Using Weighted Gene Co-Expression Network Analysis to Identify Key Genes Related to the Onset of Preeclampsia

Xinyang Shen

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Zhijian Wang (✉ wzjnfyy@163.com)

Southern Medical University Nanfang Hospital <https://orcid.org/0000-0003-2156-0835>

Zhirui Zeng

Guizhou Medical University

Zhenqin Huang

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Xiaowei Huang

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Lizi Zhang

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Shumin Pan

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Lijia Xie

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Jiaqi Li

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Lu Chen

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Xiaojing Yue

Southern Medical University Nanfang Hospital Department of Obstetrics and Gynecology

Research article

Keywords: Preeclampsia, Weighted gene co-expression network analysis, Hub gene, Placenta, Gene set enrichment analysis

Posted Date: October 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-87182/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Preeclampsia is a form of hypertension in pregnancy, which induced by complicated factors. However, the pathogenesis of the disease is unclear. The present study was aimed to discover the critical biomarkers associated with the occurrence and development of preeclampsia.

Methods: Gene data profile GSE75010 was downloaded from the Gene Expression Omnibus (GEO) database and used as discovery cohort to establish a WGCNA network determining significant modules which associated with clinical traits. Subsequently, functional enrichment analysis, pathway analysis and protein-protein interaction (PPI) network construction were performed on the core genes in significant modules to identify hub genes. Then, gene data profile GSE25906 was used as verified cohort to determine their diagnostic value of hub genes. The protein expression levels of these hub genes in preeclampsia and control placental tissues were verified using immunohistochemistry method. Finally, GSEA was performed to analyze their enrichment pathways.

Results: Total 33 co-expression modules were identified after the establishment of WGCNA, of which 4 gene modules were identified as significant modules because they were related to multiple (>3) clinical traits. Total 75 core genes in significant modules were analyzed, and results showed that they were mainly enriched in adaptive immune response (Gene Ontology term) and platelet activation (Kyoto Encyclopedia of Genes and Genomes term). Finally, a total of 5 genes including TYROBP, PLEK, LCP2, HCK, ITGAM were identified as hub genes which scored high in PPI network and had high diagnostic value. Furthermore, the protein level of these 5 genes in placental tissues of preeclampsia was lower than that of the control group. Moreover, these 5 genes were all enriched in 17 pathways, including autoimmunity pathway.

Conclusions: These 5 genes (TYROBP, PLEK, LCP2, HCK, ITGAM) may be closely related to the pathogenesis of preeclampsia, which may also help the diagnosis and therapy of preeclampsia.

Background

Preeclampsia is a common obstetric disease which occurs in 5–7% of pregnant woman causing more than 70,000 maternal deaths and 500,000 fetal deaths per year [1]. Pregnant women with preeclampsia will have elevated blood pressure after 20 weeks (systolic blood pressure > 140, diastolic blood pressure > 90), causing proteinuria or severe complications. Various studies [1], [2] demonstrated that preeclampsia significantly increases the risk of cardiovascular disease in the mother and fetus in the future. In addition to the developmental risks caused by premature birth, preeclampsia also increases the possibility of the occurrence of mental and neurological diseases in offspring [3]. At present, the American College of Obstetricians and Gynecologists (ACOG) [4] proposed that pregnant women with high risk of preeclampsia should take aspirin before 16 weeks to prevent it. However, the only treatment for preeclampsia is still delivery. Therefore, uncovering the molecular mechanism of the development of preeclampsia may help the clinical therapy.

Although the pathogenesis of preeclampsia remains known limit, various study had demonstrated that abnormal placenta formation or abnormal placental vascular changes may contribute to the development of preeclampsia [5]. With the delivery of the placenta, preeclampsia symptoms have alleviated or disappeared. Therefore, preeclampsia is considered a placental disease [1]. Therefore, most of the research on the mechanism of preeclampsia focuses on the genetic changes of the placenta. Endothelin-converting enzyme-1 (ECE-1) expression in placenta is significantly reduced in preeclampsia, suggesting that it caused poor placental vascularization by affecting the normal regulation of endothelin-1 (ET-1) in placental [6]. In addition, RGS2 (Regulator of G Protein Signaling-2) expression in placenta

of preeclampsia patients is reduced, G protein cascades activated by hormones associated with preeclampsia is enhanced. [7]. At present, with the development of technologies such as big data and high-throughput sequencing, omics has become a new method for studying differences in gene expression profiles of preeclampsia placenta [8]. Therefore, using bioinformatics to analyze the pathogenesis of preeclampsia has a strong advantage.

Bioinformatics analysis for placenta gene expression profile is an effective method commonly used to identify genes involved in the progression of placenta. Lei et al [9] found that the expression of ARRDC3 increased in preeclampsia by analyzing the difference between the expression of preeclampsia and normal placenta gene profiles in the GEO database. In addition, cell functional experiments were conducted and it was found that overexpression of ARRDC3 gene can inhibit the invasion and migration of HTR8/SVneo cells. Weighted gene co-expression network analysis (WGCNA) is an advanced bioinformatics algorithms which clusters genes that are highly related to clinical traits as module, and calculates the correlation coefficient between the modules and clinical traits [10]. Therefore, gene clusters associated with clinical traits were identified.

In present study, WGCNA analysis and related experiments were performed to identify key genes associated with the development of preeclampsia. Our present study may provide the novel biomarkers for the diagnosis of preeclampsia, as well as potential targets for the clinical therapy.

Materials And Methods

Data Collection and Preprocessing.

Gene expression profile data GSE75010 and GSE25906 were downloaded from the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) [11]. Gene expression profile GSE75010 was used as discovery cohort, while GSE25906 was used as validation cohort. The GSE75010 dataset contained the gene expression profile information of 80 preeclampsia placentas and 77 non-preeclampsia placentas using a GPL6244 platform (Affymetrix Human Gene 1.0 ST Array). The GSE25906 dataset was performed using the GPL6102 platform (Illumina human-6 v2.0 expression beadchip) and contained the gene expression profile of 23 preeclampsia and 37 normal control placentas samples. Before performing WGCNA analysis, the probe names in GSE75010 were converted to the gene names (If the gene symbol corresponded to more than one probe names, the mean value of these probes was considered the expression value of the gene). Finally, 21045 genes were annotated.

WGCNA.

Before performing WGCNA, the gene expression profile of 80 preeclampsia placentas were used to determine the good samples and good genes via a “goodSamplesGenes function” in R software (version 3.6.3). Mean FPKM = 0.5 was used as the criterion to filter the low-expression genes. Furthermore, Pearson correlation analysis was used to cluster the samples and detect outliers. The threshold for identifying outliers was set as 60, and one outlier (GSM1940547) was found and removed. Finally, total 21045 genes of 79 preeclampsia placenta samples were enrolled to construct WGCNA, as well as the corresponding clinical traits of the patients including age, body mass index (BMI), nulliparity, previous hypertension mean uterine perfusion index (PI), mean umbilical perfusion index (PI), maximum systolic blood pressure (BP), maximum diastolic blood pressure (BP), proteinuria, delivery mode, gestation days, infant gender, newborn weight z-score, neonatal intensive care unit (NICU) transfer, placental weight z-score and umbilical cord diameter. The R package “WGCNA” (<https://www.r-project.org/>; version 3.6.2) was used to construct WGCNA. In WGCNA, genes were analyzed with each other by the Pearson's correlation test and constructed a similarity matrix. The suitable soft-thresholding power was determined which ensured scale independence > 0.85

and the mean connectivity ~ 0 . The adjacency matrix was transformed into the topological matrix (TOM), and TOM-based dissimilarity (1-TOM) measure was used to cluster the genes using the hClust function. The mini size of module gene numbers set as 30 to identify the gene modules based on the means of the dynamic tree cut method.

Identification of key clinically significant gene modules and module core genes.

After the construction of WGCNA, the correlation between the gene modules and clinical traits were analyzed. The clinical traits mainly included the maternal age, BMI, nulliparity, previous hypertension, mean uterine PI, mean umbilical PI, maximum systolic blood pressure, maximum diastolic blood pressure, mode proteinuria, mode of delivery, gestation (days), infant gender, newborn weight z-score, NICU transfer, placental weight z-score, umbilical cord diameter. Clinically significant gene modules were determined which were relative to the clinical traits meeting the cutoffs as correlation score (cor) > 0.3 and P value < 0.05 . While clinically significant gene modules related to more than 3 clinical traits, the modules were considered as key clinically significant gene modules. Then, the module membership (MM) of each gene in clinically significant gene modules was calculated. Genes with MM score ≥ 0.8 was set as module core genes.

Functional enrichment analysis and KEGG analysis of module genes.

All module core genes were uploaded to the DAVID 6.8 [12] (<http://david-d.ncifcrf.gov/>) to perform the Gene Ontology (GO) functional annotation. GO functional annotation included three aspects: biological processes (BP), molecular function (MF) and cellular components (CC). Furthermore, the Kyoto encyclopedia of genes and genomes (KEGG) enrichment pathway analysis was performed on kobas3.0 [13] (<http://kobas.cbi.pku.edu.cn/>). All results met the cut-off as adjusted P-value < 0.05 and were visualized by R software.

PPI network construction and identification of hub genes.

STRING [14] (version 11.0; <https://string-db.org/>) was used to analyze the interaction between module core genes. The PPI network was visualized by Cytoscape software (version: 3.7.2; <https://cytoscape.org/>). The R software was used to count the connections between genes, the number of connections represents the degree of connectivity. The genes with the top 10%-degree score were selected as the hub genes for further analysis.

Validation of the expression of hub genes.

The expression of hub genes between 79 preeclampsia placentas and 77 non-preeclampsia placentas in gene expression profile GSE75010 was analyzed using GraphPad Prism (version 7.0; GraphPad Software, Inc.). Data was presented as mean $\pm$ standard deviation. Unpaired independent t test was used for statistical analysis, and P < 0.05 was set as cut-off to consider statistically significant.

Diagnostic value of hub genes.

The GSE25906 dataset was used as verification cohort to evaluate the sensitivity and specificity of the core genes via ROC curve analysis. SPSS 22.0 (IBM, Corps.) was used to create ROC curves. While the area under the ROC curve (AUC) was more than 0.6, the genes were considered to have prominent diagnostic value.

Tissues ethics.

The procedures of the present study were ethically approved by the Clinical Ethics Management Committee of Nanfang Hospital of Southern Medical University (NFEC-2019-133, July 26, 2019). A total of 15 normal placenta

tissues and 18 preeclampsia placenta tissues were obtained between February 2020 and May 2020. All women included in this study did not have any complications other than preeclampsia. All pregnant women included in this study signed an informed consent.

Immunohistochemical staining.

All tissues are fresh tissues obtained within half an hour after delivery, and then routinely fixed with paraformaldehyde at room temperature for at least 24 hours. Then gradually dehydrate with stepwise alcohol and dimethylbenzene, finally these tissues were embedded in paraffin and cut as 4- μ m-thick serial sections. Subsequently, after heating at 60 °C for 1 h, it was dewaxed with xylene at room temperature and then rehydrated with graded ethanol (100, 80, 60, 40%). After extracting the antigen with sodium citrate (100 mm), the sections were blocked with 3% H₂O₂ for 20 minutes, and the sections were blocked with 3% bovine serum albumin for 30 minutes at room temperature. Then we incubate all samples with primary antibodies, including TYROBP (cat. no.DF7316,Affinity Biosciences,1:100), PLEK (cat. no. A6305, ABclonal,1:200), LCP2 (cat. no. 12728-1-AP,Proteintech,1:200), HCK (cat. no. A2083, ABclonal,1:100), ITGAM (cat. no. A1581, ABclonal,1:100) for 12 h at 4 °C. Subsequently, the secondary goat anti-mouse and rabbit antibodies labeled with horseradish peroxidase immunohistochemical staining (HRP) were incubated at room temperature for 2 hours. The sections were stained with HRP-3,3'-diaminobenzidine kit and the pictures were collected with an orthographic microscope (Olympus BX51). The results of immunohistochemistry were determined by staining intensity and positive cell percentage: Stain intensity score: 0 (no staining), 1 (+), 2 (++) and 3 (+++) ; the proportion of positive cells score: 0 (0–1%), 1 (1–33%), 2 (34–66%) and 3 (67–100%); Then the values were multiplied.

Gene set enrichment analysis (GSEA).

In order to explore the potential function of these hub genes in preeclampsia, the Gene set enrichment analysis method was applied. In the GSE75010 data set, 80 preeclampsia samples were divided into two groups based on the median expression levels of these hub genes. GSEA was used to evaluate whether a set of a priori defined genes show statistically significant differences between the two biological states. In this study, GSEA4.0 software was used, and c2.cp.kegg.v7.1.symbols.gmt dataset was selected as the reference dataset. Then P value < 0.05 and NES > 1.5 was chosen as the cut-off criteria.

Results

WGCNA co-expression network construction.

The outlier sample (GSM1940547) was removed, and the remaining gene expression profile and corresponding clinical traits of 79 preeclampsia samples in GSE75010 were finally enrolled into WGCNA analysis (Fig. 1). We selected the power of $\beta = 6$ as the soft-thresholding power to achieve the optimal scale independence (> 0.85) and mean connectivity (~ 0) (Fig. 2a-b). Similarly, checking scale free topology showed that the soft-thresholding power was suitable which ensured scale free $R^2 \geq 0.85$ (Fig. 2c). Finally, thirty three co-expression modules were obtained including floral white module, light green module, maroon module, thistle2 module, cyan module, dark orange module, paleturquoise module, light pink4 module ,plum2 module, light cyan module, brown4 module, dark magenta module, purple module, navajowhite2 module, grey60 module, palevioletred3 module, sky blue3 module, ivory module, pink module, yellow green module, midnight blue module, dark turquoise module, light yellow module, bisque4 module, white module, salmon4 module, steel blue module, dark slate blue module, yellow module,

darkorange2 module, blue module, green module. Genes without co-expression relationship were set in grey module (Fig. 2d).

Identification of clinically significant modules and module core genes.

In order to clarify the relationship between preeclampsia clinical traits and gene modules, we performed Pearson analysis on each module and clinical traits. Genes in light green module were significantly associated with gestation duration ($R = 0.31$, $P = 0.006$) and nulliparity ($R = -0.4$, $P = 0.0003$). The genes of cyan module were significantly associated with umbilical cord diameter ($R = -0.39$, $P = 0.0003$). Moreover, genes in dark orange module were significantly associated with newborn weight z-score ($R = -0.39$, $P = 0.0005$), and were significantly associated with umbilical cord diameter ($R = -0.32$, $P = 0.004$). Genes in light cyan module were significantly associated with age ($R = -0.31$, $P = 0.005$), while genes in navajowhite2 module were significantly associated with BMI ($R = 0.49$, $P = 0.000004$). The genes in purple module were significantly associated with the mean umbilical PI ($R = -0.32$, $P = 0.004$), while the genes in pink modules were significantly associated with mean uterine PI ($R = 0.38$, $P = 0.0005$) and gestation days ($R = -0.41$, $P = 0.0002$). Furthermore, the floral white module, plum2 module, dark magenta module and yellow green module were closely related to various clinical traits (> 3), so these were regarded as the significant modules for subsequent analysis (Fig. 3). Through calculating the module membership (MM) of each gene in significant modules, 75 genes with $MM > 0.8$ were identified and set as module core genes (Table 1).

Table 1
Identification of module core genes

Gene	Module color
PECAM1.1	dark magenta
GUCY1B3	dark magenta
CD34	dark magenta
GJC1	dark magenta
TCF4	dark magenta
PLSCR4	dark magenta
SPTAN1	dark magenta
MEF2C	dark magenta
FLI1	dark magenta
EGFLAM	dark magenta
KL	dark magenta
ADGRL2	dark magenta
PCDHB14	dark magenta
ADAMTSL3	dark magenta
CYTH4	floral white
BTK	floral white
CYBB	floral white
CD53	floral white
CD4	floral white
ITGAM	floral white
TFEC	floral white
HNMT	floral white
TBXAS1	floral white
MS4A14	floral white
MILR1	floral white
SLC7A7	floral white
SYK	floral white
TMSB4X	floral white

Annotation: Genes in significant modules with module membership ≥ 0.8 were set as module core genes and showed in the list.

Gene	Module color
HCK	floral white
TYROBP	floral white
AOAH	floral white
CD84	floral white
NCKAP1L	floral white
CORO1A	floral white
CTSS	floral white
LRMP	floral white
CD86	floral white
TNFSF8	floral white
PIK3CG	floral white
TRPS1	floral white
LAIR1	floral white
PLEK	floral white
PRKCB	floral white
PTPRC	floral white
SH3BGRL	floral white
TLR1	floral white
TLR4	floral white
GPR65	floral white
C3AR1	floral white
FCER1G	floral white
AIF1	floral white
DOCK2	floral white
LCP2	floral white
LAPTM5	floral white
MYO1F	floral white
SAMHD1	floral white
TLR7	floral white
TLR8	floral white
Annotation: Genes in significant modules with module membership ≥ 0.8 were set as module core genes and showed in the list.	

Gene	Module color
ADAP2	floral white
ARHGAP15	floral white
RASSF4	floral white
HAVCR2	floral white
OLFML2B	plum2
SEC14L1	plum2
AFF2	plum2
HIP1.1	plum2
KCNK3	plum2
PLCG1	plum2
TMEM204	plum2
ATP1A1	yellow green
ZMYND11	yellow green
TMEM139	yellow green
THSD4	yellow green
FBN2	yellow green
SLC23A2	yellow green
Annotation: Genes in significant modules with module membership ≥ 0.8 were set as module core genes and showed in the list.	

Functional enrichment analysis and KEGG analysis for module core genes.

Module core genes were uploaded to the DAVID database for GO functional annotation. The top 10 results showed that the genes in key clinical significant modules were mainly enriched in adaptive immune response (BP, GO: 0002250), innate immune response (BP, GO: 0045087), integral component of plasma membrane (CC, GO: 0005887), inflammatory response (BP, GO: 0006954), plasma membrane (CC, GO: 0005886), B cell receptor signaling pathway (BP, GO: 0050853), toll-like receptor signaling pathway (BP, GO: 0002224), positive regulation of T cell proliferation (BP, GO: 0042102), MyD88-dependent toll-like receptor signaling pathway (BP, GO: 0002755), leukocyte migration (BP, GO: 0050900) (Fig. 4). The KEGG pathway enrichment analysis reflected that module core genes were enriched in pathways platelet activation, Fc gamma R-mediated phagocytosis, tuberculosis, natural killer cell mediated cytotoxicity, Fc epsilon RI signaling pathway, NF-kappa B signaling pathway, Toll-like receptor signaling pathway, Cell adhesion molecules (CAMs), phospholipase D signaling pathway and phagosome (Fig. 5).

PPI network construction and identification of hub genes.

Module core genes were uploaded to the String database for construction of PPI network (Fig. 6a). R was used to calculate the number of connections between genes, and the genes including PTPRC, TYROBP, PLEK, LCP2, HCK,

ITGAM, BTK, CD86 with the top 10%degree score were obtained as hub genes (Fig. 6b). These hub genes were used further analysis.

Identification the expression of core genes between preeclampsia placentas and non-preeclampsia placentas.

Furthermore, the expression of hub genes between 79 preeclampsia placentas and 77 non-preeclampsia placentas in GSE75010 were analyzed. The results showed that the expression levels of these hub genes including PTPRC, TYROBP, PLEK, LCP2, HCK, ITGAM, BTK, CD86 were significantly decreased in the preeclampsia placentas compared with the non-preeclampsia placentas (Fig. 7).

Diagnostic efficiency analysis of hub genes.

Then, the gene expression profile GSE25906 included 23 preeclampsia placentas and 37 non- preeclampsia placentas were used to detect the diagnostic efficiency of hub genes via constructing ROC curve. The results showed that TYROBP (AUC = 0.744), PLEK (AUC = 0.783), LCP2 (AUC = 0.732), HCK (AUC = 0.669), ITGAM (AUC = 0.746) exhibited remarkable diagnostic efficiency (Fig. 8).

Immunohistochemical verification results of hub genes.

In order to verify the expression of hub genes between the two groups, we performed immunohistochemistry on 18 preeclamptic placenta tissues and 15 normal control placenta tissues. There was no difference in gestational days, age, and BMI between the two groups of pregnant women (Table 2). The results showed that these hub genes (TYROBP, PLEK, LCP2, HCK, ITGAM) have lower expression in preeclampsia than the normal group (Fig. 9).

Table 2
Clinical characteristics of two groups of pregnant women

Clinical traits	group	
	Control (n = 15)	Preeclampsia (n = 18)
gestational age, days	260.40 ± 18.59	251.94 ± 23.13
birth weight, g	2847.33 ± 603.83	2247.78 ± 749.61*
placenta weight, g	584.67 ± 106.90	493.33 ± 139.58*
systolic pressure, mm Hg	111.67 ± 7.67	163.67 ± 13.58*
diastolic pressure, mm Hg	74.67 ± 4.05	102.78 ± 9.37*
Maternal age, years	29.20 ± 2.73	29.11 ± 3.85
Maternal body mass index, kg/m ²	26.13 ± 2.40	27.45 ± 2.70
Annotation: Clinical characteristics of two groups of pregnant women for immunohistochemical staining. The value was expressed as mean ± SD. The P value between the preeclampsia group and the control group was given. *P< 0.05.		

Gene set enrichment analysis (GSEA).

In order to explore the potential biological functions of hub genes in preeclampsia, GSE75010 dataset was used for GSEA analysis. The results indicated that these 5 hub genes were all enriched in the 17 pathways including T cell receptor signaling pathway, leukocyte transendothelial migration, viral myocarditis, primary immunodeficiency, FC

epsilon RI signaling pathway, prion diseases, cell adhesion molecules cams, natural killer cell mediated cytotoxicity, complement and coagulation cascades, leishmania infection, systemic lupus erythematosus, antigen processing and presentation, intestinal immune network for IgA production, pantothenate and COA biosynthesis, nod like receptor signaling pathway, pathogenic Escherichia coli infection, glutathione metabolism (Table 3). These results indicated that these 5 hub genes may be involved in the development of preeclampsia via regulating these pathways.

Table 3
Pathways enriched in 5 genes in GSEA analysis

GGene	HCK	ITGAM	LCP2	PLEK	TYROBP
Pathways					
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	2.12	2.01	1.82	1.83	1.9
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	1.99	2.05	1.81	1.75	1.97
KEGG_VIRAL_MYOCARDITIS	1.88	1.71	1.62	1.62	1.72
KEGG_PRIMARY_IMMUNODEFICIENCY	1.88	1.76	1.61	1.86	1.94
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	1.88	1.89	1.69	1.89	1.81
KEGG_PRION_DISEASES	1.86	1.76	1.76	1.81	1.83
KEGG_CELL_ADHESION_MOLECULES_CAMS	1.86	1.86	1.65	1.77	1.79
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	1.84	1.73	1.52	1.62	1.8
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	1.81	1.71	1.58	1.72	1.79
KEGG_LEISHMANIA_INFECTION	1.8	1.7	1.64	1.69	1.67
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	1.77	1.68	1.65	1.67	1.62
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	1.68	1.66	1.53	1.57	1.63
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	1.67	1.52	1.52	1.56	1.58
KEGG_PANTOTHENATE_AND_COA_BIOSYNTHESIS	1.66	1.52	1.53	1.55	1.69
KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	1.63	1.65	1.61	1.74	1.63
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	1.61	1.88	1.57	1.68	1.64
KEGG_GLUTATHIONE_METABOLISM	1.59	1.61	1.53	1.78	1.53
Annotation: The GSE75010 was used for GSEA analysis. The values in the table represent the NES value. The table listed KEGG pathways in which all genes satisfy $P < 0.05$. The number in units means the NES value of each for the pathway.					

Discussion

Preeclampsia is one of the most serious diseases that endanger the life of the mother and fetus. Many scholars were devoted to the study the pathogenesis and demonstrated that the possible pathogenesis of preeclampsia included genetic factor, angiogenic imbalances, poor immune adaptation, abnormal lipid metabolism, trophoblast invasive decreased and increased apoptosis or necrosis, abnormal placentation and excessive maternal inflammatory. Among them, most of the research evidence [15] indicates that abnormal placenta formation, and changes in the

placental gene profile have a greater impact on the trophoblast migration and invasion function, which leads to the occurrence of preeclampsia. Therefore, it is of great significance to explore the changes of placental gene profile in preeclampsia.

In the current study, GSE75010 (discovery cohort) and GSE25906 (validation cohort) were downloaded from the GEO database. Through construction of WGCNA, GO analysis, KEGG analysis, construction of PPI network, expression analysis, diagnostic value validation and IHC experiment, TYROBP, PLEK, LCP2, HCK, ITGAM were identified as hub genes.

Awamleh [16] showed that ITGAM mRNA expression is reduced in the placenta of patients with PE and IUGR. ITGAM was regulated by miR-210-5p, and participated in the occurrence of preeclampsia by inhibiting the invasion and migration of HTR8 cells. In current study, we found that ITGAM gene belongs to floral white module, which is negatively correlated with mean uterine perfusion index (PI), mean umbilical perfusion index (PI), and positively correlated with the total days of pregnancy. In addition, immunohistochemistry and expression analysis indicate that the ITGAM gene is lowly expressed in preeclampsia and has good diagnostic value. Therefore, we hypothesize that ITGAM gene is a protective factor for preeclampsia, which is consistent with previous research results.

TYROBP encodes a transmembrane signal peptide whose cytoplasmic region contains an immunoreceptor activation motif (ITAM) based on tyrosine[17]. TYROBP has been identified as a universal marker for macrophages in mouse and human tissues[18], and the preeclampsia pathogenesis is manifested by the maternal innate immune response to various proinflammatory stimuli from the placenta. LCP2 is a coding protein that encodes a substrate for the T cell antigen receptor (TCR) to activate the protein tyrosine kinase pathway[19]. Lcp2 protein expression is reduced in colonic epithelial cells of spontaneously hypertensive rat model[20]. Pan[21] proposed that TYROBP and LCP2 may be related to the pathogenesis of atherosclerosis in a high-fat diet Tibetan minipig model. Song et al conducted a biological analysis of the expression profile chips of periodontitis and normal gingival tissues and found that PLEK may interact with IRF-8 to cause periodontal inflammation, suggesting that PLEK may be related to inflammatory reactions[22]. Furthermore, PLEK is associated with the pathogenesis of abdominal aortic aneurysms, which are a type of vascular disease, and its pathogenesis may be related to oxidative stress, inflammation[23]. Therefore, PLEK is related to inflammatory response and oxidative stress response, so it may be involved in the development of preeclampsia. Hematopoietic cell kinase (HCK) is a member of the Src family of protein kinases, and is involved in various signal transduction pathways that regulate cell growth, proliferation, differentiation, migration, apoptosis[24]. In colorectal tumors, HCK is associated with prognosis and immune cell responses during local inflammation[25]. However, these genes are related to inflammatory response, immune response, cell migration and invasion, but none have been reported in preeclampsia. In our research, we found these genes belong to floral white module, which is negatively correlated with mean uterine perfusion index (PI), mean umbilical perfusion index (PI), and positively correlated with the total days of pregnancy, as well as possessing good diagnostic value. Moreover, the expression levels of these genes and the results of immunohistochemistry suggest that the placenta in preeclampsia is lower than that in the control group. Therefore, we speculate that TYROBP, PLEK, LCP2, HCK genes may be protective factors in preeclampsia, but more functional experiments are still needed to verify.

Conclusions

In summary, WGCNA analysis, GO analysis, KEGG analysis, expression analysis, diagnosis value analysis, GSEA and immunohistochemical stain were used, and 5 genes (TYROBP, PLEK, LCP2, HCK, ITGAM) which were considered playing a protective role in the development of preeclampsia. These genes may be helpful for the diagnosis and treatment of preeclampsia.

Abbreviations

GEO: Gene Expression Omnibus; WGCNA: weighted gene co-expression network analysis

PPI: protein-protein interaction; GSEA: Gene set enrichment analysis; GO: Gene Ontology term

KEGG: Kyoto Encyclopedia of Genes and Genomes term; SBP: systolic blood pressure; DBP: diastolic blood pressure; ACOG: American College of Obstetricians and Gynecologists; BMI: body mass index; PI: perfusion index; NICU: neonatal intensive care unit; ECE-1: Endothelin-converting enzyme-1; ET-1: endothelin-1; RGS2: Regulator of G Protein Signaling-2; BP: biological processes; MF: molecular function; CC: cellular components; TYROBP: transmembrane immune signaling adaptor; PLEK: pleckstrin; LCP2: lymphocyte cytosolic protein 2; HCK: HCK proto-oncogene; ITGAM: integrin subunit alpha M

Declarations

Acknowledgements

Not applicable

Funding

the research funding comes from the National Key Research and Development Program of China (grant no. 2018YFC1004104).

Availability of data and materials

All data in this study can be provided with the consent of the corresponding author.

Ethics approval and consent to participate

The medical Ethics committee of NanFang Hospital of Southern Medical University approved the study. All patients who provided samples signed an informed consent form. The present study was performed in accordance with the principles outlined in the declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Rana S, Lemoine E, Granger JP, Karumanchi SA. Preeclampsia: Pathophysiology, Challenges, and Perspectives. *Circulation research*. 2019;124(7):1094–112.
2. Alsnes IV, Vatten LJ, Fraser A, Bjorngaard JH, Rich-Edwards J, Romundstad PR, Asvold BO: **Hypertension in Pregnancy and Offspring Cardiovascular Risk in Young Adulthood: Prospective and Sibling Studies in the HUNT Study (Nord-Trondelag Health Study) in Norway.** *Hypertension (Dallas, Tex. 1979)* 2017, **69**(4):591–598.

3. Gumusoglu SB, Chilukuri ASS, Santillan DA, Santillan MK, Stevens HE. Neurodevelopmental Outcomes of Prenatal Preeclampsia Exposure. *Trends in neurosciences*. 2020;43(4):253–68.
4. **ACOG Practice Bulletin**. No. 202: Gestational Hypertension and Preeclampsia. *Obstetrics gynecology*. 2019;133(1):e1–25.
5. Raguema N, Moustadraf S, Bertagnolli M. Immune and Apoptosis Mechanisms Regulating Placental Development and Vascularization in Preeclampsia. *Frontiers in physiology*. 2020;11:98.
6. McCormick A, Kristianto J, Wang X, McIntosh J, Cruz M, Hibbard JU, Blank RD. Placental endothelin-converting enzyme-1 is decreased in preeclampsia. *Pregnancy Hypertens*. 2020;20:108–10.
7. Perschbacher KJ, Deng G, Sandgren JA, Walsh JW, Witcher PC, Sapouckey SA, Owens CE, Zhang SY, Scroggins SM, Pearson NA, et al: **Reduced mRNA Expression of RGS2 (Regulator of G Protein Signaling-2) in the Placenta Is Associated With Human Preeclampsia and Sufficient to Cause Features of the Disorder in Mice**. *Hypertension (Dallas, Tex. 1979)* 2020, **75**(2):569–579.
8. Benny PA, Alakwaa FM, Schlueter RJ, Lassiter CB, Garmire LX. A review of omics approaches to study preeclampsia. *Placenta*. 2020;92:17–27.
9. Lei D, Deng N, Wang S, Huang J, Fan C. Upregulated ARRDC3 limits trophoblast cell invasion and tube formation and is associated with preeclampsia. *Placenta*. 2020;89:10–9.
10. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
11. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*. 2002;30(1):207–10.
12. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009;4(1):44–57.
13. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic acids research*. 2006;34:W720–4. (Web Server issue).
14. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*. 2019;47(D1):D607-d613.
15. Mohamad MA, Mohd Manzor NF, Zulkifli NF, Zainal N, Hayati AR, Ahmad Asnawi AW. **A Review of Candidate Genes and Pathways in Preeclampsia-An Integrated Bioinformatical Analysis**. *Biology* 2020, 9(4).
16. Awamleh Z, Han VKM. Identification of miR-210-5p in human placentae from pregnancies complicated by preeclampsia and intrauterine growth restriction, and its potential role in the pregnancy complications. *Pregnancy Hypertens*. 2020;19:159–68.
17. Ascone G, Di Ceglie I, Walgreen B, Sloetjes AW, Lindhout E, Bot I, van de Loo FAJ, Koenders MI, van der Kraan PM, Blom AB, et al. High LDL levels lessen bone destruction during antigen-induced arthritis by inhibiting osteoclast formation and function. *Bone*. 2020;130:115140.
18. Dang D, Taheri S, Das S, Ghosh P, Prince LS, Sahoo D. Computational Approach to Identifying Universal Macrophage Biomarkers. *Frontiers in physiology*. 2020;11:275.
19. Wang X, Li JP, Chiu LL, Lan JL, Chen DY, Boomer J, Tan TH. Attenuation of T cell receptor signaling by serine phosphorylation-mediated lysine 30 ubiquitination of SLP-76 protein. *J Biol Chem*. 2012;287(41):34091–100.
20. Yang T, Li H, Oliveira AC, Goel R, Richards EM, Pepine CJ, Raizada MK. Transcriptomic signature of gut microbiome-contacting cells in colon of spontaneously hypertensive rats. *Physiol Genom*. 2020;52(3):121–32.

21. Pan Y, Yu C, Huang J, Rong Y, Chen J, Chen M. Bioinformatics analysis of vascular RNA-seq data revealed hub genes and pathways in a novel Tibetan minipig atherosclerosis model induced by a high fat/cholesterol diet. *Lipids Health Dis.* 2020;19(1):54.
22. Song L, Yao J, He Z, Xu B. Genes related to inflammation and bone loss process in periodontitis suggested by bioinformatics methods. *BMC oral health.* 2015;15:105.
23. Hinterseher I, Erdman R, Elmore JR, Stahl E, Pahl MC, Derr K, Golden A, Lillvis JH, Cindric MC, Jackson K, et al. Novel pathways in the pathobiology of human abdominal aortic aneurysms. *Pathobiol J ImmunoPathol Mol Cell Biol.* 2013;80(1):1–10.
24. Wang J, Chen R, Liu X, Shen J, Yan Y, Gao Y, Tao T, Shi J. Hck Promotes Neuronal Apoptosis Following Intracerebral Hemorrhage. *Cell Mol Neurobiol.* 2017;37(2):251–61.
25. Roseweir AK, Powell A, Horstman SL, Inthagard J, Park JH, McMillan DC, Horgan PG, Edwards J. Src family kinases, HCK and FGR, associate with local inflammation and tumour progression in colorectal cancer. *Cellular signalling.* 2019;56:15–22.

Figures

Sample dendrogram and trait heatmap

Figure 1

Sample clustering tree and clinical trait heat map in preeclampsia. The cutoff value was set as 60, and one outlier (GSM1940547) was found and removed. In this study, 16 clinical traits were described, including age, body mass index (BMI), nulliparity, previous hypertension, mean uterine perfusion index (PI), mean umbilical perfusion index (PI), maximum systolic blood pressure (BP), maximum diastolic blood pressure (BP), proteinuria, delivery mode, gestation days, infant gender, newborn weight z-score, neonatal intensive care unit (NICU) transfer, placental weight z-score and umbilical cord diameter.

Figure 2

Selection of soft threshold power in Weighted Gene Co-expression Network Analysis (WGCNA). (a) Scale independence and (b) mean connectivity of various soft-thresholding powers (β). (c) Scale-free topology when $\beta=6$. (d) Clustering tree map of GSE75010 gene. Each branch represents a gene, and the different colors below represent different co-expression modules.

Module-trait relationships

Figure 3

Identification of key modules. Heatmap for the study of the correlation between modular genes and clinical traits. Among them, the flower white module, plum2 module, deep magenta module and yellow-green module were closely related to a number of clinical traits (>3) and were regarded as important modules. BMI, body mass index; uterine PI mean uterine perfusion index; umbilical PI, mean umbilical perfusion index; max systolic BP, maximum systolic blood pressure; max diastolic BP, maximum diastolic blood pressure; ga day, gestation days; NICU transfer, neonatal intensive care unit (NICU) transfer.

Figure 4

GO analysis results of module core genes .GO Gene Ontology.

Figure 5

KEGG analysis results of module core genes. KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 6

PPI network construction and selection of hub genes. (a) Flower white, plum2, dark magenta and yellow-green modular genes for PPI network construction. (b) Genes with the top10 highest score based on the number of connections.

Figure 7

Verification of hub gene expression. Comparison of the expression levels of core genes (PTPRC, TYROBP, PLEK, LCP2, HCK, ITGAM, BTK, CD86) in the data set GSE75010 between preeclampsia and the control group. The blue square represents the genes expression level in the control group (n=77), and the red square represents the genes expression level in the preeclampsia group (n=79). *, P < 0.05.

Figure 8

Diagnostic efficiency verification of hub genes. Obtain the expression of hub genes in the GSE25906 data set and use receiver operating curve to verify the diagnostic value. The 5 genes—TYROBP, PLEK, LCP2, HCK, ITGAM—in the figure have significant diagnostic efficiency.

Figure 9

Validation results of immunohistochemical staining of hub genes. Representative images of immunohistochemical staining of TYROBP, PLEK, LCP2, HCK, ITGAM genes in control group and placenta tissues of preeclampsia.