



## OPEN ACCESS

## EDITED BY

Xiao Chang,  
Children's Hospital of Philadelphia,  
United States

## REVIEWED BY

Pengyuan Du,  
Shandong First Medical University and  
Shandong Academy of Medical Science,  
China  
Yingchao Song,  
Shandong First Medical University and  
Shandong Academy of Medical Sciences,  
China

## \*CORRESPONDENCE

Yu Xu,  
✉ xuyu01@xinhuamed.com.cn

RECEIVED 15 August 2023

ACCEPTED 30 October 2023

PUBLISHED 10 November 2023

## CITATION

Chen H, Chen E, Lu Y and Xu Y (2023),  
Identification of immune-related genes in  
diagnosing retinopathy of prematurity  
with sepsis through bioinformatics  
analysis and machine learning.  
*Front. Genet.* 14:1264873.  
doi: 10.3389/fgene.2023.1264873

## COPYRIGHT

© 2023 Chen, Chen, Lu and Xu. This is an  
open-access article distributed under the  
terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Identification of immune-related genes in diagnosing retinopathy of prematurity with sepsis through bioinformatics analysis and machine learning

Han Chen<sup>1</sup>, Enguang Chen<sup>1</sup>, Yao Lu<sup>2</sup> and Yu Xu<sup>1\*</sup>

<sup>1</sup>Department of Ophthalmology, School of Medicine, Xinhua Hospital, Shanghai Jiao Tong University, Shanghai, China, <sup>2</sup>Department of Ophthalmology, Anhui No. 2 Provincial People's Hospital, Anhui, Hefei, China

**Background:** There is increasing evidence indicating that immune system dysregulation plays a pivotal role in the pathogenesis of retinopathy of prematurity (ROP) and sepsis. This study aims to identify key diagnostic candidate genes in ROP with sepsis.

**Methods:** We obtained publicly available data on ROP and sepsis from the gene expression omnibus database. Differential analysis and weighted gene correlation network analysis (WGCNA) were performed to identify differentially expressed genes (DEGs) and key module genes. Subsequently, we conducted functional enrichment analysis to gain insights into the biological functions and pathways. To identify immune-related pathogenic genes and potential mechanisms, we employed several machine learning algorithms, including Support Vector Machine Recursive Feature Elimination (SVM-RFE), Least Absolute Shrinkage and Selection Operator (LASSO), and Random Forest (RF). We evaluated the diagnostic performance using nomogram and Receiver Operating Characteristic (ROC) curves. Furthermore, we used CIBERSORT to investigate immune cell dysregulation in sepsis and performed cMAP analysis to identify potential therapeutic drugs.

**Results:** The sepsis dataset comprised 352 DEGs, while the ROP dataset had 307 DEGs and 420 module genes. The intersection between DEGs for sepsis and module genes for ROP consisted of 34 genes, primarily enriched in immune-related pathways. After conducting PPI network analysis and employing machine learning algorithms, we pinpointed five candidate hub genes. Subsequent evaluation using nomograms and ROC curves underscored their robust diagnostic potential. Immune cell infiltration analysis revealed immune cell dysregulation. Finally, through cMAP analysis, we identified some small molecule compounds that have the potential for sepsis treatment.

**Conclusion:** Five immune-associated candidate hub genes (CLEC5A, KLRB1, LCN2, MCEMP1, and MMP9) were recognized, and the nomogram for the diagnosis of ROP with sepsis was developed.

## KEYWORDS

retinopathy of prematurity, sepsis, machine learning, immune infiltration, diagnostic value

## 1 Introduction

Sepsis is a severe systemic infection characterized by a dysregulated host response, leading to organ dysfunction to varying degrees (Singer et al., 2016). Furthermore, according to population-level studies over the past 2 decades, the occurrence of neonatal sepsis was recorded at 2.2%, with mortality rates spanning from 11% to 19% (Liu et al., 2016; Dong et al., 2020). In summary, sepsis stands as one of the primary contributors to hospital-related mortality (Rhee et al., 2019). Additionally, it imposes significant economic burden and psychological stress on patients and their families.

Retinopathy of prematurity (ROP) constituted a significant factor that led to visual impairment and blindness in children, characterized by abnormal retinal neovascularization in premature infants (Rivera et al., 2011). It was recognized as an “iatrogenic disease” following the introduction of supplemental oxygen and incubators, which made survival of premature infants possible (Dogra et al., 2017). The incidence and severity of ROP were impacted by multiple factors, such as birth weight, gestational age, oxygen exposure, bronchopulmonary dysplasia, blood transfusions, as well as several systemic risk factors (Wu et al., 2018).

Inflammation is widely recognized as a significant factor in the development of ROP. Furthermore, in the oxygen-induced retinopathy model, it has been demonstrated that systemic inflammation in newborns disrupts the development of retinal blood vessels and induces the pathological features of ROP (Tremblay et al., 2013; Hong et al., 2014). Premature infants are at increased risk of developing ROP due to incomplete development, especially in the vascular system of the eyes (Sato et al., 2009). Sepsis, as a systemic infection, can lead to vascular inflammation and endothelial dysfunction, thereby increasing the risk of ROP in premature infants (Lee and Dammann, 2012). Premature infants with low birth weight were more susceptible to infections, and there was evidence of an increased risk of ROP in these cases. Several studies have demonstrated an association between the incidence of early-onset sepsis in premature infants and an elevated risk as well as severity of ROP (Huang et al., 2019; Bonafiglia et al., 2022; Shen et al., 2022). This indicates that infections and inflammatory responses may have a vital role in the pathogenesis of ROP. Currently, the mechanisms underlying the role of inflammation in ROP accompanied by sepsis are still under investigation.

In recent years, bioinformatics analysis has found widespread applications in exploring potential pathogenic mechanisms, identify therapeutic targets, and discover small molecule drugs for various diseases. In our study, we utilized sepsis and ROP datasets downloaded from GEO to identify differentially expressed genes (DEGs) and important module genes through weighted gene co-expression network analysis (WGCNA). We conducted enrichment analysis on intersecting genes, constructed PPI networks, and applied machine learning algorithms, including LASSO, RF, and SVM-RFE. We performed immune cell infiltration analysis and Connectivity map (cMAP) analysis. Furthermore, we assessed the essential immune-related diagnostic genes for sepsis combined with ROP using nomogram and ROC curve analysis. This research aims to identify potential immune-related diagnostic biomarkers for ROP with sepsis, providing valuable insights for clinical diagnosis and treatment guidance.

## 2 Materials and methods

### 2.1 Data collection

We performed a search using “retinopathy of prematurity” and “sepsis” as keywords in the GEO database and retrieved two relevant gene expression datasets: GSE13904 and GSE218039 (Barrett et al., 2013). The microarray dataset GSE13904, comprises 18 samples from normal children and 52 samples from sepsis cases. The RNA-seq-based dataset GSE218039, consists of 16 control samples and 14 samples from ROP cases. To better illustrate our research process, we have created a flowchart in the Figure 1.

### 2.2 Data preprocessing and differentially expressed genes analysis

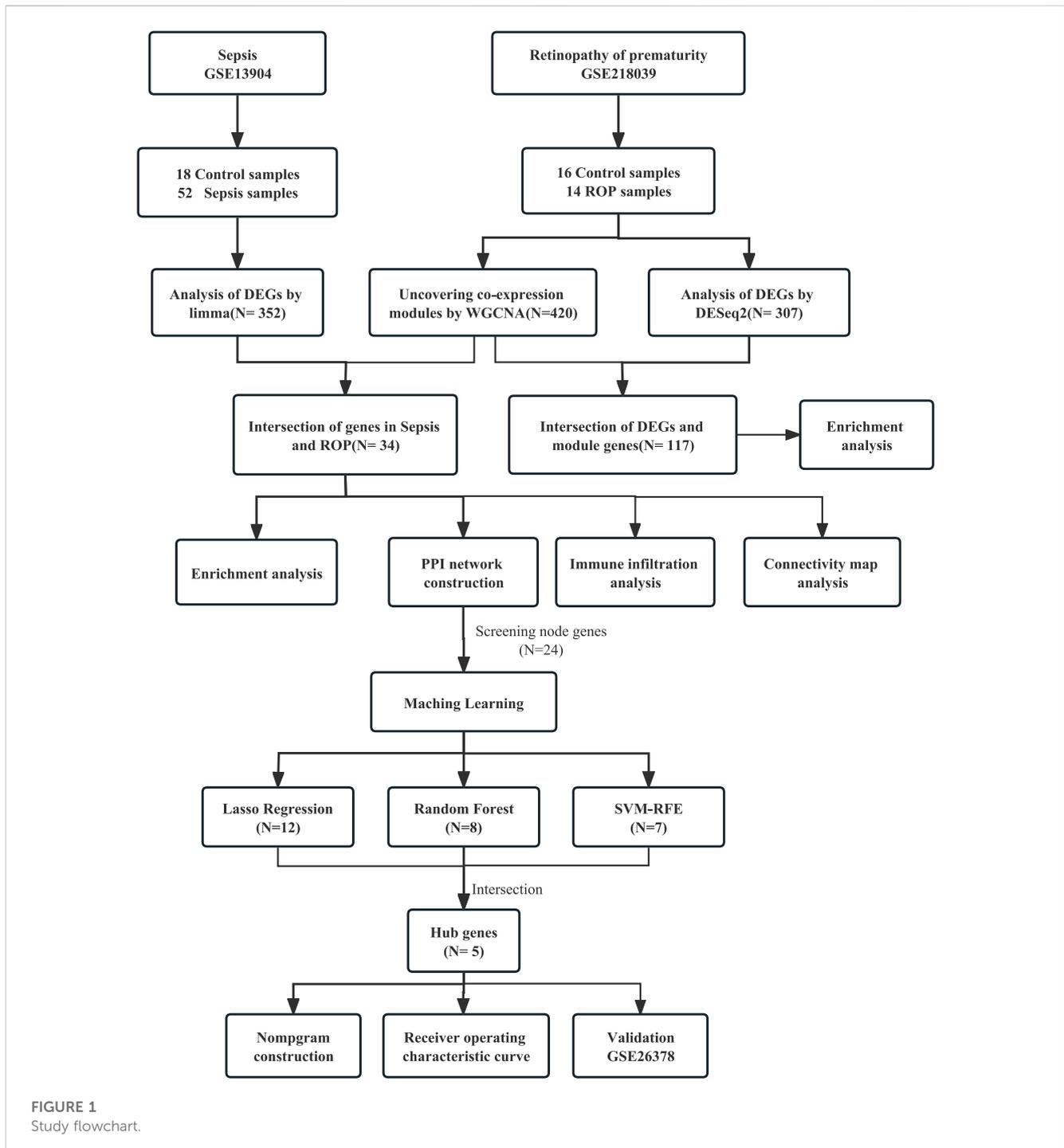
The microarray dataset for GSE13904 was downloaded from GEO and subjected to normalization using the “limma” package in R software. The “limma” package was employed to compare the gene expression profiles between normal samples and sepsis cases in order to identify DEGs. For GSE218039, we obtained the raw expression data, normalized the data, and employed the DESeq2 package in R to identify DEGs. Probe sets without corresponding gene symbols were excluded. Genes with multiple probe sets were averaged. Genes with an adjusted  $p$ -value  $< 0.05$  and a threshold of  $|\log_2 \text{Fold Change}| > 1$  were set for the analysis. Furthermore, we employed the R software package “ggplot2” to create separate volcano plots for visualizing the DEGs obtained from each of the two datasets.

### 2.3 Weighted gene correlation network analysis (WGCNA)

To investigate the correlation between genes, we constructed a gene co-expression network using the WGCNA (Langfelder and Horvath, 2008). Firstly, we calculated the median absolute deviation (MAD) for each gene and identified the top 5,000 genes. Then, we computed the adjacency using a “soft” thresholding power ( $\beta$ ) derived from co-expression similarity. The adjacency was transformed into a topological overlap matrix (TOM), and gene dissimilarity and heterogeneity were assessed. Thirdly, module detection was performed using hierarchical clustering and dynamic tree cutting. Genes that displayed comparable expression profiles were assigned to gene modules utilizing the average linkage hierarchical clustering method, ensuring a minimum gene module size ( $n = 30$ ) based on the TOM-based dissimilarity measure and gene dendrogram. Finally, employing WGCNA analysis, we identified and visualized the crucial modules associated with ROP.

### 2.4 Function enrichment analysis

To explore the functional and interactive roles of genes in biological pathways, we utilized the DAVID online database (<https://david.ncifcrf.gov>) to perform GO and KEGG pathway enrichment analyses (Kanehisa and Goto, 2000; The Gene Ontology Consortium, 2019). In this study, the results were visualized using the bioinformatics platform (<https://>



[www.bioinformatics.com.cn](http://www.bioinformatics.com.cn)). We performed two rounds of GO and KEGG analysis based on the intersection of DEGs and the key module genes of ROP, and the intersection of sepsis-related DEGs and the key module genes of ROP.

## 2.5 Protein–protein intersection network construction

We uploaded the candidate genes to the STRING database ([www.string-db.org](http://www.string-db.org)) and constructed a PPI network (Szkłarczyk

et al., 2021), with a medium confidence score of  $>0.4$ . Subsequently, we utilized the Cytoscape software to identify interacting genes for further analysis (Shannon et al., 2003).

## 2.6 Machine learning

To further identify candidate biomarkers for ROP and sepsis and establish a diagnostic model, we employed three machine learning algorithms. Lasso is a regularization method for linear regression that incorporates feature selection and model

interpretability by adding an L1 regularization term, thereby improving prediction accuracy (Yang et al., 2018). RF is an ensemble learning algorithm that constructs multiple decision trees for classification or regression (Ellis et al., 2014). The final prediction result is based on the voting or average of all decision trees, providing robustness and generalization capabilities. SVM-RFE combines support vector machines with recursive feature elimination, eliminating features with minimal impact on the classification results to obtain a simpler and superior performing model (Sanz et al., 2018). The genes identified by the overlapping of these three machine learning algorithms serve as candidate hub genes for diagnosis.

## 2.7 Nomogram construction and validation of candidate biomarkers

A nomogram for candidate hub genes was constructed using the “rms” package in R. Additionally, we utilized ROC analysis to further assess the clinical value of the candidate hub genes and the nomogram. The clinical value was determined by the area under the curve (AUC). An AUC value greater than 0.7 was deemed to indicate excellent diagnostic value.

To validate the robustness of the hub genes analysis results, we conducted validation using the microarray dataset GSE26378, which comprises 21 control samples and 82 sepsis cases. Statistical significance was determined by an unpaired *t*-test with a *p*-value <0.05.

## 2.8 Immune infiltration analysis

The relative abundance of immune-infiltrating cells was calculated using the CIBERSORT algorithm from the “IOBR” R package (Newman et al., 2015; Zeng et al., 2021). A bar plot was generated to illustrate the proportions of immune cells in different samples, and a box plot was employed to compare the different types of immune cells between the sepsis and control groups. Additionally, the “ggplot” R package was used to visualize the correlations between the hub genes and the 22 infiltrating immune cells.

## 2.9 Connectivity map (cMAP) analysis

The cMAP (<https://clue.io>) is an important gene expression database that provides valuable resources for researchers to explore the associations between genes, drugs, diseases, and small molecules (Subramanian et al., 2017). In this study, the overlapping upregulated genes from the DEGs of sepsis and the key modules of ROP were included in the cMAP online database to identify potential small molecule drugs that could be used for treatment. Finally, the top 10 small molecule compounds with the most negative correlation were determined.

## 2.10 Statistical analysis

All statistical analyses were performed using R software. The Wilcoxon or Student’s *t*-test was utilized to analyze the difference between the two groups. The correlation analysis between the

variables was determined using the Spearman’s correlation test. Statistical significance was set at a two-tailed *p* < 0.05.

# 3 Results

## 3.1 Differentially expressed genes

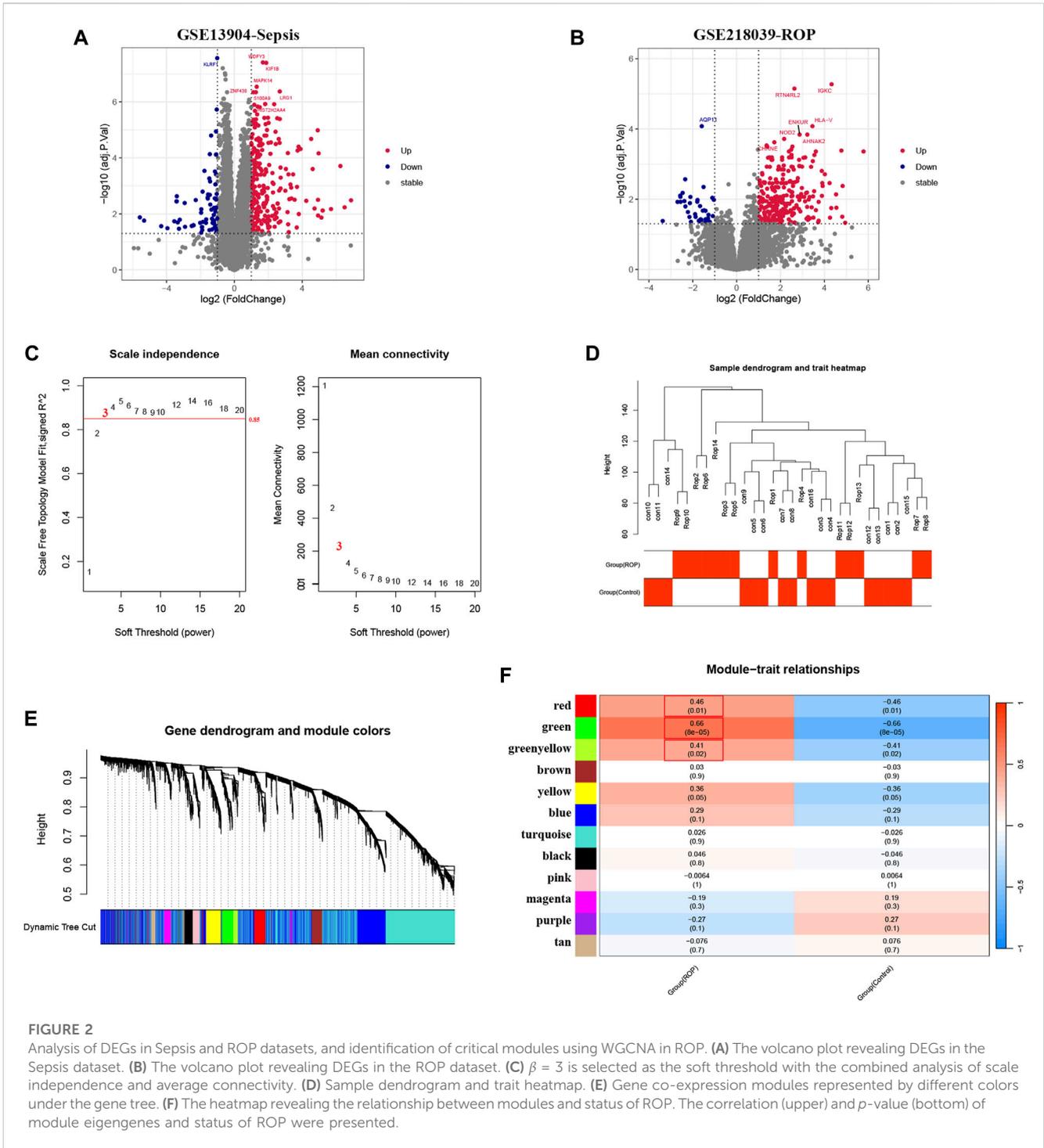
Overall, 352 DEGs were identified in the sepsis dataset, with 266 genes significantly upregulated and 86 genes significantly downregulated. As for the ROP dataset, a total of 307 DEGs were identified, with 267 genes upregulated and 40 genes downregulated. The volcano plots of DEGs are shown in Figures 2A, B.

## 3.2 Weighted gene co-expression network analysis and selection of critical modules

In GSE218039, we performed WGCNA to identify significantly correlated gene modules in ROP. In order to determine the most suitable soft threshold ensuring a scale-free network, we utilized the “pick Soft Threshold” function from the WGCNA package, which filtered power parameters within the range of 1–20. Based on a scale independence value greater than 0.85, we selected a power value ( $\beta$ ) of 3, which corresponded to a scale-free  $R^2$  of 0.88, as the optimal soft threshold (Figure 2C). Figure 2D illustrates the dendrogram of clustering for ROP and control groups. We used the “cutree” dynamic and module eigengenes functions to create a cluster diagram. This analysis resulted in the identification of 12 modules, each composed of genes exhibiting similar co-expression patterns (Figure 2E). Then, a heatmap was generated to display the correlation between ROP and modules based on Spearman correlation coefficients (Figure 2F). Notably, three modules—“red”, “green”, and “greenyellow”—displayed robust positive correlations with ROP and were therefore designated as ROP-related modules, with respective correlation coefficients and *p*-values as follows: red module ( $r = 0.46$ ,  $p = 0.01$ ), green module ( $r = 0.66$ ,  $p = 8e-05$ ), and greenyellow module ( $r = 0.41$ ,  $p = 0.02$ ). These modules consist of 167, 186, and 67 genes, totaling 420 genes in all, collectively considered as key module genes associated with ROP.

## 3.3 Functional enrichment analysis of ROP

Given the rare nature of ROP patients, we conducted an in-depth analysis of the dataset GSE218039 to explore potential mechanisms underlying ROP. Furthermore, this dataset is relatively new and has not been previously explored. We further conducted functional enrichment analysis based on the intersection of differential analysis results and key module genes from WGCNA. By intersecting the 307 DEGs and the 420 in the key module genes, a total of 117 common genes were screened (Figure 3A). As depicted in Figure 3B, Biological Process (BP) terms were predominantly enriched in “immune response”, “cell surface receptor signaling pathway”, and “inflammatory response”. In terms of Cellular Component (CC) ontology, enrichment was observed in “plasma membrane”, and “external side of plasma membrane”. Regarding Molecular Function (MF) analysis, enrichment was mainly found in

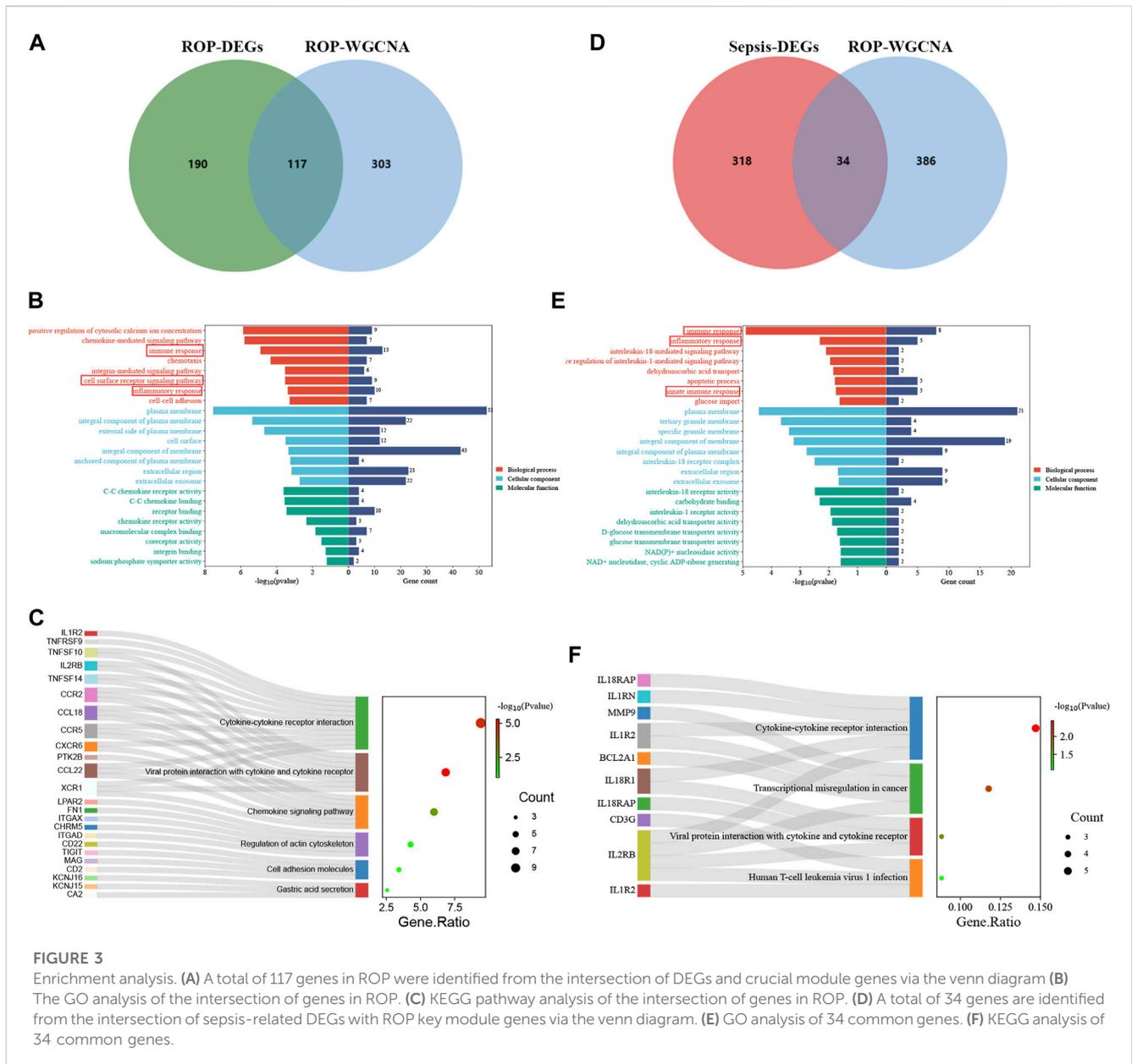


“receptor binding”, and “C-C chemokine receptor activity”. In addition, as shown in Figure 3C, KEGG analysis revealed enrichment in “Viral protein interaction with cytokine and cytokine receptor” and “Cytokine-cytokine receptor interaction”.

The enrichment analysis reveals a significant correlation between the overlapping genes in ROP and immune response and inflammation, suggesting their potential involvement in the pathogenesis of ROP. The detailed results of the GO and KEGG analyses from the DAVID database are available in Supplementary Table S1.

### 3.4 Enrichment analysis of ROP with sepsis and screening node genes via the PPI network

To explore the potential relationship between ROP-associated crucial genes and the pathogenesis of sepsis, we identified 34 genes from the intersection of sepsis-related DEGs with ROP key module genes using a Venn diagram (Figure 3D). GO analysis highlighted that these genes were primarily enriched in “immune response”, “inflammatory response”, and “innate immune response” (BP);



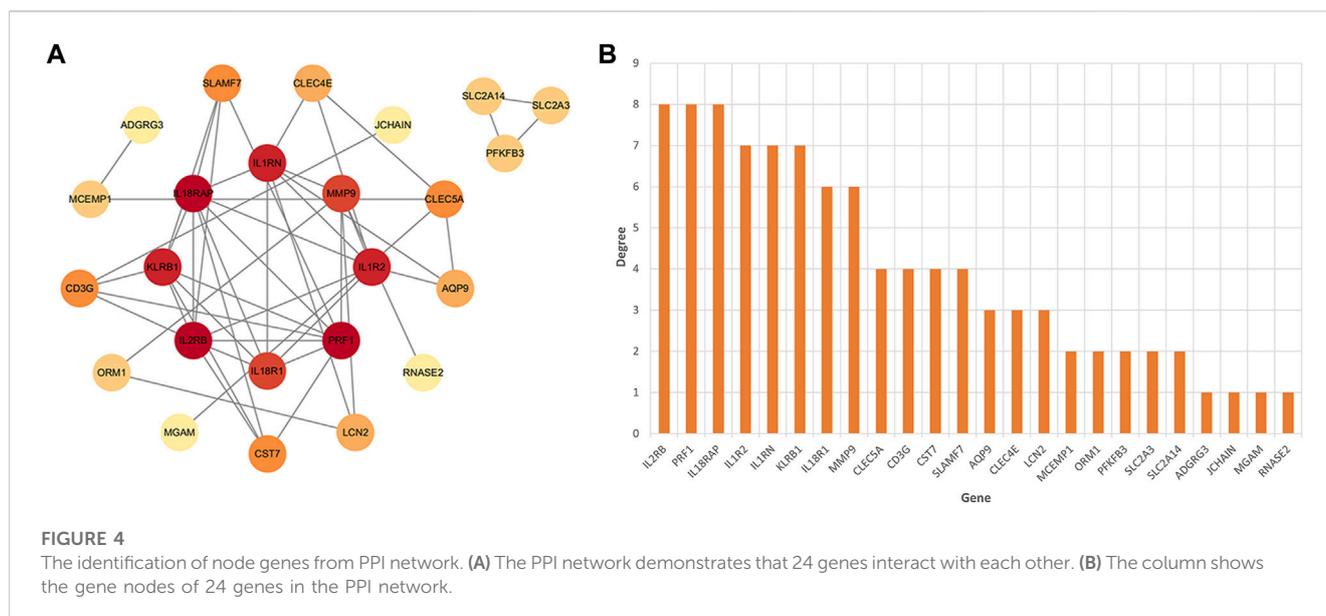
**FIGURE 3** Enrichment analysis. **(A)** A total of 117 genes in ROP were identified from the intersection of DEGs and crucial module genes via the venn diagram **(B)** The GO analysis of the intersection of genes in ROP. **(C)** KEGG pathway analysis of the intersection of genes in ROP. **(D)** A total of 34 genes are identified from the intersection of sepsis-related DEGs with ROP key module genes via the venn diagram. **(E)** GO analysis of 34 common genes. **(F)** KEGG analysis of 34 common genes.

“plasma membrane” and “interleukin-18 receptor complex” (CC); “interleukin-18 receptor activity” and “interleukin-1 receptor activity” (MF) (Figure 3E). Furthermore, based on KEGG analysis, the 34 genes showed prominent enrichment in pathways such as “Cytokine-cytokine receptor interaction” and “Viral protein interaction with cytokine and cytokine receptor” (Figure 3F, Supplementary Table S2). All the above findings strongly implicate the involvement of the immune system in these processes.

After identifying the immune-related nature of these genes, we constructed a PPI network to explore the interacting nodes among them (Figure 4A). Using Cytoscape software (Figure 4B), we identified 24 genes that interacted with each other, and these genes were sorted by their node degree.

### 3.5 Screening key genes via machine learning

Potential candidate hub genes were screened using the LASSO regression, RF algorithm, and SVM-RFE. Notably, LASSO regression revealed 12 potential candidate hub genes (Figures 5A, B). The RF algorithm ranked candidate genes based on the Mean Decrease Gini and selected the top eight genes with the highest scores (Figures 5C, D). SVM-RFE identified 7 genes as candidate biomarkers (Figure 5E). By employing a Venn diagram (Figure 5F), we intersected the results from the three algorithms, ultimately identifying five hub genes (CLEC5A, KLRB1, LCN2, MCEMP1 and MMP9) as potential candidate biomarkers (Supplementary Table S3).



### 3.6 Diagnosis value evaluation

To better diagnose and predict the five hub genes, we constructed a nomogram (Figure 6A). The calibration curve demonstrated that the predicted probabilities of our nomogram diagnostic model were in close agreement with the ideal model's predicted probabilities (Figure 6B). Additionally, decision curve analysis (DCA) was performed on the nomogram, indicating that using the nomogram model for decision-making may be beneficial for diagnosing ROP with sepsis (Figure 6C). We calculated the AUC values for both the nomogram and each hub gene as follows: nomogram (AUC: 0.991), CLEC5A (AUC: 0.9071), KLRB1 (AUC: 0.9209), LCN2 (AUC: 0.9049), MCEMP1 (AUC: 0.9466), and MMP9 (AUC: 0.9396) (Figures 6D, E). As observed, all five candidate biomarkers had AUC values greater than 0.9, and the AUC value of the nomogram was higher than that of each individual hub gene, suggesting that the nomogram has strong diagnostic value for ROP with sepsis.

Furthermore, we validated five candidate biomarkers using the GSE26378 dataset. CLEC5A, LCN2, MCEMP1, and MMP9 exhibited significantly higher expression levels in patients, while KLRB1 showed markedly lower expression in patients (Figure 6F). Subsequently, we conducted ROC curve single-factor analysis on the GSE26378 dataset, resulting in AUC values of 0.9431 for CLEC5A, 0.9332 for KLRB1, 0.9344 for LCN2, 0.9994 for MCEMP1, and 0.883 for MMP9 (Figure 6G). These findings indicate substantial diagnostic potential for these five candidate biomarkers.

### 3.7 Immune infiltration analysis

Our findings suggest that genes associated with ROP can also play a role in sepsis, particularly in immune regulation. Therefore, we proceeded with immune cell infiltration analysis to gain further insights into the involvement of the immune system in sepsis. Figure 7A displays the proportions of 22 immune cell types in

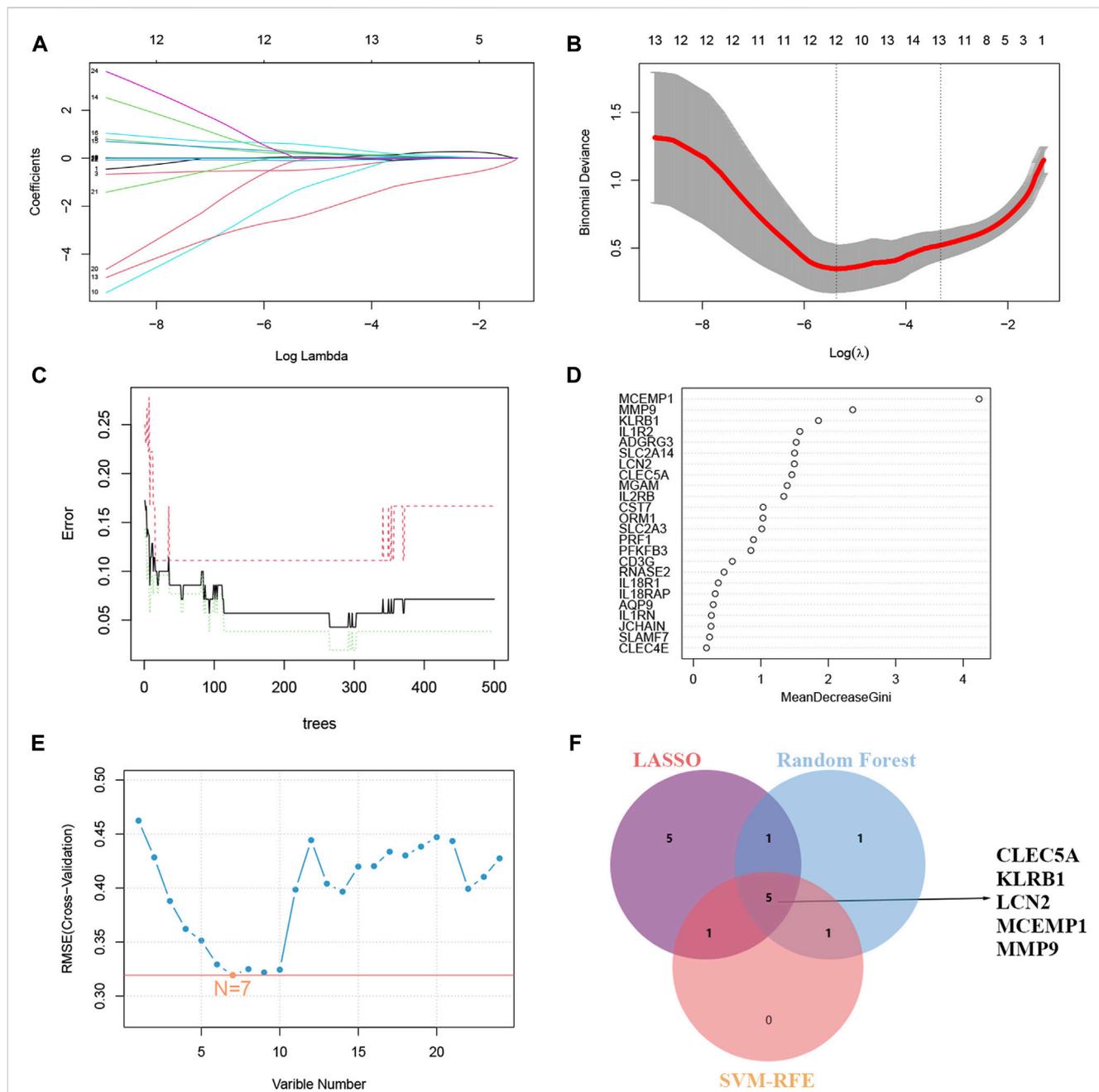
samples from the sepsis group and the normal group. The box plots demonstrate that sepsis patients have higher levels of Macrophages M0, Macrophages M2, Mast cells resting, and Mast cells activated, while lower levels of T-cells CD8<sup>+</sup>, T-cells CD4 naive, CD4 memory resting, and T-cells follicular helper (Figure 7B). Furthermore, we observed significant correlations between the five hub genes and immune cell infiltration in sepsis, as depicted in Figure 7C. In conclusion, various immune cell types exhibit varying degrees of infiltration in sepsis patients, which may provide potential targets for novel therapies.

### 3.8 Identifying core small molecule compounds for treatment

To further explore potential small-molecule drugs that may have therapeutic effects for ROP with sepsis, we inputted the upregulated genes from the 34 overlapping genes into the cMAP database. Our aim was to identify small-molecule compounds that could potentially reverse the abnormal gene expression patterns. The top-scoring compounds with the most antagonistic effects included escitalopram, harpagoside, hydroquinidine, latrepirdine, mepyramine, molsidomine, MW-STK33-3B, palonosetron, phenprobamate, and T-0070907. These compounds displaying antagonistic effects may serve as potential therapeutic drugs (Figure 8A). Figure 8B displayed chemical structures of these ten small molecule compounds.

## 4 Discussion

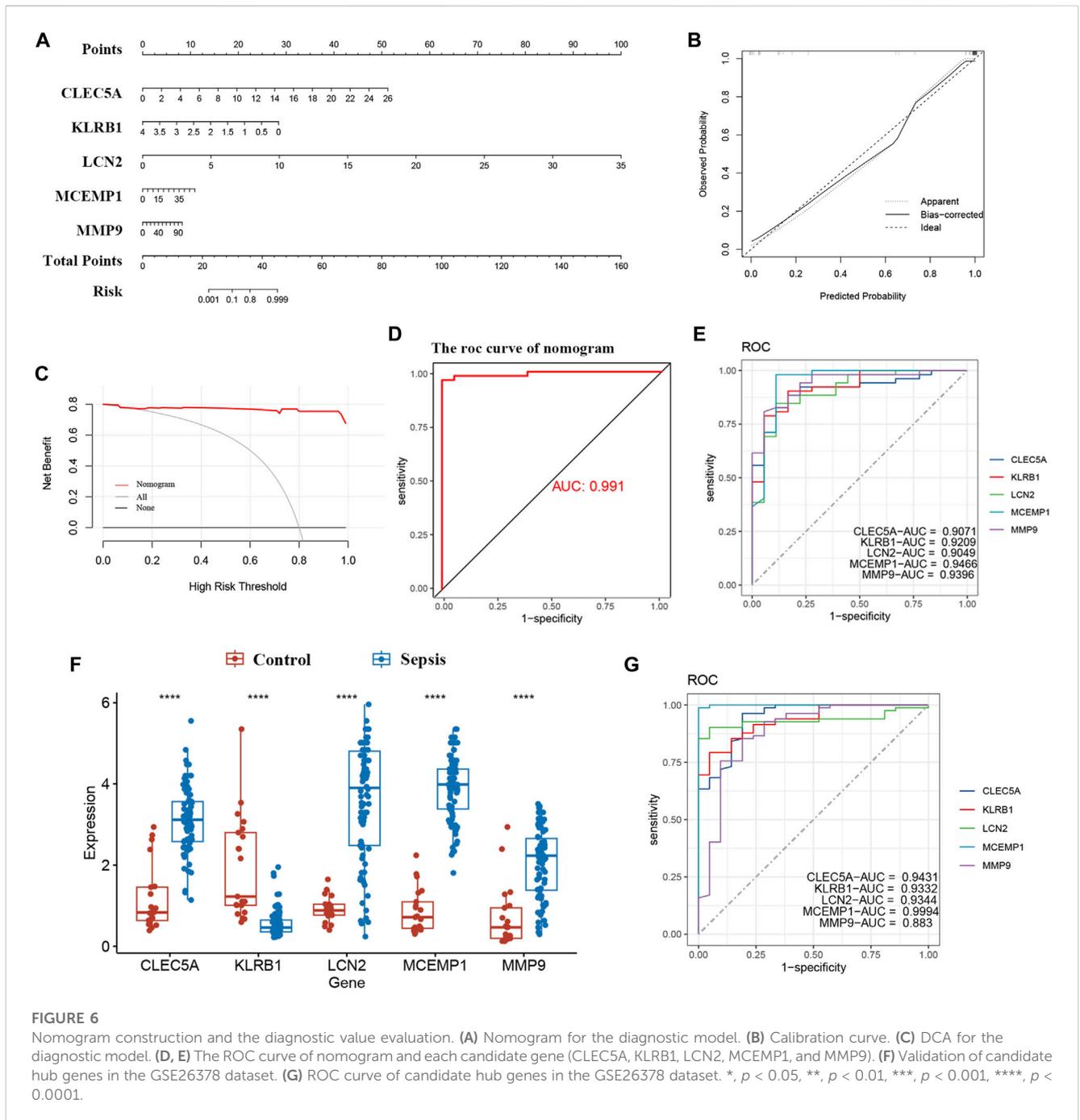
ROP is a retinal vascular regulatory disorder that affects the development of the retinal vasculature in premature infants, and it is influenced by various factors, including infection and prematurity (Huang et al., 2019; Wang et al., 2019). Sepsis increases the risk and severity of ROP at any stage, and the inflammatory factors triggered by sepsis may regulate retinal vascular development and alter



**FIGURE 5** Machine learning in screening candidate hub genes for ROP with sepsis. (A, B) The number of genes ( $n = 12$ ) corresponding to the lowest point of the curve in the LASSO model. (C) When “tree” was set to 200, the error within the model had essentially stabilized. (D) The Random Forest algorithm ranked genes based on the MDG. (E) Based on SVM-RFE to screen biomarkers. (F) Venn plot exhibiting the reliable biomarkers among LASSO, RF, and SVM-RFE.

angiogenesis (Dong et al., 2020). Alterations in retinal vascular development during the developmental process contribute to the development of ROP, making it a prominent cause of visual impairment in children (Hong et al., 2014). Currently, researches combining these two diseases is relatively rare. Accordingly, we conducted bioinformatics analysis and machine learning methods to develop a nomogram for evaluating the diagnostic performance of ROP with sepsis. Interestingly, we discovered five key immune-related candidate hub genes (CLEC5A, KLRB1, LCN2, MCEMP1, and MMP9) and built a nomogram for this purpose.

C-Type lectin receptor 5A (CLEC5A), also known as MDL-1, promotes extracellular trap formation, reactive oxygen species generation, and the production of pro-inflammatory cytokines in neutrophils (Bakker et al., 1999; Chen et al., 2017). CLEC5A is primarily expressed on myeloid cells (neutrophils, monocytes, macrophages, and dendritic cells) and its expression is regulated by the key transcription factor PU.1, indicating that CLEC5A expression is modulated by oxidative stress (Cheung et al., 2011; González-Domínguez érika et al., 2015; Cheng et al., 2016). Furthermore, CLEC5A can be considered as an innate immune checkpoint that

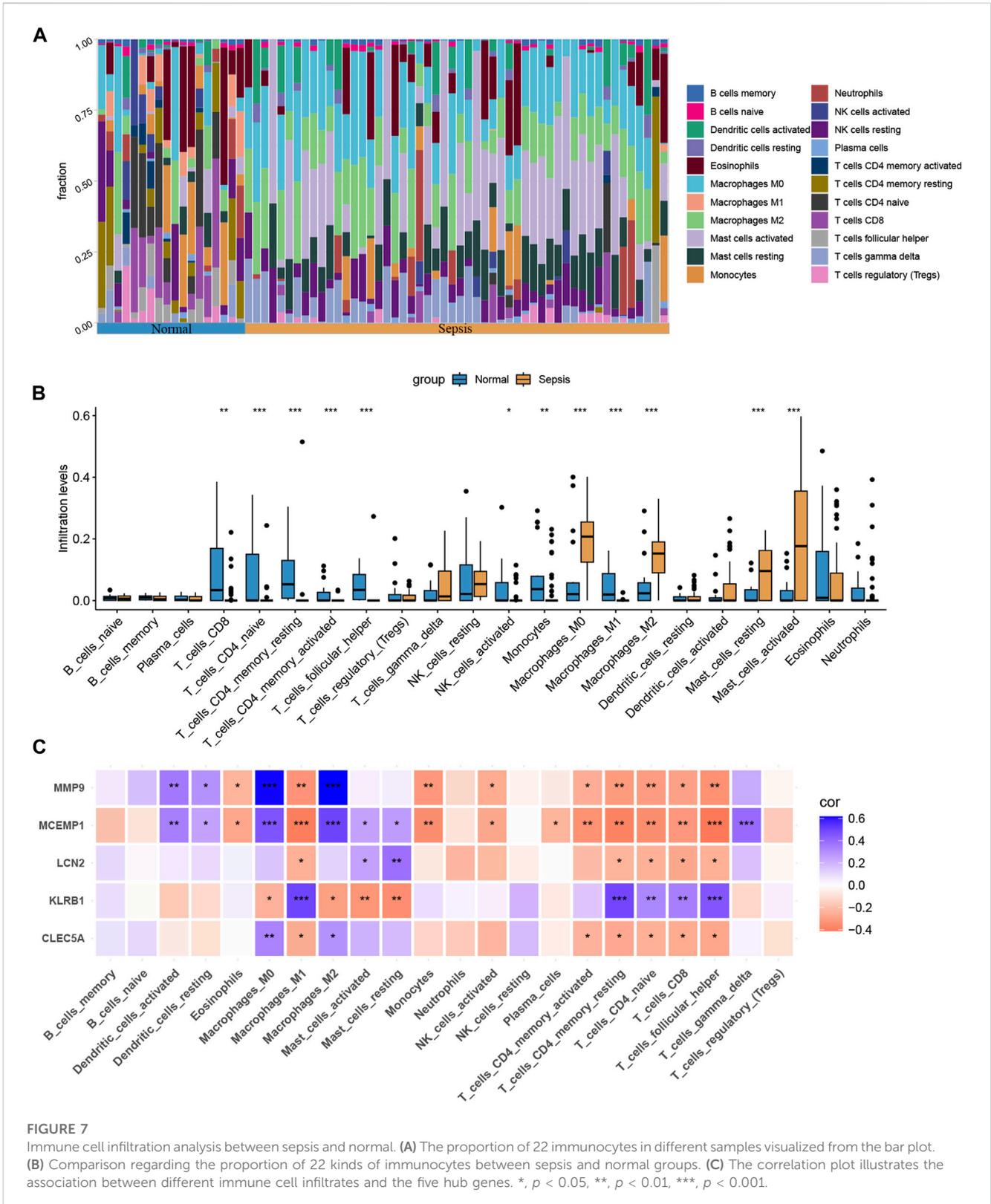


amplifies pro-inflammatory signals to facilitate the occurrence of infection or sterile inflammation (Tosiek et al., 2022). In sepsis, uncontrolled immune responses lead to harmful and potentially fatal inflammatory cascades. The involvement of CLEC5A in extracellular trap formation may exacerbate tissue damage and impede proper blood flow, potentially worsening the condition. In premature infants, the role of CLEC5A in modulating inflammatory responses in myeloid cells may contribute to abnormal retinal blood vessel development and neovascularization, leading to ROP.

Killer cell lectin-like receptor B1 (KLRB1), encodes a C-type lectin receptor found on natural killer cells and T-cells in peripheral blood, umbilical cord blood, and thymus (Giorda et al., 1990; Lanier

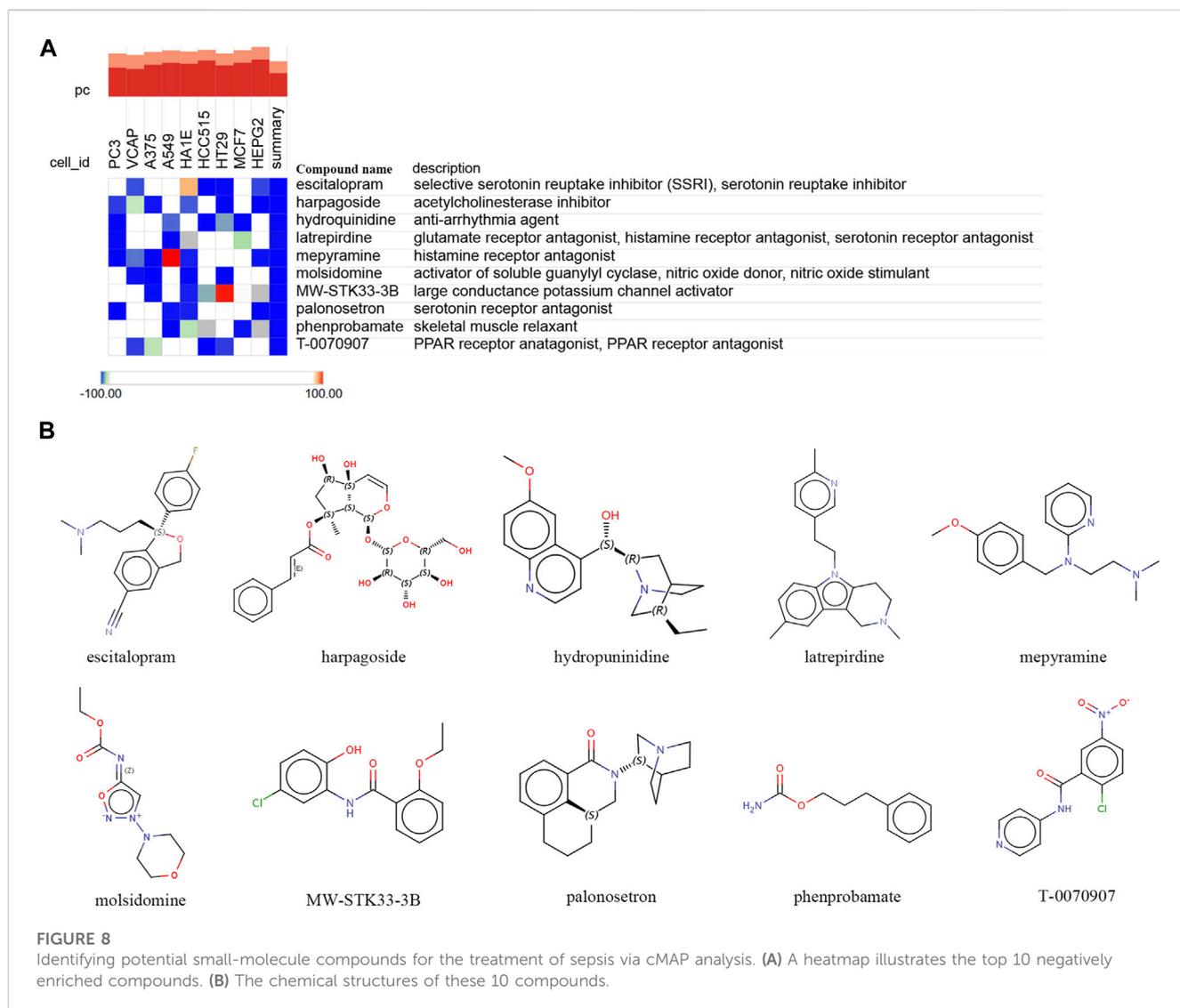
et al., 1994; Duurland et al., 2022). According to research findings, KLRB1 plays a crucial role in differentiation, particularly in dendritic cells and monocytes, and the expression of KLRB1 can serve as an indicator of NK cells involved in the pathogenesis of inflammatory diseases (Poggi et al., 1997; Kurioka et al., 2018). The increased expression of KLRB1 on NK cells and T cells may be correlated with immune cell activation and inflammatory responses in sepsis. NK cells release cytokines and participate in cytotoxicity during inflammation, potentially influencing vascular development and inflammatory responses.

Lipocalin-2 (LCN2) is an innate immune protein involved in various physiological and pathological processes, including iron



homeostasis, inflammation, and tumorigenesis (Chakraborty et al., 2012; Moschen et al., 2017). Additionally, we have found a clinical correlation between LCN2 and sepsis of intestinal origin (Lu et al., 2019). LCN2, as a transporter of lipids and iron, plays a critical regulatory role in lipid metabolism (Jaberi et al., 2021). Recent

research has revealed its involvement in lipid metabolism dysregulation in various sepsis conditions (Amunugama et al., 2021). Moreover, LCN2 may be implicated in septic cardiomyopathy by mediating lipid accumulation and influencing mitochondrial function (Liu et al., 2022).



MCCEMP1 is a transmembrane protein spanning the cell membrane and is commonly expressed by immune-related cells like mast cells and macrophages, playing a role in the pathogenesis of allergic and inflammatory diseases (Li et al., 2005; Schleinitz, 2015). Previous research has revealed that MCCEMP1 plays a crucial role in sepsis and viral infections (Raman et al., 2016; Nicolas De Lamballerie et al., 2021). Therefore, we also consider it as a diagnostic biomarker.

MMPs are a zinc enzyme family that plays a critical role in degrading and remodeling extracellular matrix proteins during normal developmental processes, as well as being involved in various physiological functions such as inflammatory responses, organ morphogenesis, and vascular formation during pathological processes (Stamenkovic, 2003). MMP9 is predominantly released by neutrophils and macrophages, and it is responsible for regulating inflammation in various tissues and diseases (Deryugina et al., 2014). Interestingly, it has been observed that MMP9 is closely associated with the human protein homolog of LCN2, known as NGAL (Kjeldsen et al., 1993). The MMP9/NGAL complex exhibits a strong correlation with TSP1, and they are found to be more actively

involved in the process of angiogenesis during orthodontic periodontal remodeling (Surlin et al., 2014). This finding suggests a potential interplay between MMP9/NGAL and TSP1 in the regulation of vascular formation. Furthermore, several studies have indicated that the expression and activity of MMPs increase during preterm birth, and they play a crucial role in fetal development, inflammatory responses, and angiogenesis (Cockle et al., 2007; Qazi et al., 2009). The imbalance between MMPs and their tissue inhibitors may also contribute to the occurrence of complications in newborns (Cockle et al., 2007). In the OIR model, we observed that systemic inhibition of metalloproteinases reduces neovascularization (Das et al., 1999). We speculate that MMP9 may be a candidate diagnostic gene for ROP patients with sepsis.

From previous research, we have come to understand that immune cells and inflammatory responses play a crucial role in the development and progression of sepsis (Riedemann et al., 2003). The core mechanism of sepsis is immune dysfunction, and as key cells in the innate immune system, macrophages play a crucial role in sepsis by performing functions such as antigen presentation and secretion of inflammatory and chemotactic factors (Epelman et al.,

2014; Luan et al., 2014; Qiu et al., 2019). Autophagy is closely linked to inflammation and immunity, and enhancing autophagy in sepsis can exert a protective effect by negatively regulating aberrant macrophage activation, reducing inflammasome activation and the release of pro-inflammatory cytokines, and influencing macrophage apoptosis (Qiu et al., 2019). The majority of studies emphasize the role of mast cells in the early stages of sepsis, where they exhibit immediate inflammatory responses and unique potential for combating infections (Gekara and Weiss, 2008; Heib et al., 2008; Ramos et al., 2010). Furthermore, we observed that inducing sepsis in mice leads to a decrease in the percentage of CD4<sup>+</sup> T-cells, while the percentages of T helper cells (Th2 and Th17) and regulatory T-cells (Treg) are upregulated (Yeh et al., 2022). In our study, we observed elevated levels of macrophages and mast cells, while the levels of CD8<sup>+</sup> T-cells, naive CD4<sup>+</sup> T-cells, memory CD4<sup>+</sup> T-cells, and follicular helper T-cells were found to be lower, which is consistent with previous research findings. Furthermore, a recent study suggested that immune cells might have played a role in the onset and progression of pediatric sepsis, which aligns with our findings (Zhang et al., 2023). By gaining a deeper understanding of the regulation mechanisms of immune cells and inflammatory responses, we can better comprehend the pathophysiology of sepsis, thus providing more comprehensive and precise guidance for the search of effective therapeutic strategies.

In recent years, significant progress has been made in identifying small molecule compounds with therapeutic potential for various diseases. Compared to other treatment methods, small molecule compounds offer advantages such as strong penetrability, high specificity, convenient administration, tunable pharmacokinetic properties, diversity, and flexibility, making them promising candidates for disease treatment. A metal gel capable of inhibiting sepsis was successfully synthesized through direct coordination interactions and Schiff base reactions between aminoglycosides, 2,2'-bipyridine-4,4'-dicarbaldehyde, and metal ions (Li et al., 2023). However, to date, there has been no research focusing on identifying potential therapeutic small molecule compounds based on sepsis gene expression characteristics. Therefore, we conducted a cMAP analysis, linking the upregulated genes in sepsis with the pathogenic genes related to ROP, and identified ten small molecule compounds (escitalopram, harpagoside, hydroquinidine, latrepirdine, mepyramine, molsidomine, MW-STK33-3B, palonosetron, phenprobamate, and T-0070907) as candidate compounds for potential sepsis treatment. In the cMAP analysis, these small molecule compounds exhibited the highest negative enrichment score, indicating that it has the potential to effectively reverse the upregulated expression of pathogenic genes in sepsis.

## 5 Conclusion

By leveraging bioinformatics analysis and machine learning algorithms, we systematically identified five immune-related candidate hub genes (CLEC5A, KLRB1, LCN2, MCEMP1, and MMP9) and provided a nomogram for diagnosing ROP with sepsis. Additionally, we emphasized the imbalance of immune cells and uncovered small molecule compounds as highly

promising candidates for sepsis treatment based on our analysis with cMAP. In conclusion, our research offered comprehensive insights into the pathophysiology and treatment of ROP with sepsis, ranging from the discovery of immune-related candidate genes, the construction of predictive models, to the identification of immune cell dysregulation and potential therapeutic drugs.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

## Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

## Author contributions

HC: Data curation, Software, Visualization. EGC: Writing—original draft, Writing—review and editing. YL: Study conception, Design. YX: Study conception, Design, Writing—review and editing.

## Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. The study was supported by the Science and Technology Commission of Shanghai Municipality (20Z11900400).

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1264873/full#supplementary-material>

## References

- Amunugama, K., Pike, D. P., and Ford, D. A. (2021). The lipid biology of sepsis. *J. Lipid Res.* 62. doi:10.1016/j.lipres.2021.100090
- Bakker, A. B. H., Baker, E., Sutherland, G. R., Phillips, J. H., and Lanier, L. L. (1999). Myeloid DAP12-associated lectin (MDL)-1 is a cell surface receptor involved in the activation of myeloid cells. *Proc. Natl. Acad. Sci.* 96, 9792–9796. doi:10.1073/pnas.96.17.9792
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193
- Bonafiglia, E., Gusson, E., Longo, R., Ficial, B., Tisato, M. G., Rossignoli, S., et al. (2022). Early and late onset sepsis and retinopathy of prematurity in a cohort of preterm infants. *Sci. Rep.* 12, 11675. doi:10.1038/s41598-022-15804-4
- Chakraborty, S., Kaur, S., Guha, S., and Batra, S. K. (2012). The multifaceted roles of neutrophil gelatinase associated lipocalin (NGAL) in inflammation and cancer. *Biochim. Biophys. Acta* 1826, 129–169. doi:10.1016/j.bbcan.2012.03.008
- Chen, S.-T., Li, F.-J., Hsu, T., Liang, S.-M., Yeh, Y.-C., Liao, W.-Y., et al. (2017). CLEC5A is a critical receptor in innate immunity against *Listeria* infection. *Nat. Commun.* 8, 299. doi:10.1038/s41467-017-00356-3
- Cheng, Y.-L., Lin, Y.-S., Chen, C.-L., Tsai, T.-T., Tsai, C.-C., Wu, Y.-W., et al. (2016). Activation of Nrf2 by the dengue virus causes an increase in CLEC5A, which enhances TNF- $\alpha$  production by mononuclear phagocytes. *Sci. Rep.* 6, 32000. doi:10.1038/srep32000
- Cheung, R., Shen, F., Phillips, J. H., McGeachy, M. J., Cua, D. J., Heyworth, P. G., et al. (2011). Activation of MDL-1 (CLEC5A) on immature myeloid cells triggers lethal shock in mice. *J. Clin. Invest.* 121, 4446–4461. doi:10.1172/JCI57682
- Cockle, J. V., Gopichandran, N., Walker, J. J., Levene, M. I., and Orsi, N. M. (2007). Matrix metalloproteinases and their tissue inhibitors in preterm perinatal complications. *Reprod. Sci.* 14, 629–645. doi:10.1177/1933719107304563
- Das, A., McLamore, A., Song, W., and McGuire, P. G. (1999). Retinal neovascularization is suppressed with a matrix metalloproteinase inhibitor. *Archives Ophthalmol.* 117, 498–503. doi:10.1001/archophth.117.4.498
- Deryugina, E. I., Zajac, E., Juncker-Jensen, A., Kupriyanova, T. A., Welter, L., and Quigley, J. P. (2014). Tissue-infiltrating neutrophils constitute the major *in vivo* source of angiogenesis-inducing MMP-9 in the tumor microenvironment. *Neoplasia* 16, 771–788. doi:10.1016/j.neo.2014.08.013
- Dogra, M. R., Katoch, D., and Dogra, M. (2017). An update on retinopathy of prematurity (ROP). *Indian J. Pediatr.* 84, 930–936. doi:10.1007/s12098-017-2404-3
- Dong, Y., Basmaci, R., Titomanlio, L., Sun, B., and Mercier, J.-C. (2020). Neonatal sepsis: within and beyond China. *Chin. Med. J. Engl.* 133, 2219–2228. doi:10.1097/CM9.0000000000000935
- Duurland, C. L., Santegoets, S. J., Abdulrahman, Z., Loof, N. M., Sturm, G., Wesselink, T. H., et al. (2022). CD161 expression and regulation defines rapidly responding effector CD4+ T cells associated with improved survival in HPV16-associated tumors. *J. Immunother. Cancer* 10, e003995. doi:10.1136/jitc-2021-003995
- Ellis, K., Kerr, J., Godbole, S., Lanckriet, G., Wing, D., and Marshall, S. (2014). A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol. Meas.* 35, 2191–2203. doi:10.1088/0967-3334/35/11/2191
- Epelman, S., Lavine, K. J., and Randolph, G. J. (2014). Origin and functions of tissue macrophages. *Immunity* 41, 21–35. doi:10.1016/j.immuni.2014.06.013
- Gekara, N. O., and Weiss, S. (2008). Mast cells initiate early anti-*Listeria* host defenses. *Cell. Microbiol.* 10, 225–236. doi:10.1111/j.1462-5822.2007.01033.x
- Giorda, R., Rudert, W. A., Vavassori, C., Chambers, W. H., Hiserodt, J. C., and Trucco, M. (1990). NKR-P1, a signal transduction molecule on natural killer cells. *Science* 249, 1298–1300. doi:10.1126/science.2399464
- González-Domínguez Érika, Samaniego, R., Flores-Sevilla, J. L., Campos-Campos, S. F., Gómez-Campos, G., Salas, A., Campos-Peña, V., et al. (2015). CD163L1 and CLEC5A discriminate subsets of human resident and inflammatory macrophages *in vivo*. *J. Leukoc. Biol.* 98, 453–466. doi:10.1189/jlb.3H1114-531R
- Heib, V., Becker, M., Taube, C., and Stassen, M. (2008). Advances in the understanding of mast cell function. *Br. J. Haematol.* 142, 683–694. doi:10.1111/j.1365-2141.2008.07244.x
- Hong, H. K., Lee, H. J., Ko, J. H., Park, J. H., Park, J. Y., Choi, C. W., et al. (2014). Neonatal systemic inflammation in rats alters retinal vessel development and simulates pathologic features of retinopathy of prematurity. *J. Neuroinflammation* 11, 87. doi:10.1186/1742-2094-11-87
- Huang, J., Tang, Y., Zhu, T., Li, Y., Chun, H., Qu, Y., et al. (2019). Cumulative evidence for association of sepsis and retinopathy of prematurity. *Medicine* 98, e17512. doi:10.1097/MD.00000000000017512
- Jaberi, S. A., Cohen, A., D'Souza, C., Abdulrazzaq, Y. M., Ojha, S., Bastaki, S., et al. (2021). Lipocalin-2: structure, function, distribution and role in metabolic disorders. *Biomed. Pharmacother.* 142, 112002. doi:10.1016/j.biopha.2021.112002
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27
- Kjeldsen, L., Johnsen, A. H., Sengelov, H., and Borregaard, N. (1993). Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *J. Biol. Chem.* 268, 10425–10432.
- Kurioka, A., Cosgrove, C., Simoni, Y., van Wilgenburg, B., Geremia, A., Björkander, S., et al. (2018). CD161 defines a functionally distinct subset of pro-inflammatory natural killer cells. *Front. Immunol.* 9, 486. doi:10.3389/fimmu.2018.00486
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559
- Lanier, L. L., Chang, C., Phillips, J. H., and Human Nkr-P1A, (1994). A disulfide-linked homodimer of the C-type lectin superfamily expressed by a subset of NK and T lymphocytes. *J. Immunol.* 153, 2417–2428. doi:10.4049/jimmunol.153.6.2417
- Lee, J., and Dammann, O. (2012). Perinatal infection, inflammation, and retinopathy of prematurity. *Seminars Fetal Neonatal Med.* 17, 26–29. doi:10.1016/j.siny.2011.08.007
- Li, K., Wang, S.-W., Li, Y., Martin, R. E., Li, L., Lu, M., et al. (2005). Identification and expression of a new type II transmembrane protein in human mast cells. *Genomics* 86, 68–75. doi:10.1016/j.ygeno.2005.03.006
- Li, H., Zhang, J., Xue, H., Li, L., Liu, X., Yang, L., et al. (2023). An injectable all-small-molecule dynamic metallogel for suppressing sepsis. *Mater. Horiz.* 10, 1789–1794. doi:10.1039/d3mh00005b
- Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., et al. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet* 388, 3027–3035. doi:10.1016/S0140-6736(16)31593-8
- Liu, W., Guo, X., Jin, L., Hong, T., Zhang, Q., Su, F., et al. (2022). Lipocalin-2 participates in sepsis-induced myocardial injury by mediating lipid accumulation and mitochondrial dysfunction. *Front. Cardiovasc. Med.* 9, 1009726. doi:10.3389/fcvm.2022.1009726
- Lu, F., Inoue, K., Kato, J., Minamishima, S., and Morisaki, H. (2019). Functions and regulation of lipocalin-2 in gut-origin sepsis: a narrative review. *Crit. Care* 23, 269. doi:10.1186/s13054-019-2550-2
- Luan, Y.-Y., Dong, N., Xie, M., Xiao, X.-Z., and Yao, Y.-M. (2014). The significance and regulatory mechanisms of innate immune cells in the development of sepsis. *J. Interferon and Cytokine Res.* 34, 2–15. doi:10.1089/jir.2013.0042
- Moschen, A. R., Adolph, T. E., Gerner, R. R., Wieser, V., and Tilg, H. (2017). Lipocalin-2: a master mediator of intestinal and metabolic inflammation. *Trends Endocrinol. Metabolism* 28, 388–397. doi:10.1016/j.tem.2017.01.003
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337
- Nicolas De Lamballerie, C., Pizzorno, A., Dubois, J., Padey, B., Julien, T., Traversier, A., et al. (2021). Human respiratory syncytial virus-induced immune signature of infection revealed by transcriptome analysis of clinical pediatric nasopharyngeal swab samples. *J. Infect. Dis.* 223, 1052–1061. doi:10.1093/infdis/jiaa468
- Poggi, A., Rubartelli, A., Moretta, L., and Zocchi, M. R. (1997). Expression and function of NKR-P1A molecule on human monocytes and dendritic cells. *Eur. J. Immunol.* 27, 2965–2970. doi:10.1002/eji.1830271132
- Qazi, Y., Maddula, S., and Ambati, B. K. (2009). Mediators of ocular angiogenesis. *J. Genet.* 88, 495–515. doi:10.1007/s12041-009-0068-0
- Qiu, P., Liu, Y., and Zhang, J. (2019). Review: the role and mechanisms of macrophage autophagy in sepsis. *Inflammation* 42, 6–19. doi:10.1007/s10753-018-0890-8
- Raman, K., O'Donnell, M. J., Czlonkowska, A., Duarte, Y. C., Lopez-Jaramillo, P., Peñaherrera, E., et al. (2016). Peripheral blood MCEMP1 gene expression as a biomarker for stroke prognosis. *Stroke* 47, 652–658. doi:10.1161/STROKEAHA.115.011854
- Ramos, L., Peña, G., Cai, B., Deitch, E. A., and Ulloa, L. (2010). Mast cell stabilization improves survival by preventing apoptosis in sepsis. *J. Immunol.* 185, 709–716. doi:10.4049/jimmunol.1000273
- Rhee, C., Jones, T. M., Hamad, Y., Pande, A., Varon, J., O'Brien, C., et al. (2019). Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. *JAMA Netw. Open* 2, e187571. doi:10.1001/jamanetworkopen.2018.7571
- Riedemann, N. C., Guo, R.-F., and Ward, P. A. (2003). Novel strategies for the treatment of sepsis. *Nat. Med.* 9, 517–524. doi:10.1038/nm0503-517
- Rivera, J. C., Sapienza, P., Joyal, J.-S., Duhamel, F., Shao, Z., Sitaras, N., et al. (2011). Understanding retinopathy of prematurity: update on pathogenesis. *Neonatology* 100, 343–353. doi:10.1159/000330174
- Sanz, H., Valim, C., Vegas, E., Oller, J. M., and Reverter, F. (2018). SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinforma.* 19, 432. doi:10.1186/s12859-018-2451-4

- Sato, T., Kusaka, S., Shimojo, H., and Fujikado, T. (2009). Vitreous levels of erythropoietin and vascular endothelial growth factor in eyes with retinopathy of prematurity. *Ophthalmology* 116, 1599–1603. doi:10.1016/j.ophtha.2008.12.023
- Schleinitz, D. (2015). Genetic determination of serum levels of diabetes-associated adipokines. *Rev. Diabet. Stud.* 12. doi:10.1900/RDS.2015.12.277
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303
- Shen, R., Embleton, N., Lyng Forman, J., Gale, C., Griesen, G., Sangild, P. T., et al. (2022). Early antibiotic use and incidence of necrotising enterocolitis in very preterm infants: a protocol for a UK based observational study using routinely recorded data. *BMJ Open* 12, e065934. doi:10.1136/bmjopen-2022-065934
- Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 315, 801–810. doi:10.1001/jama.2016.0287
- Stamenkovic, I. (2003). Extracellular matrix remodelling: the role of matrix metalloproteinases. *J. Pathology* 200, 448–464. doi:10.1002/path.1400
- Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., et al. (2017). A next generation connectivity map: L1000 platform and the First 1,000,000 profiles. *Cell* 171, 1437–1452.e17. doi:10.1016/j.cell.2017.10.049
- Surlin, P., Silosi, I., Rauten, A. M., Cojocaru, M., and Foia, L. (2014). Involvement of TSP1 and MMP9/NGAL in angiogenesis during orthodontic periodontal remodeling. *ScientificWorldJournal* 2014, 421029. doi:10.1155/2014/421029
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi:10.1093/nar/gkaa1074
- The Gene Ontology Consortium (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338. doi:10.1093/nar/gky1055
- Tosiek, M. J., Groesser, K., Pekcec, A., Zwirek, M., Murugesan, G., and Borges, E. (2022). Activation of the innate immune checkpoint CLEC5A on myeloid cells in the absence of danger signals modulates macrophages' function but does not trigger the adaptive T cell immune response. *J. Immunol. Res.* 2022, 9926305. doi:10.1155/2022/9926305
- Tremblay, S., Miloudi, K., Chaychi, S., Favret, S., Binet, F., Polosa, A., et al. (2013). Systemic inflammation perturbs developmental retinal angiogenesis and neuroretinal function. *Investigative Ophthalmol. Vis. Sci.* 54, 8125–8139. doi:10.1167/iovs.13-12496
- Wang, X., Tang, K., Chen, L., Cheng, S., and Xu, H. (2019). Association between sepsis and retinopathy of prematurity: a systematic review and meta-analysis. *BMJ Open* 9, e025440. doi:10.1136/bmjopen-2018-025440
- Wu, T., Zhang, L., Tong, Y., Qu, Y., Xia, B., and Mu, D. (2018). Retinopathy of prematurity among very low-birth-weight infants in China: incidence and perinatal risk factors. *Investigative Ophthalmol. Vis. Sci.* 59, 757–763. doi:10.1167/iovs.17-23158
- Yang, C., Delcher, C., Shenkman, E., and Ranka, S. (2018). Machine learning approaches for predicting high cost high need patient expenditures in health care. *Biomed. Eng. Online* 17, 131. doi:10.1186/s12938-018-0568-3
- Yeh, C.-L., Wu, J.-M., Yang, P.-J., Lee, P.-C., Chen, K.-Y., Yeh, S.-L., et al. (2022). Intravenous calcitriol treatment benefits the homeostasis of CD4+ T cells and attenuates kidney injury in obese mice complicated with polymicrobial sepsis. *Nutrition* 103–104, 111741. doi:10.1016/j.nut.2022.111741
- Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., et al. (2021). IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front. Immunol.* 12, 687975. doi:10.3389/fimmu.2021.687975
- Zhang, W.-Y., Chen, Z.-H., An, X.-X., Li, H., Zhang, H.-L., Wu, S.-J., et al. (2023). Analysis and validation of diagnostic biomarkers and immune cell infiltration characteristics in pediatric sepsis by integrating bioinformatics and machine learning. *World J. Pediatr.* 19, 1094–1103. doi:10.1007/s12519-023-00717-7