JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Original Article

Identification of potential biological targets in sepsis based on random forest and artificial neural network models

Chenghong OuYang[1], Huiwei He[1], Zhiying Lin[1]

[1] *Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang, 330006, China*

Abstract

Background: Machine learning and artificial neural networks are increasingly making new research progress in various human diseases.
Objective: To identify key gene markers related to sepsis based on random forest (RF) and artificial neural network (ANN) models.
Methodology: Key genes (KG) were screened through RF based on GSE9960, GSE13904, and GSE26440 datasets, and the diagnostic model of core genes was constructed using ANN analysis, which was validated.
Results: GSE9960, GSE13904, and GSE26440 were merged as the training set, and 452 differentially expressed genes (DEGs) were identified. Then, GO and KEGG enrichment analysis, as well as Cytoscape module analysis, were used to explore the mechanisms related to sepsis. Subsequently, these DEGs were further screened through RF to obtain key genes *UPP1*, *ZNF600*, *GPR160*, *TRAT1*, *C3AR1*, *NAIP*, *MAL*, and *F5*. An ANN method was used to construct a screening model. The AUC metrics obtained from the training and validation cohorts were quantitatively assessed through ROC curve analysis, yielding discriminative performance values of 0.969 and 0.792, respectively. The CIBERSORT-based immune profiling revealed a marked depletion of lymphocyte populations (including B cells and T cells) in the experimental cohort than healthy controls, with regulatory T cell subsets demonstrating a paradoxical increase in abundance. Sepsis immune suppression was the dominant process.
Conclusions: The core-gene model constructed by the RF and ANN method is a robust method for predicting, diagnosing, or treating sepsis, with nice predictive accuracy. It may be used for sepsis-related medical decisions and provide new entry routes for the treatment or prognosis improvement of sepsis.

Key words: sepsis; artificial neural network; random forest; biomarker discovery.

J Infect Dev Ctries 2026; 20(4):537-550. doi:10.3855/jidc.21442

(Received 11 February 2025 – Accepted 04 August 2025)

Introduction

Sepsis is conceptualized as a potentially fatal multiorgan system compromise arising from dysregulated host defense mechanisms against pathogenic invasion. This condition transcends mere manifestations of a generalized inflammatory cascade or immunological derangement, reflecting instead a multidimensional pathophysiological syndrome characterized by cross-organ functional perturbations. The disease pathogenesis manifests multifaceted pathobiological complexity, encompassing interrelated mechanisms extending beyond the inflammatory-immune axis dyshomeostasis to include mitochondrial dysfunction, coagulopathy, neuroendocrine-immune axis perturbations, endoplasmic reticulum stress responses, and autophagic flux abnormalities—collectively driving the progression of systemic organ failure [1].

Sepsis, a potentially life-endangering state emerging from the body's disordered reaction to infection, impacts more than 30 million individuals globally on an annual basis. It represents one of the most significant contributors to mortality in critically ill patient populations across the world [2]. Research indicates that any individual who has been infected is at risk of developing sepsis. Moreover, the incidence of sepsis accounts for as much as 1-2% among all hospitalized patients [1]. Although medical technologies such as anti-infection treatment have made great progress in recent years, the incidence rate of sepsis continues to rise. In addition, even following treatment, patients with sepsis are still confronted with long-term and grave issues like physical, psychological, and cognitive impairments [3,4]. The non-specific symptoms of sepsis patients can also lead to delayed diagnosis and intervention [5]. Early and accurate diagnosis and treatment are key points to reducing the risk of sepsis and improving the prognosis of patients, while specific biomarkers are helpful for timely diagnosis, monitoring, severity grading, and predicting clinical outcomes of sepsis. Optimal biomarkers are expected to exert an influence on multiple sepsis-related pathways. Clinically, they can be utilized to facilitate the formulation of novel treatment strategies and aid in the diagnosis of sepsis [6], as recent reviews highlight the expanding role of biomarkers beyond

traditional inflammatory indicators in guiding sepsis management. Early identifying and treating sepsis is a highly complex and multifaceted challenge, while artificial intelligence is gradually replacing humans in some of its medical decisions to improve clinical practice and patient prognosis [7]. A number of studies have made efforts to construct machine - learning models aimed at the early detection of sepsis risk, such as the random forest (RF) model, which is considered a valuable tool for predicting sepsis [7,8]. The evaluation of rapidly measured gene expression characteristics beyond the current standards will provide clinicians with diagnostic or therapeutic value [9]. The large amount of data generated through transcriptomics will offer us fresh hints for a more profound comprehension of the pathology of sepsis and the identification of sepsis biomarkers. Polymerase chain reaction (PCR) detection technology makes transcription-level biomarkers more convenient, rapid, and inexpensive [6]. Based on partially automated qPCR methods, it is possible to quickly and efficiently evaluate sepsis-related gene biomarkers with diagnostic value.

In this study, several data sets in the GEO database were downloaded and integrated into a gene expression matrix. RF as well as Artificial Neural Network (ANN) models were employed to sift out the key genes (KG) associated with sepsis. Followed by CIBERSORT analysis, and finally, the drug sensitivity analysis related to KG was carried out based on the CellMiner database.

Methodology

Research design

Datasets GSE9960 (blood, H. sapiens, GPL570), GSE13904 (whole blood, H. sapiens, GPL570), GSE26440 (blood, H. sapiens, GPL570), and GSE64457 (whole blood, H. sapiens, GPL570) were retrieved from the Gene Expression Omnibus (GEO) database (Table 1). GSE9960, GSE13904, and GSE26440 were merged as the training set, and GSE64457 served as the independent validation cohort. These datasets were first employed to screen for differentially expressed genes (DEGs). Subsequently, the identified DEGs were subjected to gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) analyses. Finally, RF classification was applied to screen for key genes, which were then used for ANN model development and verification. According to these "important genes", after calculating gene weights, an ANN model is developed and further verified. Finally, the data were consulted to analyze the infiltration of immune cells, according to the important genes obtained by RF analysis and ANN model development and verification. All the statistical data utilized are computed by R software (version 4.3.3). The overall research design and experimental process are illustrated in Figure 1.

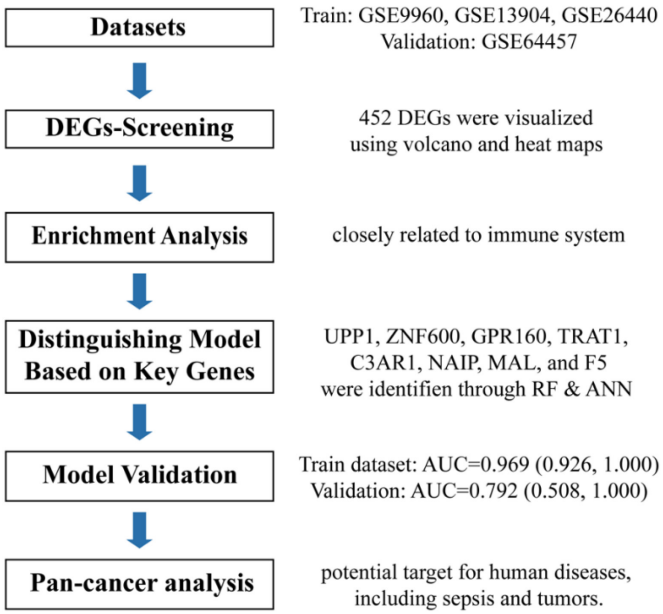
Table 1. The information of selected datasets.

Application / Datasets	Group Information	
	Normal OR Under Sepsis	Sepsis
Training model		
GSE9960	16	54
GSE13904	18	118
GSE26440	32	98
Verifying model		
GSE64457	8	15

Data downloading and processing

The datasets listed below were retrieved from the GEO database: GSE9960, GSE13904, GSE26440, and GSE64457, to obtain gene expression profile information (the detailed information can be seen in Table 1). The information in the dataset was annotated using the relevant platform file (GPL570). The first three datasets are used as training sets to construct the diagnostic model, and the fourth dataset is used as the validation set. In order to adjust batch effects, the 'removeBatchEffect' function within the R package 'Limma' is employed. Subsequently, principal component analysis (PCA) is utilized to conduct an evaluation of these adjusted batch effects.

Figure 1. Flow chart of overall research design and experimental process.



Sepsis-related DEGs and functional enrichment analysis

Differential expression profiling was conducted in the integrated training cohort using the Limma package within the R statistical environment, comparing transcriptional patterns between experimental and control groups. Stringent screening parameters were established for identifying differentially expressed genes (DEGs), requiring |log2FC| > 1.5 combined with statistical significance ($p < 0.05$). Subsequent hierarchical clustering analysis of identified DEGs was visualized through heatmap representations generated by the pheatmap package. To elucidate the functional roles and pathway associations of these DEGs in sepsis pathophysiology, comprehensive enrichment analyses were executed. This included simultaneous investigation of GO classifications and KEGG pathway databases. The analytical workflow incorporated specialized bioinformatics tools, including clusterProfiler (version 4.1.3) [10], ggplot2, and enrichplot packages, with statistical significance thresholds maintained at $p < 0.05$ for all enrichment evaluations.

Metascape analysis of differentially expressed genes and construction of PPI network

Metascape (http://metascape.org) serves as an accessible and dependable instrument for conducting functional enrichment analysis [11], with established utility in disease biomarker studies across diverse pathologies, including virus-associated cancers and infectious conditions [12]. Similarly, GO/KEGG analysis was performed using clusterProfiler as previously validated in immunological transcriptomic studies [10,13]. STRING (https://string-db.org/cgi/input.pl) has been extensively utilized for the construction of protein-protein interaction (PPI) networks [14]. Visualize the PPI network using Cytoscape (v3.9.0), then use MCODE to identify module hub genes, followed by functional analysis of the biological processes in which module hub genes are primarily involved

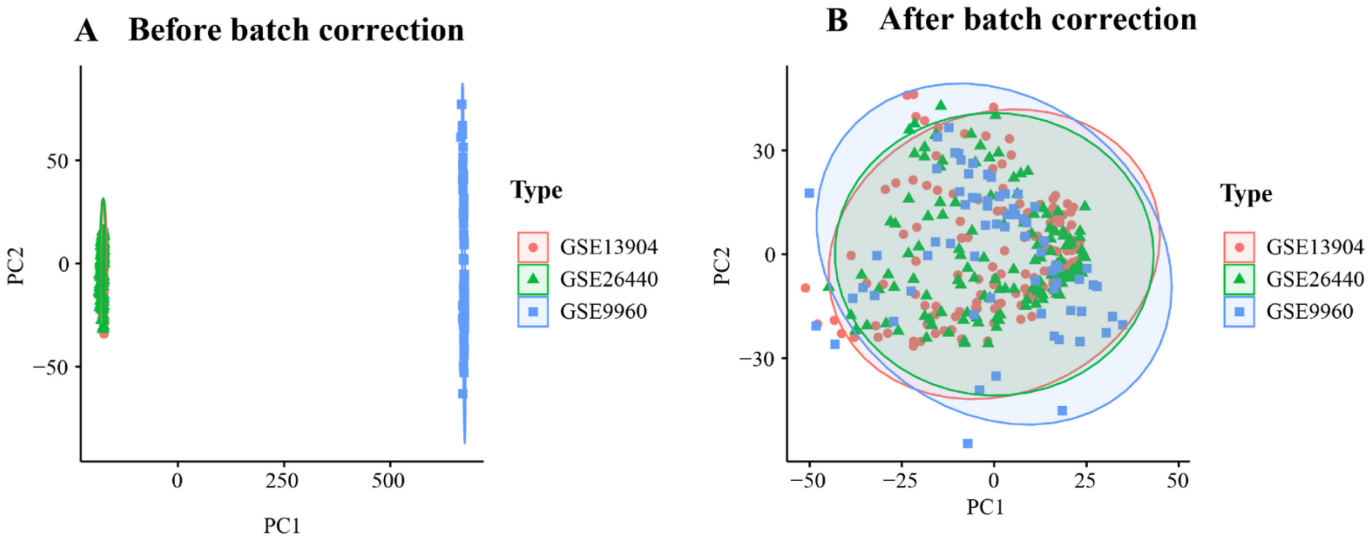
Model Gene Screening via a Random Forest Model

The feature selection of candidate biomarkers was performed through the implementation of an RF algorithm employing version 4.6.14 of the randomForest package in R. Variable importance evaluation incorporated the Gini impurity metric, which quantifies each feature's contribution to reducing node heterogeneity throughout the classification tree architecture. Initial model configuration specified 500 ensemble decision trees with a predefined randomization state (seed = 123456) to ensure computational reproducibility. Based on the parameters selected by the Gini coefficient method above, select model genes with MeanDeceaseGini greater than 2 for subsequent model construction and validation. Hierarchical clustering of selected candidate biomarkers was performed using the pheatmap package in R, with results visualized through a heatmap representation.

Model Training and Validation of Artificial Neural Networks

Based on the feature model genes obtained through

Figure 2. Data preprocessing using principal component analysis (PCA) for batch calibration of GSE9960, GSE13904, and GSE26440. (A) Before batch calibration and (B) After batch calibration.



RF screening, the merged training set is standardized according to the median principle, and then, using the R package "neuralnet", an ANN model is built based on these significant model variables. As a model parameter, the quantity of hidden neuron layers is configured to 5 neurons in a single hidden layer. The constructed ANN model's accuracy was validated using another dataset, GSE64457. The R packages "pROC" and "ggplot2" were employed to compute the AUC results of the constructed model and draw the ROC curve.

Sepsis-related immune cell subtype CIBERSORT analysis

CIBERSORT can be used to characterize cellular heterogeneity in almost any tissue, and when used in conjunction with statistical filtering, the analysis results of CIBERSORT are more reliable [13]. In the present study, the R package CIBERSORT was utilized to determine the distribution of 22 immune cell types in the sepsis patient group for further analysis.

Drug sensitivity analysis

The CellMiner database, accessible at https://discover.nci.nih.gov/cellminer/home.do/, functions as an online platform for performing an analysis of the correlation that exists between gene expression and drug sensitivity.

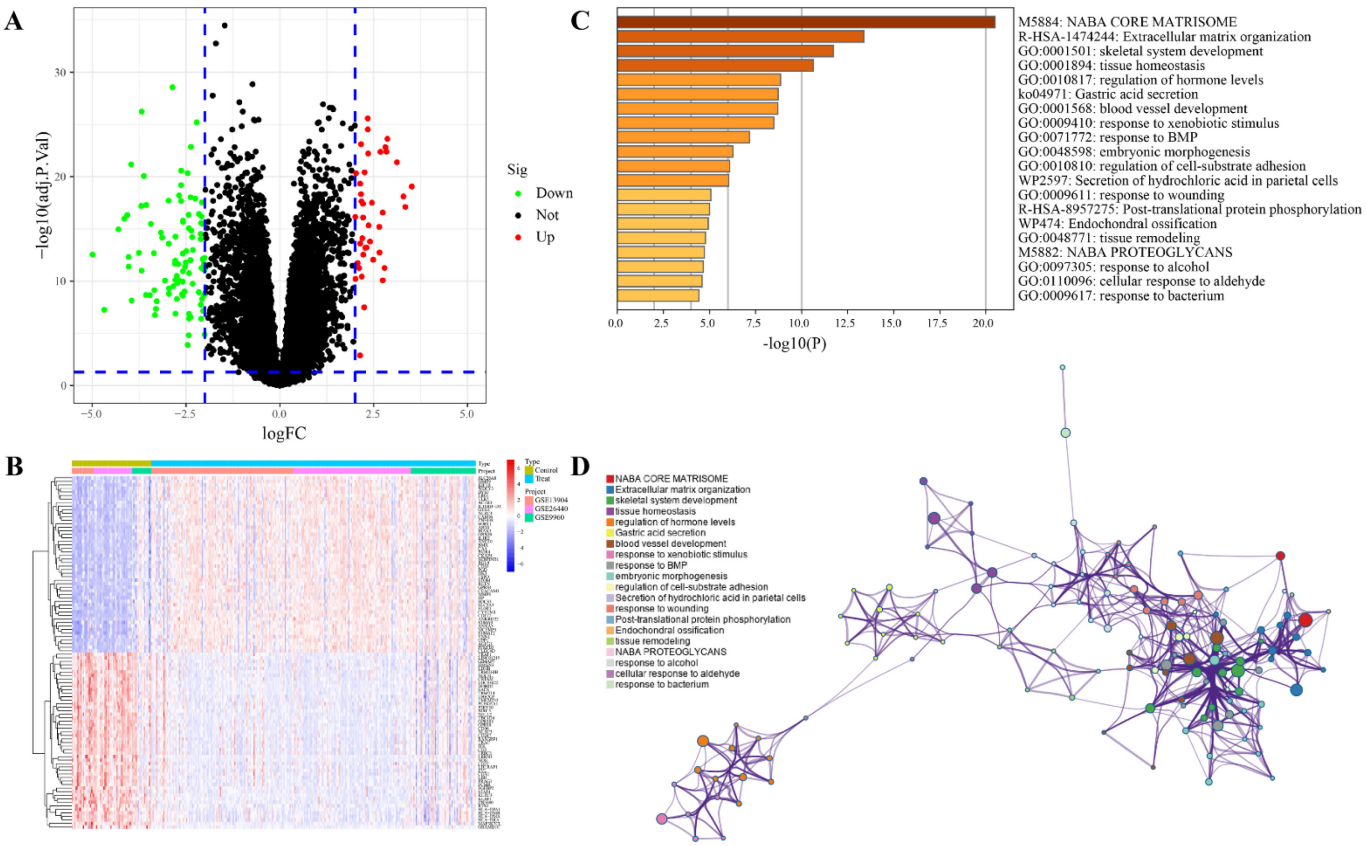
Results

DEGs screening between Sepsis patients and normal patients

Sample information of sepsis patients and normal individuals was obtained from the datasets GSE9960, GSE13904, GSE26440, and GSE64457, and gene expression profile information was further extracted based on the original grouping of the datasets. The initial three datasets are combined to form the training set, while the final dataset is utilized as the validation set to assess the model's performance. To make the analysis results more reliable, we conducted batch correction analysis before merging the first three datasets. The PCA analysis results showed that there was heterogeneity in the dataset before merging (Figure 2A), and direct merging was not considered. Therefore, the batch correction method was adopted. The corrected PCA graph (Figure 2B) showed no obvious clustering between the data. Subsequent analysis can continue.

Further comparison of gene expression profiles among the case group and the control group was

Figure 3. DEGs analysis. A-B: Differential gene screening; Metascape analysis of C-D: DEGs.

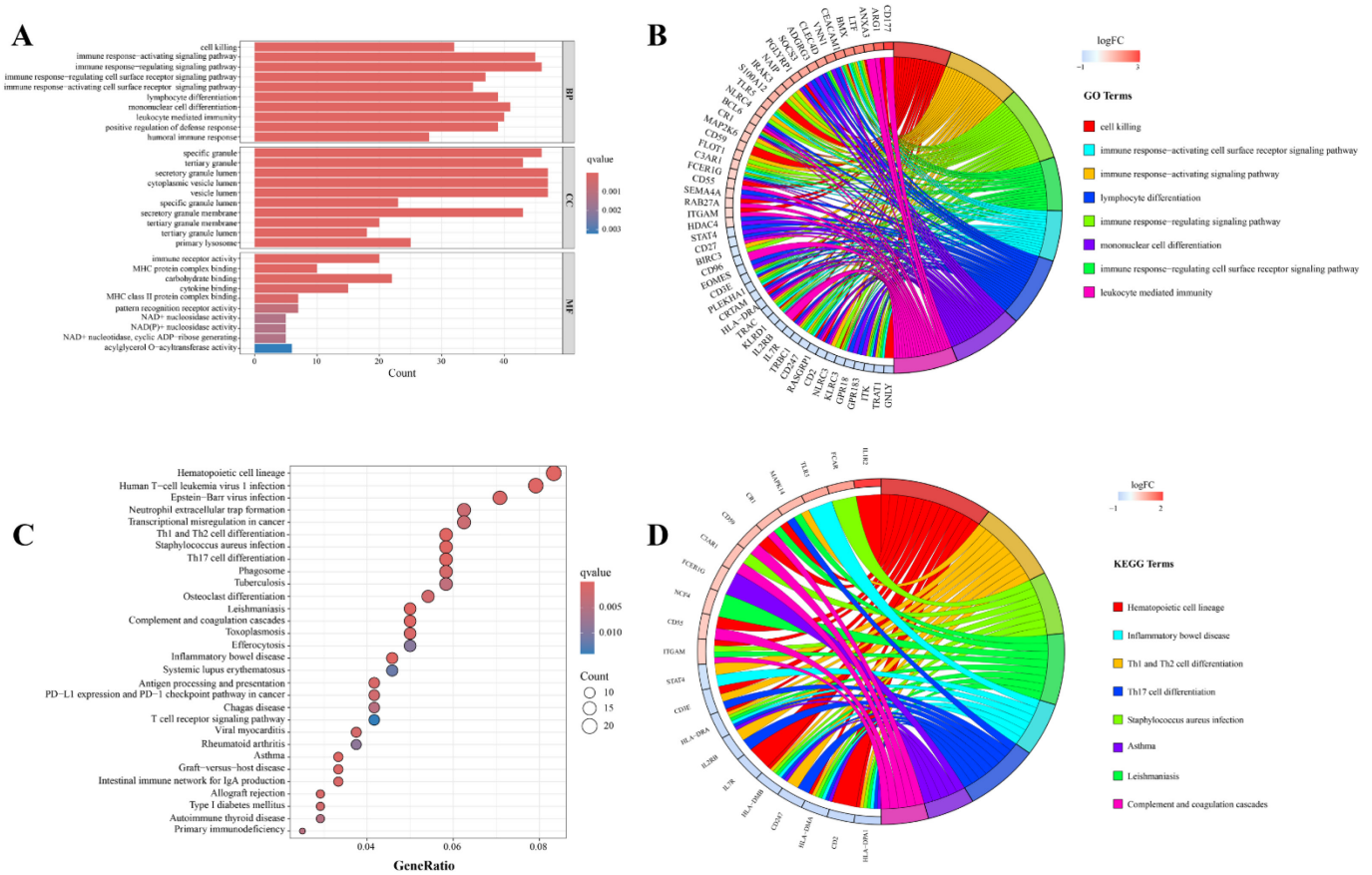


conducted to identify significantly DEGs for further analysis. The results were visualized using volcano and heat maps (Figure 3A and B), and a relatively large quantity of differentially expressed genes was discovered, amounting to a total of 452 DEGs. Clearly, these genes exert a significant influence on the differences between the case group and the control group. However, at the specific level of each gene, given the numerous functions of each gene, it is difficult to distinguish the interconnections and interactions between genes. For example, through Metascape analysis, we found that there are at least 20 possible pathways through which these DEGs may take effect (Figure 3C), and the network connections between genes are tight and varied (Figure 3D). Therefore, it is necessary to proceed to the next level of analysis.

Functional enrichment analysis of DEGs

In order to explore more profoundly the biological significance of the 452 DEGs in the pathogenesis of Sepsis, conventional GO and KEGG enrichment analyses were carried out on these DEGs with the utilization of R software. Through the GO enrichment analysis, 511 annotations were acquired, consisting of 401 BP annotations, 60 CC annotations, and 50 MF annotations. The GO enrichment analysis results show that Sepsis is closely related to some functions and molecules of the immune system (Figure 4 A-B). The enrichment of relevant BP includes some biological processes related to immune responses, such as cell killing, biochemical pathways that induce immune cell activation and response, Signal transduction pathways that are engaged in the regulation of immune responses, and those that participate in the regulation of immune responses via cell-surface receptors, differentiation of lymphocytes and monocytes, and positive regulatory mechanisms of defense responses. The enrichment of CC mainly involves the internal substances, structures, and functions of immune cells, including specific granules, tertiary granules, secretory granule cavities, cytoplasmic vesicle cavities, specific granule cavities, and primitive lysosomes. The enrichment of MF mainly describes the role of different molecules and enzyme activities in cellular function in the immune system, including immune receptor activity, binding activity of major histocompatibility complex (MHC) protein

Figure 4. Functional correlation analysis. Analyze DEGs through GO and KEGG. The results of A and B: GO are represented by bar charts and circle charts. The results of C and D: KEGG are displayed in bubble and circle plots.

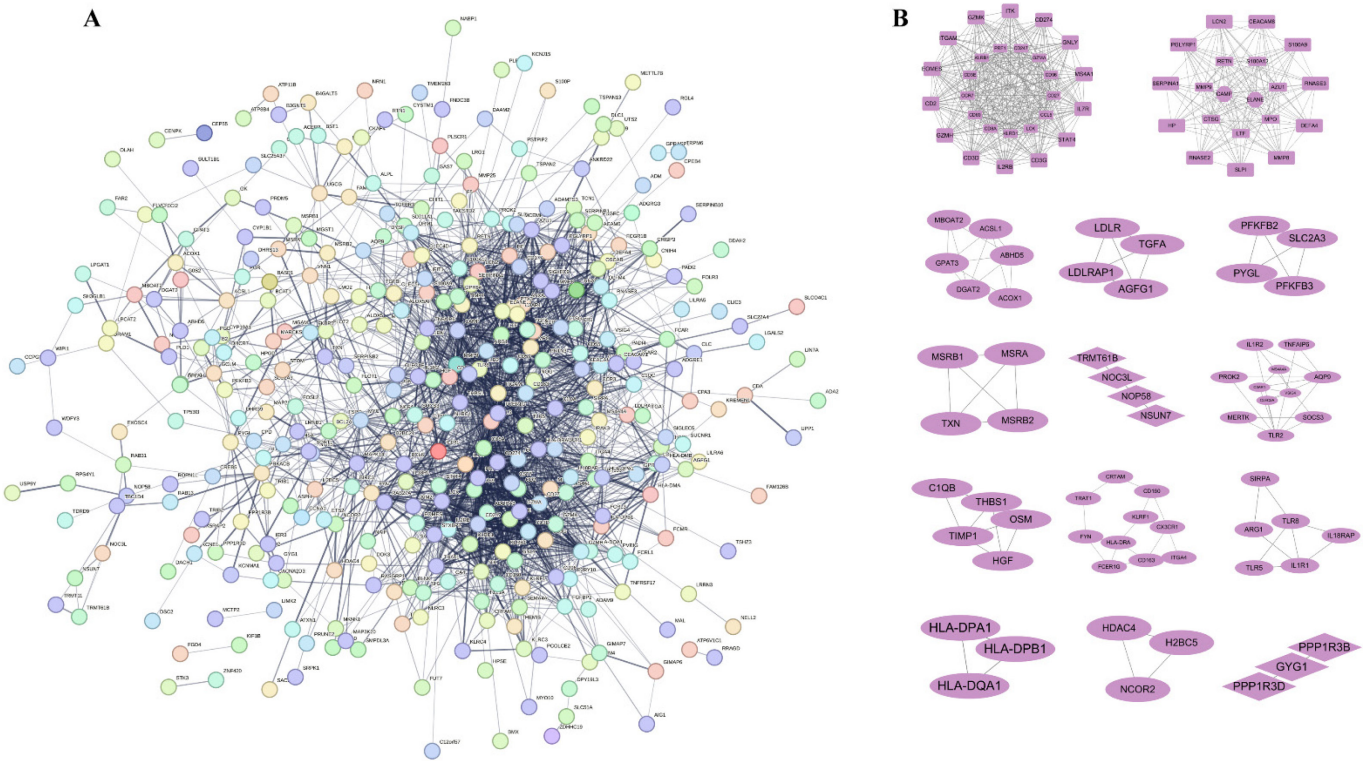


complexes, carbohydrate binding activity, cytokine binding, pattern recognition receptor activity, NAD (P) + nucleosidase activity, MHC class II protein complex binding, NAD + nucleosidase activity, and acylglycerol acyltransferase activity. The KEGG analysis enriched 34 pathways (Figure 4 C-D), which are mainly enriched in pathways related to biological processes or pathological mechanisms such as the immune system and infection, including hematopoietic cell lineage, complement and coagulation cascade reactions, Th1 and Th2 cell differentiation, inflammatory bowel disease, primary immunodeficiency, *Staphylococcus aureus* infection, and T cell receptor signaling pathways.

PPI network module analysis

To elucidate protein-level interactions among the identified DEGs, we constructed a PPI network using STRING (Figure 5A). Cytoscape analysis revealed 11 highly interconnected MCODE modules (Figure 5B), representing functional subnetworks within the sepsis-associated interactome. Functional annotation of these modules (Supplementary Table 1) demonstrate their involvement in diverse pathophysiological processes, ranging from substance metabolism to cellular structure, for instance, the facilitation of the transfer of an acyl group to an oxygen atom on the acylglycerol molecule [GOC:ai], the promotion of the transfer of acyl groups from an acyl - coa to a lysophospholipid [GOC:cjk], and any minute, fluid - filled, spherical organelle encapsulated by a membrane [GOC:mah, GOC:pz, GOC:vesicles]. etc.; From the organism internal environment to immune system related processes, such as Oxidative stress response [GOC: jl], Protein repair, Catalysis of an oxidation reduction (redox) reaction [GOC: ai], An immune response mediated through a body fluid [GOC: hb, ISBN: 0198506732], and Any process that modulates the frequency, rate, or extent of production of a cytokine [GOC: add, ISBN: 0781735149], etc.; From the regulation of genetic material to the transmission of cellular signals, such as The cellular synthesis of microrna (mirna) transcripts [GOC: dph, GOC: kmv], Any procedure that modulates the frequency, speed, or degree of the transcription of a non - protein - coding gene from a DNA template [GOC: pg]. This includes the transcription of non - protein - coding RNA from a DNA template, as well as the Notch HLH transcription pathway [GOC: dos]. Additionally, it encompasses the process of a cell receiving a signal and relaying it within the cell to trigger an immune response. As defined in Wikipedia, this is related to immune receptors, among other aspects. All gene modules comprehensively involve different aspects and processes of physiological functions or pathological diseases in the human body.

Figure 5. MCODE analysis and module function analysis of PPI network related to DEGs



Construction and validation of diagnostic models using RF and ANN

Regarding the connection between the model error and the quantity of decision trees (Figure 6A, B), 452 DEGs were introduced into an RF classifier. We chose 500 trees as the parameters for the final model, and the results indicated that the model had stable errors. Subsequently, we determined 8 model genes with MeanDeceaseGini > 2 as candidate genes for the subsequent analysis. Based on these 8 crucial variables, k-means unsupervised clustering was carried out on the merged dataset. The results, as shown in Figure 6C, demonstrated that these 6 genes could be utilized to differentiate disease samples from normal samples within the merged dataset. Among them, the ZNF600, TRAT1, and MAL genes are expressed at a low level in normal samples but are highly expressed in disease samples. The remaining genes, namely *UPP1*, *GPR160*, *C3AR1*, *NAIP*, and *F5*, have higher expression in normal samples and lower expression in disease samples. Afterwards, we further validated the 8 model genes selected by the RF using an ANN and constructed the diagnostic model, in which, as model parameters, 5 neurons in the hidden layers were configured. Based on

Figure 6. RF and ANN models used for sepsis estimation. A: The relationship between the overall error rate of RF classification and the number of trees in the case group and control group. The black curve represents the overall misclassification rate of the model, the green curve represents the misclassification rate when classifying the case group, and the red curve represents the misclassification rate when classifying the control group. B: The Gini index ranking of RF screening results. C: The distribution of key genes in the case group and control group based on RF Gini coefficient screening in the heatmap. D: A sepsis related key gene discrimination model based on ANN, with 8 neurons in the input layer, 5 neurons in the hidden layer, and 1 neuron in the output layer.

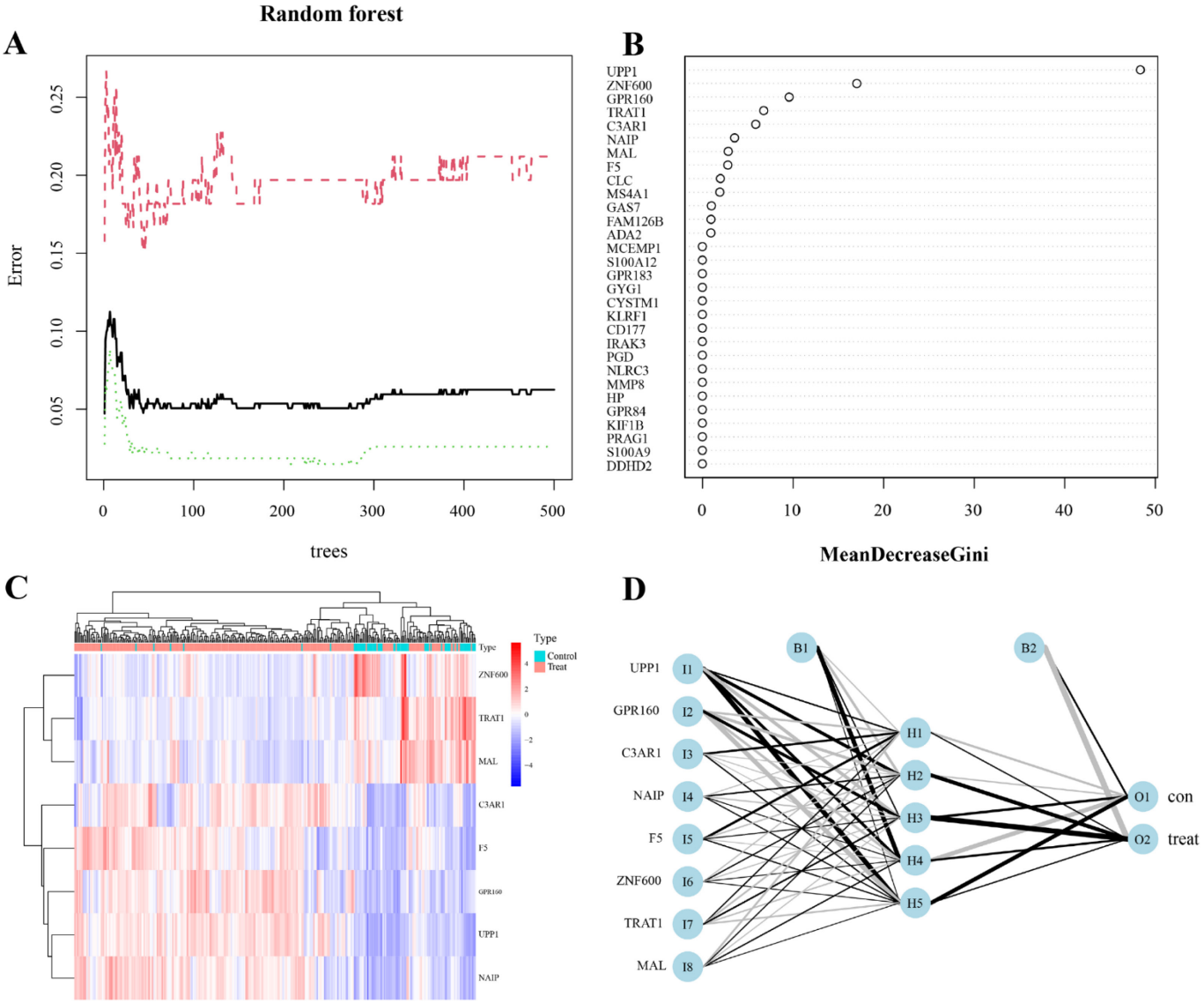
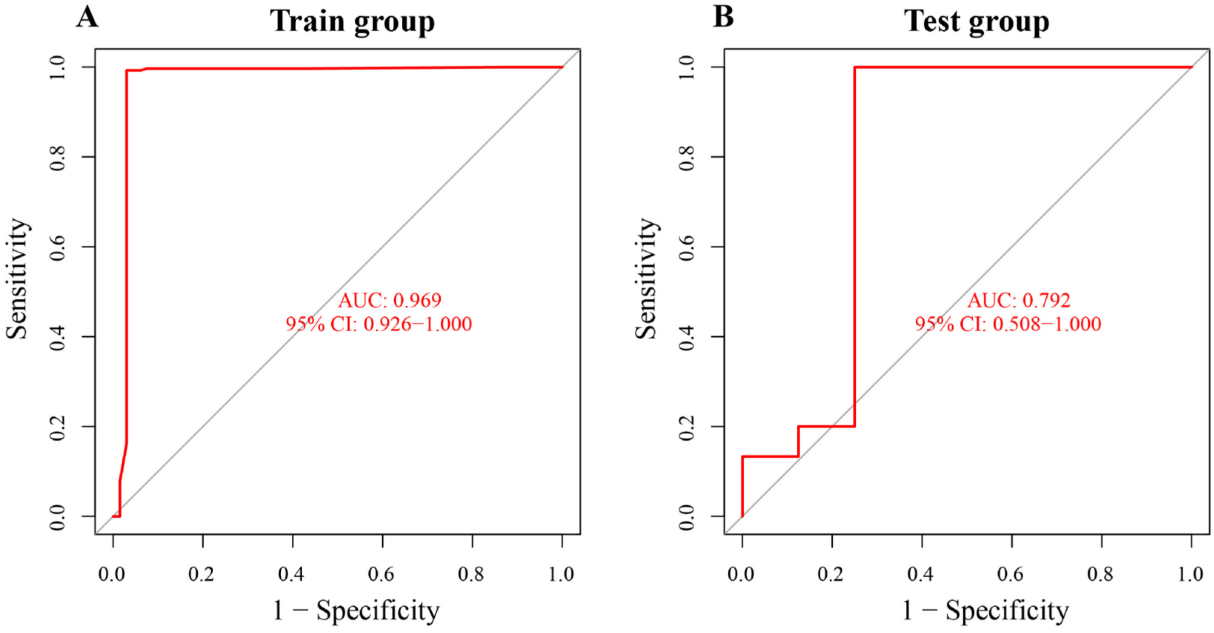


Figure 7. The diagnostic performance of RF combined with ANN in constructing the sepsis discrimination model. ROC curve analysis of (A) training (AUC = 0.969) and (B) validation (AUC = 0.792) sets.



the output outcomes of the ANN model (Figure 6D), the predicted weights of each hidden neuron layer are -1.16083, 0.53297, -1.09021, 0.03219, and -0.00965, respectively, and the weights of the classification output layer are 0.17423 and 0.05769, respectively.

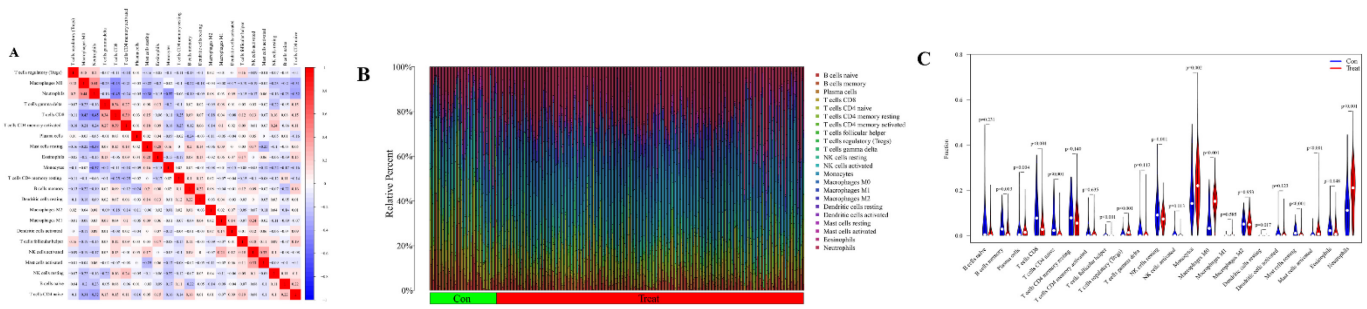
ROC validation of model performance

Then, to assess the model's ability to differentiate between normal samples and sepsis samples, ROC curves were plotted. The Area Under the Curve (AUC) values for the training set and the validation set were 0.969 ($p < 0.001$) and 0.792 ($p = 0.003$), respectively (Figure 7). A greater AUC value implies a higher degree of reliability for the model. Therefore, the sepsis identification model constructed through RF and ANN in this study has nice sensitivity and specificity, and has a reliable theoretical basis as marker genes for sepsis in the future.

CIBERSORT analysis

CIBERSORT analysis were conducted to calculate the deconvolution of immune cell infiltration for each tissue sample in the training set, and explore sepsis-related immune cell infiltration. Subsequently, the correlations among the 22 immune cells in the tissues of sepsis patients were assessed (Figure 8A). For example, there is a positive correlation between M0 macrophages and neutrophils. Moreover, $CD8^+$ T cells show a positive correlation with both γδ T cells and $CD4^+$ T cells. A positive correlation exists between eosinophils and resting mast cells. Conversely, $CD8^+$ T cells have a negative correlation with M0 macrophages as well as neutrophils. Secondly, the composition of the 22 immune cells within each sample was presented in

Figure 8. Immune cell infiltration in Sepsis patients and normal tissues. A: Evaluate the correlation of 22 immune cells in Sepsis synovial tissue. Red represents positive correlation; blue represents negative correlation. B: A bar chart showing the types and relative percentages of immune cells in each sample, with different colors representing different cell types. C: Comparison of infiltration of each type of immune cell between the case group and the control group.



the form of a percentage-based bar graph (Figure 8B). Distinct colors signify the proportion of various immune cells in each sample, with the cumulative total amounting to 1. The outcomes indicated that when compared with the control group, the proportion of monocytes and macrophages of the three types (M0, M1, and M2) in the case sample was higher, and they were the main infiltrating immune cells. The Wilcoxon test was used to statistically analyze the different infiltration extent of 22 types of immune cells in the case group and the control group. The results were visualized as the violin plot (Figure 8C); 14 immune cells had significantly different infiltration patterns in the case group and the control group ($p < 0.05$). The number of memory B cells, initial B cells, plasma cells, initial CD4+T cells, follicular helper T cells, CD8+T cells, γδ T cells, dormant NK cells, dormant dendritic cells, dormant mast cells, activated NK cells, and eosinophils were significantly reduced in the case group, while the number of monocytes, regulatory T cells, M0 macrophages, neutrophils and activated mast cells were significantly elevated in sepsis organization. And this phenomenon is closely consistent with the activation of inflammatory pathways mentioned in previous studies, such as Kanashiro et al.'s report that Bacterial invasion has the potential to directly or indirectly trigger multiple inflammatory pathways, which eventually culminate in sepsis [15]. The principal mechanisms implicated in these inflammatory pathways entail the excessive activation of neutrophils, monocytes, and macrophages. This is succeeded by the release of cytokines and pathogenic components like lipopolysaccharides. These components interact with Toll-like receptors on monocytes, thereby inducing signaling cascades and facilitating the release of pro-inflammatory cytokines, for example, TNF-α and IL-1 [16]. It was indicated that an immune system imbalance has an important impact on the occurrence of sepsis, and the distribution spectrum of immune cell infiltration may become a specific distribution spectrum marker for sepsis diagnosis, and it is also expected to quickly treat sepsis by controlling immune infiltration through these model genes.

Drug sensitivity analysis

CellMiner database results suggested that TRAT1 expression had a positive correlation with Nelarabine, Chelerythrine, XK-469, Fluphenazine, Dexamethasone Decadron, Asparaginase, and PX-316, suggesting that these drugs inhibited tumors (Figure 9). C3AR1 expression had a positive correlation with Denileukin Diftitox Ontak, Imexon, Isotretinoin, Carmustine, and Estramustine, and negatively correlated with Irofulven (Figure 9). NAIP expression had a positive correlation with Nelarabine and Chelerythrine (Figure 9G and Figure 9L). F5 expression had a positive correlation with Elesclomol (Figure 9N). Therefore, these findings could be beneficial for drug screening and personalized treatment.

Discussion

Biomarkers refer to biological molecules or characteristics that serve as indicators of physiological or pathological processes. Measuring the levels of a biomarker within a particular biological sample, be it in whole blood, plasma, serum, cell fluid, or any other form, has frequently been demonstrated to be of great importance in clinical diagnosis. Currently, a vast amount of research has been carried out to identify biomarkers for sepsis. The primary goals are to comprehend their influence, clinical implications, detection methods, and quantification mechanisms. However, as noted in recent reviews [17], the translation of most candidate biomarkers from bench to bedside remains limited, underscoring the need for more robust validation in clinical cohorts. Diagnosing sepsis at the early stages of its onset and development is of utmost importance. This implies that sepsis patients have a higher chance of receiving effective treatment and have a high chance of recovery. Simultaneously detecting multiple biomarkers for optimal diagnosis is also welcomed by everyone.

In the past decade, there has been a leap in the research of biomarkers for sepsis. Biomarkers commonly utilized in clinical diagnosis and serving as representatives are PCT and C-reactive protein (CRP). These are the most commonly used biomarkers that can be quickly detected at low cost through the use of blood samples. The sensitivity and specificity of these markers in the diagnosis of sepsis are restricted. Moreover, infections caused by specific pathogens, like viruses, and other non-infectious factors, such as surgery and immunotherapy, can also have an impact on the levels of CRP and PCT [17], highlighting the need for more pathogen-specific or pathway-targeted biomarkers in sepsis diagnosis. At present, the use of peripheral blood can detect transcriptional biomarkers at the mRNA level for diagnosing sepsis or reflecting the severity and prognosis of sepsis, while the development of transcriptomics provides a systematic approach for promoting precision medicine strategies [18]. Accordingly, exploring the potential utility of gene expression biomarkers in enhancing existing measurement tools in bedside and laboratory

environments is a promising approach. These biomarkers may help predict and support precise management strategies for sepsis, ultimately achieving the goal of changing the risk profile associated with this condition [18].

In the present study, we endeavored to identify diagnostic biomarkers specific to sepsis and to probe into the correlations between these biomarkers and tissue immune cell infiltration in sepsis patients. Numerous studies have shown that a significant portion of sepsis-related deaths are due to unresolved opportunistic infections and immunosuppressive features [16-18]. Sepsis can be regarded as a "contest" occurring between the infection and the host's immune response, where pathogens attempt to gain an advantage by obstructing various aspects of host immunity. However, sepsis can directly affect the immune system, affecting the formation, maturation, function, and death of immune cells [19].

Many reports indicate that there is a clear association between the *C3AR1*, *NAIP*, *MAL*, and *F5* hub genes and the sepsis process to some extent, and in some cases, this association is significant [20], aligning with clinical guidelines that emphasize the integration of multiple biomarkers to improve sepsis diagnostic accuracy [21-23]. *UPP1* (uridine phosphorylase 1) is very important for various metabolic functions of the body, especially for uridine metabolism [24]. Uridine metabolism is a series of finely controlled processes closely regulated by a series of enzymes, playing an important role in regulating pathological and physiological processes such as genetic material and glycogen synthesis. It has anti-inflammatory, anti-fibrotic, antioxidant, and anti-aging functions. Its metabolic disorders are also related to the progression of a series of diseases, including sepsis and sepsis-induced acute lung injury [25]. ZNF600 belongs to the zinc finger protein family, which typically binds to DNA and plays a role in gene expression regulation. Its specific biological function is not yet clear, but research reports suggest that zinc finger proteins may play a role in regulating gene expression in immune cells, particularly in cellular stress responses or inflammatory processes [26,27]. The research on zinc finger proteins and sepsis is also constantly being explored, and different zinc finger proteins are intimately associated with the developmental process of sepsis [28,29]. G protein-coupled receptors (GPCRs), as a widely present key family of cell membrane proteins, are important mediators of transmembrane signal transduction. After binding to ligands, they exhibit significant ability to activate downstream signaling pathways and participate in numerous physiological and pathological activities across mammals [30,31]. Currently, a large number of studies have discovered that GPCRs are frequently implicated in neurological pain. Moreover, neurotransmitters and regulatory molecules present in the spinal cord as well as peripheral tissues have been demonstrated to exert a vital role in the modulation of both acute and chronic pain. They achieve this by exerting diverse impacts on neurons, immune cells, and the nervous system. GPCRs include multiple categories [32], among which GPR160 belongs to the class A of rhodopsin-like receptors. GPR160 plays a significant regulatory role in sepsis-related inflammation and immune-related pain, which is consistent with recent research results that "GPR160 has a significant role in signal transduction of neuropathic pain" [33]. The complexity and diversity it possesses may have significant implications for the efficacy of sepsis therapy drugs. TRAT1, which stands for T cell receptor-associated transmembrane junction protein 1, is a gene associated with T cells. The T cell receptor (TCR) identifies antigen fragments in the form of peptides that are bound to MHC molecules. A deficiency in MHC has the potential to result in immune dysfunction. As a consequence, TRAT1 holds a significant position in the immune process [34]. This suggests that TRAT1 is associated with abnormal immune function in sepsis patients, which not only provides a reference value for sepsis diagnosis but also provides new research entry points for specific treatment of patients.

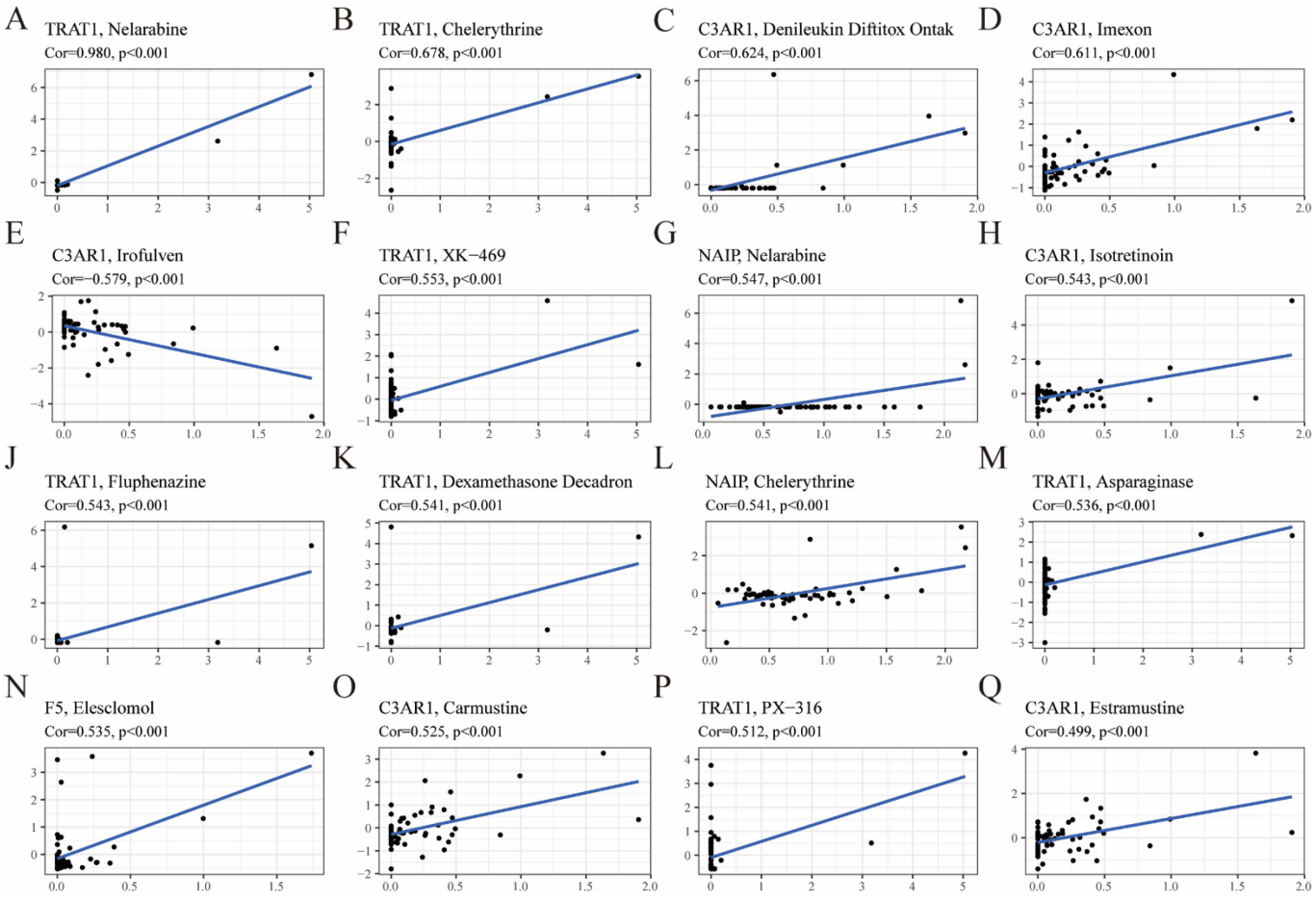
Through in-depth analysis of model genes, we found that it is consistent with the GO and KEGG enrichment results, indicating that model genes mainly affect the physiological and pathological processes of sepsis patients by influencing the body's immune system and inflammatory response. This provides us with specific directions for clinical decision-making. This work additionally furnishes a theoretical foundation for the clinical utilization of model genes. In light of this, we employed the CIBERSORT program to examine the immune infiltration process taking place during sepsis. The aim was to obtain a more profound comprehension of the influence of immune system imbalance on sepsis patients. The results indicate that various changes in immune cell infiltration may be related to the occurrence and progression of sepsis. In the fight against infection, the first line of defense begins to play a part. The T and B cells of the adaptive immune system undergo antigen-specific reactions, such as the activation of CD4 T cells, which polarizes these cells into specific helper T cell (Th) subgroups.

Each subgroup initiates CD8 T cell activation, produces pro-inflammatory cytokines, etc., to further engulf and eradicate microorganisms and fight inflammation [35-37]. Just like T cells, B cells too can experience imbalance and depletion as the infection progresses. In the context of sepsis, these processes are severely deranged. As a result, the cells of the adaptive immune system are incapable of generating proper defense responses against the infection [38]. The surface of γδ T cells has a unique TCR, which is dedicated to maintaining the immune homeostasis of the lung and intestinal epithelium, preventing invading pathogens, and thus preventing pneumonia and intestinal infections [39]. Regulatory T cells are of vital importance in the regulation of immune cells within the body's homeostasis and disease-related issues. Their main function is to inhibit the overactive immune responses that are initiated by other cells of the adaptive immune system when reacting to infections, and they also work to dampen inflammation [40].

Zheng *et al.* [41] explored specific gene sets for bacterial and fungal sepsis as biomarkers, using DEG analysis and GSVA scoring. This method involves gene screening and diagnostic assessment, but does not use RF for feature selection; instead, it is directly based on differential expression analysis. Nor was any neural network model mentioned for prediction. Unlike prior studies using single-model approaches, our work integrates RF for robust gene selection and ANN for diagnostic modeling, achieving superior AUC (0.969 training, 0.792 validation). Furthermore, we identified a novel 8-gene panel (*UPP1, ZNF600, GPR160, TRAT1, C3AR1, NAIP, MAL, F5*) and linked these to immune dysregulation via CIBERSORT and drug-repurposing opportunities via CellMiner — a multidimensional strategy unreported in existing literature.

It is crucial to acknowledge a significant limitation of this study: the findings are derived solely from computational analyses of publicly available transcriptomic datasets. While our integrated machine learning approach identified promising key genes (*UPP1, ZNF600, GPR160, TRAT1, C3AR1, NAIP, MAL, F5*) and revealed compelling immune

Figure 9. Drug sensitivity analysis using the CellMiner database.



dysregulation patterns, these results require rigorous experimental validation to confirm their biological and clinical relevance. Future work should prioritize validating the expression levels and diagnostic utility of the identified key genes using independent patient cohorts and techniques such as quantitative real-time PCR or droplet digital PCR in blood samples. Furthermore, functional validation is essential; this could involve in vitro studies, for example, gene knockdown/overexpression in relevant immune cells like monocytes or macrophages followed by assessment of cytokine production, phagocytosis, or cell survival, or in vivo models, for example, utilizing septic animal models to examine the impact of modulating these genes on survival, organ dysfunction, or the observed immune cell imbalance [42]. The predicted drug sensitivities associated with TRAT1, C3AR1, NAIP, and F5 warrant testing in cellular models of sepsis. Ultimately, prospective clinical studies are needed to evaluate the diagnostic and prognostic performance of the 8-gene panel and the ANN model in real-world settings.

In our research, it was revealed that a total of 452 DEGs were filtered out among the sepsis case group and the control group. Subsequently, GO and KEGG enrichment analyses demonstrated that all of these DEGs were closely related to the immune system. Relying on these DEGs, the RF and ANN algorithms were integrated to screen and identify diagnostic biomarkers for sepsis. Through RF screening, several model genes related to sepsis, including *UPP1*, *ZNF600*, *GPR160*, *TRAT1*, *C3AR1*, *NAIP*, *MAL*, and *F5*, were identified. Furthermore, a discrimination model was constructed using ANN, and the results of the validation set demonstrated the accuracy of the biomarkers and model based on RF and ANN screening. The results indicate that the Sepsis prediction or diagnostic model we screened and constructed as an integrated method can also demonstrate significant practicality in clinical applications.

Conclusions

In order to explore the specific KG related to sepsis, we carried out a thorough and profound analysis, examining the relevant genes and pathways. The eight key genes (*UPP1, ZNF600, GPR160, TRAT1, C3AR1, NAIP, MAL,* and *F5*) that we discovered are expected to expand our comprehension of the molecular mechanisms. Additionally, they offer more potential therapeutic targets for clinical treatment. However, further research is required to validate and develop these findings.

Availability of data and materials

All datasets are available in GEO (GSE9960, GSE13904, GSE26440, GSE64457) (https://www.ncbi.nlm.nih.gov/gds/) and CellMiner (https://discover.nci.nih.gov/cellminer/home.do/).

Corresponding author

Zhiying Lin
Jiangxi provincial People's Hospital,
The First Affiliated Hospital of Nanchang Medical College,
No.92 Aiguo Road,
Nanchang, Jiangxi Province, China.
Tel: 0791-86895550
Email: lzy1191528412@163.com

Conflict of interest

No conflict of interest is declared.

References

1. Huang M, Cai S, Su J (2019) The pathogenesis of sepsis and potential therapeutic targets. Int J Mol Sci 20: 5376. doi: 10.3390/ijms20215376.
2. Li C, Zhang J, Yang P, Wang R, Chen T, Li L (2024) The role of polymyxin B-immobilized hemoperfusion in reducing mortality and enhancing hemodynamics in patients with sepsis and septic shock: a systematic review and meta-analysis. Heliyon 10: e33735. doi: 10.1016/j.heliyon.2024.e33735.
3. Gaieski D, Edwards J, Kallan M, Carr B (2013) Benchmarking the incidence and mortality of severe sepsis in the United States. Crit Care Med 41: 1167-1174. doi: 10.1097/CCM.0b013e31827c09f8.
4. Iwashyna T, Cooke C, Wunsch H, Kahn J (2012) Population burden of long-term survivorship after severe sepsis in older Americans. J Am Geriatr Soc 60: 1070-1077. doi:10.1111/j.1532-5415.2012.03989.x.
5. Chen Q, Li R, Lin C, Lai C, Huang Y, Lu W, Li L (2023) SEPRES: intensive care unit clinical data integration system to predict sepsis. Appl Clin Inform 14: 65-75. doi: 10.1055/a-1990-3037.
6. Barichello T, Generoso JS, Singer M, Dal-Pizzol F (2022) Biomarkers for sepsis: more than just fever and leukocytosis-a narrative review. Crit Care 26: 14. doi: 10.1186/s13054-021-03862-5.
7. Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, Han B, Wang H, Duan X, Sun T (2021) A machine learning model for accurate prediction of sepsis in ICU patients. Front Public Health 9: 754348. doi: 10.3389/fpubh.2021.754348.
8. Yang Z, Cui X, Song Z (2023) Predicting sepsis onset in ICU using machine learning models: a systematic review and meta-analysis. BMC Infect Dis 23: 635. doi: 10.1186/s12879-023-08614-0.
9. Baghela A, Pena OM, Lee AH, Baquir B, Falsafi R, An A, Farmer SW, Hurlburt A, Mondragon-Cardona A, Rivera JD, Baker A, Trahtemberg U, Shojaei M, Jimenez-Canizales CE, Dos Santos CC, Tang B, Bouma HR, Cohen Freue GV, Hancock REW (2022) Predicting sepsis severity at first clinical presentation: the role of endotypes and mechanistic signatures. EBioMedicine 75: 103776. doi: 10.1016/j.ebiom.2021.103776.
10. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G (2021) ClusterProfiler 4.0: A universal enrichment tool for

interpreting omics data. Innovation (Camb) 2: 100141. doi: 10.1016/j.xinn.2021.100141.
11. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10: 1523. doi: 10.1038/s41467-019-09234-6.
12. Mazziotta C, Cervellera CF, Badiale G, Vitali I, Touzé A, Tognon M, Martini F, Rotondo JC (2023) Distinct retinoic gene signatures discriminate Merkel cell polyomavirus-positive from -negative Merkel cell carcinoma cells. J Med Virol 95: e28949. doi: 10.1002/jmv.28949.
13. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G (2021) ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) 2: 100141. doi: 10.1016/j.xinn.2021.100141.
14. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 49: D605–D612.
15. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12: 453-457. doi: 10.1038/nmeth.3337.
16. Kanashiro A, Sônego F, Ferreira RG, Castanheira FV, Leite CA, Borges VF, Nascimento DC, Cólon DF, Alves-Filho JC, Ulloa L, Cunha FQ (2017) Therapeutic potential and limitations of cholinergic anti-inflammatory pathway in sepsis. Pharmacol Res 117: 1-8. doi: 10.1016/j.phrs.2016.12.014.
17. He RR, Yue GL, Dong ML, Wang JQ, Cheng C (2024) Sepsis biomarkers: advancements and clinical applications-a narrative review. Int J Mol Sci 25: 9010. doi: 10.3390/ijms25169010.
18. Asrar R, Feras Al, Hari K, Govind B, Wael H, Joe B, Benjamin H, Praveen K, Christos K, Berit SB, Mohammed T, Zain H, Harish V, Zainab M, Maike S, Rayaz M Shriprasad D, Nasir Q, Raziya K, Amrita S, Guftar S, Javed S, Syed AZ, Rajesh P, Akash D, Ahmed Al, Amir H (2022) Applying transcriptomics for an enhanced clinical research framework, implications for an improved research strategy based on an omics approach: a scoping review. Informatics in Medicine Unlocked 44: 101419. doi: 10.1016/j.imu.2023.101419.
19. Zhang WY, Chen ZH, An XX, Li H, Zhang HL, Wu SJ, Guo YQ, Zhang K, Zeng CL, Fang XM (2023) Analysis and validation of diagnostic biomarkers and immune cell infiltration characteristics in pediatric sepsis by integrating bioinformatics and machine learning. World J Pediatr 19: 1094-1103. doi: 10.1007/s12519-023-00717-7.
20. Póvoa P, Coelho L, Dal-Pizzol F, Ferrer R, Huttner A, Conway Morris A, Nobre V, Ramirez P, Rouze A, Salluh J, Singer M, Sweeney DA, Torres A, Waterer G, Kalil AC (2023) How to use biomarkers of infection or sepsis at the bedside: guide to clinicians. Intensive Care Med 49: 142–153. doi: 10.1007/s00134-022-06956-y.
21. Bauer R, Rauch I (2020) The NAIP/NLRC4 inflammasome in infection and pathology. Mol Aspects Med 76: 100863. doi: 10.1016/j.mam.2020.100863.
22. Shukla P, Dwivedi P, Gupta PK, Mishra PR (2014) Optimization of novel tocopheryl acetate nanoemulsions for parenteral delivery of curcumin for therapeutic intervention of sepsis. Expert Opin Drug Deliv 11: 1697–1712. doi: 10.1517/17425247.2014.932769.
23. Kumpf O, Giamarellos-Bourboulis EJ, Koch A, Hamann L, Mouktaroudi M, Oh DY, Latz E, Lorenz E, Schwartz DA, Ferwerda B, Routsi C, Skalioti C, Kullberg BJ, van der Meer JW, Schlag PM, Netea MG, Zacharowski K, Schumann RR (2010) Influence of genetic variations in TLR4 and TIRAP/Mal on the course of sepsis and pneumonia and cytokine release: an observational study in three cohorts. Crit Care 14: R103. doi: 10.1186/cc9047.
24. Nwosu ZC, Ward MH, Sajjakulnukit P, Poudel P, Ragulan C, Kasperek S, Radyk M, Sutton D, Menjivar RE, Andren A, Apiz-Saab JJ, Tolstyka Z, Brown K, Lee HJ, Dzierozynski LN, He X, Ps H, Ugras J, Nyamundanda G, Zhang L, Halbrook CJ, Carpenter ES, Shi J, Shriver LP, Patti GJ, Muir A, Pasca di Magliano M, Sadanandam A, Lyssiotis CA (2023) Uridine-derived ribose fuels glucose-restricted pancreatic cancer. Nature 618: 151-158. doi: 10.1038/s41586-023-06073-w.
25. Lai K, Song C, Gao M, Deng Y, Lu Z, Li N, Geng Q (2023) Uridine alleviates sepsis-induced acute lung injury by inhibiting ferroptosis of macrophage. Int J Mol Sci 24: 5093. doi: 10.3390/ijms24065093.
26. Yao RQ, Ren C, Xia ZF, Yao YM (2021) Organelle-specific autophagy in inflammatory diseases: a potential therapeutic target underlying the quality control of multiple organelles. Autophagy 17: 385-401. doi: 10.1080/15548627.2020.1725377.
27. Zhang W, Zhangyuan G, Wang F, Jin K, Shen H, Zhang L, Yuan X, Wang J, Zhang H, Yu W, Huang R, Xu X, Yin Y, Zhong G, Lin A, Sun B (2021) The zinc finger protein Miz1 suppresses liver tumorigenesis by restricting hepatocyte-driven macrophage activation and inflammation. Immunity 54: 1168-1185.e8. doi: 10.1016/j.immuni.2021.04.027.
28. Maruyama K, Kidoya H, Takemura N, Sugisawa E, Takeuchi O, Kondo T, Eid MMA, Tanaka H, Martino MM, Takakura N, Takayama Y, Akira S, Vandenbon A, Kumagai Y (2020) Zinc finger protein st18 protects against septic death by inhibiting VEGF-A from macrophages. Cell Rep 32: 107906. doi: 10.1016/j.celrep.2020.107906.
29. Liu H, Pan D, Li P, Wang D, Xia B, Zhang R, Lu J, Xing X, Du J, Zhang X, Jin L, Jiang L, Yao L, Li M, Wu J (2023) Loss of ZBED6 protects against sepsis-induced muscle atrophy by upregulating DOCK3-mediated RAC1/PI3K/AKT signaling pathway in pigs. Adv Sci 10: e2302298. doi: 10.1002/advs.202302298.
30. Xu C, Wang Y, Ni H, Yao M, Cheng L, Lin X (2024) The role of orphan G protein-coupled receptors in pain. Heliyon 10: e28818. doi: 10.1016/j.heliyon.2024.e28818.
31. Yang H, Liu H, Chen H, Mo H, Chen J, Huang X, Zheng R, Liu Z, Feng Y, Liu F, Ge B (2016) G protein-coupled receptor160 regulates mycobacteria entry into macrophages by activating ERK. Cell Signal 28: 1145-1151. doi: 10.1016/j.cellsig.2016.05.022.
32. Wacker D, Stevens RC, Roth BL (2017) How ligands illuminate GPCR molecular pharmacology. Cell 170: 414–427. doi: 10.1016/j.cell.2017.07.009.
33. Yosten GL, Harada CM, Haddock C, Giancotti LA, Kolar GR, Patel R, Guo C, Chen Z, Zhang J, Doyle TM, Dickenson AH, Samson WK, Salvemini D (2020) GPR160 de-orphanization reveals critical roles in neuropathic pain in rodents. J Clin Invest 130: 2587-2592. doi: 10.1172/JCI133270.
34. Chen J, Yang P, Zhang Q, Chen R, Wang P, Liu B, Sun W, Jian X, Xiang S, Zhou J, Li N, Wang K, Gao C, Wen Y, Wu C,

Zhang J, Zhao Y, Yang Q, Li M, Stewart R, Sun Y, Pan D, Niu Y, Wang Z, Xu Y, Li X, He L, Li Z, Shi Y (2021) Genetic risk of clozapine-induced leukopenia and neutropenia: a genome-wide association study. Transl Psychiatry 11: 343. doi: 10.1038/s41398-021-01470-z.

35. Tham EL, Shrikant P, Mescher MF (2002) Activation-induced nonresponsiveness: a Th-dependent regulatory checkpoint in the CTL response. J Immunol 168: 1190-1197. doi: 10.4049/jimmunol.168.3.1190.
36. Martinez-Sanchez ME, Huerta L, Alvarez-Buylla ER, Villarreal Luján C (2018) Role of cytokine combinations on CD4+ T cell differentiation, partial polarization, and plasticity: continuous network modeling approach. Front Physiol 9: 877. doi: 10.3389/fphys.2018.00877.
37. Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL (2005) Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. J Immunol 136: 2348-57.
38. Hotchkiss RS, Tinsley KW, Swanson PE, Schmieg RE Jr, Hui JJ, Chang KC, Osborne DF, Freeman BD, Cobb JP, Buchman TG, Karl IE (2001) Sepsis-induced apoptosis causes progressive profound depletion of B and CD4+ T lymphocytes in humans. J Immunol 166: 6952-63. doi: 10.4049/jimmunol.166.11.6952.
39. Andreu-Ballester JC, Tormo-Calandín C, Garcia-Ballesteros C, Pérez-Griera J, Amigó V, Almela-Quilis A, Ruiz del Castillo J, Peñarroja-Otero C, Ballester F (2013) Association of γδ T cells with disease severity and mortality in septic patients. Clin Vaccine Immunol 20: 738-746. doi: 10.1128/CVI.00752-12.
40. Nedeva C (2021) Auid-orcid. Inflammation and cell death of the innate and adaptive immune system during sepsis. Biomolecules 11: 1011. doi: 10.3390/biom11071011.
41. Zheng X, Luo Y, Li Q, Feng J, Zhao C, Lu J, Luo J, Zhang J (2020) Two gene set variation index as biomarker of bacterial and fungal sepsis. Biomed Res Int 2020: 8182358. doi: 10.1155/2020/8182358.
42. Saxena J, Das S, Kumar A, Sharma A, Sharma L, Kaushik S, Kumar Srivastava V, Jamal Siddiqui A, Jyoti A (2024) Biomarkers in sepsis. Clin Chim Acta 562: 119891. doi: 10.1016/j.cca.2024.119891.

Annex – Supplementary Items

Supplementary Table 1. Enrichment Analysis of 11 MCODE modules.

Modules	Items	Results				
MCODE 1	p-value	8.12E-19	5.46E-17	8.61E-17	6.09E-16	8.10E-16
	term id	GO:0006955	GO:0045321	GO:0046649	GO:0009897	GO:0001775
	term name	immune response	leukocyte activation	lymphocyte activation	external side of plasma membrane	cell activation
MCODE 2	p-value	1.51E-29	1.80E-29	1.90E-29	1.98E-26	1.54E-25
	term id	GO:0034774	GO:0060205	GO:0031983	GO:0030141	REAC:R-HSA-6798695
	term name	secretory granule lumen	cytoplasmic vesicle lumen	vesicle lumen	secretory granule	Neutrophil degranulation
MCODE 3	p-value	1.14E-08	1.23E-07	3.25E-06	3.76E-06	3.76E-06
	term id	GO:0016411	GO:0008374	GO:0003841	GO:0042171	GO:0071617
	term name	acylglycerol O-acyltransferase activity	O-acyltransferase activity	1-acylglycerol-3-phosphate O-acyltransferase activity	lysophosphatidic acid acyltransferase activity	lysophospholipid acyltransferase activity
MCODE 4	p-value	1.35E-06	5.04E-06	1.75E-05	3.36E-05	1.20E-04
	term id	REAC:R-HSA-8856825	REAC:R-HSA-8856828	GO:0030669	GO:0045334	GO:0030665
	term name	Cargo recognition for clathrin-mediated endocytosis	Clathrin-mediated endocytosis	clathrin-coated endocytic vesicle membrane	clathrin-coated endocytic vesicle	clathrin-coated vesicle membrane
MCODE 5	p-value	4.45E-05	7.42E-05	1.56E-04	1.81E-04	3.34E-04
	term id	GO:0003873	GO:0004331	GO:0008443	GO:0006003	GO:0050308
	term name	6-phosphofructo-2-kinase activity	fructose-2,6-bisphosphate 2-phosphatase activity	phosphofructokinase activity	fructose 2,6-bisphosphate metabolic process	sugar-phosphatase activity
MCODE 6	p-value	4.40E-12	1.15E-08	4.11E-08	6.02E-08	4.92E-06
	term id	REAC:R-HSA-5676934	GO:0016667	GO:0016671	GO:0030091	WP:WP3580
	term name	Protein repair	oxidoreductase activity, acting on a sulfur group of donors	oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor	protein repair	Methionine de novo and salvage pathway
MCODE 7	p-value	1.35E-06	5.04E-06	1.75E-05	3.36E-05	1.20E-04
	term id	REAC:R-HSA-8856825	REAC:R-HSA-8856828	GO:0030669	GO:0045334	GO:0030665
	term name	Cargo recognition for clathrin-mediated endocytosis	Clathrin-mediated endocytosis	clathrin-coated endocytic vesicle membrane	clathrin-coated endocytic vesicle	clathrin-coated vesicle membrane
MCODE 8	p-value	5.05E-05	1.25E-04	5.33E-04	5.58E-04	8.04E-04
	term id	GO:0006950	GO:0006952	GO:0001817	GO:0001816	GO:0006954
	term name	response to stress	defense response	regulation of cytokine production	cytokine production	inflammatory response
MCODE 9	p-value	5.58E-05	1.43E-04	6.82E-04	0.001002502	0.001007108
	term id	GO:0031093	GO:0031091	REAC:R-HSA-6785807	REAC:R-HSA-114608	GO:0008083
	term name	platelet alpha granule lumen	platelet alpha granule	Interleukin-4 and Interleukin-13 signaling	Platelet degranulation	growth factor activity
MCODE 10	p-value	4.65E-07	7.51E-07	1.90E-06	8.38E-06	1.12E-05
	term id	GO:0098552	GO:0140375	GO:0002684	GO:0002250	GO:0007159
	term name	side of membrane	immune receptor activity	positive regulation of immune system process	adaptive immune response	leukocyte cell-cell adhesion
MCODE 11	p-value	2.08E-06	1.40E-05	1.59E-05	5.91E-05	1.22E-04
	term id	GO:0031347	GO:0002684	GO:0032101	GO:0080134	GO:0002682
	term name	regulation of defense response	positive regulation of immune system process	regulation of response to external stimulus	regulation of response to stress	regulation of immune system process
MCODE 12	p-value	1.14E-04	1.84E-04	1.84E-04	2.31E-04	2.51E-04
	term id	GO:0042613	GO:0002503	GO:0002399	GO:0042611	GO:0098553
	term name	MHC class II protein complex	peptide antigen assembly with MHC class II protein complex	MHC class II protein complex assembly	MHC protein complex	lumenal side of endoplasmic reticulum membrane
MCODE 13	p-value	8.27E-04	9.56E-04	0.001746264	0.003477439	0.006535707
	term id	GO:1902894	GO:2000629	REAC:R-HSA-157118	REAC:R-HSA-350054	GO:1902893
	term name	negative regulation of miRNA transcription	negative regulation of miRNA metabolic process	Signaling by NOTCH	Notch-HLH transcription pathway	regulation of miRNA transcription
MCODE 14	p-value	4.62E-06	4.62E-06	1.36E-05	2.03E-05	2.20E-05
	term id	GO:0009250	GO:0005978	GO:0000271	GO:0005977	GO:0044042
	term name	glucan biosynthetic process	glycogen biosynthetic process	polysaccharide biosynthetic process	glycogen metabolic process	glucan metabolic process

Supplementary Figure 1. The correlation between KG.