



## OPEN ACCESS

## EDITED BY

Yongqing Li,  
School of Medicine, Michigan Medicine,  
University of Michigan, United States

## REVIEWED BY

Shiyu Xia,  
California Institute of Technology,  
United States  
Paul J. Baker,  
National Institute of Allergy and Infectious  
Diseases (NIH), United States

## \*CORRESPONDENCE

Zhong Wang  
✉ wangzhong523@vip.163.com

## SPECIALTY SECTION

This article was submitted to  
Inflammation,  
a section of the journal  
Frontiers in Immunology

RECEIVED 22 November 2022

ACCEPTED 17 January 2023

PUBLISHED 02 February 2023

## CITATION

Wang X, Guo Z, Wang Z, Liao H, Wang Z,  
Chen F and Wang Z (2023) Diagnostic  
and predictive values of pyroptosis-  
related genes in sepsis.  
*Front. Immunol.* 14:1105399.  
doi: 10.3389/fimmu.2023.1105399

## COPYRIGHT

© 2023 Wang, Guo, Wang, Liao, Wang, Chen  
and Wang. 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.

# Diagnostic and predictive values of pyroptosis-related genes in sepsis

Xuesong Wang <sup>1</sup>, Zhe Guo<sup>1</sup>, Ziyi Wang<sup>1</sup>, Haiyan Liao<sup>1</sup>,  
Ziwen Wang<sup>1</sup>, Feng Chen<sup>1</sup> and Zhong Wang<sup>2\*</sup>

<sup>1</sup>School of Clinical Medicine, Tsinghua University, Beijing, China, <sup>2</sup>Department of General Medicine, Beijing Tsinghua Changgung Hospital affiliated to Tsinghua University, Beijing, China

**Background:** Sepsis is an organ dysfunction syndrome caused by the body's dysregulated response to infection. Yet, due to the heterogeneity of this disease process, the diagnosis and definition of sepsis is a critical issue in clinical work. Existing methods for early diagnosis of sepsis have low specificity.

**Aims:** This study evaluated the diagnostic and predictive values of pyroptosis-related genes in normal and sepsis patients and their role in the immune microenvironment using multiple bioinformatics analyses and machine-learning methods.

**Methods:** Pediatric sepsis microarray datasets were screened from the GEO database and the differentially expressed genes (DEGs) associated with pyroptosis were analyzed. DEGs were then subjected to multiple bioinformatics analyses. The differential immune landscape between sepsis and healthy controls was explored by screening diagnostic genes using various machine-learning models. Also, the diagnostic value of these diagnosis-related genes in sepsis (miRNAs that have regulatory relationships with genes and related drugs that have regulatory relationships) were analyzed in the internal test set and external test.

**Results:** Eight genes (CLEC5A, MALT1, NAIP, NLRC4, SERPINB1, SIRT1, STAT3, and TLR2) related to sepsis diagnosis were screened by multiple machine learning algorithms. The CIBERSORT algorithm confirmed that these genes were significantly correlated with the infiltration abundance of some immune cells and immune checkpoint sites (all  $P < 0.05$ ). SIRT1, STAT3, and TLR2 were identified by the DGldb database as potentially regulated by multiple drugs. Finally, 7 genes were verified to have significantly different expressions between the sepsis group and the control group ( $P < 0.05$ ).

**Conclusion:** The pyroptosis-related genes identified and verified in this study may provide a useful reference for the prediction and assessment of sepsis.

## KEYWORDS

sepsis, pyroptosis, machine learning (ML), immune landscape, MALT1 gene

## 1 Introduction

Sepsis is the systemic response to severe infection. It is the main cause of death in critically ill patients, and its incidence keeps rising (1). Sepsis can be caused by an infection that originates at multiple sites and most patients with sepsis are infected by pathogenic microorganisms. Patients with severe underlying diseases (such as diabetes, leukemia, etc.) are more likely to suffer from this condition. Sepsis is the leading cause of death in hospitalized patients (2), and previous studies have shown that 30% of inpatient mortality is caused by sepsis (3). Yet, due to the heterogeneity of the development of this disease, the diagnosis and definition of sepsis are critical issues in clinical work. In early 2016, the new definitions of sepsis changed dramatically; briefly, greater emphasis was placed on the diagnostic value of “organ dysfunction”, described as an acute increase in total Sequential Organ Failure Assessment (SOFA) score. According to the new sepsis 3.0 diagnostic criteria, sepsis is defined as “life-threatening organ dysfunction caused by the imbalance of the body’s response to infection” (4). Therefore, the diagnosis of sepsis is delayed compared with previous systemic inflammatory response syndrome (SIRS) scores. However, screening for patients at risk of developing sepsis may lead to more effective interventions that can greatly reduce the incidence of sepsis, mortality, and medical costs of infected patients (5). Therefore, early diagnosis and prevention of sepsis may be one of the main means to prevent and control sepsis in the future.

Pyroptosis is a new type of programmed cell death with lytic and pro-inflammatory characteristics. It is mediated by the cysteine aspartate-specific protein kinases 1, 4, 5, and 11, and its final step is dependent on the activity of the gasdermin family proteins that form pores in the cell membrane (6, 7). Pyroptosis is characterized by cell swelling, changes in plasma membrane permeability, and inflammatory release (8). Inflammasomes have a crucial role in the occurrence and development of sepsis (9). Studies have shown that inflammation-damaged and infected cells in sepsis patients can release pathogens through pyroptosis, which is then promptly eliminated by various immune cells (10). Therefore, pyroptosis can help to fight sepsis-induced infection to a certain extent. Some pyroptosis-related genes have been confirmed to be closely related to sepsis. For example, *GSDMD* has a vital role in sepsis by initiating pyroptosis and activating immune response (11, 12). However, the role of many pyroptosis-related genes in sepsis remains unclear.

In this study, we evaluated the diagnostic and predictive values of pyroptosis-related genes in normal and sepsis tissue and their role in the immune microenvironment using machine-learning methods.

**Abbreviations:** BP, Biological process; CC, Cellular component; DEGs, Differentially expressed genes; DGIdb, Drug-Gene Interaction Database; DNN, Deep Neural Network; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, Molecular Function; NLRP3, NLR family pyrin domain containing 3; PCC, Pearson correlation coefficient; PD-1, Programmed cell death protein - 1; PD-L, Programmed death Ligand; RF, Random Forest; RT-PCR, Quantitative Real-time polymerase chain reaction; TNF- $\alpha$ , Tumor necrosis factor- $\alpha$ .

## 2 Methods

### 2.1 Differential expression analysis of genes

The Limma package of R language was used to analyze the differential expression of whole genes (13). The filtering conditions were set as a  $P$ -value < 0.05 and an absolute value of  $\log_{2}FC > 0.2$ .

### 2.2 Construction of protein-protein interaction network

A PPI network was constructed to explore the interaction among DEGs and the intersection genes of pyroptosis-related genes. Specifically, the PPI network was obtained using protein interaction relationships in the STRING database (14) and the highest confidence interaction score was set to 0.4. Then, the PPI visualization results were obtained using Cytoscape software.

### 2.3 Gene ontology and functional enrichment analysis

Gene Ontology (GO) enrichment analysis (<http://www.geneontology.org>) (15) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis ([www.genome.jp/kegg/](http://www.genome.jp/kegg/)) (16) were used to determine the gene biological functions. The GO term is composed of three parts, i.e., biological process (BP), cellular component (CC), and molecular function (MF). Ultimately, significant pathways with  $P < 0.05$  were selected.

### 2.4 Construction of diagnostic model construction

The machine learning methods used for diagnostic model construction and the relevant package in Scikit-learn (17) were used. These packages’ specific parameter setting ranges are presented in the algorithms.

#### 2.4.1 Random Forest algorithm

The RF algorithm was applied to calculate the importance of genes (18). For each decision tree in the RF, the corresponding out-of-pocket (OOB) data was used to calculate its out-of-pocket data error. The error value could be denoted as  $error_{OOB1}$ . Next, by randomly adding noise to the out-of-pocket data  $X$ , the out-of-pocket data error  $error_{OOB2}$  could be calculated again. Finally, the calculation formula of the importance of out-of-pocket data  $X$  ( $feature(X)$ ) was obtained:

$$feature(X) = \sum (err_{OOB2} - err_{OOB1}) / N_{tree} \quad (1)$$

where  $N_{tree}$  represents the number of decision trees. When ranking feature importance, the specific parameters of RF were set as follows: the number of decision trees was selected between 100 and 1000, and the default values were used for other parameters.

## 2.4.2 Adaboost algorithm

The Adaboost algorithm was used to construct the diagnostic model. The AdaBoost algorithm is a boosting algorithm (19). In model training, each sample was given a probability (weight) that indicates that it was selected into the training set in a specific classifier. If the sample was accurately classified, the probability of it being chosen for the next training set was reduced. Conversely, the probability of being selected for the next training set was considerably improved.

In a binary classification training set  $(X=\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\})$ ,  $x_i$  and  $y_i$  represent the sample points and their corresponding binary classification results  $y_i$ . The algorithm first initializes the weights of the training set, and the weight redistribution value of the MTH weak learner is  $D_m=(w_{m1}, \dots, w_{mi}, \dots, w_{mN})$ . Here,  $w_{mi} = \frac{1}{N}$ ,  $i = 1, 2, \dots, N$ . Next, the weight redistribution values of the training set are updated for the  $M$  learners. Let  $G_m(x)$  represent the base classifier of the training set with weight distribution  $D_m$ . Then the classification error rate  $e_m$  of  $G_m(x)$  on the training set can be calculated as follows:

$$e_m = \sum_{i=1}^N W_{mi} I(G_m(x_i) \neq y_i) \quad (2)$$

the value of  $I(G_m(x_i) \neq y_i)$  is 0 or 1, representing correct and incorrect classification, respectively. Next, the weight coefficient  $\alpha_m$  of the classifier was calculated.

$$\alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \quad (3)$$

Finally, the final classifier  $G(x)$  was obtained by updating the weight redistribution value of the training set.

$$D_{m+1} = (w_{m+11}, w_{m+12}, \dots, w_{m+1N})$$

$$w_{m+1i} = \frac{w_{mi}}{Z_m} \exp(-\alpha_m y_i G_m(x_i)), i = 1, 2, \dots, N$$

$$G(x) = \text{sign}(\sum_{m=1}^M \alpha_m G_m(x)) \quad (4)$$

The specific parameters of Adaboost were set as follows: the number of decision trees was selected between 100 and 300, the learning rate was determined from two values of 0.01 and 0.1, and the default values were used for other parameters.

## 2.4.3 Logistic Regression algorithm

LR is a probabilistic nonlinear regression model that can be used for binary classification tasks. Its prediction function is shown below.

$$y = \frac{1}{1 + e^{-(w^T x + b)}} \quad (5)$$

where  $w$  and  $b$  are the weight and bias vectors, respectively. In this paper, the specific parameters of LR were set as follows: the reciprocal of the regularization strength was selected between 0.1 and 3.1, and the default values were used for other parameters.

## 2.4.4 Deep Neural Network algorithm

DNN is an extension based on perceptron. It can be understood as a neural network with many hidden layers (20). The layers of the DNN are fully connected, and any neuron in layer  $i$  must be connected to any neuron in layer  $i+1$ . The forward propagation algorithm of DNN uses several weight matrices  $W$  and bias vectors  $b$  to perform a series of nonlinear

operations with the input value vector  $x$  where the output value  $a^l$  of layer  $l$  can be expressed as the following equation.

$$a^l = \sigma(W^l a^{l-1} + b^l) (l = 2, \dots, L) \quad (6)$$

where  $W^l$  and  $b^l$  represent the  $l$ -th layer's weight matrix and bias vector, respectively;  $L$  represents the total number of layers, and  $\sigma$  represents the activation function. Next, the DNN algorithm updates  $W$  and  $b$  by backpropagation. The loss function  $J(W, b, x, y)$  in the backpropagation process can be expressed as the following equation.

$$J(W, b, x, y) = \frac{1}{2} \|a^L - y\|_2^2 \quad (7)$$

where  $a^L$  and  $y$  represent the output of the output layer and the sample, respectively. The  $W$  and  $b$  of each layer can be obtained by gradient descent. In this paper, the specific parameters of DNN were set as follows: the number of hidden layers was 2, the number of neurons in the two hidden layers was 8 or 16, the regularization parameters were selected in 0.001 and 0.01, and the learning rate was selected in 'constant' and 'adaptive.' Default values were used for other parameters.

## 2.4.5 Performance evaluation of diagnostic model

The evaluation method of the diagnostic model included calculating the model's sensitivity, specificity, and f1 score, using the following formulas:

$$\text{Sensitivity}_i = \frac{TP_i}{TP_i + FN_i} \quad (8)$$

$$\text{Specificity}_i = \frac{TN_i}{TN_i + FP_i} \quad (9)$$

$$F1\text{score}_i = \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i} \quad (10)$$

where  $TP_i$  and  $TN_i$  are the true positive and true negative of class  $i$ , respectively, and  $FP_i$  and  $FN_i$  are the false positive and false negative of class  $i$ , respectively.

## 2.5 Immune infiltration analysis

To explore the relationship between diagnosis-related genes and immune cell-related expression, the CIBERSORT algorithm was used to calculate the proportion of different immune cell types in significant samples of GSE26440 dataset ( $P < 0.05$ ) (21). Furthermore, the infiltration abundance of 22 kinds of immune cells was obtained, and the correlation between the diagnosis-related genes and the content of 22 immune cells was calculated using the Spearman correlation coefficient. In addition, based on the dataset GSE26440, we used Pearson to calculate the correlation between 47 immune test sites and diagnosis related genes. Finally, we used the external dataset GSE12904 to verify the expression of diagnostic genes.

## 2.6 Construct diagnostic gene-miRNA network

To explore the regulatory relationship between diagnostic genes and miRNAs, miRTarBase (<http://mirtarbase.cuhk.edu.cn/php/index>).

php), miRDB (<http://mirdb.org/>), and TargetScan ([http://www.targetscan.org/vert\\_72/](http://www.targetscan.org/vert_72/)) databases were queried for miRNAs that have a regulatory relationship with diagnostic genes, and the intersection genes of the three databases were selected to construct a diagnostic gene-miRNA network.

## 2.7 Identifying potential drugs

To identify potential therapeutic drugs for sepsis, Drug-Gene Interaction Database (DGIdb) version 3.0.2 (<https://www.dgldb.org>) (22) was used to analyze the interaction between the diagnostic genes and related drugs. In addition, we searched the DGIdb to predict potential drugs or molecular compounds that interact with DEG.

## 2.8 qRT-PCR

The whole blood samples of six patients with sepsis and six healthy people were collected from Beijing Tsinghua Changgung Hospital (Beijing, China). Demographic information could be found in [Table S1](#) of Supplementary materials. The screening criteria for sepsis patients were based on sepsis 3.0 (4). Whole blood samples were obtained from patients, and the quantitative real-time polymerase chain reaction (RT-PCR) was performed (Primer list could be found in [Table S2](#) of Supplementary materials). None of the patients had a history of autoimmune disorders, neoplastic diseases, or oral immunosuppressants. This study was approved by the Ethics Committee (NCT05095324).

# 3 Results

## 3.1 Microarray data set acquisition and data processing

Microarray data sets related to sepsis were downloaded from the GEO database. GSE26440 was selected as the internal training and test set and GSE13904 as the external test set. GSE26440 contains 130 whole-blood RNA samples from 98 samples with septic shock and 32 healthy controls. GSE13904 includes 158 samples with sepsis and 18 healthy controls.

GPL570-55999 platform file was used to convert gene symbols for the two GEO datasets. When multiple probes were mapped to the same gene symbol, the average value of the probes was selected as the gene expression value. Furthermore, 3837 DEGs were selected for differential expression analysis on the GSE26440 dataset. Of these, 2302 were down-regulated ([Figure 1](#), blue dots) and 1535 were up-regulated ([Figure 1](#), red dots). Next, a heat map of the top 50 up-regulated genes and the top 50 down-regulated genes among the differentially expressed genes were drawn ([Figure 1](#)).

Also, 110 genes related to pyroptosis were collected from previous studies. There were 35 intersection genes of DEGs and

pyroptosis-related genes, which could be visualized by the Venn diagram in this paper ([Figure 1](#)). The PPI network of intersecting genes is shown in [Figure 1](#). Among them, the node color's depth and area represent the number of connections with the gene. [Figure 1](#) shows the Pearson correlation coefficient (PCC) network plot among intersecting genes. Most of the 35 intersecting genes have strong correlations.

In addition, to explore the pathway information involved in the intersection genes, we drew a bubble plot of significant pathways with the BP ([Figure 2](#)), CC ([Figure 2](#)), and MF ([Figure 2](#)) entries in the GO enrichment analysis and the KEGG enrichment analysis ([Figure 2](#)), respectively.

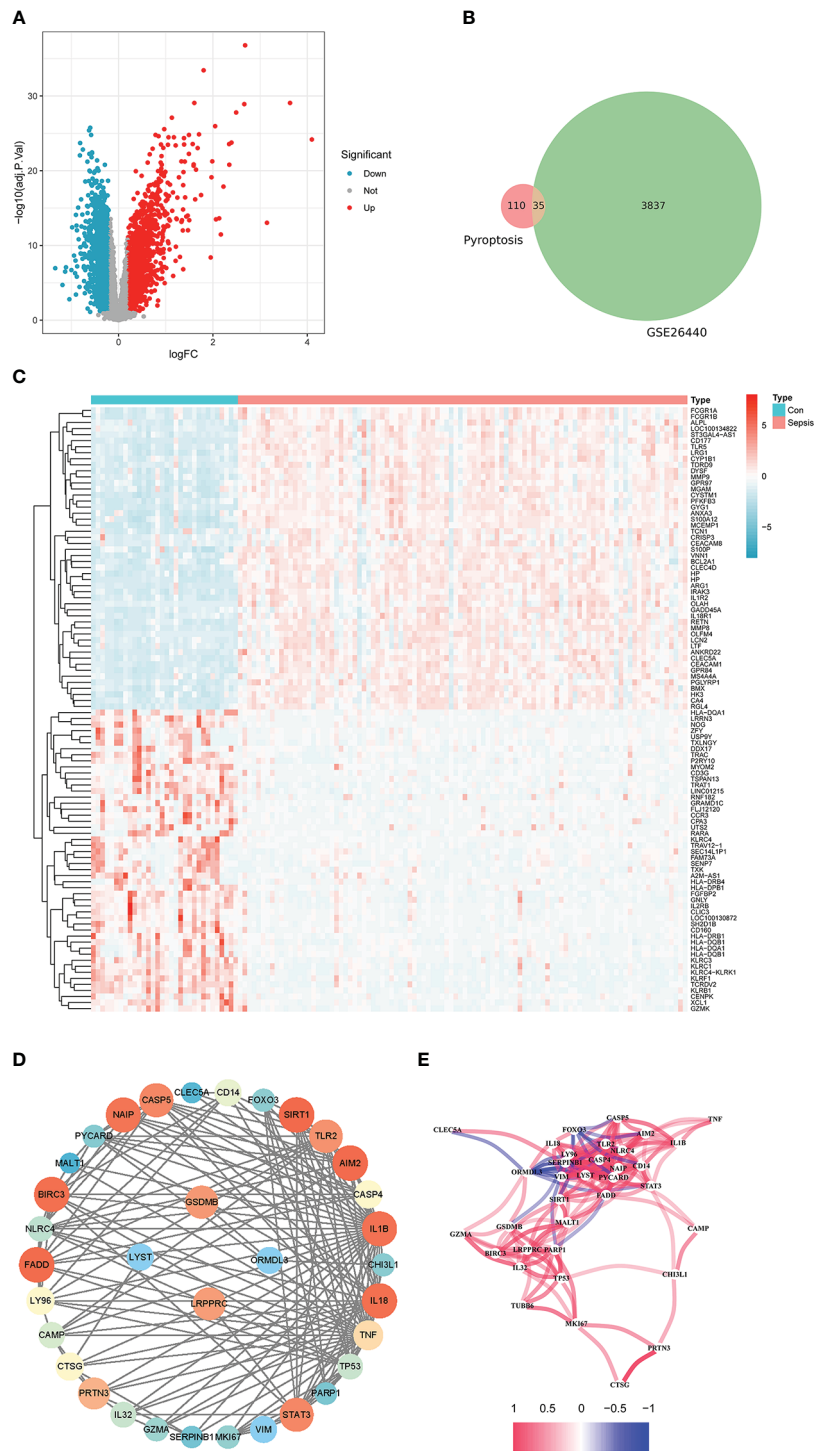
## 3.2 Gene screening and diagnostic model construction

To screen out the genes related to diagnosis from the 35 intersection genes, RF was used to rank the importance of the 35 features. The mean weight of 35 genes was 0.0285. We retained nine genes larger than the weight mean ([Figure 3A](#)). Next, we constructed sepsis diagnostic models using Adaboost, LR, and DNN classifiers to identify the pyroptosis-related genes related to diagnosis. Specifically, the top 1-top 9 genes on the training set were input to the three classifiers. After setting the corresponding classifier parameter range, the optimal parameters were selected using ten-fold cross-validation, and then the AUC results on the internal test set were obtained ([Figure 3B](#)). The AUC of the DNN classifier reached 0.9935 when using the top 9 genes. To further confirm the diagnostic power of the signature genes, the results were validated using the GSE13904 dataset ([Figure 3C](#)). The DNN classifier can reach an AUC of 0.9786 when using the top 8 genes. [Table 1](#) presents the model performance metrics when the AUC is maximized using the DNN model.

In addition, eight genes were further validated in the internal test set ([Figures 4A–H](#)) and external test set ([Figures 5A–H](#)) independently using the DNN classifier. [Table 2](#) and [Table 3](#) state the diagnostic model's performance on the internal and external test sets, respectively. LRPPRC was not considered because it presented false positives in diagnostic model construction.

## 3.3 Immune infiltration and the landscape of immune checkpoint genes in sepsis

To explore the immune infiltration of sepsis, the CIBERSORT algorithm was used to analyze the differences in peripheral blood immune cells between the significant sepsis samples and healthy control samples. The boxplots in [Figures 6A–H](#) showed the expression of eight diagnostically relevant genes in sepsis patients and healthy controls. Eight genes significantly differed between the two groups. [Figure 7A](#) shows the proportional expression of 22 immune cells in Control samples and sepsis samples. [Figure 7B](#) shows a heatmap of correlations between immune cells. Resting mast cells and resting dendritic cells had

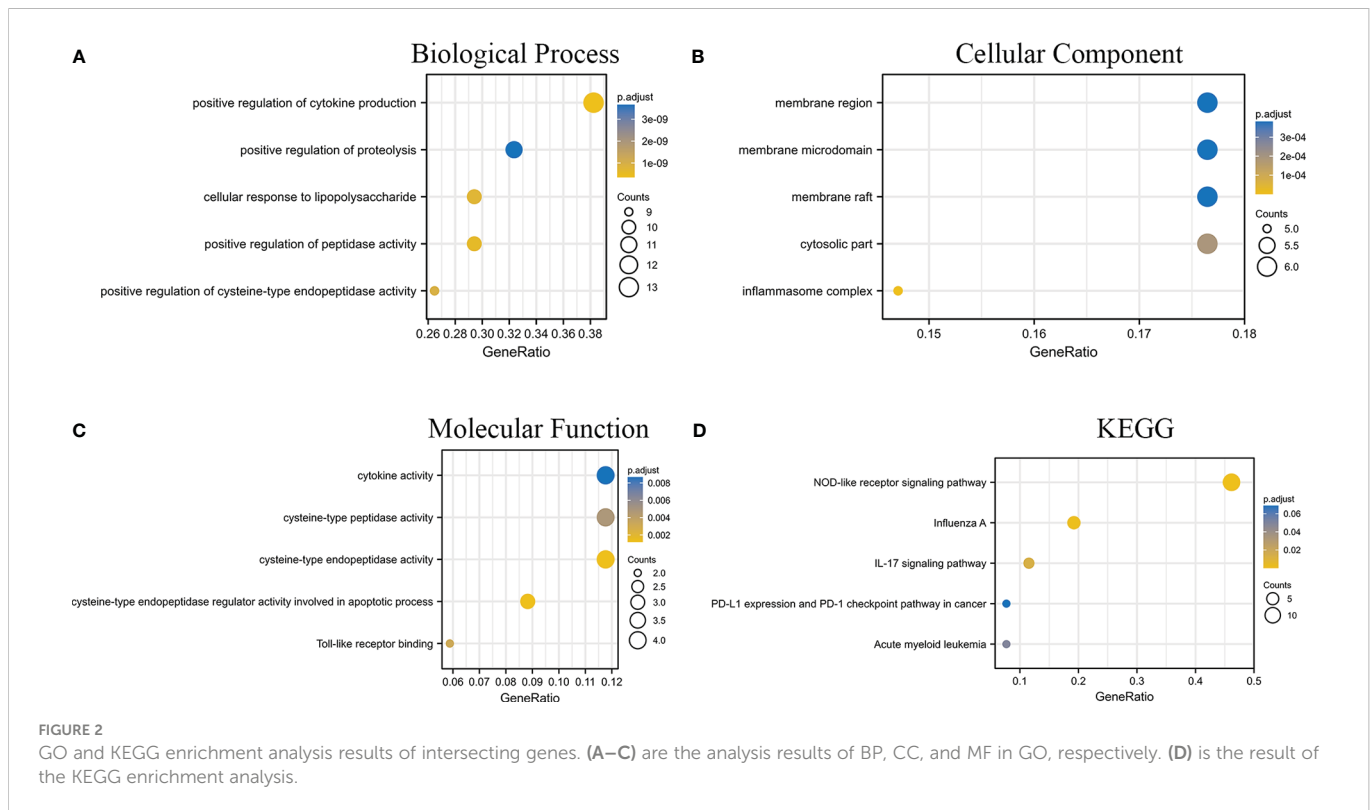


**FIGURE 1**  
 Differential analysis of the internal training set. **(A)** is the volcano plot from the difference analysis. **(B)** is a Venn diagram of pyroptosis-related genes plotted against the differentially expressed genes of the training set. **(C)** is the expression heatmap of the top 50 up-regulated and down-regulated genes by absolute logFC. **(D)** is the PPI network constructed by the intersection genes of DEGs and pyroptosis-related genes. **(E)** is a correlation network plot among intersecting genes.

the strongest positive correlation ( $r=0.94$ ). Conversely, Macrophages M0 and T cells CD8 had the strongest negative correlation ( $r=0.5$ ). **Figure 7C** shows the differential expression of different immune cell marker expression types between sepsis and controls. Among them, the expression levels of B cells naive, B cells memory, T cells CD8, activated T CD4 memory cells, T cells

follicular helper, regulatory T cells, T cells gamma delta, activated NK cells, macrophages M0, macrophages M1, resting dendritic cells, resting and cells activated mast cells, and neutrophils were significantly different between the two groups (all  $P<0.05$ ). In addition, we showed the lollipop plots of eight sepsis diagnosis-related genes (CLEC5A, MALT1, NAIP, NLR4, SERPINB1,





SIRT1, STAT3, and TLR2) associated with 22 immune cells in **Figures 7D–K**.

Next, we analyzed the association of diagnostic genes with 47 immune checkpoints. Specifically, the immune checkpoints were retained with significant associations ( $P < 0.001$ ) with eight diagnosis-related genes and a heatmap of the associations between them was created (**Figures 8A–H**). For CLEC5A, the immune checkpoints with the strongest positive/negative correlation were CD44 ( $r = 0.39$ ) and CD244 ( $r = -0.34$ ), respectively. For MALT1, the immune checkpoints with the strongest positive/negative correlation were BTLA ( $r = 0.55$ ) and BTNL2 ( $r = -0.55$ ), respectively. For NAIP, the immune checkpoints with the strongest positive/negative correlation were CD44 ( $r = 0.72$ ) and TNFRSF25 ( $r = -0.49$ ), respectively. For NLRC4, the immune checkpoints with the strongest positive/negative correlation were LAIR1 ( $r = 0.75$ ) and TNFRSF25 ( $r = -0.45$ ), respectively. For SERPINB1, the immune checkpoints with the strongest positive/negative correlation were LAIR1 ( $r = 0.75$ ) and TNFRSF25 ( $r = -0.51$ ), respectively. For SIRT1, the immune checkpoints with the strongest positive/negative correlation were BTLA ( $r = 0.47$ ) and ICOSLG ( $r = -0.68$ ), respectively. For STAT3, the immune checkpoints with the strongest positive/negative correlation were C10orf54 ( $r = 0.66$ ) and CD276 ( $r = -0.49$ ), respectively. For TLR2, the immune checkpoints with the strongest positive/negative correlation were C10orf54 ( $r = 0.71$ ) and ICOSLG/TNFRSF25 ( $r = -0.49$ ), respectively.

To determine the regulatory relationship among eight diagnosis-related genes and miRNAs, three miRNA databases were used to query the regulatory relationship among diagnosis-related genes. We took the intersection genes of the three database

results and mapped the regulatory network of diagnosis-related genes and miRNAs (**Figure S1A**). More miRNAs regulated STAT3, NAIP, and SIRT1. Furthermore, we applied DGIdb to identify potential drugs or molecular compounds that could reverse the expression of diagnostically relevant genes (**Figure S1B**). Various drugs might regulate SIRT1, STAT3, and TLR2. The inhibitory effects of some compounds on sepsis were demonstrated. Rare ginsenosides Rk1 and Rg5 have an inhibitory effect on high mobility group box 1 (HMGB1)-mediated sepsis (23). HMGB1 is a cytokine present in the late stage of sepsis. The up-regulation of SIRT1 by ethyl pyruvate (EP) may promote the deacetylation of HMGB1, thereby reducing the release of HMGB1 from lipopolysaccharide (LPS)-activated macrophages (24). Lee et al. demonstrated the inhibitory effect of ginsenoside Rh1 on HMGB1-mediated sepsis response (25). In addition, celecoxib combined with low-dose antibiotics in treating sepsis in mice can significantly reduce the bacterial load and inflammatory markers in different organs of mice (26).

### 3.4 Expression validation of diagnosis-related genes

qRT-PCR was used to verify the expression levels of eight diagnosis-related genes between the sepsis group and the control groups, Demographic information could be found in **Table S1** of **Supplementary materials**. The results showed that, except for STAT3, the expression of other genes was significantly different between the two groups (**Figure 9**), which was consistent with the results of bioinformatics analysis. Therefore, further exploration of

their roles in sepsis is of great significance for diagnosing and treating sepsis.

## 4 Discussion

Sepsis is a syndrome that often occurs due to various infections. Its incidence is increasing year by year. Therefore, early diagnosis of

sepsis is necessary. In addition, the role of pyroptosis-related genes in the occurrence and development of sepsis is still unclear. Therefore, this paper aims to explore the genes related to the diagnosis of sepsis and pyroptosis.

We performed enrichment analyses after the intersection of DEGs and pyroptosis-related genes and obtained multiple immune-related pathways. The GO and KEGG enrichment analysis showed that most of these genes were enriched in immune and tumor progression-

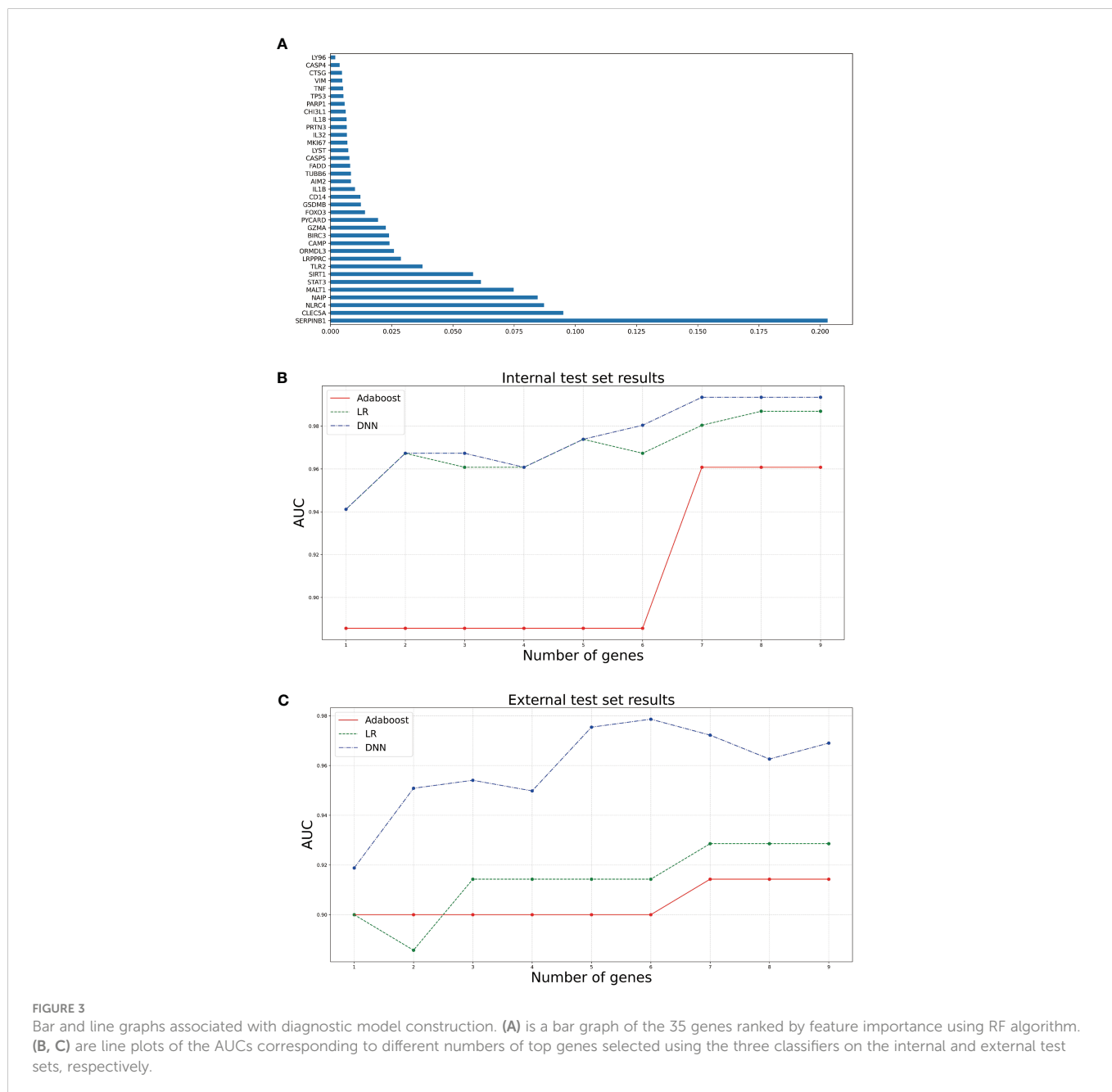


TABLE 1 AUC, sensitivity, specificity, and f1 score of the diagnostic model consisting of multiple genes on the internal and external test sets.

Test set type	Number of genes	AUC	Sensitivity	Specificity	f1 score
Internal test set	9	0.9935	1	0.8889	0.9714
External test set	6	0.9786	0.9423	1	0.9703

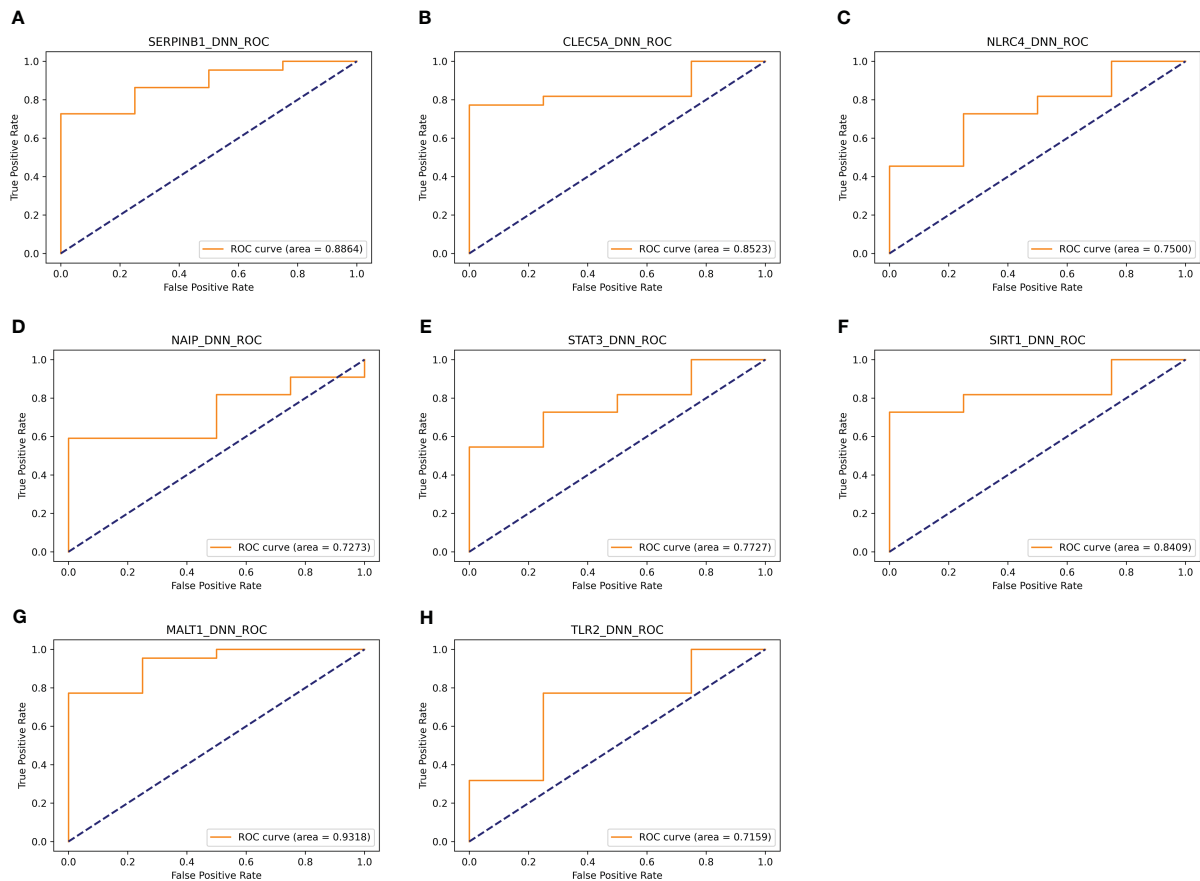


FIGURE 4 Results of separate validation of the top eight genes on the internal validation set. (A–H) are ROC curves for the top 8 genes, respectively.

related pathways. Danielski et al. systematically reviewed the NLRP3 inflammasome and its role in developing sepsis, and incorrect regulation of NF- $\kappa$ B has been associated with inflammation and autoimmunity (27). Zhou et al. found that NF- $\kappa$ B activation is involved in cardiomyocyte apoptosis, cardiomyocyte autophagy, and inflammatory cytokine release in sepsis patients (28). Toll-like receptors (TLRS) are an important class of protein molecules involved in non-specific immunity. Chen et al. systematically reviewed the complexity of using Toll-like receptors to intervene in sepsis (29).

Programmed cell death protein 1 (PD-1)/programmed death ligand-1 (PD-L1) immune checkpoint blockade has successfully treated cancer. Because the immune paralysis of sepsis involves depleted T lymphocytes, Nakamori et al. suggested that the role of PD-1/PD-L1 on innate lymphocyte function and the exosome form of PD-L1 deserves further investigation (30). In addition, pathways associated with macrophage apoptosis were identified. Tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) was a key regulator of innate immunity related to sepsis-induced acute lung injury. Yang et al. confirmed that polymorphonuclear neutrophils stimulated by TNF- $\alpha$  up-regulates the expression of NLR family pyrin domain

containing 3 (NLRP3) inflammasome through the nuclear factor- $\kappa$ B signaling pathway, which triggers macrophages for pyroptosis (31).

CLEC5A, MALT1, NAIP, NLRC4, SERPINB1, SIRT1, STAT3, and TLR2 were confirmed as sepsis biomarkers. Yang et al. performed bioinformatics analysis and identified SERPINB1 as a potential biomarker of sepsis (the expression of *SERPINB1* was up-regulated in patients with sepsis) (32). Zhang et al. used a weighted co-expression network to analyze and establish 15 mRNAs, including NLRC4, as hub genes of the mRNA-lncRNA-pathway network related to sepsis, and the expression of these genes was up-regulated in sepsis patients (33) *Salmonella typhimurium* is a Gram-negative pathogen that causes diseases ranging from gastroenteritis to systemic infections and sepsis. Related experimental analysis by Naseer et al. confirmed the response of caspase-4-mediated inflammasome of NAIP to *Salmonella* pathogenic island 1-expressing *Salmonella* (34). Their study elaborated that high expression of NAIP is sufficient to activate the inflammasome in response to infection under physiological conditions. Shayantan et al. used a machine learning approach to identify 20 differentially expressed genes using gene expression



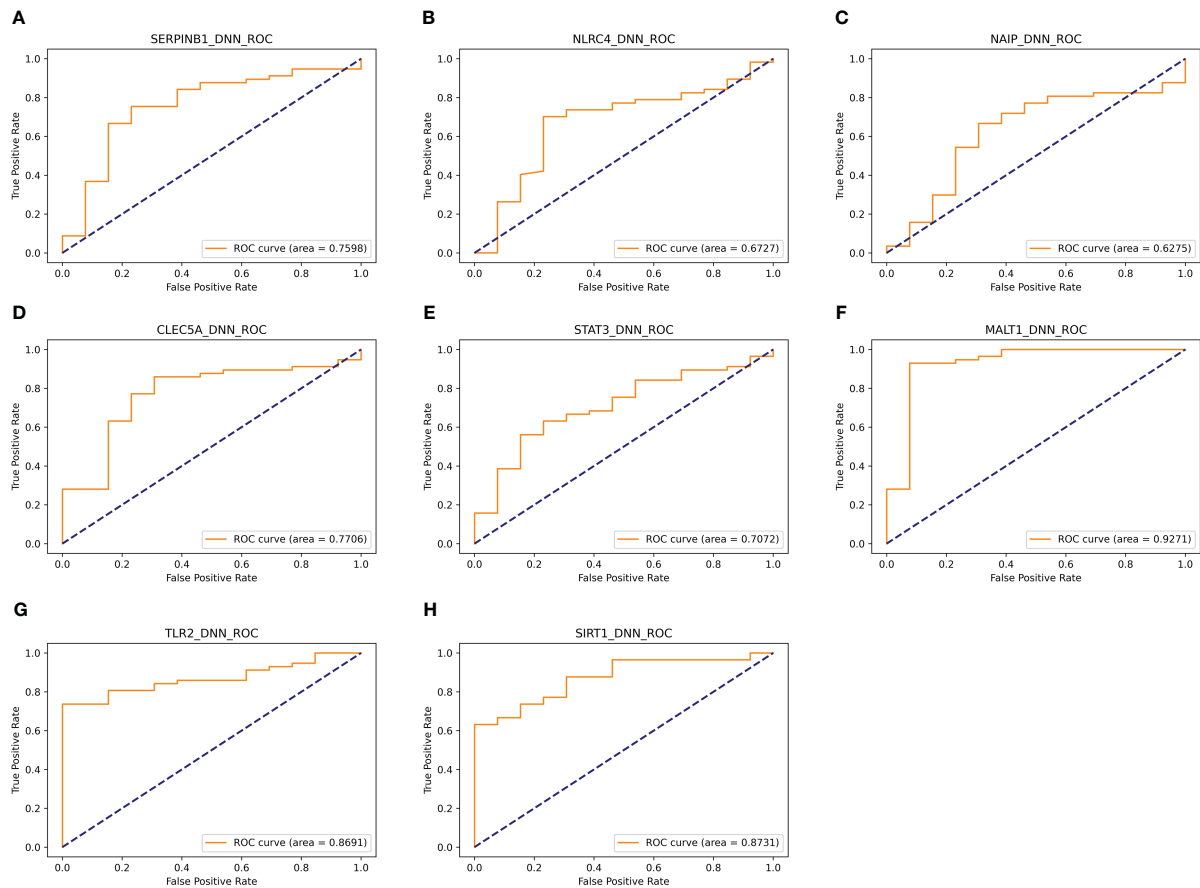


FIGURE 5 Results of separate validation of the top eight genes on the external validation set. (A–H) are ROC curves for the top 8 genes, respectively.

profiles of peripheral blood cells obtained within 24 hours of admission to the pediatric ICU (pICU) and extensive clinical data from 228 septic patients from pICU. Among them, eight genes, including CLEC5A, were confirmed to be closely related to sepsis mortality (35). Activation of CLEC5A by the dengue virus and Japanese encephalitis virus leads to the secretion of pro-inflammatory cytokines (36). In addition, many studies have shown that targeting STAT3 has a broad application prospect in treating sepsis (37). STAT3 was further up-regulated during the

transition from healthy to sepsis induction. MALT1 deficiency is a congenital immunodeficiency characterized by recurrent bacterial, viral, and fungal infections. Wang et al. demonstrated a significantly high expression of MALT1 in septic samples, and its high expression was closely related to various organ dysfunction (38). In addition, some studies have confirmed that NEAT1 alleviates sepsis-induced myocardial injury by regulating TLR2/NF- $\kappa$ B signaling pathway (39). NEAT1 can significantly down-regulate TLR2 to improve myocardial injury induced by

TABLE 2 AUC, sensitivity, specificity, and f1 score of the diagnostic model composed of a single gene on the internal test set.

Gene symbol	AUC	Sensitivity	Specificity	f1 score
SERPINB1	0.8664	0.9412	0.3333	0.8205
NLRC4	0.75	0.8889	0.25	0.8
NAIP	0.7273	0.8889	0.25	0.8
CLEC5A	0.8523	0.9444	0.375	0.85
STAT3	0.7727	0.8636	0.25	0.8636
MALT1	0.9318	0.9524	0.6	0.9302
TLR2	0.7159	0.85	0.1667	0.8095
SIRT1	0.8409	0.8571	0.2	0.8372

TABLE 3 AUC, sensitivity, specificity, and f1 score of the diagnostic model composed of a single gene on the external test set.

Gene symbol	AUC	Sensitivity	Specificity	f1 score
SERPINB1	0.7598	0.902	0.4211	0.8519
NLRC4	0.6727	0.8333	0.25	0.8108
NAIP	0.6275	0.8364	0.2667	0.8214
CLEC5A	0.7706	0.9245	0.5294	0.8909
STAT3	0.7072	0.8421	0.3077	0.8421
MALT1	0.9271	0.9808	0.6667	0.9358
TLR2	0.8691	0.8621	0.4167	0.8696
SIRT1	0.8731	0.873	0.7143	0.9167

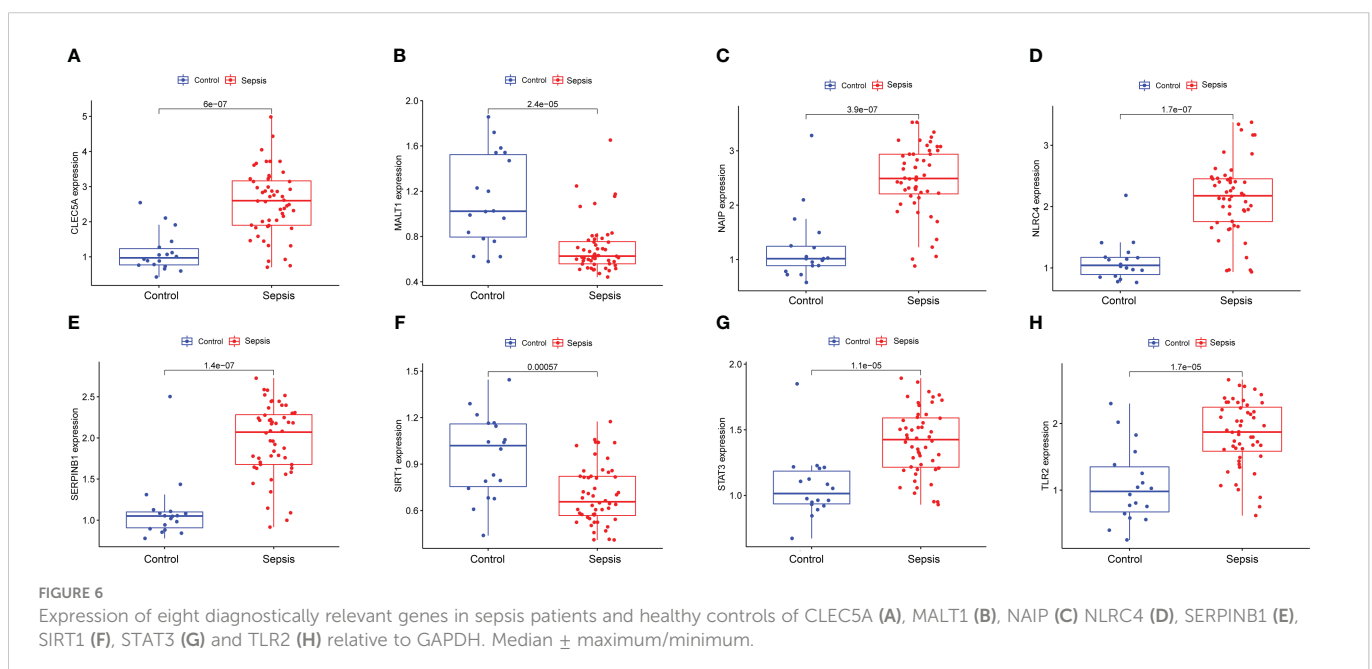
LPS in mice. Deng et al. demonstrated that SIRT1 could alleviate sepsis-induced acute kidney injury through autophagy activation mediated by Beclin1 deacetylation (40). They found that low SIRT1 expression promotes the development of acute kidney injury. In addition, PCR results showed significant differences in the expression of the other 7 genes between the sepsis samples and the normal samples ( $P < 0.05$ ).

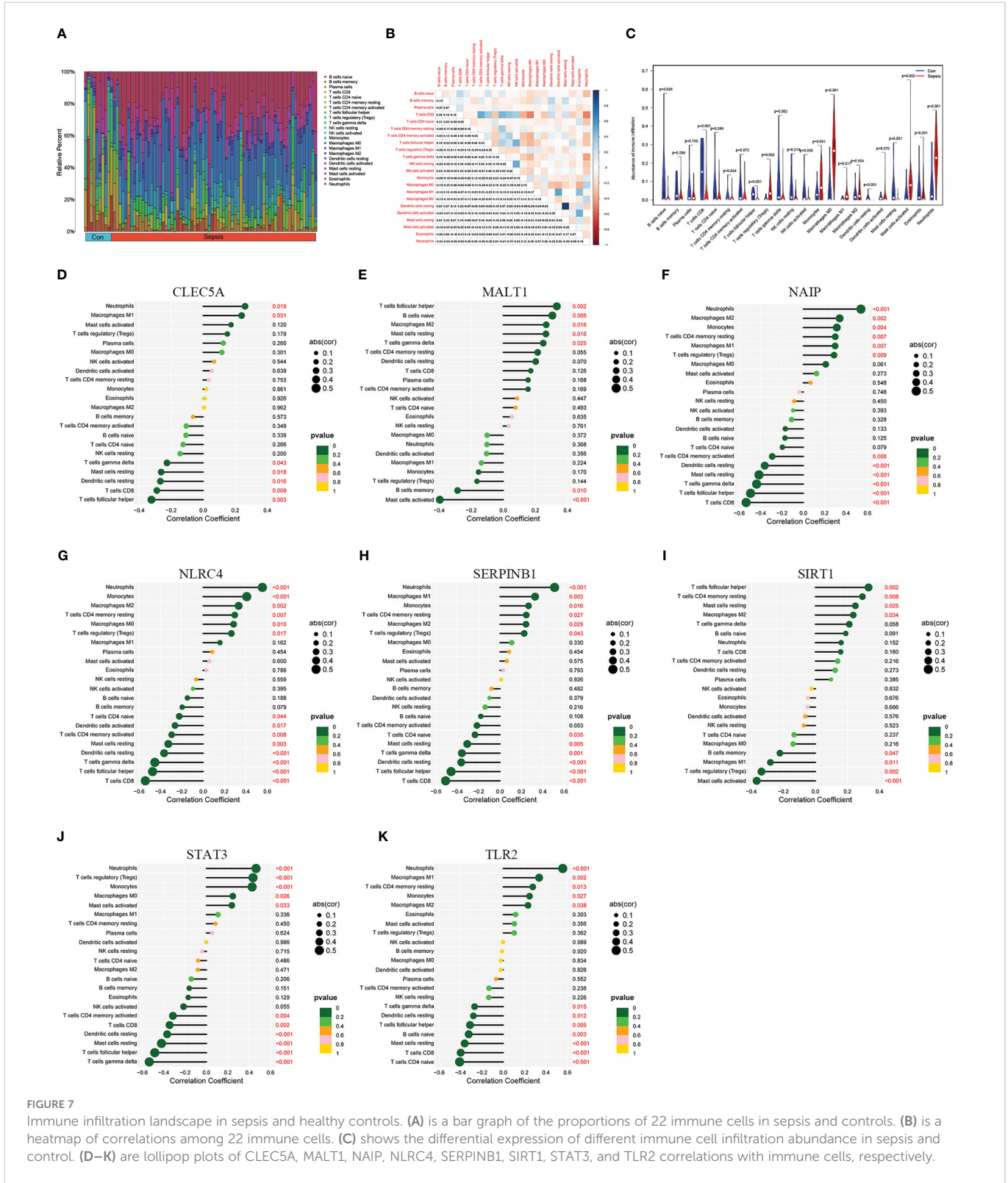
To explore the relationship between eight diagnostically relevant genes and the immune microenvironment and immune checkpoints, we applied CIBERSORT to estimate the abundance of 22 immune cells. We found that some immune cells were significantly different between the two groups. Xiao et al. elucidated the role of regulatory B cells in the pathogenesis of sepsis (41). Chaturvedi et al. showed that CD8+ T cells have a substantial predictive value for hemophagocytic lymphohistiocytosis versus early sepsis or healthy controls (42). Furthermore, another study suggested that post-sepsis immunosuppression depends on the regulation of mTOR/IFN- $\gamma$  in NK cells by NKT cells (43). M1 macrophages increase endothelial permeability and enhance p38 phosphorylation in sepsis through

PPAR- $\gamma$ /CXCL13-CXCR5 (44). Mast cells exacerbate sepsis by inhibiting phagocytosis of peritoneal macrophages (45). Neutrophils can interfere with pulmonary microcirculation in sepsis-induced acute lung injury (46).

Our analysis also showed that multiple immune checkpoints were significantly expressed in sepsis. A review by Richendrfel et al. suggested that proteoglycan 4 may enter cells through the CD44 receptor, act on the receptor, and prevent downstream processes of inflammation (47). Wang et al. showed that up-regulation of BTLA expression in bone marrow dendritic cells was associated with treatment outcomes in neonatal sepsis (48).

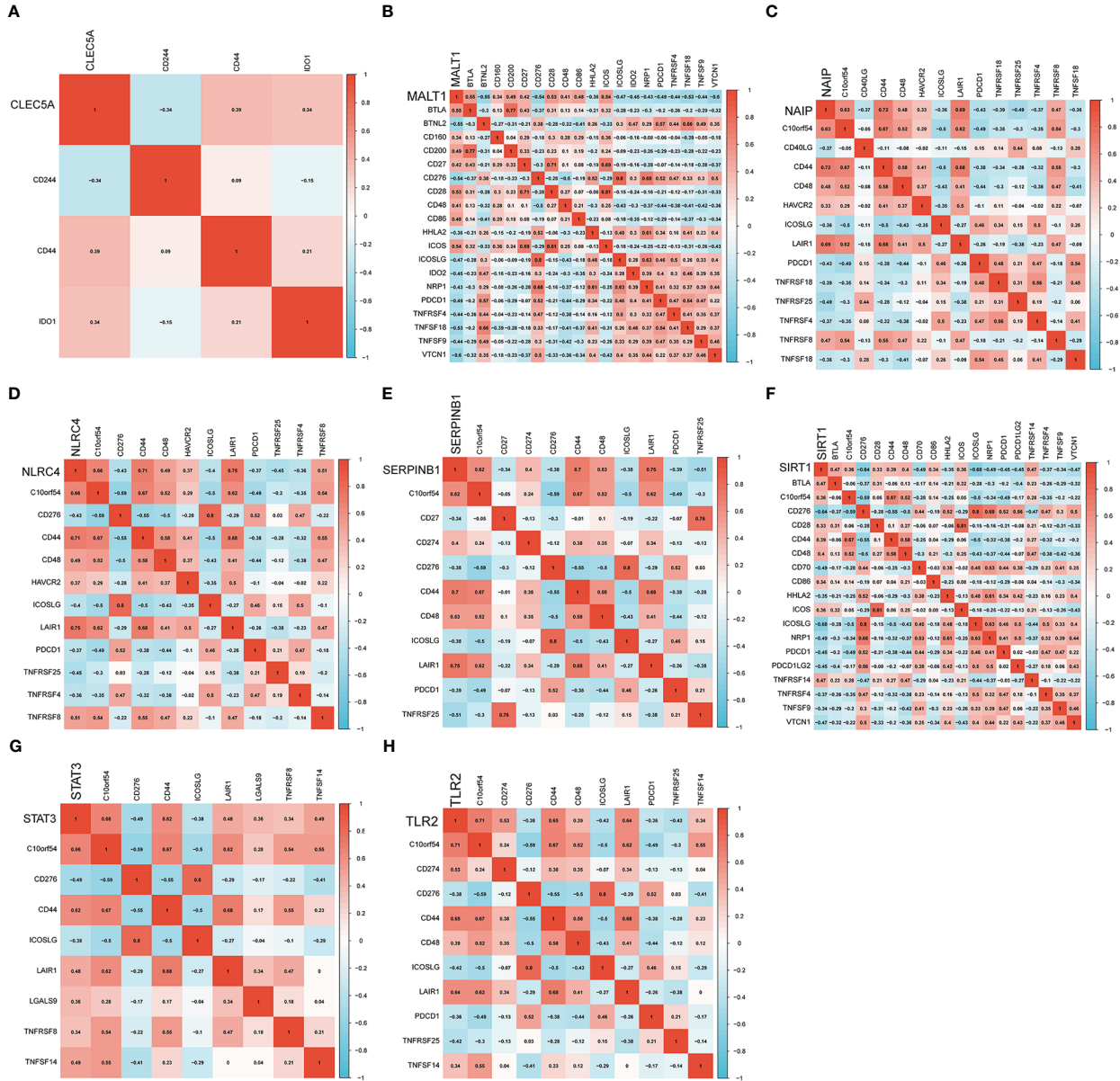
The present study has some similarities with Wang's study (49), namely, mining the role of pyroptosis-related genes in sepsis patients and constructing a diagnostic model. In addition, both these studies discussed the immune landscape of hub genes. However, the two papers differed in the following ways. Firstly, we aimed to identify genes associated with sepsis diagnosis using machine learning approaches. To this end, various machine





learning methods were used for gene importance assessment and classifier design. However, that paper focused on constructing nomogram models to evaluate the predictive value of clinical factors and genes for sepsis risk. Secondly, the regulatory

relationship of the mined genes was queried, and the regulatory network of diagnosis-related genes and miRNAs was constructed. Furthermore, DGIdb was applied to identify potential drugs or molecular compounds that could reverse the expression of



**FIGURE 8** Heatmap of the association of eight genes with immune checkpoints. (A–H) are correlation heatmaps with significant associations between diagnosis-related genes (CLEC5A, MALT1, NAIP, NLR4, SERPINB1, SIRT1, STAT3 and TLR2) and immune checkpoints.

diagnosis-related genes. Finally, experimental verification of the mined diagnosis-related genes was carried out.

This study has some limitations. First, the study has a small sample size. Therefore, we started collecting samples from patients with sepsis to investigate these findings further. Secondly, the specificity of the selected pyroptosis-related genes in diagnosing sepsis needs further experimental verification and discussion, especially comparing other inflammatory diseases and sepsis. Finally, given the complexity of the pathological process of sepsis, and to further reduce the application scenario and scope of this experimental results, the following experiments will further verify the different stages of sepsis, e.g., taking suspected sepsis (5), sepsis, and septic shock as separate experimental groups, which

can better divide the pathological process of sepsis, which we plan to perform in our next study.

### 5 Conclusion

In this paper, the potential value of diagnosis-related genes in the early diagnosis of sepsis was discussed using machine-learning methods based on the genes related to pyroptosis. In addition, various bioinformatics analysis methods were used to identify the differences between these genes in immune cells and checkpoints. However, different subtypes of sepsis may significantly differ in prognosis and survival. In

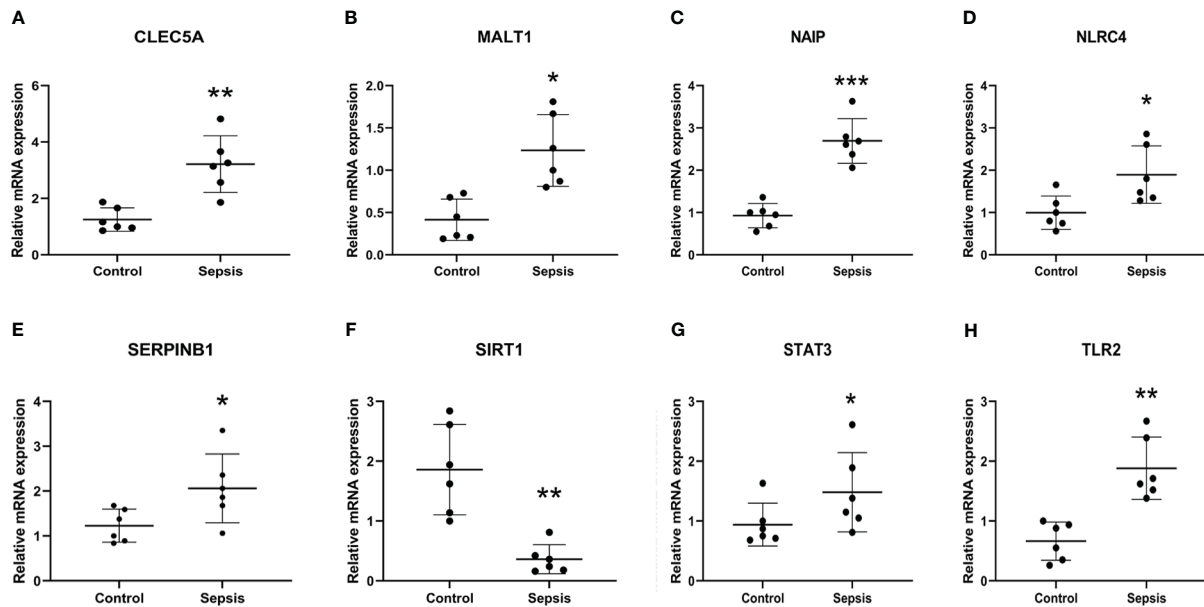


FIGURE 9

Transcriptional expression in whole blood from sepsis patients and healthy controls of CLEC5A (A), MALT1 (B), NAIP (C), NLRC4 (D), SERPINB1 (E), SIRT1 (F), STAT3 (G) and TLR2 (H) relative to GAPDH. Mean  $\pm$  standard deviation, \*\*\* $P$ <0.001, \*\* $P$ <0.01, \* $P$ <0.05 vs. the Control group.

future research, we plan to further analyze the differences between different subtypes of sepsis patients through various clustering algorithms to provide new insights for precision medicine.

## 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 studies involving human participants were reviewed and approved by Beijing Tsinghua Changgung Hospital. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

XW performed the major work and drafted the manuscript. ZG, ZiyW, HL, FC and ZiwW participated in revising the manuscript. ZhoW conceptualized, supervised, and revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This article is supported by Beijing Tsinghua Changgung Hospital.

## Acknowledgments

Over the writing of this paper, I would like to express my thanks to XY Chen who gave me support and encouragement to persist accomplish this article.

## 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/fimmu.2023.1105399/full#supplementary-material>



## References

- Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. *Intensive Care Med* (2021) 47(11):1181–247. doi: 10.1007/s00134-021-06506-y
- Abe T, Yamakawa K, Ogura H, Kushimoto S, Saitoh D, Fujishima Sc, et al. Epidemiology of sepsis and septic shock in intensive care units between sepsis-2 and sepsis-3 populations: sepsis prognostication in intensive care unit and emergency room (SPICE-ICU). *J Intensive Care* (2020) 8:44. doi: 10.1186/s40560-020-00465-0
- Rhee C, Jones TM, Hamad Y, Pande A, Varon J, O'Brien C, et al. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. *JAMA Network Open* (2019) 2(2):e187571. doi: 10.1001/jamanetworkopen.2018.7571
- CSinger M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315(8):801–10. doi: 10.1001/jama.2016.0287
- Emergency Medicine Branch Of Chinese Medical Care International Exchange Promotion A, Emergency Medical Branch Of Chinese Medical A, Chinese Medical Doctor Association Emergency Medical B, Chinese People's Liberation Army Emergency Medicine Professional C. Consensus of Chinese experts on early prevention and blocking of sepsis. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* (2020) 32(5):518–30. doi: 10.3760/cma.j.cn121430-20200514-00414
- Tan Y, Chen Q, Li X, Zeng Z, Xiong W, Li G, et al. Pyroptosis: a new paradigm of cell death for fighting against cancer. *J Exp Clin Cancer Res* (2021) 40(1):153. doi: 10.1186/s13046-021-01959-x
- Zhang Z, Zhang Y, Xia S, Kong Q, Li S, Liu X, et al. Gasdermin e suppresses tumour growth by activating anti-tumour immunity. *Nature* (2020) 579(7799):415–20. doi: 10.1038/s41586-020-2071-9
- Bertheloot D, Latz E, Franklin BS. Necroptosis, pyroptosis and apoptosis: an intricate game of cell death. *Cell Mol Immunol* (2021) 18(5):1106–21. doi: 10.1038/s41423-020-00630-3
- Kaukonen KM, Bailey M, Pilcher D, Cooper DJ, Bellomo R. Systemic inflammatory response syndrome criteria in defining severe sepsis. *New Engl J Med* (2015) 372(17):1629–38. doi: 10.1056/NEJMoal415236
- Kim M-J, Bae SH, Ryu JC, Kwon Y, Oh JH, Kwon J, et al. SESN2/sestrin2 suppresses sepsis by inducing mitophagy and inhibiting NLRP3 activation in macrophages. *Autophagy* (2016) 12(8):1272–91. doi: 10.1080/15548627.2016.1183081
- Guo H, Callaway JB, Ting JP. Inflammasomes: mechanism of action, role in disease, and therapeutics. *Nat Med* (2015) 21(7):677–87. doi: 10.1038/nm.3893
- He WT, Wan H, Hu L, Chen P, Wang X, Huang Z, et al. Gasdermin d is an executor of pyroptosis and required for interleukin-1 $\beta$  secretion. *Cell Res* (2015) 25(12):1285–98. doi: 10.1038/cr.2015.139
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for Rna-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
- Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The String Database in 2017: Quality-Controlled Protein-Protein Association Networks, Made Broadly Accessible. *Nucleic Acids Res* (2017) 45(D1):D362–D8. doi: 10.1093/nar/gkw937
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The gene ontology (GO) database and informatics resource. *Nucleic Acids Res* (2004) 32(Database issue):D258–61. doi: 10.1093/nar/gkh036
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (1999) 27(1):29–34. doi: 10.1093/nar/27.1.29
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* (2011) 12:2823–30.
- Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics* (2012) 99(6):323–9. doi: 10.1016/j.ygeno.2012.04.003
- Wang HC, Zhang LM. A novel fast training algorithm for adaboost. *J Fudan Univ* (2004) 01:27–33. doi: 10.15943/j.cnki.fdxh-jns.2004.01.006
- Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw* (2015) 61:85–117. doi: 10.1016/j.neunet.2014.09.003
- Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* (2019) 37(7):773–82. doi: 10.1038/s41587-019-0114-2
- Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, et al. Dgidb 3.0: A redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* (2018) 46(D1):D1068–D73. doi: 10.1093/nar/gkx1143
- Kim JE, Lee W, Yang S, Cho SH, Baik MC, Song GY, et al. Suppressive effects of rare ginsenosides, rkl and rg5, on hmgb1-mediated septic responses. *Food Chem Toxicol* (2019) 124:45–53. doi: 10.1016/j.fct.2018.11.057
- Kim YM, Park EJ, Kim JH, Park SW, Kim HJ, Chang KC. Ethyl pyruvate inhibits the acetylation and release of hmgb1 via effects on sirt1/stat signaling in lps-activated raw2647 cells and peritoneal macrophages. *Int Immunopharmacol* (2016) 41:98–105. doi: 10.1016/j.intimp.2016.11.002
- Lee W, Cho SH, Kim JE, Lee C, Lee JH, Baik MC, et al. Suppressive effects of ginsenoside rh1 on hmgb1-mediated septic responses. *Am J Chin Med* (2019) 47(1):119–33. doi: 10.1142/S0192415X1950006X
- Annamanedi M, Varma GYN, Anuradha K, Kalle AM. Celecoxib enhances the efficacy of low-dose antibiotic treatment against polymicrobial sepsis in mice and clinical isolates of escape pathogens. *Front Microbiol* (2017) 8:805. doi: 10.3389/fmicb.2017.00805
- Danielski LG, Giustina AD, Bonfante S, Barichello T, Petronillo F. The Nlrp3 inflammasome and its role in sepsis development. *Inflammation* (2020) 43(1):24–31. doi: 10.1007/s10753-019-01124-9
- Zhou H, Wang GQ, Luo PF, Xia ZF. Role of nuclear factor-kappa activation in sepsis-induced myocardial dysfunction. *Zhonghua Shao Shang Za Zhi* (2017) 33(12):782–4. doi: 10.3760/cma.j.issn.1009-2587.2017.12.014
- Chen F, Zou L, Williams B, Chao W. Targeting toll-like receptors in sepsis: from bench to clinical trials. *Antioxid Redox Signal* (2021) 35(15):1324–39. doi: 10.1089/ars.2021.0005
- Nakamori Y, Park EJ, Shimaoka M. Immune deregulation in sepsis and septic shock: reversing immune paralysis by targeting pd-1/pd-l1 pathway. *Front Immunol* (2020) 11:624279. doi: 10.3389/fimmu.2020.624279
- Jiao Y, Zhang T, Zhang C, Ji H, Tong X, Xia R, et al. Exosomal mir-30d-5p of neutrophils induces m1 macrophage polarization and primes macrophage pyroptosis in sepsis-related acute lung injury. *Crit Care* (2021) 25(1):356. doi: 10.1186/s13054-021-03775-3
- Yang YX, Li L. Identification of potential biomarkers of sepsis using bioinformatics analysis. *Exp Ther Med* (2017) 13(5):1689–96. doi: 10.3892/etm.2017.4178
- Zhang X, Cui Y, Ding X, Liu S, Han B, Duan X, et al. Analysis of mrna-lncrna and mrna-lncrna-pathway co-expression networks based on wgcn in developing pediatric sepsis. *Bioengineered* (2021) 12(1):1457–70. doi: 10.1080/21655979.2021.1908029
- Naseer N, Zhang J, Bauer R, Constant DA, Nice TJ, Brodsky IE, et al. Salmonella enterica serovar typhimurium induces naip/nlr4- and nlrp3/asc-independent, caspase-4-dependent inflammasome activation in human intestinal epithelial cells. *Infect Immun* (2022) 90(7):e0066321. doi: 10.1128/iai.00663-21
- Banerjee S, Mohammed A, Wong HR, Palaniyar N, Kamaleswaran R. Machine learning identifies complicated sepsis course and subsequent mortality based on 20 genes in peripheral blood immune cells at 24 h post-icu admission. *Front Immunol* (2021) 12:592303. doi: 10.3389/fimmu.2021.592303
- Chen ST, Li FJ, Hsu TY, Liang SM, Yeh YC, Liao WY, et al. Clec5a is a critical receptor in innate immunity against listeria infection. *Nat Commun* (2017) 8(1):299. doi: 10.1038/s41467-017-00356-3
- Lei W, Liu D, Sun M, Lu C, Yang W, Wang C, et al. Targeting stat3: A crucial modulator of sepsis. *J Cell Physiol* (2021) 236(11):7814–31. doi: 10.1002/jcp.30394
- Wang Y, Huang Q, He F. Aberrant blood malt1 and its relevance with multiple organic dysfunctions, t helper cells, inflammation, and mortality risk of sepsis patients. *J Clin Lab Anal* (2022) 36(4):e24331. doi: 10.1002/jcla.24331
- Wang SM, Liu GQ, Xian HB, Si JL, Qi SX, Yu YP. Lncrna neat1 alleviates sepsis-induced myocardial injury by regulating the tlr2/nf-kappa signaling pathway. *Eur Rev Med Pharmacol Sci* (2019) 23(11):4898–907. doi: 10.26355/eurrev\_201906\_18078
- Deng Z, Sun M, Wu J, Fang H, Cai S, An S, et al. Sirt1 attenuates sepsis-induced acute kidney injury via beclin1 deacetylation-mediated autophagy activation. *Cell Death Dis* (2021) 12(2):217. doi: 10.1038/s41419-021-03508-y
- Pan X, Ji Z, Xue J. Percentage of peripheral cd19+cd24hicd38hi regulatory b cells in neonatal sepsis patients and its functional implication. *Med Sci Monit* (2016) 22:2374–8. doi: 10.12659/msm.895421
- Chaturvedi V, Marsh RA, Zoref-Lorenz A, Owsley E, Chaturvedi V, Nguyen TC, et al. T-Cell activation profiles distinguish hemophagocytic lymphohistiocytosis and early sepsis. *Blood* (2021) 137(17):2337–46. doi: 10.1182/blood.2020009499
- Kim EY, Ner-Gaon H, Varon J, Cullen AM, Guo J, Choi J, et al. Post-sepsis immunosuppression depends on nkt cell regulation of mtor/ifn-gamma in nk cells. *J Clin Invest* (2020) 130(6):3238–52. doi: 10.1172/JCI128075
- Chen W, Wang Y, Zhou Y, Xu Y, Bo X, Wu J. M1 macrophages increase endothelial permeability and enhance p38 phosphorylation via ppar-gamma/cxcl13-cxcr5 in sepsis. *Int Arch Allergy Immunol* (2022) 183(9):997–1006. doi: 10.1159/000524272
- Dahdah A, Gautier G, Attout T, Fiore F, Lebourdais E, Msallam R, et al. Mast cells aggravate sepsis by inhibiting peritoneal macrophage phagocytosis. *J Clin Invest* (2014) 124(10):4577–89. doi: 10.1172/JCI75212
- Park I, Kim M, Choe K, Song E, Seo H, Hwang Y, et al. Neutrophils disturb pulmonary microcirculation in sepsis-induced acute lung injury. *Eur Respir J* (2019) 53(3). doi: 10.1183/13993003.00786-2018
- Richendrer H, Jay GD. Lubricin as a therapeutic and potential biomarker in sepsis. *Crit Care Clin* (2020) 36(1):55–67. doi: 10.1016/j.ccc.2019.08.005
- Wang WD, Yang XR, Guo MF, Pan ZF, Shang M, Qiu MJ, et al. Up-regulation of btlA expression in myeloid dendritic cells associated with the treatment outcome of neonatal sepsis. *Mol Immunol* (2021) 134:129–40. doi: 10.1016/j.molimm.2021.03.007
- Wang L, Zhang J, Zhang L, Hu L, Tian J. Significant difference of differential expression pyroptosis-related genes and their correlations with infiltrated immune cells in sepsis. *Front Cell Infect Microbiol* (2022) 12:1005392. doi: 10.3389/fcimb.2022.1005392