Machine Learning Model for Early Prediction of Sepsis Outcomes Based on Immune Response

Huiyin Deng
Third Xiangya Hospital
Jiuyi Li
Chenzhou First People's Hospital
Abid Shah
Central South University
Lingkun Huang
Chenzhou First People's Hospital
Lite Ge
Second Xiangya Hospital of Central South University
Wen Ouyang (ouyangwen133@vip.sina.com)
Third Xiangya Hospital

Research Article

Keywords: Sepsis, survival, immune-related genes, supervised machine learning

Posted Date: May 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1663047/v1

Abstract

Background: Host immune dysregulation participates in the prognosis of sepsis with high morbidity and mortality. The contribution of sepsis to alive or dead, and the early immunologic signature to which they are preventable, is unknown. Therefore, knowing the immunogenomic landscape in blood samples is of paramount importance. This study develops a machine learning model to learn signature IRGs associated with the dysregulation of the host immune in sepsis and to predict sepsis survival up to 24 h at diagnosis that may be useful for planning individualized therapies in the future.

Methods: A total of 142 sepsis patients with corresponding clinical information were retrieved and analyzed from January 1, 2022, to March 31, 2022, as a secondary analysis of public data. The variables used for analysis included demographic characteristics, clinical conditions, and the differentially expressed immune-related genes (IRGs). The machine learning methods used included logistic regression, least absolute shrinkage, and selection operator (LASSO), random forest (RF). Prediction accuracy was randomly assigned to training, test, and validation sets. The performance of these models was evaluated using the area under the receiver operating characteristic curve (AUC). We also utilized a calibration curve to explain the prediction model.

Results: Our study cohort included 142 sepsis patients (mean [SD] age, 61.8 [15.8] years; 86 [60.6%] men; 104 [73.2%] survivor) downloaded from GEO database.The prognostic model based on IRGs and SOFA scores at diagnosis performed well in sepsis survival estimations (area under the curve, 0.842; 95% CI, 0.704-0.875). This model included a total of six survival-associated IRGs. Patients assigned to the high-risk group had worse survival than patients from the low-risk group (27 deaths [38%] vs 11 deaths [15.0%]; $P < 0.001$). The cell adhesion molecules (CAMs), chemokine signaling, and antigen processing and presentation pathways were the associated pathways for survival ($P < .001$).

Conclusions: This cohort study developed a prognostic modeling tool for sepsis survival based on IRG expression profiles, and has the capacity for early prediction of sepsis outcomes via monitoring the immunogenomic landscape, and also possibly the individualized therapies for sepsis survival.

Introduction

Sepsis, a host inflammatory response that occurs due to severe fatal infection accompanying organ dysfunction (1), is a major cause of morbidity and mortality in intensive care units (ICU) (2, 3). Despite significant strides made in the medical field over the last few decades, sepsis-related deaths in hospitals continue to be high, around 20–30% worldwide each year (4–6). Since there is currently no effective cure for sepsis, there is a need to accurately identify high-risk patients at early diagnosis to reduce the persistently high mortality rate associated with this condition.

In the sepsis microenvironment, the immune cells involved in innate and adaptive immunity play an essential role. Each immunocyte is unique and has unique features and characteristics. In this regard, immunotherapy relies on the immune capacity of immunocytes. An in-depth understanding of patients'

immune status, such as their immune-related gene expression and the number of immunocytes infiltrating their bodies, would help to determine the appropriate course of immune therapy for sepsis. A great deal of research has demonstrated that the immune status of septic patients determines their prognosis and even death. Different researchers have shown several biomarkers to characterize the immunological dysfunctions of sepsis (7–10). Major histocompatibility complex (MHC) II cell surface receptor (HLA-DR) whose reduced expression could lead it to be used as a surrogate marker for monocyte energy and decreased antigen presentation; that could play a role in sepsis-induced immunosuppression and may commence secondary infections ultimately leading to death (11–13). Similarly, immune-stimulatory agents such as interleukin-7 (IL-7) (14), granulocyte macrophage-colony stimulating factor (GM-CSF) (15), and interferon-gamma (IFN-γ) (16) have been shown to recapitulate the normal functions of immunocytes such as monocyte and lymphocytes thus allowing the reversal of immunosuppression. Lately, sepsis initially thought of as a “proinflammatory state” has been changed to an “anti-inflammatory state” indicating that these states were not temporal and separate but were rather simultaneously overlapping (17, 18). Unfortunately, significant lacunae exist in our understanding regarding the pathogenesis of sepsis and the foundering of various targeted treatments directed at modifying the altered host immune response in patients having sepsis (19). Therefore, there is a compelling need for early phase sepsis trials to focus on identifying those at risk for sepsis and initiating appropriate treatment. This would have a significant impact on overall mortality, which would ideally be a readily measured biomarker.

Of late, new next-generation sequencing technologies in the field of genomics and genetics have enabled the sequence of DNA and RNA of thousands of genes and the subsequent changes in the etiology of complex diseases. Through integrated analysis, we could explore more efficient and reliable markers of sepsis among the many prognostic biomarkers. In addition, machine learning in the field of medicine is a useful utility to decipher the complex interplay of various factors in immune dysregulation in septic patients (20, 21). Using the approach of RNA sequencing, previous studies have explored the transcriptomic profiles of whole blood in sepsis, still, no combined analysis of immune-related genes that could predict sepsis prognosis has been reported.

We carried out this study to explore the potential of immune-related genes (IRGs) as biomarkers associated with sepsis risk stratification. Using bioinformatics and a machine learning approach, differentially expressed genes (DEGs) were studied to confirm the independent prediction of prognosis. Furthermore, computational analyses were performed to delineate molecular mechanisms, gene expression regulation, and immune cell infiltration. A nomogram model was constructed for the prognosis of septic patients. Lastly, we evaluated the reliability of these immune gene signatures in mortality and survival groups of septic patients using quantitative PCR (qPCR). This study aimed to provide a comprehensive immunogenomic landscape of sepsis and identify IRGs as prospective biomarkers and intervention targets for sepsis immunotherapy.

Methods

Approval for the study was provided by the Ethics Review Committee of the Third Xiangya Hospital of Central South University (No. 2020-S373). Patients were informed about the study and a consent form was procured from all the patients or their legal guardians. The protocol of the genetic association study was according to “Strengthening the Reporting of Genetic Association Studies” (STREGA) reporting guidelines.

Clinical Samples and Data Extraction

We retrieved whole blood gene-expression data of sepsis samples from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). Overall, 142 septic patients from GSE54514, GSE57065, and GSE95233 have thorough clinical parameters (i.e, age, sex, SOFA, or APACHE II, mortality) were incorporated for a comprehensive analysis. Follow-up started on the first day after diagnoses, and RNA isolation of GSE57065 and GSE95233 were processed using Affymetrix Human Genome U133 Plus 2.0 Array (GPL570), while Illumina HumanHT-12 V3.0 expression bead chip (GPL6947) for GSE54514. The Immunology Database and Analysis Portal (IMMPORT) (22) database (https://www.immport.org/shared/genelists) provides IRGs for further analysis. RNA-seq data (GSE63042) and sepsis cohort from our hospital (Xiangya Third) were external validation sets. Sepsis definition was in line with “The Third International Consensus Definitions for Sepsis and Septic Shock” (1) and the clinical information of included study is shown in Additional file 1: eTable 1.

Risk Stratification Analysis

Initially, we investigated the differential expression of IRGs linked to sepsis survival. DEGs between sepsis patients and healthy controls were analyzed using the“limma” packages in R (23) with a $|\log_2$ (fold change)$| \geq 1.5$ and an adjusted p-value < 0.05 for GSE57065 and GSE95233, and an adjusted p-value < 0.05 for GSE54514 because of the small size of DEGs in this dataset, and differently expressed IRGs were obtained after matching the 2483 IRGs from IMMPORT. The results were displayed by Venn Diagram analysis (24) and bidirectional hierarchical clustering analysis using R software (25). Univariate logistic regression analysis was applied to each different expressed IRG, and significant IRGs associated with sepsis death with a p-value < 0.05 were chosen as candidate genes. These genes were introduced in the Least Absolute Shrinkage and Selection Operator (LASSO) regression and Random Forest Algorithm (RFA) to finish further selection using the glmnet package in R [26]. Then, the survival-associated IRGs were applied in the multivariate logistic regression to construct the IRGs score and visualized through package”glmnet”, and “forestplot” in R.

IRGs score=$\sum \beta i * RNAi$

For group patients, the median of the IRGs scores was set as the threshold. There were 71 patients in the high-risk group (< median) and 71 patients in the low-risk group (≥ median).

Functional Enrichment Analysis

By making use of the bioinformatics resource Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/) (26, 27), we tried to probe the potential molecular mechanisms of differential expressed IRGs by making use of Gene Ontology (GO) (28) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (29) for enrichment analysis. Use the "clusterProfiler" R package to visualize significant enrichment results with a p-value < 0.05 including biological process (BP), cellular component (CC), and molecular function (MF) components along with the KEGG pathways. PPI was used to recognize major inter-related protein partners and to get an idea regarding the structuring of proteins into particular functional units, through Metascape (parameters: physical core, min and max network size: 3, 500 respectively) database (https://metascape.org/). Similarly, PPI networks were clustered utilizing “The molecular complex detection” (MCODE) algorithm to find any potential protein complexes as subnetworks (30).

Prognostic Model Construction

Univariate and multivariate logistic regression was used to find clinical indicators that have prognostic value. Sepsis patients were randomly divided into training or a testing set according to a 7:3 ratio. Then, the Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values of sepsis survival outcomes were plotted for clinical indicators, IRGs profile, and IRGs score in the training, testing, and external validation (GSE63042) sets. Moreover, for clinical use of our model, the nomogram prognostic model was constructed using the “rms” package of R software with the performance assessed by ROC and calibration curve.

Immune Infiltration Analysis

CIBERSORT is an analysis platform, a gene expression-based deconvolution algorithm that uses gene expression signatures to estimate the immune composition of tissues (31). We used CIBERSORT to calculate and compare the cellular fraction of 22 immune infiltrations between both the high and low-risk groups. The association among 22 immune cell types and the IRGs scores were calculated with *P*-value < 0.05 as statistically significant.

Gene Set Enrichment Analysis

For enrichment analysis, GSEA (https://www.gsea-msigdb.org/gsea/index.jsp) was used to explore the mechanisms that lead to different outcomes between the high-risk and low-risk groups using the “clusterProfiler” package. KEGG sets were chosen as the reference sets for functional enrichment analysis. The adjusted p-value at < 0.05 was considered to be statistically enriched and visualized by the R “enrichplot” package.

Quantitative Real-Time PCR Analysis

Total RNA was isolated and purified using RNA Fast 200 kit following the manufacturer’s instructions (FASTAGEN). RNA obtained was reverse-transcribed into cDNA by TransScript All-in-One First-Strand cDNA Synthesis SuperMix for further downstream analysis (TransGen Biotech). We used SYBR Green (Vazyme Biotech) on a LightCycler 480 system (Roche Diagnostics) with primers shown in Additional file

1: eTable 2 for DNA quantification. The relative expression of all genes was normalized by the expression of 18S rRNA in each sample and calculated by the $2^{-\Delta\Delta CT}$ method (32).

Statistical Analysis

All statistical analyses were performed using R (www.r-project.org/) (version 4.0.2) and GraphPad Prism 8.00 software for plotting graphs (GraphPad Software Inc.). Using an ANOVA or Student's t-test to determine the differences between the groups for continuous data, defined as mean ± standard error of the mean (SEM); besides, other data was defined as the median by applying the Mann-Whitney U test. The Chi-square test was used to evaluate categorical data that are shown as numbers (N) and percentages (%). A P-value < 0.05 was considered statistically significant.

Results

Identification of Survival-Associated IRGs in Sepsis

According to the above criteria, RNA-seq and clinical data from 142 patients (mean [SD] age, 61.8 [15.8] years; 86 [60.6%] men; 104 [73.2%] survivor) were downloaded from GEO database (Additional file 1: eTable 3). Since improving patients' outcomes remain a significant challenge for sepsis management, we aim to find out survival-associated IRGs which could be possibly utilized as a bedside standard prognostic indicator. Analysis of the differentially expressed gene (DEGs) revealed 63 differentially expressed IRGs including 25 up-regulated and 38 down-regulated between sepsis and health controls (Additional file 1: eFigure 1, Additional file 1: eTable 4). Later, univariate logistic regression analysis was used to find 10 differentially expressed IRGs that were considerably associated with survival after sepsis (Fig. 1). Through further machine learning, we identified six IRGs as survival-related IRGs (Fig. 2).

IRGs Score–Based Patient Grouping

After multivariate logistic analysis of the survival-associated IRGs (Table 1), we calculated individualized IRGs scores with coefficient-weighted expression levels of six genes with the following:

IRGs score = expression level of HLA.DPA1∗0.5201 + IL18RAP∗0.5066 + MMP9 ∗0.4911 + RNASE3∗(− 0.5141) + S100P ∗(− 0.5965) + PTX3∗(− 0.1497).

Table 1
Multivariate logistic regression analysis for survival-associated IRG

Gene	Coefficient	HR (95% CI)	*P* value
HLA.DPA1	0.5201	2.15(1.38–3.35)	0.001
IL18RAP	0.5066	2.24(1.3–3.86)	0.004
MMP9	0.4911	1.63(1.08–2.47)	0.02
RNASE3	-0.5141	0.60 (0.39–0.92)	0.018
S100P	-0.5965	0.41(0.22–0.75)	0.004
PTX3	-0.1497	0.66(0.46–0.94)	0.022

A total of 71 patients with an IRGs score higher than or equal to the median (4.47) were classified as a low-risk group while the other was a high-risk group (Fig. 3a). A marked variation was observed in the survival state between the two groups (Fig. 3b). Patients in the high-risk group had worse survival state compared with the low-risk group (27 deaths [38%] vs 11 deaths [15.0%]; $P < 0.001$). Moreover, the two groups have significant differences in terms of survival-associated IRGs expression profiles (Fig. 3c).

Functional Annotation Analysis

Functional enrichment via GO and KEGG analysis can be used to explore the biological features of differentially expressed IRGs that might help us to understand the immunological dysfunction caused by sepsis. Biological processes substantially enriched in the up-regulated IRGs were “antibacterial humoral response”, “immune response”, “innate immune response”, “defense response to Gram-positive bacterium”, and “innate immune response in mucosa” (Additional file 1: eFigure 2a, Additional file 1: eTable5). In the same fashion, significantly enriched functions in the down-regulated IRGs were “T cell receptor signaling pathway”, “T cell costimulation”, “immune response”, “T cell activation”, and “positive regulation of T cell proliferation” respectively (Additional file 1: eFigure 2b, Additional file 1: eTable5).

Moreover, KEGG pathway analysis shown that up-regulated IRGs were significantly associated with cytokine-cytokine receptor interaction ($P = 7.56E-03$) and TNF signaling pathway ($P = 1.43E-02$). However, down-regulated IRGs involve the following pathways: T cell receptor signaling pathway ($P = 1.89E-15$), graft-versus-host disease ($P = 4.30E-11$), allograft rejection ($P = 1.02E-10$), type I diabetes mellitus ($P = 2.65E-10$), and autoimmune thyroid disease ($P = 1.28E-09$) (Additional file 1: eFigure 2c, Additional file 1: eTable 5).

PPI Network and MCODE Components Analysis

Here, we constructed a PPI network to explore the interactions and the potential molecular mechanisms among the obtained IRGs (Additional file 1: eFigure 3a). Then, MCODE components analysis exhibited the enrichment of differentially expressed up-regulated IRGs in neutrophil degranulation ($Log_{10}(P) = -18.0$)

(Additional file 1: eFigure 3b), while the down-regulated genes were significantly enriched in the generation of second messenger molecules ($Log_{10}(P)$ = -30.7), chemokine receptors bind chemokines ($Log_{10}(P)$ = -10.9), antigen processing and presentation of exogenous peptide antigen via MHC class II ($Log_{10}(P)$ = -9.1) (Additional file 1: eFigure 3c).

Establish Nomogram Prognosis Model

Analysis of IRGs scores established that patients in the survivor group had significantly higher IRG scores than patients from the non-survivor groups according to the training, testing, and validation set (Fig. 4a). Compared with the SOFA or APACHE II scores, IRGs predict sepsis survival more accurately (AUC: 0.76, 0.82, and 0.728) (Fig. 4b). The univariate and multivariate logistic analysis confirmed that patients in the low-risk group had much higher survival compared with the high-risk group, and were independent prognostic factors after adjustments were made for age, sex, and SOFA or APACHE II score (Additional file 1: eTable 6). To systematically validate the prognostic value of survival-associated IRGs, a nomogram was constructed by combining IRGs score with statistical independent clinical risk factors to predict the probability of death in patients with sepsis. In the nomogram plot, the line of IRGs scores has more length than SOFA, which indicated that the IRGs scores have a greater impact on the outcomes (Fig. 4c). Figure 4d shows that the nomogram prognosis model provided perfect prognostic performance (AUC, 0.842; 95% CI, 0.704–0.875). The calibration plots showed excellent agreement between the predicted probability of outcome and actual observation by the bias-corrected line only slightly deviated from the ideal line (Fig. 4e).

Immune Cells Infiltration Analysis

To investigate the differential composition of immune cells between sepsis groups (high-risk vs low-risk), CIBERSORT was applied to evaluate the infiltration of immune cells in sepsis. As shown in Additional file 1: eFigure 4 the intergroup proportions of the 22 immune cell types were similar while that of sepsis showed a difference. Visualization of the relative proportions of the 22 immune cell types between the two groups showed (Fig. 6a, Additional file 1: eFigure 5) four immune cell types: mast cells resting (P = 0.0034), B cells naive (P = 0.006), CD4 T cells naive (P = 0.0044) and macrophages M2 (P = 5.2e-05) were abundant in the low-IRGs group than the high IRGs group, whereas dendritic cells activated (P = 0.013), B cells memory (P = 6.9e-05), CD4 T cells memory resting (P = 0.031), eosinophils (P = 0.014) were in higher proportions in the high IRGs group. The two most common immune cell types in sepsis were neutrophils and monocytes, with P = 0.35 and P = 0.18 respectively. Moreover, only neutrophils, eosinophils, mast cells resting, and NK cells resting were negatively correlated with IRGs scores. Other immunocytes like B cell, dendritic cell, macrophage, CD4 T cell, CD8 T cell, monocyte, mast cells activated, and NK cells activated were all positively correlated with the IRGs score (Fig. 6b, Additional file 1: eFigure 6).

Gene Set Enrichment Analysis (GSEA)

To further gain an insight into the signal pathways causing different outcomes in sepsis, GSEA & KEGG analysis was performed on the DEGs between the high-risk and low-risk groups (Additional file 1: eTable

7). Results show Alzheimer's disease and cell cycle pathways most positively correlated with the high-risk group, while the low-risk group was most affected by cell adhesion molecules (CAMs), chemokine signaling, and antigen processing & presentation pathways. It suggests that IRGs promote sepsis survival through the cell adhesion molecules (CAMs), chemokine signaling, and antigen processing & presentation pathways that could prove to be a prospective remedial target for sepsis (Fig. 7, Additional file 1: eTable 8).

Survival-Associated IRGs Analysis

The overtime analysis shows that these survival-associated IRGs exhibited significantly different expressions at a very early time that lasted for 5 days post-infection (Fig. 7a). To validate the expression of six survival-associated IRGs in sepsis patients, we performed real-time qPCR on whole blood samples. As shown in Fig. 7b, IL18RAP and MMP9 expression levels in the survivor group are higher than in the control group ($P < 0.01$ and $P < 0.001$, respectively). Similarly, S100P, RNASE3, and PTX are significantly higher in the nonsurvivors group than in healthy controls ($P < 0.0001$, respectively). While HLA.DPA1 expression was significantly lower in the nonsurvivor group compared to the healthy controls ($P < 0.0001$). In addition, the differential expression of these survival-associated IRGs was also validated by a GEO external database (Fig. 7c).

Discussion

Sepsis is a common cause of death in ICU characterized by an abnormal host response (33). Patients with sepsis have a highly modulated blood transcriptome, resulting in an enhanced inflammatory response, as well as early and profound changes in innate and adaptive immunity (34, 35). As sepsis manifests in the activation of the innate immune system along with the suppression of both innate and adaptive immune systems, several biomarkers have been identified that are indicative of sepsis's immune dysfunction and contribute to the prognosis of the condition (35). However, no biomarker has sufficient specificity or sensitivity to provide a benchmark for its usage in clinical practice. Few tests are available that can discriminate sepsis from other inflammatory conditions or could provide the likely outcome. Thus, identifying early IRGs in blood samples is of paramount importance. This is to elucidate the molecular mechanisms by which sepsis progresses and to provide potential targets for early diagnosis and subsequent therapeutic development. In critically ill patients, the easy availability of whole blood provides a major advantage for monitoring immunological dysfunction.

Recently, methodological advances in the field of sequencing technologies have been instrumental in sequencing the RNA expression in thousands of genes in humans offering a more detailed overview of sepsis in blood transcriptomics. More and more studies begin to focus on using machine learning to better predict the clinical outcome of sepsis. It remains to be determined which genes are expressed for assessment of immune function in sepsis that can predict two extremes of clinical recovery (survivors versus nonsurvivors). To the best of our knowledge, this is the first sepsis prediction model built to explain each prediction and to jointly analyze DEGs in blood transcript with immune-related genes in the early course of sepsis. Thus, we identified 25 and 36 IRGs in different upregulated and downregulated

expressions. Screening criteria for the 25 and 36 IRGs comprised their differential expression in any two blood samples.

Our findings support the idea that the pro-and anti-inflammatory response can occur simultaneously at the onset of sepsis and contributes to death (18). GO and KEGG analysis showed that different up-regulated expressed IRGs are all pro-inflammatory in biological processes, including immune response, innate immune defense against Gram-positive bacteria, and innate immunity in the mucosa to activating cytokine-cytokine receptor interaction and TNF signaling pathway. The down-regulated IRGs are mainly related to promoting the immunosuppressive cellular program, including T cell receptors, costimulation of T cells, immune response, T cell activation, and positive regulation of T cell proliferation through downregulation of T-cell receptor signaling pathway, Graft-versus-host disease, allograft rejection, Type I diabetes mellitus, and autoimmune thyroid disease pathway. Similarly, PPI networking shows uniform findings suggesting that immune-related gene expression signatures can define different immune response states in sepsis (35). Our findings also validated the methods utilized by prior research studies for defining a gene expression signature that could envisage individual survival, which performed poorly in our sepsis cohort (36). Based on these, we suggest a multitude of factors responsible for sepsis mortality, and in contrast with prior research, an early gene expression signature can define an individual immune response and is consistent associates with a worse prognosis.

Also, the IRGs score on the expression of the six survival-associated IRGs assisted in the grouping of septic patients as survivors or non-survivors. The nomogram model of IRGs score in combination with SOFA or APACHE II performed well in predicting mortality (28-day) of sepsis, indicated by a high C-index value of 0.81, an acceptable calibration. Our study showed that six survival-associated IRGs were differentially expressed within the first 24 h after sepsis and undergoing 5 days post-infection, which are central to the prognostic of sepsis and can define individual sepsis immune state signatures; these data largely accord with previous reports. For example, Siegler et al.(37) found that decreased transcription of HLA-DPA1 can modulate monocyte activation during sepsis. Other studies reported the implication of MMP9 in sepsis and septic shock pathogenesis (38). In addition, it is a promising novel biomarker to predict the severity and outcome of sepsis (39, 40). Similarly, IL18RAP is a subunit of the heterodimeric receptor for interleukin 18 and is reported to drive NK cell activation to impair Treg activity (41). Lee and colleagues conducted a meta-analysis and found that higher levels of PTX-3 are observed in septic patients in non-survivors compared to survivors. They concluded that high PTX-3 is also a significant predictor of mortality. As a marker of sepsis severity and predictor of mortality outcomes (42, 43), human RNase3, is a member of the RNase superfamily involved in host immunity. RNase3 exhibits immune effects through independent modes in a macrophage-cell line infection model (44). In addition, S100 proteins are of interest as mediators of calcium-associated signal transduction that change subcellular distribution in response to extracellular stimuli. They also function as chemotactic agents and may play a role in the pathogenesis of the epidermal disease, including inflammation (45).

Sepsis immunotherapy relies on an understanding of the interactions between the host response and immune cells. However, owing to sepsis heterogeneity, it can be challenging to recognize associations

between immune cell infiltration and various clinicopathological factors. In the current study, we utilized the CIBERSORT database to study the relationship between immunocyte infiltration and IRGs score. Neutrophils and eosinophils infiltration were more abundant in the high IRGs score group and negatively associated with IRGs scores. Neutrophils are critical for the early control of invading pathogens (46). Various reports show the aberrant function of neutrophils preceding the advancement of nosocomial infections (47). Those individuals who have severely reduced neutrophil functions are at risk of getting nosocomial infections (47). The reduced neutrophil function also further leads to *Pseudomonas aeruginosa* secondary infection susceptibility according to the murine model manifesting in polymicrobial sepsis (48). Moreover, extracorporeal cell therapy with donor granulocytes was found to decrease various biomarkers of sepsis and improve sepsis severity in ten patients with septic shock (49). While over numbers will result in increased disease severity and even death (50). Like neutrophils, eosinophils perform a major role in sepsis prognosis, and their abnormal activity results in poor prognosis. However, it is unclear whether eosinophils are simply a marker of disease severity or a reflection of impaired type 2 immune responses. This is because it is unclear if they are necessary for cellular repair. So, gaining an insight into the mechanism through which eosinophils are reduced in sepsis is necessary. In our study, we identified that monocytes/macrophages and dendritic cells were infiltration in the high IRGs group, and are positively associated with our IRGs score. As we know that monocytes and macrophages in circulation have important roles against bacterial invasion that may reinstate the peripheral immune response to protect the host from infection morbidity (51). Similarly, dendritic cells are activated to induce adaptive immune responses for controlling infection that gives rise to sepsis survival (52, 53). Besides, B-lymphocytes, T-lymphocytes, and NK cells were significantly infiltration in the high IRGs group, and were positive with IRGs score, which is in accord with the reduced or dysfunction of these cells may promote immunoparalysis, which is one hallmark for sepsis survival (54). In addition, to delineate the potential molecular mechanism of survivor-associated IRGs' role, GSEA-KEGG analysis suggested that the possibility of survivor-associated IRGs regulating cell adhesion molecules (CAMs), chemokine signaling, and antigen processing and presentation pathways, all having a crucial role in mediating immune cells activation, thus acting as a vital factor in guarding host response, association with sepsis survival (53, 55, 56).

Limitations

There are some limitations to this study. At first, transcriptome analyses are not reflective of the overall immune state. Moreover, our study sample size was not large due to the need for complete clinical data. In the same way, other vital contributing factors such as various causes and co-morbidities were not taken into account in our study owing to the database's limited clinical information. Hence, further larger clinical studies of our findings are warranted in the future.

Conclusions And Future Directions

There exists heterogeneity in immunological states among septic patients thereby impacting clinical immunotherapy. Modulating the septic response is essential for improving survival in sepsis. In this

genetic association study, differently expressed IRGs were systematically investigated according to their roles, related pathways, interaction network, efficacy, and clinical utility. Future studies looking at survival-associated IRGs may identify high-risk patients of sepsis and potentially may be used as relevant clinical biomarkers and targeted therapies to modulate a dysfunctional host response.

Abbreviations

AUC: Area under the curve; APACHE II: Chronic Health Evaluation II; BP: Biological process; CC: Cellular component; DEGs: Differentially expressed genes; GSEA: Gene-set enrichment analysis; GO: Gene ontology; GEO: Gene Expression Omnibus; HLA.DPA1: Human leukocyte antigen-DPA1; IL18RAP: Interleukin 18 receptor accessory protein; IRGs: Immune-related genes; KEGG: Kyoto encyclopedia of genes and genomes; LASSO: Least Absolute Shrinkage and Selection Operator; MF: Molecular function; MCODE: Molecular complex detection; MODS: Multiple organ dysfunction syndromes; MMP9: Matrix metalloprotease 9; PPI: Protein-protein interaction; PTX3: Prototypic long pentraxin 3; PCR: Quantitative polymerase chain reaction; ROC: Receiver operating characteristic; RNASE3: Ribonuclease A family member 3; SOFA: Sequential organ failure assessment scores; S100P: S100 calcium-binding protein P;

Declarations

ETHICAL APPROVAL

Study permission was given by the Ethics Committee of the Third Xiangya Hospital of Central South University. Informed consent was acquired from all patients according to the tenets of the Declaration of Helsinki.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Design, analysis, and executing lab experiments: HYD, Data Analysis and drafting of manuscript: JYL, LKH, Gathering of data and revising the manuscript: AAS, Study design and drafting manuscript: GL, WOY. All authors agree to the final version of the draft.

ACKNOWLEDGMENTS

We thank all ICU staff and participants for their helpful contributions to this study.

FUNDING

National Natural Science Foundation of China (No. 81971028).

AVAILABILITY OF DATA AND MATERIALS

The dataset is available in the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The data of Xiangya Third has been approved for limited use by the Ethics Review Committee of the Third Xiangya Hospital of Central South University (No. 2020-S373) and has not been disclosed to the public.

References

1. Singer 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.
2. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive Care Med. 2013;39(2):165–228.
3. Deutschman CS, Tracey KJ. Sepsis: current dogma and new perspectives. Immunity. 2014;40(4):463–75.
4. Friedman G, Silva E, Vincent JL. Has the mortality of septic shock changed with time. Crit Care Med. 1998;26(12):2078–86.
5. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med. 2001;29(7):1303–10.
6. Kumar G, Kumar N, Taneja A, Kaleekal T, Tarima S, McGinley E, et al. Nationwide trends of severe sepsis in the 21st century (2000–2007). Chest. 2011;140(5):1223–31.
7. Gossez M, Rimmelé T, Andrieu T, Debord S, Bayle F, Malcus C, et al. Proof of concept study of mass cytometry in septic shock patients reveals novel immune alterations. Sci Rep. 2018;8(1):17296.
8. Conway Morris A, Datta D, Shankar-Hari M, Stephen J, Weir CJ, Rennie J, et al. Cell-surface signatures of immune dysfunction risk-stratify critically ill patients: INFECT study. Intensive Care Med. 2018;44(5):627–35.
9. Pierrakos C, Vincent JL. Sepsis biomarkers: a review. Crit Care. 2010;14(1):R15.
10. Reinhart K, Bauer M, Riedemann NC, Hartog CS. New approaches to sepsis: molecular diagnostics and biomarkers. Clin Microbiol Rev. 2012;25(4):609–34.
11. Monneret G, Finck ME, Venet F, Debard AL, Bohé J, Bienvenu J, et al. The anti-inflammatory response dominates after septic shock: association of low monocyte HLA-DR expression and high interleukin-10 concentration. Immunol Lett. 2004;95(2):193–8.
12. Lukaszewicz AC, Grienay M, Resche-Rigon M, Pirracchio R, Faivre V, Boval B, et al. Monocytic HLA-DR expression in intensive care patients: interest for prognosis and secondary infection prediction. Crit Care Med. 2009;37(10):2746–52.
13. Muenzer JT, Davis CG, Chang K, Schmidt RE, Dunne WM, Coopersmith CM, et al. Characterization and modulation of the immunosuppressive phase of sepsis. Infect Immun. 2010;78(4):1582–92.

14. Shakoory B, Carcillo JA, Chatham WW, Amdur RL, Zhao H, Dinarello CA, et al. Interleukin-1 Receptor Blockade Is Associated With Reduced Mortality in Sepsis Patients With Features of Macrophage Activation Syndrome: Reanalysis of a Prior Phase III Trial. Crit Care Med. 2016;44(2):275–81.
15. Meisel C, Schefold JC, Pschowski R, Baumann T, Hetzger K, Gregor J, et al. Granulocyte-macrophage colony-stimulating factor to reverse sepsis-associated immunosuppression: a double-blind, randomized, placebo-controlled multicenter trial. Am J Respir Crit Care Med. 2009;180(7):640–8.
16. Peters van Ton AM, Kox M, Abdo WF, Pickkers P. Precision Immunotherapy for Sepsis. Front Immunol. 2018;9:1926.
17. Delano MJ, Ward PA. The immune system's role in sepsis progression, resolution, and long-term outcome. Immunol Rev. 2016;274(1):330–53.
18. Denstaedt SJ, Singer BH, Standiford TJ. Sepsis and Nosocomial Infection: Patient Characteristics, Mechanisms, and Modulation. Front Immunol. 2018;9:2446.
19. Angus DC, Carlet J. Surviving intensive care: a report from the 2002 Brussels Roundtable. Intensive Care Med. 2003;29(3):368–77.
20. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. Jama. 2018;319(13):1317–8.
21. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719–31.
22. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, et al. ImmPort: disseminating data to the public for the future of immunology. Immunol Res. 2014;58(2–3):234–9.
23. 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.
24. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.
25. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13(8):1194–202.
26. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44–57.
27. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13.
28. Oliveira GS, Santos AR. Using the Gene Ontology tool to produce de novo protein-protein interaction networks with IS_A relationship. Genet Mol Res. 2016;15(4).
29. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353-d61.
30. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.
31. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol. 2018;1711:243–59.

32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402–8.
33. van der Poll T, van de Veerdonk FL, Scicluna BP, Netea MG. The immunopathology of sepsis and potential therapeutic targets. Nat Rev Immunol. 2017;17(7):407–20.
34. Plassais J, Venet F, Cazalis MA, Le Quang D, Pachot A, Monneret G, et al. Transcriptome modulation by hydrocortisone in severe burn shock: ancillary analysis of a prospective randomized trial. Crit Care. 2017;21(1):158.
35. Xiao W, Mindrinos MN, Seok J, Cuschieri J, Cuenca AG, Gao H, et al. A genomic storm in critically injured humans. J Exp Med. 2011;208(13):2581–90.
36. Cazalis MA, Friggeri A, Cavé L, Demaret J, Barbalat V, Cerrato E, et al. Decreased HLA-DR antigen-associated invariant chain (CD74) mRNA expression predicts mortality after septic shock. Crit Care. 2013;17(6):R287.
37. Siegler BH, Altvater M, Thon JN, Neuhaus C, Arens C, Uhle F, et al. Postoperative abdominal sepsis induces selective and persistent changes in CTCF binding within the MHC-II region of human monocytes. PLoS One. 2021;16(5):e0250818.
38. Vanlaere I, Libert C. Matrix metalloproteinases as drug targets in infections caused by gram-negative bacteria and in septic shock. Clin Microbiol Rev. 2009;22(2):224–39, Table of Contents.
39. Hoffmann U, Brueckmann M, Borggrefe M. Matrix metalloproteinases and their inhibitors: promising novel biomarkers in severe sepsis? Crit Care. 2009;13(6):1006.
40. Mühl D, Nagy B, Woth G, Falusi B, Bogár L, Weber G, et al. Dynamic changes of matrix metalloproteinases and their tissue inhibitors in severe sepsis. J Crit Care. 2011;26(6):550–5.
41. Dean JW, Peters LD, Fuhrman CA, Seay HR, Posgai AL, Stimpson SE, et al. Innate inflammation drives NK cell activation to impair Treg activity. J Autoimmun. 2020;108:102417.
42. Lee YT, Gong M, Chau A, Wong WT, Bazoukis G, Wong SH, et al. Pentraxin-3 as a marker of sepsis severity and predictor of mortality outcomes: A systematic review and meta-analysis. J Infect. 2018;76(1):1–10.
43. Albert Vega C, Mommert M, Boccard M, Rimmelé T, Venet F, Pachot A, et al. Source of Circulating Pentraxin 3 in Septic Shock Patients. Front Immunol. 2018;9:3048.
44. Lu L, Wei R, Prats-Ejarque G, Goetz M, Wang G, Torrent M, et al. Human RNase3 immune modulation by catalytic-dependent and independent modes in a macrophage-cell line infection model. Cell Mol Life Sci. 2021;78(6):2963–85.
45. Eckert RL, Broome AM, Ruse M, Robinson N, Ryan D, Lee K. S100 proteins in the epidermis. J Invest Dermatol. 2004;123(1):23–33.
46. Zhou J, Nefedova Y, Lei A, Gabrilovich D. Neutrophils and PMN-MDSC: Their biological role and interaction with stromal cells. Semin Immunol. 2018;35:19–28.
47. Stephan F, Yang K, Tankovic J, Soussy CJ, Dhonneur G, Duvaldestin P, et al. Impairment of polymorphonuclear neutrophil functions precedes nosocomial infections in critically ill patients. Crit

Care Med. 2002;30(2):315–22.
48. Rada B. Interactions between Neutrophils and Pseudomonas aeruginosa in Cystic Fibrosis. Pathogens. 2017;6(1).
49. Altrichter J, Sauer M, Kaftan K, Birken T, Gloger D, Gloger M, et al. Extracorporeal cell therapy of septic shock patients with donor granulocytes: a pilot study. Crit Care. 2011;15(2):R82.
50. 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.
51. Teh YC, Chooi MY, Liu D, Kwok I, Lai GC, Ayub Ow Yong L, et al. Transitional premonocytes emerge in the periphery for host defense against bacterial infections. Sci Adv. 2022;8(9):eabj4641.
52. Hotchkiss RS, Monneret G, Payen D. Immunosuppression in sepsis: a novel understanding of the disorder and a new therapeutic approach. Lancet Infect Dis. 2013;13(3):260–8.
53. Lu J, Sun K, Yang H, Fan D, Huang H, Hong Y, et al. Sepsis Inflammation Impairs the Generation of Functional Dendritic Cells by Targeting Their Progenitors. Front Immunol. 2021;12:732612.
54. Schrijver IT, Théroude C, Roger T. Myeloid-Derived Suppressor Cells in Sepsis. Front Immunol. 2019;10:327.
55. Wong MM, Fish EN. Chemokines: attractive mediators of the immune response. Semin Immunol. 2003;15(1):5–14.
56. Kaufmann SH, Schaible UE. Antigen presentation and recognition in bacterial infections. Curr Opin Immunol. 2005;17(1):79–87.

Figures

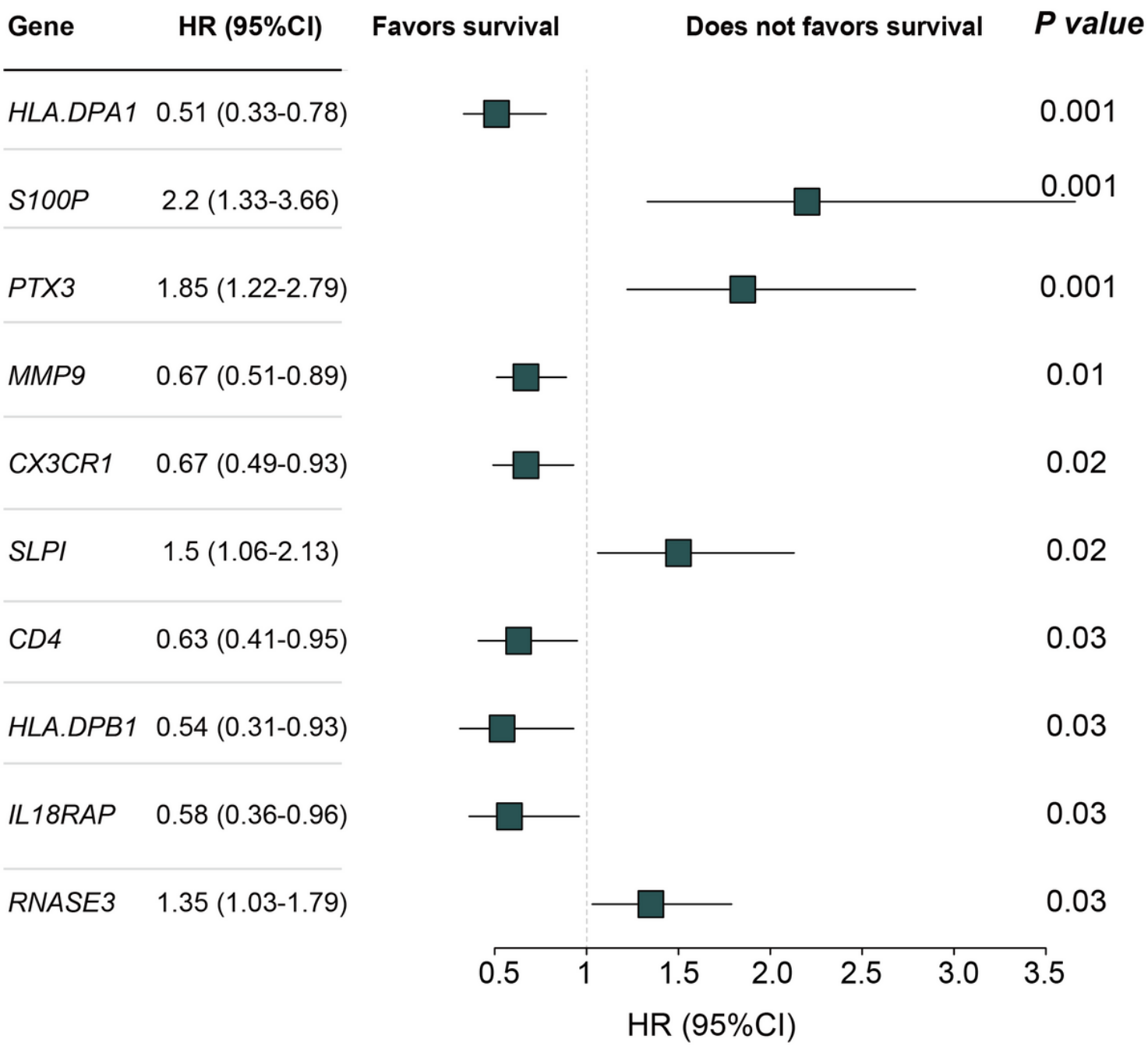

Figure 1

Univariate Logistic Regression Analysis of Differentially Expressed Immune-Related Genes. HR indicates hazard ratio.

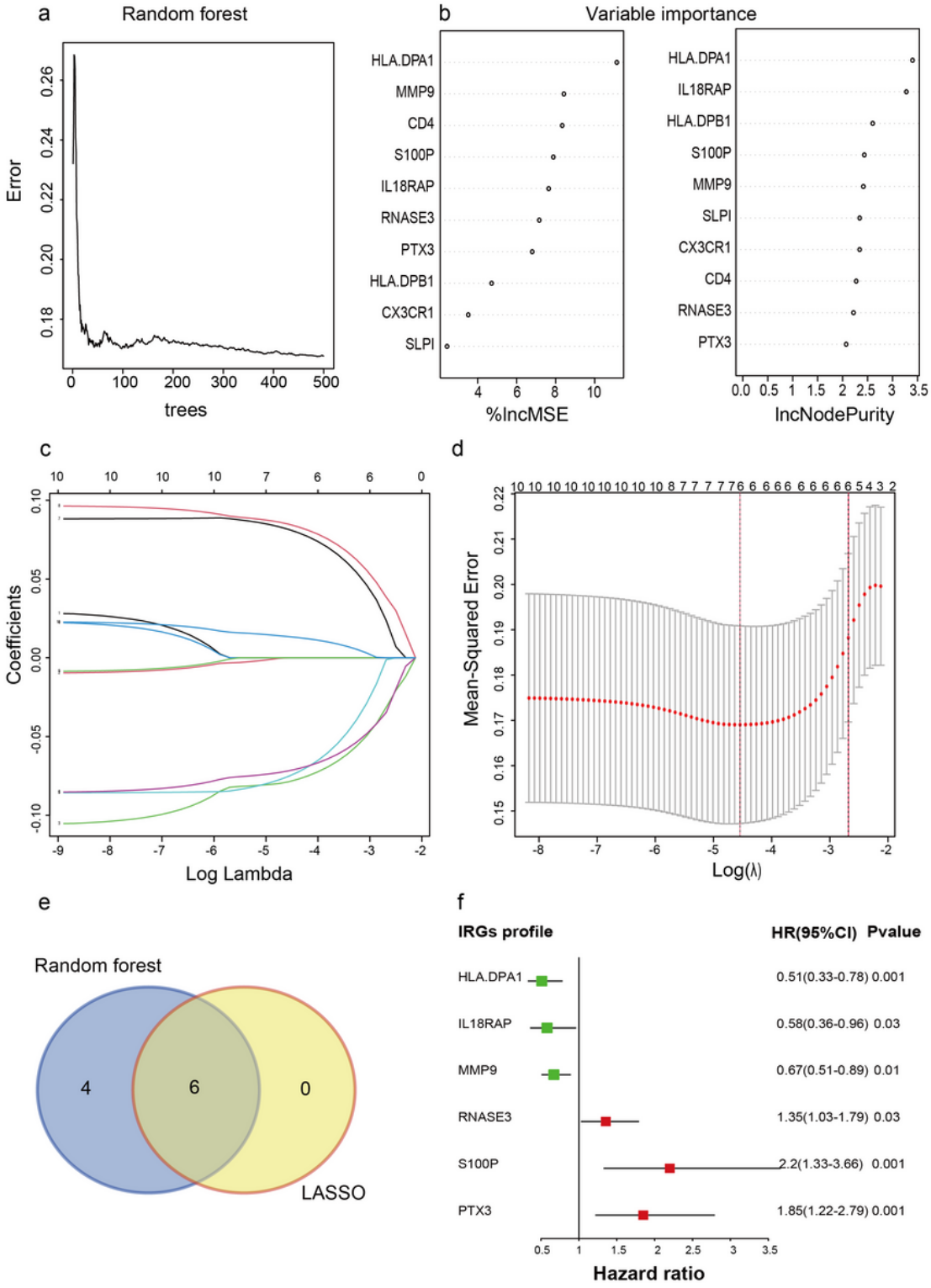

Figure 2

Survival-Associated Immune-Related Genes (IRGs) Selected From Machine Learning. a-b Random forest algorithm for survival-associated IRGs. c-d LASSO regression for survival-associated IRGs. e Venn diagram indicating that six genes were identified as survival-associated IRGs. f Univariate logistic regression analysis for sepsis survival. Green: protect factor; red: risk factor.

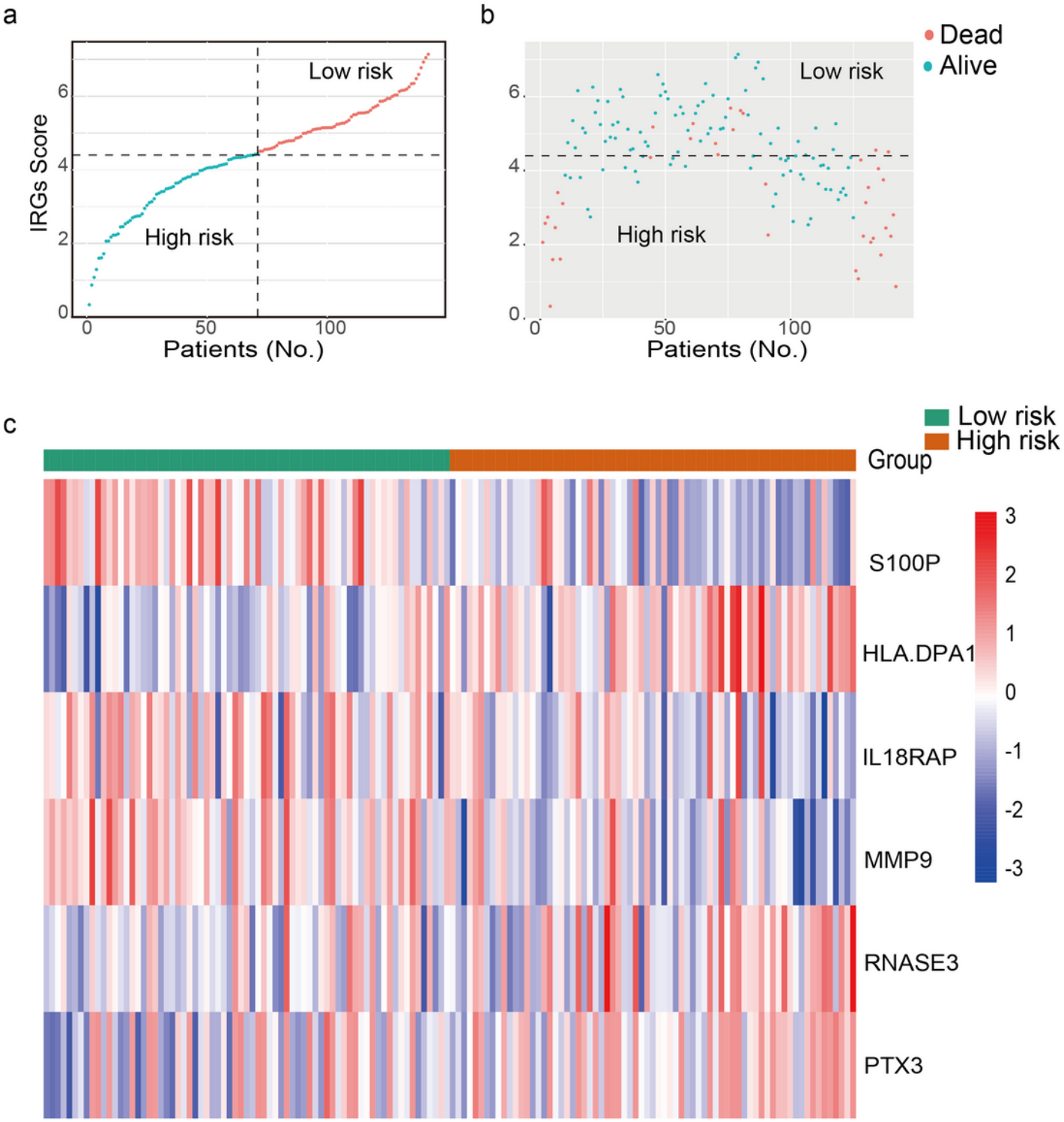

Figure 3

Clinical Outcomes Based on IRG's score. a The rank of IRGs score and distribution of groups. B Survival state in high- and low-risk groups. c Heatmap of expression profiles of the survival-associated IRGs.

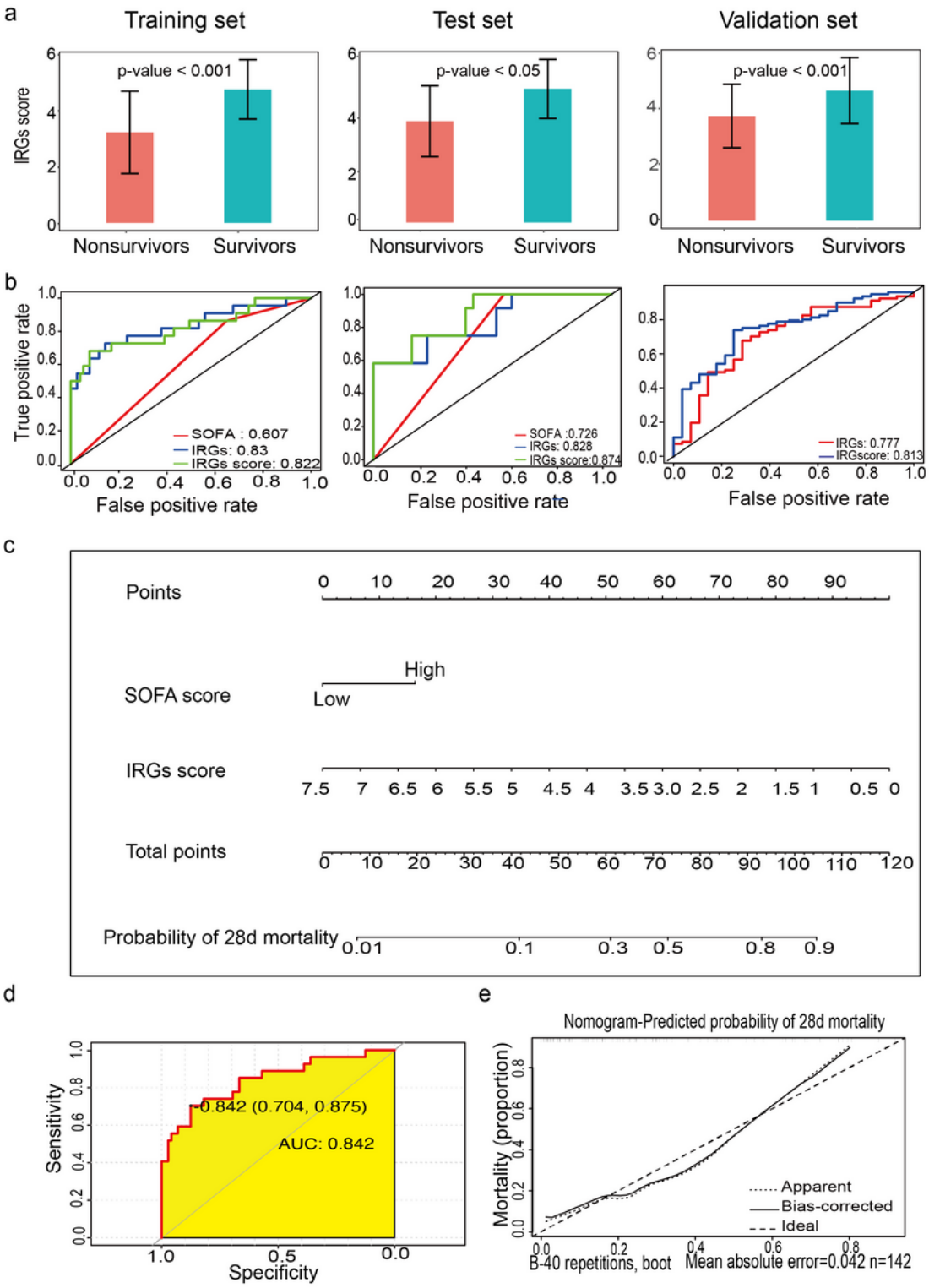

Figure 4

Estimation of Survival (1-Mortality) in Septic Patients. an IRGs score in survivor and nonsurvivor groups of sepsis. b The ROC curve for sepsis live. X-axis, false positive. Y-axis, true positive. **p < 0.01; ***p < 0.001; ****p < 0.0001. c Nomogram to predict the risk of mortality of septic patients. When using it, drawing a vertical line from each variable to the points axis for the score, then the points for all the parameters were added, and finally, a line from the total points axis was drawn to correspond to the

possibility of mortality at the bottom. d ROC curve for mortality. e Calibration curve prediction of mortality for internal validation

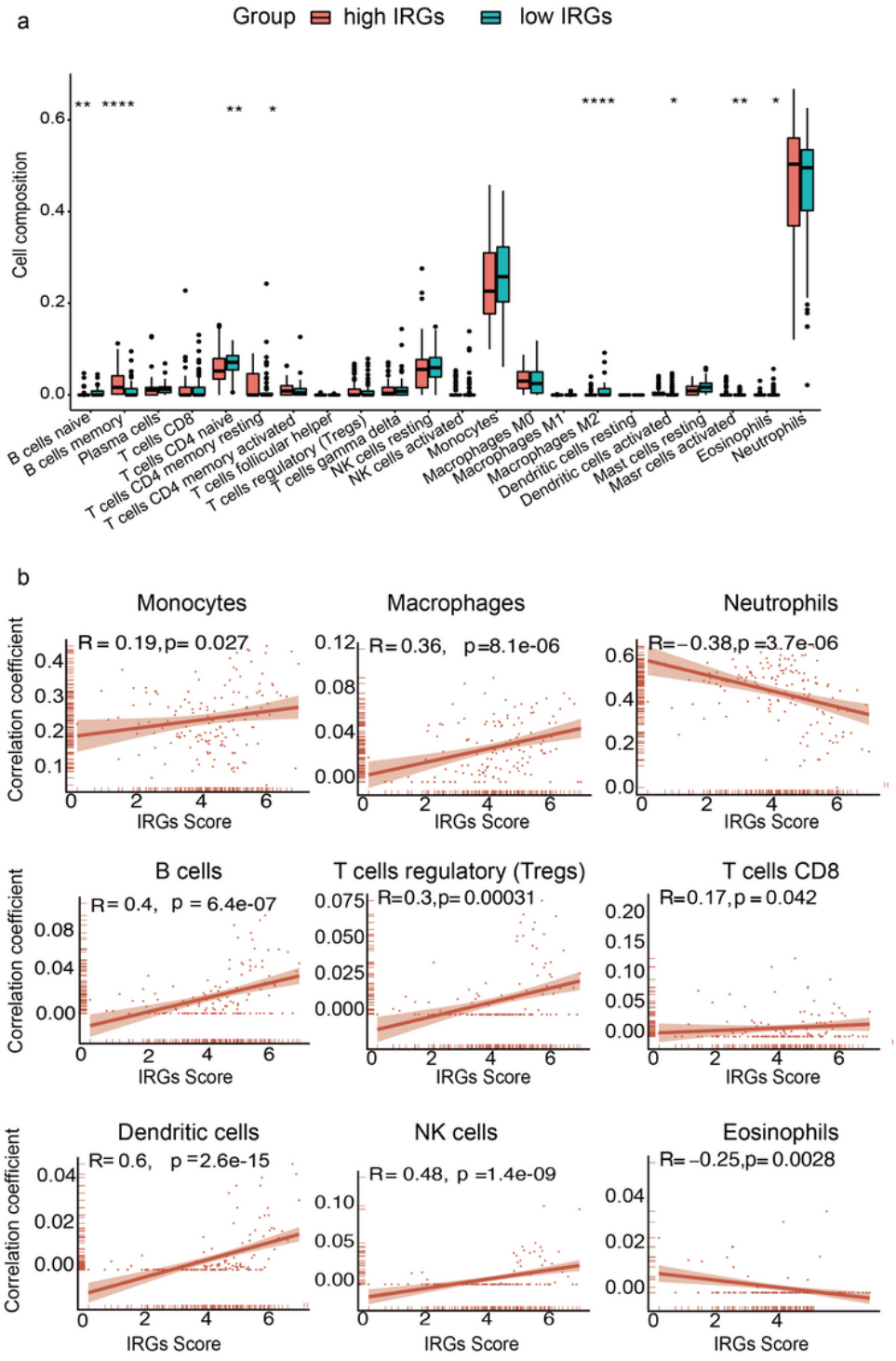

Figure 5

Immunocyte infiltration.

a The proportions of the 22 immune cell types between the high- and low- IRGs score groups. b The association between immune cell types and IRGs score. Dots indicate individual data points; line, trend; p-value < 0.05 are considered significant.

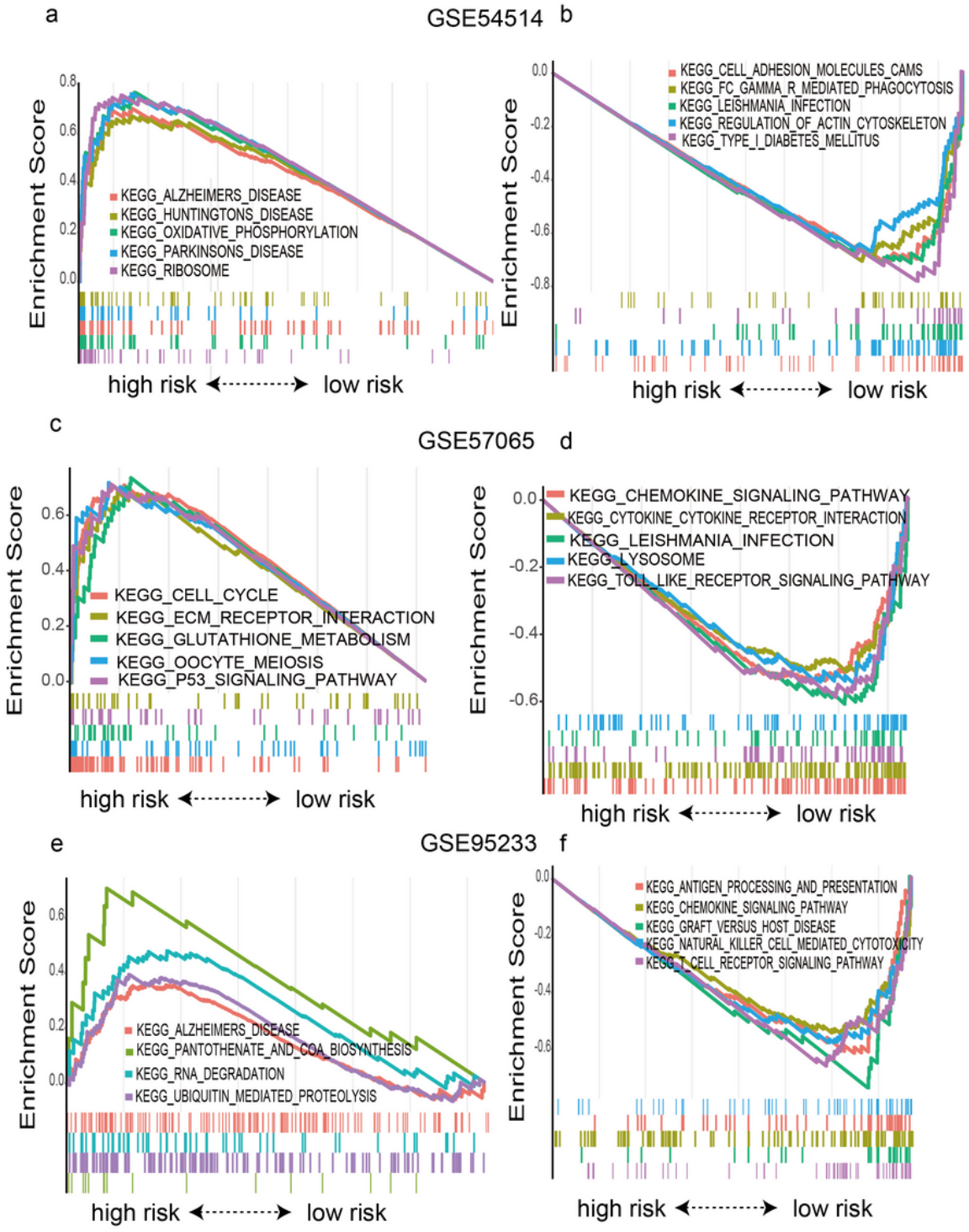

Figure 6

Gene Set Enrichment Analysis (GSEA) Between Groups. (a, c, e) GSEA pathways were enriched in samples with a high-risk group from GSE54514 (a), GSE57065 (c), and GSE95233 (e). (b,d,f) GSEA pathways were enriched in samples with low-risk groups from GSE54514 (b), GSE57065 (d), and GSE95233 (f). FDR, false discovery rate. Gene sets with p-value < 0.05 and FDR < 0.05 are considered as significant.

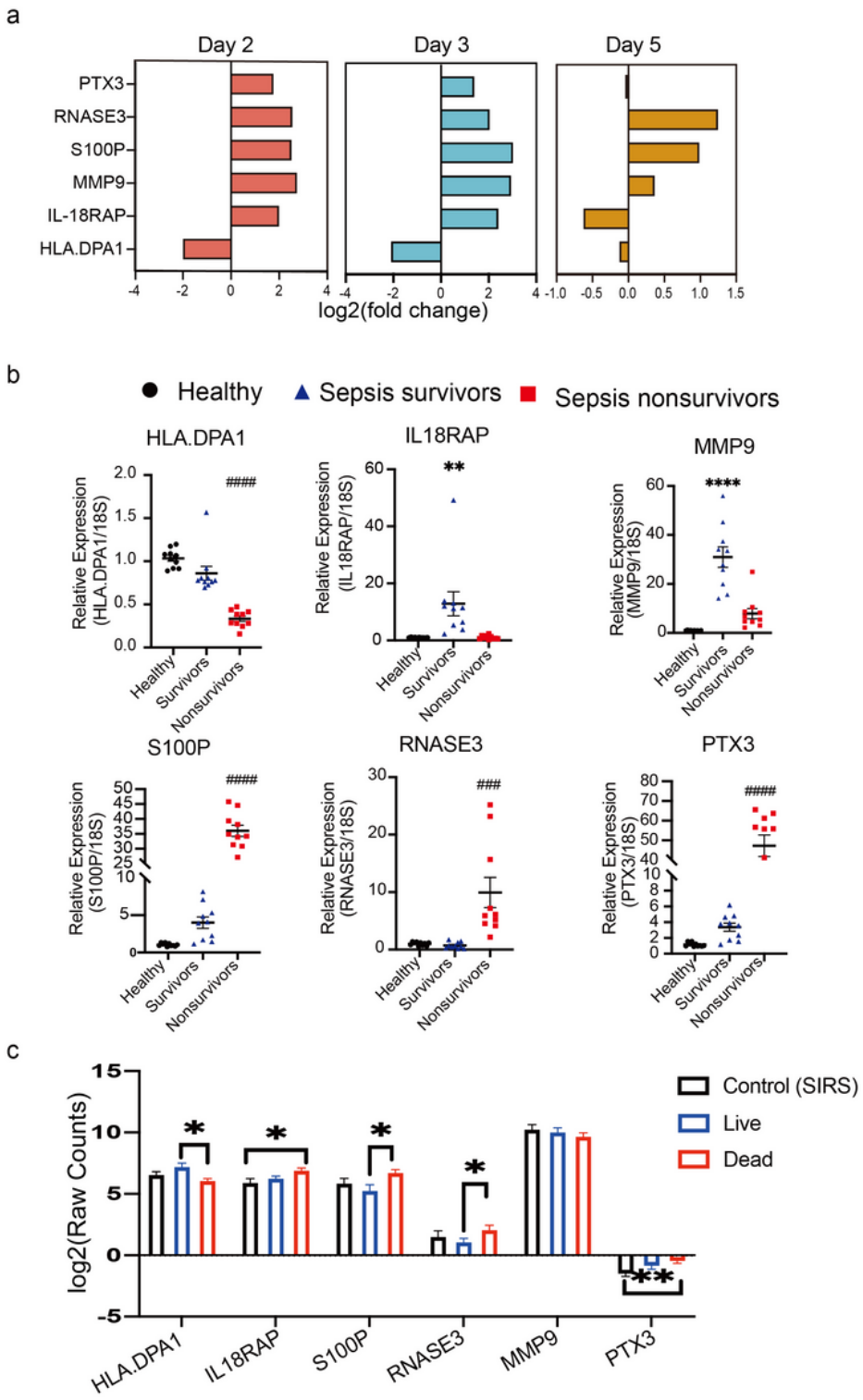

Figure 7

Survival-Associated IRGs Analysis in Septic Patiens**.** a The $\log_2$(FoldChange) of survival-associated IRGs between sepsis patients and healthy controls. b The relative expression of survival-associated IRGs comparison between sepsis survivor patients / sepsis nonsurvior patients and healthy controls, along with the expression of 18S rRNA as an internal standard. c The $\log_2$(Raw Counts) of survival-associated IRGs in a external dataset (GSE63042). The asterisks indicate a significant statistical p-value calculated by one-way ANOVA followed by the Tukey test (*p < 0.05; **p < 0.01; ***p < 0.001, and ****p < 0.0001 for comparing sepsis survivor patients with healthy controls; #p < 0.05; ##p < 0.01; ###p < 0.001, and ####p < 0.0001 for comparing sepsis nonsurvivor patients with healthy controls).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- eFigure1.png
- eFigure2.png
- eFigure3.png
- eFigure4.png
- eFigure5.png
- eFigure6.png
- eTable1.xlsx
- eTable2.xlsx
- eTable3.xlsx
- eTable4.xlsx
- eTable5.xlsx
- eTable6.xlsx
- eTable7.xlsx
- eTable8.xlsx
- Graphicabstract.jpeg