International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

https://gssrr.org/index.php/JournalOfBasicAndApplied/index

Identification of Epithelial-mesenchymal Transition-related Genes and Potential Biomarkers for Bronchopulmonary Dysplasia using Bioinformatical Analysis

Xiaohong He[a], Zongli Zhang[b], Shibing Xi[c*]

[a,c]*Department of Pediatrics, Affiliated Hubei University of Medicine, Shiyan, Hubei,442000, China*

[b]*Institute of Pediatric Diseases, Affiliated Taihe Hospital of Hubei University of Medicine, Shiyan, Hubei, 442000, China*

[c]*Department of Pediatrics, Affiliated Taihe Hospital of Hubei University of Medicine, Shiyan, Hubei, 442000, China.*

[a]*Email: 903066040@qq.com,* [b]*Email: zzl@hbmu.edu.cn*

[c]*Email: xishibing2009@163.com*

Abstract

Bronchopulmonary dysplasia (BPD) is a prevalent and serious disease among preterm newborns. Epithelial-mesenchymal transition (EMT) represents the biological process where cells switch from an epithelial to mesenchymal phenotype. Hyperoxia induces EMT in alveolar epithelial cells, which can affect alveolar development. However, the specific molecular mechanism of EMT in BPD remains incompletely unclear. In this study, the GSE108754 and GSE32472 datasets were merged and de-batched, with external validation from GSE188944 and GSE51039. Differentially Expressed Genes (DEGs) were screened, focusing on those associated with EMT using the GeneCards database, resulting in the identification of 213 overlapping genes. Weighted gene co-expression network analysis (WGCNA) revealed that the turquoise module was strongly correlated with BPD, identifying 33 hub genes related to EMT in BPD patients.

Received: 5/1/2025
Accepted: 6/22/2025
Published: 7/4/2025

---

* *Corresponding author.*

Utilizing support vector machine (SVM) and random forest recursive feature elimination (RF-RFE) methods, 17 potential biomarkers were assessed. Risk prediction performance was evaluated through a nomogram model and receiver operating characteristic (ROC) analysis. Notably, FLOT2, AQP9, SEMA4A, and CXCR1 emerged as vital genes, with RT-qPCR validation indicating their mRNA levels significantly increased in BPD. In summary, four biomarkers, FLOT2, AQP9, SEMA4A, and CXCR1, were potentially critical in BPD occurrence and progression, shedding novel light on diagnosing and treating BPD.

Keywords: Bronchopulmonary dysplasia; Epithelial-mesenchymal transition; Weighted gene co-expression network analysis; Machine learning; Bioinformatic analysis.

1. Introduction

Bronchopulmonary dysplasia (BPD) represents a chronic respiratory disorder commonly observe in premature infants[1]. As medical advancements lead to increased survival rates for very low-birth-weight infants, BPD rates have also increased[2]. Patients with BPD usually face a grim prognosis, with the potential for long-term effects extending into adulthood, including severe neurodevelopmental disorders and compromised lung function, bringing great burdens on families and society[3]. The histopathology and pathogenic mechanisms of BPD are complex and incompletely comprehended, and there is a lack of safe and efficient prevention or clinical treatments[4]. As a result, it is urgently necessary to delve into the BPD pathogenic mechanism and explore potential treatments.

Epithelial-mesenchymal transition (EMT) is a complex biological procedure, involving the epithelial polarity reversal and mesenchymal characteristics acquisition by epithelial cells[5, 6]. EMT is a key player in embryonic development, tissue fibrosis wound healing, and cancer metastasis[7], accompanied by elevated protein and mRNA expression of N-cadherin and α-Smooth muscle actin(α-SMA), and decreased expression of E-cadherin. An increasing number of evidence suggests that EMT may be beneficial for the development of BPD by influencing alveolar epithelial cell structure and function. For example, EMT in alveolar epithelial cells can result in alveolar wall thickening and lumen reduction, influencing gas exchange efficiency[8]. A study observed that hyperoxia-induced differentiation of lung type II cells into fibroblasts through EMT impairs alveolar development[6]. Yang and his colleagues have reported that EMT also influences BPD development[9]. However, the precise relationship between BPD and EMT needs to be further investigated.

FLOT2, AQP9, SEMA4A, and CXCR1 were analyzed in this study using bioinformatics methods to examine their role in the progression of BPD. Four distinct biomarkers were discovered and assessed for their diagnostic efficacy in BPD patients. Furthermore, potential signaling pathways related to the progression of BPD were clarified. Moreover, the findings of this study offer a novel insight into the association of EMT with BPD.

2. Materials and methods

2.1. Data collection

GSE108754[10], GSE32472[11], GSE188944[12], and GSE51039[13] consisted of four gene expression microarray datasets were obtained in the GEO database (http://www.ncbi.nlm.nih.gov/geo/). Microarray profiles

for blood specimens from BPD neonates are available in the GSE108754, GSE32472, and GSE188944 datasets, while expression data from mice exposed to hyperoxia is included in the GSE51039 dataset. Detailed information on the datasets is displayed in Table 1.

Table 1: Summary of BPD microarray datasets from the GEO database

Series	Platform	BPD samples	NonBPD samples	Organism	Data type
GSE108754	GPL13497	5	6	Homo sapiens	Microarray
GSE32472	GPL6244	182	112	Homo sapiens	Microarray
GSE188944	GPL14951	6	17	Homo sapiens	Microarray
GSE51039	GPL6246	6	6	Homo sapiens	Microarray

2.2. Data processing and differentially expressed genes (DEGs) overlapping gene analysis

The 'sva' package's combat function was used to eliminate batch effects[14] of GSE108754 and GSE32472 mRNA microarray datasets, which were merged and normalized. On this basis, 'limma' package in R was adopted for identifying DEGs[15] upon the thresholds of $p < 0.05$ and log FC > 0.3. External validation was performed using the GSE188944 and GSE51039 datasets. The GeneCards database, which provides detailed information on human genes[16, 17], was applied to identify genes associated with EMT that had a relevance score greater than 1. 'Venn diagram' package was utilized to identify intersected genes of the merged datasets with EMT-related genes. In addition, R software 'pheatmap' and 'ggplot2' packages were used for visualizing Heatmap and volcano plots for DEGs. The identified overlapping genes were selected for further analysis.

2.3. Immune infiltration analysis via CIBERSORT

CIBERSORT a method which can be applied to estimate the cell composition of intricate tissues using gene expression data[18, 19]. Based on the combined data set, CIBERSORT algorithm was adopted for predicting the abundance of 22 infiltrating immune cells in each sample. Afterwards, R software 'ggplot2' and 'ggpubr' were used for visualizing and comparing immune cell infiltration degrees between BPD and normal samples.

2.4. Construction of weighted gene co-expression networks

Weighted Gene Co-expression Network Analysis (WGCNA) refers to an approach that categorizes genes in microarray samples into modules in accordance with their correlation. By associating these modules with trait samples, relevant modules for targeted therapy or biomarkers can be detected[20]. 'WGCNA' package was utilized to build the co-expression network in this study[20, 21] and modules most relevant to Epithelial-Mesenchymal Transition (EMT) were identified. Initially, outliers were detected by generating the sample-clustering tree. Then, ‘WGCNA' package calculated gene co-expression matrices based on Pearson's correlation coefficients among gene pairs. A soft threshold power (β) was chosen to ensure independence and average connectivity degree within co-expression modules. Using the scale-free network, the soft threshold of 15 was

utilized for constructing the scale-free co-expression network. To explore the relationship between gene feature of modules and the disease phenotype, a dendrogram was created using hierarchical clustering. The module with the highest correlation coefficient and smallest p-value was detected as the disease characteristic. Gene significance (GS) of the trait of every gene and module membership (MM) within one module was determined. MM reflects the relationship of genes with module eigengenes, implying that genes in the modules were reliable. Core genes were filtered based on GS > 0.5 and MM > 0.8, with modules of $p < 0.05$ suggesting statistical significance. The gene intersection within GS, MM, and key modules yielded EMT-related hub genes for further investigation.

2.5. Gene selection using random and forest support vector machine

To identify feature genes, two machine-learning techniques were used including random forest (RF) and support vector machine (SVM). RF refers to a learning algorithm integrating multiple decision trees[22]. In this analysis, the 'randomForest' and 'caret' packages were utilized to construct the RF model, followed by the calculation of feature importance scores using the RF classifier. The recursive Feature Elimination (RFE) algorithm applies a classifier to rank features and iteratively eliminates the least significant ones until the desired number of features is reached[23]. Using the RFE method, we determined the importance and ranking of each gene. SVM is a machine learning algorithm proposed by Vapnik and his colleagues deriving classification insights from a limited number of samples[24]. The 'e1071' package was employed to establish the SVM-RFE to identify genes with minimal cross-validation error.

2.6. Nomogram model construction

Based on the 'rms' package, a nomogram model was constructed using the candidate genes identified by two machine learning algorithms. The expression of candidate genes was visualized using the R packages 'ComplexHeatmap', 'ggpubr', and 'ggplot2'.

2.7. Validation of candidate genes

ROC curves were constructed to evaluate effectiveness of candidate genes in combined dataset on disease diagnosis with 'pROC' package. The accuracy of these gene predictions was validated in the independent GSE188944 and GSE51039 datasets. Then, to measure accuracy and diagnostic capability, the area under the ROC curve (AUC) was employed.

2.8. Functional and pathway enrichment analysis

DAVID (https://david.ncifcrf.gov) offers a wide range of functional annotation tools for investigators to explore biological significance for a large gene repertoire[25]. This platform was used to perform functional enrichment analysis using Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The GO terms were categorized into the following three classes: biological process (BP), molecular function (MF), and cellular component (CC). Subsequently, for the purpose of visualizing the GO/KEGG data, the 'ggplot2' package in R software was adopted. A Benjamini-adjusted p-value of 0.05 was selected as the

threshold for determining significant GO terms.

2.9. Construction of protein-protein interaction network

The Search Tool for the Retrieval of Interacting Genes (STRING) database (Version 12.0, http://string-db.org/) was used to perform the PPI network analysis. Network visualization was conducted with Cytoscape (Version 3.9.1)[26]. The CytoHubba plugin within Cytoscape was employed to determine protein node degrees and identify significant modules. Various topology methods including Maximal Clique Centrality (MCC)[27], Edge Percolated Component (EPC)[27], Density of Maximum Neighborhood Component (DMNC)[28], Maximum Neighborhood Component (MNC)[27, 28]and Closeness[27] were utilized. Hub genes were selected by the top ten nodes.

2.10. Unsupervised clustering analysis

Samples were classified into different subtypes by expression of candidate genes. R software 'ConsensusClusterPlus'[29] was used for analyzing samples in experimental group and for determining clustering K value. The cumulative distribution function of K for different values was evaluated to acquire the most reliable clustering analysis results[30]. In addition, R software 'ConsensusClusterPlus' was also adopted for identifying molecular subtypes based on the expression level of candidate genes. Uniform Manifold Approximation and Projection (UMAP) was employed as an unsupervised dimension reduction technique[31]. Finally, the R software 'umap' was employed to identify BPD clusters.

2.11. Cell culture and treatment

Human alveolar basal epithelial cells (A549, Abiowell, AW-CCH011, China) were cultivated in Dulbecco's modified Eagle medium (DMEM, Cytiva, SH30243.01, USA) including 10% fetal bovine serum (FBS, absin, abs983, China), at 37 °C with 5% carbon dioxide (CO_2). The A549 cells were classified into control and hyperoxia-treated groups. Next, cells in the hyperoxia group were cultured with 85% O2 in the medium for 48 h.

2.12. Real-time fluorescence quantitative PCR (RT-qPCR)

TRIzol reagent (TIANGEN, DP424, China) was used to extract total RNA in cultivated A549 cells. Subsequently, total RNA was reverse transcribed into cDNA utilizing the HiScript Q RT SuperMix for qPCR (Vazyme, R122-01, China), in line with protocol of the manufacturer. Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Q712-02, China) was adopted for quantitative real-time PCR analysis. Candidate gene mRNA levels were calculated through the 2-ΔΔCt method following normalization to β-Actin mRNA levels. Table 2 lists the specific primers.

Table 2: Primers used in RT-qPCR

Gene	Feverse (5′–3′)	Reverse (5′–3′)
β-Actin	AAACTGGAACGGTGAAGGTG	AGAGAAGTGGGGTGGCTTTT
FLOT2	TTGCTGACTCTAAGCGAGCC	TCCACGGCAATCTGTTTCTTG
AQP9	GAAGAGCAGCTTAGCGAAAGA	ACAGCCACATCCAAGGACAAT
SEMA4A	TGGATGGGATGCTCTATTCTGG	GCGGAGGAAGTTGTCGGTC

2.13. Statistical analysis

Statistical analysis was performed with R software 4.3.1 and GraphPad Prism 9.0.0 software. Comparisons between the two groups were performed by adopting the Wilcoxon test or Student's t-test. $P < 0.05$ was thought to be of statistical significance.

3. Results

3.1. Identification of significant overlapping genes

To enhance the comprehensibility of this study, the data processing workflow was depicted (Fig. 1). Since GSE108754 and GSE32472 were conducted on different platforms, with different samples and timeframes, we combined and standardized samples from both datasets using the 'sva' R package. This approach was aimed at increasing the sample size and mitigating batch effects. DEGs between BPD patients and healthy individuals were analyzed with the 'limma' package. The combined microarray datasets showed totally 891 DEGs, including 481 substantially up-regulated and 410 significantly down-regulated genes in BPD relative to the control group. These findings were visually represented in a volcano plot (Fig. 2A) and heatmap (Supplementary Fig. 1). Furthermore, we retrieved EMT-related genes from GeneCards, causing 6960 EMT-associated genes. By applying a relevance score threshold of >1, a total of 3403 EMT-associated genes were detected. Ultimately, by intersecting the 891 DEGs with the 3403 EMT-associated genes, 213 overlapping genes were identified using VennDiagram (Fig. 2B).

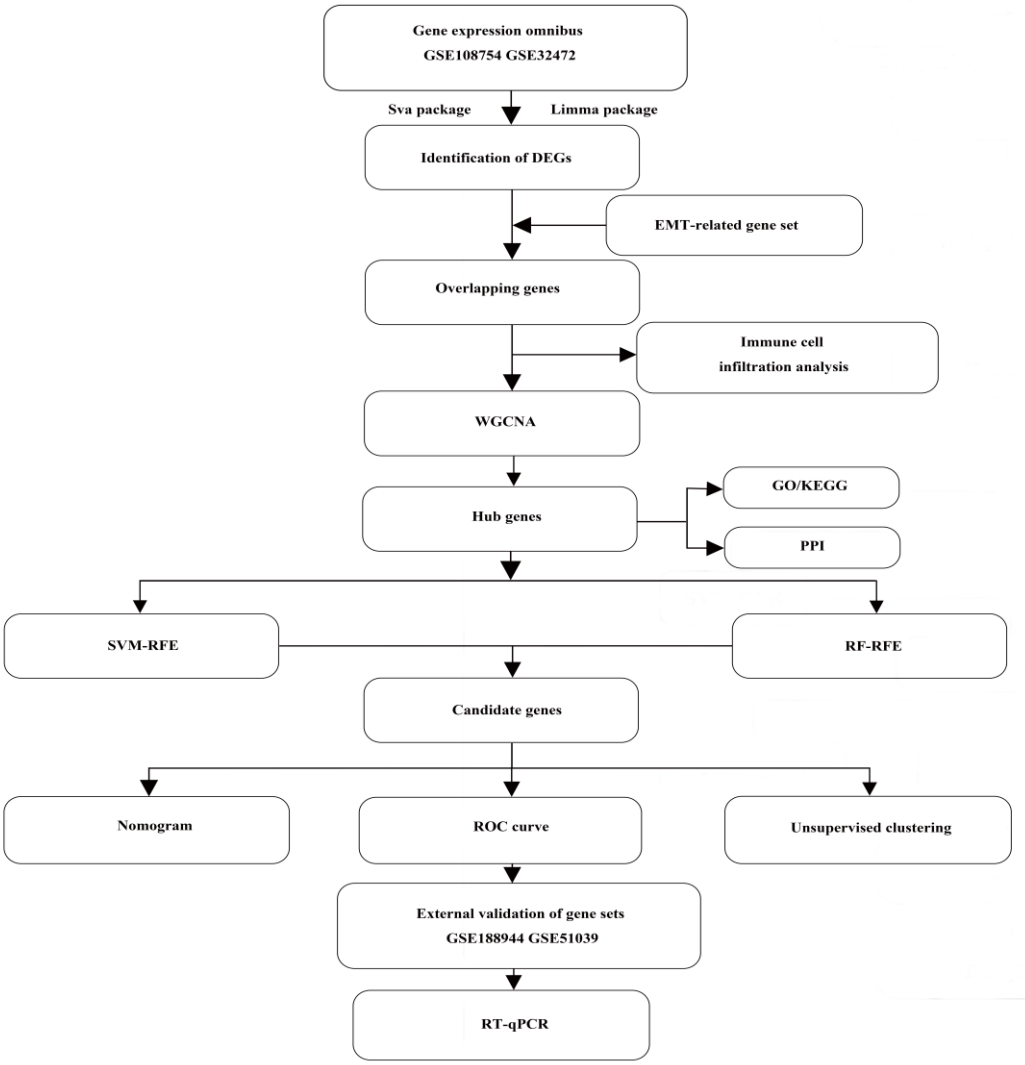


Figure 1: Workflow for the whole study.

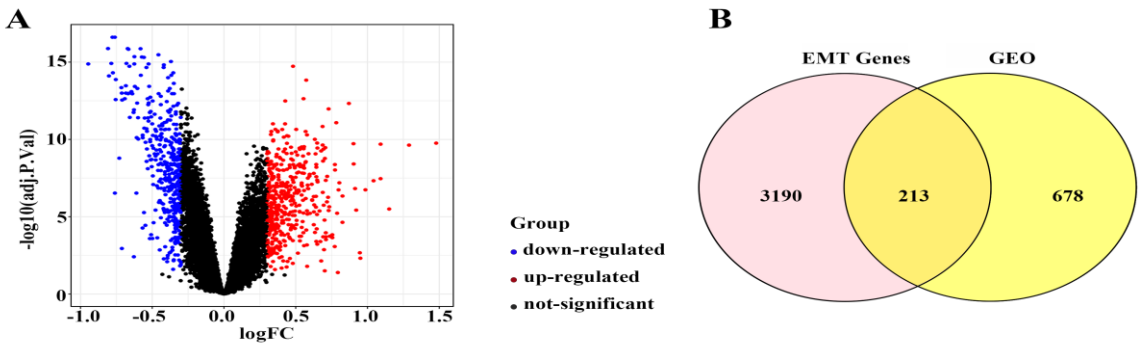


Figure 2: Overlapping genes analysis. (A) Differential gene expressions of BPD are shown in a volcano plot. The X-axis represents logFC and the Y-axis represents -log (adj.*P*.Val). Blue: down-regulated expression genes; Red: up-regulated expression genes; Black: no significant change expression genes. (B)Venn diagram showed the overlapping results of BPD from GEO and GeneCards. A total of 213 EMT-related genes are defined as overlapping genes.

3.2. Evaluation on immune cell infiltration into BPD

With the purpose of investigating the immunological microenvironment of BPD, the relative abundance of 22 immune cell types was estimated using the CIBERSORT algorithm. Then, this study compared the infiltration of these immune cells between non-BPD and BPD samples. The results were visually represented through boxplots and bar charts. Figs. 3A and 3B show the proportions of immune cell infiltration in BPD samples and non-BPD samples. Our analysis indicated that BPD samples exhibited higher infiltration levels of T cells follicular helper, natural killer (NK) cells activated, M0 macrophages, M1 macrophages, resting Mast cells, activated Mast cells, and neutrophils compared with non-BPD samples. By contrast, the infiltration levels of T cells CD8, CD4 naive, CD4 memory resting, regulatory T cells (Tregs), resting NK cells, and activated Dendritic cells were lower in BPD samples. No statistically significant variations were observed in the infiltration levels of the remaining immune cell types between the two groups. These findings suggest a distinct immune microenvironment in BPD compared with non-BPD individuals.

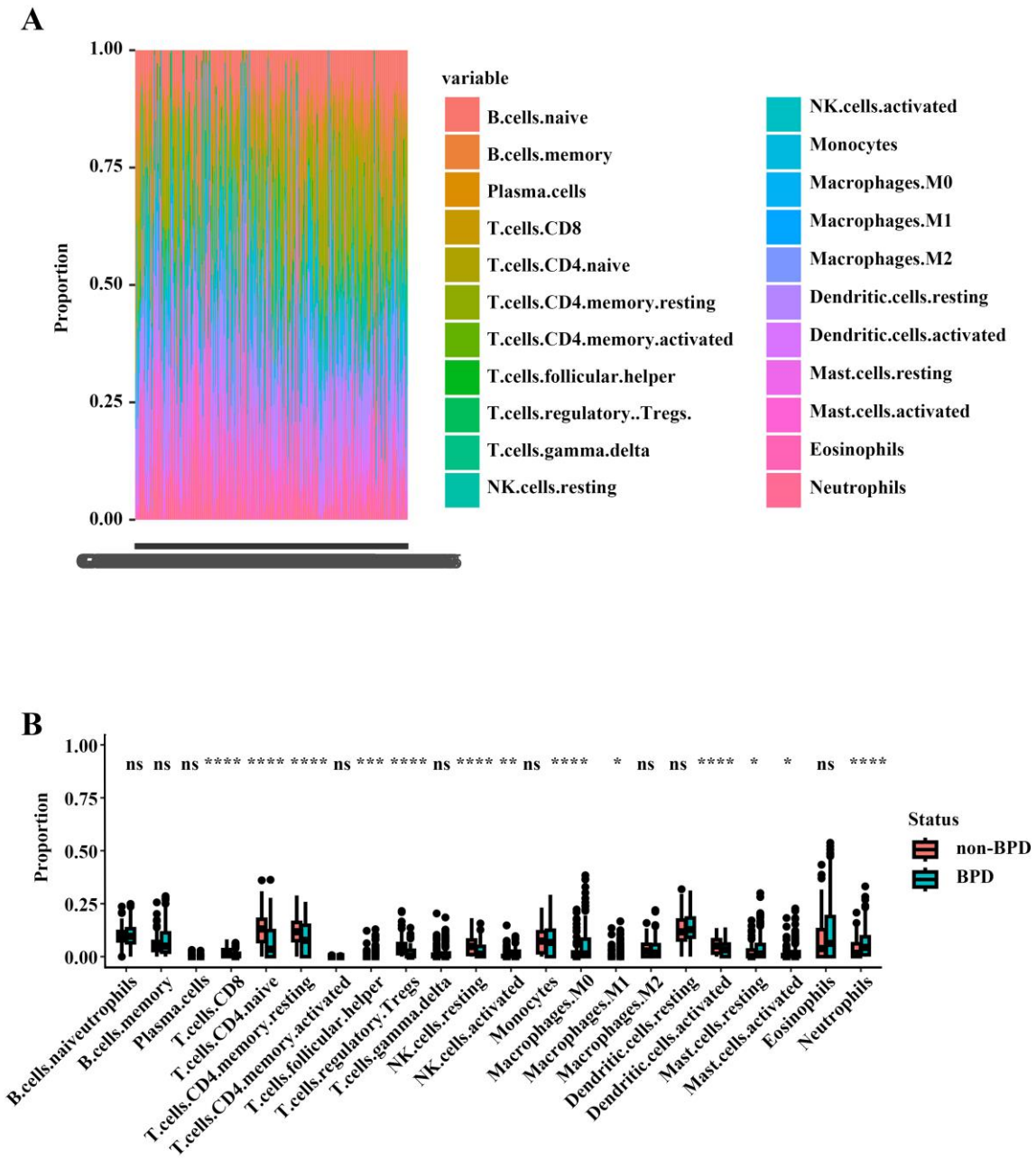


Figure 3: Microenvironmental immune infiltration analysis. (A) bar graph of different types of immune cells in each sample. (B) Comparison of 22 immune cell subtypes between patients in BPD and controls. data are presented as mean ± SD. *P < 0.05, **P < 0.01, ***P < 0.001, ****P<0.0001.

3.3. Construction of the WGCNA network

Through Weighted Gene Co-expression Network Analysis (WGCNA) technique, co-expression modules were built by analyzing 213 genes which were excluded during sample clustering (Supplementary Fig. 2). The process involved step-by-step network construction and module clustering. Based on a scale-free R^2 value of 0.71, a soft-thresholding power of fifteen was chosen (Fig. 4A). The network displayed low average connection and strong independence at this power rating (Fig. 4A), contributing to the use of β = 15 for hierarchical

clustering. Genes with similar expression patterns were grouped into modules using average link clustering (Fig. 4B). By adopting the topological overlap matrix and hierarchical average linkage clustering, distinct gene modules were identified in both the normal and BPD groups, including turquoise, gray, blue, and red modules (Fig. 4C). Based on 119 genes and the highest positive correlation (r=0.4, p=4e-13) with EMT, module turquoise is a vital module for additional research. A significant correlation was observed between the module's eigengene-based connectivity (MM) and gene significance (GS) (cor=0.62, p=5.5e-14) (Fig. 4D). With thresholds of GS > 0.3 and MM > 0.8, 70 and 38 key genes were identified in the turquoise module, respectively. The intersection of genes across modules, GS, and MM revealed 33 hub genes for subsequent analysis.

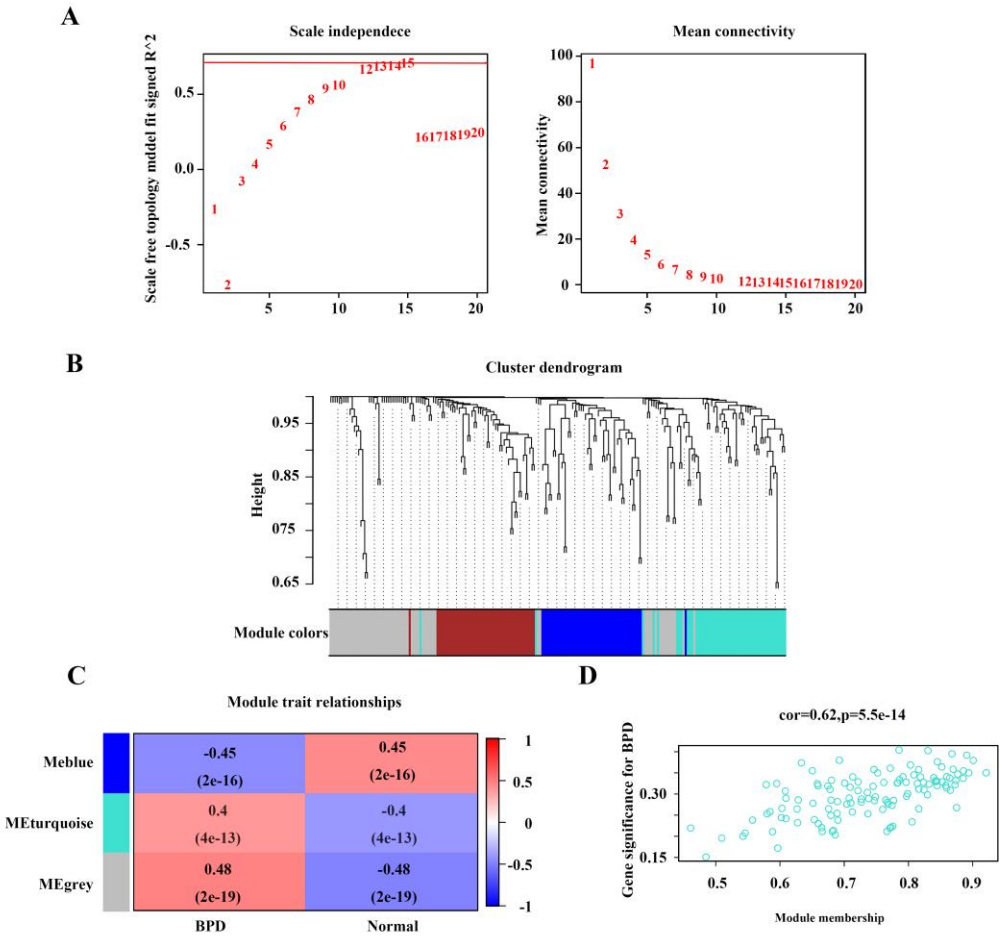


Figure 4: Co-expression network construction and gene module identification. (A) Scale independence and mean connectivity analysis for various soft threshold powers. (B) Clustering dendrograms of genes. The different colors below indicate different co-expression modules. (C) Module–trait relationship. Each row represents a module eigengene and each column represents a trait. Each cell includes the corresponding correlation and p-value. (D) Correlation between gene signatures for BPD and module membership in the turquoise module.

3.4. Identification of feature genes by machine learning algorithms

R software was utilized for machine learning analysis on 33 hub genes, which led to the RF algorithm's discovery of 28 feature genes. The random forest model was optimized with 500 trees, with the minimum error achieving at 74 trees (Fig. 5A). Subsequent evaluation determined the optimal number of trees to be 28 (Fig. 5B). Visualization of feature gene importance was performed using the 'ggplot' package, revealing the top five genes as REPS2, ALOX5, BCL6, SLC2A3, and SVIL (Fig. 5C). In addition, 20 feature genes were identified through the SVM-RFE algorithm, with the final optimal set determined to be 20 genes (Fig. 5D). Combining the results of both algorithms yielded totally seventeen feature genes, including ALOX5, BCL6, SVIL, FLOT2, PRKCD, MMP9, MAPK14, ABHD5, PPP4R1, CD63, CEACAM3, ANXA3, ARG1, AQP9, CXCR1, SEMA4A, and C5AR1(Fig. 5E).

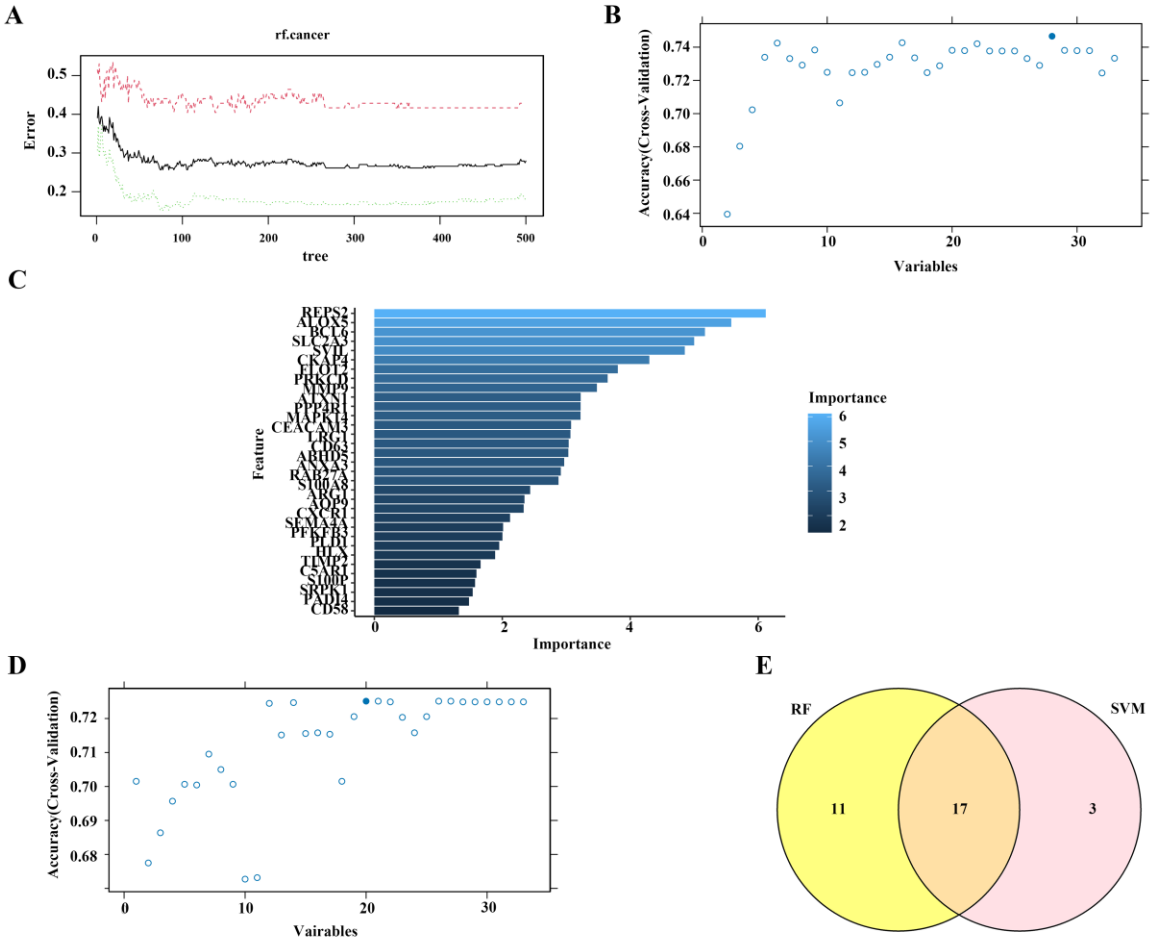


Figure 5: Screening for feature genes. (A) The influence of the number of decision trees on the error rate. The x-axis represents the number of decision trees, and the y-axis indicates the error rate. (B) Biomarker signature gene expression validation by support vector machine recursive feature elimination (RF–RFE) algorithm selection. (C) The top 5 relatively important genes were selected. (D) A plot of gene selection via SVM-RFE algorithm. (E) Two algorithmic Venn diagram screening genes.

3.5. Construction and Evaluation of a Nomogram

After removing the genes that had previously been investigated, 10 potential genes were chosen by a

combination of machine learning and literature review screening. Then, a nomogram model for BPD risk prognosis was then developed utilizing the 'rms' package. As depicted in Fig. 6A, the top three risk factors identified in the model were CXCR1, FLOT2, and AQP9. Subsequently, boxplots and heatmaps were generated using the 'ggpubr' and 'ComplexHeatmap' packages, revealing that all 10 genes were upregulated in BPD samples and statistically significant ($P < 0.0001$) (Figs. 6B and C). Further analysis focused on the genes CXCR1, FLOT2, AQP9, BCL6, PPP4R1, and SEMA4A for future studies.

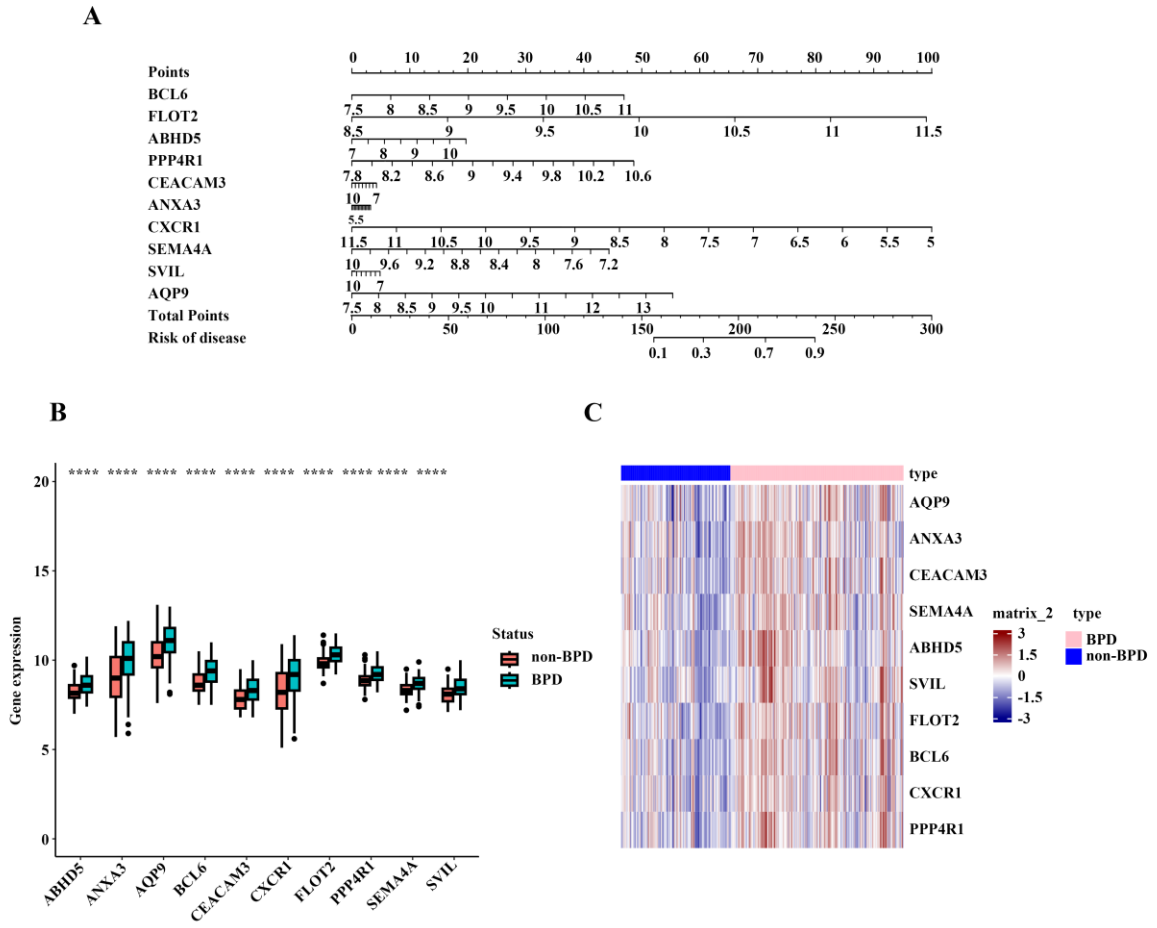


Figure 6: Construction of Nomogram. (A) The nomogram was constructed based on the training cohort. (B-C) Expression of candidate genes in BPD samples.

3.6. ROC Curve Analysis of Candidate Genes

In this study, ROC curve analysis was conducted and visualized using the 'ggplot2' package to enhance the diagnostic potential of candidate genes. In the combined dataset, all six diagnostic genes exhibited AUC values above 0.6 (Fig. 7A). The results of the validation process using the GSE188944 datasets showed that the AUC values of FLOT2, SEMA4A, CXCR1, and AQP9 were 0.755, 0.775, 0.676, and 0.686, respectively (Fig. 7B), while the values of BCL6 and PPP4R1 were 0.588 and 0.48, respectively (Supplementary Fig. 3). Besides, ROC curves were produced for genes in the GSE51039 validation dataset whose AUC values were greater than 0.7 (Supplementary Fig. 4). These findings indicate promising diagnostic capabilities for FLOT2, AQP9, SEMA4A,

and CXCR1 as potential biomarkers for BPD, requiring further validation in future investigations.

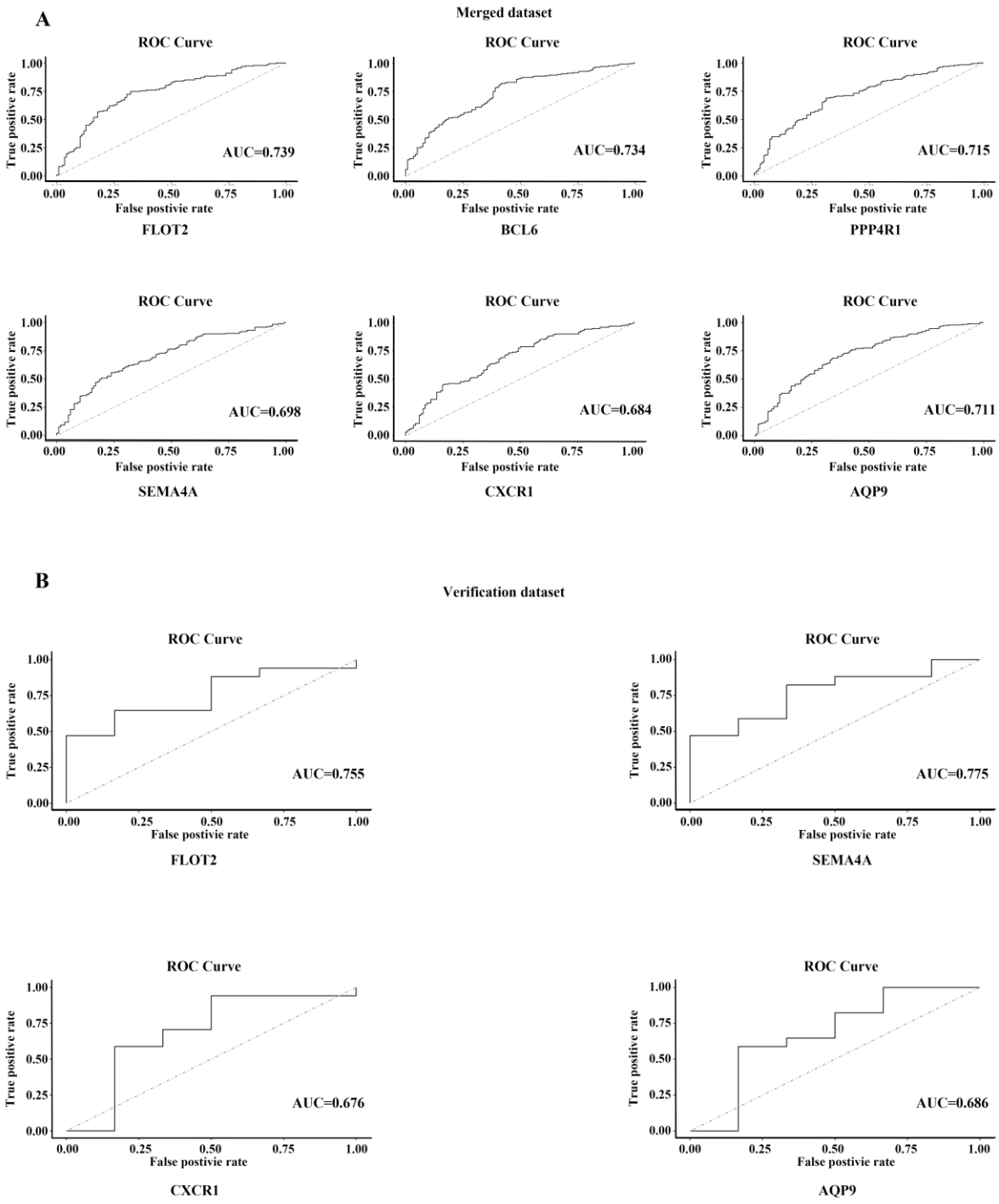


Figure 7: Analysis of the disease-predicting abilities of candidate genes. (A) ROC curve analysis of candidate genes in the merged dataset. (B) ROC curve analysis of hub genes in the verification dataset GSE188944.

3.7. GO and KEGG Analyses on Hub Genes

Utilizing DAVID database, all 33 hub genes were analyzed to identify the GO and KEGG pathways. Next, the

top ten highly ranked genes were then visualized by "ggplot" package in R. Based on enrichment analyses of the 33 hub genes, the top three biological processes (BPs) terms included chemotaxis, response to an organic substance, and defense response to a bacterium (Fig. 8A). Within the cellular component (CC) groups, the hub genes were predominantly associated with plasma membrane, extracellular region, and extracellular exosome (Supplementary Fig. 5A). Furthermore, calcium ion binding, protein binding, and calcium-dependent protein binding were shown to be the main molecular functions (MFs) associated with BPD (Supplementary Fig. 5B). The KEGG pathway analysis indicated significant enrichment of the 33 hub genes in Neutrophil extracellular trap formation, Proteoglycans in cancer, and IL-17 signaling pathways (Fig. 8B).

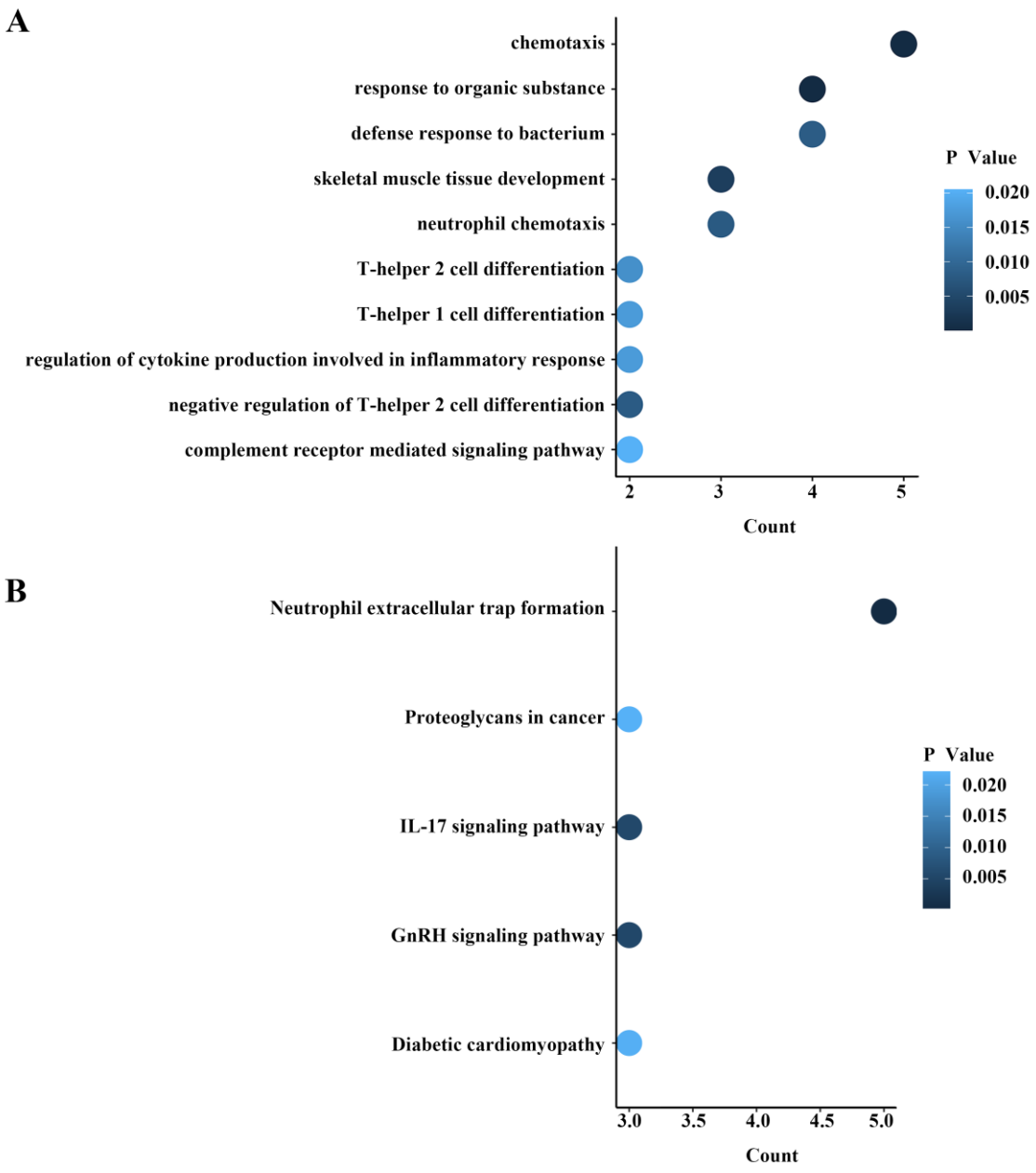


Figure 8: GO and KEGG pathway enrichment analyses were performed on 33 hub genes in BPD. (A) GO biological process (BP)enrichment results. (B) KEGG pathway enrichment results.

3.8. PPI network

To examine the interactions among hub genes, the STRING database was used for creating PPI networks, which were then visualized with Cytoscape software. Fig. 9A illustrates a PPI network with 33 nodes and 76 edges representing hub genes. Afterwards, the data was analyzed using the CytoHubba plugin, assessing the degree of association between hub genes using five distinct calculation methods: EPC, MCC, DMNC, MNC, and Closeness (Fig. 9B-F and Table 3).

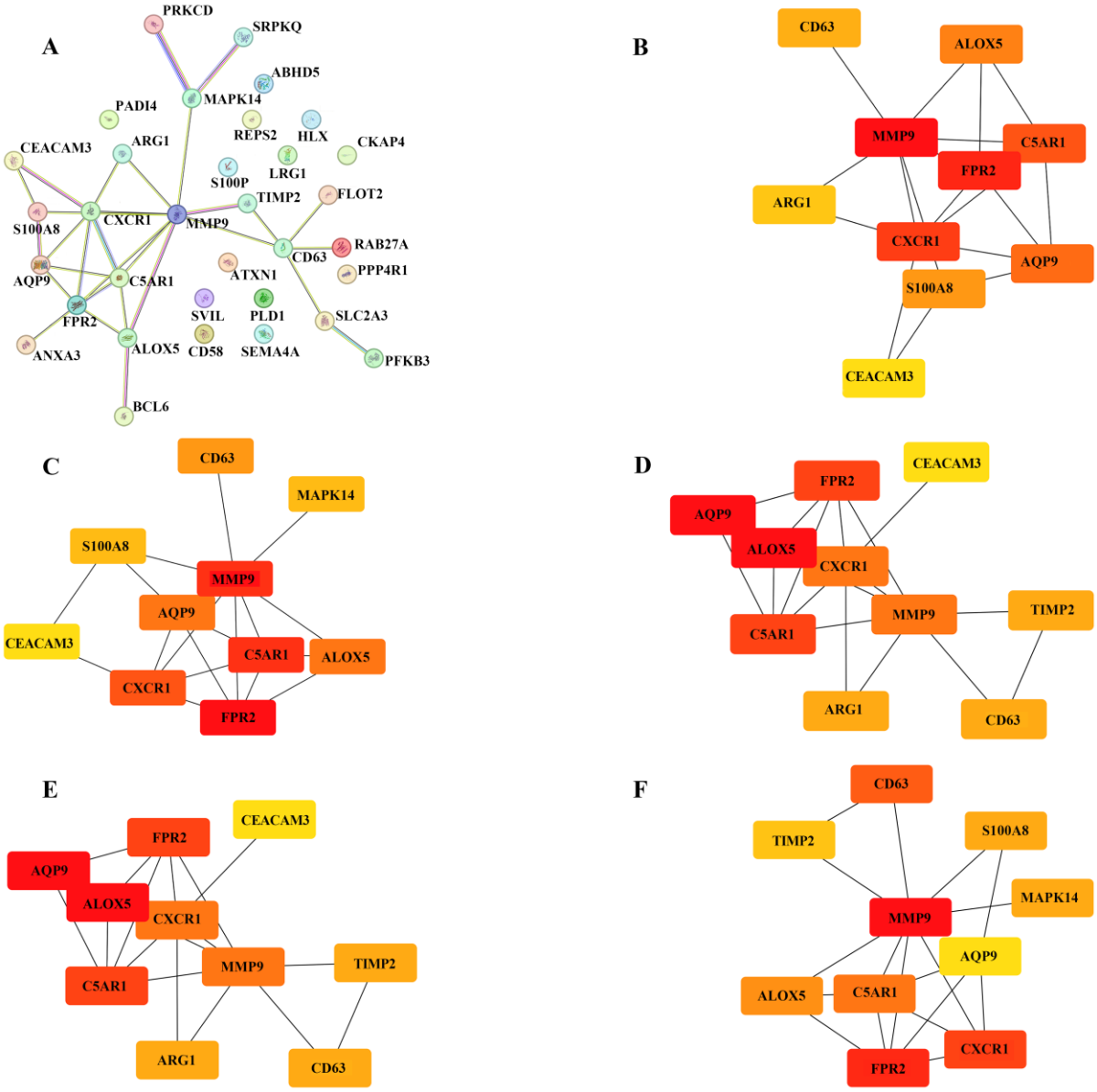


Figure 9: PPI network and interaction analysis of hub genes between the BPD and the nonBPD group.(A)PPI network of common hub genes with 33 nodes and 76 edges which was generated by Cytoscape. (B–F) The hub genes were identified using five topological analysis methods (B) EPC, (C) MCC, (D) DMNC, (E) MNC, and (F)Closeness with cytoHubba. As shown, the node color represented the score, with deeper and redder nodes meaning higher scores.

Table 3: The 5 methods of topology analysis in CytoHubba for the hub genes in BPD

Method	EPC	MCC	DMNC	MNC	Closeness
	MMP9	FPR2	ALOX5	MMP9	MMP9
	CXCR1	MMP9	AQP9	FPR2	FPR2
	FPR2	C5AR1	FPR2	C5AR1	CXCR1
	C5AR1	CXCR1	C5AR1	CXCR1	CD63
Gene	AQP9	ALOX5	MMP9	ALOX5	C5AR1
	ALOX5	AQP9	CXCR1	AQP9	ALOX5
	S100A8	CD63	CD63	CD63	S100A8
	CD63	S100A8	ARG1	ARG1	MAPK14
	ARG1	MAPK14	TIMP2	TIMP2	TIMP2
	TIMP2	CEACAM	CEACAM3	CEACAM3	AQP9

3.9. Classification of candidate genes into two subtypes by unsupervised clustering

Based on the candidate genes selected above, we utilized the 'ConsensusClusterPlus' method for analyzing BPD sample subtypes. In addition, the consistency matrix heat map demonstrated that the samples could be clearly divided into two different classes based on the results (Fig. 10A). Two clusters were identified in a non-linear UMAP space (Fig. 10B).

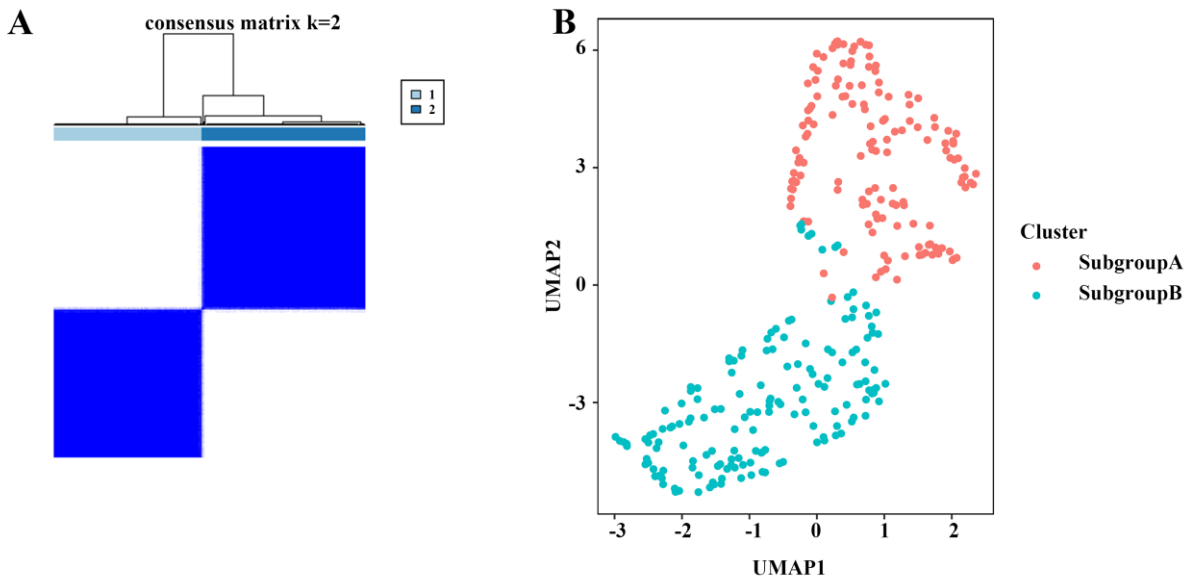


Figure 10: The expression of candidate genes was divided by unsupervised clustering analysis into two subtype samples. (A) Heat map of clusters when consensus K= 2. (B) Data clusters in UMAP spaces.

3.10. Validation of the expression of candidate genes in A549 cells

After 48 hours of hyperoxia treatment, N-Cadherin and α-SMA mRNA levels were upregulated in A549 cells, as shown by RT-qPCR experiments, suggesting hyperoxia-induced EMT (Fig. 11A). Subsequently, the expression patterns of four candidate genes were examined. Fig. 11B illustrates the upregulation of FLOT2, AQP9, and SEMA4A mRNA in hyperoxia-treated A549 cells. However, CXCR1 primers produced ambiguous results with multiple peaks, which could prevent verification of the data.

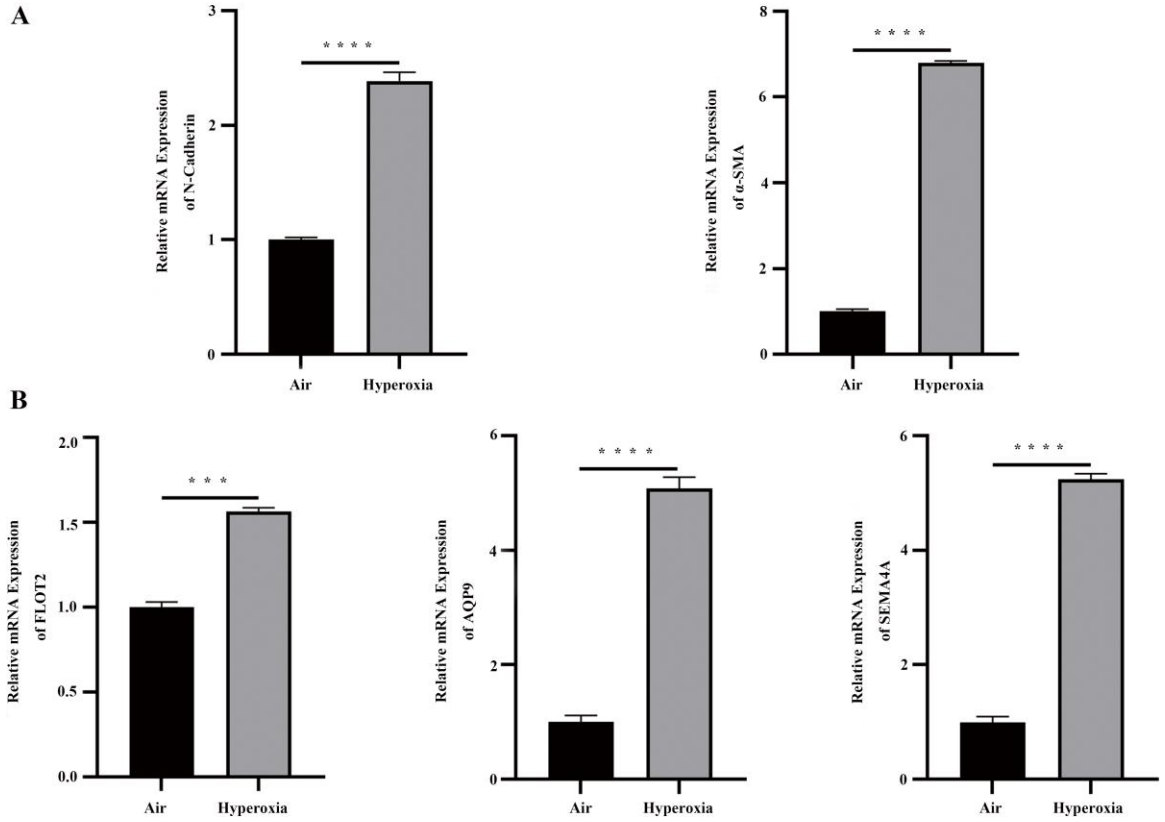


Figure 11: Validation of EMT markers and their candidate genes in A549 cells. (A) N-Cadherin and α-SMA mRNA levels in hyperoxia-treated A549 cells. (B) Candidate genes (FLOT2, AQP9, SEMA4A) mRNA levels in hyperoxia-treated A549 cells. *p<0.05, **p<0.01, ***p<0.001, ****p<0.001.

4. Discussion

BPD has emerged as a prominent pulmonary disease among premature infants, with significant potential for risk assessment and prevention in its treatment[32]. Baraldi and his colleagues emphasized the correlation between BPD development and adrenocortical insufficiency[33]. With Bioinformatics technology, five disease signature genes have been recognized as potential targets for the identification and treatment of BPD[30]. Despite this progress, limited research has investigated the effect of EMT on BPD. Therefore, this study aimed to investigate diagnostic biomarkers within the EMT for BPD. The focus on EMT is crucial considering its role as a key mechanism in disease regulation. According to earlier studies, EMT exerts a vital role in controlling the growth of cancer and predicting a patient's susceptibility to chemotherapy medications[34-36]. Targeting EMT could

potentially provide an effective treatment strategy for a variety of diseases. EMT of lung epithelial cells is associated with BPD pathogenesis, implying its importance in interfering with alveolar development and lung injury repair[9, 37]. Therefore, it is imperative to identify reliable biomarkers to establish the relationship between EMT and BPD. Our preliminary exploration into EMT-related diagnostic biomarkers has identified four potential candidates, exhibiting promising diagnostic value.

Flotillins (FLOTs) are a group of membrane-associated proteins that are part of the larger family of proteins with the stomatin-prohibitin-flotillin-HflK/C (SPFH) domain[38]. Among these proteins, FLOT2 is vital for some biological processes, including apoptosis, actin rearrangement, cell adhesion and migration, and signal transduction[39, 40]. Its increased expression is connected with tumor development and metastasis within several cancers, which included breast cancer, nasopharyngeal carcinoma, renal cancer, gastric cancer, and malignant melanoma[41]. According to studies, FLOT2 inhibits apoptosis and activates certain oncogenic signaling pathways to enhance cell invasion and growth[42]. Furthermore, studies have demonstrated that AEC II cells are capable of transdifferentiating into mesenchymal cells through EMT, which is connected to illnesses like pulmonary fibrosis and brought on by triggers including hypoxia and hyperoxia[37]. Moreover, the molecular and metabolic processes of BPD and its evolution are significantly influenced by programmed cell death (PCD) mechanisms like necrosis, apoptosis, ferroptosis, and autophagy[43]. The paper 'Flotillin-2 Modulates Fas-Signaling Mediated Apoptosis following Hyperoxia in Lung Epithelial Cells' [44] increases the possibility of a connection between FLOT2 and BPD, highlighting the protein’s possible role in BPD diagnosis.

Transmembrane mosaic proteins called aquaporins (AQPs) are extensively dispersed on membranes of cells or organelles in a variety of organs and tissues[45]. Aquaporin 9 (AQP9) is a member of AQP family, serves as the water-selective membrane channel, and exerts a role in cell migration, tumor growth, angiogenesis, immune response, and bactericidal activity[46]. AQP9 expression can be detected in tumor cells and immune cells including neutrophils, macrophages, and T cells[47, 48]. Neutrophils makes a vital impact on the lungs of BPD and are associated with symptom severity[49]. Immune infiltration analysis revealed high expression of neutrophils in BPD, conforming to previous studies. Increased AQP9 expression in BPD suggests potential clinical diagnostic significance.

Semaphorins are a class of proteins that were first discovered in the neurological system as axon guidance factors. They have been associated with the development of the heart, kidney, and immune system, among other organs[50]. Semaphorin 4A (SEMA4A) is a protein of 761 amino acids related to several physiopathological processes including cancer, immunological response, and angiogenesis[51, 52]. It has been indicated that SEMA4A controls stromal cell IL-10 production. In concert with IL-10, SEMA4A induces the EMT, facilitating tumor cell invasion[53]. In addition, SEMA4A is essential for T cell activation and differentiation as well as for the control of the Th1/Th2 immune response[54]. In the context of BPD, a condition characterized by various lung abnormalities, SEMA4A may also be implicated, consistent with previous research findings.

The CXCR1 chemokine receptor, as a seven-transmembrane G-coupled protein receptor involved in inflammatory signaling, recruits and activates leukocytes by binding two ligands, CXCL6 and CXCL8[55]. Numerous cells, including fibroblasts, neutrophils, or vascular endothelial cells, express this receptor[56].

CXCR1 stimulates gastric cancer growth, invasion, migration, and metastasis both in vitro and in vivo[57]. In cases of androgen-independent prostate cancer, depletion of CXCR1 has been shown to decrease angiogenic capacity and tumor growth[58]. Moreover, it has been found that breast cancer stem cells may target CXCR1[59]. Many cellular responses, including cytoskeleton reorganization, morphological alterations, migration, integrin production, reactive oxygen species (ROS) generation, and phagocytosis, are triggered by the activation of CXCR1[60]. This receptor contributes to angiogenesis by playing a vital function in cell migration[61]. Furthermore, data indicate that several cardiovascular diseases (CVDs), like ischemic disorders, atherosclerosis, cardiac remodeling, and hypertension, may be correlated with CXCR1 receptors[61]. Through increasing the synthesis of reduced glutathione (GSH), CXCR1 knockout preserves endothelial cells' redox equilibrium. It lowers lung neutrophil influx and ROS release, which contributes to mitigating hyperoxia-induced acute lung injury (ALI)[62]. Therefore, CXCR1 holds clinical significance in the diagnosis of BPD.

Recent data suggests the vital impact of immune response on BPD pathology[49]. Macrophage levels are significantly higher in BPD patients[63]. T lymphocytes exert an essential role in chronic lung disease occurrence and development among prematurely born infants[64]. CD8+ T cells are widely suggested to facilitate BPD immunopathology while increasing the respiratory morbidity risk[65]. A neutrophil count is the risk factor associated with new BPD occurrence among newborns[66]. From the analysis of 213 overlap genes using the CIBERSORT approach, differences in various immune cells, including NK cells activated, T cells follicular helper, M0 macrophage cells, M1 macrophage cells, neutrophils, activated mast cells, and resting mast cells, were discovered in control compared with BPD samples. These results indicate the potential involvement of immune cells in the pathogenesis of BPD. However, this study still has the certain limitations. The small sample size necessitates a large cohort for validating and strengthening the results. In addition, the analysis is less accurate and thorough because there are no in vivo or further in vitro investigations to confirm gene expression levels. Further validations should be carried out to enhance our result credibility.

5. Conclusion

To conclude, FLOT2, AQP9, SEMA4A, and CXCR1 are identified as potential biomarkers for BPD. In addition, we have created a nomogram model and ROC based on EMT-related biomarkers to predict BPD. It is necessary to perform further research on the functions and regulatory mechanisms of key genes in BPD. The results shed more light on BPD mechanisms and their treatment from an EMT perspective.

Acknowledgments

We thank all authors for their contributions to this study.

6. Conflict of Interest

None declared.

7. Funding

The study was funded by the Innovative Research Program for Graduates of Hubei University of Medicine (No. YC2023031).

References

[1] W. H. Northway, Jr., R. C. Rosan, and D. Y. Porter, "Pulmonary disease following respirator therapy of hyaline-membrane disease. Bronchopulmonary dysplasia," (in eng), *N Engl J Med*, vol. 276, no. 7, pp. 357-68, Feb 16 1967.

[2] Y. Regin, A. Gie, A. Eerdekens, J. Toelen, and A. Debeer, "Ventilation and respiratory outcome in extremely preterm infants: trends in the new millennium," (in eng), *Eur J Pediatr*, vol. 181, no. 5, pp. 1899-1907, May 2022.

[3] J. Y. Islam, R. L. Keller, J. L. Aschner, T. V. Hartert, and P. E. Moore, "Understanding the Short- and Long-Term Respiratory Outcomes of Prematurity and Bronchopulmonary Dysplasia," (in eng), *Am J Respir Crit Care Med*, vol. 192, no. 2, pp. 134-56, Jul 15 2015.

[4] O. D. Saugstad, C. P. Speer, and H. L. Halliday, "Oxygen saturation in immature babies: revisited with updated recommendations," (in eng), *Neonatology*, vol. 100, no. 3, pp. 217-8, 2011.

[5] L. Fu *et al.*, "Reactive oxygen species-evoked endoplasmic reticulum stress mediates 1-nitropyrene-induced epithelial-mesenchymal transition and pulmonary fibrosis," (in eng), *Environ Pollut*, vol. 283, p. 117134, Aug 15 2021.

[6] R. Kalluri and R. A. Weinberg, "The basics of epithelial-mesenchymal transition," (in eng), *J Clin Invest*, vol. 119, no. 6, pp. 1420-8, Jun 2009.

[7] P. Debnath, R. S. Huirem, P. Dutta, and S. Palchaudhuri, "Epithelial-mesenchymal transition and its transcription factors," (in eng), *Biosci Rep*, vol. 42, no. 1, p. BSR20211754, Jan 28 2022.

[8] K. K. Kim *et al.*, "Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix," (in eng), *Proc Natl Acad Sci U S A*, vol. 103, no. 35, pp. 13180-5, Aug 29 2006.

[9] H. Yang *et al.*, "Epithelial-mesenchymal transitions in bronchopulmonary dysplasia of newborn rats," (in eng), *Pediatr Pulmonol*, vol. 49, no. 11, pp. 1112-23, Nov 2014.

[10] X. Gong, J. Qiu, G. Qiu, and C. Cai, "Adrenomedullin regulated by miRNA-574-3p protects premature infants with bronchopulmonary dysplasia," (in eng), *Biosci Rep*, vol. 40, no. 5, pp. BSR20191879,May 29 2020.

[11] J. J. Pietrzyk *et al.*, "Gene expression profiling in preterm infants: new aspects of bronchopulmonary dysplasia development," (in eng), *PLoS One*, vol. 8, no. 10, p. e78585, 2013.

[12] X. Wang *et al.*, "Epigenome-wide association study of bronchopulmonary dysplasia in preterm infants: results from the discovery-BPD program," (in eng), *Clin Epigenetics,* vol. 14, no. 1, p. 57, Apr 28 2022.

[13] K. Lingappan, C. Srinivasan, W. Jiang, L. Wang, X. I. Couroucli, and B. Moorthy, "Analysis of the transcriptome in hyperoxic lung injury and sex-specific alterations in gene expression," (in eng), *PLoS One,* vol. 9, no. 7, p. e101581, 2014.

[14] J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey, "The sva package for removing batch effects and other unwanted variation in high-throughput experiments," (in eng), *Bioinformatics,* vol. 28, no. 6, pp. 882-3, Mar 15 2012.

[15] M. E. Ritchie *et al.*, "limma powers differential expression analyses for RNA-sequencing and microarray studies," (in eng), *Nucleic Acids Res,* vol. 43, no. 7, p. e47, Apr 20 2015.

[16] M. Safran *et al.*, "GeneCards Version 3: the human gene integrator," (in eng), *Database (Oxford),* vol. 2010, p. baq020, Aug 5 2010.

[17] G. Stelzer *et al.*, "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses," (in eng), *Curr Protoc Bioinformatics,* vol. 54, pp. 1.30.1-1.30.33, Jun 20 2016.

[18] A. M. Newman *et al.*, "Robust enumeration of cell subsets from tissue expression profiles," (in eng), *Nat Methods,* vol. 12, no. 5, pp. 453-7, May 2015.

[19] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, and A. A. Alizadeh, "Profiling Tumor Infiltrating Immune Cells with CIBERSORT," (in eng), *Methods Mol Biol,* vol. 1711, pp. 243-259, 2018.

[20] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," (in eng), *BMC Bioinformatics,* vol. 9, p. 559, Dec 29 2008.

[21] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," (in eng), *Stat Appl Genet Mol Biol,* vol. 4, p. Article17, 2005.

[22] Y. Hu, J. Han, S. Ding, S. Liu, and H. Wang, "Identification of ferroptosis-associated biomarkers for the potential diagnosis and treatment of postmenopausal osteoporosis," (in eng), *Front Endocrinol (Lausanne),* vol. 13, p. 986384, 2022.

[23] G. Karami, M. Giuseppe Orlando, A. Delli Pizzi, M. Caulo, and C. Del Gratta, "Predicting Overall Survival Time in Glioblastoma Patients Using Gradient Boosting Machines Algorithm and Recursive Feature Elimination Technique," (in eng), *Cancers (Basel),* vol. 13, no. 19, p.4976, Oct 4 2021.

[24] V. N. Vapnik, "An overview of statistical learning theory," (in eng), *IEEE Trans Neural Netw,* vol. 10,

no. 5, pp. 988-99, 1999.

[25] B. T. Sherman *et al.*, "DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis," (in eng), *BMC Bioinformatics*, vol. 8, p. 426, Nov 2 2007.

[26] D. Otasek, J. H. Morris, J. Bouças, A. R. Pico, and B. Demchak, "Cytoscape Automation: empowering workflow-based network analysis," (in eng), *Genome Biol*, vol. 20, no. 1, p. 185, Sep 2 2019.

[27] N. Bhattacharyya *et al.*, "CDK1 and HSP90AA1 Appear as the Novel Regulatory Genes in Non-Small Cell Lung Cancer: A Bioinformatics Approach," (in eng), *J Pers Med*, vol. 12, no. 3, p. 393, Mar 4 2022.

[28] C. Y. Lin, C. H. Chin, H. H. Wu, S. H. Chen, C. W. Ho, and M. T. Ko, "Hubba: hub objects analyzer--a framework of interactome hubs identification for network biology," (in eng), *Nucleic Acids Res*, vol. 36, no. Web Server issue, pp. W438-43, Jul 1 2008.

[29] M. D. Wilkerson and D. N. Hayes, "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking," (in eng), *Bioinformatics*, vol. 26, no. 12, pp. 1572-3, Jun 15 2010.

[30] M. Jia *et al.*, "Identification and validation of cuproptosis related genes and signature markers in bronchopulmonary dysplasia disease using bioinformatics analysis and machine learning," (in eng), *BMC Med Inform Decis Mak*, vol. 23, no. 1, p. 69, Apr 14 2023.

[31] V. Grollemund *et al.*, "Development and validation of a 1-year survival prognosis estimation model for Amyotrophic Lateral Sclerosis using manifold learning algorithm UMAP," (in eng), *Sci Rep*, vol. 10, no. 1, p. 13378, Aug 7 2020.

[32] S. H. Wang and P. N. Tsao, "Phenotypes of Bronchopulmonary Dysplasia," (in eng), *Int J Mol Sci*, vol. 21, no. 17, p.6112,Aug 25 2020.

[33] E. Baraldi *et al.*, "Untargeted Metabolomic Analysis of Amniotic Fluid in the Prediction of Preterm Delivery and Bronchopulmonary Dysplasia," (in eng), *PLoS One*, vol. 11, no. 10, p. e0164211, 2016.

[34] A. Dongre and R. A. Weinberg, "New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer," (in eng), *Nat Rev Mol Cell Biol*, vol. 20, no. 2, pp. 69-84, Feb 2019.

[35] R. Y. Huang, V. Y. Chung, and J. P. Thiery, "Targeting pathways contributing to epithelial-mesenchymal transition (EMT) in epithelial ovarian cancer," (in eng), *Curr Drug Targets*, vol. 13, no. 13, pp. 1649-53, Dec 2012.

[36] C. Li *et al.*, "Exploration of epithelial-mesenchymal transition-related lncRNA signature and drug

sensitivity in breast cancer," (in eng), *Front Endocrinol (Lausanne)*, vol. 14, p. 1154741, 2023.

[37] H. Y. Lu *et al.*, "ILC2 influence the differentiation of alveolar type II epithelial cells in bronchopulmonary dysplasia mice," (in eng), *J Leukoc Biol*, vol. 114, no. 6, pp. 604-614, Nov 24 2023.

[38] H. Yokoyama, S. Fujii, and I. Matsui, "Crystal structure of a core domain of stomatin from Pyrococcus horikoshii Illustrates a novel trimeric and coiled-coil fold," (in eng), *J Mol Biol*, vol. 376, no. 3, pp. 868-78, Feb 22 2008.

[39] A. Banning, A. Tomasovic, and R. Tikkanen, "Functional aspects of membrane association of reggie/flotillin proteins," (in eng), *Curr Protein Pept Sci*, vol. 12, no. 8, pp. 725-35, Dec 2011.

[40] E. A. R. Morris *et al.*, "Flotillins control zebrafish epiboly through their role in cadherin-mediated cell-cell adhesion," (in eng), *Biol Cell*, vol. 109, no. 5, pp. 210-221, May 2017.

[41] Y. L. Wang, W. J. Yao, L. Guo, H. F. Xi, S. Y. Li, and Z. M. Wang, "Expression of flotillin-2 in human non-small cell lung cancer and its correlation with tumor progression and patient survival," (in eng), *Int J Clin Exp Pathol*, vol. 8, no. 1, pp. 601-7, 2015.

[42] T. Song, Z. Hu, J. Liu, and W. Huang, "FLOT2 upregulation promotes growth and invasion by interacting and stabilizing EphA2 in gliomas," (in eng), *Biochem Biophys Res Commun*, vol. 548, pp. 67-73, Apr 9 2021.

[43] X. Deng *et al.*, "Molecular mechanisms of cell death in bronchopulmonary dysplasia," (in eng), *Apoptosis*, vol. 28, no. 1-2, pp. 39-54, Feb 2023.

[44] S. Wei, H. G. Moon, Y. Zheng, X. Liang, C. H. An, and Y. Jin, "Flotillin-2 modulates fas signaling mediated apoptosis after hyperoxia in lung epithelial cells," (in eng), *PLoS One*, vol. 8, no. 10, p. e77519, 2013.

[45] Y. Qian *et al.*, "AQP9 suppresses hepatocellular carcinoma cell invasion through inhibition of hypoxia-inducible factor 1α expression under hypoxia," (in eng), *J Gastroenterol Hepatol*, vol. 35, no. 11, pp. 1990-1997, Nov 2020.

[46] B. Nico and D. Ribatti, "Role of aquaporins in cell migration and edema formation in human brain tumors," (in eng), *Exp Cell Res*, vol. 317, no. 17, pp. 2391-6, Oct 15 2011.

[47] A. Holm, K. E. Magnusson, and E. Vikström, "Pseudomonas aeruginosa N-3-oxo-dodecanoyl-homoserine Lactone Elicits Changes in Cell Volume, Morphology, and AQP9 Characteristics in Macrophages," (in eng), *Front Cell Infect Microbiol*, vol. 6, p. 32, 2016.

[48] T. Karlsson, M. Glogauer, R. P. Ellen, V. M. Loitto, K. E. Magnusson, and M. A. Magalhães, "Aquaporin 9 phosphorylation mediates membrane localization and neutrophil polarization," (in eng), *J*

Leukoc Biol, vol. 90, no. 5, pp. 963-73, Nov 2011.

[49] M. Heydarian, C. Schulz, T. Stoeger, and A. Hilgendorff, "Association of immune cell recruitment and BPD development," (in eng), *Mol Cell Pediatr,* vol. 9, no. 1, p. 16, Aug 2 2022.

[50] U. Yazdani and J. R. Terman, "The semaphorins," (in eng), *Genome Biol,* vol. 7, no. 3, p. 211, 2006.

[51] E. Nkyimbeng-Takwi and S. P. Chapoval, "Biology and function of neuroimmune semaphorins 4A and 4D," (in eng), *Immunol Res,* vol. 50, no. 1, pp. 10-21, May 2011.

[52] D. Ito and A. Kumanogoh, "The role of Sema4A in angiogenesis, immune responses, carcinogenesis, and retinal systems," (in eng), *Cell Adh Migr,* vol. 10, no. 6, pp. 692-699, Nov 2016.

[53] X. Liu *et al.*, "SEMA4A promotes prostate cancer invasion: involvement of tumor microenvironment," (in eng), *J Cancer,* vol. 14, no. 14, pp. 2633-2643, 2023.

[54] H. Kayama *et al.*, "BATF2 prevents T-cell-mediated intestinal inflammation through regulation of the IL-23/IL-17 pathway," (in eng), *Int Immunol,* vol. 31, no. 6, pp. 371-383, May 21 2019.

[55] C. Molczyk and R. K. Singh, "CXCR1: A Cancer Stem Cell Marker and Therapeutic Target in Solid Tumors," (in eng), *Biomedicines,* vol. 11, no. 2, p. 576, Feb 16 2023.

[56] K. H. Susek, M. Karvouni, E. Alici, and A. Lundqvist, "The Role of CXC Chemokine Receptors 1-4 on Immune Cells in the Tumor Microenvironment," (in eng), *Front Immunol,* vol. 9, p. 2159, 2018.

[57] J. Wang *et al.*, "CXCR1 promotes malignant behavior of gastric cancer cells in vitro and in vivo in AKT and ERK1/2 phosphorylation," (in eng), *Int J Oncol,* vol. 48, no. 5, pp. 2184-96, May 2016.

[58] N. Shamaladevi, D. A. Lyn, D. O. Escudero, and B. L. Lokeshwar, "CXC receptor-1 silencing inhibits androgen-independent prostate cancer," (in eng), *Cancer Res,* vol. 69, no. 21, pp. 8265-74, Nov 1 2009.

[59] P. A. Ruffini, "The CXCL8-CXCR1/2 Axis as a Therapeutic Target in Breast Cancer Stem-Like Cells," (in eng), *Front Oncol,* vol. 9, p. 40, 2019.

[60] H. Ha, B. Debnath, and N. Neamati, "Role of the CXCL8-CXCR1/2 Axis in Cancer and Inflammatory Diseases," (in eng), *Theranostics,* vol. 7, no. 6, pp. 1543-1588, 2017.

[61] K. Dhayni, K. Zibara, H. Issa, S. Kamel, and Y. Bennis, "Targeting CXCR1 and CXCR2 receptors in cardiovascular diseases," (in eng), *Pharmacol Ther,* vol. 237, p. 108257, Sep 2022.

[62] H. Qin *et al.*, "Targeting CXCR1 alleviates hyperoxia-induced lung injury through promoting glutamine metabolism," (in eng), *Cell Rep,* vol. 42, no. 7, p. 112745, Jul 25 2023.

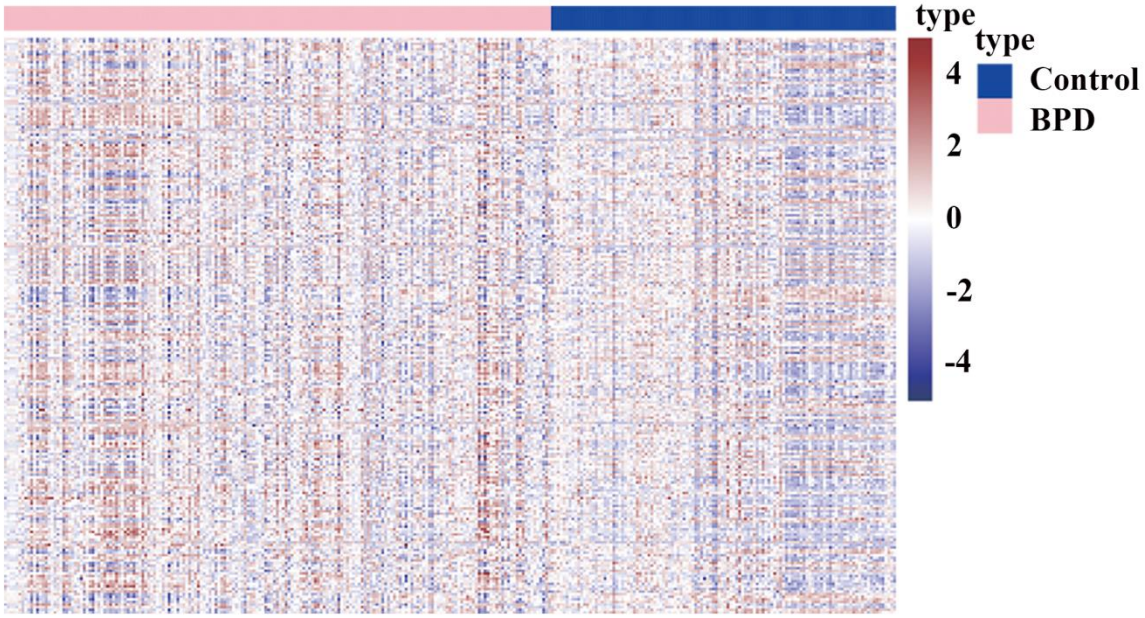
[63] J. D. Jacobson, W. E. Truog, and D. R. Benjamin, "Increased expression of human leukocyte antigen-DR on pulmonary macrophages in bronchopulmonary dysplasia," (in eng), *Pediatr Res*, vol. 34, no. 3, pp. 341-4, Sep 1993.

[64] G. Toldi, H. Hummler, and T. Pillay, "T Lymphocytes, Multi-Omic Interactions and Bronchopulmonary Dysplasia," (in eng), *Front Pediatr*, vol. 9, p. 694034, 2021.

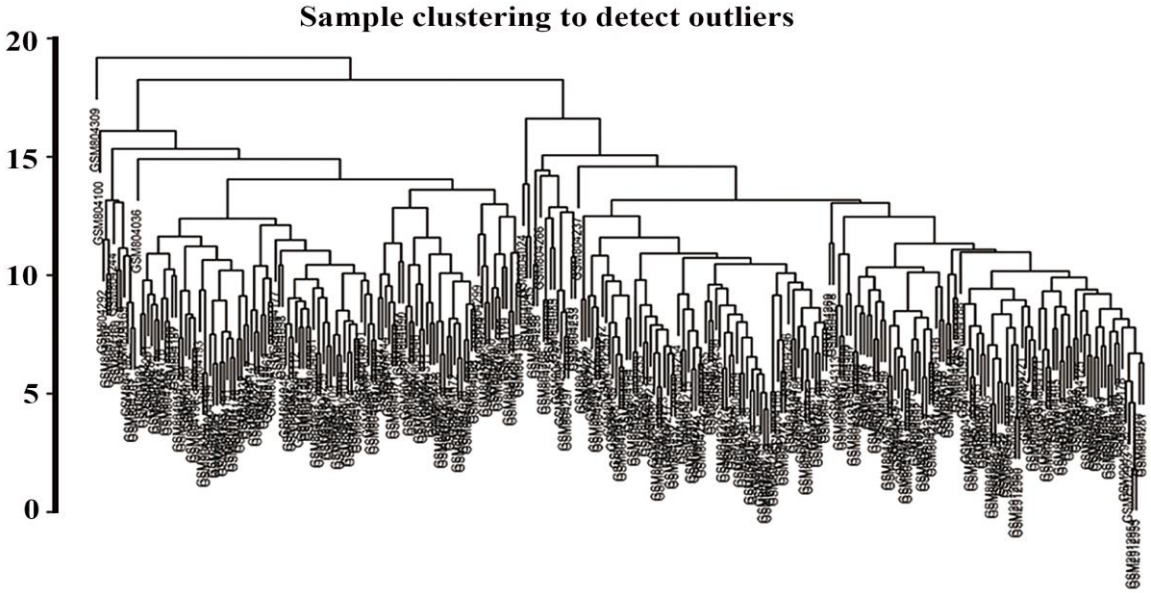
[65] K. M. Scheible *et al.*, "T cell developmental arrest in former premature infants increases risk of respiratory morbidity later in infancy," (in eng), *JCI Insight*, vol. 3, no. 4, p. e96724, Feb 22 2018.

[66] B. I. Kim *et al.*, "Increase in cord blood soluble E-selectin and tracheal aspirate neutrophils at birth and the development of new bronchopulmonary dysplasia," (in eng), *J Perinat Med*, vol. 32, no. 3, pp. 282-7, 2004.

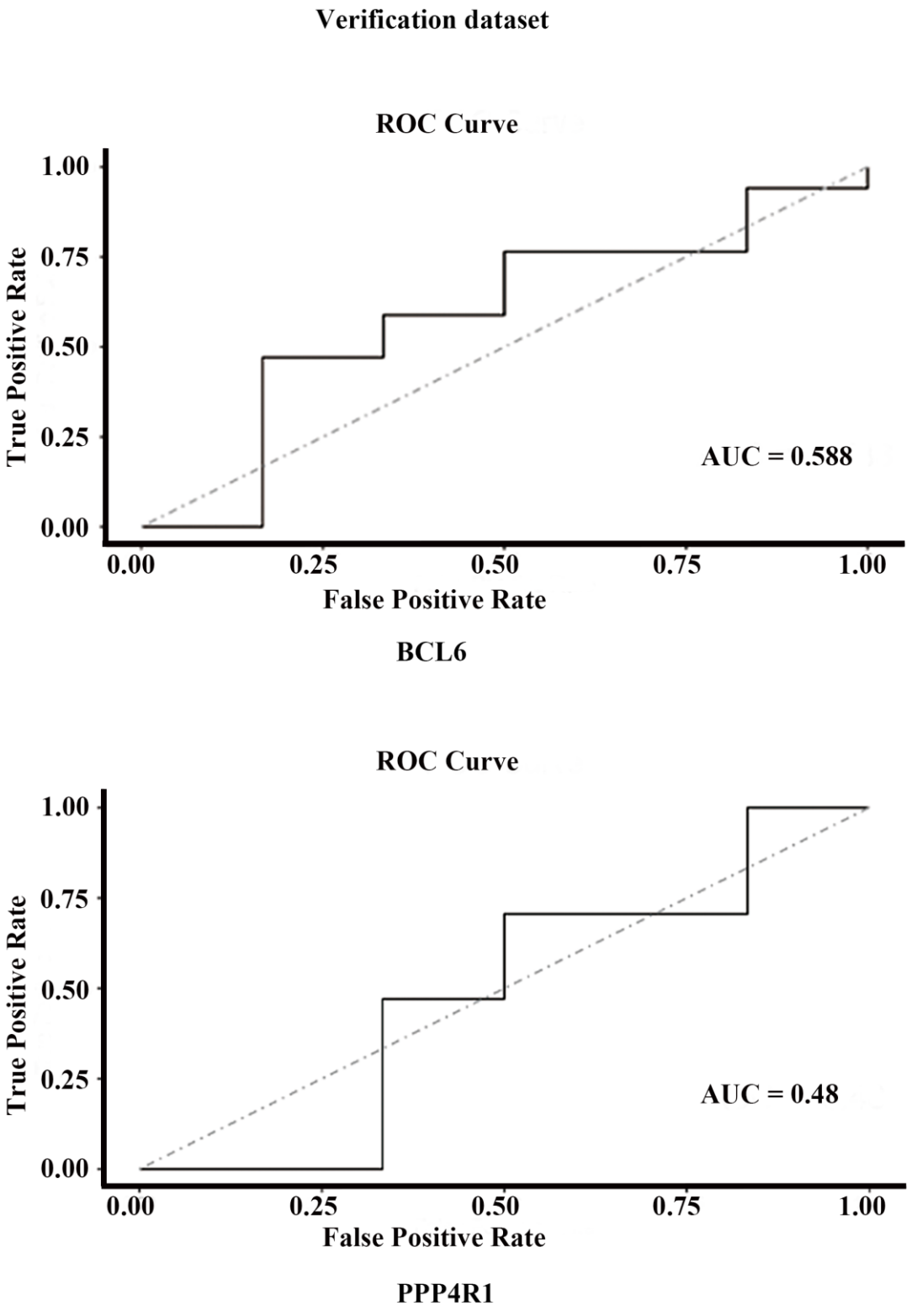
Appendix



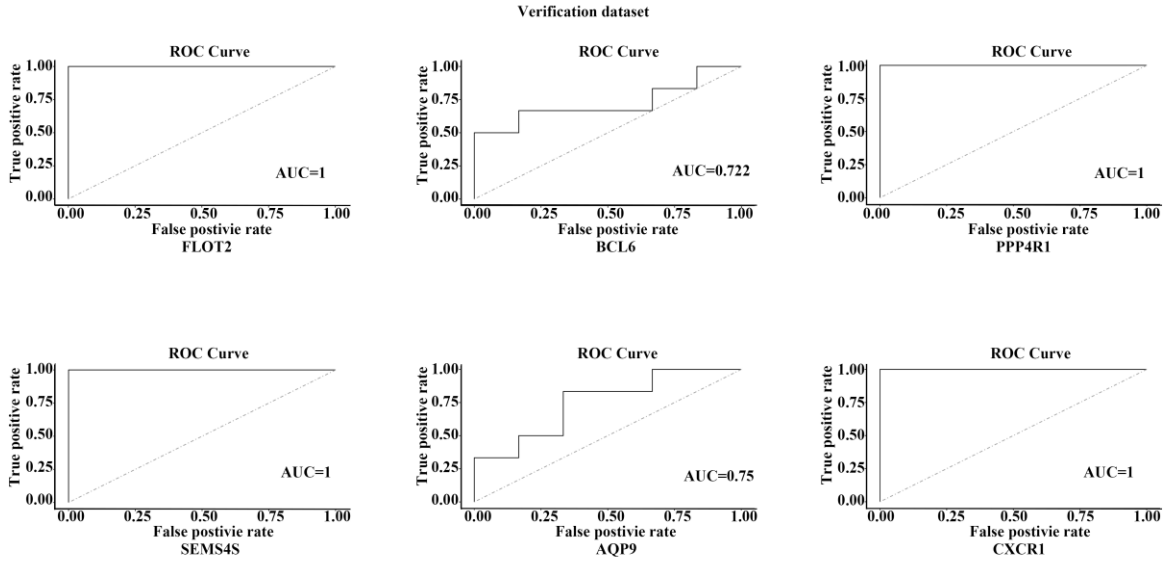
Supplementary Figure 1: Differential expressed genes analysis. Heatmap displaying different expressions of the DEGs. Red indicates a higher expression and blue represents a lower expression.



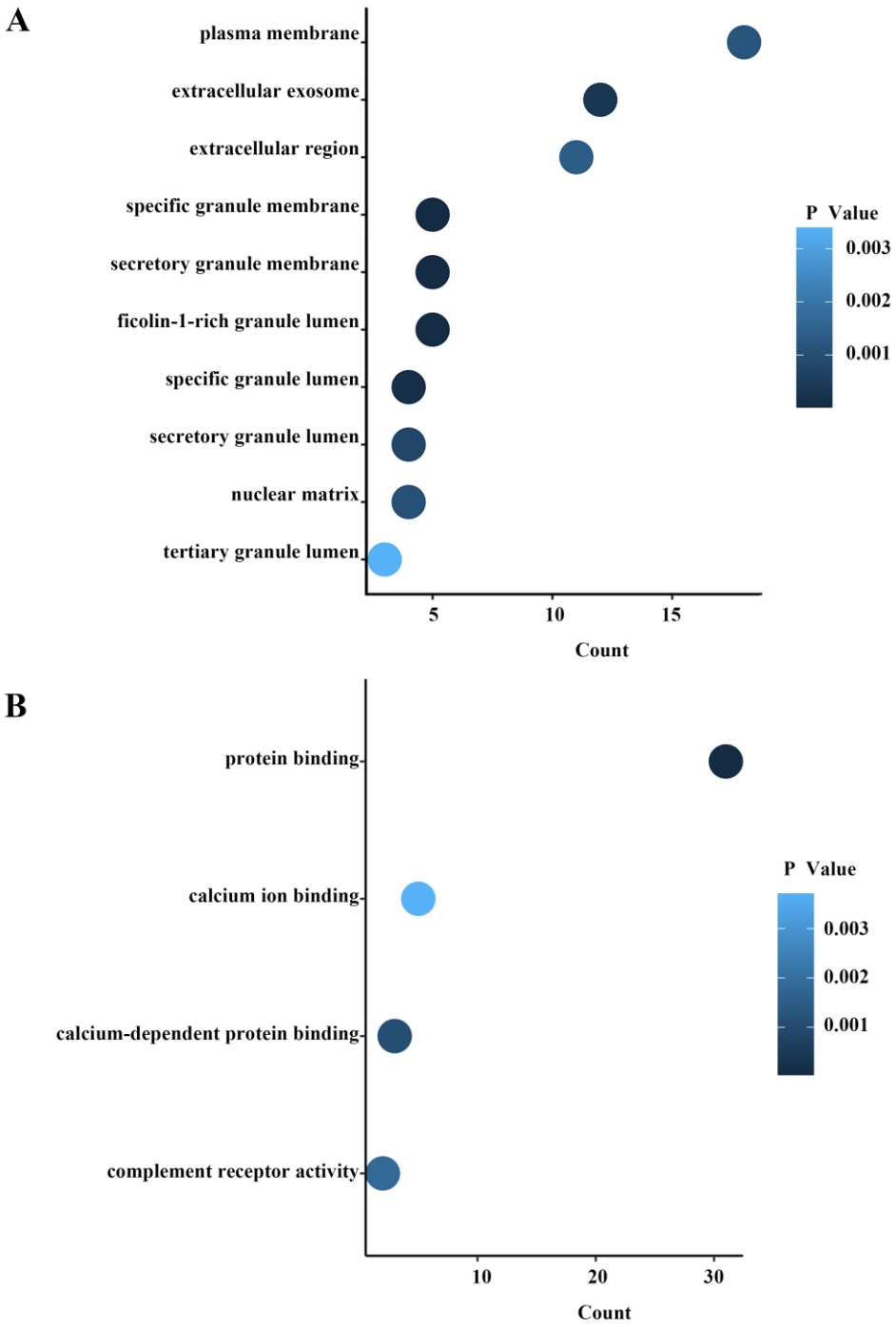
Supplementary Figure 2: Co-expression network construction and gene module identification. clustering dendrogram of 305 samples.



Supplementary Figure 3: Analysis of the disease-predicting abilities of candidate genes. ROC curve analysis of hub genes in the verification dataset 188944.



Supplementary Figure 4: Analysis of the disease-predicting abilities of candidate genes. ROC curve analysis of hub genes in the verification dataset GSE51039.



Supplementary Figure 5: GO and KEGG pathway enrichment analyses were performed on 33 hub genes in BPD. (A) GO cellular component (CC) enrichment results. (B) GO molecular function (MF) enrichment results.