors," the downstream signaling pathways of neuronal inflammation and immunity, and "ADORA2B mediated anti-inflammatory cytokines production." The dysregulated genes in the red module of the PFC participate in the "MAPK signaling pathway" and downstream pathways of molecular metabolism.

The enriched results revealed identical BPs and pathways both in region-specific DEGs and overlapping DEGs (**Tables 2–4** and **Figure 6**), which indicate common biological changes in the HIP and PFC with regard to coping with MDD, including PIP3-activated AKT signaling, Ephrin signaling and

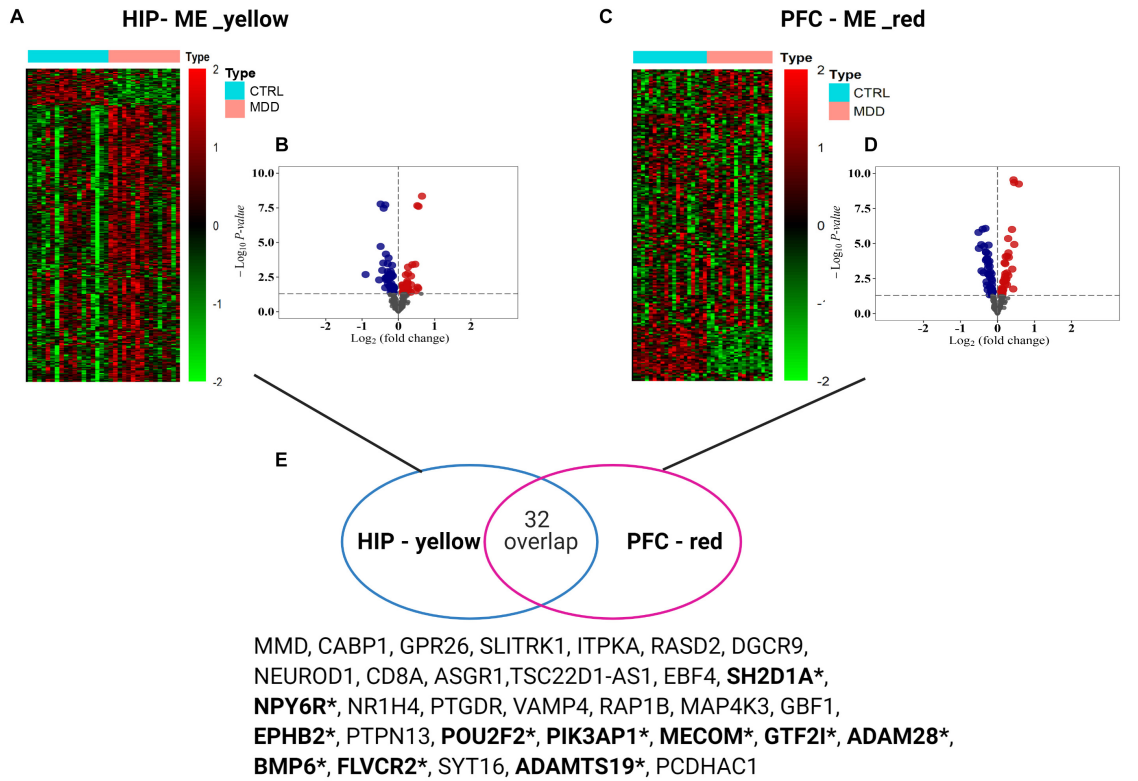


FIGURE 5 | (A,C) Heatmaps for key modules. The heatmaps of the modular genes expression levels between groups used the data normalization method of *Z*-score standardization. **(B,D)** Volcano plots for key modules. Gray dots mean genes with no significant difference, red color present up-regulated genes, and blue color dots indicate down-regulated genes **(E)** Venn diagram of gene intersection between two modules, *means the overlapping significant differentially expressed genes.

transcription-related downstream signaling pathways. These pathways are involved in the critical BPs of the immune response, and participate in the regulation of the MAPK cascade, neuronal differentiation and synaptic plasticity.

Hub Gene Selection and Verification

In total, the PPI network annotated by STRING includes 31 points, and the top-10 hub genes were identified though the *cytoHubba* algorithm, among which seven genes in the HIP, two genes in the PFC and EPHB2 in both regions were considered to have strong connections and more primary biological functions than the other genes in the network (**Figure 7A** and **Table 5**). The heatmaps were constructed for the related biological functions and signaling pathways for the hub genes (**Figure 7B** and **Table 6**). Neuroactive ligand-receptor interaction is the most significantly enriched pathway for these hub genes (ADRA1D, DRD1, PTGDR), and the process of "long-term synaptic potentiation" was regulated by the most hub genes (DRD1, EPHB2, VAMP2, GNB5). In addition, GNRH2, DRD1, ADRA1D, VAMP2, and SYCE1 were significantly upregulated and GNB5 and PTGDR were downregulated in the HIP of MDD patients compared with healthy controls (**Figure 7C**). BUB1B was significantly downregulated and LMNB1 was significantly upregulated in the PFC of MDD patients. EPHB2, which is involved in the regulation of synaptic enhancement, cation channel activity, neuron projection retraction, central nervous system neuron development and MAPKK activity, was downregulated in both the HIP ($P = 0.0009$) and PFC ($P = 0.0048$) of MDD patients (**Figures 7C,D**).

To verify our findings, we used different gene chips to analyze the expression of the hub genes. For the HIP (**Figure 8A**), 4 hub genes could be annotated in the GSE42546 gene set. The expression of GNB5 ($P = 0.037$), VAMP2 ($P = 0.0025$), and SYCE1 ($P = 0.031$) were significantly upregulated in the HIPs of MDD patients, which is consistent with the results of the analysis set. EPHB2 showed a non-significant decreased expression trend compared to healthy controls in the validation set ($P = 0.42$). Additionally, three hub genes (BUB1B, EPHB2, LMNB1) in the PFC could be annotated in the GSE12654 gene set (**Figure 8B**). BUB1B was significantly downregulated ($P = 0.0037$), and LMNB1 was significantly upregulated in MDD ($P = 0.011$), but EPHB2 expression did not differ compared with the control ($P = 0.087$).

DISCUSSION

Building on increasing evidence of the effects of depression on the brain, a previous study revealed that depression is associated with widespread network dysconnectivity rather than aberrant

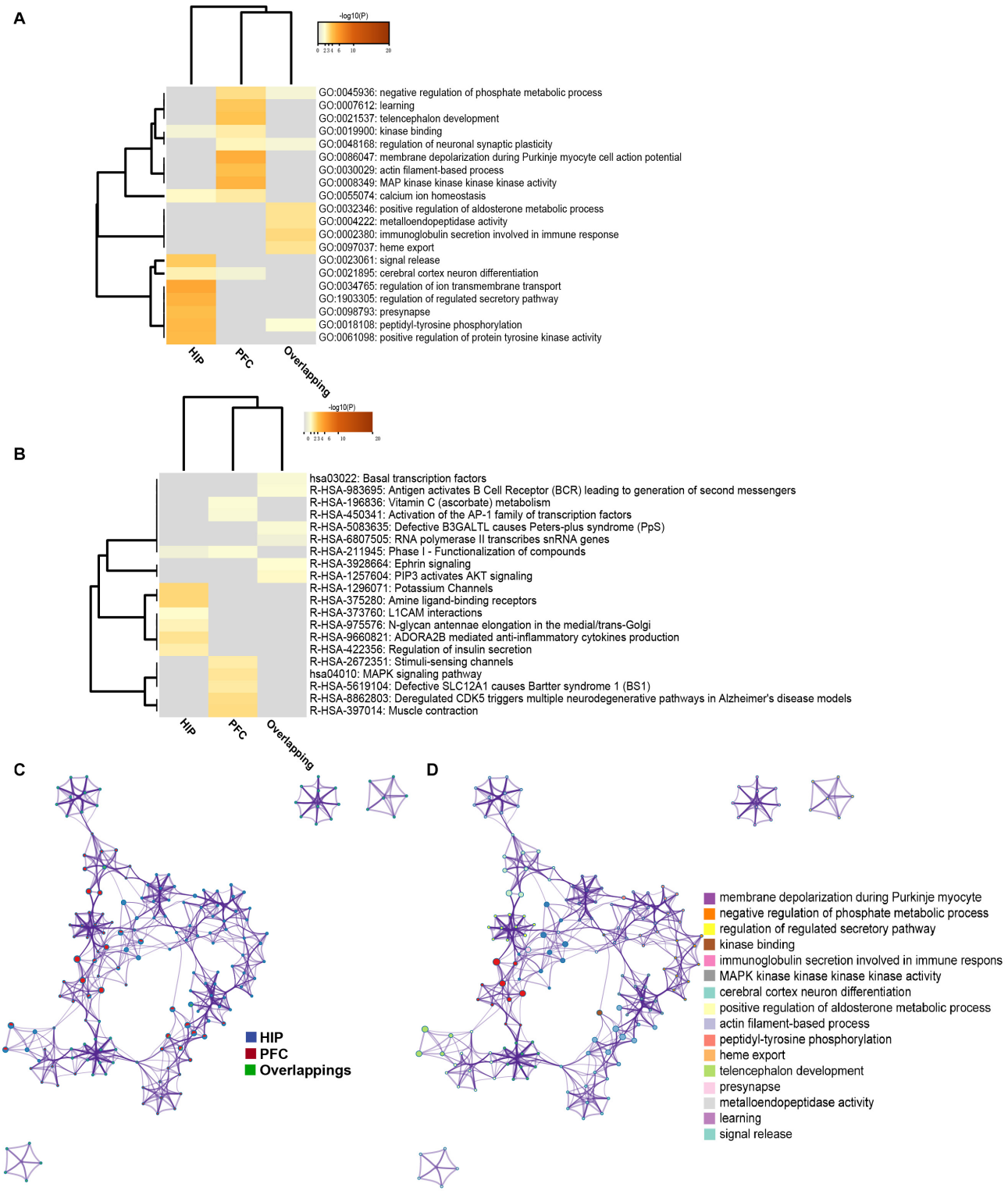


FIGURE 6 | (A) GO enrichment heatmap for regional specific DEGs (HIP, PFC) and overlapping DEGs. **(B)** KEGG enrichment heatmap for regional specific DEGs (HIP, PFC) and overlapping DEGs. Gray indicates terms have not been enriched, and the shade of orange represents the degree of enrichment. **(C)** Network of enriched terms represented as pie charts, where pies are color-coded based on the identities of the gene lists. **(D)** Network of enriched terms for HIP and PFC, colored by cluster name, where nodes that share the same cluster ID are typically close to each other.

TABLE 2 | GO enrichment for key modules.

GO_ID	Description	*P*-value		
Co-terms		**(*HIP*)**	**(*PFC*)**	***(Overlapping)***
GO:0002380	Immunoglobulin secretion involved in immune response			0.0005
GO:0032346	Positive regulation of aldosterone metabolic process			0.0009
GO:0097037	Heme export			0.0009
GO:0004222	Metalloendopeptidase activity			0.0009
GO:0001602	Pancreatic polypeptide receptor activity			0.0018
GO:0043409	Negative regulation of MAPK cascade			0.0031
GO:0036312	Phosphatidylinositol 3-kinase regulatory subunit binding		0.01791	0.0045
GO:0018108	Peptidyl-tyrosine phosphorylation	3.3E-05		0.0116
GO:0046580	Negative regulation of Ras protein signal transduction	0.01252		0.0228
GO:0048168	Regulation of neuronal synaptic plasticity		0.00399	0.0232
GO:0045936	Negative regulation of phosphate metabolic process		0.00067	0.0285
GO:2000474	Regulation of opioid receptor signaling pathway	0.00664	0.00361	
GO:0048167	Regulation of synaptic plasticity	0.00366	0.00482	
GO:0055074	Calcium ion homeostasis	0.00551	0.00174	
GO:0098662	Inorganic cation transmembrane transport	0.001	0.00243	
GO:0021895	Cerebral cortex neuron differentiation	0.00312	0.0442	
GO:0009410	Response to xenobiotic stimulus	0.01562	0.01514	
GO:0031749	D2 dopamine receptor binding	0.01651	0.009	
GO:1905702	Regulation of inhibitory synapse assembly	0.01323	0.0072	
GO:0019900	Kinase binding	0.03684	0.00204	
Hippocampus				
GO:0021853	Cerebral cortex GABAergic interneuron migration	0.00016		
GO:0023061	Signal release	0.00016		
GO:0071398	Cellular response to fatty acid	0.00091		
GO:0008422	Beta-glucosidase activity	0.00016		
GO:0098793	Presynapse	6E-05		
GO:0034765	Regulation of ion transmembrane transport	4.6E-06		
GO:0061098	Positive regulation of protein tyrosine kinase activity	3.6E-05		
GO:1903305	Regulation of regulated secretory pathway	1.7E-05		
GO:0019905	Syntaxin binding	0.00209		
GO:0034373	Intermediate-density lipoprotein particle remodeling	0.00332		
Prefrontal cortex				
GO:0007612	Learning		0.00013	
GO:0021537	Telencephalon development		0.0001	
GO:0021549	Cerebellum development		0.00091	
GO:0051899	Membrane depolarization		0.0006	
GO:0030029	Actin filament-based process		7.1E-05	
GO:0008349	MAP kinase kinase kinase kinase activity		1.9E-05	
GO:0086047	Membrane depolarization during Purkinje myocyte cell action potential		9.5E-06	
GO:0004597	Peptide-aspartate beta-dioxygenase activity		0.00181	
GO:0052593	Tryptamine:oxygen oxidoreductase (deaminating) activity		0.00361	
GO:0005668	RNA polymerase transcription factor SL1 complex		0.00361	

responses of individual brain regions (Li et al., 2018). Both structural and functional abnormalities in different areas have been found in MDD (Park and Friston, 2013; Dai et al., 2019). For example, abnormal functional neural circuitry under chronic stress conditions and decreased volumes of many brain regions have been noted in patients with depression (Korgaonkar et al., 2014). Furthermore, prior studies have revealed the importance of communication from the PFC to the HIP; neural circuitry disinhibition is the main cause of cognitive deficits in depression. Thus, the current study investigated gene signatures in distinct brain areas differentially affected by MDD given that genetic factors have been confirmed to play roles in MDD development (Bast et al., 2017).

First, we used a GSEA method to analyze overall gene data of the HIP and PFC in a holistic manner. The GSEA revealed that the involved genes in the PFC are mainly related to synapse generation and molecular metabolism. For example, the glutathione metabolism pathway is associated with genes that

TABLE 3 | Pathway enrichment for key modules.

Pathway_ID	Description	*P*-value		
Co-terms		**(*HIP*)**	**(*PFC*)**	***(Overlapping)***
R-HSA-1257604	PIP3 activates AKT signaling			0.00603
R-HSA-3928664	Ephrin signaling			0.00855
R-HSA-983695	Antigen activates B Cell Receptor (BCR) leading to generation of second messengers			0.01435
R-HSA-5083635	Defective B3GALTL causes Peters-plus syndrome (PpS)			0.01658
hsa03022	Basal transcription factors			0.02013
R-HSA-6807505	RNA polymerase II transcribes snRNA genes			0.03291
R-HSA-211945	Phase I – Functionalization of compounds	0.04864	0.01575	
Hippocampus				
R-HSA-375280	Amine ligand-binding receptors	0.00037		
R-HSA-1296071	Potassium Channels	0.00039		
R-HSA-9660821	ADORA2B mediated anti-inflammatory cytokines production	0.00105		
R-HSA-422356	Regulation of insulin secretion	0.00225		
R-HSA-975576	*N*-glycan antennae elongation in the medial/trans-Golgi	0.00337		
R-HSA-451326	Activation of kainate receptors upon glutamate binding	0.00507		
R-HSA-373760	L1CAM interactions	0.00735		
R-HSA-190374	FGFR1c and Klotho ligand binding and activation	0.00994		
R-HSA-1299503	TWIK related potassium channel (TREK)	0.00994		
R-HSA-8873719	RAB geranylgeranylation	0.01983		
R-HSA-391908	Prostanoid ligand receptors	0.02953		
R-HSA-8984722	Interleukin-35 Signaling	0.03918		
R-HSA-375281	Hormone ligand-binding receptors	0.03918		
R-HSA-1170546	Prolactin receptor signaling	0.04873		
Prefrontal cortex				
hsa04010	MAPK signaling pathway		0.00114	
R-HSA-5619104	Defective SLC12A1 causes Bartter syndrome 1 (BS1)		0.00181	
R-HSA-8862803	Deregulated CDK5 triggers multiple neurodegenerative pathways in Alzheimer's disease models		0.00072	
R-HSA-141405	Inhibition of the proteolytic activity of APC/C required for the onset of anaphase by mitotic spindl		0.03552	
R-HSA-5620916	VxPx cargo-targeting to cilium		0.03726	
R-HSA-450341	Activation of the AP-1 family of transcription factors		0.01791	
R-HSA-196836	Vitamin C (ascorbate) metabolism		0.01436	
R-HSA-1475029	Reversible hydration of carbon dioxide		0.02146	
R-HSA-9008059	Interleukin-37 signaling		0.03726	
R-HSA-2672351	Stimuli-sensing channels		0.0166	
R-HSA-210993	Tie2 Signaling		0.03202	
R-HSA-388844	Receptor-type tyrosine-protein phosphatases		0.03552	
hsa00360	Phenylalanine metabolism		0.03027	

are downregulated in MDD samples. Previous studies reported that glutathione is essential for cellular functions and plays a key role in redox balance *in vivo* mainly by eliminating free radicals to protect cells from oxidative stress (Bi et al., 2020). The current GSEA results for the PFC are consistent with previous clinical data showing that cortical glutathione is dysregulated in MDD. In addition, mounting evidence indicates that glutathione deficits are associated with increased inflammation and oxidative stress in MDD (Lapidus et al., 2014; Lindqvist et al., 2017). The GSEA results for the HIP showed that genes related to neuronal plasticity and serotonin secretion were downregulated and that genes related to the NF-κB signaling pathway and the TGF-β signaling pathway were also dysregulated in MDD samples. Several lines of evidence indicate that TGF-β pathway signaling is necessary for dopaminergic neuron development and function (Tesseur et al., 2017), and TGF-β1 is correlated with pathology in late-onset Alzheimer's disease and with MDD susceptibility (Zhang K. et al., 2020). The overall results of GSEA indicate that in the context of MDD, the transcriptomic changes in the PFC lead to the abnormal regulation of ion metabolism, while the transcriptomic changes in the HIP lead to abnormalities in the immune and neurotransmitter systems.

To identify specific genes closely related to the progression of depression, we performed WGCNA of these brain areas. Among the genes in the co-expression module of HIP and PFC under major depression conditions, more differences and less gene expression pattern overlap were found, and the HIP yellow module showed more DEGs than the PFC red module. Few overlapping genes were found between regions, indicating decreased expression consistency of HIP and PFC in MDD.

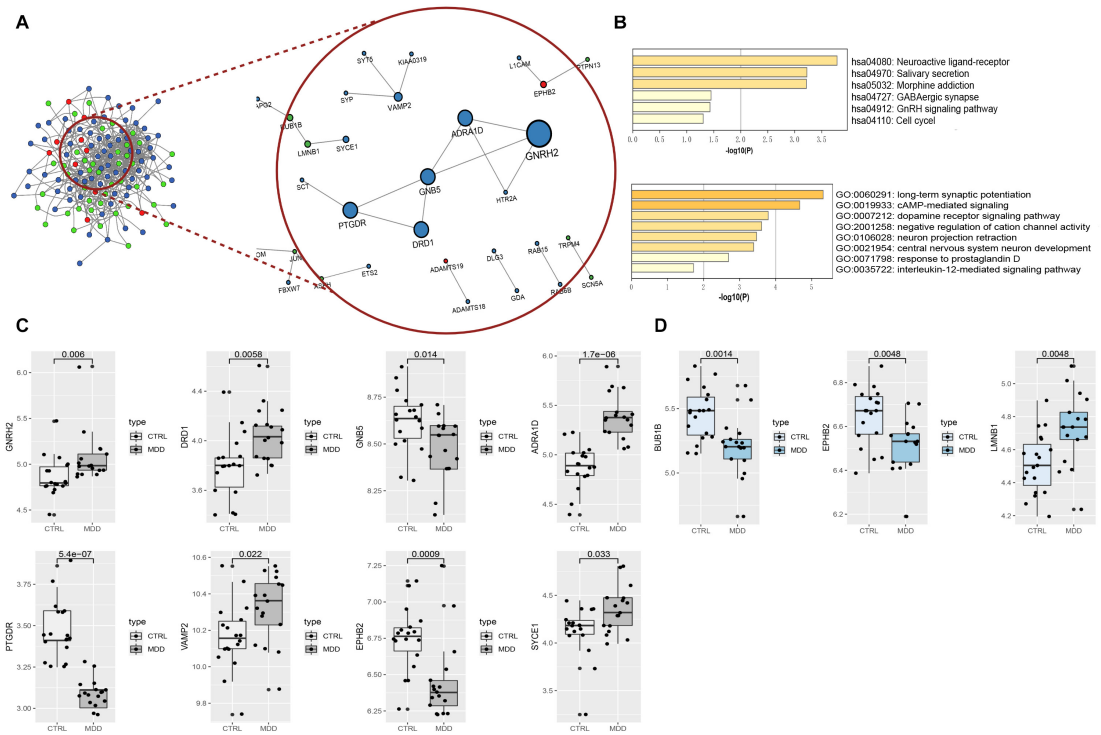


FIGURE 7 | (A) PPI network and hub components identified in the HIP and PFC region. Colored by Counts (full connection which according to STRING's minimum confidence of 0.15); nodes with 0.7 STRING high-confidence nodes only keep in the red circle that shows the high-confidence PPI network and the hub genes identified by *cytoHubba*. The larger the nodes of genes with high coefficients; the red dots represent the common genes of the two brain regions, blue nodes mean HIP genes, and green nodes indicate PFC genes. **(B)** Heatmaps of hub gene enrichment results (pathway and GO biological process). **(C)** Boxplots of hub genes expression in HIP of the analysis set (GSE53987). **(D)** Boxplots of hub genes expression in PFC of the analysis set (GSE53987).

TABLE 4 | Overlapping DEGs of key modules.

	PFC		*HIP*	
Gene	**Fold Change**	**adj.P.Val**	**Fold Change**	**adj.P.Val**
ADAM28	−1.22718	0.0012	−1.4067	1.44E-05
ADAMTS19	−1.24018	0.0005	−1.1426	0.0129
EPHB2	−1.10922	0.0349	−1.2243	0.0011
FLVCR2	−1.18999	0.0084	−1.1668	0.0120
GTF2I	−1.41651	0.0003	−1.3687	0.0008
MECOM	−1.42665	3.96E-05	−1.4524	0.0039
NPY6R	−1.1682	0.0088	−1.1726	0.0002
PIK3AP1	−1.22638	0.0226	−1.2861	0.0172
POU2F2	−1.19051	0.0007	−1.2575	0.0050
SH2D1A	−1.14261	0.0069	−1.3322	2.80E-08
BMP6	1.31198	2.73E-05	1.27219	0.0024

adj.P.Val, adjust P-value.

TABLE 5 | Hub node's information.

Gene	Score	Region-specific	Dysregulated
GNRH2	12	HIP	Up
DRD1	6	HIP	Up
GNB5	6	HIP	Down
ADRA1D	6	HIP	Up
PTGDR	6	HIP	Down
VAMP2	3	HIP	Up
BUB1B	2	PFC	Down
EPHB2	2	both	Down
LMNB1	2	PFC	Up
SYCE1	2	HIP	Up

With regard to the connections of functions, the two regional co-expression modules also showed heterogeneity. The results point to specific signaling pathways and BP clusters in the PFC and/or HIP. Both brain regions are enriched in the BP of the regulation of neuronal synaptic plasticity; however, the genes in the HIP are still more enriched in processes related to ion transport and signal release, while the depression-related genes in the PFC are significantly enriched in processes related to membrane depolarization and brain tissue development. The enhancement of synaptic potency and the plasticity of neural circuits are considered the primary cellular mechanism of learning and memory. Notably, the PFC has been related to cognitive and executive functions, and PFC dysfunction is indeed a pathological mechanism of depression. In addition, in primates, the HIP is located near the center of the brain and is one of the main limbic brain regions that stores memories and regulates cortisol production; however, its most important

function is in adult neurogenesis. Different BPs and signaling pathways are affected in these brain regions in the context of depression due to the physiological distinction between the HIP and PFC. Previous constructive research on schizophrenia suggested that there are few overlapping genes and BPs in the HIP and PFC, which means that the regional consistency in schizophrenia is reduced (Collado-Torres et al., 2019). Our results are consistent with previous studies and provide further molecular biological evidence of the decreased regional connectivity in the depressive state.

Notably, MDD-related genes in both regions were found to be involved in Ephrin signaling, possibly due to the structural changes in neuronal connections in neurological diseases, from synaptic changes to the loss or reconnection of entire axon bundles. Ephrin ligands and their Eph receptors are membrane-bound molecules that mediate the axon guidance through cell-cell contacts, and play a vital role in the guidance of neuronal growth cones, synaptic plasticity, and neuron migration (Torii et al., 2009; Klein and Kania, 2014). Previous studies demonstrated that Ephrin type-B receptor 2 (EPHB2) in the hypothalamus was significantly increased and interacted with the accumulated NMDAR subunit GluN2A in LPS-induced depressive phenotype mice (Wu et al., 2019), and the ratios of p-EphA4/EphA4 and p-ephexin1/ephexin1 in the PFC and HIP were increased in the social frustration depression mouse model (Zhang et al., 2017). In addition, the "MAPK signaling pathway" was significantly enriched in PFC, and "negative regulation of MAPK cascade" and "negative regulation of Ras protein signal transduction" were significantly enriched in the HIP-PFC network. These pathways and biological progresses are involved in the classic brain-derived neurotrophic factor (BDNF)-MAPK pathway. Stress factors alter BDNF activity and influence the BDNF-Ras-MAPK pathway, impairing neuronal cell survival and neuroplasticity, thereby resulting in depression symptoms (Easton et al., 2006). Previous studies have demonstrated that the Ras/MAPK pathway can regulate synaptic plasticity and affect the expression of hippocampal neuronal plasticity-related proteins in mice with chronic depression-like symptoms induced by morphine withdrawal (Zhu et al., 2002; Jia et al., 2013). Overactivation of the Ras/MAPK pathway impacts memory and learning behaviors in rats. Although preceding studies have reported the potential roles of MAPK signaling in depression (Aguilar-Valles et al., 2018), the downstream cascade targets of the MAPK pathway and the roles of their interactions in functional neural circuits in the regulation of depression are still unclear.

The top-10 genes were selected as hub genes from among the DEGs based on the STRING annotation and the *cytoHubba* plugin. The selected genes are involved in BPs such as long-term synaptic potentiation (LTP) and neuron projection retraction, which are representative BPs in the mechanism of depression. LTP and long-term inhibition (LTD) are two opposite forms of synaptic plasticity that contribute to fine-tuning neural connections to store information in the brain. Neural circuit information is disrupted during depression, resulting in learning and memory deficits associated with impaired synaptic plasticity (Adhikari et al., 2010; Padilla-Coreano et al., 2016; Bast et al., 2017). In one study, Zheng and Zhang (2015) clarified the potential causes of synaptic plasticity deficits in the HIP-PFC network during depression and recorded local field potentials (LFPs) in two brain regions of rats under normal and chronic unpredictable stress (CUS). The authors found that impaired synaptic plasticity in the ventral CA1 (vCA1)–medial PFC (mPFC) pathway was reflected in weakened theta coupling and theta-gamma cross-frequency coupling of LFPs in the depressed state. EPHB2 was identified as a hub gene that was significantly downregulated in both the HIP and PFC of depressed patients (analysis set). Even though it did not show significant differential expression in the validation set, many previous studies reported that the receptor tyrosine kinase EphB2 is inactivated in neuropsychiatric disorders including depression and memory disorders. Moreover, EPHB2-knockdown mice exhibit depression-like behaviors and spatial memory deficiency compared with wild-type mice (Zhen et al., 2018). In mice affected by chronic social stress, the expression level of EphB2 and its downstream molecules in mPFC are reduced, and the specific cleavage of the EphB2 receptor can increase sensitivity to stress and induce depression-like behavior (Bouzioukh et al., 2007). The synapses in the HIP of EphB2 mutant mice had normal morphology, but synaptic plasticity was defective and the LTP was attenuated in EphB2(−/−) hippocampal slices (Grunwald et al., 2001).

According to previous reports, HIP-PFC functional circuit interactions mainly occur in the theta frequency range and reflect the processes of working memory. In addition, the active processes in the two brain regions are associated with enhanced oscillatory activities (Spellman et al., 2015); thus, impaired functional neural circuits affect cognitive and memory processes in depression. In addition to discovering only a few overlapping molecular pathways in the HIP-PFC circuit, indicating weak regional coherence, we found that the overlapping pathways may be regularized responses between the HIP-PFC neural circuits that underlie the critical mechanism involved in the pathogenesis of MDD. Therefore, based on the identified gene characteristics of the two brain regions, further research could be performed on the molecular mechanism of this abnormal neural circuit in depression, particularly with regard to the overlapping pathways and several BPs that are closely associated with the HIP-PFC circuit.

The current study has certain limitations. First, most of the current evidence suggests that several brain regions are involved in MDD, including the PFC, HIP, amygdala, thalamus, and STR, and both structural and functional abnormalities in these areas have been found to occur in depression. There are also abnormalities in other neural circuits, such as the basal ganglia-thalamus-cortex circuit, that could be involved in the development of depression; however, we only analyzed the HIP-PFC neural network in this paper, given the few human disease samples and difficulty in analyzing each neural network or neural circuit extensively. WGCNA analysis associates clustered genes with a variety of external clinical information, and the features of clinical depression include the occurrence of anxiety, cognitive symptoms, and sleep phase disorder. This

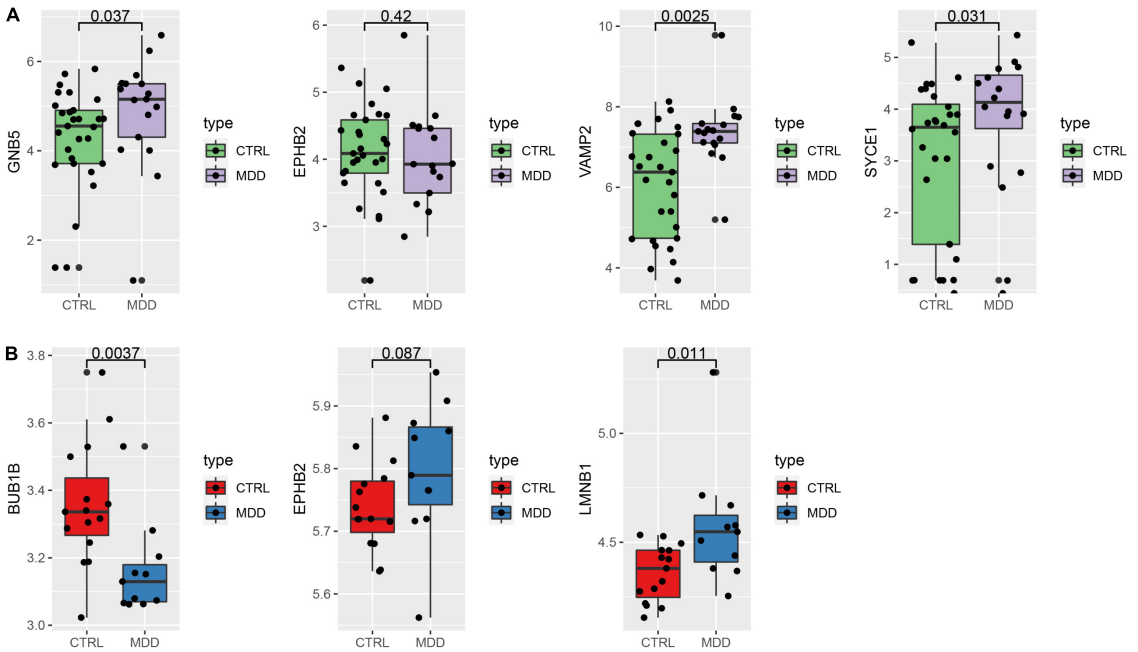


FIGURE 8 | Verify the expression of the hub genes. **(A)** Hub genes expression in the validation set GSE42546 (HIP) **(B)** hub gene expression in the validation set GSE12654 (PFC).

TABLE 6 | Enrichment results for hub genes.

Term	Description	*P*-value	Symbols
Pathway			
hsa04080	Neuroactive ligand-receptor interaction	0.000164	ADRA1D, DRD1, PTGDR
hsa04970	Salivary secretion	0.000596	ADRA1D, VAMP2
hsa05032	Morphine addiction	0.000609	DRD1, GNB5
hsa04727	GABAergic synapse	0.03554	GNB5
hsa04912	GnRH signaling pathway	0.037128	GNRH2
hsa04110	Cell cycle	0.049748	BUB1B
Biological progress			
GO:0060291	Long-term synaptic potentiation	4.83E-06	DRD1, EPHB2, VAMP2, GNB5
GO:0019933	cAMP-mediated signaling	5.05E-05	ADRA1D, DRD1, PTGDR
GO:0007212	Dopamine receptor signaling pathway	0.000162	DRD1, GNB5, VAMP2
GO:2001258	Negative regulation of cation channel activity	0.000248	EPHB2, GNB5
GO:0106028	Neuron projection retraction	0.00041	EPHB2, DRD1
GO:0021954	Central nervous system neuron development	0.000495	DRD1, EPHB2
GO:0071798	Response to prostaglandin D	0.00205	PTGDR, VAMP2
GO:0035722	Interleukin-12-mediated signaling pathway	0.019125	LMNB1
GO:0051301	Cell division	0.024569	BUB1B, SYCE1
GO:0051129	Negative regulation of cellular component organization	0.037388	BUB1B, EPHB2
GO:0051389	Inactivation of MAPKK activity	0.000821	EPHB2

study was conducted based on the current limited clinical information, which may lead to the potential external clinical features of other gene modules not being discovered. This study analyzed the two brain regions of MDD based on the gene expression level, therefore, we verified the DEGs from key modules, however, the analysis and verification of the connection between gene expression and function requires further exploration. Last, we used other gene datasets for verification, but several of the hub genes were not expressed or had no expression differences between the groups. The different annotation platforms and sample sizes might have caused this discrepancy.

Above all, the present study shows the gene expression characteristics of MDD and reveals common and unique

molecular features and patterns in the HIP-PFC network. Our results may provide novel clues from the gene function perspective to explain the pathogenic mechanism of MDD and to aid in the development of drug interventions for depression. Further research is needed to confirm our findings and to investigate the mechanisms of gene regulation of different neural networks in depression.

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/s.

AUTHOR CONTRIBUTIONS

NY, QM, and JC conceived of and designed the study. NY, KT, LH, XL, XD, and HG collected and processed the data. NY and KT wrote the manuscript. All the authors contributed to manuscript revision and read and approved the final submitted version. QM and JC take primary responsibility for communication with the journal and editorial office during the submission process, throughout peer review and during publication.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (Nos. 81630104, 81973748, 81803998, and 82003914), Huang Zhendong Research Fund for Traditional Chinese Medicine of Jinan University, and the Key-Area Research and Development Program of Guangdong Province (No. 2020B1111100001).

ACKNOWLEDGMENTS

The authors gratefully appreciate the reviewers and editors for their constructive suggestions to improve the manuscript, and sincerely thank the contributors of the GSE53987, GSE42546, and GSE12654 datasets at NCBI Gene Expression Omnibus. Besides, NY would like to thank Dr. Chuiguo Hoang for his helpful comments throughout the development of this article and thank the section members in PRI of Kyoto University for their comments and helpful discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.565749/full#supplementary-material

REFERENCES

Adhikari, A., Topiwala, M. A., and Gordon, J. A. (2010). Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety. *Neuron* 65, 257–269. doi: 10.1016/j.neuron.2009.12.002

Aguilar-Valles, A., Haji, N., De Gregorio, D., Matta-Camacho, E., Eslamizade, M. J., Popic, J., et al. (2018). Translational control of depression-like behavior via phosphorylation of eukaryotic translation initiation factor 4E. *Nature communications* 9, 2459. doi: 10.1038/s41467-018-04883-5

Altaf-Ul-Amin, M., Shinbo, Y., Mihara, K., Kurokawa, K., and Kanaya, S. (2006). Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC bioinformatics* 7:207. doi: 10.1186/1471-2105-7-207

Bast, T., Pezze, M., and Mcgarrity, S. (2017). Cognitive deficits caused by prefrontal cortical and hippocampal neural disinhibition. *British journal of pharmacology* 174, 3211–3225. doi: 10.1111/bph.13850

Belzung, C., Willner, P., and Philippot, P. (2015). Depression: from psychopathology to pathophysiology. *J Current opinion in neurobiology* 30, 24–30. doi: 10.1016/j.conb.2014.08.013

Bi, W. K., Shao, S. S., Li, Z. W., Ruan, Y. W., Luan, S. S., Dong, Z. H., et al. (2020). FSHR ablation induces depression-like behaviors*. *Acta pharmacologica Sinica* doi: 10.1038/s41401-020-0384-8

Bouzioukh, F., Wilkinson, G. A., Adelmann, G., Frotscher, M., Stein, V., and Klein, R. (2007). Tyrosine phosphorylation sites in ephrinB2 are required for hippocampal long-term potentiation but not long-term depression. *J the Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 27, 11279–11288. doi: 10.1523/jneurosci.3393-07.2007

Carreno, F. R., Donegan, J. J., Boley, A. M., Shah, A., Deguzman, M., Frazer, A., et al. (2016). Activation of a ventral hippocampus-medial prefrontal cortex pathway is both necessary and sufficient for an antidepressant response to ketamine. *Molecular psychiatry* 21, 1298–1308. doi: 10.1038/mp.2015.176

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome*. *J Bmc Systems Biology* S11, doi: 10.1186/1752-0509-8-s4-s11

Collado-Torres, L., Burke, E. E., Peterson, A., Shin, J., Straub, R. E., Rajpurohit, A., et al. (2019). Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia 8*. 103, 203.e–216.e. doi: 10.1016/j.neuron.2019.05.013

Lindqvist, D., Dhabhar, F. S., James, S. J., Hough, C. M., Jain, F. A., Bersani, F. S., et al. (2017). Oxidative stress, inflammation and treatment response in major depression. *Psychoneuroendocrinology* 76, 197–205. doi: 10.1016/j.psyneuen.2016.11.031

Dai, L., Zhou, H., Xu, X., and Zuo, Z. (2019). Brain structural and functional changes in patients with major depressive disorder: a literature review. *PeerJ* 7, e8170. doi: 10.7717/peerj.8170

Galfalvy, H., Zalsman, G., Huang, Y. Y., Murphy, L., Rosoklija, G., Dwork, A. J., et al. (2013). A pilot genome wide association and gene expression array study of suicide with and without major depression. *The world journal of biological psychiatry* 14, 574–582. doi: 10.3109/15622975.2011.597875

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy–analysis of Affymetrix GeneChip data at the probe level. *J Bioinformatics* 20, 307–315. doi: 10.1093/bioinformatics/btg405

Grunwald, I. C., Korte, M., Wolfer, D., Wilkinson, G. A., Unsicker, K., Lipp, H. P., et al. (2001). Kinase-independent requirement of EphB2 receptors in hippocampal synaptic plasticity*. 32, 1027–1040. doi: 10.1016/s0896-6273(01)00550-5

Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in medicine* 9, 811–818. doi: 10.1002/sim.4780090710

Iwamoto, K., Kakiuchi, C., Bundo, M., Ikeda, K., and Kato, T. (2004a). Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *Molecular psychiatry* 9, 406–416. doi: 10.1038/sj.mp.4001437

Iwamoto, K., Kakiuchi, C., Bundo, M., Ikeda, K., and Kato, T. (2004b). Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. *J Molecular Psychiatry* 9, 406–416.

Easton, J. B., Royer, A. R., and Middlemas, D. S. (2006). The protein tyrosine phosphatase, Shp2, is required for the complete activation of the RAS/MAPK pathway by brain-derived neurotrophic factor. *Journal of neurochemistry* 97, 834–845. doi: 10.1111/j.1471-4159.2006.03789.x

Jia, W., Liu, R., Shi, J., Wu, B., Dang, W., Du, Y., et al. (2013). Differential Regulation of MAPK Phosphorylation in the Dorsal Hippocampus in Response to Prolonged Morphine Withdrawal-Induced Depressive-Like Symptoms in Mice. *PloS one* 8:e66111. doi: 10.1371/journal.pone.0066111

Klein, R., and Kania, A. (2014). Ephrin signalling in the developing nervous system. *J Current Opinion in Neurobiology* 27, 16–24. doi: 10.1016/j.conb.2014.02.006

Kohen, R., Dobra, A., Tracy, J. H., and Haugen, E. (2014). Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness. *J Translational Psychiatry* 4, e366–e366.

Korgaonkar, M. S., Fornito, A., Williams, L. M., and Grieve, S. M. (2014). Abnormal structural networks characterize major depressive disorder: a connectome analysis. *Biological psychiatry* 76, 567–574. doi: 10.1016/j.biopsych.2014.02.018

Kroes, R. A., Panksepp, J., Burgdorf, J., Otto, N. J., and Moskal, J. R. (2006). Modeling depression: social dominance-submission gene expression patterns in rat neocortex. *Neuroscience* 137, 37–49. doi: 10.1016/j.neuroscience.2005.08.076

Kuo, D. C., Tran, M., Shah, A. A., and Matorin, A. (2015). Depression and the Suicidal Patient. *Emergency medicine clinics of North America* 33, 765–778. doi: 10.1016/j.emc.2015.07.005

Riglin, L., Eyre, O., Ak, Thapar, Stringaris, A., Leibenluft, E., Pine, D. S., et al. (2019). Identifying Novel Types of Irritability Using a Developmental Genetic Approach. *The American journal of psychiatry* 176, 635–642. doi: 10.1176/appi.ajp.2019.18101134

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

Lanz, T. A, Reinhart, V., Sheehan, M. J., Rizzo, S. J., Sukoff, B., Susan, E., et al. (2019). Postmortem transcriptional profiling reveals widespread increase in inflammation in schizophrenia: a comparison of prefrontal cortex, striatum, and hippocampus among matched tetrads of controls with subjects diagnosed with schizophrenia, bipolar or major depressive disorder*. 9, 1–13.

Lapidus, K. A., Gabbay, V., Mao, X., Johnson, A., Murrough, J. W., Mathew, S. J., et al. (2014). In vivo (1)H MRS study of potential associations between glutathione, oxidative stress and anhedonia in major depressive disorder. *Neuroscience letters* 569, 74–79. doi: 10.1016/j.neulet.2014.03.056

Li, B. J., Friston, M., Mody, Wang, H. N., Lu, H. B., and Hu, D. W. (2018). A brain network model for depression: From symptom understanding to disease intervention*. *CNS neuroscience therapeutics* 24, 1004–1019. doi: 10.1111/cns.12998

Li, H., Wei, D., Sun, J., Chen, Q., Zhang, Q., and Qiu, J. (2015). Brain structural alterations associated with young women with subthreshold depression. *Scientific reports* 5, 9707. doi: 10.1038/srep09707

Mamdani, F., Rollins, B., Morgan, L., Myers, R. M., Barchas, J. D., Schatzberg, A. F., et al. (2015). Variable telomere length across post-mortem human brain regions and specific reduction in the hippocampus of major depressive disorder. *Translational psychiatry* 5, e636. doi: 10.1038/tp.2015.134

Padilla-Coreano, N., Bolkan, S. S., Pierce, G. M., Blackman, D. R., Hardin, W. D., Garcia-Garcia, A. L., et al. (2016). Direct Ventral Hippocampal-Prefrontal Input Is Required for Anxiety-Related Neural Activity and Behavior. *Neuron* 89, 857–866. doi: 10.1016/j.neuron.2016.01.011

Park, H. J., and Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science* 342, 1238411. doi: 10.1126/science.1238411

Perlis, R. H., Brown, E., Baker, R. W., and Nierenberg, A. A. (2006). Clinical features of bipolar depression versus major depressive disorder in large multicenter trials. *The American journal of psychiatry* 163, 225–231. doi: 10.1176/appi.ajp.163.2.225

Rangaraju, S., Dammer, E. B., Raza, S. A., Rathakrishnan, P., Xiao, H., Gao, T., et al. (2018). Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. *Molecular neurodegeneration* 13, 24. doi: 10.1186/s13024-018-0254-8

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature protocols* 14, 482–517. doi: 10.1038/s41596-018-0103-9

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*. 43, e47. doi: 10.1093/nar/gkv007

Saldanha, A. J. (2004). Java Treeview–extensible visualization of microarray data. *Bioinformatics* 20, 3246–3248. doi: 10.1093/bioinformatics/bth349

Spellman, T., Rigotti, M., Ahmari, S. E., Fusi, S., Gogos, J. A., and Gordon, J. A. (2015). Hippocampal-prefrontal input supports spatial encoding in working memory. *Nature* 522, 309–314. doi: 10.1038/nature14445

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. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Tesseur, I., Nguyen, A., Chang, B., Li, L., Woodling, N. S., Wyss-Coray, T., et al. (2017). Deficiency in Neuronal TGF-β Signaling Leads to Nigrostriatal Degeneration and Activation of TGF-β Signaling Protects against MPTP Neurotoxicity in Mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 37, 4584–4592. doi: 10.1523/jneurosci.2952-16.2017

Torii, M., Hashimoto-Torii, K., Levitt, P., and Rakic, P. (2009). Integration of neuronal clones in the radial cortical columns by EphA and ephrin-A signalling. *J Nature* 461, 524–528. doi: 10.1038/nature08362

Wu, Y., Wei, Z., Li, Y., Wei, C., Li, Y., Cheng, P., et al. (2019). Perturbation of Ephrin Receptor Signaling and Glutamatergic Transmission in the Hypothalamus in Depression Using Proteomics Integrated With Metabolomics. 13, 1359. doi: 10.3389/fnins.2019.01359

Kajiyama, Y., Iijima, Y., Chiba, S., Furuta, M., Ninomiya, M., Izumi, A., et al. (2010). Prednisolone causes anxiety- and depression-like behaviors and altered expression of apoptotic genes in mice hippocampus. *Progress in neuro-psychopharmacology biological psychiatry* 34, 159–165. doi: 10.1016/j.pnpbp.2009.10.018

Zhang, G., Xu, S., Yuan, Z., and Shen, L. (2020). Weighted Gene Coexpression Network Analysis Identifies Specific Modules and Hub Genes Related to Major Depression. *Neuropsychiatric disease treatment* 16, 703–713. doi: 10.2147/ndt.S244452

Zhang, J.-C., Wei, Y., Youge, Q., Mayumi, N., Chao, D., Chun, Y., et al. (2017). Increased EphA4-ephexin1 signaling in the medial prefrontal cortex plays a role in depression-like phenotype*. 7, 1–14.

Zhang, K., Yang, C., Chang, L., Sakamoto, A., Suzuki, T., Fujita, Y., et al. (2020). Essential role of microglial transforming growth factor-β1 in antidepressant actions of (R)-ketamine and the novel antidepressant TGF-β1. *Translational psychiatry* 10, 32. doi: 10.1038/s41398-020-0733-x

Zhang, Z., Liu, R., Jin, R., Fan, Y., Li, T., Shuai, Y., et al. (2019). Integrating Clinical and Genetic Analysis of Perineural Invasion in Head and Neck Squamous Cell Carcinoma. *Frontiers in oncology* 9:434. doi: 10.3389/fonc.2019.00434

Zhen, L., Shao, T., Luria, V., Li, G., Li, Z., Xu, Y., et al. (2018). EphB2 Deficiency Induces Depression-Like Behaviors and Memory Impairment: Involvement of NMDA 2B Receptor Dependent Signaling. 9, 862. doi: 10.3389/fphar.2018.00862

Zheng, C., and Zhang, T. (2015). Synaptic plasticity-related neural oscillations on hippocampus-prefrontal cortex pathway in depression. *Neuroscience* 292, 170–180. doi: 10.1016/j.neuroscience.2015.01.071

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications* 10, 1523. doi: 10.1038/s41467-019-09234-6

Zhu, J. J., Qin, Y., Zhao, M., Van Aelst, L., and Malinow, R. (2002). Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell* 110, 443–455. doi: 10.1016/s0092-8674(02)00897-8

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