



Identification of differentially expressed genes, transcription factors, microRNAs and pathways in neutrophils of sepsis patients through bioinformatics analysis

Yukai Zheng[#], Lu Peng[#], Zhijie He, Zijun Zou, Fangyi Li, Canxia Huang^{*}, Weichao Li^{*}

Department of ICU, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong Province, P.R. China

[#]Co-first authors of the article and contribute equally to the article

ARTICLE INFO

Original paper

Article history:

Received: July 25, 2021

Accepted: November 09, 2021

Published: December 15, 2021

Keywords:

Sepsis; neutrophil; miRNA; bioinformatics

ABSTRACT

Sepsis has been recognized to be a life-threatening organ dysfunction caused by the dysregulation of the host response to infections. Our work aims to screen key biomarkers related to neutrophils in sepsis using bioinformatics analysis. For this purpose, the microarray datasets related to neutrophils in sepsis patients were downloaded from the Gene Expression Omnibus (GEO) database. According to the Bayesian test, the Limma package in R was used to screen differentially expressed genes (DEGs). Then, DEGs were uploaded to the DAVID online diagnostic tool for subsequent Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment on the selected DEGs. Next, protein-protein interaction (PPI) network was established based on the selected DEGs using the STRING website and the Cytoscape software. Furthermore, according to the function of the iRegulon plug-in in Cytoscape, our study further predicts and established regulatory networks related to transcription factors and regulatory genes. In addition, the miRWalk2.0 database was used to search for miRNA-DEG pairs, associated with the conduction of intersections of miRNAs predicted by TargetScan, Miranda, miRDB and RNA22 databases. Then, these miRNA-DEG pairs were also displayed in the form of a regulatory network through Cytoscape. Finally, two datasets were selected to verify the screened genes, regulatory factors, and miRNAs, to plot receiver operating characteristics (ROC) curves and compute the area under the curve (AUC) values. The results showed that AKT1, MMP9, ARG1, ETS1 targeting AKT1, and has-miR-124-3p targeting RPS6KA5 may have diagnostic value for patients with sepsis and septic shock. While further experimental studies are required to confirm their role in septic neutrophils.

DOI: <http://dx.doi.org/10.14715/cmb/2021.67.5.53>

Copyright: © 2021 by the C.M.B. Association. All rights reserved.



Introduction

Sepsis causes life-threatening organ dysfunction caused by an imbalanced response of the host to infections. It has a high incidence, and the condition remains one of the leading causes of death worldwide. It is estimated that there are over 19 million sepsis (previously severe sepsis) cases and 5 million sepsis-related deaths each year (1, 2). It has been recognized to be a major health burden worldwide with considerable economic consequences (3). It is of great practical significance for exploring the molecular mechanism of sepsis to benefit the development of novel therapeutic approaches for this disease.

Neutrophils are the most abundant immune cells in sepsis and the main source of IL-1 β and IL-18(4). Neutrophils are considered to be major players in the host's immune response to sepsis (5). Prior research has documented the critical role of regulating neutrophils in

sepsis treatment. For example, inducing neutrophil autophagy may enhance the production of neutrophil extracellular traps (NETs), thereby protecting mice from the lethality induced by sepsis (6).

Accumulated evidence has confirmed the role of important genes and key regulatory miRNAs in the pathogenesis and development of sepsis. For instance, in critical illness and sepsis, high levels of miR-133a are related to the severity of the disease and predict adverse outcomes in critically ill patients (7). Despite the above studies, there is still a lack of comprehensive understanding of the genes and miRNAs that are involved in neutrophils during sepsis. Furthermore, bioinformatics combines computer science and biological science to screen various molecules and signaling pathways, conduct data mining, and study diseases at the molecular level. Generally, bioinformatic analysis is performed first to identify potential molecules associated

^{*}Corresponding author. E-mail: liwch6@mail.sysu.edu.cn; huangcx5@mail.sysu.edu.cn
Cellular and Molecular Biology, 2021, 67(5): 405-423

with diseases, followed by investigation and validation in clinical trials. It has been widely applied in the study of sepsis. Several biomarkers have previously been identified and tested in the treatment of sepsis. However, there are few studies on the miRNA in neutrophils during sepsis. Therefore, this study aims to conduct bioinformatics analysis on the gene expression profile in neutrophils of sepsis.

Materials and methods

Data acquisition and preliminary processing

Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) database, a publicly accessible genomics database on PubMed, contains high-throughput gene expression data, chips and microarray data. It was visited to verify the difference of gene expressions in neutrophils between patients with sepsis and healthy subjects using "sepsis" or "sepsis shock" and "neutrophil" as searching terms. A total of 520 gene chip sets were obtained using "Homo sapiens" in Filters. The retrieval program was provided as follows: (("neutrophils"[MeSH Terms] OR neutrophil [All Fields]) OR "neutrophils"[MeSH Terms]) AND (("sepsis"[MeSH Terms] OR sepsis[All Fields]) OR ("shock, septic"[MeSH Terms] OR septic shock[All Fields])) OR (("sepsis"[MeSH Terms] OR sepsis[All Fields]) AND ("shock"[MeSH Terms] OR shock[All Fields]))).

In this study, the inclusion criteria of the datasets we selected were: 1) gene expression dataset; and 2) dataset composed of the normal control group and sepsis or raw data obtained from cases in the septic shock group. At the same time, the exclusion criteria were: 1) datasets with non-encoding RNA and DNA methylation sequence data; and 2) sample data from non-sepsis datasets of patients and non-neutrophils. Finally, 3 gene datasets were selected for data analysis (GSE64457, GSE123729 and GSE100150) from the GEO database, and 2 gene expression profiles (GSE100159 and GSE101639) as the verification datasets. Meanwhile, neutrophil samples were selected from sepsis patients and normal subjects.

The microarray data of GSE64457 was obtained based on GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), GSE123729 based on GPL21970 platform ([HuGene-2_0-st] Affymetrix Human Gene 2.0 ST Array [CDF: hugene20st_Hs_ENTREZG.cdf, Brainarray version

20.0.0]), and GSE100150 based on GPL6884 platform (Illumina HumanWG-6 v3.0 expression beadchip). Among them, GSE64457 contained transcription data of neutrophils isolated from peripheral blood of 9 patients with sepsis and 8 healthy volunteers, GSE123729 included 15 sepsis patients and 11 normal subjects, and GSE100150 provided the data of 35 sepsis samples and 12 normal samples.

In the next step, we downloaded the original data of the microarray, the GPL platform files, including CEL files and gene probes information. Based on data in the GPL file, the probe information was converted into a corresponding gene symbol. All data can be obtained through GEO data. At the same time, the whole process of this research did not involve any experimental behavior on humans or animals, with no ethical approval required.

Identifying differentially expressed genes (DEGs)

First, the affy extension package (<https://bioconductor.org/packages/3.3/bioc/html/affy.html>) in R software (version 4.0.2, <https://www.r-project.org/>) was used to read the gene expression chip data-CEL format data and process them into an expression matrix. The matrix data were further processed, including background correction, normalization, etc., to form a robust multi-array average (RMA). In order to identify DEGs, the limma (linear model of microarray data) package in R software (<https://bioconductor.org/packages/release/bioc/html/limma.html>) was selected to analyze and calculate corresponding data of each group (sepsis group and normal group). Value with $P < 0.05$ and $|\text{Log}_2\text{FC}| > 1$ was determined as the threshold. False discovery rate (FDR) was controlled by the Benjamini-Hochberg method. The DEGs data were uploaded to the Venn diagram website (<http://bioinformatics.psb.ugent.be/webtools/Venn>) to get the Venn diagrams of the DEGs from the three microarray datasets. All DEGs were displayed in the volcano map made by R software.

Gene Ontology (GO) Analysis and KEGG Pathway Enrichment Analysis

GO-based function enrichment analysis is a bioinformatics method including three common analyses of biological processes (BP), molecular

functions (MF), and cellular components (CC). Through previous analysis, the overlapped DEGs among the three groups were submitted to the Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8, <https://david.ncifcrf.gov>) with the species limited as “Homo sapiens” for BP, MF and CC analysis in GO analysis, and KEGG pathway analysis. DAVID provides a comprehensive set of functional annotation tools for investigators to have a functional interpretation of large lists of genes. DAVID can identify enriched biological themes, particularly GO terms. Besides, GO can provide a systematic language for the description of the attributes of genes and their product. KEGG is such a database that can benefit the understanding of the advanced functions and utility of biological systems (e.g., the cell, the organism and the ecosystem) based on molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. $P < 0.05$ meant that the difference was statistically significant.

Protein-Protein Interaction (PPI) Network and Hub Gene Identification

In order to further explore the protein interaction relationships of DEGs, a PPI network was generated firstly by using the STRING database (V11.0, <http://string-db.org>). STRING can not only establish PPI Networks but also is available for functional enrichment analysis. We set the minimum required interaction score as >0.4 , and the rest followed the default settings. In addition, Cytoscape software (<http://www.cytoscape.org/>, version 3.8.2) was used to visualize the data exported from STRING to construct the PPI network. By using the CytoHubba program in Cytoscape, the values of degree centrality (DC), betweenness centrality (BC), and closeness centrality (CC) between genes in the PPI network were analyzed to find hub genes. Besides, the MCODE App in Cytoscape software was used to perform module analysis for the established network.

Prediction of transcription factor (TF)-DEG pairs and construction of a regulatory network

In order to predict the TFs in the PPI network, the iRegulon app in Cytoscape was used to process the constructed PPI network, with the normalized enrichment score (NES) >3 as the standard. Finally, a

TF-DEG regulatory network was constructed, after which Cytoscape software was applied to visualize the predicted TF-DEG pairs. Genes related to key pathways were selected and processed by the miRWalk2.0 database

(<http://zmf.umm.uniheidelberg.de/apps/zmf/mirwalk2/generetsysself.html>) to predict their targeted miRNAs. In order to verify the accuracy of the results, TargetScan, Miranda, miRDB and RNA22 databases were also used for the intersection. Then, the screened targets were downloaded from the miRWalk2.0 database. By locating the target genes of miRNAs to the previously calculated DEGs, the miRNA-DEGs regulatory network related to neutrophils in sepsis patients were obtained and displayed in Cytoscape.

Receiver operating characteristic (ROC) analysis

To evaluate whether the DEGs, TFs, and miRNAs screened above were of important diagnostic value for sepsis or septic shock, package "PROC" in R software was utilized to perform ROC analysis on each gene. Meanwhile, the area under the ROC curve (AUC) was calculated to evaluate the accuracy of DEGs, TFs, and miRNAs in the diagnosis of sepsis/septic shock through the values of AUC. In other words, when the value of AUC was >0.8 , the gene screened might have a better value in diagnosing sepsis/septic shock patients from normal controls.

Results and discussion

Identification of DEGs

After the integration of 3 microarray data related to neutrophils of sepsis patients and normal subjects, the obtained DEGs were crossed with the screening of 603 DEGs in GSE64457, 695 in GSE123729, and 3,158 in GSE100150. [Figures S1(1), S1(2)-(6), S2(1), S2(2)-(6), S3(1) and S3(2)-(6)]. The DEGs of these three datasets are shown in the form of a volcano diagram (Fig 1. A-C). The overlap of the three microarray data contained 116 genes, as displayed in Fig 1.D, including 18 downregulated genes and 98 upregulated genes in the sepsis group when compared with those in the normal group.

GO and KEGG pathway enrichment analyses

GO functional analysis and KEGG pathway enrichment analysis were performed on the screened

DEGs with upregulation and downregulation, including BP, CC, MF and KEGG pathways. The first 5 items of each category were listed in Tables 1-4. The terms for enrichment of upregulated genes included innate immune response (BP, $p=1.45E-04$), extracellular exosome (CC, $p=1.06E-05$), cysteine-type endopeptidase inhibitor activity (MF, $p=1.44E-02$) and Legionellosis (pathway, $p=7.70E-03$). Moreover, the first 5 terms of the enrichment of downregulated genes were lymphocyte homeostasis (BP, $p=1.80E-02$), whole membrane (CC; $p=2.23E-02$) and MAPK signaling pathway (pathway; $p=8E-03$).

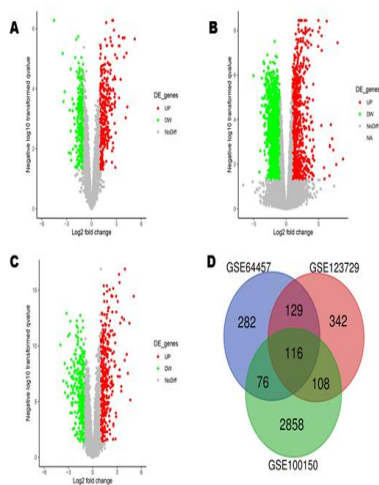


Figure 1. Identification of differentially expressed genes (DEGs). (A–C) The volcano plot of differential expression analysis between sepsis samples and normal controls, respectively. X-axis represents the fold change (log₂scale) and Y-axis represents the p-value (-log₁₀scale). (D) Venn diagram of the three lists of DEGs. DEGs-differentially expressed genes.

Table 1. Biological processes(BP) in GO(Gene ontology) analysis of DEGs, upregulated genes and downregulated genes.

A.DEGs			
Term	Function	Count	P-value
GO:0045087	innate immune response	13	2.49E-05
GO:0042742	defense response to bacterium	8	4.69E-05
GO:0006809	nitric oxide biosynthetic process	4	7.46E-05
GO:0006915	apoptotic process	14	8.55E-05
GO:0071222	cellular response to lipopolysaccharide	7	9.86E-05
B.Upregulated			
Term	Function	Count	P-value
GO:0045087	innate immune response	11	1.45E-04
GO:0042742	defense response to bacterium	7	1.63E-04
GO:0071222	cellular response to	6	4.30E-04

GO:0010042	lipopolysaccharide response to manganese ion	3	4.58E-04
GO:0006809	nitric oxide biosynthetic process	3	2.32E-03

C.Downregulated			
Term	Function	Count	P-value
GO:0002260	lymphocyte homeostasis	3	1.80E-02
GO:0002376	immune system process	9	1.86E-02
GO:0043281	regulation of cysteine-type endopeptidase activity involved in apoptotic process	4	1.86E-02
GO:0080134	regulation of response to stress	7	1.86E-02
GO:1901214	regulation of neuron death	4	1.86E-02

Table 2. Cellular components (CC) in GO (Gene ontology) analysis of DEGs, upregulated genes and downregulated genes.

A.DEGs			
Term	Function	Count	P-value
GO:0070062	extracellular exosome	35	3.99E-05
GO:0016021	integral component of membrane	46	3.69E-03
GO:0031093	platelet alpha granule lumen	4	4.58E-03
GO:0016020	membrane	24	6.39E-03
GO:0005615	extracellular space	17	7.49E-03

B.Upregulated			
Term	Function	Count	P-value
GO:0070062	extracellular exosome	32	1.06E-05
GO:0031093	platelet alpha granule lumen	4	2.70E-03
GO:0005615	extracellular space	16	2.94E-03
GO:0016020	membrane	21	6.11E-03
GO:0016021	integral component of membrane	38	9.22E-03

C.Downregulated			
Term	Function	Count	P-value
GO:0098805	whole membrane	7	2.23E-02
GO:0098588	bounding membrane of organelle	7	4.59E-02

Table 3. Molecular functions (MF) in GO (Gene ontology) analysis of DEGs, upregulated genes.

A.DEGs			
Term	Function	Count	P-value
GO:0004674	protein serine/threonine kinase activity	8	1.00E-02
GO:0004869	cysteine-type endopeptidase inhibitor activity	3	1.96E-02
GO:0004896	cytokine receptor activity	3	2.18E-02
GO:0003873	6-phosphofructo-2-kinase activity	2	2.51E-02
GO:0005524	ATP binding	17	2.60E-02

B.Upregulated			
Term	Function	Count	P-value
GO:0004869	cysteine-type endopeptidase inhibitor activity	3	1.44E-02
GO:0003873	6-phosphofructo-2-kinase activity	2	2.14E-02
GO:0004331	fructose-2,6-bisphosphate 2-phosphatase activity	2	2.67E-02
GO:0004908	interleukin-1 receptor activity	2	3.71E-02
GO:0005518	collagen binding	3	4.16E-02

PPI Network and Hub Gene Identification

The PPI network of DEGs was constructed by using Cytoscape, involving 116 nodes and 244 interactions (Fig 2. A). Fig 2. B shows the PPI

network displayed by Cytoscape. The hub genes were found by the CytoHubba program in Cytoscape. According to the scores of DC, BC and CC, 10 hub genes were determined with the highest network connectivity. The top 10 hub genes identified were AKT serine/threonine kinase 1 (AKT1), matrix metalloproteinase (MMP)9, arginase 1 (ARG1), MMP8, C3AR1, LCN2, QSOX1, TLR5, MAPK14 and FCAR (Table 5). The gene with the largest number of node networks (n=13) was AKT1. The 3 PPI network modules fulfilled the cutoffs of the degree cutoff (2), the maximum depth (100), the κ -core (2) and node score cutoff (0.2) (Fig 2. C-E, Table 6). Among them, genes of module 1 (the highest score, score=10.105) were enriched in Staphylococcus aureus infection pathway ($p=4.62E-02$) (Table 7.A); genes of module 2 (score=3.333) were enriched in Glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate ($p=5.81E-03$) (Table 7.B); genes of module 3 (score=3.333) were enriched in cGMP-PKG signaling pathway ($p=7.40E-03$) and Pentose phosphate pathway ($p=2.50E-02$) (Table 7.C).

The GO and KEGG pathway enrichment analyses of the genes in 3 modules (Table 7) showed that the genes in module 3 containing the hub gene AKT1 were mainly linked to postsynapse, cytosol, and cGMP-PKG signaling pathway.

Table 4. KEGG(The Kyoto Encyclopedia of Genes and Genomes) pathway analysis of DEGs, upregulated and downregulated DEGs(differentially expressed genes). When each category has more than 5 enriched terms, select the first five terms based on the P-value.

A.DEGs				
Pathway ID	Name	Count	P-value	Genes
hsa04621	NOD-like receptor signaling pathway	5	1.55E-03	CARD6,NLRC4, NLRP1, MAPK14, NAIP
hsa04010	MAPK signaling pathway	9	1.81E-03	MEF2C,RPS6KA5,ZAK, GADD45A, MKNK1, AKT1, ARRB1, MAPK14, MAP2K6
hsa04668	TNF signaling pathway	6	2.65E-03	RPS6KA5, AKT1, MAPK14, MMP9, IL18R1, MAP2K6
hsa05134	Legionellosis	4	1.25E-02	CR1, NLRC4, TLR5, NAIP
hsa05321	Inflammatory bowel disease (IBD)	4	1.97E-02	IL18RAP,IL4R, TLR5, IL18R1
B.Upregulated				
Pathway	Name	Count	P-	Genes

ID		value		
hsa05134	Legionellosis	4	7.70E-03	CR1, NLRC4, TLR5, NAIP
hsa04621	NOD-like receptor signaling pathway	4	8.52E-03	CARD6,NLRC4,MAPK14, NAIP
hsa05321	Inflammatory bowel disease (IBD)	4	1.23E-02	IL18RAP,IL4R, TLR5, IL18R1
hsa00051	Fructose and mannose metabolism	3	2.41E-02	PFKFB2, HK3, PFKFB3
hsa04640	Hematopoietic cell lineage	4	2.77E-02	CR1, IL4R, IL1R2, CD44

C.Downregulated				
Pathway ID	Name	Count	P-value	Genes
hsa04010	MAPK signaling pathway	4	8E-03	AKT1,ARRB1,MEF2C,RPS6KA5

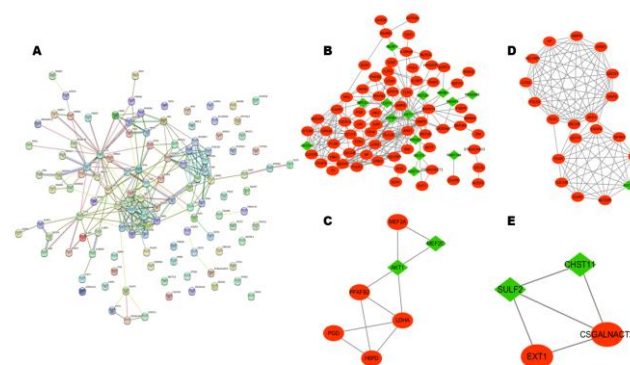


Figure 2. PPI network analysis. (A) PPI network of the three lists of DEGs obtained from STRING database, involving 116 nodes and 244 interactions. (B) The PPI network was generated through Cytoscape software using data from STRING. (C-E) Module analysis of the PPI network. A total of 3 Modules were obtained satisfying the cutoffs of the degree cutoff (2), the maximum depth (100), the κ -core (2) and node score cutoff (0.2). PPI-Protein-protein interaction.

Table 5. In the protein-protein interaction (PPI) network, the top ten genes with the highest degree of interaction.

Gene Name	Degree
AKT1	23
MMP9	20
ARG1	16
MMP8	16
C3AR1	13
LCN2	13
QSOX1	13
TLR5	12
MAPK14	12
FCAR	12

In the PPI(protein-protein interaction) network, the top ten hub genes are all listed in the table above. AKT1-AKT serine/threonine kinase 1;MMP9-matrix metalloproteinase 9;ARG1-arginase 1; MMP8- matrix metalloproteinase 8; C3AR1-complement C3a receptor 1;LCN2-lipocalin 2;QSOX1-quiescin sulfhydryl oxidase 1;TLR5-toll-like receptor 5;MAPK14-mitogen-activated protein kinase 14;FCAR-Fc fragment of IgA receptor.

Table 6. The 3 modules from the protein-protein interaction (PPI) network fulfilled the criteria of MCODE score \geq 4 and the number of nodes \geq 4.

Cluster	Score	Nodes	Edges	Node IDs
1	10.105	20	96	ARG1,PGLYRP1,C3AR1,STOM,TCN1,ORM1,CLEC4D,MMP9,HP,GPR84,FCAR,HVCN1,CD177,CKAP4,QSOX1,MMP8,OSCAR,MCEMP1,FOLR3,LCN2.
2	3.333	4	5	EXT1,CHST11,SULF2,CSGALNACT2
3	3.333	7	10	MEF2A,H6PD,PFKFB2,PGD,LDHA,MEF2C,AKT1

Score = Density \times number of Nodes.

Table 7A. Perform functional and pathway enrichment analysis on the genes in module1. When more than 5 enrichment items are found in each category, the first 5 enrichment items are selected according to the P-value.

A. Biological processes (BPs)				
Term	Name	Count	P-value	Genes
GO:005090	leukocyte migration	3	7.43E-03	C3AR1, MMP9, CD177
GO:000963	response to herbicide	2	7.48E-03	ARG1, LCN2
GO:007020	protein homodimerization	2	2.33E-02	ARG1, LCN2
GO:003598	endodermal cell differentiation	2	2.86E-02	MMP8, MMP9
GO:001004	response to zinc ion	2	3.79E-02	ARG1, HVCN1
B. Cellular components (CCs)				
Term	Name	Count	P-value	Genes
GO:000557	extracellular region	11	1.24E-06	ORM1, TCN1, HP, LCN2, QSOX1, FOLR3, PGLYRP1, MMP8, MMP9, FCAR, OSCAR
GO:000561	extracellular space	9	3.15E-05	ORM1, TCN1, ARG1, HP, LCN2, STOM, QSOX1, MMP8, MMP9
GO:007006	extracellular exosome	11	1.81E-04	ORM1, ARG1, HP, LCN2, STOM, QSOX1, PGLYRP1, MMP9, CKAP4, CD177, OSCAR
GO:007256	blood microparticle	3	1.08E-02	ORM1, HP, STOM
C. Molecular functions (MF)				
Term	Name	Count	P-value	Genes
GO:000425	serine-type endopeptidase activity	3	2.66E-02	HP, MMP8, MMP9

D. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways

Term	Name	Count	P-value	Genes
hsa05150	Staphylococcus aureus infection	2	4.62E-02	C3AR1, FCAR

Table 7B. Perform functional and pathway enrichment analysis on the genes in module2. When more than 5 enrichment items are found in each category, the first 5 enrichment items are selected according to the P-value.

A. Biological processes (BPs)				
Term	Name	Count	P-value	Genes
GO:0002063	chondrocyte development	2	2.68E-03	CHST11, SULF2
GO:0030206	chondroitin sulfate biosynthetic process	2	4.46E-03	CHST11, CSGALNACT2
B. Cellular components (CCs)				
Term	Name	Count	P-value	Genes
GO:0000139	Golgi membrane	3	3.08E-03	EXT1, CHST11, CSGALNACT2
C. Molecular functions (MF)				
Term	Name	Count	P-value	Genes
hsa00532	Glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate	2	5.81E-03	CHST11, CSGALNACT2

Table 7C. Perform functional and pathway enrichment analysis on the genes in module3. When more than 5 enrichment items are found in each category, the first 5 enrichment items are selected according to the P-value.

A. Biological processes (BPs)				
Term	Name	Count	P-value	Genes
GO:0006089	lactate metabolic process	2	1.79E-03	PFKFB2, LDHA
GO:0006098	pentose-phosphate shunt	2	3.93E-03	H6PD, PGD
GO:0051149	positive regulation of muscle cell differentiation	2	8.55E-03	MEF2A, MEF2C
GO:0006096	Glycolytic process	2	1.21E-02	PFKFB2, LDHA
GO:0055114	oxidation-reduction process	3	1.69E-02	LDHA, H6PD, PGD
B. Cellular components (CCs)				
Term	Name	Count	P-value	Genes
GO:0098794	postsynapse	2	7.22E-03	MEF2C, AKT1
GO:0005829	cytosol	5	1.20E-02	PFKFB2, MEF2C, LDHA, AKT1, PGD
C. Molecular functions (MF)				
Term	Name	Count	P-value	Genes
GO:0033613	activating transcription factor binding	2	7.80E-03	MEF2A, MEF2C
GO:0042826	histone deacetylase binding	2	3.57E-02	MEF2A, MEF2C
D. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways				
Term	Name	Count	P-value	Genes
hsa05150	Staphylococcus aureus infection	2	4.62E-02	C3AR1, FCAR
hsa04022	cGMP-PKG signaling pathway	3	7.40E-03	MEF2A, MEF2C, AKT1
hsa00030	Pentose	2	2.50E-02	H6PD, PGD

phosphate pathway
Table 8. Pathways enriched in genes involved in the transcriptional regulatory network.

Term	Name	Count	P-value	Gene
hsa04010	MAPK signaling pathway	9	1.13E-05	MEF2C, RPS6KA5, ZAK, GADD45A, MKNK1, AKT1, ARRB1, MAPK14, MAP2K6
hsa04668	TNF signaling pathway	6	1.02E-04	RPS6KA5, AKT1, MAPK14, MMP9, IL18R1, MAP2K6
hsa05164	Influenza A	5	7.30E-03	AKT1, JAK2, MAPK14, AGFG1, MAP2K6
hsa04066	HIF-1 signaling pathway	4	8.92E-03	HK3, PFKFB3, MKNK1, AKT1
hsa05205	Proteoglycans in cancer	5	1.18E-02	HGF, AKT1, MAPK14, MMP9, CD44
hsa05145	Toxoplasmosis	4	1.19E-02	AKT1, JAK2, MAPK14, MAP2K6
hsa04722	Neurotrophin signaling pathway	4	1.63E-02	RPS6KA5, SORT1, AKT1, MAPK14
hsa05169	Epstein-Barr virus infection	4	1.70E-02	AKT1, MAPK14, CD44, MAP2K6
hsa04621	NOD-like receptor signaling pathway	3	2.60E-02	NLRC4, NLRP1, MAPK14
hsa04664	Fc epsilon RI signaling pathway	3	3.72E-02	AKT1, MAPK14, MAP2K6
hsa04917	Prolactin signaling pathway	3	4.03E-02	AKT1, JAK2, MAPK14
hsa05152	Tuberculosis	4	4.44E-02	CD74, AKT1, JAK2, MAPK14

Prediction of TF-DEG pairs and construction of network

The regulatory network of TF-DEG by Cytoscape was constructed, in which TF ETS proto-oncogene 1 (*ETS1*) targeted 39 DEGs (such as *AKT1*) (Fig 3). In addition, a pathway analysis was performed on the genes involved in the TF-DEG regulatory network. The top 10 enriched pathways included MAPK signaling pathway ($p=1.13E-05$), tumor necrosis factor (TNF) signaling pathway ($p=1.02E-04$), Influenza A ($p=7.30E-03$), HIF-1 signaling pathway ($p=8.92E-03$), Proteoglycans in cancer ($p=1.18E-02$), Toxoplasmosis ($p=1.19E-02$), Neurotrophin signaling pathway ($p=1.63E-02$), Epstein-Barr virus infection ($p=1.70E-02$), NOD-like receptor signaling pathway ($p=2.60E-02$), and Fc epsilon RI signaling pathway ($p=3.72E-02$) (Table 8).

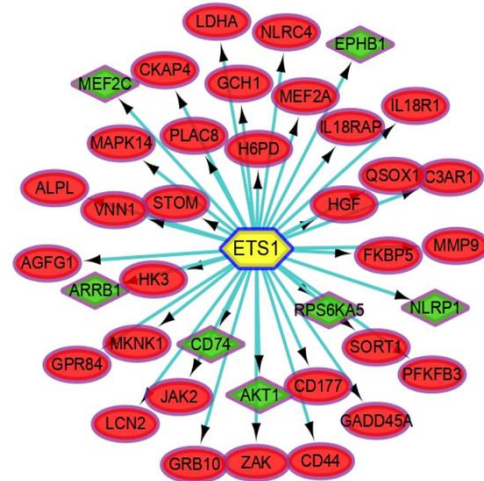


Figure 3. TF-DEG regulatory network identified from the PPI network. The red circles, green diamonds and yellow hexagons represent upregulated genes, downregulated genes and TFs, respectively. The blue-green arrow indicates the regulation relationship of TF-DEG. TF-transcription factor; DEGs-differentially expressed genes; PPI-Protein-protein interaction.

Through the iRegulon App in Cytoscape and verified by TargetScan, Miranda, miRDB and RNA22 databases, has-miR-124-3p was screened to be related to neutrophil function in sepsis. Then the targets of has-miR-124-3p were downloaded from the miWalk2.0 database. Among these target genes, 9 genes (7 upregulated genes and 2 downregulated genes, such as ribosomal protein S6 kinase alpha-5 and *RPS6KA5*) were found to be differentially expressed. In addition, Fig 4 shows the miRNA-DEG network constructed by Cytoscape.

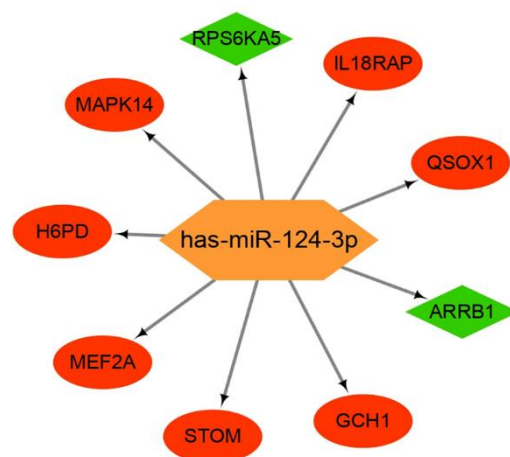


Figure 4. The miRNA-DEG regulatory network and targets of has-miR-124-3p. The red ovals, green diamonds and

orange hexagons represent upregulated genes, downregulated genes and miRNAs, respectively. miRNA-microRNA; DEGs-differentially expressed genes.

ROC analysis

According to the results of ROC curve analysis to verify the diagnostic significance of the screened genes and miRs in sepsis, MMP9 (AUC=0.955), ETS1 (AUC=0.950), RPS6KA5 (AUC=0.945), and ARG1 (AUC=0.917) were found to have the ability to clearly distinguish between sepsis patients and normal controls. However, AKT1 revealed a poor ability to recognize sepsis and corresponding controls (AUC=0.602) (Fig 5. A). Similarly, our findings showed that has-miR-124-3p had a significant diagnostic value of distinguishing septic shock (AUC=0.889) and sepsis (AUC=0.833) samples from normal controls (AUC=0.889) (Fig 5. B-D).

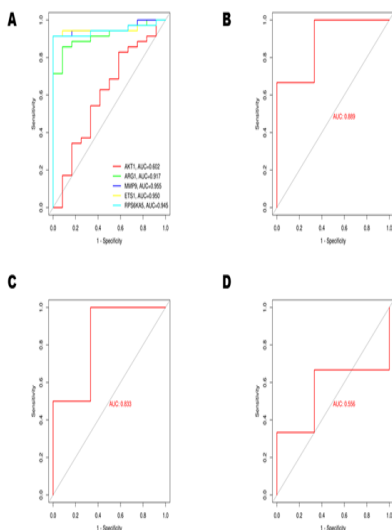


Figure 5. ROC curve analysis of 3 key genes (ARG1, MMP9, RPS6KA5), transcription factor (ETS1) and miRNA (has-miR-124-3p) to reveal corresponding sensitivity and significance for the diagnosis of sepsis/septic shock. X-axis. (A) ROC curves of AKT1, ARG1, MMP9, RPS6KA5, and ETS1 in patients with sepsis. (B) ROC curve of has-miR-124-3p in normal controls versus septic shock patients. (C) ROC curve of has-miR-124-3p in normal controls versus sepsis patients. (D) ROC curve of has-miR-124-3p in patients with sepsis versus those with septic shock. The AUC value >0.8 indicates that the gene may have a diagnostic value for sepsis. ROC-receiver operating characteristics; AUC-area under the ROC curve.

In our study, 116 DEGs were identified on the basis of sepsis and normal samples from the screened datasets, including 98 upregulated genes and 18 downregulated genes. The PPI network was constructed involving these DEGs. Based on DC, BC and CC scores, AKT1, MMP9 and ARG1 were identified to be the hub genes. In addition, 3 modules were identified from the PPI network. Furthermore, we also constructed the ETS1-DEG and the has-miR-124-3p regulatory networks.

As one of the most versatile kinase families, AKT1 (also known as PKB) serine-threonine kinase is a key regulator of cell survival, proliferation, metabolism and migration. For instance, Ursula et al. constructed a mouse model and demonstrate that the overexpression of AKT1 in lymphocytes can prevent sepsis-induced apoptosis, cause Th1 cytokine propensity, and improve survival when compared with wild-type littermates(8). Meanwhile, AKT1 phosphorylation could also be stimulated by lncRNA MALAT1 based on the recruitment of EZH2 and decrease of BRCA1 expression, thus accelerating inflammatory response to aggravating sepsis (9). Furthermore, in terms of MMP-9, it has been reported that the administration of inhibitors to inhibit the trimeric proteoform of MMP-9 selectively would increase the presence of neutrophils in the inflammation triggering site, suggesting a possible beneficial role of MMP-9 trimers in inhibiting excessive neutrophil migration or anti-inflammatory effect (10). Reckens et al. found that MMP-9(-/-) mice developed more severe damage to distant organs during infection, implying that MMP-9 may be significantly involved in the effective host response to *E. coli* peritonitis (11). In addition, Shi, H. et al. demonstrated in their study that through the upregulation of the expression of ARG1 and inducible nitric oxide synthase (iNOS) via Arctigenin, the immunosuppression activity of MDSC could be promoted on M1 macrophage polarization(12). Uhel F. et al found that genes (e.g., MMP8 and ARG1) involved in MDSC inhibitory function could be upregulated in the peripheral blood of patients with sepsis. Besides, the high level of ARG1 was clearly associated with subsequent nosocomial infections (13). GO enrichment analysis in our study showed that AKT1 and MMP9 were enriched in TNF signaling pathway. These evidences indicate that

AKT1, MMP9 and ARG1 may all play important roles in neutrophils during sepsis.

Furthermore, in another study carried out by He, H., et al., aging-related inflammation could promote senescence of hepatic stellate cells through the $TNF\alpha \rightarrow ETS1 \rightarrow IL-27ra$ pathway (14). Shiu, Y. T. and E. A. Jaimes' experimental evidence suggested that ETS1 plays a role in vascular and kidney damage (15). Moreover, ETS1 is involved in regulating the expression of cytokines, growth factors and adhesion molecules. The research of Hua, P., et al. showed that ETS1 is a TF that can transactivate the expression of fibronectin induced by Ang II. In addition, the phosphorylation of ETS1 induced by Ang II involves both the ERK-MAPK pathway and PI3K/Akt pathway (16). In our study, pathway analysis indicated that genes regulated by ETS1 were enriched in MAPK signaling pathway, TNF signaling pathway, Influenza A and HIF-1 signaling pathway. Among them, AKT1 was involved in the above signaling pathways. In the constructed TF-DEG network, AKT1 was the target gene of ETS1, indicating that ETS1 targeting AKT1 may play a role in septic neutrophils through MAPK and TNF signaling pathways.

Concerning the value of miR-124-3p, Liang, Y., et al. found that miR-124-3p treatment could alleviate the degree of lung injury caused by endotoxin (such as alveolar hemorrhage, and interstitial edema), accompanied by a significant reduction in the levels of IL-1 β , IL-6 and TNF-alpha. Overexpression of miR-124-3p can significantly inhibit lipopolysaccharide-induced p65 expression in NR8383 cells and in mice and cell apoptosis in NC8383 cells, suggesting a protective role of miR-124-3p by targeting p65 (17). Besides, in a high-quality RCT study on out-of-hospital cardiac arrest, miR-124-3p levels can be used as a predictive tool for neurological prognosis and survival after out-of-hospital cardiac arrest (18). At the same time, it has been demonstrated that RPS6KA5 could regulate protein synthesis and affect cell division, proliferation and differentiation (19, 20). In our study, in the miRNA-DEG regulatory network we constructed, RPS6KA5 was targeted by has-miR-124-3p, implying that has-miR-124-3p targeting RPS6KA5 may participate in mediating the functions of neutrophils in sepsis.

Conclusions

To sum up, 116 DEGs were identified from sepsis samples. AKT1, MMP9, and ARG1 are the hub genes. AKT1, MMP9, ARG1, ETS1 targeting AKT1, and has-miR-124-3p targeting RPS6KA5 may have potential values in the diagnosis of sepsis and septic shock. Nevertheless, our study do have some limitations that deserve attention. Firstly, it shall be noted that findings in our study were speculation and assumptions based on bioinformatics analysis. There is an absence of clinical experiments for validation, which may affect the reliability of our study and shall be carried out in the future. In addition, there may be some methodological overlaps considering the emergence of bioinformatic analysis at present. While findings in our study are still valuable for the diagnosis and potential interventions of sepsis. While their roles in neutrophils of sepsis still need further experimental studies to confirm.

Funding

The work in this paper was not funded.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Acknowledgements

We acknowledge the public database of GEO database for providing their platforms and contributors for uploading their meaningful datasets. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. Our study was carried out based on open source data, and hence there were no relevant ethical issues and other conflicts of interest.

Acknowledgement

This project was supported by the PSAU (Grant Number 2019/03/10211).

Interest conflict

The authors declare no conflict of interest.

References

1. Rudd KE, Kissoon N, Limmathurotsakul D et al. The global burden of sepsis: barriers and potential solutions. *Curr Mol Biol* 2018; 22(1): 232.
2. Fleischmann C, Scherag A, Adhikari NK et al. Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. *Am J Respir Crit Care Med* 2016; 193(3): 259-272.
3. Singer M, Deutschman CS, Seymour CW et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *Jama* 2016; 315(8): 801-810.
4. Liu L, Sun B. Neutrophil pyroptosis: new perspectives on sepsis. *Cell Mol Life Sci* 2019; 76(11): 2031-2042.
5. Hotchkiss RS, Monneret G, Payen D. Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy. *Cell Mol Biol* 2013; 13(12): 862-874.
6. Park SY, Shrestha S, Youn YJ et al. Autophagy Primes Neutrophils for Neutrophil Extracellular Trap Formation during Sepsis. *Am J Respir Crit Care Med* 2017; 196(5): 577-589.
7. Tacke F, Roderburg C, Benz F et al. Levels of circulating miR-133a are elevated in sepsis and predict mortality in critically ill patients. *Crit Care Med* 2014; 42(5): 1096-1104.
8. Bommhardt U, Chang KC, Swanson PE et al. Akt decreases lymphocyte apoptosis and improves survival in sepsis. *J Immunol* 2004; 172(12): 7583-7591.
9. Yong H, Wu G, Chen J et al. lncRNA MALAT1 Accelerates Skeletal Muscle Cell Apoptosis and Inflammatory Response in Sepsis by Decreasing BRCA1 Expression by Recruiting EZH2. *Mol Ther Nucleic Acids* 2020; 19: 97-108.
10. Nuti E, Rossello A, Cuffaro D et al. Bivalent Inhibitor with Selectivity for Trimeric MMP-9 Amplifies Neutrophil Chemotaxis and Enables Functional Studies on MMP-9 Proteoforms. *Cells* 2020; 9(7).
11. Renckens R, Roelofs JJ, Florquin S et al. Matrix metalloproteinase-9 deficiency impairs host defense against abdominal sepsis. *J Immunol* 2006; 176(6): 3735-3741.
12. Shi H, Dong G, Yan F et al. Arctigenin Ameliorates Inflammation by Regulating Accumulation and Functional Activity of MDSCs in Endotoxin Shock. *Inflammation* 2018; 41(6): 2090-2100.
13. Uhel F, Azzaoui I, Grégoire M et al. Early Expansion of Circulating Granulocytic Myeloid-derived Suppressor Cells Predicts Development of Nosocomial Infections in Patients with Sepsis. *Am J Respir Crit Care Med* 2017; 196(3): 315-327.
14. He H, Xu P, Zhang X et al. Aging-induced IL27Ra signaling impairs hematopoietic stem cells. *Blood* 2020; 136(2): 183-198.
15. Shiu YT, Jaimes EA. Transcription Factor ETS-1 and Reactive Oxygen Species: Role in Vascular and Renal Injury. *Antioxidants (Basel)* 2018; 7(7).
16. P, Feng W, Rezonzew G, Chumley P, Jaimes EA. The transcription factor ETS-1 regulates angiotensin II-stimulated fibronectin production in mesangial cells. *Am J Physiol Renal Physiol* 2012; 302(11): F1418-1429.
17. Liang Y, Xie J, Che D et al. MiR-124-3p helps to protect against acute respiratory distress syndrome by targeting p65. *Biosci Rep* 2020; 40(5).
18. Devaux Y, Dankiewicz J, Salgado-Somoza A et al. Association of Circulating MicroRNA-124-3p Levels With Outcomes After Out-of-Hospital Cardiac Arrest: A Substudy of a Randomized Clinical Trial. *JAMA Cardiol* 2016; 1(3): 305-313.
19. Reyes D, Ballaré C, Castellano G et al. Activation of mitogen- and stress-activated kinase 1 is required for proliferation of breast cancer cells in response to estrogens or progestins. *Oncogene* 2014; 33(12): 1570-1580.
20. Gawrzak S, Rinaldi L, Gregorio S et al. MSK1 regulates luminal cell differentiation and metastatic dormancy in ER(+) breast cancer. *Nat Cell Biol* 2018; 20(2): 211-221.

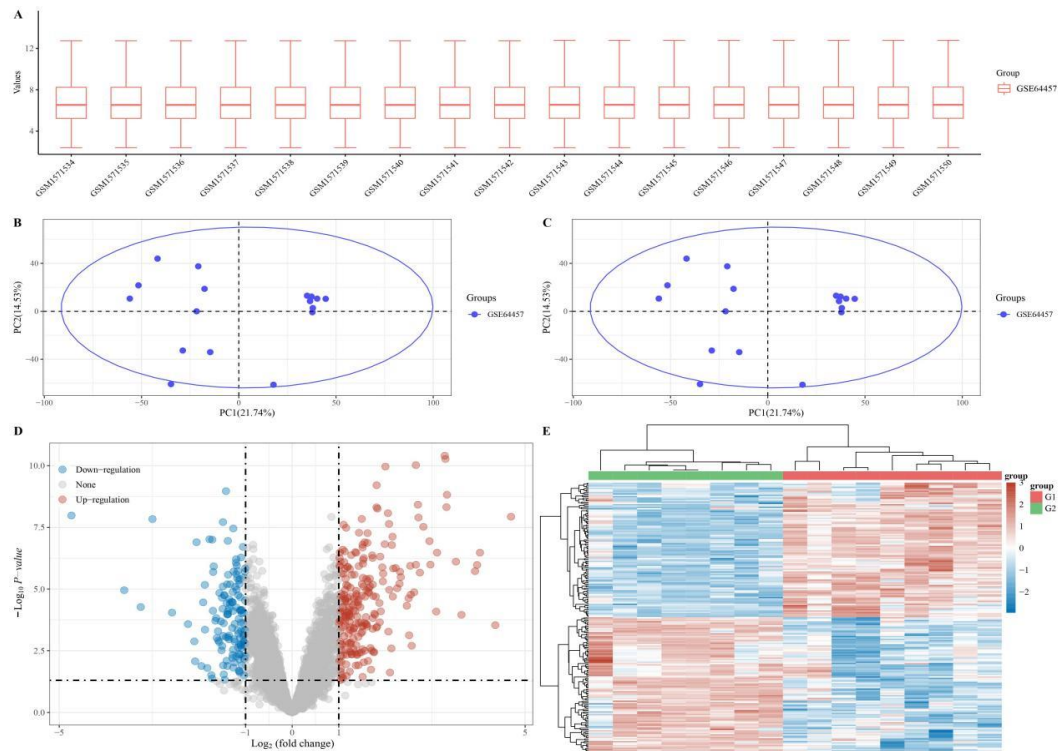
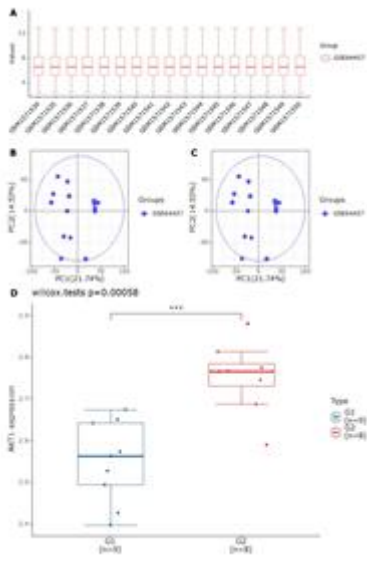
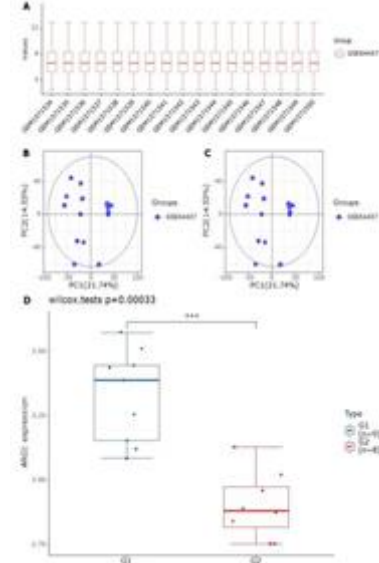


Figure S1(1). The data of the GSE64457 dataset was divided into 2 groups (sepsis patients and normal adults) for analysis, with preliminary processing and analysis performed. A: Box plot after data standardization, different colors represent different datasets. B: PCA results before batch removal for multiple datasets. Different colors represent different datasets; C: PCA results after batch removal, which can be used as a batch of data for subsequent analysis; D: Volcano plots using fold-change values and adjusted P-value. The red point represents the over-expressed mRNAs and the blue point indicates the down-expressed mRNAs with statistical significance. E: Hierarchical clustering analysis of mRNAs, which were differentially expressed in neutrophils between sepsis patients and normal controls.

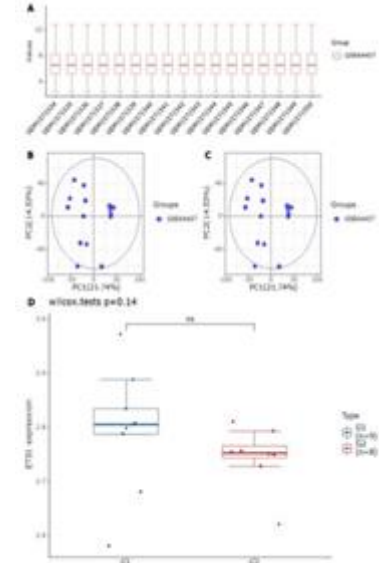
AKT1



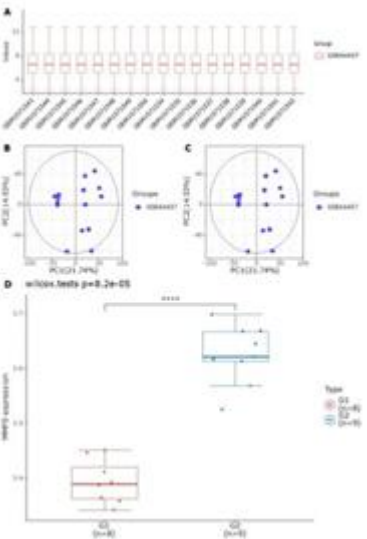
ARG1



ETS1



MMP9



RPS6KA5

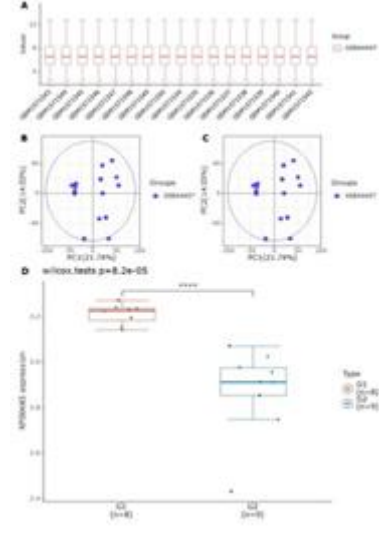


Figure S1(2)-(6). Comparison of AKT1, ARG1, ETS1, MMP9 and RPS6KA5 gene expression in the GSE64457 dataset in sepsis group and normal group (sepsis patients and normal adults). A: Box plot after data standardization, different colors represent different datasets. B: PCA results before batch removal for multiple datasets. Different colors represent different datasets. C: PCA results after batch removal, which can be used as a batch of data for subsequent analysis; D: The expression distribution of AKT1, ARG1, ETS1, MMP9 and RPS6KA5 in each group. The horizontal axis represents different groups of samples, and the vertical axis represents the gene expression distribution. Different colors represent different groups, and the upper left corner represents the p-value test method to indicate statistical significance.

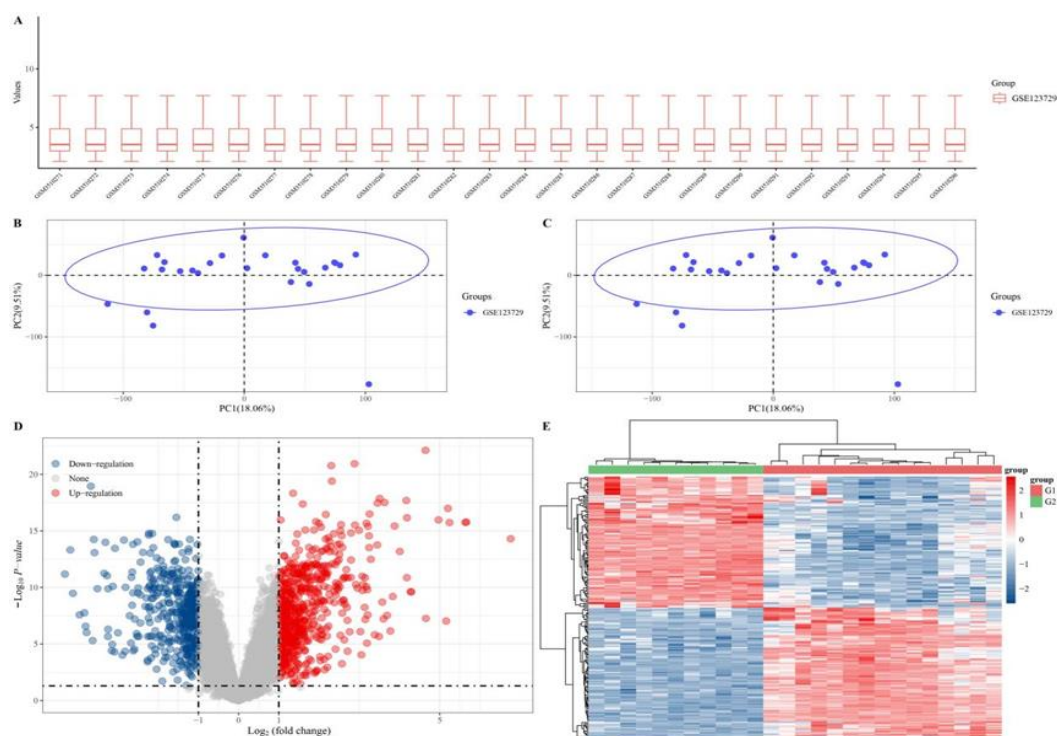
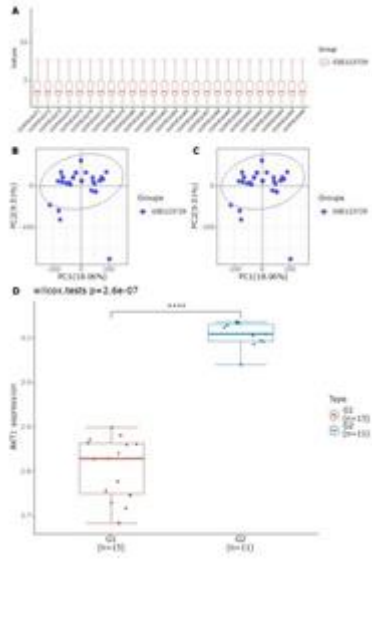
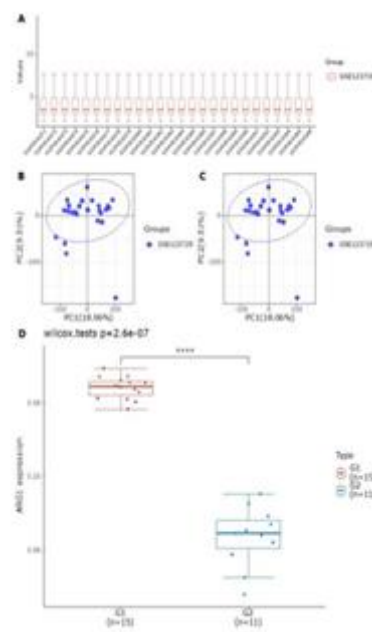


Figure S2(1). The data of the GSE123729 dataset was divided into 2 groups (sepsis patients and normal adults) for analysis, with preliminary processing and analysis performed. A: Box plot after data standardization, different colors represent different datasets. B: PCA results before batch removal for multiple datasets. Different colors represent different datasets; C: PCA results after batch removal, which can be used as a batch of data for subsequent analysis; D: Volcano plots using fold-change values and adjusted P values. The red point represents the over-expressed mRNAs and the blue point indicates the down-expressed mRNAs with statistical significance. E: Hierarchical clustering analysis of mRNAs, which were differentially expressed in neutrophils between sepsis patients and normal controls.

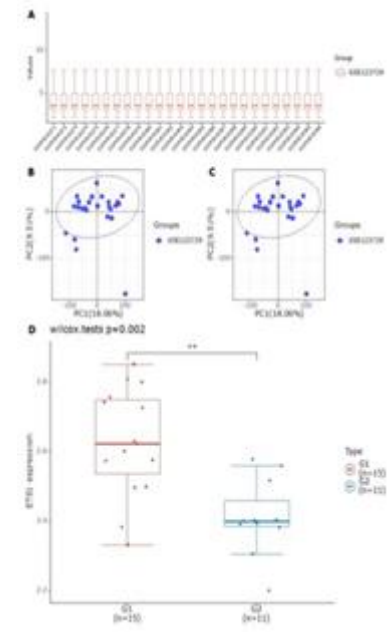
AKT1



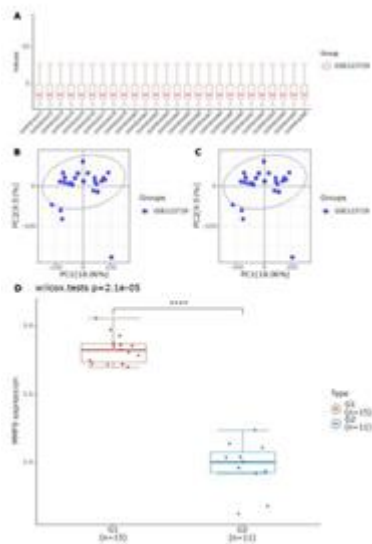
ARG1



ETS1



MMP9



RPS6KA5

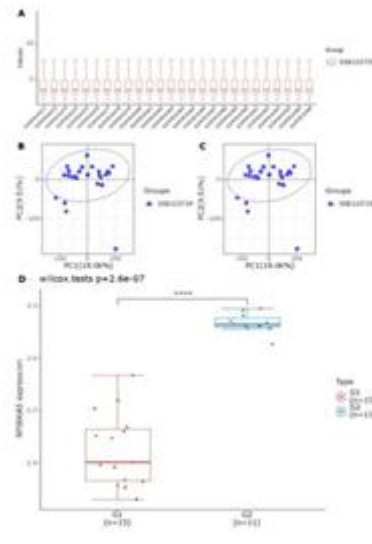


Figure S2(2)-(6). Comparison of AKT1, ARG1, ETS1, MMP9 and RPS6KA5 gene expression in the GSE123729 dataset in sepsis group and normal group (sepsis patients and normal adults). A: Box plot after data standardization, different colors represent different datasets. B: PCA results before batch removal for multiple datasets. Different colors represent different datasets. C: PCA results after batch removal, which can be used as a batch of data for subsequent analysis; D: The expression distribution of AKT1, ARG1, ETS1, MMP9 and RPS6KA5 in each group. The horizontal axis represents different groups of samples, and the vertical axis represents the gene expression distribution. Different colors represent different groups, and the upper left corner represents the p-value test method to indicate statistical significance.

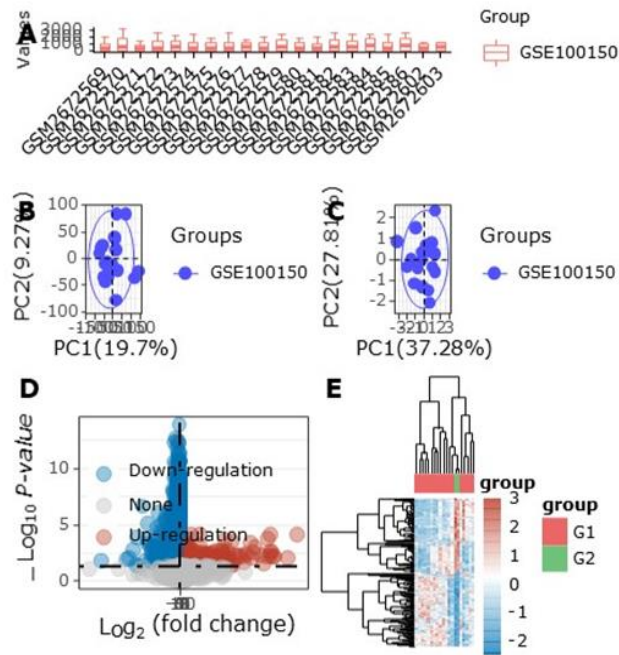
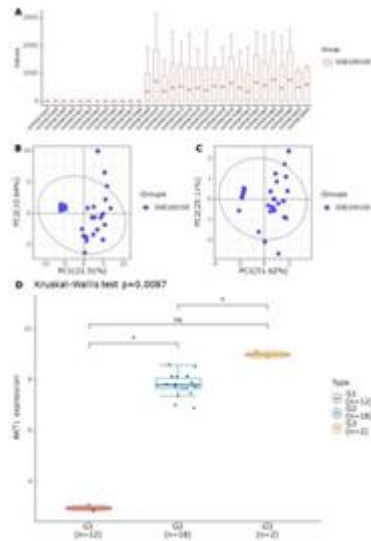
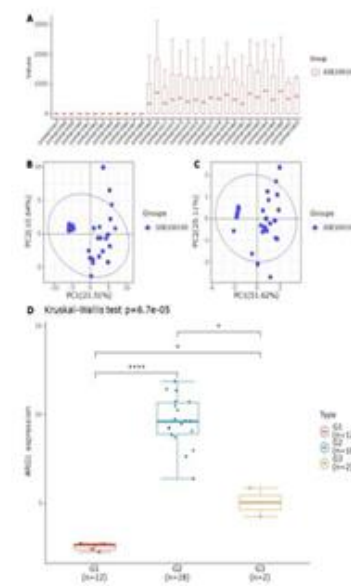


Figure S3(1). The data of the GSE100150 dataset was divided into 2 groups (sepsis patients and normal adults) for analysis, with preliminary processing and analysis performed. A: Box plot after data standardization, different colors represent different datasets. B: PCA results before batch removal for multiple datasets. Different colors represent different datasets; C: PCA results after batch removal, which can be used as a batch of data for subsequent analysis; D: Volcano plots using fold-change values and adjusted P values. The red point represents the over-expressed mRNAs and the blue point indicates the down-expressed mRNAs with statistical significance. E: Hierarchical clustering analysis of mRNAs, which were differentially expressed in neutrophils between sepsis patients and normal controls.

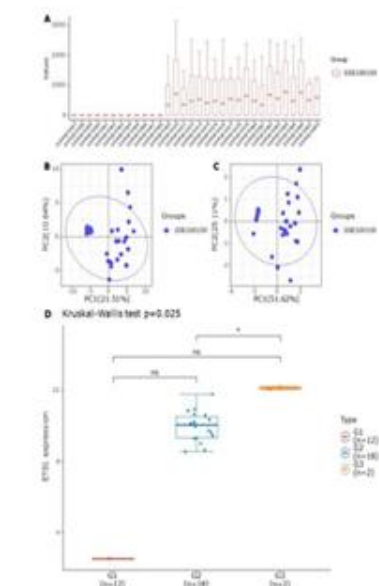
AKT1



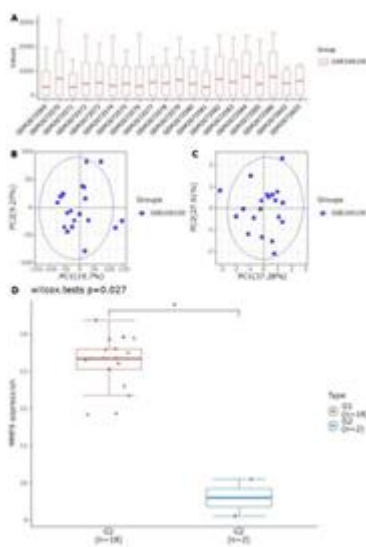
ARG1



ETS1



MMP9



RPS6KA5

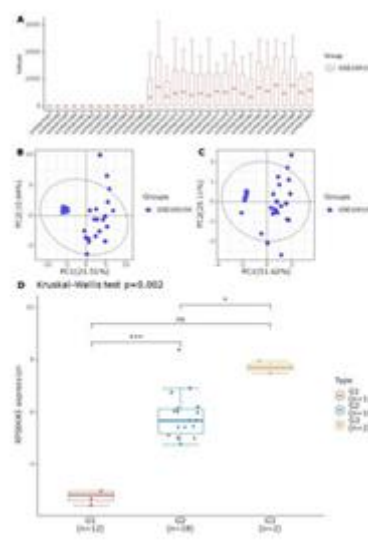


Figure S3(2)-(6). Comparison of AKT1, ARG1, ETS1, MMP9 and RPS6KA5 gene expression in the GSE100150 dataset in sepsis group and normal group (sepsis patients and normal adults). A: Box plot after data standardization, different colors represent different datasets. B: PCA results before batch removal for multiple datasets. Different colors represent different datasets. C: PCA results after batch removal, which can be used as a batch of data for subsequent analysis; D: The expression distribution of AKT1, ARG1, ETS1, MMP9 and RPS6KA5 in each group. The horizontal axis represents different groups of samples, and the vertical axis represents the gene expression distribution. Different colors represent different groups, and the upper left corner represents the p-value test method to indicate statistical significance.