scientific reports

OPEN

Advancing sepsis diagnosis and immunotherapy machine learning-driven identification of stable molecular biomarkers and therapeutic targets

Weichuan Xiong[1,2], Yian Zhan[1,2], Rui Xiao[3]✉ & Fangpeng Liu[1,2]✉

Sepsis represents a significant global health challenge, necessitating early detection and effective treatment for improved outcomes. While traditional inflammatory markers facilitate the diagnosis of sepsis, the aspect of immune suppression remains poorly addressed. This study aimed to identify critical immune-related genes (IIRGs) associated with sepsis through genomic analysis and machine learning techniques, thereby enhancing diagnostic and treatment response predictions. Analyses of two extensive datasets were conducted, identifying significant immune genes using the ESTIMATE algorithm, Weighted Gene Correlation Network Analysis (WGCNA), and five machine learning methods. Prediction models were constructed and validated using six machine learning algorithms, achieving high accuracy (AUC > 0.75). Eleven key IIRGs were identified as active in immune pathways, such as the JAK-STAT signaling pathway, and were significantly correlated with immune cell infiltration in sepsis. Additionally, drug sensitivity analysis indicated that IIRGs correlated with responses to anticancer drugs. These results underscore the potential of these genes in enhancing sepsis diagnosis and treatment, highlighting the imperative for further validation across diverse populations.

Keywords Sepsis, Immune-related genes, Machine learning, Diagnostic framework, Therapeutic targets

Sepsis, a life-threatening condition triggered by a dysregulated immune response to infection, continues to pose a significan global health challenge due to its high prevalence and mortality rates. A 2020 Lancet study titled “Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study” reported approximately 48.9 million cases and 11 million related deaths worldwide in 2017[1]. These alarming figures significantly exceed previous estimates, underscoring the urgent need to address sepsis as a primary cause of death and a critical public health issue. Rapid and accurate diagnosis, coupled with effective treatment strategies, are crucial for enhancing patient survival.

Sepsis arises from a chaotic immune response involving both innate and adaptive systems, leading to excessive immune activation, widespread inflammation, and tissue damage that may result in organ failure. Current research focuses on the identification of biomarkers to enhance the early diagnosis and prognosis of sepsis[2–4]. Commonly investigated biomarkers include *C-reactive protein (CRP)*, *Procalcitonin (PCT)*, *Interleukin-6 (IL-6)*, *Tumor Necrosis Factor-α (TNF-α)*, *IL-1β*, *soluble Triggering Receptor Expressed on Myeloid cells-1 (sTREM-1)*, endothelial markers, and MicroRNA. Although these biomarkers show promise, they lack the accuracy required for dependable clinical use. Further research and refinement are necessary to develop markers that provide both sensitivity and specificity for sepsis diagnosis[4,5].

As medical technology evolves, machine learning (ML) models[6,7] and related technologies are becoming pivotal tools for disease detection and prediction. Jiang et al.[8] identified diagnostic genes and molecular mechanisms of Alzheimer’s disease using ML algorithms. Wang et al.[9] applied deep learning to predict the

[1]Jiangxi Provincial Key Laboratory of Respiratory Diseases, Jiangxi Institute of Respiratory Diseases, The Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, China. [2]China-Japan Friendship Jiangxi Hospital, National Regional Center for Respiratory Medicine, Nanchang 330200, Jiangxi, China. [3]The Department of Critical Care Medicine, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, China. ✉email: viefaciginmisp@mail.com; ndyfy08810@ncu.edu.cn

association between circRNA and diseases, while other researchers[10,11] have utilized artificial intelligence to identify tumor-related biomarkers. These data-driven algorithms process vast quantities of clinical and biological data, detecting complex patterns that may elude human experts[12]. ML models have proven to offer earlier and more accurate detection of sepsis than traditional methods, facilitating prompt interventions and potentially enhancing patient outcomes[13]. Notably, algorithms such as Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) have shown exceptional efficacy in predicting sepsis onset. One model, SepsisFinder, has demonstrated the ability to detect sepsis earlier than conventional models such as NEWS2 and GBDT under comparable sensitivity settings[14,15]. ML models are also being explored for their capability to predict sepsis-related mortality, with a systematic review and meta-analysis[16] confirming their significant potential in this domain. For instance, a focused meta-analysis in ICU settings revealed that RF and XGBoost were particularly effective in predicting sepsis outcomes. These findings emphasize the crucial role of ML in advancing sepsis diagnosis and treatment.

Our study aims to extend these developments by integrating diverse ML algorithms with extensive patient datasets, to further enhance the accuracy and reliability of sepsis prediction. We are particularly interested in IIRGs and are developing multi-classifier ML models to improve diagnostic precision. In our preprint (DOI: https://doi.org/10.21203/rs.3.rs-4306022/v1), we explore the use of ML-derived biomarkers from immune sources and examine the role of immune cells in sepsis pathogenesis. By assessing the significance of each gene, we aim to elucidate their mechanisms and associations with drug responsiveness. This approach is designed to advance early diagnosis, tailor treatment strategies, and ultimately improve patient care in sepsis management.

Materials and methods

Data download

Utilizing the R package GEOquery (version 2.68.0)[17], expression data were retrieved from the GEO database for datasets GSE154918 (n = 105) and GSE134347 (n = 298)[18]. The GSE154918 dataset consists of gene expression profiles from 56 patients diagnosed with sepsis and 49 healthy controls, totaling 105 samples, and was used as the validation dataset for this project. This dataset enables comprehensive comparison of gene expression patterns between patients with sepsis and healthy individuals. Similarly, the GSE134347 dataset includes gene expression data from 215 sepsis patients and 83 healthy controls, totaling 298 samples, and served as the training dataset for this project. This dataset provides extensive resources for analyzing the genetic basis of sepsis, contrasting the expression profiles of affected individuals with those of healthy donors.

The data platform for dataset GSE154918 was GPL20301 Illumina HiSeq 4000 (Homo sapiens); the data platform for dataset GSE134347 was GPL17586 Affymetrix Human Transcriptome Array 2. All samples in dataset GSE154918 and GSE134347 were corrected for batch effects prior to further analysis. See Table S1 for specific dataset information.

To identify IIRGs, the ImmPort database was utilized, available at [https://www.immport.org/home]. ImmPort serves as a critical resource in immunology, providing a platform for the aggregation, organization, and dissemination of research data. It supports the life sciences research community by facilitating the archival and exchange of scientific data through advanced information technology. This database offers a robust repository for both research and clinical data, ensuring long-term, sustainable storage. From this database, a list of 1,509 unique IIRGs was meticulously compiled and cross-verified. For detailed information on these genes, refer to Supplementary Table S2.

During the data integration process, the issue of missing data was carefully addressed to minimize its potential impact on analysis results. Gene expression data were standardized prior to analysis to ensure comparability among different samples.

For batch effects, the R package asva was employed to process the data and eliminate batch effects between datasets, ensuring uniform data distribution and allowing research findings to more accurately reflect real biological differences.

Estimation

The Estimation (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) algorithm is utilized to assess the purity of tumor samples. Developed by MD Anderson Cancer, it requires only a simple gene expression matrix to infer the levels of immune cells, stromal cells, and tumor purity. For this study, the sepsis dataset GSE134347 was input into the ESTIMATE algorithm to calculate the immune score, stroma score, ESTIMATE score, and tumor purity for each sample. The differences in these four scores between sepsis and non-sepsis subgroups were tested using the Wilcoxon test n. $P < 0.05$ was considered statistically significant.

WGCNA algorithm was used to identify disease-related genes in the sepsis dataset

To isolate Sepsis-Related Genes (SRGs), the initial step involved applying the WGCNA[19]. WGCNA aims to identify modules of co-expressed genes, elucidate the relationship between gene networks and immunity, and pinpoint key genes within these networks. The 'pickSoftThreshold' function was used to determine the optimal soft threshold, which was set at 5, facilitating the construction of scale-free networks. Topological matrices were generated, followed by hierarchical clustering. Setting a minimum gene count of 50 for each module, we dynamically sliced and identified gene modules, computed module Eigengenes, and used these Eigengenes to establish inter-module correlations and perform further hierarchical clustering. Modules with correlations above 0.25 were merged, resulting in a total of 22 distinct modules. The relationship between these modules and clinical features was analyzed using Pearson or Spearman correlation analysis.

To identify immune-related differentially expressed genes (IRDEGs) associated with sepsis, the SRGs identified through WGCNA were intersected with the IIRGs from the ImmPort database. To visualize these genes, the R package 'pheatmap' (version 1.0.12) was employed to create an expression heatmap.

Screening for important IIRGs in the sepsis dataset

To refine the identification of important genes within IRDEGs, five prevalent ML algorithms were employed: Elastic Net, LASSO regression, RF, Boruta, and XGBoost decision trees. Elastic Net, a linear regression model, incorporates both L1 and L2 norm regularization in its training. LASSO regression introduces a penalty term to reduce overfitting and enhance generalizability, implemented using the 'glmnet' R package. The outcomes of LASSO regression are visually represented through diagnostic model diagrams and variable trace plots, demonstrating their effectiveness in identifying important genes.

RF uses ensemble learning to integrate multiple decision trees, handling nonlinear relationships and complex interactions effectively. It is robust, performing well with missing data and noise, and evaluates the importance of all genes comprehensively. RF aggregates predictions from multiple trees, with the final decision derived through majority vote, executed via the 'caret' package.

Boruta, a feature selection method gaining popularity, identifies all features correlated with the dependent variable, regardless of their impact on a specific model's cost function. This method ensures that no important gene features are missed and is suitable for feature screening purposes, especially when complete correlation rather than specific model adaptation is of concern. The 'Boruta' package was applied to achieve this.

XGBoost, a gradient boosting algorithm, builds its model iteratively, each time adding a CART tree that fits the residual differences from the previous trees' predictions. XGBoost is particularly suitable for processing large-scale data. It optimizes each iteration through the forward distribution algorithm to improve the accuracy and robustness of the model. In our analysis, XGBoost effectively identified complex gene interactions, thereby improving the classification performance of the final mode, facilitated by the 'xgboost' package. Additionally, each ML algorithm was fine-tuned using Cross-Validation (CV) for hyperparameter optimization, ensuring model performance enhancement. To bolster robustness, the optimization was repeated ten times for each resampling, each time with a different random seed.

When the five different ML models were used to predict the dataset of sepsis samples, they all demonstrated high predictive performance (such as AUC, C-index, and F1-score), indicating that the features we screened are effective and possess good predictive ability in practical applications. The complementarity between the five ML methods helps to identify more important feature genes, which showed consistent importance in different models. This ensured that the features selected were not only important but also highly stable, improving the robustness and accuracy of the final analysis.

Ultimately, to achieve stable results, genes identified by all five ML algorithms were consolidated as the final set of important IIRGs for our ensuing predictive model development.

The diagnostic model of IIRGs in the sepsis dataset

In our effort to develop a sophisticated sepsis response classification model, IIRGs were trained using six prevalent ML algorithms: Naive Bayes (NB), Conditional Inference RF (cforest), LogitBoost (an advanced form of logistic regression), Gradient Boosting Machine (GBM), Model Averaged Neural Network (avNNet), and Penalized Discriminant Analysis (pda). For all these algorithms, cross-validation (CV) was meticulously employed for hyperparameter tuning, aiming to enhance the model's performance and accuracy. To ensure the robustness of our models, this optimization process was diligently repeated ten times, with a unique random seed for each iteration of resampling.

Following the construction of classifiers using these diverse algorithmic models, a thorough analysis was conducted through the validation dataset GSE154918. This step was crucial in determining the most effective algorithm in terms of classification performance within the validation dataset. Subsequently, the algorithm that demonstrated the best classification efficacy was selected for the final assembly of our sepsis prediction model. This approach underscores our commitment to precision and reliability in developing a model adept at predicting sepsis with high accuracy.

Enrichment analysis

The utilization of GO analysis is a prevalent method for conducting comprehensive functional enrichment studies. This analysis encompasses three key areas: Molecular Function (MF), Biological Process (BP) and Cellular Component (CC)[20]. Additionally, the KEGG database is extensively employed for its vast repository of information on genomes, biological pathways, diseases, and pharmaceuticals[21]. To explore the potential mechanisms underlying the actions of the identified crucial IIRGs, the 'clusterProfiler' package in R (version 4.8.3) was leveraged. This package facilitated our detailed exploration through GO annotation analysis and KEGG pathway enrichment analysis. In our study, a False Discovery Rate (FDR) threshold of less than 0.05 was set as the benchmark for statistical significance, ensuring the reliability and relevance of our findings.

CIBERSORT

CIBERSORT, accessible at [https://cibersortx.stanford.edu/], utilizes linear support vector regression, a sophisticated statistical technique in ML[22]. Available as both an R package and a web-based application, it excels in deconvolving expression matrices of various human immune cell subtypes. The tool effectively evaluates the infiltration status of immune cells in sequenced samples, using a gene expression signature set that represents 22 distinct immune cell subtypes. In our study, we used the CIBERSORT algorithm to assess immune cell infiltration in a composite dataset of tumor samples. We also employed the Wilcoxon test to analyze variations in immune cell infiltration among sepsis and non-sepsis subgroups, setting a P value of less than 0.05 as the threshold for statistical significance.

ssGSEA immunoinfiltration analysis

The single-sample gene set enrichment analysis (ssGSEA) algorithm was implemented to precisely quantify the relative abundance of immune cell infiltration[23]. Initially, labels were assigned to various infiltrated immune cell types, including Activated CD8 T cell, Gamma delta T cell, Natural killer cell, and Regulatory T cell. The ssGSEA analysis calculated enrichment scores, which indicated the relative abundance of each immune cell type within individual samples. For graphical representation, the ggplot2 package (version 3.4.2) was used to illustrate the distribution patterns in sepsis and control groups. Additionally, the Wilcoxon test was employed to determine the differences in immune cell infiltration between the sepsis and non-sepsis subgroups, establishing a P-value threshold of less than 0.05 to denote statistical significance.

Correlation analysis of IIRGs and immune infiltration in the sepsis dataset

To further explore the role of IIRGs in sepsis, our study extended its analysis to the correlation between the expression of IIRGs and immune infiltration in sepsis patients, specifically within the sepsis dataset GSE134347. We also investigated the relationship between the expression of IIRGs and immune checkpoints in these patients. For this correlation analysis, Pearson correlation analysis was our primary analytical method. The 'ggcorrplot' R package (version 0.1.4.1) was utilized to create detailed correlation loop diagrams, enhancing our understanding of the interactions and potential impact of these IIRGs in the context of sepsis and endometriosis.

PPI network analysis (STRING)

The STRING database, known for its extensive mapping of both established and speculative Protein–Protein Interaction (PPI)[24], was utilized as a key resource in our study. We used this database to construct a PPI network for the essential genes identified. The parameters for this construction were carefully set at coefficients of 0.4, 0.7, and 0.9 to ensure optimal specificity and relevance. Data from the STRING database were exported and visualized using Cytoscape[25], a sophisticated tool for complex network analysis and visualization. Additionally, to explore the central components of this network, the CytoHubba plug-in[26] was employed for an in-depth analysis of the hub genes. This approach enabled the unraveling of the complex web of interactions among key proteins and the identification of pivotal genes.

Drug sensitivity analysis

The complex genomic alterations in various cancers significantly influence clinical treatment responses, often acting as reliable biomarkers for drug efficacy. In this context, the Genomics of Drug Sensitivity in Cancer (GDSC) database (accessible at www.cancerRxgene.org) is recognized as the most extensive publicly available resource, providing valuable insights into drug sensitivity and molecular indicators of drug response in cancer cells. Utilizing the pRRophetic algorithm, we predicted the sensitivity of patients, grouped by different clinical variables, to prevalent anti-cancer drugs or small molecular compounds. This prediction was based on the analysis of expression matrices from the dataset and involved calculating IC50 values. Comparative group comparison graphs were utilized to present our findings effectively, allowing a clear and detailed visual representation of the results and facilitating a better understanding of drug response dynamics in cancer treatments.

Statistical analysis

Data processing and statistical analysis in our study were meticulously conducted using R software (available at [https://www.r-project.org/], version 4.0.2). For continuous variables across two groups, statistical significance was determined for variables with normal distribution through the independent Student's t-test. For variables not adhering to normal distribution, the Mann–Whitney U test (also known as the Wilcoxon rank-sum test) was used to analyze differences. For categorical variables, the Chi-square test or Fisher's exact test was relied upon to evaluate statistical significance between two groups. In all instances, statistical P values were considered from a bilateral perspective, with a threshold of $P<0.05$ established to denote statistical significance. In this study, P values were not adjusted for multiple comparisons, reflecting more accurately the significance level of each test[27]. When constructing the model, we controlled for relevant variables such as clinical characteristics and minimized the influence of confounding factors, ensuring the reliability of the results. This rigorous approach ensures a comprehensive and accurate assessment of the data, adhering to the highest standards of statistical analysis.

Results

ESTIMATE score distribution between patients with sepsis and those without sepsis

According our technology roadmap (Fig. S1), the ESTIMATE algorithm was initially employed to calculate four immune-related metrics for sepsis: immune score, stroma score, ESTIMATE score, and tumor purity[28]. These scores were compared between sepsis and non-sepsis patient groups. As depicted in the heatmap of Fig. S2-A, a notable disparity was observed in the distribution of these four immune-related scores between sepsis patients and normal sample. Notably, sepsis patients exhibited elevated stroma scores (Anova test, $P<2.2\text{e}{-16}$, Fig. S2-B) and tumor purity (Anova test, $P=7.9\text{e}{-09}$, Fig. S2-C), while showing reduced ESTIMATE scores (Anova test, $P=7.9\text{e}{-09}$, Fig. S2-D) and immune scores (Anova test, $P<2.2\text{e}{-16}$, Fig. S2-E). This analysis provides valuable insights into the immunological landscape of sepsis, highlighting significant differences in immune and stromal components.

Identification and correlation analysis of IIRG modules in sepsis

A WGCNA was conducted on the sepsis dataset to identify gene modules associated with sepsis immunity. The analysis revealed that the optimal soft threshold was 5, achieving the lowest mean connectivity (Fig. S3-A,B). Figure S3-C displayed the gene clustering numbers, with various modules differentiated by distinct colors. Subsequent steps involved identifying gene modules in relation to the four immune-related scores.

The correlation heatmap between different gene modules and the four immune-related scores is shown in Fig. S3-C, revealing intriguing correlations; for instance, matrix scores exhibited the strongest association with the darkgreen module. Conversely, immune scores were most closely linked with the grey module. Both the ESTIMATE score and tumor purity showed the highest correlation with the black module. These findings offer a nuanced understanding of the relationships between specific gene modules and key immune-related scores in the context of sepsis.

In our analysis of gene clustering within different color-coded modules, a notable similarity in expression patterns among genes grouped in the same colored module was observed (Fig. 1A). Further exploration into the inter-modular relationships revealed relatively low correlation levels between different modules, as illustrated in Fig. 1B. We then focused on detailing the heatmap of correlations between these diverse colored modules and sepsis. This analysis brought to light that the darkmagenta module demonstrated the most significant negative correlation with sepsis ($r = -0.78$), whereas the brown module exhibited the most substantial positive correlation with the disease ($r = 0.7$), as shown in Fig. 1C. Subsequently, detailed scatter plots were presented illustrating the correlation between the brown module and its associated genes. This scatter plot analysis revealed a significant correlation ($P < 0.05$, Fig. 1D). Based on these findings, the genes within the brown module were ultimately selected as the final identified IIRGs. This decision was grounded in the strong correlation these genes exhibited with sepsis, underscoring their potential importance in understanding the disease's immunological aspects.

Expression differences and biological pathway enrichment analysis of IIRGs in sepsis

To enhance our understanding of IIRGs in sepsis, we first sourced a set from the ImmPort database. These were intersected with disease-related genes identified through WGCNA, and the intersection was depicted in a Venn diagram (Fig. 2A). This process led to the identification of 108 IRDEGs specifically expressed in the context of sepsis (Table S3). To compare their expression patterns visually, we utilized the R-package 'pheatmap' to create a heatmap. Displayed in Fig. 2B, this heatmap showed distinct expression patterns of these 108 genes between sepsis and non-sepsis patients. Furthermore, our analysis revealed that genes significantly overexpressed in sepsis patients were predominantly enriched in biological pathways such as Osteoclast Differentiation, B Cell Receptor Signaling Pathway, Th17 Cell Differentiation, and T Cell Receptor Signaling Pathway. Conversely, genes markedly upregulated in healthy patients showed significant enrichment in the Chemokine Signaling Pathway, Th17 Cell Differentiation, JAK-STAT Signaling Pathway, PD-L1 Expression, and the PD-1 Checkpoint Pathway in Cancer, among other pathways. These findings offer crucial insights into the distinct immunological landscapes characterizing sepsis patients compared to healthy individuals.

Functional enrichment analysis of IRDEGs reveals

To elucidate the potential molecular mechanisms underpinning the IRDEGs, we performed both GO and KEGG functional enrichment analyses on the 108 IRDEGs. The KEGG analysis identified significant enrichment of these genes in pathways associated with cancer immunity. Notable pathways include Osteoclast Differentiation, Th17 Cell Differentiation, Cytokine–Cytokine Receptor Interaction, B Cell Receptor Signaling Pathway, T Cell Receptor Signaling Pathway, and the JAK-STAT Signaling Pathway, as shown in Fig. S4-A and Table S4. Furthermore, the GO functional enrichment analysis highlighted their significant roles in critical biological processes. These include the Cytokine-Mediated Signaling Pathway, Positive Regulation of Cytokine Production, Leukocyte Mediated Immunity, Immune Receptor Activity, Growth Factor Receptor Binding, and the T Cell Receptor Complex, as depicted in Fig. S4-B and Table S5. These findings provide insight into the roles of these 108 IRDEGs, particularly in contributing to key immune pathways and processes.

IIRG features identified by ML algorithm and displayed in a Venn diagram

To further identify critical features in IRDEGs, we utilized five common ML algorithms: Elastic Net, LASSO regression, RF, Boruta, and XGBoost decision trees. LASSO regression identified 53 important genetic features (Fig. 3A); Elastic network identified 38 important genetic features (Fig. 3B); RF identified 108 important genetic features (Fig. 3C); 61 important gene features were identified by Boruta algorithm (Fig. 3D); and XGBoost identified 20 important genetic features (Fig. 3E). As illustrated in the Venn diagram (Fig. 3F), the five ML algorithms collectively identified 11 IIRGs, which we subsequently designated as marker genes.

High-performance sepsis prediction model developed using six ML algorithms

Subsequently, we employed six different ML algorithms to develop sepsis prediction models. The results demonstrated that all six prediction models achieved high AUC values (Fig. 4A), and the models' C-index and F1-scores were also high (Fig. 4B), indicating that the prediction model we developed exhibits high prediction performance.

Independent dataset validates ML algorithm model for sepsis prediction

The predictive performance of our model was validated using an independent sepsis dataset. In the GSE154918 dataset, models constructed using six different ML algorithms demonstrated relatively high AUC values, all greater than 0.75, with the pda model reaching 0.901 (Fig. 5A, Table S6). However, these models showed relatively low C-index and F1-scores (Fig. 5B), suggesting that our model provides good and stable predictive performance for sepsis.

Importance and contribution of genetic features in different models

The contribution of 11 IIRGs was investigated across various models. In the NB model, gene MAPK14 was identified as the most contributive to sample prediction (Fig. S5a-A). In the LogitBoost model, gene IL10 made the largest contribution (Fig. S5a-B); in the GBM model, gene IL21R was the most influential (Fig. S5a-C); in

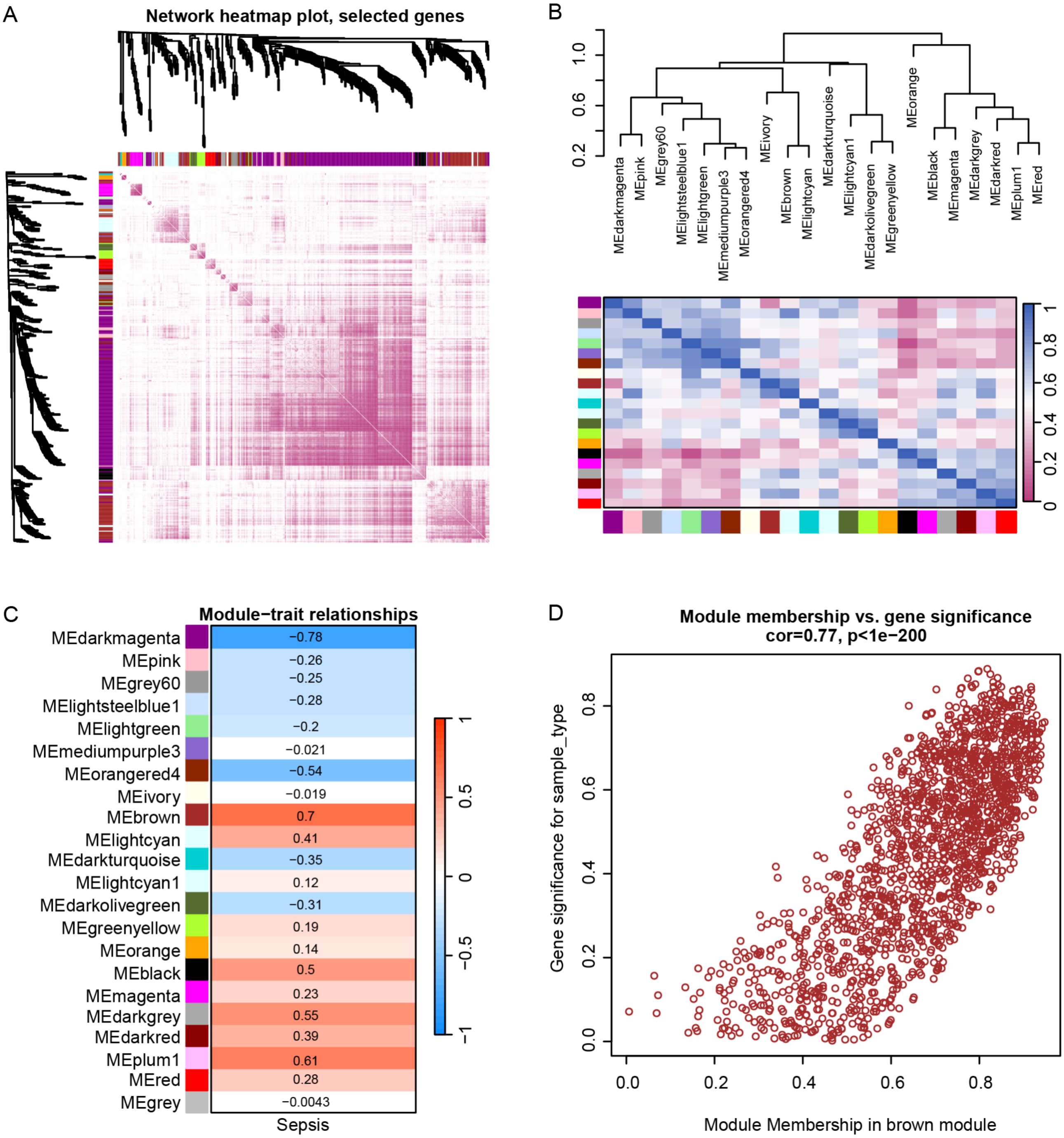

Fig. 1. Correlation analysis of the most relevant modules in the sepsis dataset. (**A**) Clustering network diagram between the genes of different color modules; (**B**) heat maps of correlations between different modules, with blue representing high correlations and purple representing low correlations; (**C**) heat maps of correlation between different colors and sepsis, with red representing positive correlation and blue representing negative correlation; And (**D**) scatter plots of correlations in the brown gene module.

the cforest model, gene MAPK14 again played a significant role (Fig. S5a-D). In the avNNet model, gene JAK2 was the most contributive (Fig. S5a-E). For the pda model, gene MAPK14 was found to be the most influential (Fig. S5a-F).

In the NB model, gene JAK2 made the largest contribution (Fig. S5b-A). In the LogitBoost model, gene SOCS3 was the most contributive (Fig. S5b-B); in the GBM model, gene NCR3 played the most significant role (Fig. S5b-C). In the cforest model, gene MAPK14 was identified as the most influential (Fig. S5b-D). In the avNNet model, gene JAK2 made the largest contribution (Fig. S5b-E). In the pda model, gene MAPK14 was the most significant contributor (Fig. S5b-F).

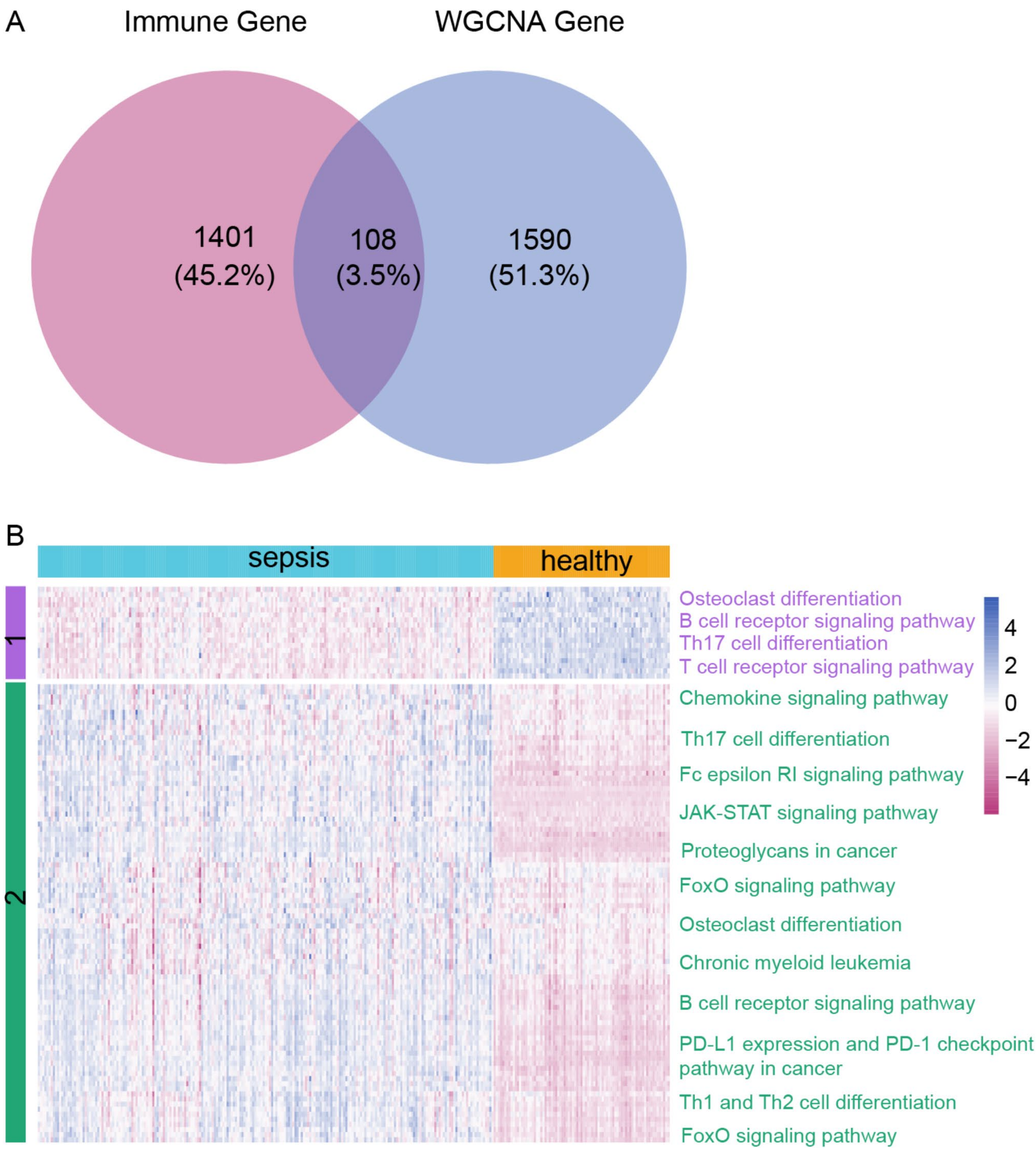

Fig. 2. Identification of IRDEGs in the sepsis dataset. (**A**) Intersection diagram of disease-related genes and immune-related genes identified by WGCNA; and (**B**) heat maps of expression of sepsis associated immune genes between sepsis and normal patients. Blue represents high expression and purple represents low expression. IRDEGs, immune-related differentially expressed genes; WGCNA, weighted gene correlation network analysis.

Relationship analysis between IIRGs and immune infiltration

To explore the association between the 11 IIRGs and immune infiltration in sepsis, we utilized CIBERSORT (Table S7) and ssGSEA (Table S8) methodologies. The CIBERSORT analysis revealed significant differences in immune cell profiles between sepsis patients and healthy controls (Fig. 6A). Specifically, immune cells such as T cells CD4 memory resting, T cells CD8, NK cells resting, B cells naive, and T cells CD4 naive were predominantly observed in healthy individuals, whereas Neutrophils, T cells regulatory (Tregs), Macrophages M0, and Monocytes were notably more abundant in sepsis patients.

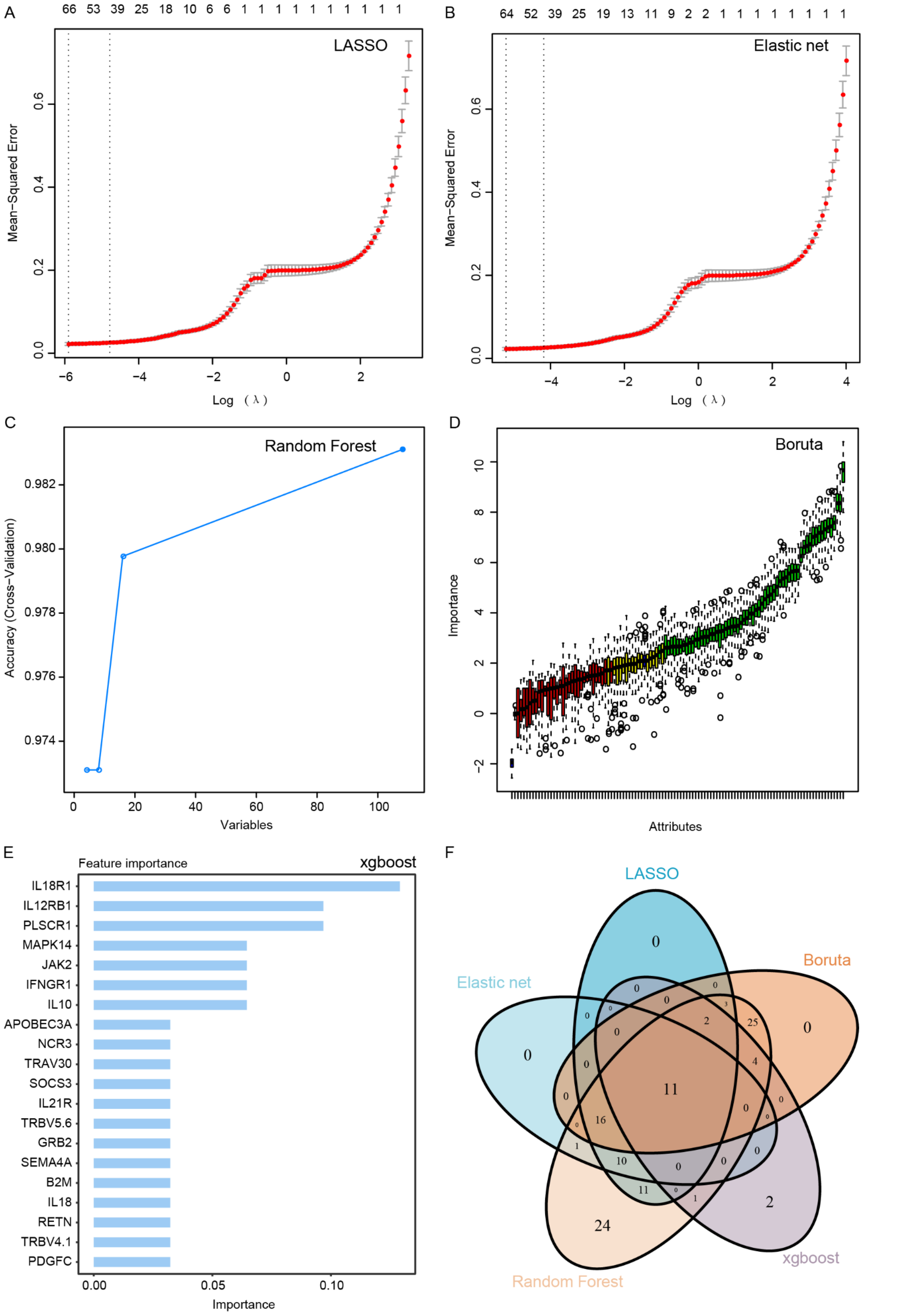

Fig. 3. Screening important immune-related features in the sepsis dataset. (**A**) LASSO regression screening for important features; (**B**) elastic network screening important features; (**C**) RF screening important features; (**D**) Boruta screening for important features; (**E**) XGBoost algorithm filters important features. LASSO, Least absolute shrinkage and selection operator.

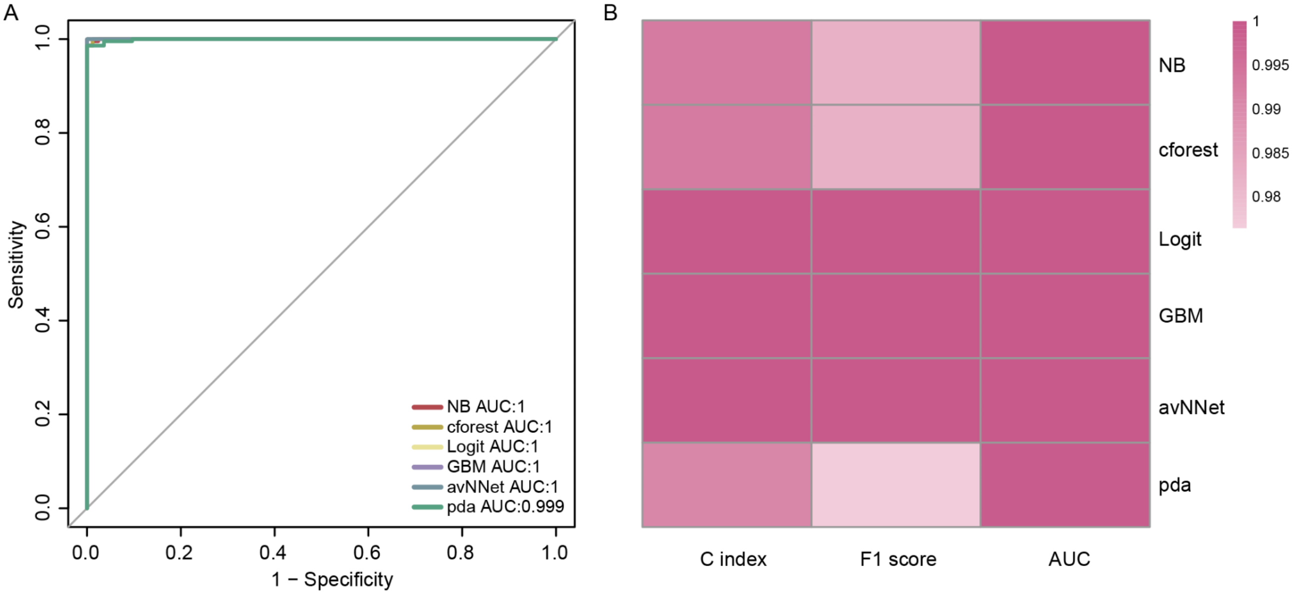

Fig. 4. Construction of sepsis prediction model. (**A**) ROC curve of sepsis prediction model constructed by different machine learning algorithms; and (**B**) C-index and F1-score of sepsis prediction models built using different machine learning algorithms. ROC, receiver operating characteristic.

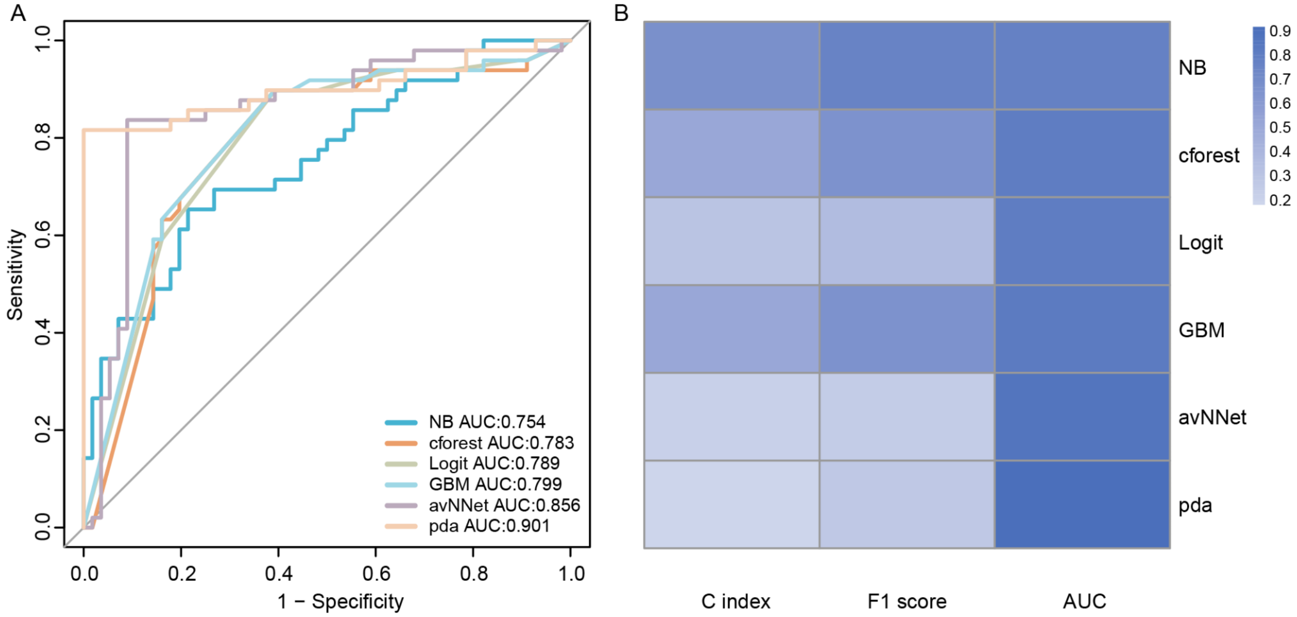

Fig. 5. Verifying the sepsis prediction model in the GSE154918 dataset. (**A**) ROC curve of sepsis prediction model constructed by different machine learning algorithms in GSE154918 dataset; (**B**) C-index and F1-score of sepsis prediction models constructed using different machine learning algorithms in the GSE154918 dataset. ROC, receiver operating characteristic.

The ssGSEA analysis indicated marked disparities in immune infiltration between sepsis patients and healthy individuals (Fig. 6B). In sepsis patients, immune cells such as Neutrophils, Th2 cells, Macrophages, Mast cells, iDC, DC, Th17 cells, Treg cells, and aDC were significantly enriched. Conversely, immune cells like NK CD56bright cells, TNK cells, Th1 cells, NK CD56dim cells, and Tfh cells were more prevalent in healthy patients. These findings provide valuable insights into the immune landscape of sepsis, highlighting the distinct immune cell profiles in sepsis patients compared to healthy individuals and emphasizing the potential roles of these 11 IIRGs in modulating immune responses in sepsis.

In our study, we explored the correlation between the expression of 11 IIRGs and the levels of immune infiltration in patients, visualizing these relationships through correlation circles. Within the CIBERSORT

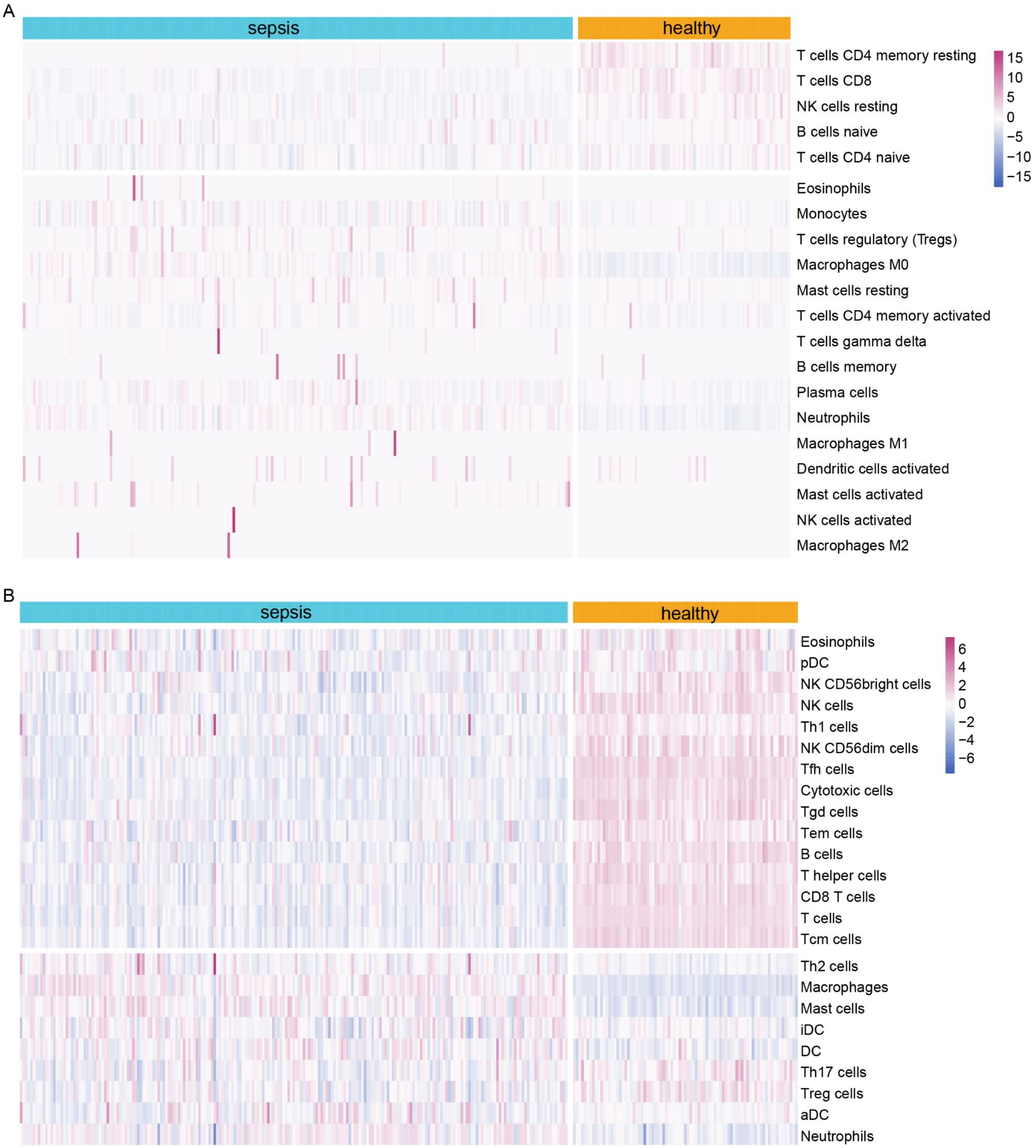

Fig. 6. Analysis of immune infiltration in sepsis and healthy patients in the sepsis dataset. (**A**) Distribution of immune cell infiltration between sepsis and non-sepsis patients in CIBERSORT algorithm; and (**B**) distribution of immune cell infiltration between sepsis and non-sepsis patients in ssGSEA algorithm, with purple representing high infiltration and blue representing low infiltration.ssGSEA, single-sample gene-set enrichment analysis.

analysis, a significant correlation was observed between the expression of these 11 IIRGs and various immune cells, notably T cell CD8, T cells CD4 memory resting, Macrophages M0, and Neutrophils (Fig. S6-A). Similarly, the ssGSEA analysis showed a strong correlation between the expression of these genes and immune cells such as T cell CD8, B cells, Cytotoxic cells, Macrophages, NK cells, and T cells (Fig. S6-B).

Analysis of the relationship between IIRGs and immune checkpoints

Furthermore, we investigated the interaction between the expression of these 11 IIRGs and immune checkpoint genes. Initially, the expression patterns of immune checkpoint genes were compared between sepsis patients and healthy samples. This comparison revealed a notable difference in the expression of these genes between the two groups, suggesting a potential impact of immune factors on the progression of sepsis (Fig. 7A). Subsequent analysis focused on the correlation between the expression of these 11 IIRGs and immune checkpoint genes, again represented through correlation circles. Notably, genes such as IL21R, NCR3, and TRAV30 from our characteristic set exhibited a significant positive correlation with the expressions of most immune checkpoint genes (Pearson correlation analysis, $P < 0.05$), including HLA-DPB1, HLA-DPA1, HLA-DQB2, HLA-DRB1, HLA-DQA1 (Fig. 7B). Conversely, the other IIRGs in the study often displayed a negative correlation with the expression levels of a similar spectrum of immune checkpoint-related genes (Pearson correlation analysis, $P < 0.05$), as indicated in Fig. 7B. These results underscore the intricate relationships between IIRGs and immune checkpoints, providing critical insights into their influential roles within the immune responses characteristic of sepsis.

Protein interaction network analysis of important immunity-related genes

The interactions among the 11 identified IIRGs were analyzed within the STRING database (Fig. S7), revealing that these genes exhibited strong interactions with each other. Notably, the genes JAK2 and IL10 showed a high degree of connectivity within the network, indicating their strong interactions with other genes.

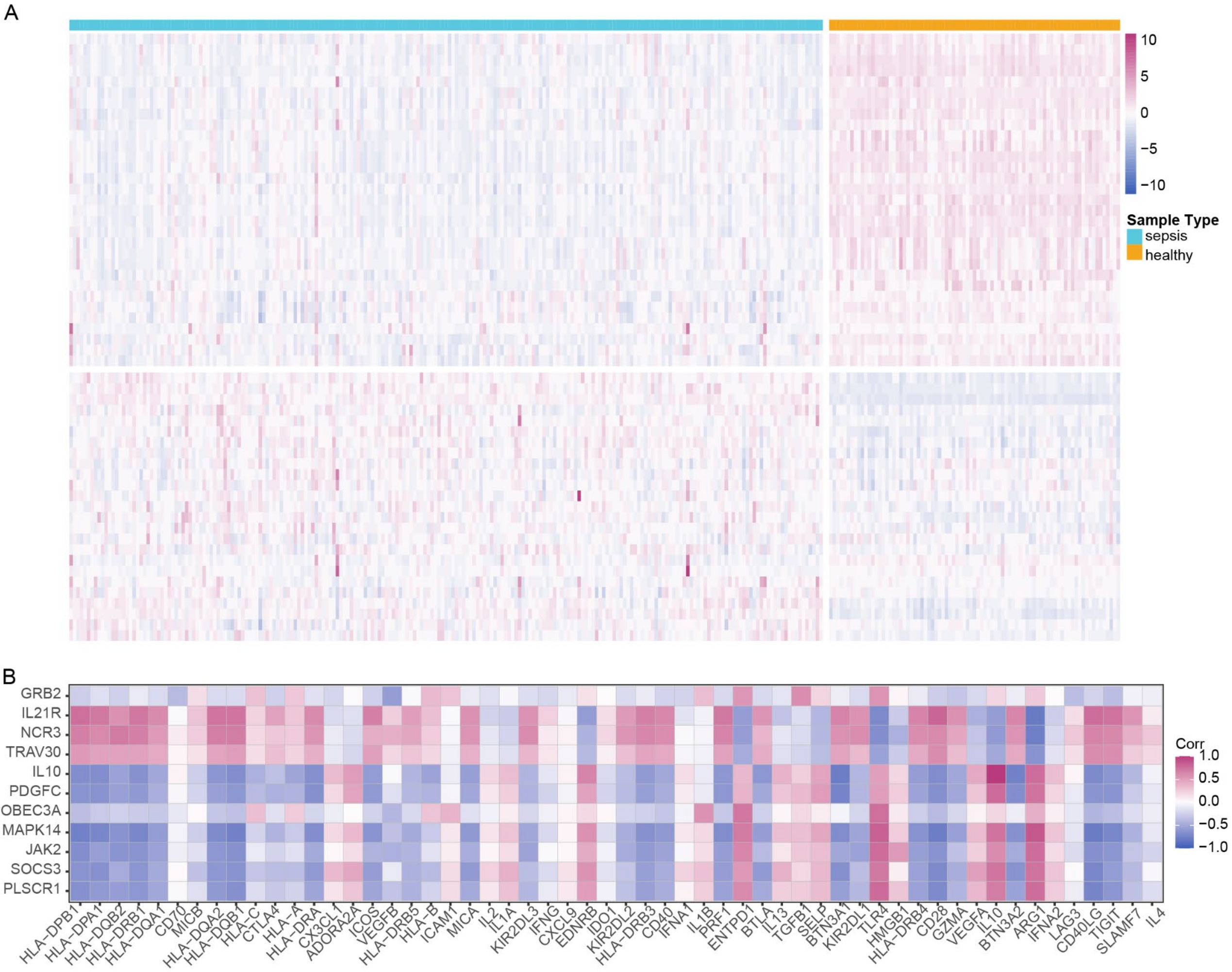

Fig. 7. Correlation analysis between IIRGs and immune checkpoint related genes in the sepsis dataset. (**A**) Heat maps of the expression of immune checkpoint related genes in the sepsis training dataset in patients with sepsis and patients without sepsis, with blue representing low expression and purple representing high expression. (**B**) The correlation between the expression of IIRGs in the sepsis training dataset and immune checkpoint-related genes, blue represents negative correlation and purple represents positive correlation. IIRGs, important immune-related genes.

Association analysis of IIRGs and drug sensitivity

To investigate the relationship between the 11 IIRGs and drug sensitivity, the pRRophetic package was used to calculate the IC50 values for 14 drugs from the GSE134347 sepsis dataset in the CCLE database. The analysis revealed that, with the exception of PD.0325901, PF2341066, and PHA.665752, the IC50 values of the remaining drugs were significantly different between sepsis and healthy patients ($P < 0.05$, Fig. 8A), suggesting a distinct difference in drug efficacy between these groups. Additionally, a correlation analysis was conducted between the 11 IIRGs and the IC50 of drug sensitivity, showing significant associations (Fig. 8B). For example, the expression of GRB2 gene was strongly and positively correlated with the IC50 of Erlotinib ($r = 0.51$, $P < 0.05$), yet it displayed a negative correlation with the IC50 of PD.0325901 ($r = -0.61$, $P < 0.05$).

Discussion

In a comprehensive study spanning 2005 to 2014 across 27 academic hospitals, there was observed a significant increase in septic shock incidences[29,30], rising from 12.8 to 18.6 cases per 1,000 hospital admissions, while mortality rates decreased slightly from 55 to 51%[31]. This upward trend is attributed to factors[32–35] such as aging populations, increased immunosuppression, and the prevalence of multi-drug resistant infections, underscoring the ongoing challenge of sepsis as a critical global health concern[13,36]. Despite traditional inflammatory markers being crucial in diagnosing various sepsis types, a significant gap in research remains regarding immune exhaustion in septic patients[37,38], which could result in either under-treatment or overtreatment[39–41].

In response to these challenges, our team has developed an innovative multi-biomarker model using ML techniques. This model successfully identifies 11 key IIRGs., enhancing the ability to detect sepsis and predict drug treatment responsiveness. This advancement not only facilitates the early identification of septic patients but also aids in evaluating their immune status, thereby laying the groundwork for more tailored and precise treatment approaches. Additionally, our research highlights the potential effectiveness of immune checkpoint blockade therapy in treating sepsis, marking a significant stride in the field.

In our preprint (DOI: https://doi.org/10.21203/rs.3.rs-4306022/v1), RNA-seq data from GSE154918 was utilized and sepsis-related disease expression genes were identified through WGCNA. By intersecting these genes with IIRGs obtained from the ImmPort database (https://www.immport.org/shared/home), we pinpointed 108 immune-related disease expression genes linked to sepsis. Using five different commonly used ML algorithms for CV, we identified 11 key IIRGs. Further, we leveraged six prevalent ML methods to sift through these genes, aiming to find the algorithm with the best classification performance in our validation dataset, leading to the development of a new predictive model. This model demonstrated precision in identifying septic patients, thereby validating the predictive value of these 11 IIRGs in sepsis. The genes identified are *GRB2, IL21R, NCR3, TRAV30, IL10, PDGFC, APOBEC3A, MAPK14, JAK2, SOCS3,* and *PLSCR1*. A comprehensive literature review revealed that six of these genes—*GRB2, IL21R, NCR3, TRAV30, PDGFC,* and *PLSCR1*—had not been previously associated with sepsis. In-depth exploration of the roles of these genes in sepsis poses a significant challenge[42]. However, such research could potentially reveal novel therapeutic targets or disease mechanisms, offering fresh perspectives and theoretical foundations for developing new treatments for sepsis.

The IIRGs we identified are involved in multiple key immune signaling pathways, such as cytokine signaling pathways, T cell receptor signaling pathways, and B cell receptor signaling pathways. Cytokines play a crucial regulatory role in the inflammatory responses associated with sepsis. These pathways influence the activation and differentiation of immune cells, thus regulating the immune status in patients with sepsis. For example, Th17 cells enhance the body's pathogen clearance capability by promoting the activation of other immune cells through the secretion of cytokines such as IL-17[43].

These IIRGs potentially play a significant role in regulating immunity and inflammation. Among the identified IIRGs, certain genes exhibit critical biological functions. IL10, an anti-inflammatory cytokine, may shield the host by dampening inflammatory responses in sepsis and preventing a persistent inflammatory state[44]. The expression of IL-10 is inversely associated with the prognosis of sepsis patients, indicating its protective role in the progression of the disease, although excessive IL-10 can lead to immune escape and foster infection[45]. In our constructed PPI network, IL10 shows a high degree of connectivity, and a significant correlation exists between IL-10 and MAPK14, along with related signaling pathways. Research has identified MAPK14 as a key regulator of IL-10[46], thus marking it as a potential sepsis treatment target. Additionally, another study recognized MAPK14 as crucial within a new mitochondrial-related gene signature for diagnosing sepsis[47]. This study established a diagnostic model featuring MAPK14, underscoring its diagnostic significance, which aligns with our findings. Our PPI network analysis further confirms the linkage of MAPK14 with pathways involving PDGFRB, SOS2, IL-10, and GRB2, highlighting MAPK14 as a promising target for sepsis therapy.

JAK2 is implicated in the immune response by mediating cytokine signaling[48]. Its abnormal expression in sepsis may result in an inadequate immune response to infection and diminish the capacity to eliminate pathogens. Additionally, JAK2 is thought to regulate leukocyte differentiation and function, potentially affecting the inflammatory microenvironment.

Through our PPI network analysis, a high degree of connectivity was observed between JAK2 and GRB2 within a network composed of 11 IIRGs. However, research on the interaction between JAK2 and GRB2 has predominantly been focused in oncology, with limited exploration in the context of sepsis.

The GRB2 protein is pivotal[49] in various critical biological processes, including cell growth, proliferation, metabolism, embryonic development, and the differentiation of cancer cells. It plays a crucial role in the signal transduction of immune cells[50], with the function of Gab2 being essential in the progression of various cancers, often associated with thrombosis and inflammation. However, the specific role of GRB2 in sepsis has not been highlighted. Therefore, investigating the interaction between GRB2 and JAK2 in sepsis could lead to groundbreaking discoveries, providing new perspectives and strategies for treating sepsis.

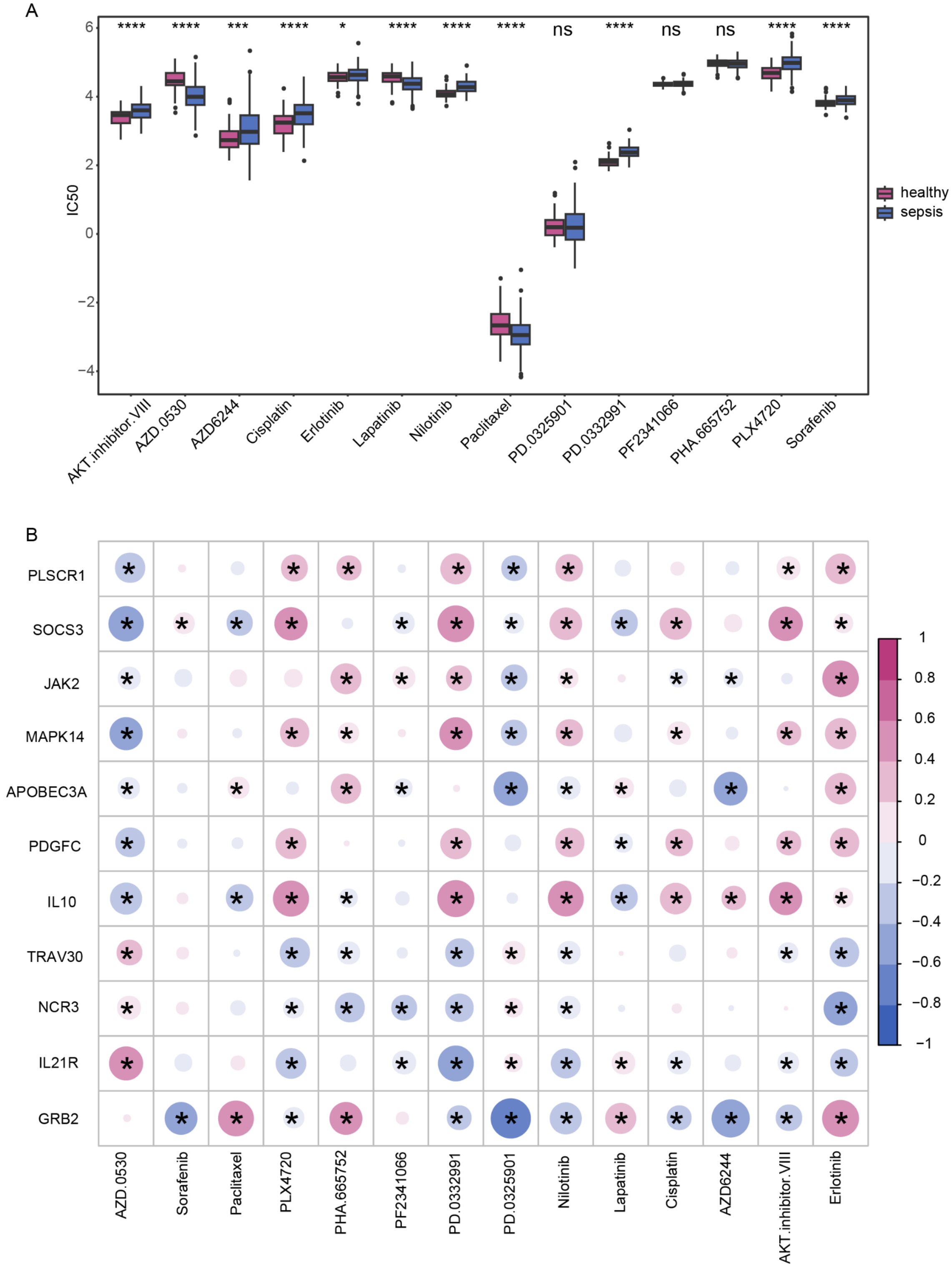

Fig. 8. Drug sensitivity analysis of the sepsis dataset. (**A**) Box diagram of drug IC50 distribution between sepsis patients and healthy patients, ns representing $P \geq 0.05$, * representing $P < 0.05$, *** representing $P < 0.001$. (**B**) Correlation circle diagram of 11 IIRGs and drug IC50, purple represents positive correlation, blue represents negative correlation, * represents $P < 0.05$, the darker the color, the stronger the correlation. IIRGs, important immune-related genes.

NCR3, a natural killer (NK) cell receptor, shows upregulation in sepsis, indicating the significance of NK cells in controlling infection and clearing tumor cells[51]. However, the activation of NCR3 may also promote systemic inflammatory responses, and its dual role in sepsis warrants further exploration[52].

Through KEGG enrichment analysis, it was found that these IIRGs are predominantly involved in cytokine signaling pathways, T cell receptor signaling pathways, etc., and are closely associated with inflammation and immune regulation. Sepsis patients often experience severe cytokine storms, and the abnormal expression of related genes may be a critical factor leading to imbalanced immune responses. The immunosuppressive characteristics of sepsis are linked to the dysfunction of T cells and B cells[53]. A deeper exploration of the role of IIRGs in the activation, proliferation, and functional regulation of T cells and B cells will aid in understanding the immune escape mechanisms in sepsis.

Alterations in the expression levels of key immune infiltration genes play a critical role in the onset, progression, and treatment of sepsis[54]. Utilizing the CIBERSORT algorithm, we analyzed the relationship between the expression of these pivotal genes and immune cell infiltration in septic patients. Our findings indicated a significant negative correlation between *MAPK14*, *JAK2*, *SOCS3*, and *PLSCR1* with T cell CD8 immune cells, whereas *IL21R* and *NCR3* were positively correlated with T cell CD8. These results are crucial for understanding the immune response characteristics in septic patients. Additionally, *SOCS3, MAPK14*, and *IL-10* showed a positive correlation with Macrophages M0. Analysis of these key genes and their impact on immune infiltration using the ssGSEA algorithm revealed that genes like *MAPK14, JAK2, SOCS3, PLSCR1, IL-10,* and *PDGFC* were positively correlated with Macrophages, while *IL21R* and *NCR3* demonstrated negative correlations with Macrophages and positive correlations with various NK and T cells, reflecting their diverse roles in immune regulation. Further research into these IIRGs, especially those not previously directly linked to sepsis such as *IL21R, NCR3, PLSCR1, is deemed highly significant*[55].

Sepsis and cancer share many pathophysiological characteristics, with immune suppression mechanisms in both involving dysfunctions in myeloid and lymphoid cells, ultimately leading to impaired antibacterial phagocytosis and antitumor cytotoxicity[56]. Recent studies have suggested that antitumor drugs might alleviate lung damage during acute sepsis[57], and the occurrence of sepsis might also reduce the risk of certain cancers[58]. Exploring the immune-related aspects between sepsis and cancer represents a novel area of investigation. Our study discovered that these 11 IIGs are closely linked with immune checkpoints, crucial in regulating the immune system, particularly in tumor immune evasion[59]. This suggests the presence of tumor-like immune escape mechanisms in sepsis[60–62]. Previous research has indicated that anti-*PD-L1* peptides could improve survival rates in mice infected with fungi[62], pointing to a role for immune checkpoint pathways in immune suppression during the later stages of sepsis. Through KEGG functional enrichment analysis, pathways related to cancer immunity were found to be significantly enriched in sepsis, suggesting a potential overlap in immune regulation between these two diseases. Surprisingly, we also discovered that the expression of these IIGs is closely associated with the sensitivity and resistance to antitumor drugs, with the *GRB2* gene showing a significant positive correlation with paclitaxel[63]. While studies have demonstrated that the *GRB2* gene could increase sensitivity to paclitaxel in non-small cell lung cancer, its role in enhancing sensitivity to paclitaxel through the activation of the *MAPK* pathway in sepsis is yet to be reported. This study not only validates the efficacy of IIRGs as biomarkers for diagnosing sepsis but also reveals their significant correlation with genes related to immune checkpoints. Modulating the expression or function of these genes could help restore normal immune responses or improve drug responsiveness in patients with sepsis, potentially increasing survival rates. These findings might pave the way for new potential therapeutic targets in sepsis treatment, laying the groundwork for the development of more effective drug interventions for this disease.

While the expression of our constructed 11 characteristic genes shows good and stable predictive performance in identifying sepsis, closely correlating with patients' immune infiltration status and drug sensitivity, several limitations must be acknowledged and addressed in subsequent research. Firstly, our study primarily utilized datasets from GSE154918 (n = 105) and GSE134347 (n = 298). Although the sample sizes of these datasets are relatively large, they may not adequately represent all populations and clinical scenarios, potentially limiting the generalizability of our results. Despite existing literature[64,65] supporting the reliability of these datasets in sepsis research, future studies should include more datasets to enhance the applicability of the results. Secondly, despite our model's demonstrated good predictive performance, given the nature of sepsis as a potentially fatal disease, additional evaluation indicators are required to assess the model and improve the accuracy and scientific rigor of the study. Additionally, although the model showed high accuracy and applicability across multiple datasets, it has not yet been tested on clinical samples, which could limit its practical clinical use. These genes still need experimental verification or clinical evaluation to refine the accuracy and scientific rigor of the study. In the future, we will consider relevant experimental validation or clinical evaluations to enhance the reliability and comprehensiveness of our research results. Moreover, our data analysis was based on data from public databases, which might contain errors or missing information. Without strict quality control measures, the reliability of our analysis could be compromised. Therefore, more prospective and mechanistic studies are needed to further validate and refine these findings.

In our study, we identified 11 important genes associated with sepsis, and the expression patterns of these genes can serve as potential biomarkers for early diagnosis and disease stratification. In the future, standardized detection methods could be developed to quickly assess the expression of these biomarkers in clinical samples. This approach could enable doctors to identify sepsis patients early based on gene expression characteristics, thereby initiating appropriate treatment promptly.

Our study established a variety of ML models (such as conditional inference RF, naive Bayes, etc.) and verified their high predictive performance in sepsis classification. These models can be further refined with actual clinical data to optimize parameters and algorithms to suit specific patient groups. By integrating these predictive models with clinical information systems, doctors can swiftly calculate the sepsis risk score upon hospital admission and formulate personalized treatment plans based on this.

In the future, to effectively integrate these biomarkers and predictive models into clinical workflows, preclinical and clinical validation, along with multidisciplinary collaboration and education, are essential to enhance their application in sepsis management.

Sepsis is a complex and rapidly evolving condition, and the introduction of ML tools risks leading to an over-reliance on models by medical staff[66]. While AI can analyze vast amounts of data and identify potential patterns, the diagnostic process should still prioritize the doctor's clinical intuition and judgment. For instance, if the model inaccurately assesses the patient's status, it may lead to unnecessary treatments or delay the correct diagnosis[67]. Medical staff must maintain a balance between AI recommendations and clinical experience, and strengthen the verification and cross-checking of model results to ensure that patients receive the most appropriate treatment[68].

The integration of AI tools in medical practice necessitates a clear ethical framework, which involves not only the ethical relationships between the model's developers, medical providers, and patients but also the broader social impact. The development of comprehensive ethical guidelines can assist medical institutions in addressing ethical challenges that may arise when deploying these tools[69].

Conclusion

This investigation has successfully identified 11 IIRGs linked to sepsis through the application of genomic analysis combined with machine learning techniques. These genes, including GRB2, IL21R, and NCR3, are identified as potential biomarkers for early sepsis diagnosis, offering profound insights into the immune processes that underpin the condition. The predictive models developed from these genes have shown high levels of accuracy and stability, highlighting their potential applicability in clinical settings to improve detection and treatment of sepsis. Moreover, functional enrichment and immune infiltration analyses underscore the pivotal role these genes play in modulating immune responses within sepsis contexts. Furthermore, a notable correlation between these IIRGs and drug sensitivity highlights the prospects for personalized medicine, suggesting customized treatment options for sepsis patients. This research emphasizes the significance of discerning immune-related genetic markers in sepsis, laying down a theoretical framework for the formulation of novel diagnostic and therapeutic avenues that confront a significant global health challenge. Future research should focus on validating these results across varied populations and delving deeper into the mechanistic functions of these genes to optimize their clinical efficacy in managing sepsis.

Data availability

All data generated in this study are publicly available in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/). Specifically datasets GSE154918 and GSE134347. All materials used in this study are commercially available or have been described in detail in the Methods section. Any additional data that support the findings of this study are available from the corresponding author upon request. The source code is available at https://www.jianguoyun.com/p/DW_ArZEQi9maDRjCtukFIAAe.

Received: 24 September 2024; Accepted: 4 March 2025
Published online: 11 March 2025

References

1. Rudd, K. E. et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *Lancet* **395**, 200–211 (2020).
2. Cohen, M. & Banerjee, D. Biomarkers in sepsis: A current review of new technologies. *J. Intensive Care Med.* **39**, 399 (2023).
3. Oikonomakou, M. Z., Gkentzi, D., Gogos, C. & Akinosoglou, K. Biomarkers in pediatric sepsis: A review of recent literature. *Biomark. Med.* **14**, 895–917 (2020).
4. Zeng, Z., Peng, Y. & Yuan, Z. Research advances of sepsis biomarkers. *Zhonghua Shao Shang za zhi = Zhonghua Shaoshang Zazhi = Chin. J. Burns* **39**, 679–684 (2023).
5. Barichello, T., Generoso, J. S., Singer, M. & Dal-Pizzol, F. Biomarkers for sepsis: More than just fever and leukocytosis—A narrative review. *Crit. Care* **26**, 14 (2022).
6. Khanh Le, N. Q., Nguyen, Q. H., Chen, X., Rahardja, S. & Nguyen, B. P. Classification of adaptor proteins using recurrent neural networks and PSSM profiles. *BMC Genom.* **20**, 966. https://doi.org/10.1186/s12864-019-6335-4 (2019).
7. Kha, Q. H., Le, V. H., Hung, T. N. K., Nguyen, N. T. K. & Le, N. Q. K. Development and validation of an explainable machine learning-based prediction model for drug-food interactions from chemical structures. *Sensors* **23**, 3962. https://doi.org/10.3390/s23083962 (2023).
8. Jiang, L. et al. Identification of diagnostic gene signatures and molecular mechanisms for non-alcoholic fatty liver disease and Alzheimer's disease through machine learning algorithms. *Clin. Chim. Acta Int. J. Clin. Chem.* **557**, 117892. https://doi.org/10.1016/j.cca.2024.117892 (2024).
9. Wang, Y. et al. Collaborative deep learning improves disease-related circRNA prediction based on multi-source functional information. *Brief. Bioinformat.* **24**, bbad069. https://doi.org/10.1093/bib/bbad069 (2023).
10. Rydzewski, N. R. et al. Machine learning & molecular radiation tumor biomarkers. *Semin. Radiat. Oncol.* **33**, 243–251. https://doi.org/10.1016/j.semradonc.2023.03.002 (2023).
11. Takefuji, Y. Artificial intelligence universal biomarker prediction tool. *J. Thromb. Thrombolysis* **57**, 341–343. https://doi.org/10.1007/s11239-023-02930-7 (2024).
12. Yang, Z., Cui, X. & Song, Z. Predicting sepsis onset in ICU using machine learning models: A systematic review and meta-analysis. *BMC Infect. Dis.* **23**(1), 635 (2023).
13. Goh, K. H. et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat. Commun.* **12**, 711 (2021).
14. Pepic, I. et al. Early detection of sepsis using artificial intelligence: A scoping review protocol. *Syst. Rev.* **10**, 1–7 (2021).
15. Valik, J. K. et al. Predicting sepsis onset using a machine learned causal probabilistic network algorithm based on electronic health records data. *Sci. Rep.* **13**, 11760 (2023).

16. Zhang, Y., Xu, W., Yang, P. & Zhang, A. Machine learning for the prediction of sepsis-related death: A systematic review and meta-analysis. *BMC Med. Informat. Decis. Mak.* **23**, 283 (2023).
17. Davis, S. & Meltzer, P. S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847. https://doi.org/10.1093/bioinformatics/btm254 (2007).
18. Malmstrom, E. et al. The long non-coding antisense RNA JHDM1D-AS1 regulates inflammatory responses in human monocytes. *Front. Cell. Infect. Microbiol.* **12**, 934313. https://doi.org/10.3389/fcimb.2022.934313 (2022).
19. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformat.* **9**, 559. https://doi.org/10.1186/1471-2105-9-559 (2008).
20. Gene Ontology, C. Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **43**, D1049-1056. https://doi.org/10.1093/nar/gku1179 (2015).
21. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
22. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457. https://doi.org/10.1038/nmeth.3337 (2015).
23. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112. https://doi.org/10.1038/nature08460 (2009).
24. Szklarczyk, D. et al. The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368. https://doi.org/10.1093/nar/gkw937 (2017).
25. Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. https://doi.org/10.1101/gr.1239303 (2003).
26. Chin, C. H. et al. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**(Suppl 4), S11. https://doi.org/10.1186/1752-0509-8-S4-S11 (2014).
27. Song, C. et al. DUSP6 protein action and related hub genes prevention of sepsis-induced lung injury were screened by WGCNA and Venn. *Int. J. Biol. Macromol.* **279**, 135117. https://doi.org/10.1016/j.ijbiomac.2024.135117 (2024).
28. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. https://doi.org/10.1038/ncomms3612 (2013).
29. Luhr, R., Cao, Y., Soederquist, B. & Cajander, S. Trends in sepsis mortality over time in randomised sepsis trials: A systematic literature review and meta-analysis of mortality in the control arm, 2002–2016. *Crit. Care* **23**, 1–9 (2019).
30. Chicco, D. & Jurman, G. Survival prediction of patients with sepsis from age, sex, and septic episode number alone. *Sci. Rep.* **10**, 17156 (2020).
31. Kadri, S. S. et al. Estimating ten-year trends in septic shock incidence and mortality in United States academic medical centers using clinical data. *Chest* **151**, 278–285 (2017).
32. Kaukonen, K.-M., Bailey, M., Suzuki, S., Pilcher, D. & Bellomo, R. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000–2012. *JAMA* **311**, 1308–1316 (2014).
33. Esper, A. M. & Martin, G. S. Extending international sepsis epidemiology: The impact of organ dysfunction. *Crit. Care* **13**, 1–3 (2009).
34. Blanco, J. et al. Incidence, organ dysfunction and mortality in severe sepsis: A Spanish multicentre study. *Crit. Care* **12**, 1–14 (2008).
35. Harrison, D. A., Welch, C. A. & Eddleston, J. M. The epidemiology of severe sepsis in England, Wales and Northern Ireland, 1996 to 2004: Secondary analysis of a high quality clinical database, the ICNARC Case Mix Programme Database. *Crit. Care* **10**, 1–10 (2006).
36. Wu, M., Du, X., Gu, R. & Wei, J. Artificial intelligence for clinical decision support in sepsis. *Front. Med.* **8**, 665464 (2021).
37. Wang, W. & Liu, C.-F. Sepsis heterogeneity. *World J. Pediatr.* **19**, 1–9 (2023).
38. Davenport, E. E. et al. Genomic landscape of the individual host response and outcomes in sepsis: A prospective cohort study. *Lancet Respir. Med.* **4**, 259–271 (2016).
39. Liu, D. et al. Sepsis-induced immunosuppression: Mechanisms, diagnosis and current treatment options. *Mil. Med. Res.* **9**, 1–19 (2022).
40. Hamers, L., Kox, M. & Pickkers, P. Sepsis-induced immunoparalysis: Mechanisms, markers, and treatment options. *Minerva Anestesiol.* **81**, 426–439 (2015).
41. Hotchkiss, R. S., Monneret, G. & Payen, D. Immunosuppression in sepsis: A novel understanding of the disorder and a new therapeutic approach. *Lancet Infect. Dis.* **13**, 260–268 (2013).
42. Yu, C. & Huang, Q. Towards more efficient and robust evaluation of sepsis treatment with deep reinforcement learning. *BMC Med. Inform. Decis. Mak.* **23**, 1–10 (2023).
43. Jia, L. et al. *Porphyromonas gingivalis* aggravates colitis via a gut microbiota-linoleic acid metabolism-Th17/Treg cell balance axis. *Nat. Commun.* **15**, 1617. https://doi.org/10.1038/s41467-024-45473-y (2024).
44. de Souza, S. et al. Interleukin-10 signaling in somatosensory neurons controls CCL2 release and inflammatory response. *Brain Behav. Immun.* **116**, 193–202. https://doi.org/10.1016/j.bbi.2023.12.013 (2024).
45. Stirm, K. et al. Tumor cell-derived IL-10 promotes cell-autonomous growth and immune escape in diffuse large B-cell lymphoma. *Oncoimmunology* **10**, 2003533. https://doi.org/10.1080/2162402X.2021.2003533 (2021).
46. Ge, J. et al. IL-10 delays the degeneration of intervertebral discs by suppressing the p38 MAPK signaling pathway. *Free Radic. Biol. Med.* **147**, 262–270 (2020).
47. Hao, S. et al. Identification and validation of a novel mitochondrion-related gene signature for diagnosis and immune infiltration in sepsis. *Front. Immunol.* **14**, 1196306 (2023).
48. Zanders, L. et al. Sepsis induces interleukin 6, gp130/JAK2/STAT3, and muscle wasting. *J. Cachexia Sarcopenia Muscle* **13**, 713–727. https://doi.org/10.1002/jcsm.12867 (2022).
49. Ijaz, M. et al. The role of Grb2 in cancer and peptides as Grb2 antagonists. *Protein Peptide Lett.* **24**, 1084–1095. https://doi.org/10.2174/0929866525666171123213148 (2018).
50. Yablonski, D. Bridging the Gap: Modulatory roles of the Grb2-family adaptor, gads, in cellular and allergic immune responses. *Front. Immunol.* **10**, 1704. https://doi.org/10.3389/fimmu.2019.01704 (2019).
51. Gutierrez-Iniguez, C. et al. Unraveling the non-fitness status of NK cells: Examining the NKp30 receptor and its isoforms distribution in HIV/HCV coinfected patients. *Mol. Immunol.* **172**, 9–16. https://doi.org/10.1016/j.molimm.2024.05.010 (2024).
52. Di Vito, C. et al. Persistence of KIR(neg) NK cells after haploidentical hematopoietic stem cell transplantation protects from human cytomegalovirus infection/reactivation. *Front. Immunol.* **14**, 1266051. https://doi.org/10.3389/fimmu.2023.1266051 (2023).
53. Liu, Q. et al. Mendelian randomization and transcriptomic analysis reveal the protective role of NKT cells in sepsis. *J. Inflamm. Res.* **17**, 3159–3171. https://doi.org/10.2147/JIR.S459706 (2024).
54. Perner, A. et al. Sepsis: Frontiers in diagnosis, resuscitation and antibiotic therapy. *Intensive Care Med.* **42**, 1958–1969 (2016).
55. Georgescu, A. M. et al. Evaluation of TNF-α genetic polymorphisms as predictors for sepsis susceptibility and progression. *BMC Infect. Dis.* **20**, 1–11 (2020).
56. Mirouse, A. et al. Sepsis and cancer: An interplay of friends and foes. *Am. J. Respir. Crit. Care Med.* **202**, 1625–1635 (2020).
57. Xu, L., Hu, G., Xing, P., Zhou, M. & Wang, D. Paclitaxel alleviates the sepsis-induced acute kidney injury via lnc-MALAT1/miR-370-3p/HMGB1 axis. *Life Sci.* **262**, 118505 (2020).

58. Liu, Z., Mahale, P. & Engels, E. A. Sepsis and risk of cancer among elderly adults in the United States. *Clin. Infect. Dis.* **68**, 717–724 (2019).
59. Han, Y., Liu, D. & Li, L. PD-1/PD-L1 pathway: Current researches in cancer. *Am. J. Cancer Res.* **10**, 727 (2020).
60. Nakamori, Y., Park, E. J. & Shimaoka, M. Immune deregulation in sepsis and septic shock: Reversing immune paralysis by targeting PD-1/PD-L1 pathway. *Front. Immunol.* **11**, 624279 (2021).
61. Sari, M. I. & Ilyas, S. The expression levels and concentrations of PD-1 and PD-L1 proteins in septic patients: A systematic review. *Diagnostics* **12**, 2004 (2022).
62. Zhang, T., Yu-Jing, L. & Ma, T. Role of regulation of PD-1 and PD-L1 expression in sepsis. *Front. Immunol.* **14**, 1029438 (2023).
63. Moghbeli, M., Taghehchian, N., Akhlaghipour, I., Samsami, Y. & Maharati, A. Role of forkhead box proteins in regulation of doxorubicin and paclitaxel responses in tumor cells: A comprehensive review. *Int. J. Biol. Macromol.* **248**, 125995 (2023).
64. Wang, L., Zhang, J., Zhang, L., Hu, L. & Tian, J. Significant difference of differential expression pyroptosis-related genes and their correlations with infiltrated immune cells in sepsis. *Front. Cell. Infect. Microbiol.* **12**, 1005392. https://doi.org/10.3389/fcimb.2022.1005392 (2022).
65. Diao, Y. et al. A simplified machine learning model utilizing platelet-related genes for predicting poor prognosis in sepsis. *Front. Immunol.* **14**, 1286203. https://doi.org/10.3389/fimmu.2023.1286203 (2023).
66. Bates, D. W. & Syrowatka, A. Harnessing AI in sepsis care. *Nat. Med.* **28**, 1351–1352. https://doi.org/10.1038/s41591-022-01878-0 (2022).
67. Liang, C. et al. Glucocorticoid therapy for sepsis in the AI era: a survey on current and future approaches. *Comput. Struct. Biotechnol. J.* **24**, 292–305. https://doi.org/10.1016/j.csbj.2024.04.020 (2024).
68. Nicolaou, A. et al. An overview of explainable AI studies in the prediction of sepsis onset and sepsis mortality. *Stud. Health Technol. Inform.* **316**, 808–812. https://doi.org/10.3233/SHTI240534 (2024).
69. Yang, H. S. Machine learning for sepsis prediction: Prospects and challenges. *Clin. Chem.* **70**, 465–467. https://doi.org/10.1093/clinchem/hvae006 (2024).

Acknowledgements

Not applicable.

Author contributions

W.X.: Conceptualization, data curation, formal analysis, writing—original draft, writing—review & editing; Y.Z.: Formal analysis, software; R.X.: Supervision, validation, visualization; F.L.: Methodology, project administration, resources. All authors reviewed the manuscript.

Funding

None.

Declarations

Competing interests

The authors declare no competing interests.

Consent for publication

All listed authors consent to the submission, and all data are used with the consent of the person generating the data.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-93010-8.

Correspondence and requests for materials should be addressed to R.X. or F.L.

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.