Article

Discovery Medicine 2026; 38(214): 3670–3686
https://doi.org/10.24976/Discov.Med.202638214.279

Machine Learning-Based Identification of RNA Processing-Related Diagnostic Biomarkers and Molecular Subtypes in Pediatric Sepsis

Min Huang[1], Lingling Zeng[1,*]

[1]Department of Pediatrics, The Youyang Tujia and Miao Autonomous County People's Hospital of Chongqing, 409800 Chongqing, China
*Correspondence: zll25286@163.com (Lingling Zeng)
Submitted: 21 July 2026 Revised: 5 August 2026 Accepted: 7 September 2026 Published: 25 November 2026

Background: Pediatric sepsis is a life-threatening condition with high mortality, underscoring the need for novel diagnostic biomarkers. Dysregulated RNA processing contributes to sepsis pathogenesis, highlighting the need to investigate its association with pediatric sepsis.

Methods: Microarray data and machine learning were used to identify diagnostic biomarkers. Microarray data from Gene Expression Omnibus (GEO) were analyzed via Weighted Gene Co-expression Network Analysis (WGCNA), and four algorithms (Boruta, Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine - Recursive Feature Elimination (SVM-RFE), extreme gradient boosting (XGBoost)) screened feature genes. Diagnostic performance was evaluated using a nomogram, receiver operating characteristic (ROC) curve analysis, and decision curve analysis (DCA). Immune infiltration was assessed with Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT), molecular subtyping with ConsensusClusterPlus, and pathways through gene set enrichment analysis (GSEA).

Results: Five diagnostic biomarkers (*TDRD9*, *SRSF8*, *EXOSC4*, *DPH3*, *USB1*) were identified. The nomogram showed promising diagnostic performance (area under the receiver operating characteristic curve (AUC) >0.86), and exploratory consistency patterns were observed in three related pediatric sepsis external datasets. Immune analysis revealed neutrophil/monocyte dominance. Molecular subtyping in the training cohort revealed two potential subtypes with distinct immune profiles, which warrant further validation in independent cohorts.

Conclusion: Collectively, we developed and internally evaluated a five-gene signature associated with RNA processing, which showed high discrimination between sepsis samples and healthy controls in the training cohort. Furthermore, exploratory subtype analysis performed in the training cohort revealed two distinct immune-related clusters, providing preliminary insights into the heterogeneity of sepsis.

Keywords: pediatric sepsis; diagnostic biomarkers; RNA processing; subtypes; machine learning; immune landscape

Introduction

Pediatric sepsis is a critical, life-threatening condition in which organ impairment results from an excessive or unregulated host immune response to an infection [1]. It is not merely a bloodstream infection; its core pathology lies in the fact that children, particularly infants and young children with immature immune systems, mount an excessively intense or dysregulated defensive response to invading pathogens [2]. This escalates into a "cytokine storm" that damages the host's own tissues and organs [3]. The onset of this disease is often insidious and its progression rapid, with clinical manifestations that can be non-specific, such as lethargy, tachypnea, or mottled skin. If not recognized early and managed promptly, it can rapidly deteriorate into septic shock and multiple organ failure within hours, making it a leading global threat to child mortality [4,5]. Therefore, a deep understanding of its pathogenesis, recognition of its early signs, and the initiation of time-critical, bundled management are crucial for improving outcomes. Current therapeutic strategies include controlling the source of infection with antibiotics and modulating the host's immune response [6]. Despite advances in treatment approaches, the overall prognosis remains poor, especially for infant patients, due to the rapid disease progression [7,8]. This highlights the urgent need for novel biomarkers and more effective therapies to improve outcomes for patients facing this challenging disease.

RNA processing is an essential step in gene expression and includes processes such as 5′ capping, splicing, and 3′ polyadenylation. During this process, the primary RNA transcript synthesized from a DNA template undergoes trimming and chemical modifications to become a mature, biologically functional RNA molecule [9,10,11]. Key processes include 5′ capping, 3′ polyadenylation, and RNA splicing [12,13]. These processing steps collectively play

Copyright: © 2026 The Author(s). Published by Discovery Medicine. This is an open access article under the CC BY 4.0 license.
Note: *Discovery Medicine* stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

vital roles: they protect the RNA from rapid degradation by cytoplasmic nucleases, facilitate its efficient transport from the nucleus to the cytoplasm, and are essential for efficient protein translation on ribosomes [14,15]. Aberrant RNA processing is often associated with various diseases, particularly in sepsis. Research on CLP mice revealed nine types of differentially alternative splicing events in the liver, including A3SS, A5SS, and ES, which are primarily involved in lipid metabolism and redox pathways. These splicing patterns were partially reversed following dichloroacetic acid treatment. Consequently, targeting splicing regulation represents a potential strategy to alleviate septic liver injury [16]. In sepsis, neutrophils show increased lncRNA GSEC levels, which in turn augment the release of inflammatory cytokines including TNF-α and IL-1β through upregulating PFKFB3 transcription and glycolytic flux [17]. Overall, RNA processing goes beyond being simply a "tweak" after transcription—it is an essential layer of gene regulatory control. Its malfunction can cause immune dysregulation and subsequently foster diseases like sepsis. Thus, a better understanding of how RNA processing correlates with sepsis outcome could facilitate the creation of novel treatment strategies.

This study aims to comprehensively identify and evaluate RNA processing-related genes (RPRGs)-based diagnostic biomarkers in pediatric sepsis using an integrative bioinformatics pipeline. Applying this pipeline, we will clarify the immune profiles linked to these biomarkers and probe their potential for differentiating patient subgroups that may have clinical relevance. Our work is expected to uncover novel mechanistic insights into sepsis and to potentially contribute to the development of more refined clinical approaches.

Materials and Methods

Data Collection

Microarray datasets generated using the same GPL570 platform were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) (GSE66099, GSE26378, GSE26440, GSE13904), with GSE66099 used as the training set (Sepsis: 199, Control: 47). GSE26378 (Sepsis: 82, Control: 21), GSE26440 (Sepsis: 98, Control: 32), and GSE13904 (Sepsis: 158, Control: 18) were selected for external evaluation and classified as exploratory datasets. In the aforementioned datasets, we applied uniform sample selection criteria to ensure phenotypic homogeneity across groups and to avoid interference from non-infectious inflammation in biomarker discovery. Specifically: Sepsis group (case group): Defined as pediatric patients with sepsis or septic shock meeting the diagnostic criteria established by the International Pediatric Sepsis Consensus Conference. Control Group: Defined as healthy children without any infection or inflammatory disease. Systemic inflammatory response syndrome (SIRS) Group (Exclusion Group): Defined as pediatric patients who met the diagnostic criteria for SIRS but had no clear evidence of infection; these patients were excluded to eliminate the confounding effects of non-infectious systemic inflammation. All Affymetrix microarray datasets were preprocessed using the robust multi-array average (RMA) algorithm via the affy package, including background correction, quantile normalization, and $\log_2$ transformation. Probe-to-gene mapping was performed using the hgu133plus2.db annotation package, and for multiple probes mapping to the same gene, the mean expression value was used. The same processing pipeline was applied to all datasets to minimize batch effects. All four datasets were derived from multi-center pediatric sepsis studies where whole-blood samples were collected within 24 hours of patient admission to the Pediatric Intensive Care Unit (PICU) or within 24 hours of meeting sepsis criteria, capturing the early acute phase of the host response. To assess sample-level and patient-level overlap, we conducted a sample overlap analysis across four research cohorts based on GSM accession numbers (**Supplementary Fig. 1A**). Additionally, a patient-level overlap analysis was performed across the four datasets using GEO sample metadata and relation fields. Patient identifiers were extracted and mapped from the available metadata, and samples that could not be reliably mapped were separately flagged. The results indicated that there was substantial patient-level overlap at the patient level between GSE66099 and the other three datasets (**Supplementary Fig. 1B**, **Supplementary Table 1**). Therefore, given the confirmed patient-level overlap, the external results were presented only as an exploratory consistency analysis and no longer serve as evidence of independent validation. We retrieved 1026 RPRGs associated with the Gene Ontology term "RNA processing" (GO:0006396) from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/index.jsp) following previously established methods [18].

Weighted Gene Co-Expression Network Analysis (WGCNA)

Construction of the weighted gene co-expression network was carried out using the "WGCNA" package (v1.74). Hierarchical clustering was applied to the top 25% of genes with the greatest variance. To achieve a scale-free topology fit of $R^2 = 0.85$, we chose a soft threshold of $\beta = 20$, which enabled the partitioning of similarly expressed genes into modules. The correlation between each module and the clinical phenotype was then evaluated using Pearson's coefficients, and the module with the most significant association was considered the pivotal module.

Differential Expression Analysis

Differential expression analysis was performed using the limma package (v3.68.5). The expression matrix was

collapsed by avereps() to merge duplicate genes, and a design matrix and contrast (Sepsis vs Control) were constructed. Linear models were fitted with lmFit(), followed by contrasts.fit() and eBayes() for empirical Bayes adjustment. Multiple testing was corrected by the Benjamini-Hochberg method. Genes with $|\log_2 FC| > 1$ and adj.*p*.Val < 0.05 were defined as differentially expressed genes (DEGs).

Gene Set Variation Analysis (GSVA)

We performed GSVA using the single-sample Gene Set Enrichment Analysis (ssGSEA) algorithm to evaluate the impact of RPRGs in pediatric sepsis. This analysis assessed expression differences between sepsis and control groups using the GSE66099 training set and was also performed in three exploratory datasets: GSE26378, GSE26440, and GSE13904.

Diagnostic Biomarker Screening Through Machine Learning

To optimize the screening of candidate diagnostic biomarkers, this study employed an integrated strategy utilizing multiple machine learning algorithms. Specifically, four distinct machine learning algorithms were applied. Four methods were used for feature selection: Boruta, which employed random forest with permutation testing ($p < 0.05$) to identify relevant features; Least Absolute Shrinkage and Selection Operator (LASSO) ($\alpha = 1$, λ optimized by 10-fold CV), to retain features with nonzero coefficients; and Support Vector Machine - Recursive Feature Elimination (SVM-RFE), implemented using the rfe function from the caret R package (v7.0-1) with a linear kernel. The main parameters were method = "svmLinear", C = 1, and 10-fold cross-validation (method = "cv", number = 10). Finally, extreme gradient boosting (XGBoost) was applied with fixed parameters (max.depth = 2, eta = 1, nrounds = 2, objective = "binary:logistic") to rank feature importance. Candidate diagnostic biomarkers were ultimately defined as gene signatures consistently selected across all four algorithms.

The five genes identified through the integrated feature selection strategy were subsequently used to construct a multivariate binary logistic regression model in the training set (GSE66099). Sepsis status (sepsis = 1, control = 0) was set as the dependent variable, while the expression levels of the five genes were input as independent variables. The expression values used in the logistic regression model were on the same $\log_2$-transformed RMA-normalized scale as the data used for downstream prediction. No additional scaling or normalization was performed on the data prior to logistic regression. The logistic regression model was fitted using the glm function in R with the argument family = binomial (). The predicted probability of sepsis was calculated according to the following logistic regression equation:

$$\text{logit}(p) = \beta_0 + \beta_1 \times USB1 + \beta_2 \times TDRD9 + \beta_3 \times SRSF8 + \beta_4 \times EXOSC4 + \beta_5 \times DPH3$$

where (p) represents the predicted probability of sepsis, and (β_0) is the model intercept. The predicted probability is then calculated using the following formula:

$$p = 1/(1 + e^{-\text{logit}(p)})$$

To assess and correct for potential optimism in model performance, internal bootstrap validation was performed using 1000 bootstrap resamples of the training cohort. Importantly, the complete model-development pipeline was repeated independently within each bootstrap iteration rather than applying a fixed five-gene signature. In each iteration, samples were resampled with replacement from the original training cohort with the same sample size as the training cohort. Boruta, LASSO, SVM-RFE, and XGBoost were then reapplied to the bootstrap sample, and genes consistently selected by all four algorithms were retained for model construction. A logistic regression model was subsequently fitted using the genes selected within that bootstrap sample. Model performance was evaluated in the corresponding out-of-bag (OOB) samples by calculating the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity. The bootstrap estimates were used to quantify and correct for optimism and to obtain 95% confidence intervals for model performance.

Diagnostic Model Evaluation

The logistic regression model was subsequently visualized as a nomogram using the lrm and nomogram functions implemented in the "rms" R package (v8.1-1). Calibration curves were generated to assess the agreement between the predicted and observed probabilities, and decision curve analysis (DCA) was performed to evaluate the potential clinical net benefit of the model. ROC curves were generated based on the predicted probabilities from the logistic regression model. For the training cohort, model performance was evaluated using the five-gene model fitted to the complete GSE66099 cohort. For the exploratory datasets, the regression coefficients and intercept derived from GSE66099 were directly applied to GSE26378, GSE26440, and GSE13904 without model refitting. AUC, sensitivity, and specificity, together with their corresponding 95% confidence intervals, were calculated for each datasets.

Immune Infiltration and Correlation Analysis

Immune infiltration between control and patient groups was assessed via Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBER-

SORT) (implemented in the "IOBR" package (v2.2.3)), while the "GSVA" package (v4.3.3) was used for ssGSEA to compare immune cell scores statistically. Subsequently, correlations between feature genes and both immune cells and immune-related functions were explored.

Subtype Identification Based on Diagnostic Genes

Consensus clustering was performed using the ConsensusClusterPlus package with the following parameters: maxK = 10, reps = 1000, pItem = 0.8, pFeature = 1.0, clusterAlg = 'km', distance = 'euclidean', and seed = 1234. The optimal number of clusters (k = 2) was determined by evaluating the cumulative distribution function (CDF) curves, delta area plot, consensus heatmap clarity, and principal component analysis (PCA) visualization. Quantitative analysis of expression differences in the feature genes among subtypes was performed, along with differential expression analysis (adjusted p-value < 0.05 and $|\log_2 FC| > 1$), to delineate subtype-specific molecular characteristics.

Gene Set Enrichment Analysis and Pathway Analysis

We retrieved the gene set subsets "c5.go.v2024.1.Hs.symbols" and "c2.cp.kegg_legacy.v2024.1.Hs.symbols.gmt" from MSigDB. Enrichment analysis of the feature genes was performed using the "GSEA" package, while GSEA software (v4.3.3; Broad Institute of MIT and Harvard, Cambridge, MA, USA) was employed to compare pathway activation across pediatric sepsis subtypes. Additionally, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were conducted with the "clusterProfiler" R package (v4.20.3).

Establishment of miRNA-mRNA and TF-based Regulatory Networks

Using NetworkAnalyst (https://www.networkanalyst.ca/) with ENCODE data, we systematically screened upstream transcriptional regulators of the feature genes. miRTarBase (https://mirtarbase.cuhk.edu.cn/) was employed to identify miRNA-target pairs for constructing the mRNA-miRNA regulatory network. All networks were visualized using Cytoscape (v3.10; Cytoscape Consortium, San Diego, CA, USA).

Statistical Analysis

All analyses were performed using relevant packages in R (version 4.4.0; R Foundation for Statistical Computing, Vienna, Austria). Data distribution was assessed using the Shapiro-Wilk normality test. As most gene expression data did not follow a normal distribution, differences between two groups were evaluated using the two-sided Wilcoxon rank-sum test. For multiple comparisons in differential expression analysis, the Benjamini-Hochberg (BH) method was applied to control the false discovery rate (FDR). For other comparisons, unadjusted p-values are reported with clear annotations, and all statistical tests were two-sided. Data were primarily visualized using the "ggplot2" package. Spearman's method was used to assess correlations between two variables, while Pearson correlation was employed in WGCNA. For the exploratory immune infiltration, ssGSEA, and gene–immune cell correlation analyses, nominal statistical significance was defined using the raw two-sided p value < 0.05. Given the exploratory nature of these analyses, raw p values were retained in the main figures to identify potentially informative immune-related signals, while Benjamini–Hochberg FDR-adjusted p values were additionally calculated and reported in the **Supplementary Tables 9–13** to assess the robustness of these findings under multiple-testing correction. Statistical significance was determined at a p value < 0.05.

Results

Weighted Gene Co-Expression Network Analysis

WGCNA was performed on the training set using the top 25% most variable genes. The network was constructed with a soft threshold of 20 (R^2 = 0.85), as shown in Fig. 1A, and the corresponding dendrogram was generated in Fig. 1B. Five modules were identified, each color-coded, with genes within the same module being highly interconnected. The blue module exhibited the strongest positive association with sepsis (cor = 0.64, $p = 7 \times 10^{-30}$) and a negative association with controls (cor = –0.64, $p = 7 \times 10^{-30}$) (Fig. 1C). This module contained 902 core genes. Correlation analysis between module membership and gene significance for these core genes yielded a significant positive result (cor = 0.66, $p < 2.2 \times 10^{-16}$) (Fig. 1D; **Supplementary Table 2**).

Screening of Diagnostic Biomarkers

The intersection of RPRGs and genes from the Blue module yielded 23 RNA processing-related sepsis diagnostic genes (Fig. 2A). Functional enrichment analysis (GO and KEGG) of these 23 candidate genes revealed significant enrichment in RNA processing-related pathways such as "regulation of RNA splicing" and "RNA splicing" (Fig. 2B, **Supplementary Tables 3,4**). Additionally, correlation analysis of the candidate genes revealed significant and strong correlations among them ($p < 0.05$) (Fig. 2C). These genes exhibited significant differences in expression levels between sepsis patients and the control group, with notably higher expression in sepsis patients except for SRSF8 ($p < 0.05$) (Fig. 2D). GSVA scoring based on the 23 candidate genes in the training set revealed significantly higher scores in sepsis patients compared to the control group ($p = 2 \times 10^{-16}$) (Fig. 2E). This result was also observed in three exploratory datasets ($p < 0.05$) (Fig. 2F–H). Differential expression analysis performed on the training set revealed 1863 significant DEGs (**Supplementary Table 5**), and their intersection with the WGCNA blue-module genes

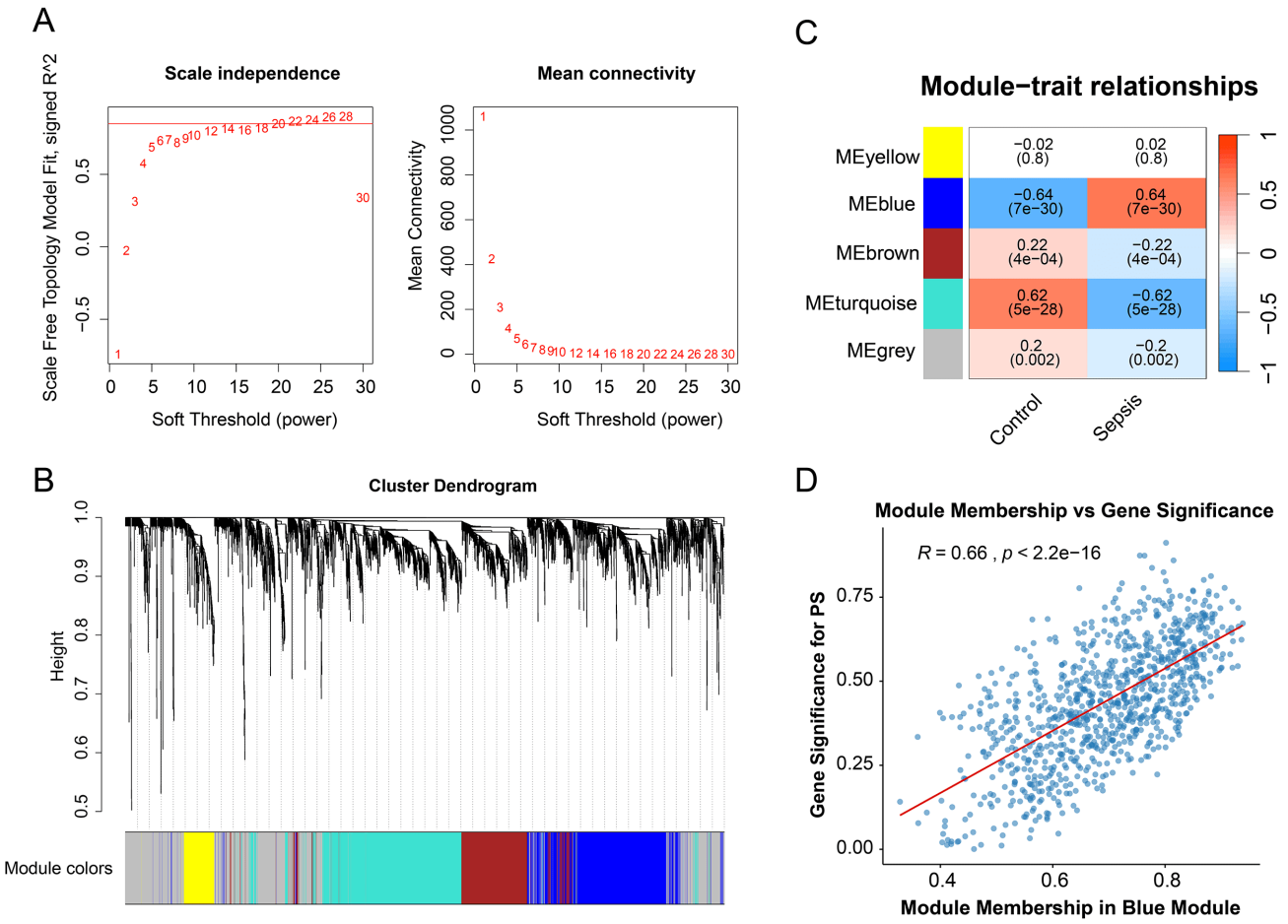

Fig. 1. Identification of key module genes. (A) WGCNA soft threshold analysis. (B) WGCNA sample module clustering diagram. (C) WGCNA Module-Trait relationship heatmap. (D) WGCNA blue module scatter plot. WGCNA, Weighted Gene Co-expression Network Analysis.

and RPRGs produced 13 differentially expressed RPRGs (DERPRGs) (Fig. 2I).

Diagnostic Biomarker Screening Via Machine Learning

To further screen key variables from the aforementioned 13 hub genes, 4 machine learning algorithms were comprehensively applied—SVM-RFE, LASSO regression, Boruta, and XGBoost. Under each algorithm, feature selection was performed based on the feature importance metrics of random forest (for Boruta), non-zero coefficients under the optimal regularization parameter of LASSO, feature importance scores from XGBoost, and the optimal feature subset identified by SVM-RFE (Fig. 3A–E). Ultimately, we identified five feature genes that were consistently selected by all four algorithms and designated them as robust diagnostic biomarkers for subsequent modeling and validation (Fig. 3F).

Model Development and Exploratory Evaluation

A multivariable binary logistic regression model was constructed using the five feature genes (*USB1*, *TDRD9*, *SRSF8*, *EXOSC4*, and *DPH3*). The final prediction model was expressed as follows:

$$\text{Logit}(P) = -30.4543 + 2.3358 \times USB1 + 0.6542 \times TDRD9 + 0.3357 \times SRSF8 + 1.4874 \times EXOSC4 + 0.817 \times DPH3$$

Based on this model, a nomogram was constructed using the *rms* package (Fig. 4A). The optimal cutoff value was 0.823, with a sensitivity of 93.97% and a specificity of 100.00% (**Supplementary Table 6**). The calibration curves demonstrated a high consistency between the model-predicted probabilities and the actual observed probabilities (Fig. 4B). DCA demonstrated that the nomogram has strong predictive potential (Fig. 4C). DCA demonstrated a positive net benefit of the nomogram across a range of threshold probabilities in all three exploratory datasets (GSE13904, GSE26378, and GSE26440), further supporting its discriminatory performance in the evaluated datasets

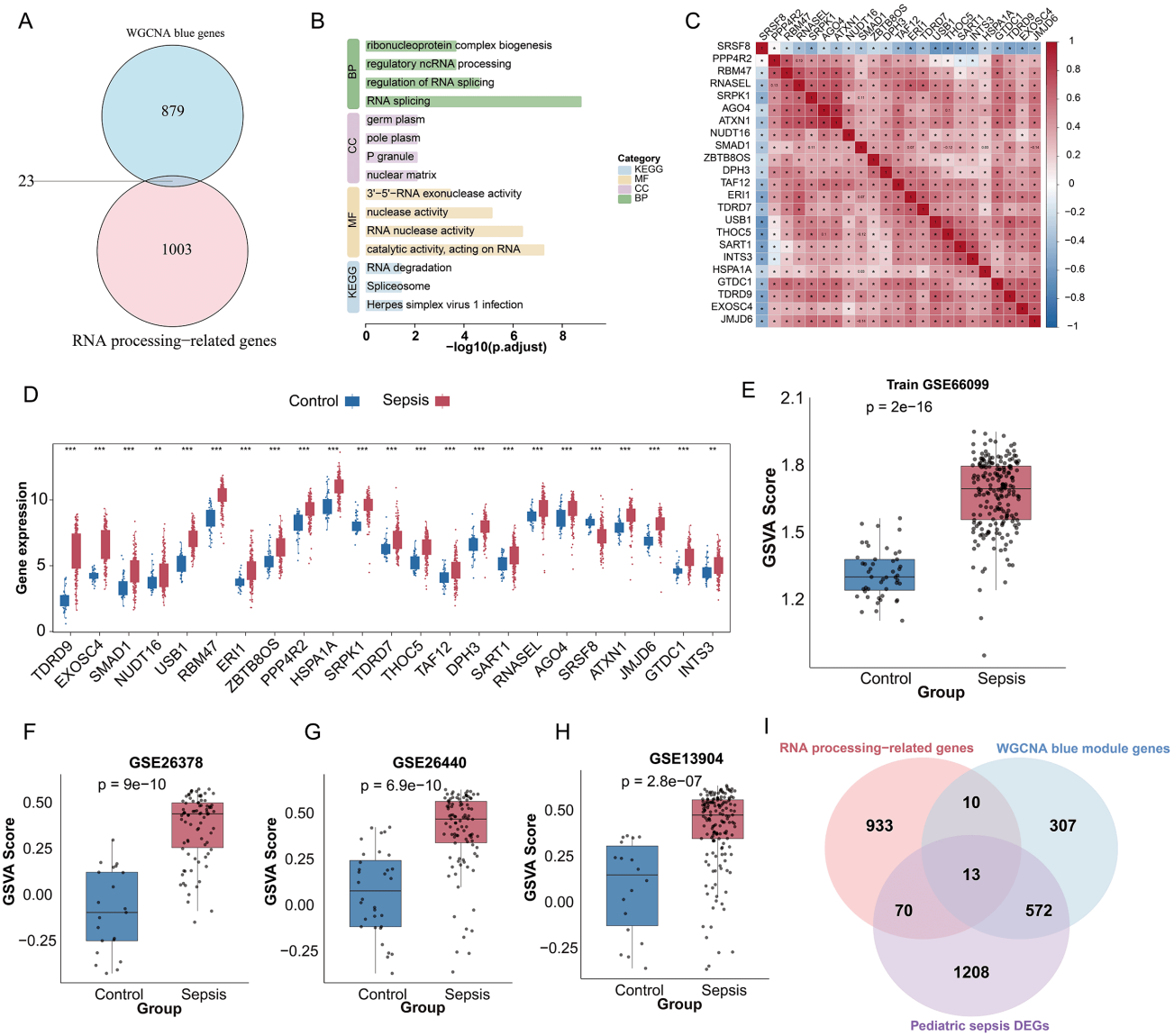

Fig. 2. Screening of differentially expressed RNA processing-related genes (DERPRGs). (A) Overlap of RPRGs with blue-module genes. (B) Functional annotation of the 23 candidates genes via Gene Ontology (GO)/Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment. (C) Correlation matrix heatmap for the 23 candidate genes. (D) Differential expression of candidate genes between controls (n = 47) and sepsis patients (n = 199). (E) Gene Set Variation Analysis (GSVA) box plot comparing control and sepsis groups in the training set (Control: n = 47; Sepsis: n = 199). GSVA box plots for three datasets GSE26378 (Control: n = 21, Sepsis: n = 82) (F), GSE26440 (Control: n = 32, Sepsis: n = 98) (G), and GSE13904 (Control: n = 18, Sepsis: n = 158) (H). (I) Intersection of RPRGs, Blue module genes, and sepsis DEGs. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

(**Supplementary Fig. 2A–C**). Furthermore, calibration curves showed agreement between predicted and observed probabilities in these exploratory datasets (**Supplementary Fig. 2D–F**). Furthermore, ROC curve analysis demonstrated that the overall nomogram model ("ALL") achieved an outstanding AUC of 0.987 in the training set, while each individual feature gene also exhibited high predictive accuracy (AUC range: 0.865–0.965; Fig. 4D). To assess the robustness of the diagnostic model and correct for potential optimism, we performed internal validation using 1000 bootstrap resamples. After 1000 bootstrap iterations, the optimism-corrected mean AUC was 0.976 (95% CI: 0.937–0.996), with a mean sensitivity of 0.989 (95% CI: 0.923–1.000) and a mean specificity of 0.919 (95% CI: 0.836–0.986) (**Supplementary Fig. 3**, **Supplementary Table 7**). The diagnostic performance of both the combined model and individual biomarkers was further evaluated in three exploratory datasets: GSE26378 (overall AUC = 0.991; Fig. 4E), GSE26440 (overall AUC = 0.986; Fig. 4F), and GSE13904 (overall AUC = 0.962; Fig. 4G). In addition to the overall AUC, 95% confidence intervals for sensitivity and specificity were also calculated for each

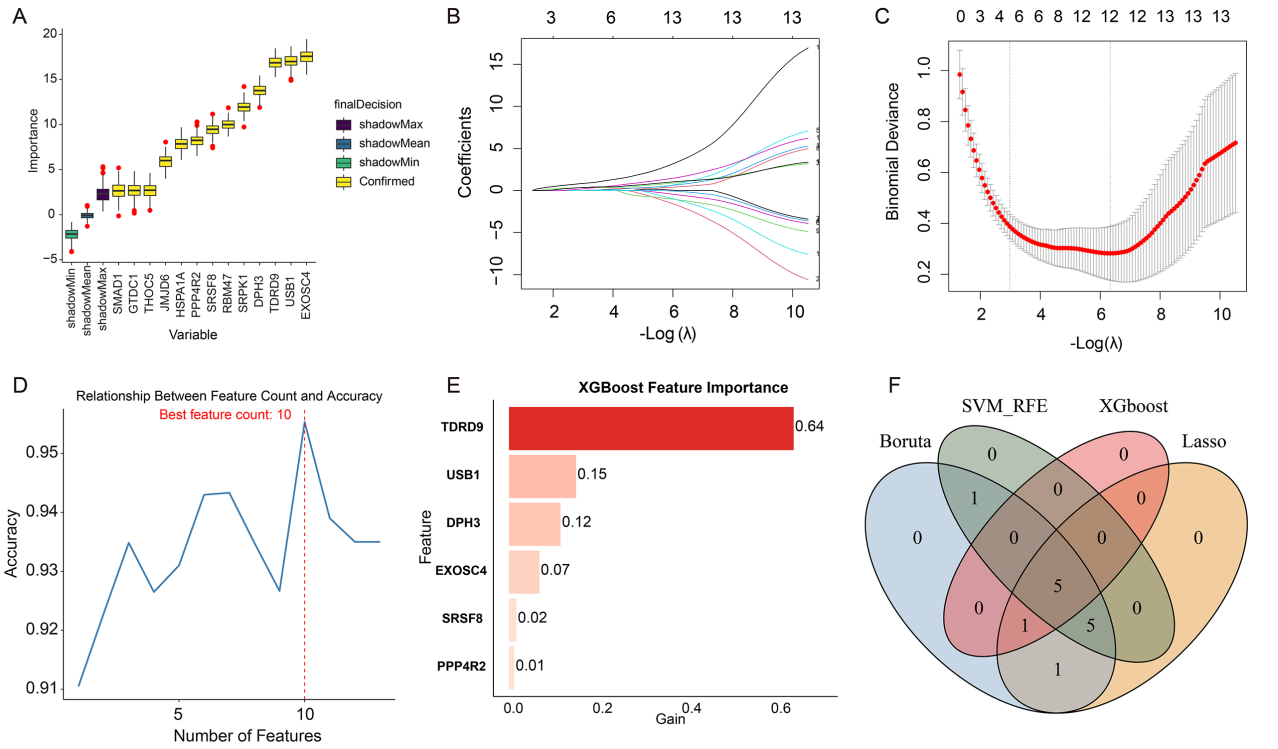

Fig. 3. Signature gene screening. (A) Box plot of variable importance analysis using the Boruta algorithm. (B) Least Absolute Shrinkage and Selection Operator (LASSO) coefficient profiles for diagnostic feature selection. (C) Coefficient distribution plot generated for the sequence of log(λ) in the LASSO model. (D) Line chart of Support Vector Machine–Recursive Feature Elimination (SVM-RFE) feature selection. (E) Top-ranked genes by feature importance scores in the extreme gradient boosting (XGBoost) algorithm. (F) Intersection analysis of the four algorithms.

datasets and are summarized in **Supplementary Table 8**. Comparison of feature gene expression between sepsis patients and controls in the training set showed that USB1, TDRD9, EXOSC4, and DPH3 were significantly upregulated, whereas SRSF8 was downregulated (Fig. 4H). Similar expression patterns were observed in the overlapping exploratory datasets (Fig. 4I–K).

Immune Infiltration Analysis

Sepsis and control groups displayed significantly distinct immune infiltration profiles (Fig. 5A). Monocytes and neutrophils were more abundant in sepsis, whereas naive B cells, resting mast/NK cells, and resting/naive CD4 T cells were enriched in controls ($p < 0.05$, Fig. 5B, **Supplementary Table 9**). ssGSEA confirmed higher control-group scores for B cells, $CD8^+$ T cells, immune checkpoints, and NK cells, and higher sepsis scores for macrophages and neutrophils ($p < 0.05$, Fig. 5C,D, **Supplementary Table 10**). Correlation analysis showed that USB1, TDRD9, EXOSC4 and DPH3 were positively correlated with neutrophils and M0 macrophages, but negatively correlated with resting NK cells, $CD8^+$ T cells, naive B cells, resting/-naive CD4 T cells. In contrast, SRSF8 exhibited opposite correlation patterns with these cell types ($p < 0.05$, Fig. 5E–I, **Supplementary Table 11**).

Single Gene Enrichment Analysis

In the GO enrichment analysis, nearly all feature genes (*USB1*, *TDRD9*, *SRSF8*, *EXOSC4*, and *DPH3*) were significantly enriched in various neutrophil-related granule components, including secretory granule membrane, specific granule, specific granule membrane, tertiary granule, and tertiary granule membrane (Fig. 6A–E). KEGG pathway enrichment analysis further highlighted immune-related functions. Most genes (*USB1*, *TDRD9*, *SRSF8*, *EXOSC4*, *DPH3*) were enriched in Fc gamma R-mediated phagocytosis and lysosome pathways (Fig. 6A–E).

Subtype Analysis of Sepsis Based on Diagnostic Signature Genes

Using the five diagnostic genes, we performed consensus cluster analysis within the GSE66099 discovery cohort and stratified these sepsis patients into two exploratory molecular subtypes: Cluster 1 (118 samples) and Cluster 2 (81 samples) (Fig. 7A–C). PCA confirmed the clear separation of these subtypes (Fig. 7D). Expression profiling revealed that *USB1*, *TDRD9*, *EXOSC4*, and *DPH3* were significantly elevated in Cluster 1, while SRSF8 showed the opposite trend ($p < 0.05$, Fig. 7E). Regarding immune infiltration, ssGSEA indicated that Cluster 2 was enriched in B cells, $CD8^+$ T cells, and NK cells, whereas macrophages

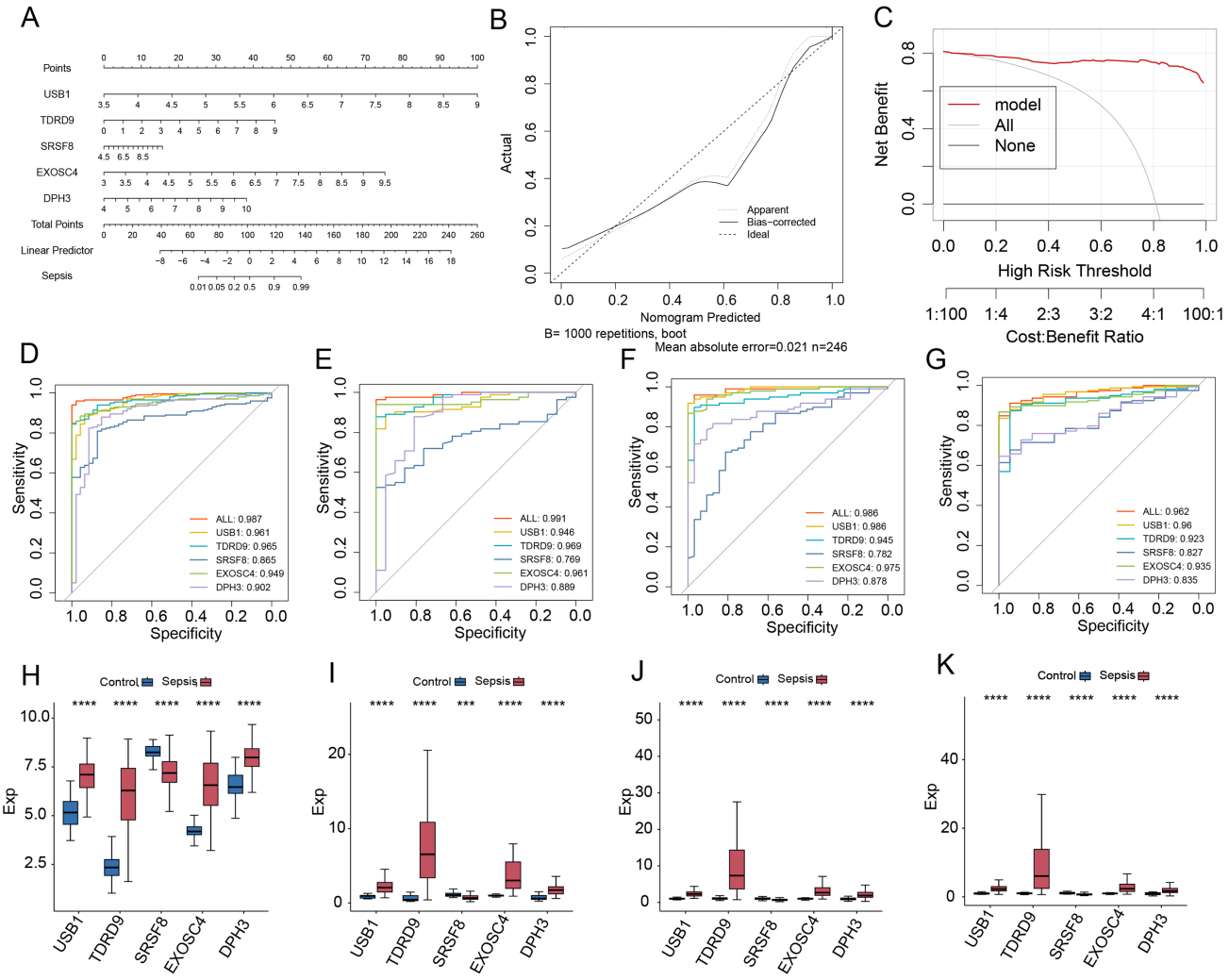

Fig. 4. Diagnostic model development and exploratory evaluation. (A) Nomogram of feature genes in the diagnostic model. (B) Calibration curve of the nomogram. (C) Decision curve analysis (DCA) curve of the nomogram. (D) Receiver operating characteristic (ROC) curves evaluating the diagnostic performance of the overall nomogram model ("ALL") and individual feature genes in the training set (Control: n = 47; Sepsis: n = 199). ROC curves of the overall model and individual feature genes in exploratory datasets GSE26378 (Control: n = 21, Sepsis: n = 82) (E), GSE26440 (Control: n = 32, Sepsis: n = 98) (F), and GSE13904 (Control: n = 18, Sepsis: n = 158) (G). (H) Box plots of feature gene expression in sepsis and control groups in the training set. Box plots of feature gene expression in sepsis and control groups in exploratory datasets GSE26378 (I), GSE26440 (J), and GSE13904 (K). The exploratory datasets contain substantial patient-level overlap with the training cohort and therefore do not constitute independent external validation. ***: $p < 0.001$, ****: $p < 0.0001$.

and neutrophils predominated in Cluster 1 ($p < 0.05$, Fig. 7F, **Supplementary Table 12**). CIBERSORT analysis also demonstrated higher M0 macrophage and monocyte levels in Cluster 1 ($p < 0.05$, Fig. 7G, **Supplementary Table 13**). A total of 432 upregulated and 154 downregulated genes were identified between the two clusters ($|\log_2 FC| > 1$, $p.adj < 0.05$, Fig. 7H).

Subtype Gene Enrichment Analysis

Functional annotation of upregulated genes revealed GO enrichment in immune activation-related terms (myeloid leukocyte activation, inflammatory response regulation, leukocyte activation, macrophage/neutrophil activation) and KEGG enrichment in phagocytosis-related (Fc gamma R-mediated), lysosomal, TNF, and complement/coagulation pathways (Fig. 8A,B). Conversely, downregulated genes were enriched in erythrocyte processes (differentiation, homeostasis, development, gas transport) and myeloid cell development/differentiation via GO, and in hematopoietic cell lineage, cofactor/porphyrin/nitrogen metabolism, and primary immunodeficiency via KEGG (Fig. 8C,D). GSEA demonstrated that Cluster 1 was primarily involved in meiotic division, cell cycle inhibition, ATPase activity, and mitochondrial distri-

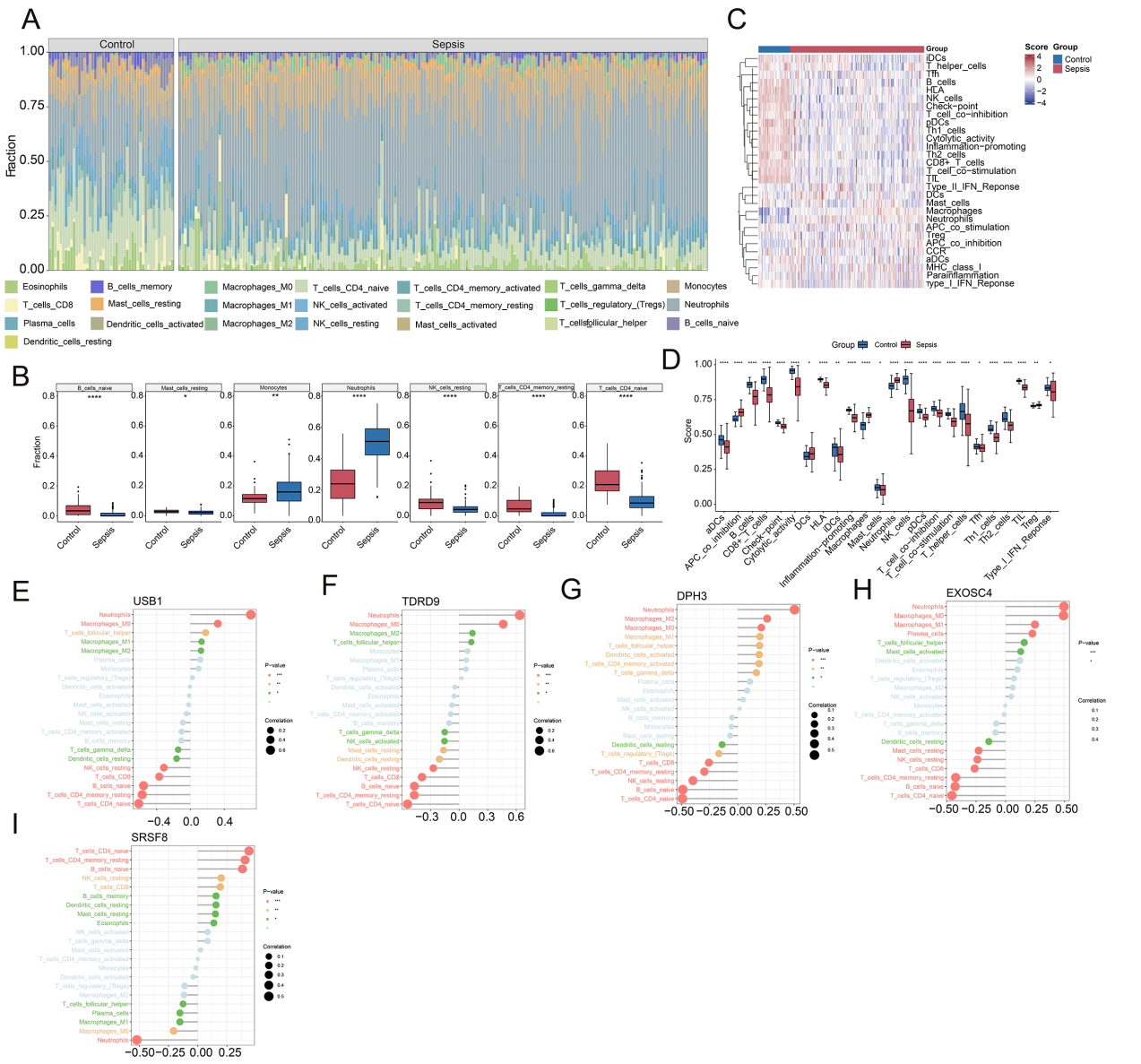

Fig. 5. Immune microenvironment characterization. (A) Distribution of immune cell types in control (n = 47) and sepsis (n = 199) samples. (B) Differential distribution of immune cell types between sepsis (n = 199) and control (n = 47) groups. (C) Heatmap of immune infiltration levels between control (n = 47) and sepsis (n = 199) patients. (D) Differences in immune infiltration levels between control (n = 47) and sepsis (n = 199) group. Correlation between *USB1* (E), *TDRD9* (F), *DPH3* (G), *EXOSC4* (H), *SRSF8* (I) gene and immune cell infiltration levels. The *p*-values shown in the main figure represent nominal (raw, unadjusted) *p*-values, with $p < 0.05$ considered nominally significant for these exploratory immune-related comparisons. Benjamini–Hochberg FDR-adjusted *p*-values for the corresponding comparisons are provided in **Supplementary Tables 9–11**. *: $p < 0.05$, **: $p < 0.01$, ****: $p < 0.0001$.

bution, whereas Cluster 2 was associated with connective tissue remodeling, peptidyl-L-cysteine S-palmitoylation, dendritic cell differentiation, and ER-cytosol transport (Fig. 8E,F).

Construction of Signature Gene-Associated Regulatory Networks

In the TFs regulatory network, *KLF9*, *KDM5A*, and *PPARG* were identified as key TFs interacting with multiple diagnostic genes, suggesting their potential roles in transcriptional regulation (Fig. 9A). Additionally, the five target genes (*USB1*, *TDRD9*, *EXOSC4*, *SRSF8*, and *DPH3*) were found to be regulated by a complex network of microRNAs (Fig. 9B).

Discussion

Pediatric sepsis presents major hurdles in both early detection and treatment owing to its multifaceted pathogenesis. By integrating WGCNA with multiple machine-

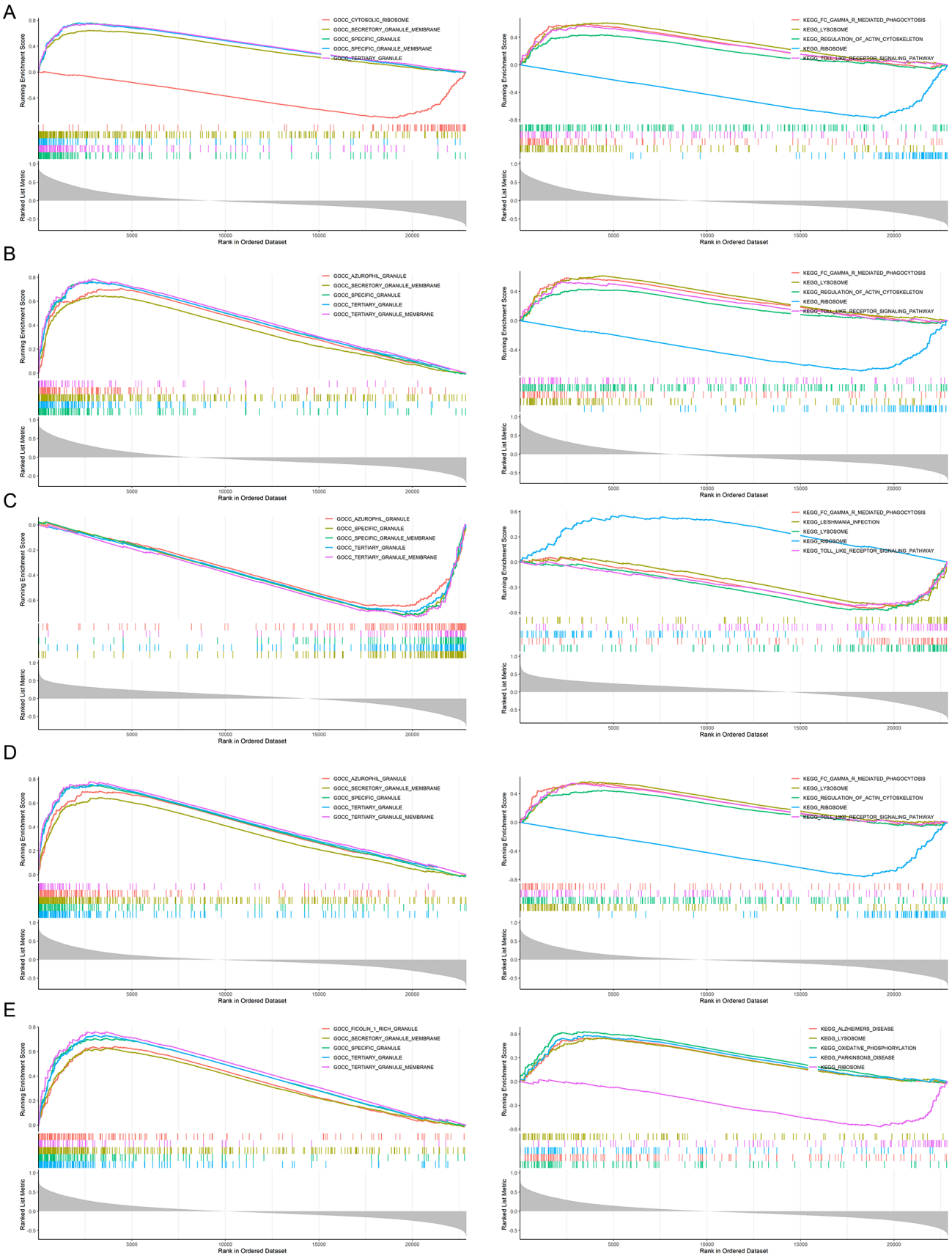

Fig. 6. Enrichment analysis of the five genes. GO (left) and KEGG (right) enrichment of the feature genes *USB1* (A), *TDRD9* (B), *SRSF8* (C), *EXOSC4* (D), and *DPH3* (E).

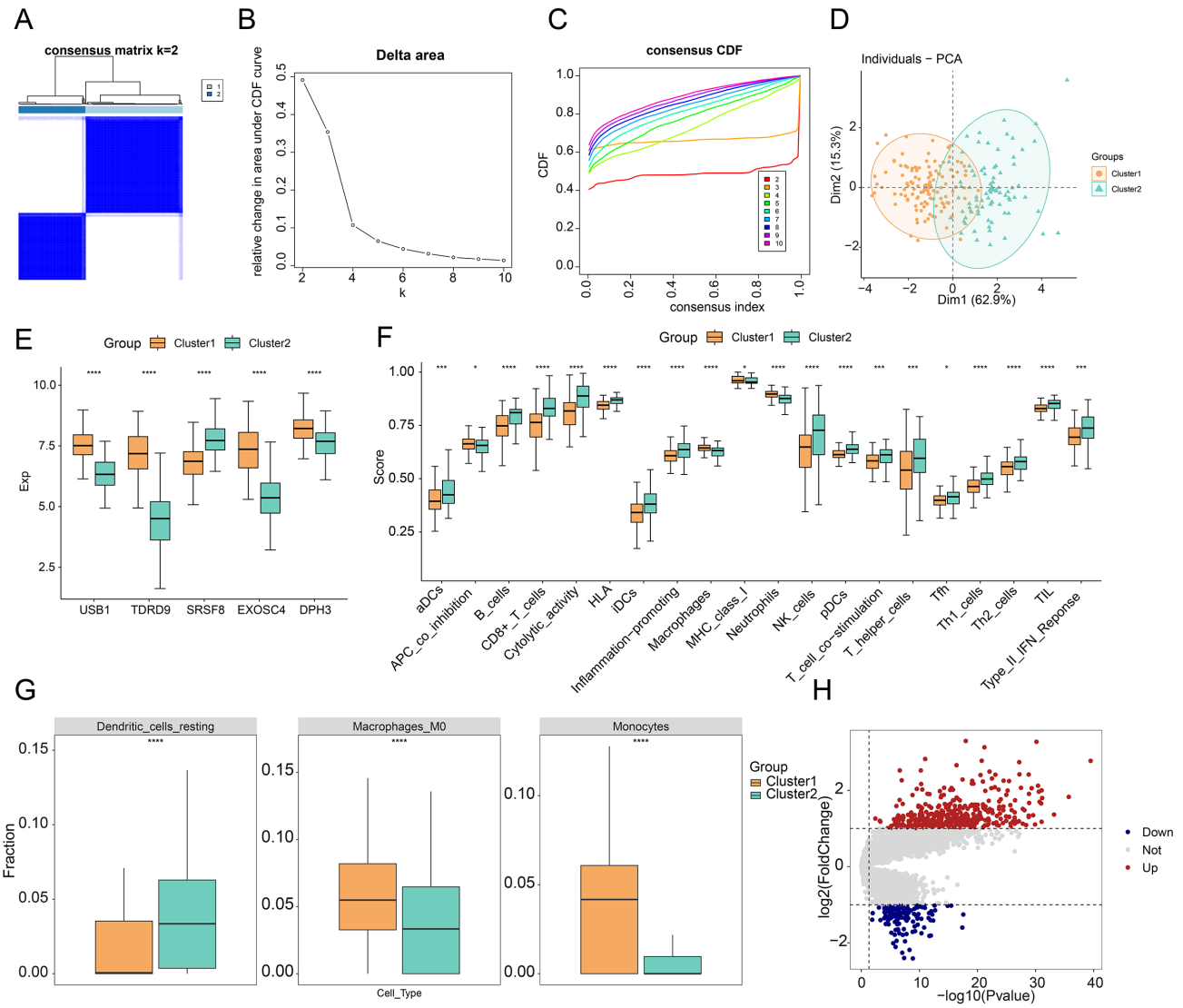

Fig. 7. Subtype analysis based on five diagnostic signature genes in the training set (total n = 199). (A) Consensus clustering matrix (k = 2), Cluster 1 (n = 118) and Cluster 2 (n = 81). (B) Delta area. (C) Cumulative distribution function (CDF) curves. (D) Principal component analysis (PCA) plot of different subtypes in the training set. (E) Expression of the five signature genes across different subtypes. (F) Box plots comparing immune cell/functional activity between subtypes. (G) Box plots comparing immune cell infiltration differences between subtypes. (H) Volcano plot of differentially expressed genes (DEGs) between subtypes. The *p*-values shown in the main figure represent nominal (raw, unadjusted) *p*-values, with $p < 0.05$ considered nominally significant for these exploratory comparisons. Benjamini–Hochberg FDR-adjusted *p*-values are provided in **Supplementary Tables 12,13**. *: $p < 0.05$, ***: $p < 0.001$, ****: $p < 0.0001$.

learning approaches, we uncovered a novel group of RPRGs that serve as candidate diagnostic biomarkers. Concurrently, immune profiling revealed distinct molecular subtypes among sepsis patients, highlighting a tight link between RPRG-related signals and compromised immune responses. These insights contribute to a better understanding of the underlying mechanisms and support the future design of diagnostic and therapeutic strategies targeting RNA processing and immune dysfunction.

Our study identified five diagnostic genes (*TDRD9*, *SRSF8*, *EXOSC4*, *DPH3*, *USB1*) that may play important roles in the pathogenesis and progression of pediatric sepsis. *TDRD9* (Tudor domain-containing protein 9) is a protein-coding gene that encodes an ATP-dependent RNA helicase and belongs to the Tudor domain-containing protein family [19]. Although a previous study in neonatal sepsis reported that *TDRD9* expression levels were negatively correlated with M0/M2 macrophage infiltration [20], our analysis in pediatric sepsis revealed that *TDRD9* was positively correlated with M0 macrophage infiltration and neutrophils. This discrepancy may stem from differences in cohort demographics or disease stages between neonatal

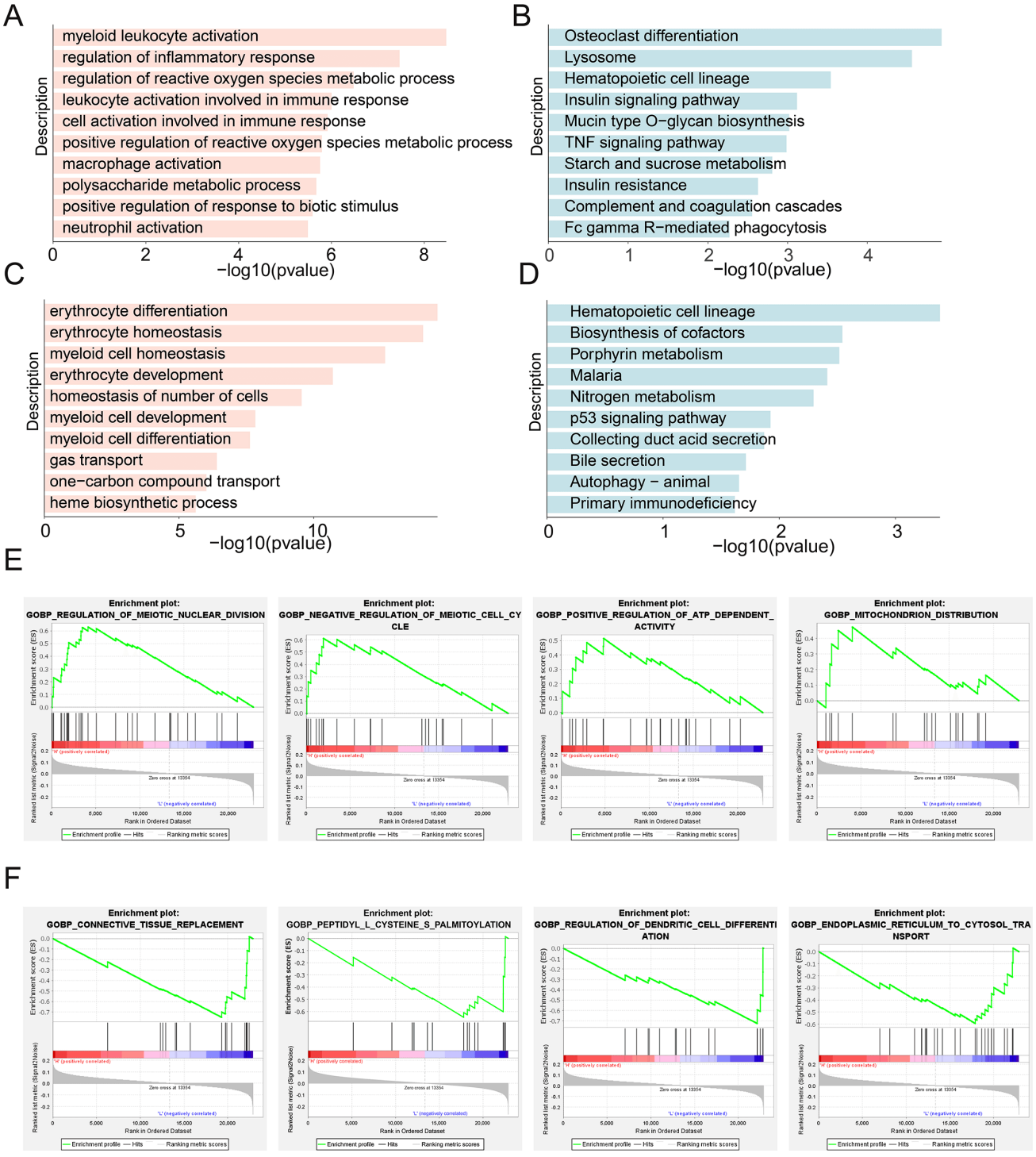

Fig. 8. Gene enrichment analysis. GO (A) and KEGG (B) enrichment analysis of upregulated DEGs. GO (C) and KEGG (D) enrichment analysis of downregulated DEGs. GSEA enrichment results for Cluster 1 (E) and Cluster 2 (F).

and pediatric sepsis. Nevertheless, *TDRD9* may contribute to immune dysregulation in sepsis by modulating innate immune activation or macrophage polarization [20]. *SRSF8* (serine/arginine-rich splicing factor 8) is a protein-coding gene located on chromosome 11q21 and belongs to the SR (serine/arginine-rich) splicing factor family [21]. Splicing imbalance is a crucial aspect of inflammatory gene regulation during the acute phase of sepsis [22]. Research indicates that *SRSF8* directly binds to the CLK1 exon 4 enhancer and recruits the activation complex, promoting its inclusion and making it a key positive regulator of this alternative splicing event [23]. Therefore, *SRSF8* likely plays a significant role in sepsis as a splicing factor. The protein encoded by the *EXOSC4* gene is a key component of

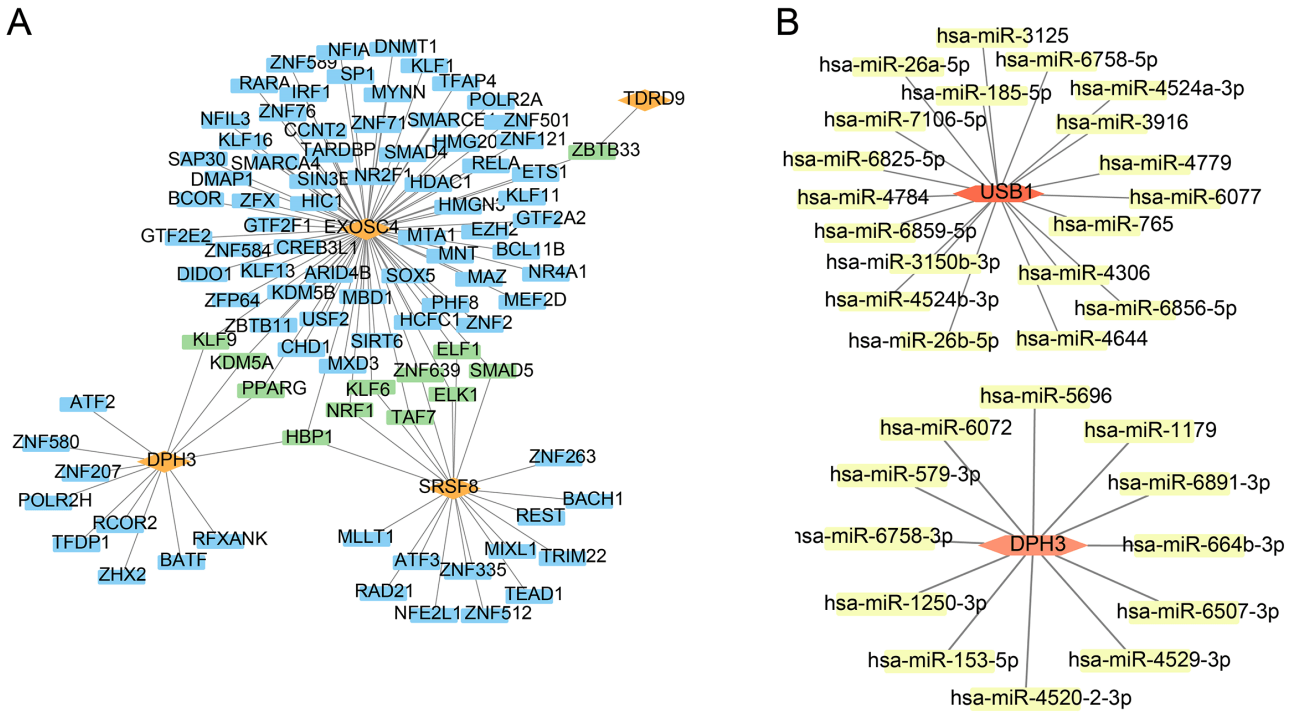

Fig. 9. Construction of regulatory networks based on signature genes. (A) NetworkAnalyst-predicted Transcription factor (TF)-gene interaction network. (B) miRNA-mRNA regulatory network constructed from the signature genes.

intracellular RNA metabolism and quality control, regulating the processing, maturation, and degradation of various RNAs through its participation in the RNA exosome complex [24]. *EXOSC4* was significantly upregulated in sepsis samples and, together with *EXOSC5*, was co-enriched in the RNA degradation pathway, suggesting that it may participate in the pathological process of sepsis by regulating RNA degradation [25]. *DPH3* encodes an 82-amino acid zinc finger protein that catalyzes the diphthamide modification of histidine in eEF2, thereby ensuring translational fidelity and influencing sensitivity to diphtheria/P. aeruginosa exotoxin A-mediated ADP-ribosylation [26]. These five diagnostic genes may play important roles in pediatric sepsis pathogenesis through their modulation of RNA metabolism and immune responses.

To further strengthen our findings, it is important to note that the identified RNA processing-related diagnostic biomarkers also exhibit altered expression and functional relevance in other inflammatory and autoimmune diseases. For instance, *EXOSC4* has recently been identified as a novel autoantigen significantly upregulated in Rheumatoid Arthritis-associated Interstitial Lung Disease, highlighting its broader involvement in autoimmune pathogenesis [27]. Similarly, *DPH3* is overexpressed in tissues of patients with Rheumatoid Arthritis and Polycystic Ovary Syndrome, acting as a core gene that bridges metabolic and systemic inflammatory dysregulation [28]. *TDRD9* functions as a crucial hub gene implicated in the progression of Kawasaki disease, an acute systemic immune vasculitis [29]. Furthermore, while *USB1* mutations are classically associated with Poikiloderma with Neutropenia, recent studies have revealed that this condition is heavily characterized by a systemic inflammatory cytokine profile and impaired monocyte plasticity [30]. Finally, as a member of the Serine/Arginine-Rich Splicing Factor family, *SRSF8* may represent part of a broader mechanism involving alternative splicing dysregulation, which has been implicated as a key contributor to various systemic autoimmune diseases [31]. Collectively, the involvement of these markers across diverse inflammatory conditions highlights their potential roles in immune regulation and further supports their robust diagnostic potential in Pediatric sepsis.

This study also found that neutrophil and monocyte counts were significantly higher in the sepsis group than in the control group, indicating a hyperinflammatory phenotype. Consistent with these findings, previous studies have shown that neutrophils are central drivers of immunothrombosis and organ dysfunction in sepsis. Specifically, their overactivation leads to an excessive release of Neutrophil Extracellular Traps (NETs), which are key factors in endothelial injury [32]. Monocytes can rapidly differentiate into macrophages/dendritic cells, clearing pathogens and secreting pro-inflammatory factors, thereby contributing to anti-infection and tissue repair [33,34]. However, persistent expansion can lead to excessive release of TNF-α and IL-6 or induce immunosuppressive subsets, driving disease progression [35]. In contrast, the control group showed increased numbers of naive B cells, naïve CD4 T cells, rest-

ing memory CD4 T cells, and resting NK cells, reflecting an intact lymphocyte reservoir. Naive B cells and naïve CD4 T cells are antigen-inexperienced cells, and their maintenance suggests well-preserved thymic and bone marrow output, supporting effective responses to novel antigens [36,37]. Resting memory CD4 T cells represent a "memory reservoir" retained after previous antigen exposure and can rapidly initiate secondary responses upon re-infection [38].

Furthermore, RNA processing mechanisms play central regulatory roles in the immunopathology of pediatric sepsis. Diagnostic genes such as *TDRD9*, *SRSF8*, and *EXOSC4* are collectively involved in RNA metabolic regulation. Notably, the universal enrichment of these biomarkers in neutrophil granule pathways (tertiary/specific granules) provides a mechanistic explanation for the hallmark pathological feature of sepsis—neutrophil overactivation [39]. The strong correlations of *USB1*, *TDRD9* and *EXOSC4* with M0 macrophage infiltration suggest that these genes may be associated with pro-inflammatory phenotypes through post-transcriptional reprogramming. M0 macrophages can rapidly release IL-1β, TNF-α, and CXCL8 via the TLR-NF-κB axis, promoting M1 polarization and amplifying inflammatory cascades, thereby exacerbating systemic inflammation [40]. The distinct expression patterns of these biomarkers provide further evidence of immune heterogeneity in pediatric sepsis. Specifically, Cluster 1 represented a hyperinflammatory phenotype dominated by innate immune activation, whereas Cluster 2 exhibited relatively preserved adaptive immunity. It is worth noting that the two subtypes identified in our study were derived from a single training cohort and have not been externally validated. Although these findings raise the hypothesis that tailored immunomodulatory approaches could be considered for each subtype, such speculation requires validation through prospective clinical studies with treatment and outcome data. At present, our classification primarily provides a foundation for understanding sepsis heterogeneity and designing future stratified clinical trials, rather than offering direct therapeutic recommendations.

KDM5A, as a hub transcriptional regulator, may play a crucial role in sepsis by regulating multiple downstream genes. During sepsis, cytokine-induced activation of STAT3, particularly by IL-6 and IL-10, upregulates SOCS3 and anti-apoptotic proteins, suppressing the release of pro-inflammatory factors such as TNF-α and IL-1β. Meanwhile, *STAT3* activation can reduce HLA-DR expression and promote the expansion of myeloid-derived suppressor cells [41]. This mechanism may provide anti-inflammatory protection during the early stages of sepsis; however, persistent activation can exert a "double-edged sword" effect by inducing immunosuppression and increasing susceptibility to secondary infections [42]. *KDM5A* has also been reported to remove H3K4me3 marks from the promoters of *PTEN* and *SOCS1*, potentially disrupting PI3K/AKT-mediated negative feedback and sustaining STAT3 activation, thereby contributing to prolonged inflammatory responses [43].

Although this study successfully identified a set of RPRG diagnostic biomarkers associated with pediatric sepsis through multi-level bioinformatics analysis and machine learning algorithms, and further classified molecular subtypes with distinct immune characteristics, several key limitations remain. First, the molecular subtypes were derived solely from the GSE66099 discovery dataset without independent subtyping validation in external cohorts. Consequently, these subtypes should be considered exploratory, and their reproducibility, stability, and potential clinical utility as stratification tools require prospective validation across larger, multi-center pediatric sepsis cohorts. Second, due to the inherent limitations of retrospective public databases, detailed clinical outcome information (such as mortality rates or disease severity scores) was largely unavailable. Therefore, this study primarily highlights the diagnostic utility of these biomarkers. Future prospective clinical cohorts with comprehensive follow-up are urgently needed to further evaluate their prognostic value and associations with patient outcomes. Lastly, although the identified RPRGs showed significant correlations with immune features, including neutrophil and macrophage infiltration, experimental validation is still lacking to clarify their specific biological functions and molecular mechanisms. Future studies should validate the expression levels and diagnostic performance of these genes through *in vivo* and *in vitro* experiments, including qPCR-based verification and functional assays. Moreover, as our analysis was based solely on transcriptomic data, the relationship between the mRNA levels of these biomarkers and their corresponding protein abundance remains unverified. Given the complexity of post-transcriptional and translational regulation, future studies incorporating targeted proteomics and wet-lab validation approaches (e.g., Western blotting or ELISA) are required to confirm the protein-level expression of these identified genes. Furthermore, the lack of longitudinal data limits our understanding of the dynamic expression trajectories of these biomarkers across different disease stages (e.g., 0–6 h, 6–24 h, and >24 h). Since our models were developed using samples collected within 24 hours of admission, their diagnostic applicability is currently restricted to the early acute phase, highlighting the need for future time-series investigations. Finally, publicly available pediatric sepsis microarray datasets exhibit patient-level overlap, and the corresponding results should therefore be regarded as exploratory. These datasets were derived from related pediatric sepsis cohorts, used the same microarray platform (GPL570), and had similar inclusion criteria. Therefore, independent validation in newly recruited populations using clinically applicable testing platforms is still required.

An additional limitation concerns the composition of the control group. In this study, healthy children were

used as the control group, and patients with non-infectious SIRS were excluded to minimize potential confounding from non-infectious systemic inflammation. Consequently, the high diagnostic performance observed in the training set and exploratory datasets primarily reflects the ability of the five-gene signature to discriminate pediatric sepsis from healthy controls. It remains unclear whether these biomarkers can differentiate sepsis from non-infectious SIRS or other clinically relevant inflammatory or infectious diseases. Therefore, the current findings should be regarded as preliminary evidence of diagnostic discrimination rather than as definitive evidence for widespread clinical application. Future prospective, multicenter studies that include patients with non-infectious SIRS and other clinically relevant differential diagnoses are needed to more rigorously evaluate the sensitivity, specificity, and added diagnostic value of this signature in real-world clinical settings. Another limitation is the variation in diagnostic performance across external cohorts, particularly the relatively lower sensitivity observed in some datasets, which may reflect differences in patient characteristics, disease severity, sampling time, and technical factors. Therefore, the five-gene signature remains a candidate research model whose clinical diagnostic utility has not been established. Future prospective, multicenter studies with standardized protocols are needed to further validate and recalibrate the model.

Conclusion

In summary, we developed and internally evaluated a five-gene signature associated with RNA processing, which showed high discrimination between sepsis samples and healthy controls in the training cohort. Furthermore, exploratory subtype analysis in the training cohort stratified patients into two distinct subgroups with different immune infiltration profiles, providing preliminary insights into the immunopathological mechanisms of sepsis. These findings warrant further validation in independent prospective cohorts.

Availability of Data and Materials

The data and materials in the current study are available from the corresponding author on reasonable request.

Author Contributions

Conception and design: MH. Administrative support: MH. Provision of study materials: MH, LLZ. Collection and assembly of data: MH, LLZ. Data analysis and interpretation: MH. MH and LLZ have been involved in drafting the manuscript and both authors have been involved in revising it critically for important intellectual content. Both authors gave final approval of the version to be published. Both authors have participated sufficiently in the work to take public responsibility for appropriate portions of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.24976/Discov.Med.202638214.279.

References

[1] Schlapbach LJ, Watson RS, Sorce LR, Argent AC, Menon K, Hall MW, et al. International Consensus Criteria for Pediatric Sepsis and Septic Shock. JAMA. 2024; 331: 665–674. https://doi.org/10.1001/jama.2024.0179

[2] Burgunder L, Heyrend C, Olson J, Stidham C, Lane RD, Workman JK, et al. Medication and Fluid Management of Pediatric Sepsis and Septic Shock. Paediatric Drugs. 2022; 24: 193–205. https://doi.org/10.1007/s40272-022-00497-z

[3] Briassouli E, Syrimi N, Ilia S. Hyperferritinemia and Macrophage Activation Syndrome in Septic Shock: Recent Advances with a Pediatric Focus (2020-2025). Children (Basel, Switzerland). 2025; 12: 1193. https://doi.org/10.3390/children12091193

[4] Menon K, Schlapbach LJ, Akech S, Argent A, Biban P, Carrol ED, et al. Criteria for Pediatric Sepsis-A Systematic Review and Meta-Analysis by the Pediatric Sepsis Definition Taskforce. Critical Care Medicine. 2022; 50: 21–36. https://doi.org/10.1097/CCM.0000000000005294

[5] Sanyaolu A, Patidar R, Ayodele O, Marinkovic A, Desai P. Pediatric Sepsis: The Importance of Understanding Criteria for Diagnosis. Pediatric Annals. 2022; 51: e405–e408. https://doi.org/10.3928/19382359-20220803-07

[6] Hidalgo JL, Kumar VK, Akech SO, Myatra SN, Jacob ST, Kortz T, et al. The Sepsis Chain of Survival: A Comprehensive Framework for Improving Sepsis Outcomes. Critical Care Medicine. 2025; 53: e1886–e1892. https://doi.org/10.1097/CCM.0000000000006796

[7] Dong R, Liu W, Weng L, Yin P, Peng J, Chen Y, et al. Temporal trends of sepsis-related mortality in China, 2006-2020: a population-based study. Annals of Intensive Care. 2023; 13: 71. https://doi.org/10.1186/s13613-023-01166-1

[8] Fleischmann-Struzek C, Goldfarb DM, Schlattmann P, Schlapbach LJ, Reinhart K, Kissoon N. The global burden of paediatric and neonatal sepsis: a systematic review. The Lancet. Res-

piratory Medicine. 2018; 6: 223–230. https://doi.org/10.1016/S2213-2600(18)30063-8

[9] Neil CR, Seiler MW, Reynolds DJ, Smith JJ, Vaillancourt FH, Smith PG, et al. Reprogramming RNA processing: an emerging therapeutic landscape. Trends in Pharmacological Sciences. 2022; 43: 437–454. https://doi.org/10.1016/j.tips.2022.02.011

[10] Szeto RA, Tran T, Truong J, Negraes PD, Trujillo CA. RNA processing in neurological tissue: development, aging and disease. Seminars in Cell & Developmental Biology. 2021; 114: 57–67. https://doi.org/10.1016/j.semcdb.2020.09.004

[11] Modestov A, Buzdin A, Suntsova M. Unveiling RNA Editing by ADAR and APOBEC Protein Gene Families. Frontiers in Bioscience (Landmark Edition). 2025; 30: 26298. https://doi.org/10.31083/FBL26298

[12] Zuniga G, Frost B. Selective neuronal vulnerability to deficits in RNA processing. Progress in Neurobiology. 2023; 229: 102500. https://doi.org/10.1016/j.pneurobio.2023.102500

[13] Welsh SA, Gardini A. Genomic regulation of transcription and RNA processing by the multitasking Integrator complex. Nature Reviews. Molecular Cell Biology. 2023; 24: 204–220. https://doi.org/10.1038/s41580-022-00534-2

[14] Wu G, Rouvière JO, Schmid M, Heick Jensen T. RNA 3'end tailing safeguards cells against products of pervasive transcription termination. Nature Communications. 2024; 15: 10446. https://doi.org/10.1038/s41467-024-54834-6

[15] Shine M, Gordon J, Schärfen L, Zigackova D, Herzel L, Neugebauer KM. Co-transcriptional gene regulation in eukaryotes and prokaryotes. Nature Reviews. Molecular Cell Biology. 2024; 25: 534–554. https://doi.org/10.1038/s41580-024-00706-2

[16] Abuduaini B, Jiyuan Z, Rehati A, Liang Z, Yunlin S. Regulation of Alternative Splicing of Lipid Metabolism Genes in Sepsis-Induced Liver Damage by RNA-Binding Proteins. Inflammation. 2024; 47: 1952–1968. https://doi.org/10.1007/s10753-024-02017-2

[17] Liu D, Sun W, Zhang D, Yu Z, Qin W, Liu Y, et al. Long noncoding RNA GSEC promotes neutrophil inflammatory activation by supporting PFKFB3-involved glycolytic metabolism in sepsis. Cell Death & Disease. 2021; 12: 1157. https://doi.org/10.1038/s41419-021-04428-7

[18] Hu B, Zhang Y, Jiang B, Li A. A Bioinformatics Analysis on the Subtypes of Hepatocellular Carcinoma Related to RNA Processing Genes to Reveal Prognosis and Immune Microenvironment Features. Digestive Diseases and Sciences. 2025; 70: 4174–4189. https://doi.org/10.1007/s10620-025-09261-y

[19] Wang W, Feng Y, Dong J, Zhou Z, Jing J, Li Z, et al. A Novel Compound Heterozygous Mutation in TDRD9 Causes Oligozoospermia. Reproductive Sciences (Thousand Oaks, Calif.). 2024; 31: 3413–3419. https://doi.org/10.1007/s43032-024-01665-x

[20] Liao W, Xiao H, He J, Huang L, Liao Y, Qin J, et al. Identification and verification of feature biomarkers associated with immune cells in neonatal sepsis. European Journal of Medical Research. 2023; 28: 105. https://doi.org/10.1186/s40001-023-01061-2

[21] Zhang Y, Yu X, Sun R, Min J, Tang X, Lin Z, et al. Splicing factor arginine/serine-rich 8 promotes multiple myeloma malignancy and bone lesion through alternative splicing of CACYBP and exosome-based cellular communication. Clinical and Translational Medicine. 2022; 12: e684. https://doi.org/10.1002/ctm2.684

[22] Tuerdimaimaiti D, Abuduaini B, Kang S, Jiao J, Li M, Madeniyati W, et al. Genome-wide identification and functional analysis of dysregulated alternative splicing profiles in sepsis. Journal of Inflammation (London, England). 2023; 20: 31. https://doi.org/10.1186/s12950-023-00355-w

[23] Shkreta L, Delannoy A, Toutant J, Chabot B. Regulatory interplay between SR proteins governs *CLK1* kinase splice variants production. RNA (New York, N.Y.). 2024; 30: 1596–1607. https://doi.org/10.1261/rna.080107.124

[24] Taniue K, Tanu T, Shimoura Y, Mitsutomi S, Han H, Kakisaka R, et al. RNA Exosome Component EXOSC4 Amplified in Multiple Cancer Types Is Required for the Cancer Cell Survival. International Journal of Molecular Sciences. 2022; 23: 496. https://doi.org/10.3390/ijms23010496

[25] Dong L, Li H, Zhang S, Yang G. miR-148 family members are putative biomarkers for sepsis. Molecular Medicine Reports. 2019; 19: 5133–5141. https://doi.org/10.3892/mmr.2019.10174

[26] Zhang Y, Su D, Dzikovski B, Majer SH, Coleman R, Chandrasekaran S, et al. Dph3 Enables Aerobic Diphthamide Biosynthesis by Donating One Iron Atom to Transform a [3Fe-4S] to a [4Fe-4S] Cluster in Dph1-Dph2. Journal of the American Chemical Society. 2021; 143: 9314–9319. https://doi.org/10.1021/jacs.1c03956

[27] Yao C, Gong X, Shang B, Dou D, Liu W, Xu H, et al. Identification EXOSC4 as a novel autoantigen of interstitial lung disease in rheumatoid arthritis. Journal of Translational Medicine. 2025; 23: 765. https://doi.org/10.1186/s12967-025-06667-0

[28] Ji Y, Xia P, Wang Y, Liu Y, Wang F, Sun F, et al. Shared pathogenesis in polycystic ovaries and rheumatoid arthritis: an analysis of key genes and pathways. Frontiers in Genetics. 2025; 16: 1554139. https://doi.org/10.3389/fgene.2025.1554139

[29] Ba H, Zhang L, Peng H, He X, Lin Y, Li X, et al. Identification of Hub Biomarkers and Immune and Inflammation Pathways Contributing to Kawasaki Disease Progression with RT-qPCR Verification. Journal of Immunology Research. 2023; 2023: 1774260. https://doi.org/10.1155/2023/1774260

[30] Parajuli P, Craig DB, Gadgeel M, Bagla S, Wright RE, 3rd, Chu R, et al. Defective monocyte plasticity and altered cAMP pathway characterize USB1-mutated poikiloderma with neutropenia Clericuzio type. British Journal of Haematology. 2024; 204: 683–693. https://doi.org/10.1111/bjh.19128

[31] Li D, Yu W, Lai M. Towards understandings of serine/arginine-rich splicing factors. Acta Pharmaceutica Sinica. B. 2023; 13: 3181–3207. https://doi.org/10.1016/j.apsb.2023.05.022

[32] Zhang H, Wang Y, Qu M, Li W, Wu D, Cata JP, et al. Neutrophil, neutrophil extracellular traps and endothelial cell dysfunction in sepsis. Clinical and Translational Medicine. 2023; 13: e1170. https://doi.org/10.1002/ctm2.1170

[33] Hung HC, Tsai SF, Chou HW, Tsai MJ, Hsu PL, Kuo YM. Dietary fatty acids differentially affect secretion of pro-inflammatory cytokines in human THP-1 monocytes. Scientific Reports. 2023; 13: 5511. https://doi.org/10.1038/s41598-023-32710-5

[34] Ashraf MU, Kang MH, Lim YT, Yang S, Bae YS. Egr2-dependent Mo-DCs regulate immunosuppressive NK cells, promoting neutrophil-mediated protective immunity against acute Listeria infection. Molecules and Cells. 2025; 48: 100287. https://doi.org/10.1016/j.mocell.2025.100287

[35] Dash SP, Gupta S, Sarangi PP. Monocytes and macrophages: Origin, homing, differentiation, and functionality during inflammation. Heliyon. 2024; 10: e29686. https://doi.org/10.1016/j.heliyon.2024.e29686

[36] Gómez-Manríquez J, Hernández-Bello J, Muñoz-Valle JF, Sifuentes-Franco S, Graciano-Machuca O, Morales-Núñez JJ. B cell development: transcriptional regulation and immunological mechanisms in homeostasis. Frontiers in Immunology. 2025; 16: 1593338. https://doi.org/10.3389/fimmu.2025.1593338

[37] Sun L, Su Y, Jiao A, Wang X, Zhang B. T cells in health and disease. Signal Transduction and Targeted Therapy. 2023; 8: 235. https://doi.org/10.1038/s41392-023-01471-y

[38] Asada N, Ginsberg P, Paust HJ, Song N, Riedel JH, Turner JE, et al. The integrated stress response pathway controls cy-

tokine production in tissue-resident memory CD4$^+$ T cells. Nature Immunology. 2025; 26: 557–566. https://doi.org/10.1038/s41590-025-02105-x

[39] Kwok AJ, Allcock A, Ferreira RC, Cano-Gamez E, Smee M, Burnham KL, et al. Neutrophils and emergency granulopoiesis drive immune suppression and an extreme response endotype during sepsis. Nature Immunology. 2023; 24: 767–779. https://doi.org/10.1038/s41590-023-01490-5

[40] Dousdampanis P, Aggeletopoulou I, Mouzaki A. The role of M1/M2 macrophage polarization in the pathogenesis of obesity-related kidney disease and related pathologies. Frontiers in Immunology. 2025; 15: 1534823. https://doi.org/10.3389/fimmu.2024.1534823

[41] Williamson L, Ayalon I, Shen H, Kaplan J. Hepatic STAT3 inhibition amplifies the inflammatory response in obese mice during sepsis. American Journal of Physiology. Endocrinology and Metabolism. 2019; 316: E286–E292. https://doi.org/10.1152/ajpendo.00341.2018

[42] Zanders L, Kny M, Hahn A, Schmidt S, Wundersitz S, Todiras M, et al. Sepsis induces interleukin 6, gp130/JAK2/STAT3, and muscle wasting. Journal of Cachexia, Sarcopenia and Muscle. 2022; 13: 713–727. https://doi.org/10.1002/jcsm.12867

[43] Qu L, Yin T, Zhao Y, Lv W, Liu Z, Chen C, et al. Histone demethylases in the regulation of immunity and inflammation. Cell Death Discovery. 2023; 9: 188. https://doi.org/10.1038/s41420-023-01489-9