scientific reports

Check for updates

OPEN

Neutrophil-based single-cell sequencing combined with transcriptome sequencing to explore a prognostic model of sepsis

Hao Zhang[1,2], Simiao Chen[1,2], Yiwen Wang[1,2], Ran Li[1,2], Qingwei Cui[1,2], Mengmeng Zhuang[1,2] & Yong Sun[1,2]✉

Sepsis is a life-threatening condition influenced by various factors. Although gene expression profiling has offered new insights, accurately assessing patient risk and prognosis remains challenging. We utilized single-cell and gene expression data of sepsis patients from public databases. The Seurat package was applied for preprocessing and clustering single-cell data, focusing on neutrophils. Lasso regression identified key genes, and a prognostic model was built. Model performance was evaluated using Receiver Operating Characteristic (ROC) curves, and further analyses, including immune cell infiltration, Gene Set Enrichment Analysis (GSEA), and clinical correlation, were conducted. Several neutrophil subtypes were identified with distinct gene expression profiles. A prognostic model based on these profiles demonstrated strong predictive accuracy. Risk scores were significantly correlated with clinical features, immune responses, and key signalling pathways. This study provides a comprehensive analysis of sepsis at the molecular level. The prognostic model shows promise in predicting patient outcomes, offering potential new strategies for diagnosis and treatment.

Keywords Prognostic models, Single-cell sequencing, Neutrophils, Sepsis, Key genes

Sepsis is a life-threatening organ dysfunction due to a dysregulated host response to infection[1]. It is currently believed that sepsis can further develop into severe sepsis, septic shock, and multiple organ dysfunction syndrome (MODS). These are defined as follows: severe sepsis as sepsis and sepsis-induced multiple organ dysfunction or tissue perfusion deficiency; and septic shock as sepsis-induced[2–4]. The principal symptoms of sepsis include chills, fever, panic, shortness of breath, and altered mental status. Severe sepsis is a reaction to any kind of microorganism, not primarily the entry of microorganisms into the bloodstream, but due to local inflammation inducing dysfunction in distant organs. The pathogenesis of sepsis is characterised by the interconnection and influence of various systems within the organism. The imbalanced inflammatory response and systemic organ failure due to coagulation dysfunction permeate the entire pathophysiology of sepsis, exerting a greater impact on the structure, metabolism, apoptosis, and specific functions of cells within the immune system. The coagulation system and the systemic organs are also affected, and gene expression plays a role in this. The immune status reflects the strong heterogeneity of sepsis, but the complexity of the pathophysiological mechanisms of sepsis, in which various organ systems interact with each other through inflammatory responses, still needs to be actively explored[5]. Globally, sepsis affects more than 19 million individuals annually, with 6 million fatalities and a mortality rate exceeding a quarter of all deaths. Furthermore, approximately 3 million of the surviving patients have cognitive dysfunction[6,7]. A study demonstrated that the incidence and mortality rates of sepsis and infectious shock in China are considerably higher than those in North America and European countries. Additionally, the morbidity and mortality rates of sepsis and infectious shock in mainland China are extremely high[8]. A number of case reports have demonstrated that patients presenting with severe clinical manifestations of sepsis or septic shock may have negative microbiological tests[9], which makes early recognition and diagnosis of sepsis challenging. Currently, sepsis has a poor prognosis and its pathogenesis remains poorly understood.

[1]Department of Burn Surgery, The Affiliated Huaihai Hospital of Xuzhou Medical University, Xuzhou 221004, Jiangsu Province, China. [2]Department of Burn Surgery, The 71st Group Army Hospital of PLA, Xuzhou 221004, Jiangsu Province, China. ✉email: sunyong_97@163.com

The prognosis of patients with sepsis can be significantly improved by early recognition of the condition[10]. Therefore, early and accurate recognition of sepsis, an appropriate assessment of the severity of the disease, and the timely administration of an individualised treatment plan are essential for the diagnosis and treatment of sepsis.

The rapid development of molecular biology technology, particularly the advent of high-throughput sequencing technology, has led to the widespread adoption of transcriptome sequencing (RNA-Seq) as a common experimental tool. However, the inherent limitation of transcriptome sequencing, which yields an average of gene expression across a population of cells, has prompted the maturation of single-cell RNA-Seq technology based on high-throughput sequencing. Single-cell sequencing technology is revolutionising biomedical research and clinical practice due to its ability to comprehensively characterise cells in complex tissues. In recent years, single-cell histological sequencing technology has addressed many of these questions and revolutionised the stem cell field[11–13]. This commenced with the development of the inaugural single-cell RNA sequencing technology in 2009[14,15], which has been investigated by laboratories in a multitude of biological disciplines, including neurobiology, developmental biology, immunology and tumour biology[16]. Furthermore, the significance of single-cell sequencing is expanding from basic research on life activities to the diagnosis and treatment of various diseases, providing novel diagnostic and therapeutic concepts for numerous critical diseases.

In this study, we selected complete single-cell sequencing data of some sepsis patients from the GEO database, constructed prognostic correlation models and tested their predictive efficacy after bioinformatics analysis. Concurrently, we performed clinical index correlation analysis, immune cell infiltration analysis, GSEA analysis and nomogram model construction in order to further explore the correlation between risk scores and multiple biological processes and signalling pathways, and to better understand the mechanisms of sepsis development. In conclusion, this study offers a novel approach to the clinical diagnosis and treatment of sepsis, as well as a comprehensive insight into the pathogenesis of sepsis.

Materials and methods

Data sources

The Gene Expression Omnibus (GEO) database, managed by the National Center for Biotechnology Information (NCBI), is a widely used repository for gene expression data. Single-cell RNA sequencing data for the dataset GSE167363 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE167363) were obtained from the NCBI GEO public database[17]. This included data from 12 samples with complete single-cell expression profiles, which were downloaded for further analysis. Additionally, the Series Matrix File for dataset GSE65682 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65682) was retrieved from GEO, along with the corresponding annotation file GPL13667 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL13667), providing expression profile data for a total of 802 patients[18–33].

Single-cell analysis

Firstly, expression profiles were imported into the Seurat package[34], and samples with aberrant expression were excluded based on the Unique Molecular Indentifier(UMI, a short sequence of randomised or specific nucleotides)counts, gene counts, and mitochondrial gene percentage captured for each cell. The data underwent sequential normalisation, and linear downscaling by principal components analysis (PCA, a linear dimensionality reduction algorithm and a commonly used data preprocessing method) to identify the optimal PC counts through the fragmentation plot. The positional relationship between each cluster was obtained through non-linear downscaling of Uniform Manifold Approximation and Projection (UMAP, a nonlinear dimensionality reduction and visualisation algorithm). Cellular annotations were performed using the CellMarker and PanglaoDB databases, as well as a literature search to identify the cell types present in the corresponding tissues and the corresponding marker genes for cell annotation.

Model construction and prognosis

A set of prognosis-related genes was identified and subjected to Lasso regression analysis to construct additional prognostic models[35]. The expression values of each gene were incorporated into a risk score formula for each patient, with the respective gene expression weighted by its corresponding Lasso regression coefficient. The risk score formula is as follows: Risk Score = (Gene 1 Expression × Coefficient 1) + (Gene 2 Expression × Coefficient 2) + . + (Gene n Expression × Coefficient n). Using the median risk score as the threshold, patients were categorized into low-risk and high-risk groups. Kaplan-Meier analysis was then applied to assess survival differences between these groups, and the log-rank test was used for comparison. The predictive power of the risk scores for patient prognosis was evaluated through Lasso regression and stratified analysis, while ROC curves were employed to assess the accuracy of the model predictions.

Validation of key genes

Using cecum ligation and perforation(CLP) to establish an animal model of sepsis in mice, and using Quantitative Real-time PCR (qRT-PCR), experiments were designed to verify the differentiation of the above key genes in sepsis patients.All the following experiments were approved by the University Laboratory Animal Ethics Committee of Xuzhou Medical University.All the following experiments were performed in accordance with relevant guidelines and regulations.All experiments described below were conducted in accordance with the ARRIVE guidelines.

Animals
C57BL/6 male mice (25–30 g), 20 in total, were sourced from the Animal Experiment Centre of Xuzhou Medical University. The experimental design was approved by the University Laboratory Animal Ethics Committee of Xuzhou Medical University.

Primary reagents

- 1% Pentobarbital sodium injection.
- 75% Alcohol solution.
- Aline (medicine).
- Red Cell's Lysis Buffer.
- TRnaZol Reagent.
- Chloroform.
- DEPC water.
- FastKing RT Kit(With gDNase).
- SYBR Green kit.

Surgical procedure
Preoperative setup Mice for the experiment have been kept in a well-ventilated, constant temperature (22 °C) environment for more than 1 week. Before the start of the experiment, the mice used were fasted for 12 h with unlimited water intake. The mice were randomly divided into two groups: sepsis group and control group, 10 animals in each group.

Surgical operation The animals in the two groups were treated according to the following operation:

1. Anaesthesia: animals in both groups were injected with sodium pentobarbital via intraperitoneal injection at a dose of 50 mg/kg body weight, and after the animals were successfully anaesthetised, the abdominal skin was disinfected and prepared; No chloroform was used during the anesthesia;
2. Preoperative disinfection: the animals in both groups were fixed on the operating table and the abdominal surgical area was disinfected three times with 75% alcohol;
3. Surgical operation: a mid-abdominal incision was made to open the abdominal cavity for about 2 cm, searching for the cecum, carefully separating its distal end from the mesentery of the large intestine, and then adopting different treatments for the two groups: the sepsis group ligated the distal 1/2 of the cecum with a sterile 4-gauge silk thread, and punctured through the centre of the distal end of the ligated cecum with a sterile No. 9 needle, gently squeezing out a small amount of intestinal contents, avoiding damage to blood vessels as much as possible, and then pushing the cecum back into the abdominal cavity, and closing the abdominal cavity. Then the cecum was pushed back into the abdominal cavity and the incision was closed; in the control group, after separating the mesentery of the distal cecum and the large intestine, the cecum was pushed back into the abdominal cavity and the incision was closed; the abdominal suture was sterilised.

Postoperative care Resuscitate animals by injecting prewarmed normal saline (37 °C; 5 ml per 100 g body weight) subcutaneously.Place mice or rats back in cages in a temperature-controlled room (22 °C) with 12-h light and dark cycles and monitor them every 6 h.When the mice in the sepsis group showed symptoms such as fever, significantly accelerated heart rate and respiratory rate, increased secretions from the oral and nasal cavities, depression, lethargy, curling up, erect hair, less movement, refusal to eat or less food, and secretions from the corners of the eyes, whole blood of the animals in the two groups was drawn with an anticoagulant tube and set aside.

RNA extraction from whole blood and quantification by qRT-PCR technique
The RNA corresponding to the key genes was extracted from the leukocyte precipitation of the above spare whole blood, reverse transcription was performed to synthesise cDNA, and qRT-PCR was performed to quantify the expression and record the data.The specific experimental operations are as follows:

RNA extraction Firstly, fresh blood samples were shaken well and placed on ice. An appropriate amount of blood was drawn into an RNase-free plastic centrifuge tube and 3 times the volume of Red Cell's Lysis Buffer was added, the plastic centrifuge tube containing the mixture was shaken vigorously on a shaker for 30 s, mixed thoroughly and placed on ice for 10 min. The centrifuge tube containing the mixture was centrifuged at 10,000 r/min for 1 min at 4 °C, the upper layer of liquid was removed and the bottom sediment (leukocyte precipitate) was taken. The upper layer of liquid was removed and the precipitate (leukocyte precipitate) was taken from the bottom of the tube. Add 1 ml of TRnaZol Reagent per 200 µl of blood collected, add TRNzol reagent to the collected leukocyte precipitate, mix by blowing and then shake well, leave on ice for 5 min, centrifuge at 12,000 r/min at 4 °C for 10 min, and remove the upper layer of clear liquid. According to the fixed ratio of TRnaZol: chloroform = 5:1, add chloroform into the colourless aqueous phase, then shake well and mix well, leave on ice for 10 min, centrifuge at 12000r/min at 4℃ for 10 min, you can see three layers in the centrifuge tube: the upper layer is a colourless aqueous phase; the middle is the precipitate of DNA; the lower layer is a pink phenol-chloroform organic phase. Transfer the upper layer to a new centrifuge tube, add 0.5 ml of isopropanol for every 1 ml of TRnaZol, add pre-cooled isopropanol to the colourless aqueous phase and mix well with shaking, let it stand on ice for 10 min, and then centrifuge it at 12,000 r/min at 4 °C for 10 min, discard the supernatant in the centrifuge tube, keep the white precipitate at the bottom of the tube, and perform a process of leaving it to

air-dry for 30 s to remove the The white precipitate was retained at the bottom of the tube and air-dried for 30 s to remove the residual liquid in the tube. Add appropriate amount of pre-cooled 75% anhydrous ethanol to the white precipitate, blow repeatedly until the precipitate is completely dissolved, repeatedly invert and shake well, centrifuge at 12000r/min at 4℃ for 10 min, carefully discard the supernatant in the tube, and use a pipette gun to suck up the rest of the supernatant, keep the precipitate at the bottom of the tube and blow-dry it until it becomes semi-transparent, then add appropriate amount of DEPC water to dissolve it, and blow repeatedly until fully dissolved. Repeatedly blow several times until fully dissolved, then proceed to the next step of the experiment or placed in the refrigerator at -80 ℃ for storage.

RNA reverse transcription to synthesise cDNA The extracted RNA was placed on ice, and the concentration and purity of the extracted RNA was measured using an ultra-micro nucleic acid detector. DEPC water was used as a blank control, and the ultra-micro nucleic acid detector was calibrated for 3 times, and then the concentration of the extracted RNA was measured by taking 1 μL of each sample and noted down.The FastKing RT Kit (With gDNase) kit was used for reverse transcription of RNA to synthesise cDNA.Naturally thaw the RNA template and the reagents (5gDNA Buffer, FQ-RT Primer Mix, FastKing RT Primer Mix, 10King RT Buffer, RNase-Free ddH20) in the FastKing RT Kit (WithgDNase) in a room temperature environment. Place on ice quickly after completion. All five reagent solutions should be shaken with a vortex shaker and briefly centrifuged before use.Follow the instructions in the kit to formulate the above reagents into a Mix system (10×King RT Buffer 2 μL, FastKing RT Primer Mix 1 μL, FQ-RT Primer Mix 2 μL, and RNase-Free ddH20 replenished to 10 μL), invert the mixing and shake it to use it, and then the formulated Mix system was Dispense into each 200 μL nuclease-free EP tube. Configure the Mix (5×gDNA Buffer 2 μL, Total RNA 2 μg, RNase-Free ddH20 replenished to 10 μl) according to the gDNA Removal Reaction System in the instructions included in the kit, mix upside down and shake well, then briefly centrifuge and place at 42 °C, incubate for 3 min, and then place on ice. Add the Mix system into the reaction solution after gDNA removal reaction, mix well and shake briefly, incubate at 42 ‘C for 15min, 95 ’C for 3min, then the cDNA can be obtained for subsequent experiments or frozen in the refrigerator at -20℃.

Quantitative real-time PCR (qRT-PCR) The SYBR Green kit was used for qRT-PCR. The Real-Time PCR reaction system (2×SuperReal PreMix Plus 5 μL, cDNA template 1 μL, forward primer 0.2 μL, reverse primer 0.2 μL, RNase-Free ddH20 3.6 μL) was set up according to the instructions of the kit. The above reagents were thawed naturally at room temperature, thoroughly inverted and shaken, and then briefly centrifuged. After the reaction system was configured and centrifuged, the reaction system was put into the Roche LightCycler 480 real-time quantitative fluorescence PCR instrument and the amplification programme was set as follows: 1×95℃ for 15 min, 50×(95℃ for 10s, 60℃ for 30s). The data obtained by real-time PCR were analysed by the $2^{-\Delta\Delta CT}$ relative quantification method.

1.5 The PCR results of different genes in control and sepsis groups were counted and plotted separately.

Correlation analysis of clinical indicators

The correlation of clinical characteristics is an important study because it allows a more profound comprehension of the patient's condition and prognosis. To analyse the risk score correlation with clinical characteristics in sepsis data, clinical correlation analysis is principally based on the correlation between patients' survival status, age and gender.

Immune cell infiltration analysis

CIBERSORT is a method for characterising the cellular composition of complex tissues from gene expression profiles, enabling large-scale analysis of RNA mixtures for cellular biomarkers and therapeutic targets[36,37]. The method is based on the principle of support vector regression and back-convolution analysis of the expression matrix of immune cell subtypes. It contains 547 biomarkers distinguishing 22 human immune cell phenotypes, including T-cell, B-cell, plasma cell, and myeloid cell subpopulations. In this study, patient data were analysed using the CIBERSORT algorithm, which was employed to infer the relative proportions of the 22 immune-infiltrating cells and to correlate gene expression as well as immune cell content.

GSEA analysis

GSEA is a widely used computational method for identifying differentially expressed genes with specific common characteristics. Additionally, GSEA can be employed to analyse related genes with the same characteristics within a given dataset[38–41]. Patients were divided into high- and low-risk groups according to the model, and the signalling pathway differences between the two groups were further analysed by GSEA. The background gene sets were downloaded from the MsigDB database as annotated gene sets for the subtype pathways, which were then subjected to differential expression analysis of the pathways between subtypes. The significantly enriched gene sets (with an adjusted p-value of less than 0.05) were ranked according to the concordance score.

Nomogram model construction

Nomograms[42] are based on regression analysis, which is based on risk scores and clinical symptoms. These are then plotted on the same plane using scaled line segments at a certain scale to express the interrelationships between the variables in a predictive model. The nomogram is a graphical tool designed to quickly approximate complex calculations without the need for a computer or calculator. It was invented in the 19th century and was popular before calculators and computers became commonplace[43]. The predictive values are calculated by constructing a multifactor regression model, assigning a score to each value level of each influencing factor in

the model based on the degree of contribution of each influencing factor to the outcome variable (the magnitude of the regression coefficients), and then summing up the individual scores to obtain a total score.

Clinical applications of nomogram modelling

A total of 56 patients admitted to the ICU for the treatment of sepsis diagnosed in the Affiliated Huaihai Hospital of Xuzhou Medical University between June 2023 and June 2024 were selected for the study. This study was reviewed and approved by the Ethics Committee of the Affiliated Huaihai Hospital of Xuzhou Medical University in January 2023(LL-2023YX04).All research was performed in accordance with relevant guidelines/regulations and informed consent was obtained from all participants and/or their legal guardians. The risk score model was employed to calculate the risk score of each patient, with age and gender substituted into the nomogram model. In accordance with the formulae of the prediction model, a patient was considered to be alive when the survival rate was $\geq 70\%$. The predicted and actual survival of the patients at 14 days and 28 days were recorded. The sensitivity, specificity and accuracy of the model were calculated based on the prediction results. A four-cell table was constructed, comprising the actual and predicted survival of patients at 14 and 28 days. The chi-square test was then employed to ascertain whether there was a statistically significant difference between the probabilities of the two groups.

Statistical analyses

Survival curves were generated by kaplan-meier method and compared by log-rank. Multivariate analyses were performed using the cox proportional risk model. All statistical analyses were performed using the R language (version 4.3.0) and were statistically significant at $p < 0.05$.

Results

The annotation of the cell

Expression profiles were first read in via the Seurat package, in which we filtered cells based on the total number of UMIs per cell, the number of genes expressed, and the percentage of mitochondrial reads per cell. Where outliers were defined as three median absolute deviations (MAD) from the median. And low expressing cells were screened out based on violin plots, scatter plots (nFeature_RNA > 200 & percent.mt <= 3*MAD & nFeature_RNA <= 3*MAD & nCount_RNA <= 3*MAD & percent. ribo <= 3*MAD), followed by doublet filtering using the DoubletFinder package[44], retaining a total of 50,944 cells, filtered violin plots, scatter plots (Supplementary Fig. 1A-B), and the 10 genes with the highest standard deviation among them are shown (Supplementary Fig. 1C). The data were sequentially processed by standardisation, homogenisation, PCA, and harmony analysis (Supplementary Fig. 1D-F), and the positional relationship between each cluster was obtained by UMAP analysis, and a total of 24 subclusters were obtained (Fig. 1.A). In this study, each subtype was further annotated and all clusters were annotated to B cells, CD14 Monocytes, FCGR3A monocytes, Dendritic Cells, Neutrophils, CD4 T cells, CD8 T cells, NK cells, Erythroid precursors, and Platelets, the 10 cell classes (Fig. 1.B). bubble plots of classical markers for the 10 cells (Fig. 1.C) and histograms of the proportions of cells corresponding to subgroups or samples (Fig. 1.D).

Secondary clustering of the cells

A total of 1817 cells were extracted from neutrophils. Of these, 74 cells were extracted from the control group and 1743 cells were extracted from the disease group. The neutrophils were then subjected to secondary clustering (Fig. 2.A-E), with PC = 10. The neutrophils were then subjected to secondary annotation, and the cells were annotated by the known and highly expressed cellular markers in this study. Neutrophils were subsequently annotated to five cell classes: CSF3R Neutrophils, MMP9 Neutrophils, LYZ Neutrophils, S100A9 Neutrophils, and CST3 Neutrophils (Fig. 2.F-G). Differential analysis of gene expression levels was subsequently performed for neutrophils. The differential gene screening conditions were as follows: *p-value adjusted (p_val_adj) < 0.05* and absolute value of log fold change (logFC) > 1. A total of 167 differential genes were obtained (Fig. 2.H).

Construction of prognostic models

In this study, clinical information of sepsis patients was collected from the GEO database, and prognostic genes in sepsis were screened by Cox one-way regression using the differentially related genes obtained in the previous step. The results showed that a total of 12 prognostic-related genes were screened by Cox one-way regression (pvalue < 0.01) (Fig. 3.A). The prognostic genes were screened according to the prognostic genes in sepsis by lasso regression feature selection algorithm. We randomly divided the sepsis dataset in the GEO database of processed patients with survival data into training and test sets in a ratio of 4:1, and the best risk score value corresponding to each sample was obtained for subsequent analyses after analysing the data by lasso regression (Fig. 3.B-D). Risk Score = CCL5 × (-0.332980575178139) + SELL × (-0.298814565902524) + RPS24 × 0.161868307092205 + ASPH × 0.236358808152168 + PPP1R15A × 0.255888626367807 + PPDPF × 0.321902222535904. The patients will be divided into high and low risk groups based on risk scores and analysed using Kaplan-Meier curves. The overall survival (OS) of the high-risk group was significantly lower than that of the low-risk group in both the training set and test set (Fig. 3.E-F). Furthermore, the results of the ROC curves in both the training set and test set indicated that the model exhibited superior validation efficacy (Fig. 3.G-H).

Validation of key genes

In this study, the key genes in whole blood of mice in sepsis group and control group were quantified. The results(Fig. 4.A) indicated that the relative expression of CCL5 and SELL in mice of the sepsis group was significantly lower than that of the control group; the relative expression of RPS24, ASPH, PPP1R15A and PPDPF in mice of the sepsis group was significantly higher than that of the control group ($p<0.05$).

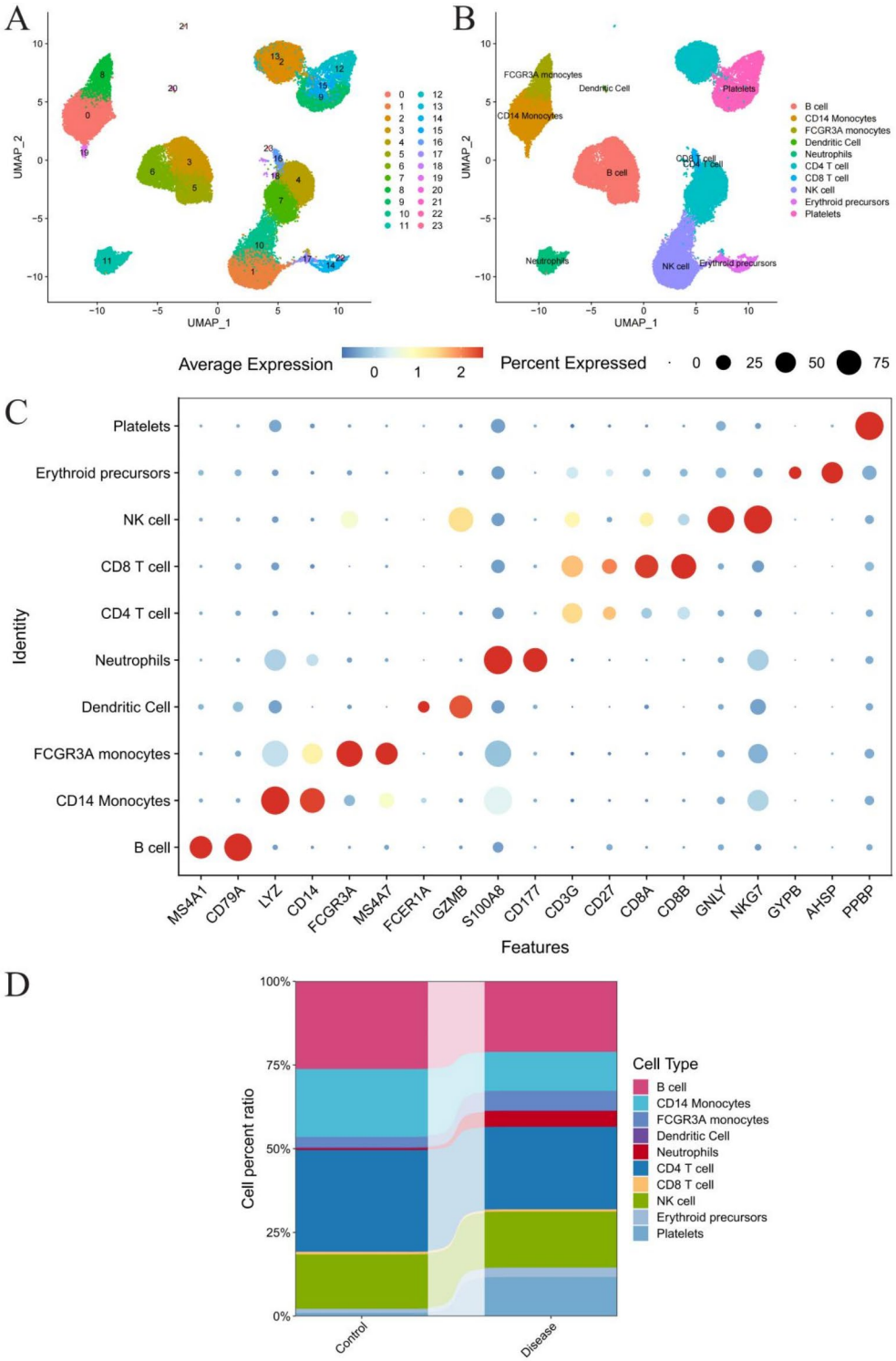

Fig. 1. Cells were annotated based on expression profiles. (**A**) We classified the cells into 24cluster by UMAP algorithm based on the important components available in PCA. (**B**) Cellular annotation profile of 24 clusters, 24 clusters were annotated into 10 cell types, i.e. B cells, CD14 Monocytes, FCGR3A monocytes, Dendritic Cells, Neutrophils, CD4 T cells, CD8 T cells, NK cells, Erythroid precursors, Platelets. (**C**) Bubble diagram of Doplot of 10 cells with cell markers. (**D**) Difference in the percentage of content of the 10 cells in the two groups of samples.

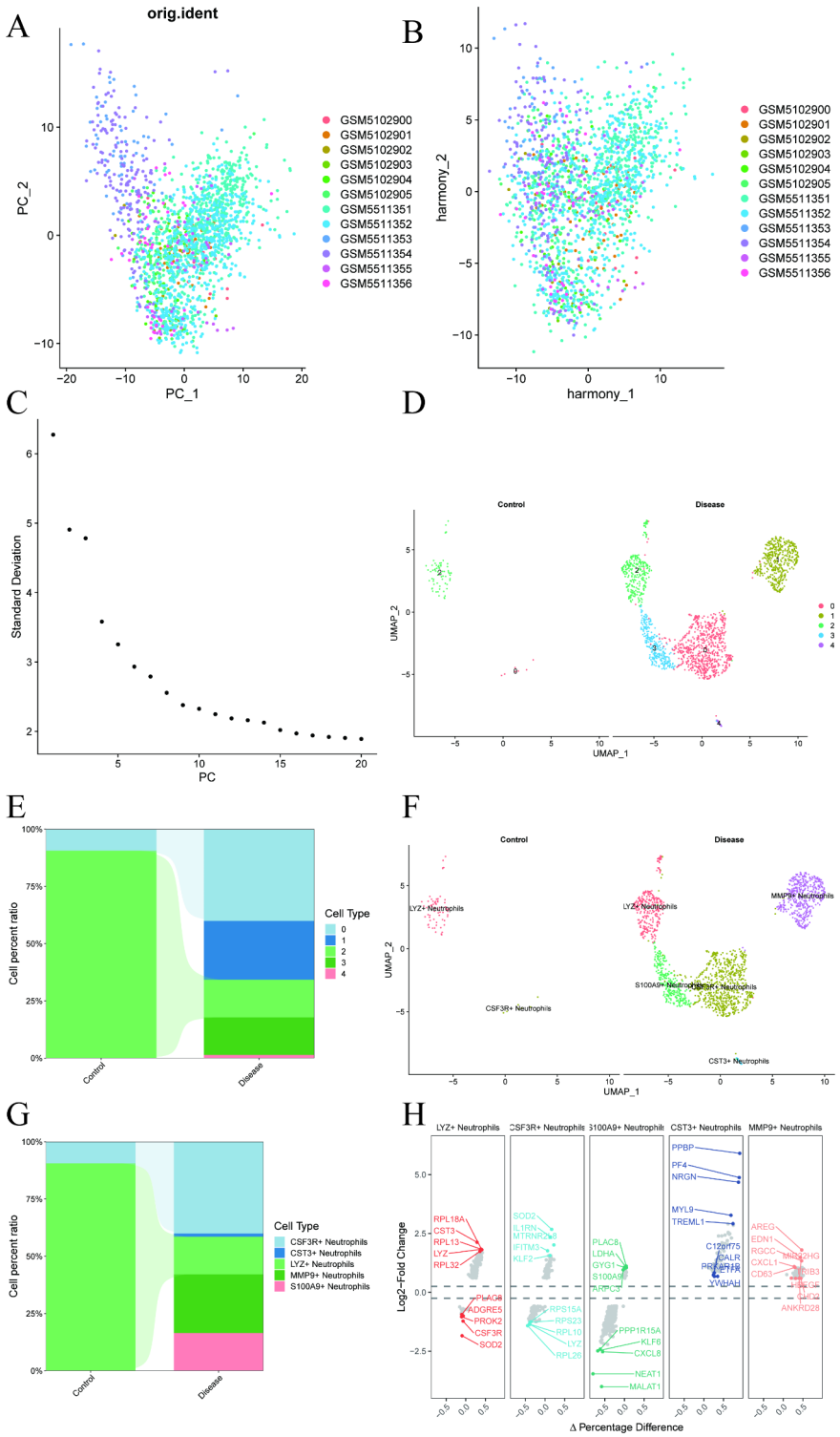

Fig. 2. Cells were subjected to secondary clustering. (**A**,**B**) Display of the secondary clustering PCA and the distribution of PCs, dots represent cells and colours represent samples. (**C**) Variance ranking plot for each PC of the secondary clustering. (**D**) Secondary clustering divides the cells into 5 clusters by the UMAP algorithm based on the significant components available in the PCA. (**F**) Cellular annotation status of the 5 clusters, 5 clusters were annotated into 5 cell types, i.e. CSF3R Neutrophils, MMP9 Neutrophils, LYZ Neutrophils, S100A9 Neutrophils, CST3 Neutrophils. (**E**–**G**) Difference in the content percentage of the 5 cell types in the two groups of samples. (**H**) Differential volcano plots of the five cell types.

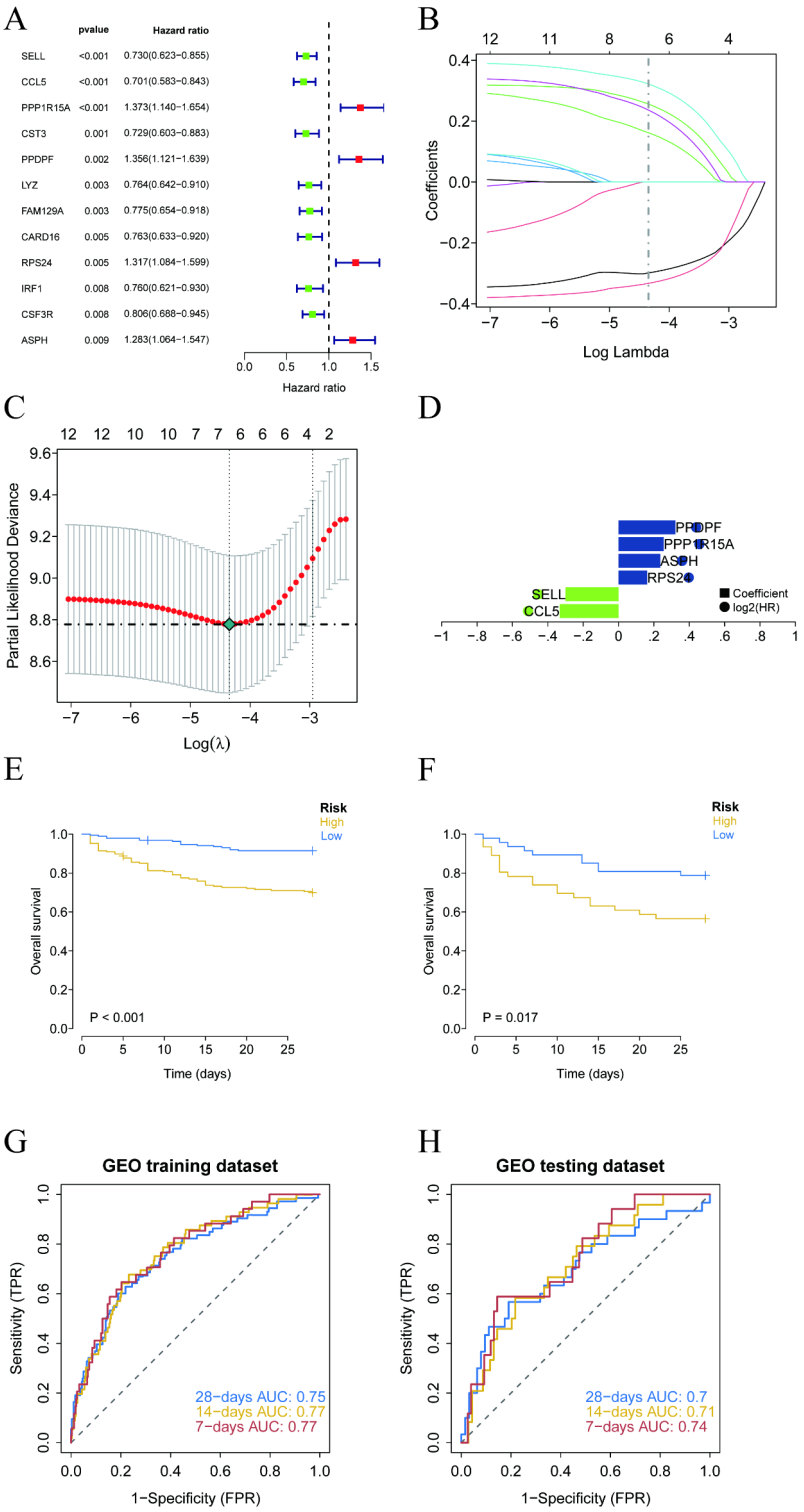

Fig. 3. Constructing prognostic models. (**A**) One-factor forest plot of prognostic genes. (**B**) LASSO coefficient distribution of prognostic genes and gene combinations at minimum lambda values. (**C**) Ten-fold cross-validation of tuning parameter selection in the LASSO model to determine the minimum lambda value. (**D**) Coefficients of Lasso genes. (**E**,**F**) Survival curves for the training set vs. test set models. (**G**,**H**) ROC curves (7-14-28 days) of the models in the training set vs. the test set.

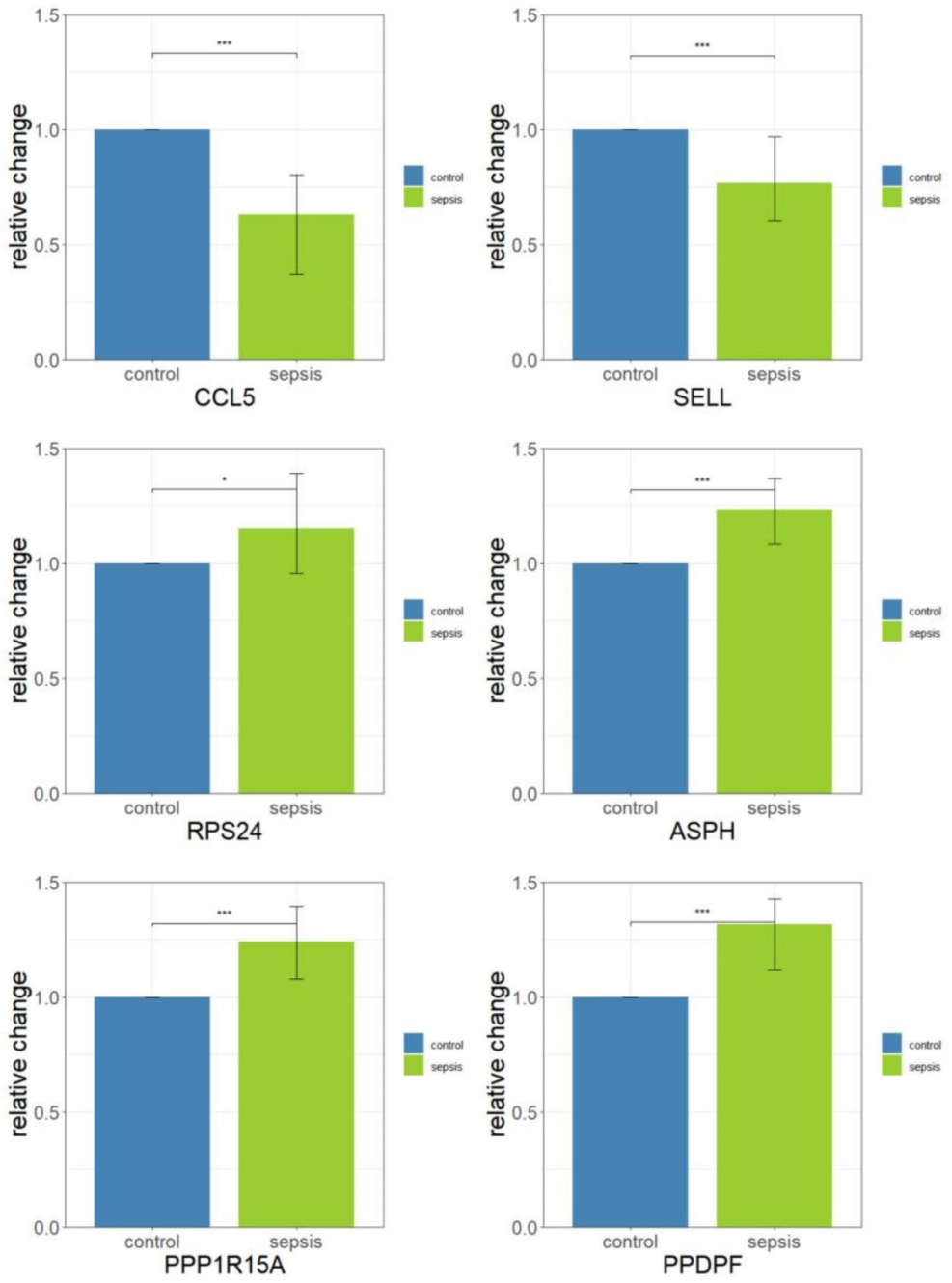

Fig. 4. A Validating key genes. (A) Comparative relative expression of key genes.

Correlation of risk scores with clinical indicators

In this study, the size of the clinical indicator values was used to divide the samples corresponding to the risk score values into different groups. The results of each clinical indicator grouping were presented through box-and-line diagrams (Fig. 5.A-C). The distribution of the risk score values was found to be significant between the groups by the rank-sum test in the clinical indicator of Fustat ($p\text{-}value < 0.05$). This indicates that the risk scores obtained from the modelling analysis have good applicability for the typing of sepsis patient samples.

Immune infiltration analysis

The microenvironment is mainly composed of a combination of immune cells, extracellular matrix, multiple growth factors, inflammatory factors and specific physicochemical features, which significantly influences the diagnosis of the disease and the sensitivity to clinical treatment. We further explored the potential molecular mechanisms by which risk scores influence sepsis progression by analysing the relationship between risk scores and immune infiltration in a sepsis dataset, demonstrating the percentage of immune cell content in each patient

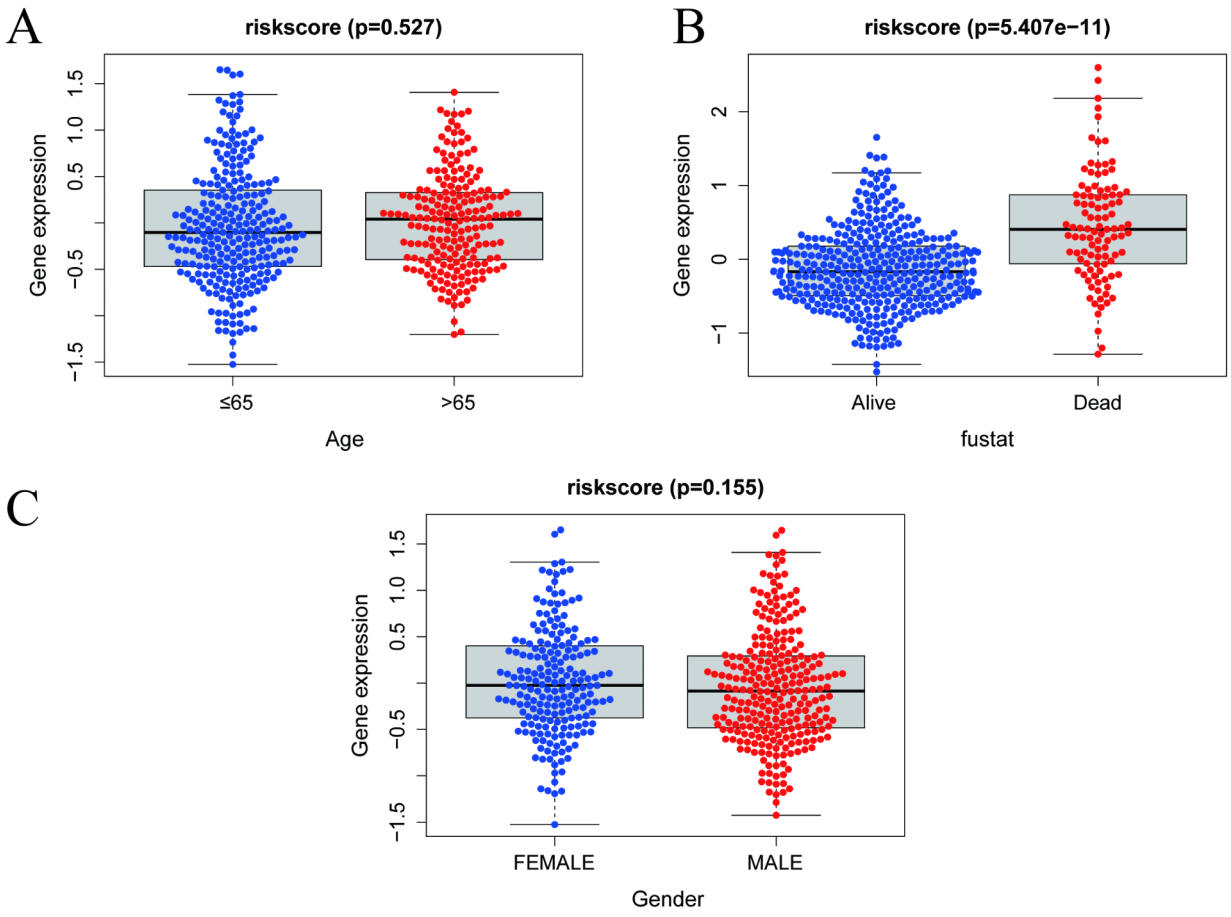

Fig. 5. Exploring the correlation between risk scores and clinical indicators. (**A**–**C**) Correlation analysis of risk scores with clinical indicators.

and the correlation between immune cells (Fig. 6.A-B). In this study, we further compared the difference in immune cell content between the low-risk and high-risk groups, and showed that Monocytes and T cells CD8 were significantly lower in the high-risk group; Macrophages M0, Mast cells resting, NK cells resting, T cells CD4 memory activated, T cells CD4 naive, etc. were significantly higher (Fig. 6.C). We then went on to explore the relationship between risk scores and immune cells and found that risk scores were significantly positively correlated with Eosinophils, Macrophages M0, Macrophages M2, NK cells resting, T cells CD4 naive, Dendritic cells resting, Mast cells resting significantly positively, and Neutrophils significantly negatively (Fig. 6.D).

Correlation of risk scores with GSEA analysis and metabolic pathways

We proceeded to investigate the specific signalling pathways involved in the high- and low-risk related models with a view to exploring the potential molecular mechanisms by which risk scores affect disease progression. The GSEA results indicated that the pathways involved were ferroptosis, nitrogen metabolism, and phagosome (Fig. 7A). The molecular action network between the pathways is shown in Fig. 7B. In this study, we examined the relationship between risk scores and metabolic pathways. The metabolic pathway heatmap revealed that the high and low expression groups exhibited notable differences in amino acid metabolism-relevant signatures, lipid metabolism-relevant signatures, drug metabolism-relevant signatures, other metabolism signatures, and C3-specific metabolism signatures, with higher activity differences observed in the high-risk groups (Fig. 7C).

Construction and clinical application of multifactorial and Nomogram models for risk score sheets

In this study, risk score was found to be an independent prognostic factor for sepsis patients through univariate and multivariate analyses (Fig. 8.A-B). Then the samples were divided into two groups of high risk and low risk through the median value of risk score values and their results through regression analysis were presented in the form of column line graphs, where the results of logistic regression analysis showed that risk score values contributed significantly to the scoring process of the nomogram prediction model in all our samples (Fig. 8.C). A further prediction analysis was conducted for the two periods of sepsis, 14 and 28 days, respectively (Fig. 8.D). The results of the clinical application of the nomogram prediction model (Tables 1 and 2) demonstrated that the sensitivity of the model was 84.09% and 89.74% for 14 and 28 days, respectively, with a specificity of 100% and

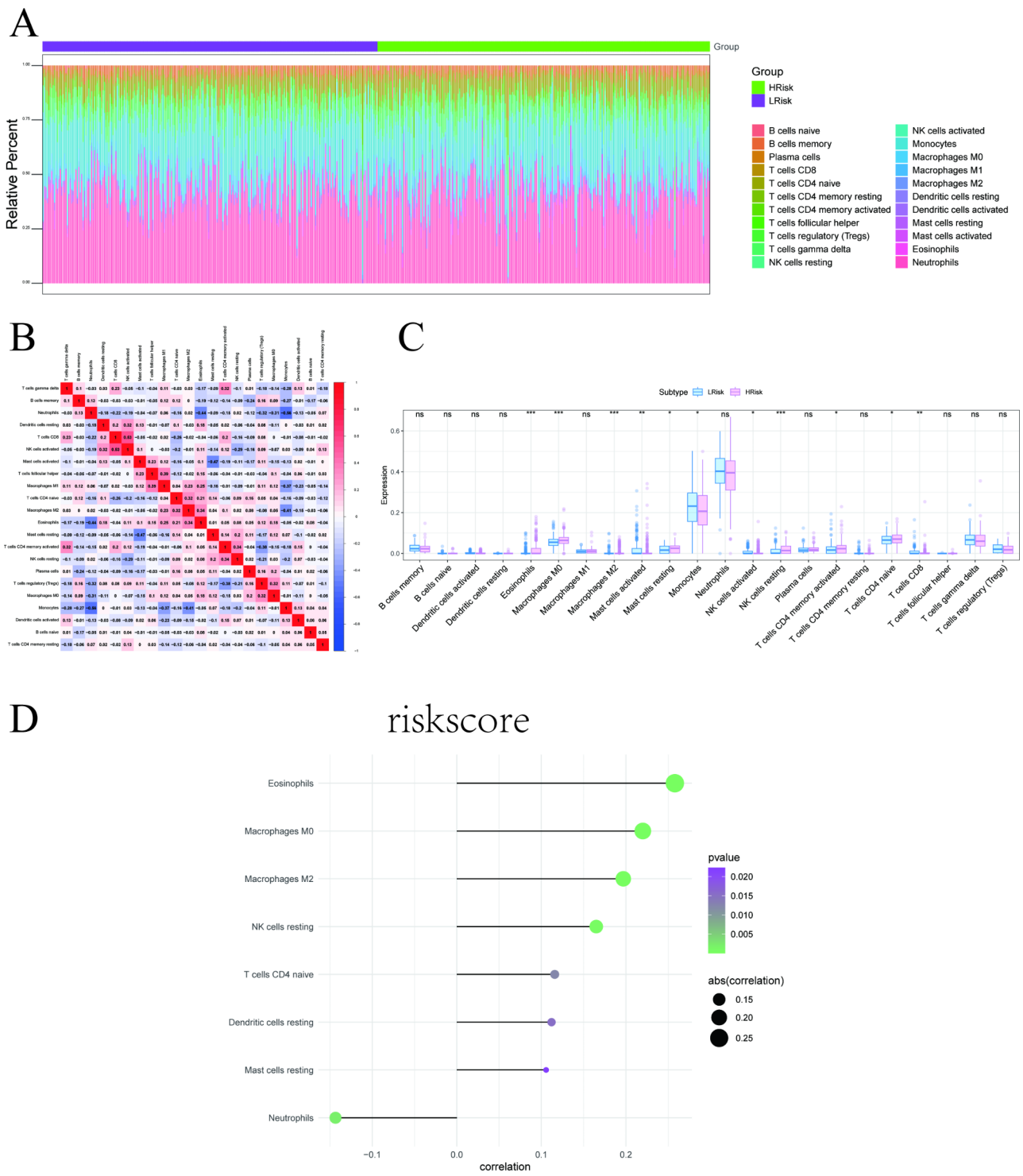

Fig. 6. Perform immuno-infiltration analysis. (**A**) Relative percentages of 22 immune cell subpopulations in patients in the high-risk and low-risk groups. (**B**) Pearson correlation between the 22 immune cells, blue indicates negative correlation and red indicates positive correlation. (**C**) Differences in immune cell content between patients in the high-risk and low-risk groups, with blue indicating low-risk patients and pink indicating patients in the high-risk group. (**D**) Correlation of immune cell content between patients in the high-risk and low-risk groups.

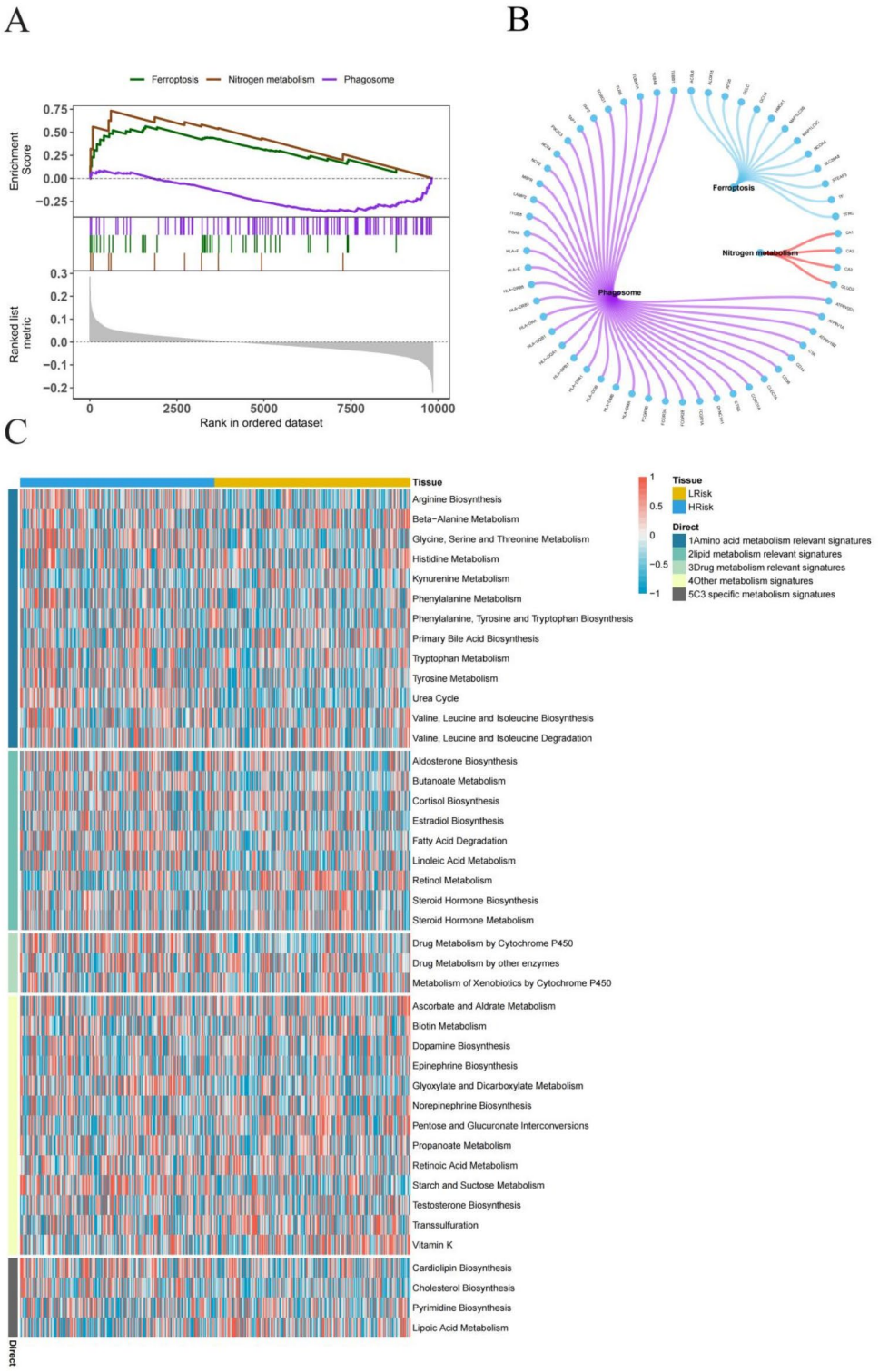

Fig. 7. Exploring the correlation of risk scores with GSEA analysis and metabolic pathways. (**A**,**B**) KEGG signalling pathways[45–47] involved in risk scoring, as well as pathway regulation and genes involved. (**C**) Heatmap of risk score with metabolic pathways.

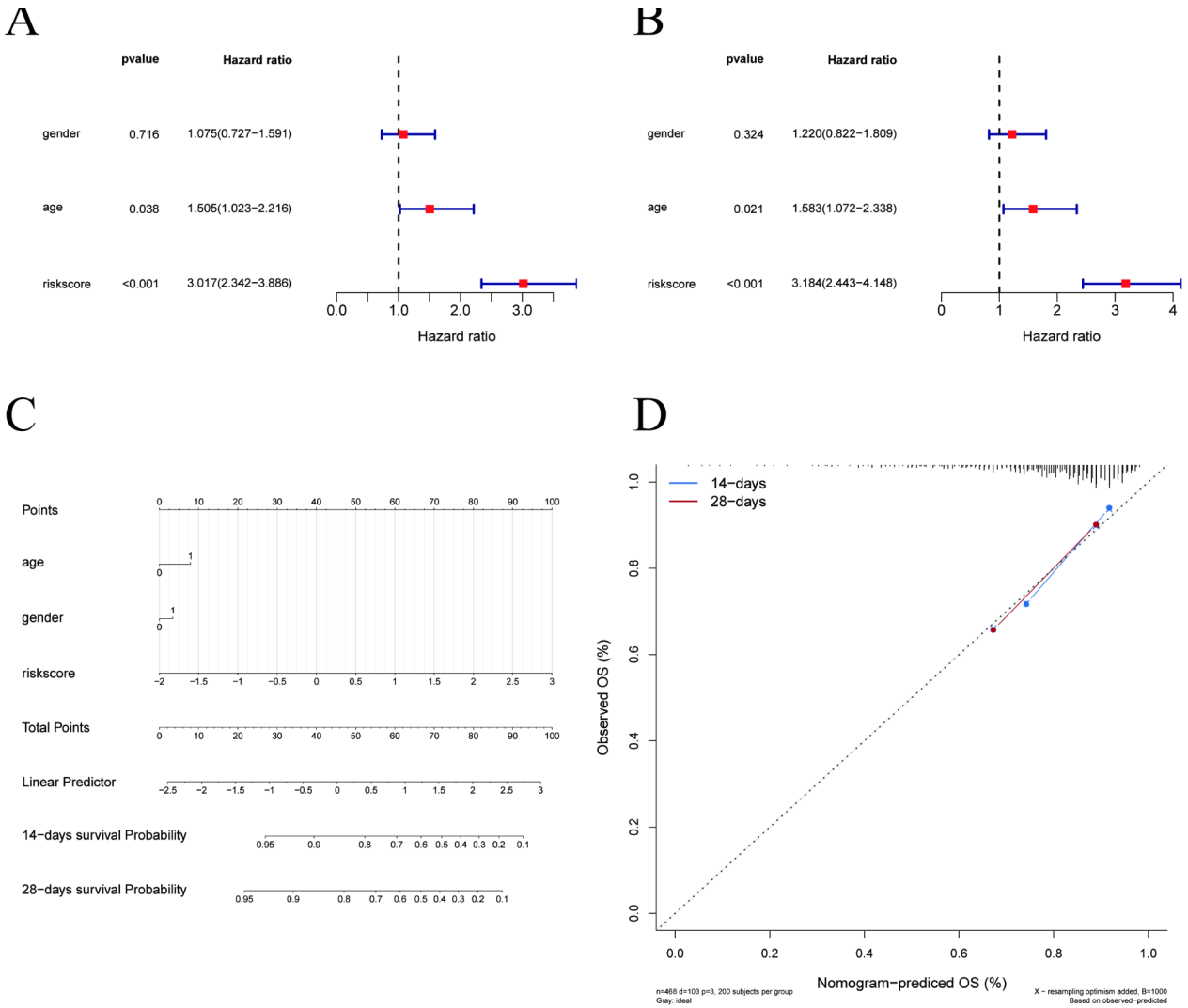

Fig. 8. Constructing Nomogram models using risk scores and clinical indicators. (**A**,**B**) Risk scores were risk factors in the unifactorial and multifactorial Cox regression analyses and were statistically different. (**C**) The values of different clinical indicators and the expression distribution of risk scores have different levels of contribution throughout the scoring process. (**D**) Predictive analyses of OS status at both 14 and 28 days.

14-day	Survivors	Deaths	Totals
Actuality	44	12	56
Prediction	37	19	56
Totals	81	31	112

Table 1. Actual and predicted survival of 56 patients with sepsis at 14 days after diagnosis (the clinical application of the nomogram prediction model).

accuracies of 87.5% and 92.86%, respectively. The chi-square values for 14 days and 28 days were calculated as 2.186 and 0.637, respectively, using the special formula for the chi-square test with four table data.

Correlation of key genes with disease progression genes

The genes associated with sepsis were obtained from the GeneCards database (https://www.genecards.org/). The top 20 genes were extracted and the intergroup expression differences of disease genes were analysed. This revealed that TNF, IL10, IL1B, HMGB, CD14, and CD163 were differentially expressed between the two groups of patients. The expression of SERPINE1, TREM1, MYD88, F5, IL18, ELANE, MIF, MYLK, ITGAM, LCN2, IL1RN, and ICAM1 was found to be significantly different between the two groups of patients. Subsequently, a correlation analysis was conducted between risk scores and disease-related genes. The results demonstrated

28-day	Survivors	Deaths	Totals
Actuality	39	17	56
Prediction	35	21	56
Totals	74	38	112

Table 2. Actual and predicted survival of 56 patients with sepsis at 28 days after diagnosis (the clinical application of the nomogram prediction model).

a significant correlation between the expression levels of risk scores and disease-related genes, with SELL exhibiting a significant positive correlation with TLR2 (cor = 0.661) and PPDPF exhibiting a significant negative correlation with MYD88 (cor=-0.387) (Fig. 9.A).

Expression profiles of single cells

In this study, the expression of six model genes in B cells, CD14 monocytes, FCGR3A monocytes, dendritic cells, neutrophils, CD4 T cells, CD8 T cells, NK cells, erythroid precursors, platelets, which are the most important genes in the world, was demonstrated (Fig. 10.A-B). Ten cells were shown (Fig. 10.A-B). Subsequently, the exhaustion factor scores and cytokine scores were obtained from the GeneCards database (https://www.genecards.org/). The risk scores were then analysed in relation to the cytokine scores versus exhaustion factor scores using the Wilcoxon test, which demonstrated a significant difference between the two groups (Fig. 10.C-D).

Discussion

Sepsis is a disease with high morbidity and mortality and is the leading cause of death in the intensive care unit (ICU). Epidemiological studies limited to ICU patients and focusing on the most critically ill are likely to underestimate the incidence of sepsis and overestimate the mortality from sepsis, with overseas data suggesting that more than half of sepsis patients are treated in general wards, and the proportion is likely to be as high as 86% in domestic hospitals[48–51]; and most efforts to define sepsis and calculate the burden of sepsis have come from large retrospective studies of hospitalised patients in high-income countries (HICs)[52]. However, population-based epidemiological studies are limited by the high mobility of the Chinese population, and the accuracy of disease incidence calculations is severely compromised[53]. Based on these two points, high-quality epidemiological studies on sepsis in China are very limited. Prognostic analyses are important objects of epidemiological studies. In 2018, Wei Feng et al. used a retrospective study to explore the epidemiological characteristics and prognostic influencing factors of sepsis, and concluded that the prognosis of severe sepsis patients with concomitant advanced age, hospital-acquired infections, multiorgan failure, and septic shock was poor[53]. For patients with severe sepsis who survive and are discharged from the hospital, there is a lack of research data on the long-term quality of life of patients surviving severe sepsis because there is no sound system of long-term follow-up of patients surviving sepsis in China, and patients are missed for various reasons. In summary, there is no clear system for the prognostic management of sepsis at this stage, and constructing a prognostic system for patients with severe sepsis is particularly important for reducing the sepsis mortality rate. In this study, we used public databases to analyse the single-cell data and serial gene expression data of some sepsis patients, selected neutrophil differentially expressed genes, in order to construct a prognostic correlation model and evaluate the predictive performance of the model, and at the same time, using the risk scoring model, we carried out a number of correlation analyses.

In order to obtain the change characteristics of neutrophil genes in sepsis patients, we extracted single-cell sequencing data from the GEO database (GSE167363) using the single-cell data file. This data was subjected to quality control, dimensionality reduction and clustering, and 10 cell categories were subsequently identified through cellular annotation. The neutrophils were found to be the second most prevalent cell type. The neutrophils were primarily annotated to obtain five categories, as illustrated in Fig. 2. E-G. The proportion of CSF3R neutrophils and LYZ neutrophils in the two groups exhibited a significant difference, with a notable decline in the disease group. Additionally, new MMP9 neutrophils, S100A9 neutrophils and CST3 neutrophils were identified. Subsequently, 167 differential genes were obtained from the differential gene expression analysis of neutrophils in order to prepare for the screening of prognosis-related genes.

Following differential analysis of 167 differential genes in neutrophils, 12 prognosis-related genes were identified through Cox one-way regression. Additionally, six characteristic genes in sepsis were identified through lasso regression feature selection algorithm. The correlation between the characteristic genes and sepsis was further confirmed using a comparison of the relative expression of the characteristic genes in animal models of sepsis and controls.The patients who had been processed with survival data in the GEO database were randomly divided into the training set and the test set in a ratio of 4:1. The best risk score value corresponding to each sample was obtained after analysis by lasso regression and divided into high-risk and low-risk groups on the basis of this. The survival curves of both the training and test sets demonstrated that the OS of the low-risk group was significantly higher than that of the high-risk group. Furthermore, the results of the ROC curves of the training and test sets (with AUC values ranging from 0.7 to 0.77) indicated that they had a good predictive effect on prognosis. In addition, the risk score model was correlated with clinical indicators. The results of the rank sum test indicated that the distribution of risk scores differed significantly between groups in the clinical indicator of survival status ($p < 0.05$). This suggests that the risk scores obtained from the modelling analysis are applicable for the typing of sepsis patient samples.

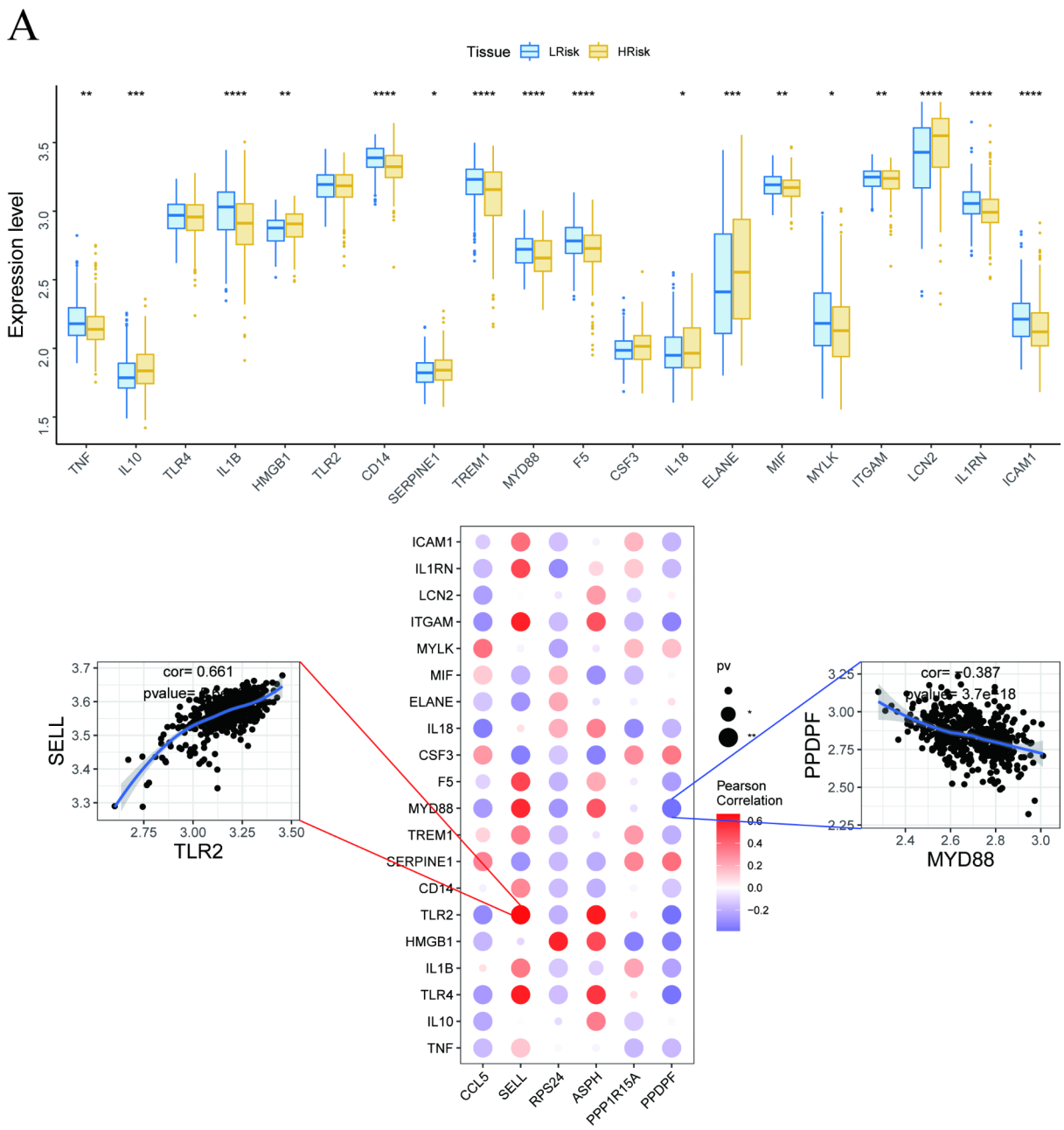

Fig. 9. Exploring the correlation between key genes and disease progression genes. (**A**) The upper panel shows the expression differences of disease-regulated genes, with blue indicating low-risk patients and yellow indicating high-risk patients. The lower panel shows the pearson correlation analysis between key genes and disease genes, blue indicates negative correlation and red indicates positive correlation.

Immune infiltration analysis is performed to clarify the composition of immune cells within the human microenvironment and thus which immune cells play an important role in the development of the disease. Although the pathological mechanisms of sepsis are complex and not fully understood, it is clear that immune cells and the immune microenvironment play an important role in the development of sepsis[54]. During the onset and development of sepsis, the immune status of the organism is not static; it may start in an immunologically activated state, enter an immunosuppressed state as the disease progresses, or remain in an immunologically disturbed state from the beginning to the end of the disease[55].We found that the proportion of immune cells in the high-risk group and the low-risk group differed significantly, such as monocytes and T cells CD8 were significantly lower in the high-risk group; macrophages M0, resting mast cells, resting NK cells, T cells CD4 memory activated, T cells CD4 naive, etc. were significantly increased. There was also a significant difference in the correlation between risk scores and immune cells, e.g. risk scores were significantly positively correlated

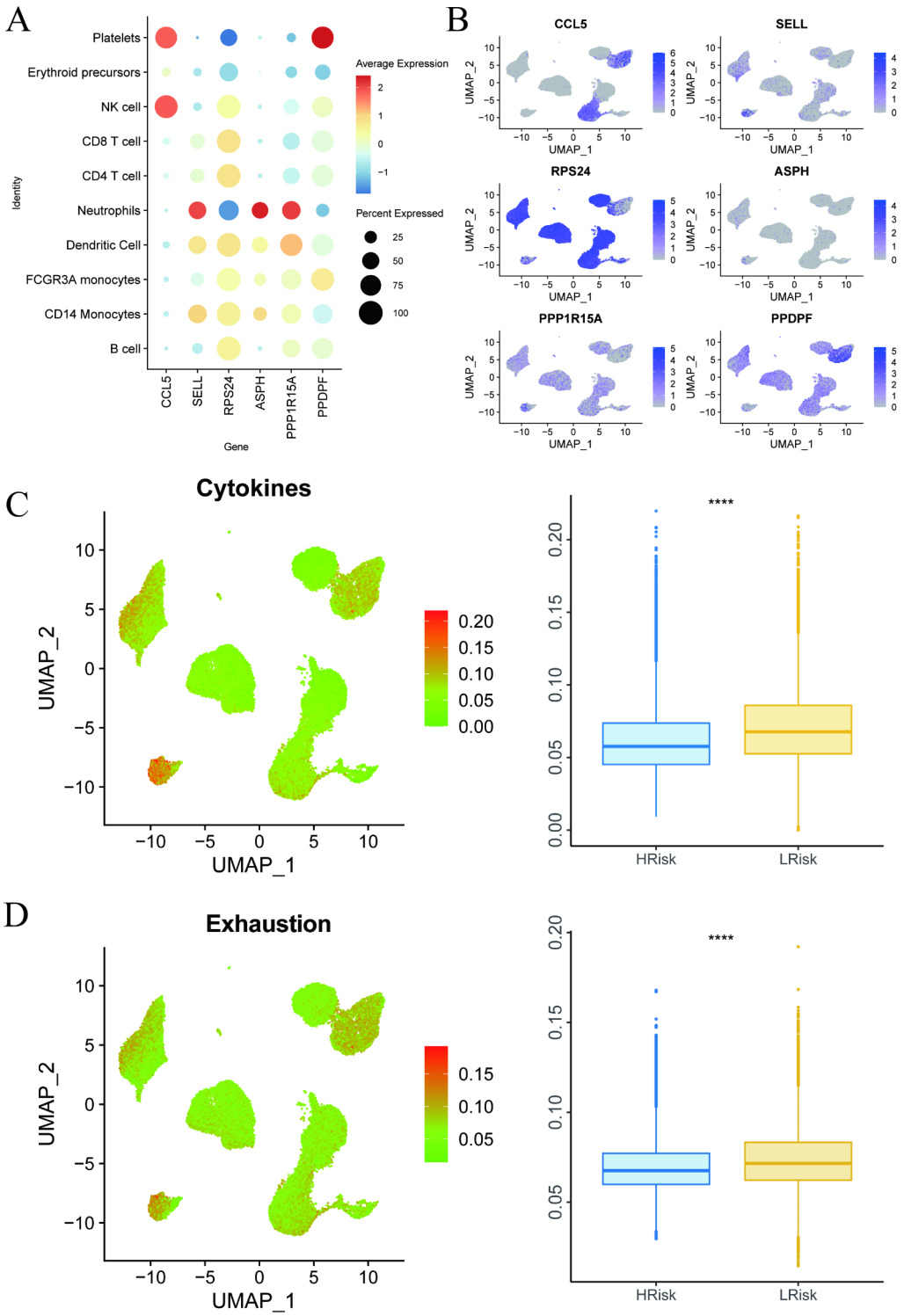

Fig. 10. Expression profiles in single cells. (**A**,**B**) Expression profiles of key genes in single cells. (**C**) Expression profile of cytokines in single cells. (**D**) Expression profile of failure factors in single cells.

with eosinophils, macrophages M0, macrophages M2, NK cells resting, T cells CD4 naive, dendritic cells resting, mast cells resting and neutrophils were significantly negatively correlated. The results of the analysis showed that immune cells play an important role in sepsis. To further illustrate the role of neutrophils in sepsis, we analysed the pathways in the high-risk and low-risk sepsis groups by GSEA, and the results showed that the pathways involved in sepsis were mainly in three directions: Ferroptosis, Nitrogen metabolism and Phagosome, and we further investigated the metabolic pathways, in which amino acid metabolism-related signalling pathways were

more active in the high-risk group.Qi Chen et al. 2022 recruited 63 sepsis patients and 43 normal controls from a prospective multicentre cohort and constructed a molecular network based on metabolomics and proteomics data to study key molecules, depicting the key immune metabolic pathways.Amino acid-related pathways, including phenylalanine metabolism, tyrosine metabolism and tryptophan biosynthesis, were elucidated as important mechanisms in sepsis[56]. It is further demonstrated that amino acid metabolism-related pathways play a key role in the development of sepsis.

To screen for independent prognostic factors, we performed univariate and multivariate Cox analyses of clinical characteristics and risk scores, and found that the risk score was an independent prognostic factor for patients with sepsis. The risk score was included in the column-line graphical model together with the clinical characteristics, and the calibration curve showed that the model had a high predictive power. The results therefore suggest that the risk score is an independent prognostic factor and that the column line graph has a high predictive effect for OS in sepsis patients at both 14 and 28 days.

To investigate whether the risk score is associated with sepsis-related genes, we used the GeneCards database to obtain sepsis-related genes and found that 17 of the top 20 genes were significantly correlated with the risk score, among which SELL was significantly and positively correlated with TLR2 (cor = 0.661) and PPDPF was significantly and negatively correlated with MYD88 (cor=-0.387), indicating that the risk score has a good correlation with sepsis-related genes. Similarly, the risk score was highly correlated with the failure factor score and the cytokine score.

Conclusions

By integrating single-cell sequencing and transcriptome sequencing data, we conducted multiple bioinformatics analyses and developed a new predictive prognostic model for OS in patients with sepsis. This model can be used to predict the probability of survival in patients with sepsis. Additionally, the risk score is a potential independent prognostic factor that is closely related to the immune microenvironment and clinical characteristics, among other factors. In conclusion, the present study can be used as a reliable predictor of sepsis outcome, thereby opening up new avenues for targeted treatment of sepsis in the future.

Data availability

Sequence data that support the findings of this study have been deposited in the Gene Expression Omnibus database with the primary accession code GSE167363， GSE65682 and GPL13667.

Received: 15 August 2024; Accepted: 21 November 2024
Published online: 02 December 2024

References

1. Singer, M. et al. The Third International Consensus definitions for Sepsis and septic shock (Sepsis-3). *JAMA* **315**(8), 801–810. https://doi.org/10.1001/jama.2016.0287 (2016).
2. Lingling, S. H. I. et al. Progress in the study of pathophysiological mechanisms of sepsis[J]. *Chin. J. Hosp. Infect. 2016* **26**(08):1914–1916.
3. Levy, M. M. et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med. ;31(4):1250–1256. doi: https://doi.org/10.1097/01.CCM.0000050454.01978.3B (2003).
4. Schürholz, T. Sepsis und Multiorganversagen - Therapeutische Möglichkeiten [Sepsis and multiple organfailure–potential therapeutic interventions]. *Anasthesiol. Intensivmed. Notfallmed. Schmerzther* **50**(2), 132–140. https://doi.org/10.1055/s-0041-100302 (2015).
5. Qing, Z. H. U. et al. Progress in the study of pathophysiological mechanisms of sepsis[J]. *Chin. J. Hosp. Infect. 2022* **32**(16):2551–2555.
6. Prescott, H. C., Angus, D. C. Postsepsis Morbidity. *JAMA* **319**(1):91. doi:https://doi.org/10.1001/jama.2017.19809 (2018).
7. Cao, Y. et al. Chinese guidelines for emergency treatment of sepsis/septic shock. *J. Clin. Emerg. Med.* **19**(09):567–588. https://doi.org/10.13201/j.issn.1009-5918.2018.09.001(2018)
8. Liu, Y. C. et al. Frequency and mortality of sepsis and septic shock in China: a systematic review and meta-analysis. BMC Infect Dis. 22(1):564. doi: https://doi.org/10.1186/s12879-022-07543-8 (2022).
9. GUO Wei. Treatment options and prognosis of sepsis[J]. *China Community Physician (Medical Speciality)* **12**(12), 4–5 (2010).
10. Zhang, H., Wang, C. & Yang, N. Diagnostic performance of machine-learning algorithms for sepsis prediction: an updated meta-analysis. *Technol. Health Care Published Online June* **22**. https://doi.org/10.3233/THC-240087 (2024).
11. Sun, D. et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat. Biotechnol.* **40**(4), 527–538. https://doi.org/10.1038/s41587-021-01091-3 (2022).
12. Zhang, Q. et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* **179**, 829–845 (2019).
13. Yofe, I., Dahan, R. & Amit, I. Single-cell genomic approaches for developing the next generation of immunotherapies. *Nat. Med.* **26**, 171–177 (2020).
14. Wen, L. & Tang, F. Recent advances in single-cell sequencing technologies. *Precis Clin. Med.* **5**(1), pbac002. https://doi.org/10.1093/pcmedi/pbac002 (2022).
15. Tang, F. et al. mRNA-Seq wholetranscriptome analysis of a single cell. *Nat. Methods* **6**, 377–382. https://doi.org/10.1038/nmeth.1315 (2009).
16. Proserpio, V., Duval, C., Falvo, V., Donati, G. & Oliviero, S. Single-cell sequencing for everybody. *Methods Mol. Biol.* **2421**, 217–229. https://doi.org/10.1007/978-1-0716-1944-5_15 (2022).
17. Qiu, X. et al. Dynamic changes in human single-cell transcriptional signatures during fatal sepsis. *J. Leukoc. Biol.* **110**(6), 1253–1268 (2021).
18. Scicluna, B. P. et al. A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *Am. J. Respir Crit. Care Med.* **192**(7), 826–835 (2015).
19. Scicluna, B. P. et al. Molecular Biomarker to assist in diagnosing Abdominal Sepsis upon ICU admission. *Am. J. Respir Crit. Care Med.* **197**(8), 1070–1073 (2018).
20. Scicluna, B. P. et al. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir Med.* **5**(10), 816–826 (2017).
21. van Vught, L. A. et al. Comparative analysis of the Host Response to community-acquired and hospital-acquired Pneumonia in critically ill patients. *Am. J. Respir Crit. Care Med.* **194** (11), 1366–1374 (2016).

22. van Vught, L. A. et al. Association of diabetes and diabetes treatment with the host response in critically ill sepsis patients. *Crit. Care* **20**(1), 252 (2016).
23. Claushuis, T. A. et al. Thrombocytopenia is associated with a dysregulated host response in critically ill sepsis patients. *Blood* **127**(24), 3062–3072 (2016).
24. Cheng, S. C. et al. Broad defects in the energy metabolism of leukocytes underlie immunoparalysis in sepsis. *Nat. Immunol.* **17**(4), 406–413 (2016). PMID: 26950237.
25. Bos, L. D. J. et al. Understanding heterogeneity in Biologic phenotypes of Acute Respiratory Distress Syndrome by Leukocyte expression profiles. *Am. J. Respir Crit. Care Med.* **200** (1), 42–50 (2019).
26. Wiewel, M. A. et al. The host response in critically ill sepsis patients on statin therapy: a prospective observational study. *Ann. Intensive Care* **8**(1), 9 (2018).
27. Uhel, F. et al. Matrix metalloproteinase-8: a useful biomarker to refine the diagnosis of community-acquired pneumonia upon intensive care unit admission? *Crit. Care* **23**(1), 226 (2019).
28. Huson, M. A. et al. The impact of HIV Co-infection on the genomic response to Sepsis. *PLoS One*. **11**(2), e0148955 (2016).
29. Uhel, F. et al. Mortality and host response aberrations associated with transient and persistent acute kidney injury in critically ill patients with sepsis: a prospective cohort study. *Intensive Care Med.* **46**(8), 1576–1589 (2020).
30. Leite, G. G. F. et al. Combined Transcriptome and Proteome Leukocyte's Profiling Reveals Up-Regulated Module of Genes/Proteins Related to Low Density Neutrophils and Impaired Transcription and Translation Processes in Clinical Sepsis. Front Immunol. 12:744799. (2021).
31. Peters-Sengers, H. et al. Source-specific host response and outcomes in critically ill patients with sepsis: a prospective cohort study. *Intensive Care Med.* **48**(1), 92–102 (2022).
32. Michels, E. H. A. et al. Association between age and the host response in critically ill patients with sepsis. *Crit. Care* **26**(1), 385 (2022).
33. van Amstel, R. B. E. et al. Uncovering heterogeneity in sepsis: a comparative analysis of subphenotypes. *Intensive Care Med.* **49**(11), 1360–1369 (2023).
34. Integrated analysis of multimodal single-cell data. *Cell* **184**(13):3573–3587e29. doi: https://doi.org/10.1016/j.cell.2021.04.048. (2021).
35. Friedman, J., Tibshirani, R. & Hastie, T. Regularization paths for generalized Linear models via Coordinate Descent. _Journal Stat. Software_ *33* (1), 1–22. https://doi.org/10.18637/jss.v033.i01 (2010).
36. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**(5), 453–457. https://doi.org/10.1038/nmeth.3337 (2015).
37. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* **1711**:243–259. doi: https://doi.org/10.1007/978-1-4939-7493-1_12 (2018).
38. Keller, A., Backes, C. & Lenhof, H. P. Computation of significance scores of unweighted gene set Enrichment analyses. *BMC Bioinform.* 8;290. (2007).
39. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
40. Oh, S. J., Ahn, J. Y. & Chung, D. H. Comparison of invariant NKT Cells with conventional T cells by using Gene Set Enrichment Analysis (GSEA). *Immune Netw.* **11**(6), 406–411. https://doi.org/10.4110/in.2011.11.6.406 (2011).
41. clusterProfiler. An R package for comparing biological themes among gene clusters. *OMICS* **16**(5), 284–287. https://doi.org/10.1089/omi.2011.0118 (2012).
42. rms. Regression Modeling Strategies_. R package version 6.6-0. https://CRAN.R-project.org/package=rms
43. Park, S. Y. Nomogram: An analogue tool to deliver digital knowledge. *J. Thorac. Cardiovasc. Surg.* **155**(4), 1793. https://doi.org/10.1016/j.jtcvs.2017.12.107 (2018).
44. DoubletFinder Doublet Detection in single-cell RNA sequencing data using Artificial Nearest neighbors. *Cell. Syst.* **8**(4), 329–337e4. https://doi.org/10.1016/j.cels.2019.03.003 (2019).
45. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
46. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
47. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
48. Wei, J. I. A. N. G. & Bin, D. U. Current status of sepsis epidemiology in China[J]. *J. Graduate Med.* **32**(01):5–8 .https://doi.org/10.16571/j.cnki.1008-8199.2019.01.002 (2019).
49. Angus, D. C. et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associ ated costs of care. *Crit. Care Med.* **29**(7), 1303–1310 (2001).
50. Esteban, A. et al. Sepsis incidence and outcome: contrasting the intensive care unit with the hospital ward. *Crit. Care Med.* **35**(5), 1284–1289 (2007).
51. Zhou, J. et al. Population-based epidemiology of sepsis in a subdistrict of Beijing. *Crit. Care Med.* **45**(7), 1168–1176 (2017).
52. Chiu, C. & Legrand, M. Epidemiology of sepsis and septic shock. *Curr. Opin. Anaesthesiol.* **34**(2), 71–76. https://doi.org/10.1097/ACO.0000000000000958 (2021).
53. Feng, W. E. I. et al. Analysis of epidemiological characteristics and prognostic factors influencing severe sepsis in ICU[J]. *Chin. J. Hosp. Infect.* **28**(10):1469–1471 (2018).
54. van der Poll, T., van de Veerdonk, F. L., Scicluna, B. P. & Netea, M. G. The immunopathology of sepsis and potential therapeutic targets. *Nat. Rev. Immunol.* **17**(7), 407–420. https://doi.org/10.1038/nri.2017.36 (2017).
55. Yueqing, D. O. N. G. & YAO Yongming. Mechanisms of cellular immune disorders in sepsis[J]. *China Emerg. Med. Crit. Illn.* **2004**(10):636–638.
56. Chen, Q. et al. Integrative analysis of metabolomics and proteomics reveals amino acid metabolism disorder in sepsis [published correction appears in J Transl Med. ;20(1):366. doi: https://doi.org/10.1186/s12967-022-03548-8]. J Transl Med. 2022;20(1):123. Published 2022 Mar14. doi:10.1186/s12967-022-03320-y. (2022).

Author contributions

H.Z. wrote the main manuscript text and prepared Tables 1, and 2 and figures. S.C. and Y.W. prepared Figs. 1, 2, 3, 4, 5, 6, 7, 8 and 9. R.L. and M.Z. are in charge of surveys and data collection. Q.C. and H.Z. are in charge of the animal experiments. Y.S. is responsible for obtaining funding, modifying ideas and overseeing. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-80791-7.

Correspondence and requests for materials should be addressed to Y.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.