Original Research

Co-expression modules construction by WGCNA and identify potential hub genes and regulation pathways of postpartum depression

Zhifang Deng[1,†], Wei Cai[2,†], Jue Liu[1], Aiping Deng[1], Yuan Yang[3], Jie Tu[1], Cheng Yuan[4], Han Xiao[3,*], Wenqi Gao[3,*]

[1]*Department of Pharmacy, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, 430030 Wuhan, Hubei, China,* [2]*Department of Hematology, Wuhan Children's Hospital (Wuhan Maternal and Child Healthcare Hospital), Tongji Medical College, Huazhong University and Technology, 430015 Wuhan, Hubei, China,* [3]*Institute of Maternal and Child Health, Wuhan Children's Hospital (Wuhan Maternal and Child Healthcare Hospital), Tongji Medical College, Huazhong University and Technology, 430015 Wuhan, Hubei, China,* [4]*Department of Radiation and Medical Oncology, Zhongnan Hospital, Wuhan University, 430071 Wuhan, Hubei, China*

TABLE OF CONTENTS

1. Abstract

Purpose: The purpose of our present study was to, for the first time, identify key genes associated with postpartum depression (PPD) and discovery the potential molecular mechanisms of this condition. **Methods**: First, microarray expression profiles GSE45603 dataset were acquired from the Gene Expression Omnibus (GEO) in National Center for Biotechnology Information (NCBI). The weighted gene co-expression network analysis (WGCNA) was performed to identify the top three modules from differentially expressed genes (DEGs). Furthermore, cross-validated differential gene expression analysis of the top three modules and DEGs was used to identify the hub genes. Gene set enrichment analysis (GSEA) was conducted to identify the potential functions of the hub genes. We conducted a Receiver Operator Characteristic (ROC) curve to verify the diagnostic efficiencies of the hub genes. Lastly, GSE44132 dataset was used to search the association between the methylation profiles of the hub genes and susceptibility to PPD. **Results**: Altogether, 8979 genes were identified as DEGs for WGCNA analysis. The turquoise, yellow, and green functional modules were the most significant modules related to PPD development after WGCNA analysis. The enrichment analysis results of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway demonstrated that hub genes in the three modules were mainly enriched in the neurotrophin signaling pathway, chemokine signaling pathway, Fcγ receptor-mediated phagocytosis, and Mitogen-activated protein kinase (MAPK) signaling pathway. Eight genes (HNRNPA2B1, IL10, RAD51, UBA52, NHP2, RPL13A, FBL, SPI1) were identified as "real" hub genes from cross-validation data of the three modules and DEGs, and possessed diagnostic value in PPD. The GSEA suggested that "OLFACTORY_TRANSDUCTION", "BUTANOATE_METABOLISM", "MELANOMA", "AMINOACYL_TRNA_BIOSYNTHESIS", and "LYSINE_DEGRADATION" were all crucial in the development of PPD. Highly significant differentially methylated

positions in the three genes (HNRNPA2B1, RPL13A and UBA52) were identified in the GSE44132. **Conclusion**: Using WGCNA analysis of GEO data, our present study, for the first time, may contribute to elucidate the pathophysiology of PPD and provide potential diagnostic biomarkers and therapeutic targets for postpartum depression.

2. Introduction

Postpartum depression (PPD) is one of the most prominent mood disorders affecting 15% of parturient women [1]. Its clinical symptoms typically occur within two weeks after delivery, with patients reporting persistent and severe moodiness, including negative emotions (e.g., anxiety, sadness, hopelessness and/or worthlessness), low energy, and social withdrawal [2]. PPD may result in both short-term and long-term negative effects on neonatal and infant development, cognition, emotion, and behavior. Previous studies have shown that maternal depression is a risk factor for higher rates of premature and low-birth-weight babies, infant malnutrition and stunting, and infant diarrhoeal, which can lead to mothers' reluctance to care for their children and impair normal mother-child relationships. And it's becoming a major public health problem [3].

PPD is the result of the combined influence of multiple factors. Previous studies have reported hypotheses regarding the physiopathologic mechanisms of PPD, such as the abnormal changes of hormones, neurotransmitters, inflammatory factors [4], among others. The molecular pathways implicated in the PPD was involved in genetic inheritance [5], HPA axis [6], serotonin transporter [7], and GABA receptor [8]. However, there is no exact evidence to support any of the above conclusions, and the physiopathologic mechanisms of PPD remain largely obscure.

Despite the severity of PPD on adverse infant and maternal outcomes, early diagnosis would help establish mechanisms for the prevention and early cure of PPD. Therefore, it is necessary to identify reliable diagnostic biomarkers of PPD. In practical work, it is difficult to acquire brain tissue for clinical testing. Peripheral blood mononuclear cells (PBMCs) share mRNA expression patterns in brain tissues [9]. In both PBMCs and the prefrontal cortex, the expression levels of many genes and biological processes are similar [10]. On the one hand, mounting evidence has indicated that gene expression changes in PBMCs may intervene molecular alterations in the region of brain [11]. On the other hand, gene expression of peripheral lymphocytes may be affected by the central nervous system (CNS) via neurotransmitters, cytokines, and hormones [12]. Therefore, PBMCs can serve as an effective tissue to search the gene expression in PPD.

With the development of high-throughput expression microarray technology, gene expression profiling has become a powerful tool for in-depth study of the pathogenesis and diagnostic biomarkers of mental disorders at the genomic level [13]. For example, microarrays have been used to detect differences in gene expression characteristics between healthy controls and PPD patients [14]. Many microarray-based studies have focused only on screening for differentially expressed genes (DEGs) without identifying their links. Genes with similar expression patterns may be functionally related. Weighted gene co-expression Network analysis (WGCNA) algorithm can divide genes into different modules according to the similarity of gene co-expression in multiple samples. This produces a set of genes that function similarly, allowing selection of disease-related biomarkers and pathways. In the present study, we, for the first time, used WGCNA to construct a co-expression network of genes and determined key modules and hub genes associated with PPD.

3. Materials and methods

3.1 Study design

The design of this study was exhibited in **Supplementary Fig. 1**. We downloaded the mRNA microarray expression profile dataset GSE45603 from GEO, a free and publicly available database. This dataset was used to screen DEGs, then construct co-expression networks using WGCNA, and finally to identify hub genes associated with PPD.

There were 210 samples from PBMCs in the GSE45603 dataset, from which 21 samples were finally selected, containing 5 healthy control samples (GSM1110484, GSM1110493, GSM1110498, GSM1110515, GSM1110539) and 16 PPD samples (GSM1110374, GSM1110378, GSM1110380, GSM1110400, GSM1110403, GSM1110409, GSM1110412, GSM1110424, GSM1110433, GSM1110447, GSM1110461, GSM1110477, GSM1110505, GSM1110506, GSM1110518, GSM1110547), the samples of preconception, 1st, 2nd, and 3rd trimester of pregnancy and postpartum were not included in present study.

3.2 Screen differential expression genes

GEO2R was used to analyze DEGs between PPD and healthy control samples, $p < 0.05$ was defined as DEGs. GEO Database interactive online analysis tool for further analysis (R, version 3.5.2, https://www.r-project.org). DEGs were defined as $\log_2 FC < -0.5$ (down-regulated gene) and $\log_2 FC > 0.5$ (up-regulated gene), and the difference was statistically significant if the adjusted p value < 0.05.

3.3 Functional and pathway enrichment analysis

Gene annotation with functions involving cellular components (CC), molecular functions (MF), and biological pathway (BP) were analyzed by Gene ontology (GO) [15, 16]. The key MF, BP, CC and pathways of statistically

significant DEGs were visualized ($p < 0.05$). The function of KEGG was to address genomes and biological pathways related to diseases ($p < 0.01$). To improve the accuracy of the analysis and identify the potential function of the hub gene, GSEA analysis was carried out (GSEA version 4.0.3, http://software.broadinstitute.org/gsea/index.jsp). p-value < 0.05 was used as truncation standard. The top 15 pathways of yellow module gene in KEGG were selected.

3.4 PPI network and module analysis

The results from the STRING (version 11.5, https://www.string-db.org/) [17] were analyzed and structured using Cytoscape software (Cytoscape version 3.9.0, https://cytoscape.org/). The MCODE plugin was employed to identify modules of the PPI network. Two modules were selected (score >8). We found there were 8979 DEGs in GSE45603 after analyse by GEO2R. Then, we set several "strict" threshold (log$_2$FC $<$–0.5 and log$_2$FC >0.5 were used as a cutoff, with the p value < 0.05) for identify "real" DEGs which were meet our requirements. 673 "real" DEGs were imported into the STRING to obtain their interaction, which were imported into Cytoscape software. CytoHubba was used to calculate gene value [18], was used to screen hub genes based on the degree algorithm, and the top 30 genes were defined as hub genes. We entered the 30 hub genes into NetworkAnalyst, a visual analytic platform for exhibition of the PPI networks [19]. In PPI network, genes with combination score ≥ 0.9 and connectivity ≥ 10 were also defined as hub genes. The top 30 hub genes with the highest ranking was presented in Table 1.

3.5 Construction of weighted gene co-expression networks

There were 8979 DEGs after analysis by GEO2R. WGCNA (Version 1.6.0, https://git.bioconductor.org/packages/GmicR) was performed using 8979 DEGs [20]. The adjacency matrix was transformed into topological overlap matrix (TOM) by one-step method, and the minimum value was 30. Hierarchical clustering was used to generate hierarchical clustering tree [21, 22]. The co-expressed gene modules were generated by hierarchical clustering tree with different colors, and the module structure was displayed by topological overlapping matrix.

3.6 Identification of clinically significant modules and hub genes in key modules

Correlations between modular characteristic genes (MEs) and clinical features were calculated to identify relevant modules. Gene significance (GS) was measured as the log10 transformation of the p-value (GS = lg p) of the linear regression between gene expression and clinical information. Modular significance (MS) was the average GS of all genes in a module. Overall, of all the selected modules, the module with the absolute first rank in MS was considered to be relevant for clinical features. In our study, the genes in key modules were imported into Cytoscape and screened

Table 1. The expression analysis of the top 30 hub genes with the highest ranking.

Gene symbol	log$_2$FC	p-value	Score
UBA52	0.56	1.21×10^{-2}	72
HNRNPA2B1	–0.847	6.80×10^{-4}	43
IL10	–0.783	8.90×10^{-3}	39
NHP2	0.581	2.52×10^{-3}	35
CAT	–1.04	1.77×10^{-3}	34
CDC5L	–0.517	3.39×10^{-2}	34
PTGES3	–0.748	3.87×10^{-3}	34
RPL23	–0.847	4.84×10^{-2}	34
SIRT1	–0.512	3.18×10^{-2}	34
RPS3A	–0.725	2.28×10^{-2}	33
CCT3	0.503	3.58×10^{-3}	32
RPS28	–0.787	2.82×10^{-2}	32
ALDH18A1	0.797	7.78×10^{-4}	31
ATM	–0.821	4.56×10^{-3}	31
PABPC1	–1.2	7.00×10^{-3}	31
ACTR3	–0.907	8.00×10^{-3}	30
RPL13A	0.538	2.57×10^{-3}	30
RPL37A	–0.511	4.49×10^{-2}	30
FBL	0.741	2.34×10^{-3}	29
RPL10A	0.538	5.93×10^{-3}	29
RPS15A	–0.607	1.36×10^{-2}	29
SPI1	–0.73	2.77×10^{-2}	29
CCL5	0.536	2.14×10^{-2}	28
RPS29	0.515	3.66×10^{-2}	28
CCR7	0.582	2.04×10^{-2}	27
HSPE1	–1.13	3.65×10^{-2}	27
KPNB1	–0.516	2.26×10^{-2}	27
RPS27	–0.551	2.97×10^{-2}	27
RAD51	–0.501	1.97×10^{-2}	26
RPS26	1.28	2.79×10^{-3}	26

out according to degree. Key modules and two genes were overlapped using Venn diagrams.

3.7 Distributions of Hub Genes and Methylation Analysis

The distribution of all DEGs in GSE45603 was identified, and the functional similarity between proteins was evaluated by GOSemSim package (version 2.18.1, http://bioconductor.org/packages/release/bioc/html/GOSemSim.html) [23]. Then, the methylation of the hub gene was detected. Genome-wide models of prenatal DNA methylation analysis data (GSE44132) were also downloaded. This dataset includes 55 samples. Among these, 25 samples without depression (GSM1079474, GSM1079478, GSM1079480, GSM1079481, GSM1079482, GSM1079483, GSM1079484, GSM1079485, GSM1079486, GSM1079488, GSM1079491, GSM1079492, GSM1079493, GSM1079494, GSM1079495, GSM1079497, GSM1079498, GSM1079499, GSM1079501, GSM1079503, GSM1079504, GSM1079505, GSM1079507, GSM1079510, GSM1079528) and 11 samples with
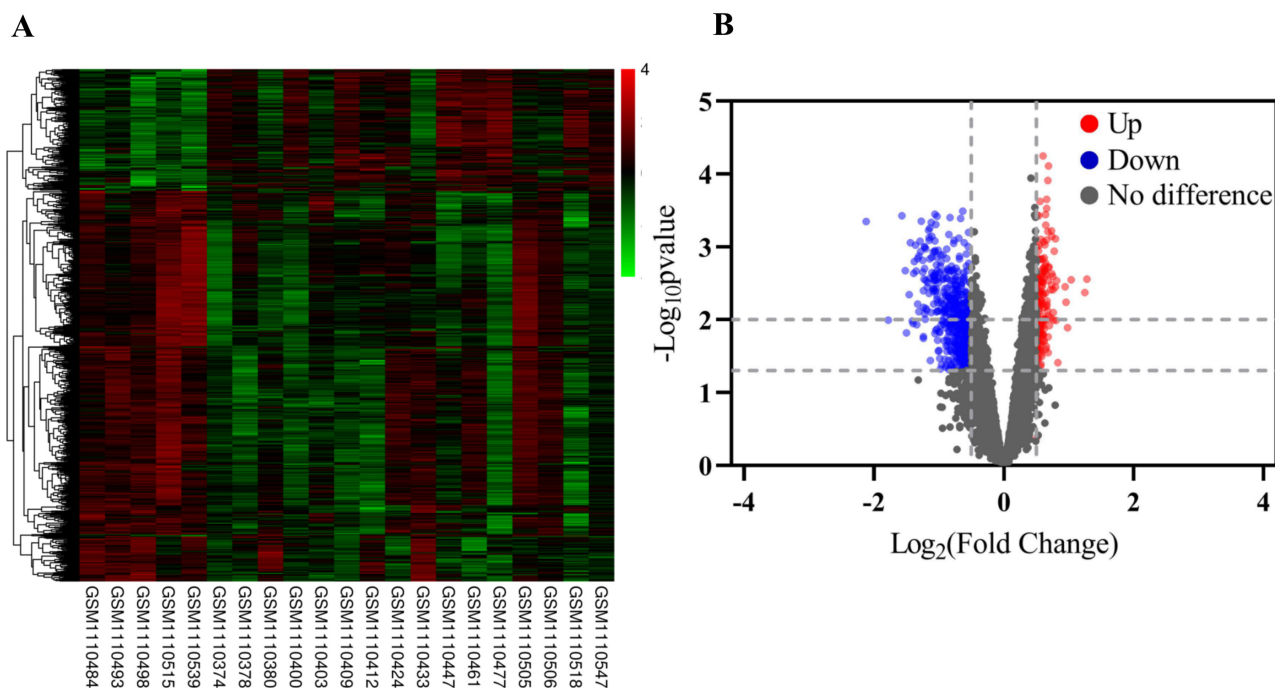
A

B

Fig. 1. The DEGs from GSE43056. (A) Heat map of differentially expressed genes. (B) Volcano plot of gene expression. Red means upregulated DEGs; green means downregulated DEGs; black means no difference.

PPD were selected (GSM1079475, GSM1079476, GSM1079477, GSM1079479, GSM1079487, GSM1079489, GSM1079490, GSM1079496, GSM1079500, GSM1079502, GSM1079506). The raw data were analyzed using "minfi" package (version 1.38.0, http://www.bioconductor.org/packages/release/bioc/html/minfi.html) in R and were converted into a score, referred to as a beta value.

3.8 Screening of diagnostic biomarkers

Receiver operating characteristic (ROC) analysis combines sensitivity and specificity to comprehensively evaluate diagnostic accuracy or discriminant effect. The "pROC" package (version 1.68.1, http://www.bioconductor.org/packages/release/bioc/html/ROC.html) [24] was applied to evaluate the diagnostic value of hub genes and to screen out the diagnostic biomarkers of PPD.

3.9 Statistical Analysis

All statistical analysis were measured as mean $\pm$ standard deviation. R software (version 3.5.2, https://www.r-project.org/) was utilized to measure the data.

4. Results

4.1 Identification of DEGs

There was a total of 21 PBMCs samples in this study, which included 5 healthy control and 16 PPD samples. There were 8979 DEGs. Then, the cutoff criteria were adjusted p value < 0.05, log FC >0.5, or log FC <-0.5. A total of 673 DEGs, including 163 upregulated genes and 510 downregulated genes, were explored after analyzing GSE45603. The expression level is displayed in Fig. 1A,B.

4.2 GO Function and KEGG pathway enrichment analysis

GO function annotation and KEGG pathway enrichment analysis were conducted to obtain more comprehensive knowledge of the selected DEGs. As for BP, the upregulated DEGs were mainly implicated in the regulation of translation, rRNA processing, regulation of immune response, and SRP-dependent cotranslational protein targeting the membrane. Mitochondrion, T cell receptor complex, ribosome, mitochondrial inner membrane, and cytosolic large ribosomal subunits were involved in the CC. Changes in MF were mainly enriched in the structural constituents of ribosomes, methyltransferase activity, poly (A) RNA binding, protein binding, and mRNA binding (**Supplementary Fig. 2A**). The downregulated DEGs were mainly responsible for reciprocal meiotic recombination, antigen processing, and presentation of exogenous peptide antigen via MHC class I, TAP-independent, positive regulation of long-term synaptic potentiation, positive regulation of macroautophagy, and negative regulation of gene expression in BP. Cytoplasmic, cytosolic, phagocytic vesicle membrane, NADPH oxidase complex, and focal adhesion in CC. Protein binding, four-way junction DNA binding, superoxide-generating NADPH oxidase activator activity, superoxide-generating NADPH oxidase activity, and enzyme binding in MF (**Supplementary Fig. 2B**).
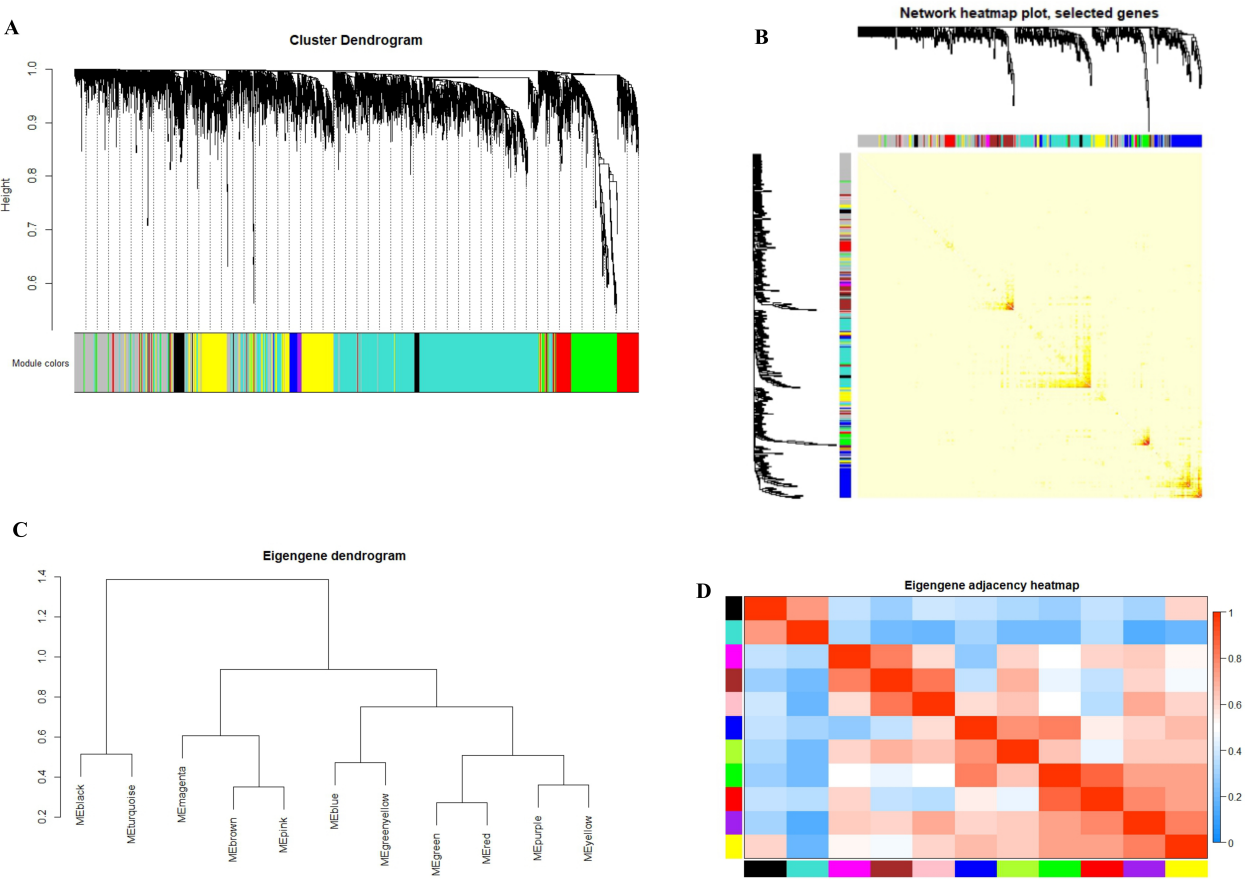
Fig. 2. Construction of co-expression modules by the WGCNA. (A) The cluster dendrogram of genes in GSE45603. Branches of the cluster dendrogram of the most connected genes gave rise to eleven gene co-expression modules. Every gene represents a line in the hierarchical cluster. The distance between two genes is shown as the height on the y-axis. (B) Interaction relationship analysis of 400 selected co-expression genes. Different colors of the horizontal axis and vertical axis represent different modules. The brightness of yellow in the middle represents the degree of connectivity of different modules. There was no significant difference in interactions among different modules, indicating a high degree of independence among these modules. (C) Hierarchical clustering of module hub genes that summarize the modules yielded in the clustering analysis. (D) Heatmap plot of the adjacencies in the hub gene network.
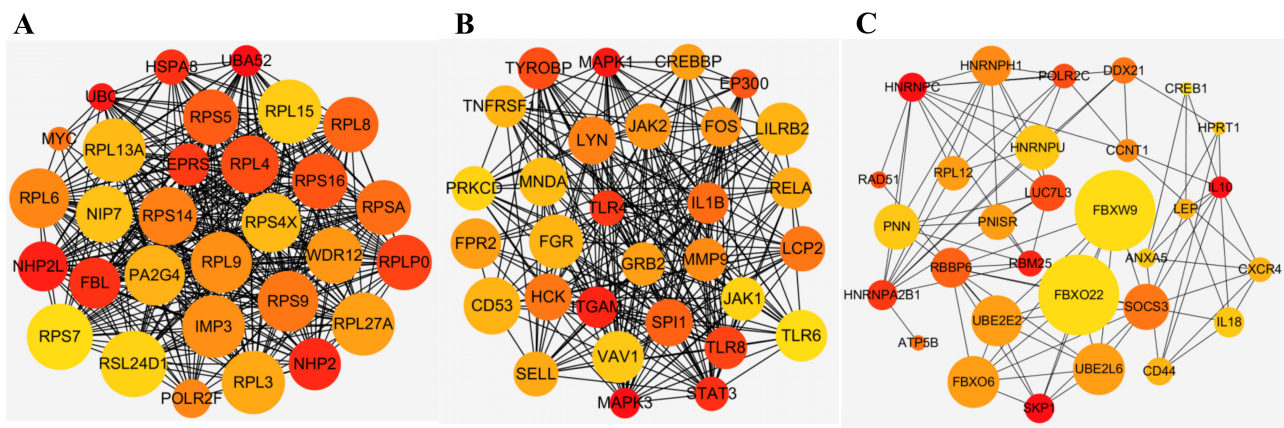

Fig. 3. Top 30 genes in (A) turquoise module, (B) yellow module, and (C) green module. The size of the node represents the clustering coefficient. The color of the node indicates its degree, the bigger the node, the greater the number of connections it has, blue, yellow and orange color represent big, middle and small degrees respectively. (A) Hub genes in turquoise module. (B) Hub genes in yellow module. (C) Hub genes in green module.
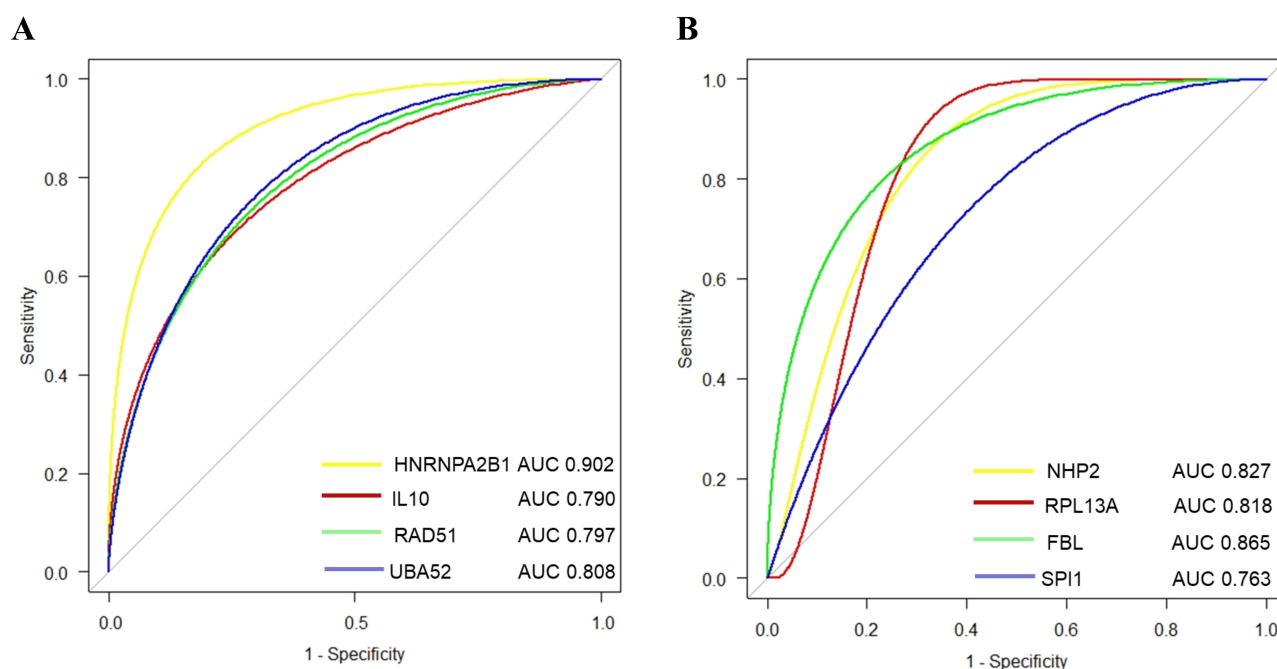
A

B

Fig. 4. Validation of diagnostic value for the hub genes in PPD. (A,B) ROC curve of the hub genes for PPD diagnosis.

The enrichment analysis results of KEGG pathway demonstrated that upregulated DEGs were mainly involved in ribosome enrichment, biosynthesis of amino acids, oxidative phosphorylation, ribosome biogenesis in eukaryotes, and hematopoietic cell lineage (**Supplementary Fig. 2C**). The downregulated DEGs were mainly enriched in phagosomes, osteoclast differentiation, antigen processing and presentation, glucagon signaling pathway, and Forkhead box O3 (FoxO) signaling pathway (**Supplementary Fig. 2D**).

4.3 Module Screening from the PPI network

The PPI network consisted of 571 nodes and 2213 edges, and module A (score = 17.263, including 20 nodes and 164 edges; **Supplementary Fig. 3A**), and B (score = 8.5, including 37 nodes and 135 edges; **Supplementary Fig. 3B**) with score >8 was identified from the PPI network. To establish DEGs from interactive PPI networks, these genes were imported into the STRING tool, the data from STRING was imported into Cytoscape (**Supplementary Fig. 4A**), and 30 hub genes were identified from the PPI network by cytoHubba (**Supplementary Fig. 4B**) and visualized by NetworkAnalyzer (**Supplementary Fig. 4C**).

4.4 Weighted co-expression network construction and key modules selection

Firstly, we checked the data quality in GSE45603. All samples were taken for subsequent analysis (**Supplementary Fig. 5**). DEGs containing 8979 genes was analyzed using WGCNA, and modules containing highly related genes were identified. Based on

the approximate scale-free topology criterion, soft threshold power was 14 (scale-free R2 = 0.86) was optimized (**Supplementary Fig. 6A** and **Supplementary Fig. 6B**). There were 12 modules (Fig. 2A), 1921 genes in turquoise module, 704 genes in yellow module, and 447 genes in green module. The 2736 genes that could not be included in any module were placed into the gray module and identified as non-co-expressed.

4.5 Correlation between modules and identification of key modules

The 400 genes were selected to draw a network heat map (Fig. 2B). Our results so far showed that the turquoise, yellow, green, and blue modules were independently verified and showed a high degree of independence between modules and the relative independence of the genes expressed in each module. The 11 modules were divided into two main clusters: one consists of two modules (turquoise, black), while the other consists of nine modules (magenta, brown, pink, blue, greenish-yellow, green, red, purple and yellow modules; Fig. 2C). Finally, heat maps were drawn according to the connectivity of characteristic genes to visualize the results (Fig. 2D).

4.6 Identification and distribution of hub genes

The turquoise, yellow, and green modules were visualized by Cytoscape, and the top 30 genes were screened out by sorting node degree candidate genes for further analysis (Fig. 3A–C). Cross-validation of the data from these three modules and DEGs revealed 3 genes (HNRNPA2B1, IL10, and RAD51) in both DEGs and the green module, 4 genes (UBA52, NHP2, RPL13A, FBL) in both DEGs
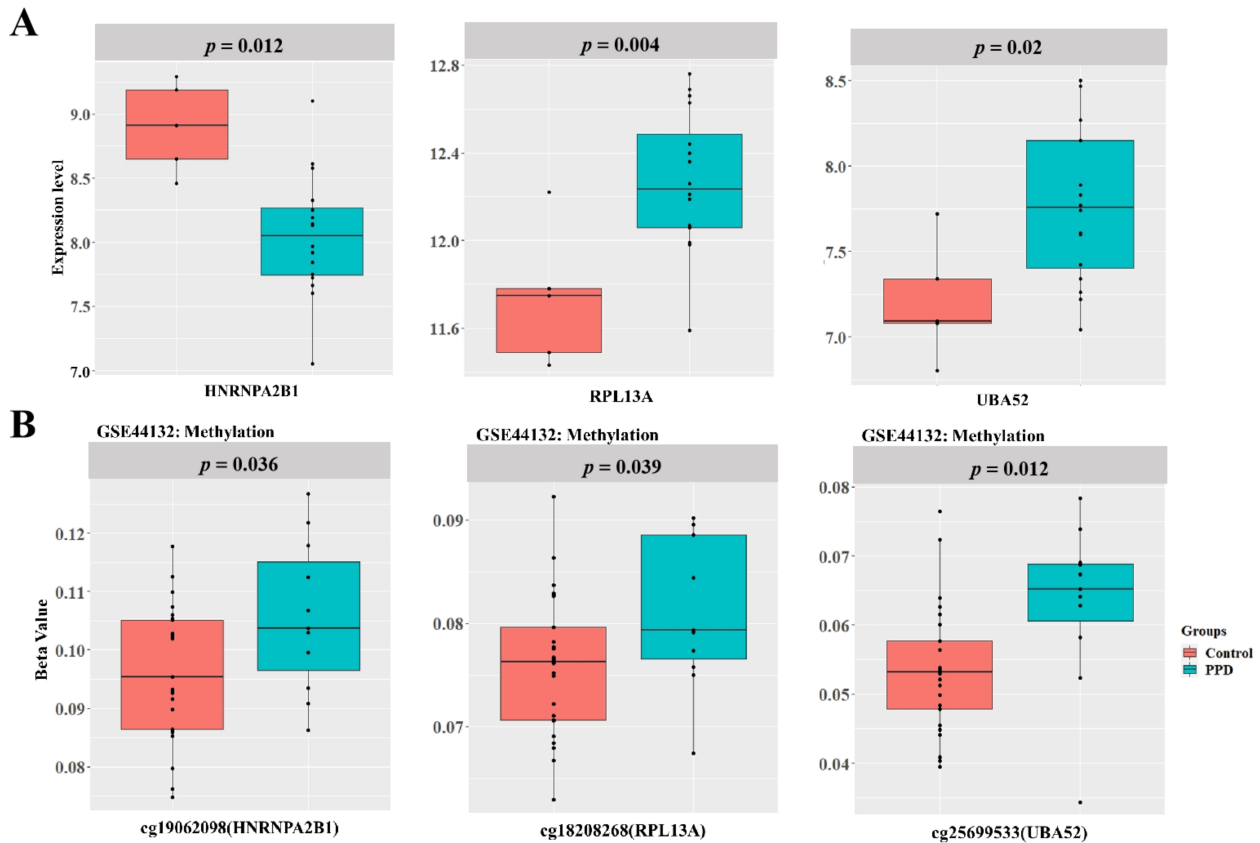
Fig. 5. The expression and methylation levels of HNRNPA2B1, RPL13A and UBA52. (A) The expression levels of the three genes (HNRNPA2B1, RPL13A and UBA52) were correlated with PPD (based on microarray data of GSE44132). (B) The methylation levels of HNRNPA2B1, RPL13A and UBA52 (based on microarray data of GSE44132).

and the turquoise module, and 1 gene (SPI1) in both DEGs and the yellow module (**Supplementary Fig. 7A**). These 8 genes were observed based on the average functional similarity (**Supplementary Fig. 7B**).

4.7 Screening of diagnostic biomarkers

The diagnostic efficiencies of 8 hub genes were evaluated by ROC curve analysis (Fig. 4). The AUC of ROC curves were 0.902, 0.790, 0.797, and 0.808 for HNRNPA2B1, interleukin 10 (IL-10), RAD51, and ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52), respectively (Fig. 4A). The AUC of ROC curves were 0.827, 0.818, 0.865, and 0.763 for NHP2, ribosomal protein L13a (RPL13A), fibrillarin (FBL), and Spi-1 proto-oncogene (SPl1), respectively (Fig. 4B). AUC values within the range of 0.700–0.900 were considered to have moderate accuracy. Thus, HNRNPA2B1, IL10, RAD51, UBA52, NHP2, RPL13A, FBL, and SPI1 were the hub genes with higher diagnostic value.

4.8 Expression of the hub genes after methylation analysis

Using GSE44132 dataset, we investigated the relationship between methylation levels of 8 genes and PPD susceptibility, and explored the possible mechanism of PPD occurrence. Data from 36 whole blood samples (11 PPD samples and 25 healthy controls) passed quality control indicators and were analyzed. In the GSE45603 dataset, HNRNPA2B1 was down-regulated, and RPL13A and UBA52 were significantly up-regulated in the PPD group, compared with the control group (Fig. 5A). It's worth noting that, highly differentiated methylation sites were found in HNRNPA2B1, RPL13A, and UBA52 genes: cg19062098 (p-value = 0.036), cg18208268 (p-value = 0.039), and cg25699533 (p = 0.012) (Fig. 5B).

4.9 GSEA analysis of hub genes

Because HNRNPA2B1, RPL13A, and UBA52 had highly significantly differentiated methylated positions, we selected the three genes for follow-up studies. To identify the potential functions of these hub genes, GSEA was conducted to identify enriched biological processes in the samples. Geneset enrichment analysis (GSEA) for gene sets related to HNRNPA2B1, RPL13A and UBA52
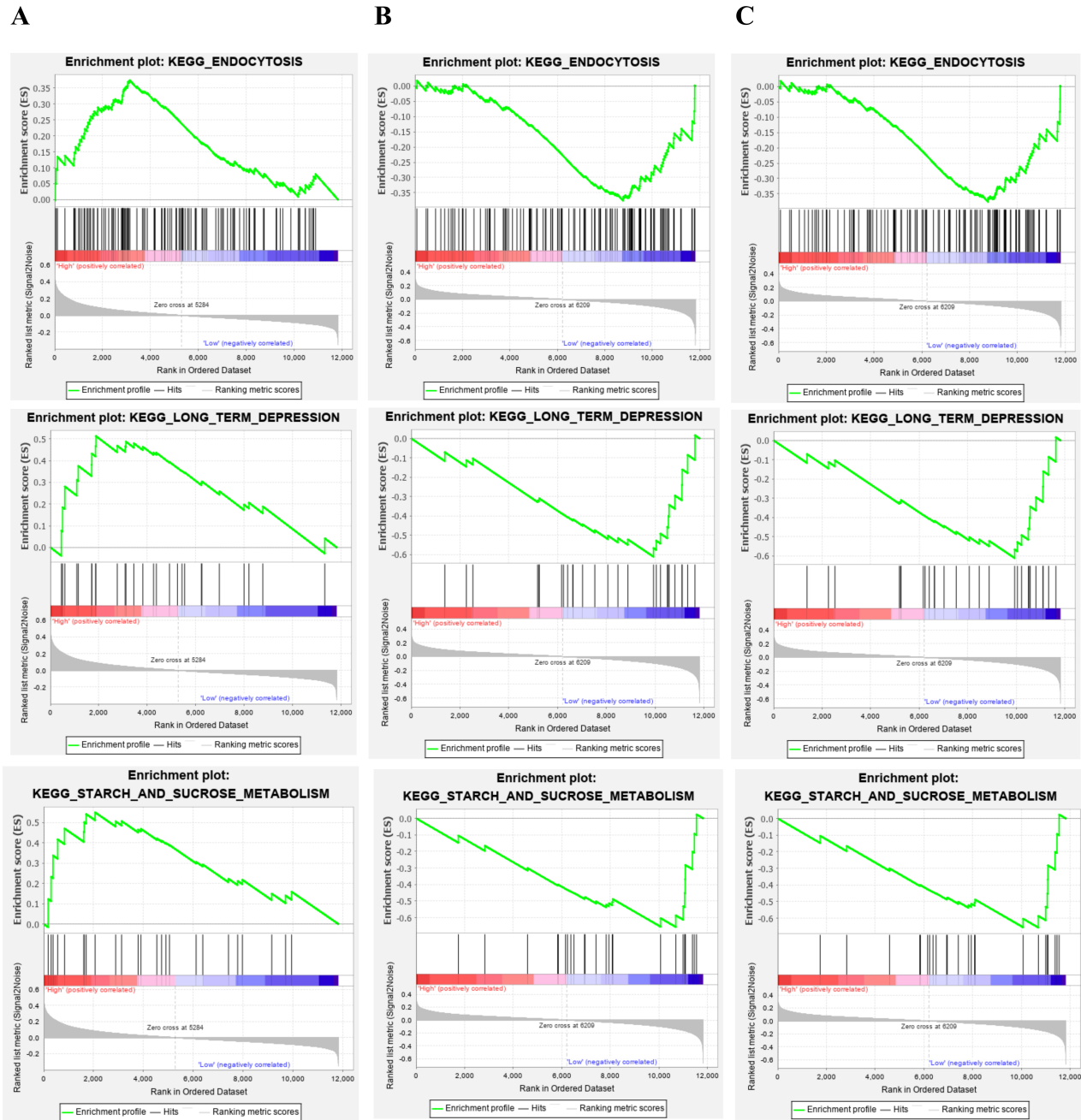
Fig. 6. **Gene set enrichment analysis (GSEA) for gene sets related to HNRNPA2B1, RPL13A and UBA52 expressions.** (A) HN-RNPA2B1 was lowly expressed in the gene set of "OLFACTORY_TRANSDUCTION", "BUTANOATE_METABOLISM", and "MELANOMA". (B) RPL13A was highly expressed respectively in the gene sets of "AMINOACYL_TRNA_BIOSYNTHESIS", "LYSINE_DEGRADATION" and "BUTANOATE_METABOLISM". (C) UBA52 was highly expressed in the gene sets of "AMINOACYL_TRNA_BIOSYNTHESIS", "LY-SINE_DEGRADATION" and "BUTANOATE_METABOLISM".

expressions. Our results found that, compared with healthy control, HNRNPA2B1 was down-regulated and RPL13A and UBA52 were up-regulated in the PPD samples, respectively (Fig. 5A), thus we considered that lowly expression of HNRNPA2B1 and highly expressions of RPL13A and UBA52 were involved in which pathways. Results suggested that HNRNPA2B1 was lowly expressed in

the gene set of "OLFACTORY_TRANSDUCTION", "BUTANOATE_METABOLISM", and "MELANOMA" (Fig. 6A). RPL13A and UBA52 were highly expressed in the gene sets of "AMINOACYL_TRNA_BIOSYNTHESIS", "LYSINE_DEGRADATION" and "BU-TANOATE_METABOLISM" (Fig. 6B,C).

5. Discussion

PPD is a major public health concern, affecting 10 to 15 percent of mothers worldwide [25]. Investigations are being performed to develop a non-invasive and quantitative clinical test to support PPD diagnosis, nevertheless, no specific or sensitive biomarkers in whole blood are currently available for the diagnosis of PPD [26, 27]. Thus, we performed several bioinformatic methods to identify the exact pathological mechanisms and effective diagnostic biomarkers of PPD. In the first part of our study, we determined the hub genes and pathways associated with PPD within whole blood samples. A total of strictly screened 673 DEGs (|LogFC| >0.5 was used as a cutoff, with the p value < 0.05 regarded as statistically significant, 163 upregulated genes and 510 downregulated genes) and 30 hub genes were explored from the GSE45603 dataset. Next, we applied a new approach, WGCNA, to validate gene co-expression modules related to the molecular mechanisms underlying PPD. GEO2R identified 8979 genes which were analyzed to construct a co-expression network and generate a highly correlated co-expression gene group. There were three highly correlated modules (turquoise, yellow, and green modules) obtained from the WGCNA analysis. We regarded the top three hub genes (HnRNPA2B1, FBL, and SPI1), both in the PPI network and co-expression network, as "real" hub genes, which may function as important regulators in the pathogenesis of PPD.

Recently, it has been widely accepted that endocrine [28], immune response [29], and genetic and epigenetic factors [30], are all involved in the pathogenesis of PPD. Our study found that genes in the three modules identified by WGCNA (turquoise, green, and yellow modules) were mainly enriched in innate immune response, signal transduction, intracellular signal transduction, inflammatory response, and viral processes. The green module was enriched in the regulation of transcription (DNA-templated), gene expression, response to ionizing radiation, and RNA processing.

According to the cross-validation of data from the three modules and DEGs, eight genes acted as high degree genes. Among these eight genes, we were interested in the HnRNPA2B1, Fibrillarin (FBL), and Spi-1 proto-oncogene (SPI1). HnRNPA2B1, an RNA-binding protein and one member of the heterogeneous nuclear ribonucleoprotein family, initiates and amplifies the innate immune response. HnRNPA2B1 dimerization is required for nucleocytoplasmic translocation and initiation of IFN-α/β expression [31]. Furthermore, hnRNPA2B1 is also involved in the pathogenesis of several devastating neurodegenerative diseases. Mutations in prion-like domains of hnRNPA2B1 play a causal role in multisystem proteinopathy (MSP) and may be involved in the pathogenesis of amyotrophic lateral sclerosis (ALS) [32]. Fibrillarin (FBL), a molecular marker of transcriptionally active RNA polymerase I, catalyzes the 2'-O-methylation of ribosomal RNAs. FBL plays an important role in ribosome biogenesis, particularly in the methylation of ribosomal RNAs and rDNA histones [33, 34]. Recent work has indicated that FBL may be an important component of ribosome size, life duration in multicellular organisms, and ribosomal RNA production, all of which are correlated with age in healthy humans [35, 36]. Spi-1 proto-oncogene (SPI1), the E26 transformation-specific (ETS) family transcription factor PU.1, serves as a master regulator of myeloid and lymphoid development and is primarily expressed in monocytes/macrophages, neutrophils, mast cells, B cells, and early erythroblasts [37, 38]. The function of SPI1 is closely related to microglial viability and phagocytic capability. The targets of SPI1 demonstrate a significant relationship with the pathways defined as "Endocytosis", "Fcγ receptor-mediated phagocytosis", "Chemokine signaling pathway", and "MAPK signaling pathway" [38, 39]. UBA52, a ubiquitin-ribosomal fusion gene and located in chromosome 19, is a major source of ubiquitin protein for covalent modification of proteinaceous substrates recycled by ubiquitin-proteasome system (UPS) [40]. UBA 52 encodes a fusion protein comprising ubiquitin at the N-terminus and RPL40 at the C-terminus [41]. Upon translation, ubiquitin and RPL40 are immediately cleaved from the translated product [42]. RAD51 has the function of discovering and invading homologous DNA sequences, and can be repaired accurately and timely [43]. NHP2 gene is a member of the H/ACA snoRNPs (small nucleolar ribonucleoproteins) gene family. snoRNPs are involved in various aspects of rRNA processing and modification [44]. RPL13A encodes a member of the ribosomal protein L13P family and is a component of the 60S subunit. The protein encoded as a component of the IFN-γ activated translational inhibitor (gait) complex also plays a role in inhibiting inflammatory genes [45]. However, the functions of RAD51, NHP2, and RPL13A in PPD did not study in detail. Interestingly, the pathway enrichment analysis indicated that genes in the three modules were enriched in "neurotrophin signaling pathway" and "chemokine signaling pathway", which partly coincided with SPI1 targets. The neurotrophin signaling pathway includes several neurotrophic factors, largely a family of secreted proteins that benefit in the growth, survival, and differentiation of neurons. Due to its beneficial and therapeutic effects on the neuronal physiology, researchers have intensively studied neurotrophic factors for decades [43]. Recently, neurotrophic factors have been considered as having significant links with the pathophysiology of neurological and neuropsychiatric disorders. Emerging evidence has demonstrated that brain-derived neurotrophic factor (BDNF), a well-known neurotrophic factor, is involved in PPD and its treatments. Clinical studies have observed a link between PPD symptoms and single nucleotide polymorphisms (SNPs) in BDNF [44]. Postpartum female mice induced by chronic unpredictable stress exhibited low

expression of BDNF mRNA in the mPFC region [45]. In the serum of patients with PPD at admission and during development, BDNF levels were lower than healthy individuals. Fluoxetine, a classic treatment for women with PPD, may mediate its antidepressant effect in PPD by upregulating BDNF expression [46]. Second, to protect the fetus against pathogens and receive the semi-allogenic markers as the fetus, the immune system of the mother would go through a tremendous adaptation during pregnancy [47]. Inflammatory cytokine levels may obviously increase in the periphery during cervical ripening [48]. Furthermore, since the mother's immune system recovers to the previous non-pregnant state after delivery, subsequently, the change in inflammatory cytokine levels may affect mood [49]. Brewster *et al.* [50] reported that there was a close relationship between chemokines and PPD, which is manifested by increased levels of pro-inflammatory cytokines, such as IL-18 and C-X-C motif chemokine 1 (CXCL1). PPD was positively correlated with multiple genes in the immune response of an immobilization stress-induced mouse PPD model. The neurotrophin signaling pathway and chemokine signaling pathway may be deeply involved in the mechanisms of PPD; however, it remains a major challenge that required continued exploration.

There were two previous studies to find biomarkers of PPD based on GSE45603 dataset. Lauren *et al.* [51] reported that the objective of this study was to independently replicate their previously published prediction model of postpartum depression and women without a history of psychiatric disorders, and to further investigate the DNA methylation status of postpartum depression biomarkers associated with changes in pregnancy hormone levels and the timing of major hormonal changes. Maria *et al.* [52] focused on whether abnormal patterns of circulating levels of cytokines and chemokines may offer a suitable biomarker for disease development and/or therapeutic response. The objective of the present study and the previous two study were both searching for biomarkers of PPD. However, there were several differences among these studies. Our present study only observed the postpartum mental state and healthy controls. Lauren *et al.* collected different gestation process samples and used their own model to identify biomarker, their key point was changes in hormone levels. The emphasis of Maria' study was chemokines levels. Our present study did not classify genes, we put all DEGs into WGCNA and obtained hub genes, and the functions of genes were involved in various aspects of cell.

There were several limitations to our present study. First, in order to comprehensively determine dysfunctions related to PPD, both brain tissue and blood samples would need to be integrated. However, our present study did not perform an analysis of brain tissue. Second, only one microarray was included in the computerized analysis, and likely leading to one-sided results and a high false-positive rate. Third, our present study only performed data mining and data analysis and did not conduct any experiments to validate the results. Finally, there were only 5 healthy control samples and 16 PPD samples selected for analysis; thus, it is necessary to increase the sample size to validate the accuracy of the results in a follow-up study.

6. Conclusions

Based on our knowledge, we for the first time used the system biology-based WGCNA method to predict several potential biological pathways and diagnostic biomarkers involved in PPD. WGCNA and co-expression network analysis identified key biological processes and signaling pathways, especially the neurotrophin signaling pathway, chemokine signaling pathway, Fcγ receptor-mediated phagocytosis, and the MAPK signaling pathway, which may contribute to the elucidation of the pathogenesis and progression of PPD. The potential diagnostic biomarkers included HNRNPA2B1, IL10, RAD51, UBA52, NHP2, RPL13A, and FBL. Finally, in-depth molecular biological experiments are required to determine the exact functions of the biological pathways and diagnostic biomarkers of PPD.

7. Author contributions

GW designed this study. GW and XH provided administrative support. DZ and CW collected the data, writed the manuscript. LJ, DA, YY, TJ, and YC analyzed the data. All the authors approved the final version of manuscript.

8. Ethics approval and consent to participate

Not applicable.

9. Acknowledgment

Not applicable.

10. Funding

11. Conflict of interest

The authors declare no conflict of interest.

12. References

[1] Villegas L, McKay K, Dennis C, Ross LE. Postpartum depression among rural women from developed and developing countries: a systematic review. The Journal of Rural Health. 2011; 27: 278–288.

[2] Beckwith B. Boston Women's Health Book Collective. Women & Health. 1985; 10: 1–7.

[3] Lobel M, Dunkel-Schetter C, Scrimshaw SC. Prenatal maternal stress and prematurity: a prospective study of socioeconomically disadvantaged women. Health Psychology. 1992; 11: 32–40.

[4] Zhang Z, Hong J, Zhang S, Zhang T, Sha S, Yang R, *et al*. Postpartum estrogen withdrawal impairs hippocampal neurogenesis and causes depression- and anxiety-like behaviors in mice. International Society of Psychoneuroendocrinology. 2016; 66: 138–149.

[5] Payne JL, Maguire J. Pathophysiological mechanisms implicated in postpartum depression. Frontiers in Neuroendocrinology. 2019; 52: 165–180.

[6] Brummelte S, Galea LA. Postpartum depression: Etiology, treatment and consequences for maternal care. Hormones Behavior. 2016; 77: 153–166.

[7] Shapiro GD, Fraser WD, Séguin JR. Emerging risk factors for postpartum depression: serotonin transporter genotype and omega-3 fatty acid status. Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie. 2012; 57: 704–712.

[8] Mody I. GABAAR Modulator for Postpartum Depression. Cell. 2019; 176: 1.

[9] Liew C, Ma J, Tang H, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. The Journal of Laboratory and Clinical Medicine. 2006; 147: 126–132.

[10] Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2006; 141B: 261–268.

[11] Fan H, Sun X, Niu W, Zhao L, Zhang Q, Li W, *et al*. Altered microRNA Expression in Peripheral Blood Mononuclear Cells from Young Patients with Schizophrenia. Journal of Molecular Neuroscience. 2015; 56: 562–571.

[12] Gladkevich A, Kauffman HF, Korf J. Lymphocytes as a neural probe: potential for studying psychiatric disorders. Progress in Neuro-Psychopharmacology; Biological Psychiatry. 2004; 28: 559–576.

[13] Iwamoto K, Kakiuchi C, Bundo M, Ikeda K, Kato T. Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. Molecular Psychiatry. 2004; 9: 406–416.

[14] Pan D, Xu Y, Zhang L, Su Q, Chen M, Li B, *et al*. Gene expression profile in peripheral blood mononuclear cells of postpartum depression patients. Scientific Reports. 2018; 8: 10139.

[15] Li N, Li L, Chen Y. The Identification of Core Gene Expression Signature in Hepatocellular Carcinoma. Oxidative Medicine and Cellular Longevity. 2018; 2018: 1–15.

[16] Sherman BT, Huang DW, Tan Q, Guo Y, Bour S, Liu D, *et al*. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics. 2007; 8: 426.

[17] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, *et al*. THE STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Research. 2017; 45: D362–D368.

[18] Chin C, Chen S, Wu H, Ho C, Ko M, Lin C. CytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Systems Biology. 2014; 8: S11.

[19] Xia J, Gill EE, Hancock REW. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nature Protocols. 2015; 10: 823–844.

[20] Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. Journal of Biopharmaceutical Statistics. 2010; 20: 281–300.

[21] Kogelman LJ, Falkenberg K, Halldorsson GH, Poulsen LU, Worm J, Ingason A, et al. Comparing migraine with and without aura to healthy controls using RNA sequencing. Cephalalgia. 2019;39(11):1435–1444.

[22] Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010; 26: 976–978.

[23] Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatrica. 2007; 96: 644–647.

[24] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, *et al*. PROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12: 77.

[25] Sockol LE, Epperson CN, Barber JP. Preventing postpartum depression: a meta-analytic review. Clinical Psychology Review. 2013; 33: 1205–1217.

[26] Segman RH, Goltser-Dubner T, Weiner I, Canetti L, Galili-Weisstub E, Milwidsky A, *et al*. Blood mononuclear cell gene expression signature of postpartum depression. Molecular Psychiatry. 2010; 15: 93–100.

[27] Bian Y, Yang L, Zhao M, Li Z, Xu Y, Zhou G, *et al*. Identification of Key Genes and Pathways in Post-traumatic Stress Disorder Using Microarray Analysis. Frontiers in Psychology. 2019; 10: 302.

[28] Bloch M, Schmidt PJ, Danaceau M, Murphy J, Nieman L, Rubinow DR. Effects of gonadal steroids in women with a history of postpartum depression. The American Journal of Psychiatry. 2000; 157: 924–930.

[29] Skalkidou A, Sylvén SM, Papadopoulos FC, Olovsson M, Larsson A, Sundström-Poromaa I. Risk of postpartum depression in association with serum leptin and interleukin-6 levels at delivery: a nested case–control study within the UPPSAT cohort. International Society of Psychoneuroendocrinology. 2009; 34: 1329–1337.

[30] Sun H, Kennedy PJ, Nestler EJ. Epigenetics of the depressed brain: role of histone acetylation and methylation. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology. 2013; 38: 124–137.

[31] Wang L, Wen M, Cao X. Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses. Science. 2019; 365: eaav0758.

[32] Martinez FJ, Pratt GA, Van Nostrand EL, Batra R, Huelga SC, Kapeli K, *et al*. Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. Neuron. 2016; 92: 780–795.

[33] Kim HJ, Kim NC, Wang Y, Scarborough EA, Moore J, Diaz Z, *et al*. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature. 2013; 495: 467–473.

[34] Shubina MY, Musinova YR, Sheval EV. Proliferation, cancer, and aging-novel functions of the nucleolar methyltransferase fibrillarin? Cell Biology International. 2018; 42: 1463–1466.

[35] Loza-Muller L, Rodríguez-Corona U, Sobol M, Rodríguez-Zapata LC, Hozak P, Castano E. Fibrillarin methylates H2a in RNA polymerase i trans-active promoters in Brassica oleracea. Frontiers in Plant Science. 2015; 6: 976.

[36] Buchwalter A, Hetzer MW. Nucleolar expansion and elevated protein translation in premature aging. Nature Communications. 2017; 8: 328.

[37] McKercher SR, Torbett BE, Anderson KL, Henkel GW, Vestal DJ, Baribault H, *et al*. Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. The EMBO Journal. 1996; 15: 5647–5658.

[38] Beers DR, Henkel JS, Xiao Q, Zhao W, Wang J, Yen AA, *et al*. Wild-type microglia extend survival in PU.1 knockout mice with familial amyotrophic lateral sclerosis. Proceedings of the National Academy of Sciences. 2006; 103: 16021–16026.

[39] Satoh J, Asahina N, Kitano S, Kino Y. A Comprehensive Profile of ChIP-Seq-Based PU.1/Spi1 Target Genes in Microglia. Gene Regulation and Systems Biology. 2014; 8: 127–139.

[40] Mao J, O'Gorman C, Sutovsky M, Zigo M, Wells KD, Sutovsky P. Ubiquitin a-52 residue ribosomal protein fusion product 1

(Uba52) is essential for preimplantation embryo development. Biology Open. 2018; 7: bio035717.

[41] Zhou Q, Hou Z, Zuo S, et al. LUCAT1 promotes colorectal cancer tumorigenesis by targeting the ribosomal protein L40-MDM2-p53 pathway through binding with UBA52. Cancer Science. 2019; 110:1194–1207.

[42] Webb GC, Baker RT, Coggan M, Board PG. Localization of the human UBA52 ubiquitin fusion gene to chromosome band 19p13.1-p12. Genomics. 1994; 19: 567–569.

[43] Bonilla B, Hengel SR, Grundy MK, Bernstein KA. RAD51 Gene Family Structure and Function. Annual Review of Genetics. 2020; 54: 25–46.

[44] Benyelles M, O'Donohue M, Kermasson L, Lainey E, Borie R, Lagresle-Peyrou C, et al. NHP2 deficiency impairs rRNA biogenesis and causes pulmonary fibrosis and Høyeraal–Hreidarsson syndrome. Human Molecular Genetics. 2020; 29: 907–922.

[45] Lee J, Harris AN, Holley CL, Mahadevan J, Pyles KD, Lavagnino Z, et al. Rpl13a small nucleolar RNAs regulate systemic glucose metabolism. The Journal of Clinical Investigation. 2016; 126: 4616–4625.

[46] Règue-Guyon M, Lanfumey L, Mongeau R. Neuroepigenetics of Neurotrophin Signaling: Neurobiology of Anxiety and Affective Disorders. Progress in Molecular Biology and Translational Science. 2018; 158: 159–193.

[47] Figueira P, Malloy-Diniz L, Campos SB, Miranda DM, Romano-Silva MA, De Marco L, et al. An association study between the Val66Met polymorphism of the BDNF gene and postpartum depression. Archives of Women's Mental Health. 2010; 13: 285–289.

[48] Liu J, Meng F, Dai J, Wu M, Wang W, Liu C, et al. The BDNF-FoxO1 Axis in the medial prefrontal cortex modulates depressive-like behaviors induced by chronic unpredictable stress in postpartum female mice. Molecular Brain. 2020; 13: 91.

[49] Tan X, Du X, Jiang Y, Botchway BOA, Hu Z, Fang M. Inhibition of Autophagy in Microglia Alters Depressive-Like Behavior via BDNF Pathway in Postpartum Depression. Frontiers in Psychiatry. 2018; 9: 434.

[50] Brewster JA, Orsi NM, Gopichandran N, McShane P, Ekbote UV, Walker JJ. Gestational effects on host inflammatory response in normal and pre-eclamptic pregnancies. European Journal of Obstetrics, Gynecology, and Reproductive Biology. 2008; 140: 21–26.

[51] Fransson E, Dubicke A, Byström B, Ekman-Ordeberg G, Hjelmstedt A, Lekander M. Negative emotions and cytokines in maternal and cord serum at preterm birth. American Journal of Reproductive Immunology. 2012; 67: 506–514.

[52] Creti L, Libman E, Rizzo D, Fichten CS, Bailes S, Tran D, et al. Sleep in the Postpartum: Characteristics of first-Time, Healthy Mothers. Sleep Disorders. 2017: 8520358.

[53] Bränn E, Fransson E, White RA, Papadopoulos FC, Edvinsson A, Kamali-Moghaddam M, et al. Inflammatory markers in women with postpartum depressive symptoms. Journal of Neuroscience Research. 2020; 98: 1309–1321.

[54] Osborne L, Clive M, Kimmel M, Gispen F, Guintivano J, Brown T, et al. Replication of Epigenetic Postpartum Depression Biomarkers and Variation with Hormone Levels. Neuropsychopharmacology. 2016; 41: 1648–1658.

[55] Petralia MC, Mazzon E, Fagone P, Falzone L, Bramanti P, Nicoletti F, et al. Retrospective follow-up analysis of the transcriptomic patterns of cytokines, cytokine receptors and chemokines at preconception and during pregnancy, in women with postpartum depression. Experimental and Therapeutic Medicine. 2019; 18: 2055–2062.

Send correspondence to:

Wenqi Gao, Institute of Maternal and Child Health, Wuhan Children's Hospital (Wuhan Maternal and Child Healthcare Hospital), Tongji Medical College, Huazhong University and Technology, 430015 Wuhan, Hubei, China, E-mail: gwq1103@ctgu.edu.cn

Han Xiao, Institute of Maternal and Child Health, Wuhan Children's Hospital (Wuhan Maternal and Child Healthcare Hospital), Tongji Medical College, Huazhong University and Technology, 430015 Wuhan, Hubei, China, E-mail: tjxiaohan@hust.edu.cn

[†] These authors contributed equally.