

RESEARCH

Open Access



# Identification and validation of autophagy-related genes in sepsis based on bioinformatics studies

Dong-po Wei<sup>1†</sup>, Wei-wei Jiang<sup>2†</sup>, Chang-xing Chen<sup>1†</sup>, Zi-yang Chen<sup>1</sup>, Fang-qing Zhou<sup>1</sup>, Yu Zhang<sup>1\*</sup> and Jian Lu<sup>1,3\*</sup>

## Abstract

We identified 14 key genes associated with mitochondrial autophagy in sepsis through differential analysis of the dataset and then analysed the identified genes for functional enrichment. The analysis of key genes and deeper analysis of key genes by molecular typing, Weighted Gene Correlation Network Analysis (WGCNA) and ceRNA were also carried out. We have also validated these key genes with clinical data. Finally, sepsis diagnostic models are constructed by combining key genes with machine learning methods. In addition, we discuss the importance of the immune system in sepsis and its relationship with signature genes, which opens up new directions for studying the role of the immune system in sepsis. Overall, our study adds new ideas to the diagnosis and treatment of sepsis.

**Keywords** Sepsis, Autophagy, Weighted gene correlation Network analysis, Gene expression omnibus dataset

## Introduction

The incidence of sepsis in developed countries ranges between 66 and 300 per 100,000 individuals, with a death rate of 27–36%. In 2017, the number of reported sepsis cases in the United States increased to 437 per 100,000 individuals [1]. In addition to the high fatality rate, people who survive sepsis can face severe consequences,

with limitations for their daily living activities. While the inflammatory responses that are a hallmark of sepsis morbidity are reasonably well-described pathological features of this disease, the mechanisms that drive both metabolic dysfunction and multi-organ disorder or failure remain elusive [2].

The main biomarkers found for sepsis diagnosis are thrombomodulin [3], vascular endothelial glycocalyx [4, 5], von Willebrand Factor (vWF) [6], nitric oxide [7], endothelin [8] and HMGB1 protein [9]. Although there are several studies on sepsis pathogenesis, to date, no diagnostic biomarkers or targeted treatment drugs have proven comprehensive [8, 10]. Nonetheless, growing evidence show that sepsis pathogenesis involves apoptosis and autophagy [11–13]. Autophagy comprises four sequential steps: autophagy initiation, autophagosome creation, autophagosome integration, and autophagosome decomposition [14]. Autophagy is key for mitochondrial health, eliminating senescent, impaired, compromised, or overloaded mitochondria [15]. The knockdown of the autophagy gene ATG5 in mouse

<sup>†</sup>Dong-po Wei, Wei-wei Jiang and Chang-xing Chen have contributed equally to this work. Dong-po Wei is the primary author, and Wei-wei Jiang should be regarded as a co-first author.

\*Correspondence:

Yu Zhang  
shui0424@126.com

Jian Lu  
lujian@live.cn

<sup>1</sup> Department of Critical Care Medicine, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, 85 Wujin Road, Shanghai 200080, China

<sup>2</sup> Department of Emergency and Critical Care Medicine, Changzheng Hospital, Naval Medical University, Shanghai, China

<sup>3</sup> Department of Critical Care Medicine, Shanghai United Family Hospital, Shanghai, China



macrophages (lysosome-mediated system) limits acute toxin induced liver disease and its fatality, by repressing the proliferation of inflammasome-dependent interleukin-1 $\beta$  (IL-1 $\beta$ ) [16]. Enhanced autophagy has been reported to improve survival in mouse models of sepsis, through a net-dependent regime [17, 18]. Nevertheless, autophagy-associated genes (ARGs) in sepsis remain largely unknown and require further investigation, which will provide biomarkers for this disease.

Bioinformatics analysis is increasingly being used to select and identify the underlying critical routes of several gene pathways and even biological markers. Several studies developed predictive models based on whole-genome analysis of gene expression profiles during sepsis [19–21]. However, clinical diagnostic models for sepsis often involve dozens of genes and are therefore not easily implemented in clinical practice. Our current research focuses on the identification of mitochondria-associated autophagy genes on the basis of which diagnose models can be developed to better diagnose sepsis in the clinical setting. We screened differentially expressed genes (DEGs) between the sepsis and normal groups from three human datasets, crossed these genes with genes related to mitochondrial autophagy, resulting in the selection of 14 key genes and functional enrichment analysis and deeper analysis of these genes and we used two machine learning methods (Random Forest, RF) and Artificial Neural Network (ANN) to construct a diagnostic model for sepsis. Finally, clinical specimens were collected and quantitative real-time polymerase chain reaction (PCR) was performed to validate these genes.

## Materials and methods

### Data acquisition

Three human datasets for sepsis were downloaded from gene expression omnibus (GEO) [22] using the R package GEOquery [23]: GSE154918 [24] (control=40, sepsis=20, excluding incomplete infection cases, septic shock cases, and follow-up cases, a total of 45 cases); GSE32707 [25] (control=34, sepsis=58, excluding 31 acute respiratory distress syndrome (ARDS) cases); and GSE54514 [26] (control=36, sepsis=127, all included in the study). Our dataset only selects data from *Homo sapiens* to ensure the comparability of the data. And the tissue sources selected are all peripheral blood (whole blood, PBMC) samples, excluding studies involving specific tissues (such as liver, lung, etc.) or cell lines. Secondly, the sequencing platforms selected are all Illumina chip sequencing to ensure the uniformity of the data and avoid systematic deviations caused by different technical platforms (Table 1). And datasets with a larger sample size are preferred to improve the reliability of

statistical analysis. Datasets with severe data missing or without standardization are also excluded. For data with batch effects, a standardized method is subsequently used for correction.

### Differential analysis and selection of mitochondria autophagy related genes

We collected mitochondrial autophagy-related genes (MRGs) from the MSigDB (<https://www.gseamsigdb.org/gsea/msigdb>) [27], Reactome database (<https://reactome.org/>), [28] and published literature respectively [29]. MSigDB (34 genes): ATG5, ATG7, BCL2L1, BNIP3, BNIP3L, DAPK1, FUNDC1, KEAP1, MAP1LC3A, MAP1LC3B, MAP1LC3C, MFN1, MFN2, MIEF1, MIEF2, OPTN, PINK1, PRKN, RAB7A, RHOT1, RHOT2, SQSTM1, TBK1, TOMM22, ULK1, UBE2D2, UBE2E3, VDAC1, etc. Reactome database (29 genes): BECN1, BNIP3, DNMI1L, FIS1, FUNDC1, GABARAP, GABARAPL1, LC3, MAP1LC3, MFN1, MFN2, OPTN, PINK1, PRKN, RHOT1, RHOT2, RUBCN, RUFY3, SLC25A4, TOMM20, TOMM70, etc. Published literature (9 genes): ATP6V0A1, CSNK2A1, FBXW7, MARCH5, PGAM5, SIRT3, STUB1, YME1L1, USP30. After final merging and deduplication, a total of 42 mitophagy-related genes were obtained. For specific information, see Table S1. Differential analysis was conducted to obtain separately expressed mitochondrial autophagy-related genes, grouped using the R package “limma”. In the field of gene expression analysis, limma, a statistical package, is used to normalize gene expression levels across various studies and different datasets so that they can be compared. This method provides a robust standardization method, particularly suitable for cross-sample normalization of microarray and RNA-Seq data. It outperforms other normalization methods in improving the comparability of different datasets. Genes with  $\log_{2}FC > 0$  and  $P$  value  $< 0.05$  were defined as differentially expressed mitochondrial autophagy-associated up-regulated genes, and those with  $\log_{2}FC < 0$  and  $P$  value  $< 0.05$  were determined as differentially expressed mitochondrial autophagy-associated down-regulated genes.

### CIBERSORT

CIBERSORT (<https://cibersort.stanford.edu/>) is a tool for deconvolution of the expression matrix of human immune cell subtypes based on the principle of linear support vector regression [30], allowing the assessment of the infiltration status of immune cells in sequenced samples, through a gene expression profile of 22 known immune cell types. Right now, it's the go-to immune subtype analysis tool, able to accurately figure out how much of each immune cell type is in mixed samples, and it works really well with transcriptome data. In this

study, the immune cell infiltration status of the combined dataset was evaluated by the CIBERSORT algorithm in Immuno-Oncology Biological Research (IOBR) package and spearman correlations were calculated for the inter-relationships between the various immune cells [31]. We obtained a set of 28 genes used to identify different types of tumor-infiltrating immune cells from published studies [32], which contained a range of human cell isoforms, including CD8<sup>+</sup> T cells, dilated cells, giant phages, and regulatory T cells. The enrichment fraction computed by ssGSEA in the "GSVA" software package was used to represent the degree of infiltration of each immune cell type in each sample [33].

#### Functional (GO) and pathway enrichment (KEGG) analyses

Gene ontology (GO) profiling is a common approach to study gene expression in accordance to their cell function or location [34] and is generally carried out at three levels: biological process (BP), molecular function (MF), and cellular component (CC) [35]. Kyoto Encyclopedia of Gene and Genomes (KEGG) is a database that stores information about the genome, biological pathways, illnesses, and medicines [36]. ClusterProfiler, a tool for bioinformatics, integrates a variety of functional analysis methods, and is efficient in enrichment analysis, offering effective visualization of the results. The R package "clusterProfiler" was used for GO functional annotation analysis and pathway and disease enrichment analysis [37, 38], and the R package "Pathview" was used to visualize the KEGG pathway results with significant differences [39]. Statistical significance was set at  $P < 0.05$ .

#### GSEA and GSVA

Gene Set Enrichment Analysis (GSEA), an enrichment method proposed by the Broad Institute Research [40], is a statistical method used to determine whether a set of pre-defined genes shows statistically significant differences between two conditions and is commonly applied to evaluate changes in pathways and biological processes. GSEA works well in cases with few genes that are expressed differently, as it can detect small but important changes in pathways. To study the biological process variation among the normal and septic groups, the reference gene set "c2.cp.kegg.v7.4.entrez.gmt" was downloaded from the MSigDB database [27] and employed using the GSEA method in the R package "clusterProfiler" for enrichment analysis and data visualization. Statistical significance was set at  $P < 0.05$ . Gene Set Variance Analysis (GSVA) is a non-parametric, unmonitored method for assessing the degree of enrichment of different metabolic pathways in the microarray transcriptome. In research, (GSVA) can effectively identify differences among samples and is suitable for the analysis of immune-related

pathways [41]. To study changes in the bioprocesses between both sets of samples, we performed GSVA, using the R package GSVA, based on a gene expression spectrum dataset [42] calculated by uploading the referred gene set "c2.cp.kegg.v7.4.entrez.gmt", obtained from the MSigDB database. The R package "limma" [43] was used to combine the enrichment scores of each sample in every pathway, to filter out significantly different pathways between groups. The enrichment results from GSVA were visualized using the R package "pheatmap", based on heatmaps. Statistical significance was set at  $P < 0.05$ .

#### Development of a diagnostic model

Two machine-learning approaches, Random Forest (RF) and Artificial Neural Network (ANN) were used to construct a diagnostic model for sepsis. RF is an ensemble learning method composed of multiple decision trees. Using multiple decision trees for combined predictions can effectively improve the model's accuracy. RF performs random sampling on the training sample set and feature vectors, and obtains the final classification result through voting. Therefore, RF has a faster learning speed, is robust against noisy data, and can analyze complex interactions of classification features. ANN is a branch of machine learning capable of distinguishing hidden linear and nonlinear relationships in high-dimensional and complex datasets. Its ability to learn from examples, tolerate faults, and predict nonlinear data makes it one of the most commonly used machine learning methods. ANN consists of an input layer, hidden layers, and an output layer. The number of neurons in the input layer represents the number of variables describing the evaluated features, while the neurons in the output layer are the dependent variables. The number of hidden layers and neurons depends on the data size and the complexity of the relationship between the input and output layers. Each neuron in the hidden and output layers connects to all neurons in the previous layer using specific numerical weights. We employed the R package "randomForest" [44] to perform a random survival analysis of mitochondrial autophagy-related differentially expressed genes (MARDEGs) and selected genes with an importance greater than 0. Next, we constructed diagnostic models using the R packages "NeuralNetTools" and "neuralnet" [45]. To assess the behaviour of the model, the R package "pROC" [46] was used to plot receiver operating characteristic (ROC) curves and to calculate the area under the curve (AUC).

#### WGCNA

WGCNA is the process of defining commonly expressed gene blocks, probing the relationship between gene

networks and phenotype, and examining the central genes in the web, this method helps identify important co-expressed gene networks and pinpoint key genes linked to the onset and progression of sepsis [47]. A soft trigger value of nine was determined using the “pick soft” threshold function. A dimensionless network was constructed based on the soft trigger value, followed by topology matrix and hierarchical clustering. Eigengenes were calculated using the minimum number of genes (50 modules), and the combined shear height of the modules was set to 0.2. Correlations between modules were constructed from the module eigengenes, and stratified clusters were used to understand the correlations between models and between gene modules and clinical characteristics, using Spearman association analysis. Genes from the modules most relevant to sepsis were intersected with differentially expressed mitochondrial ARGs to identify key genes associated with mitochondrial autophagy in sepsis.

#### ceRNA Network construction

ceRNA is a new form of modulating gene expression, involving a wider range of RNA molecules, including mRNAs, pseudogenes, long-stranded uncoding RNAs, and miRNAs, providing new perspectives on transcriptome studies, and increasing the understanding of biological phenomena [48]. In this study, we constructed ceRNA networks using the following steps: first, we used the miRcode (<http://www.mircode.org/>) database to predict miRNAs likely to be bound to key interest genes, for the construction of miRNA-mRNA interaction pairs; second, we used the same database to predict the Long non-coding RNAs (lncRNAs) that the above miRNAs might bind to, for the determination of miRNA-lncRNA interaction pairs; finally, by integrating the miRNA-mRNA and miRNA-lncRNA interaction pairs, we built a ceRNA regulatory network of mRNA-miRNA-lncRNA for key genes in sepsis.

#### Molecular typing

Consensus Clustering [49] is a resampling-based algorithm, identifying each gene, its subgroup, and verifying the reasonableness of the clusters. This is used for classifying septic samples based on immune infiltration traits and finding potential subtypes. We used the R package “ConsensusClusterPlus” [50] to perform consistent clustering of the dataset using key genes associated with mitochondrial autophagy in sepsis to facilitate the differentiation between different subtypes of sepsis samples. In this process, 80% of the total sample was drawn in 1000 replicates with `clusterAlg = “km,” distance = “euclidean”`.

#### Correlation analysis of immune cells infiltration

ssGSEA estimates the number of immune infiltrative cells and their specific immune response activity. This method is applicable for gene set enrichment analysis for a single sample and can effectively identify changes in immune cells during disease. Through this algorithm we obtained 28 gene sets used to identify different tumor-infiltrating immune cell types, including various human immune cell subtypes, such as CD8<sup>+</sup> T cells, dendritic cells, macrophages, and regulatory T cells [33]. The enrichment scores, calculated by ssGSEA analysis using the R software package “GSVA”, were used to represent the degree of infiltration of each immune cell type within each sample. The immune infiltrate gene set was obtained from the literature [51] and analyzed through ssGSEA using the R software package “GSVA”. Correlations between key sepsis-associated genes and immune-invading cells were determined by Spearman-related analysis. Heatmaps were visualized using the R software package “ggplot2”.

#### RNA Extraction and qRT-PCR

Samples for the sepsis group (n=10) were provided from patients diagnosed with sepsis according to Sepsis 3.0 within 24 h of admission, and samples for the control group (n=10) were provided from healthy people. All the blood came from the clinical laboratory of Shanghai First People’s Hospital after use. This study was approved by the Medical Ethics Committee of Shanghai First People’s Hospital (2,024,112). Whole blood was drawn and added to the erythrocyte lysate according to the manufacturer’s instructions, and after completion of erythrocyte lysis, total RNA was extracted using TRIzol reagent (Invitrogen), and after purification, the total RNA concentration was determined by absorbance at 260 nm. Reverse transcription using the Biotnt Reverse Transcription Kit (cDNA first strand synthesis kit, Room 502, No.98 Luo Jin Road, Minhang District, Shanghai, China). Fluorescent quantitative PCR was performed on ABI ViiA7 (Applied Biosystems, Foster City, California, USA) using SYBR GREEN PCR MIX (cDNA first strand synthesis kit, Room 502, No.98 Luo Jin Road, Minhang District, Shanghai, China). Primer sequences used for amplification: ATG12, forward: 5’ TAG TCC AGT AGA GGG CAG TCC 3’, reverse: 5’ GCT TGC TCT CCT GGC TAG ATA 3’, ATG7, forward: 5’ CTG GTC ATC AAT GCT GCT TTG 3’, reverse: 5’ ACA GGG TGG TTT GGA CAC AAG 3’, ATG9A, forward: 5’ CCT TGA CCT CTT CTT CTC TCG 3’, reverse: 5’ TAG TGA AGG CAA CCA CAA AGA 3’, HIF1A, forward: 5’ AGA AAC CAC CTA TGA CCT GCT 3’, reverse: 5’ CGA CTG AGG AAA GTC TTG CTA 3’, HTRA2, forward: CTG CTAAGCGGCGACACGTA, reverse: ATCCTCAGC GTTGCGATGTCT, HUWE1, forward: GTTTCTGAG

GGATTGGAACACT,reverse:GTGGAGATTCAA  
GCACCGT,MAP1LC3A,forward:5' AGC ACC CCA  
GCA AAA TC 3',reverse:5' CAA AAA CTT GGT CTT  
GTC CAG 3',MAP1LC3B,forward:5' GTT GTT ACG  
GAA AGC AGC AGT 3',reverse:5' GAA GGC AGA  
AGG GAG TGT G 3',PGAM5,forward:GGAGCAGGC  
TGA ACTCACT,reverse:GTCATAGACGAATGG  
ACGATTT,SQSTM1,forward:5' AGT CGG ATA ACT  
GTT CAG GAG 3',reverse:5' ATT CTG GCA TCT  
GTA GGG A 3',TOMM20,forward:5' CAC TTT CCC  
TCC ATT TGT TAC 3',reverse:5' CCA CAT ACA TTC  
CTC AGG TT 3',TOMM40,forward:5' ACT ACC ACT  
TCG GGG TCA CAT A 3',reverse:5' GAG CGT TGA  
GAC TGC CAC T 3',UBA52,forward:CTTCAACCCT  
CGTGCTGTCAACT,reverse:TCTTCTTGGGACGCA  
GGTTGT,VPS13C,forward:TGGTGACACTACATC  
ATCCTTG,reverse:AGAGTCTGGTCAGCAGGA  
CTATC,PCR conditions were 95°C for 10 min, followed  
by 40 amplification cycles of 95°C for 15 s and 60°C for  
1 min. The relative mRNA expression was calculated  
using the comparative threshold cycling (Ct) method.

### Statistical methods

R program (<https://www.r-project.org/>, version 4.2.0) was used for data computation and statistical analysis [52]. For comparisons between two groups, those with normal distribution were analyzed using independent Student's t-tests to estimate statistical significance. Comparisons of conditions without normal distribution were conducted by the Mann–Whitney U-test (Wilcoxon rank sum test). The non-parametric testing method Wilcoxon rank-sum test is ideal for small sample sizes. It is better at detecting differences in distribution between groups without being affected by the assumption of normality in data distribution. All statistical *P* values were bilateral, and  $P < 0.05$  was considered statistically significant.

## Results

### Analysis flow chart

The flowchart illustrates the complete research framework for obtaining data from the GEO database (GSE54514, GSE154918, and GSE32707), integrating and analyzing it. First, the three datasets are merged, and differential expression analysis is conducted to screen for key differential genes related to sepsis (MRDEGs). Subsequently, these genes are used for enrichment analysis (GO and KEGG) and immune infiltration analysis, while GSEA and GSVA are employed to explore potential biological pathways. The study further classifies immune subtypes based on consensus clustering and explores key gene modules in conjunction with WGCNA (Weighted Gene Co-expression Network Analysis). Additionally, predictive models are constructed using Random Forest

(RF) and Artificial Neural Networks (ANN), followed by ROC curve analysis to evaluate model performance. At the same time, the ceRNA (competitive endogenous RNA) network is integrated to study gene regulatory mechanisms, comprehensively revealing the molecular characteristics of sepsis (Fig. 1).

### Data set pre-processing

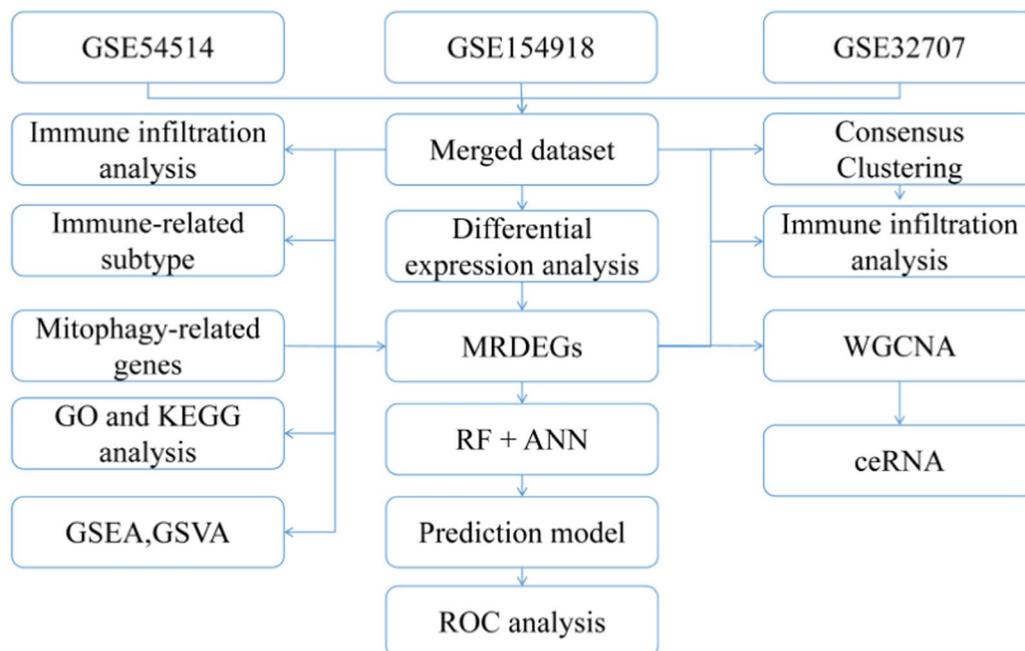
The ischemia–reperfusion expression profile data GSE154918, GSE32707, and GSE54514 were downloaded from GEO. Datasets were standardized using the “normalizeBetweenArrays” feature in the R software package “limma”, while the three expression profile data were batch calibrated and merged using the “Combat” feature in the R software package “sva” [53]. Results are presented through box-line plots and the first two principal component point plots of PCA. After normalization, we observed similar distributions of gene expression profiles between the three datasets (Fig. 2A and B). The batch effects of the three datasets were corrected (Fig. 2C and D), allowing for subsequent analysis.

### Correlation analysis of immunoinfiltration

The CIBERSORT algorithm was used to calculate the levels of 22 immune cell types in the different samples, shown by plotting stacked bar graphs (Fig. 3A). Removal of immune cells with an abundance of 0 (resting dendritic cells) and a Wilcoxon test algorithm were used to allow comparison of immune cell infiltrates between the control and sepsis conditions. Results showed 11 types of immune cells (B cells naïve, plasma cells, resting T CD4 memory cells, regulatory T cells (Tregs) gamma delta T cells, resting NK cells, monocytes, M0 macrophages, resting mast cells, activated mast cells, eosinophils, and neutrophils) significantly expressed in the two groups. The infiltration level of resting T CD4 memory cells, resting NK cells, and neutrophils was significantly increased in the sepsis group (Fig. 3C). This suggests that the above mentioned immune cells might play an essential role in sepsis. Correlation analysis in the sepsis dataset revealed that gamma delta T cells and M0 macrophages ( $R = 0.26$ ,  $P < 0.05$ ) showed a distinct correlation at higher infiltration levels in the sepsis group, whereas resting T CD4 memory cells and resting NK cells showed correlation at lower infiltration levels ( $R = 0.36$ ,  $P < 0.05$ ). Furthermore, neutrophil and monocyte counts ( $R = -0.49$ ,  $P < 0.05$ ) showed a significant positive correlation, and resting NK cells and gamma delta T cells ( $R = -0.47$ ,  $P < 0.05$ ) showed a significant negative correlation (Fig. 3B).

### Immunological signature subtype analysis

We calculated the abundance of 28 immune cell infiltrates in disease-group samples using the ssGSEA algorithm



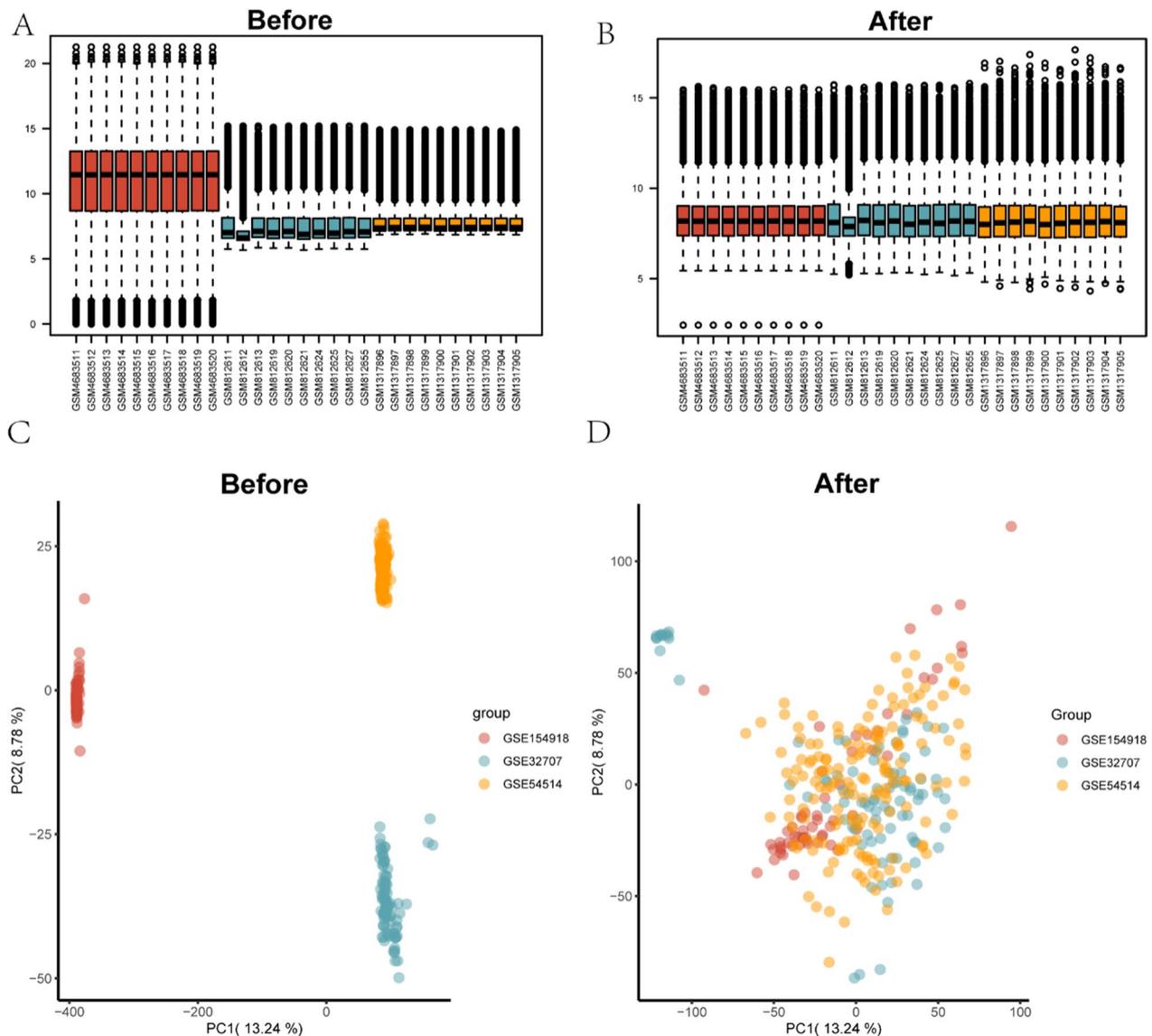
**Fig. 1** Analysis flow chart. M.DEGs, Mitophagy-related differentially expressed genes. WGCNA, weighted gene correlation network analysis. RF, random forest. ANN, artificial neural network. GO, gene ontology. KEGG, kyoto encyclopedia of genes and genomes. ROC, receiver operating characteristic. GSEA, gene set enrichment analysis. GSVA, gene set variation analysis

and then performed immune profiling of sepsis samples using unsupervised consistent agglomerative clustering, based on the immune cell infiltration abundance matrix. With this analysis we identified two different subtypes (Fig. 4A), with cluster1 containing 93 disease group samples and cluster2 containing 112 disease group samples. Figure 4B and C show the cumulative distribution function (CDF) plots of the consistent clustering results for different numbers of clusters and the area under the CDF curve delta plots. As shown in Fig. 4A, the most consistent clustering results were obtained when  $k=2$  was used as the clustering number for supervised clustering. We then performed principal component analysis (PCA) on the expression matrix of the dataset for the two disease subtype samples, and the PCA clustering results showed significant differences between the two disease subtype samples (Fig. 4D). Expression of mitochondrial autophagy-associated genes in the two different subtypes are shown by grouped comparison plots, volcano plots, and heatmaps (Fig. 4E and F), with  $|\log_{2}FC| > 0.2$  and  $P < 0.05$  as thresholds for statistical differences. This analysis showed 4 upregulated (MAP1LC3B, HIF1A, MFN2, and ATG7) and 11 downregulated genes (UBA52, SRC, TOMM7, SQSTM1, PHB, VDAC1, OPTN, UBB, TOMM22, TOMM40, and RPS27A). We also analyzed

the variation of immune cell imbibition in the two characterizing subtypes, revealing that most immune courts differed clearly in both subtypes (Activated B cells, Activated CD4<sup>+</sup> T cells, Activated CD8<sup>+</sup> T cells, Central memory CD4<sup>+</sup> T cells, Effector memory CD4<sup>+</sup> T cells, Effector memory CD8<sup>+</sup> T cells, Immature B cells, Memory B cells, Regulatory T cells, T follicular helper cell, Activated dendritic cell, CD56 dim NK cell, Eosinophil, Macrophage, Mast.cell, Myeloid-derived suppressor cell (MDSC), NK cell, and neutrophil) (Fig. 4G,  $P < 0.05$ ).

#### Differential expression and enrichment analyses

To analyze the differential gene expression profiles between control and sepsis groups, we performed a differential analysis of the combined dataset, taking the intersection of mitochondrial autophagy-related genes with sepsis DEGs. We found 8 differentially up-regulated genes in the sepsis group (TOMM20, HUWE1, SQSTM1, TOMM40, HIF1A, UBA52, PGAM5, and VPS13C) and 6 genes differentially expressed down-regulated in the sepsis group (HTRA2, ATG12, ATG7, MAP1LC3B, ATG9A, and MAP1LC3A). To analyze the biological functions associated with these differentially expressed mitochondrial autophagy, we characterized the 14 genes in terms of: BP, MF, CC, and biological routes (Tables 2



**Fig. 2** Comparison of gene expression distribution before and after bulk calibration for the GSE154918, GSE32707, and GSE54514 datasets. Box plots of gene expression distribution **A** before and **B** after bulk calibration; PCA **C** before and **D** after bulk calibration. PCA, principal component analysis

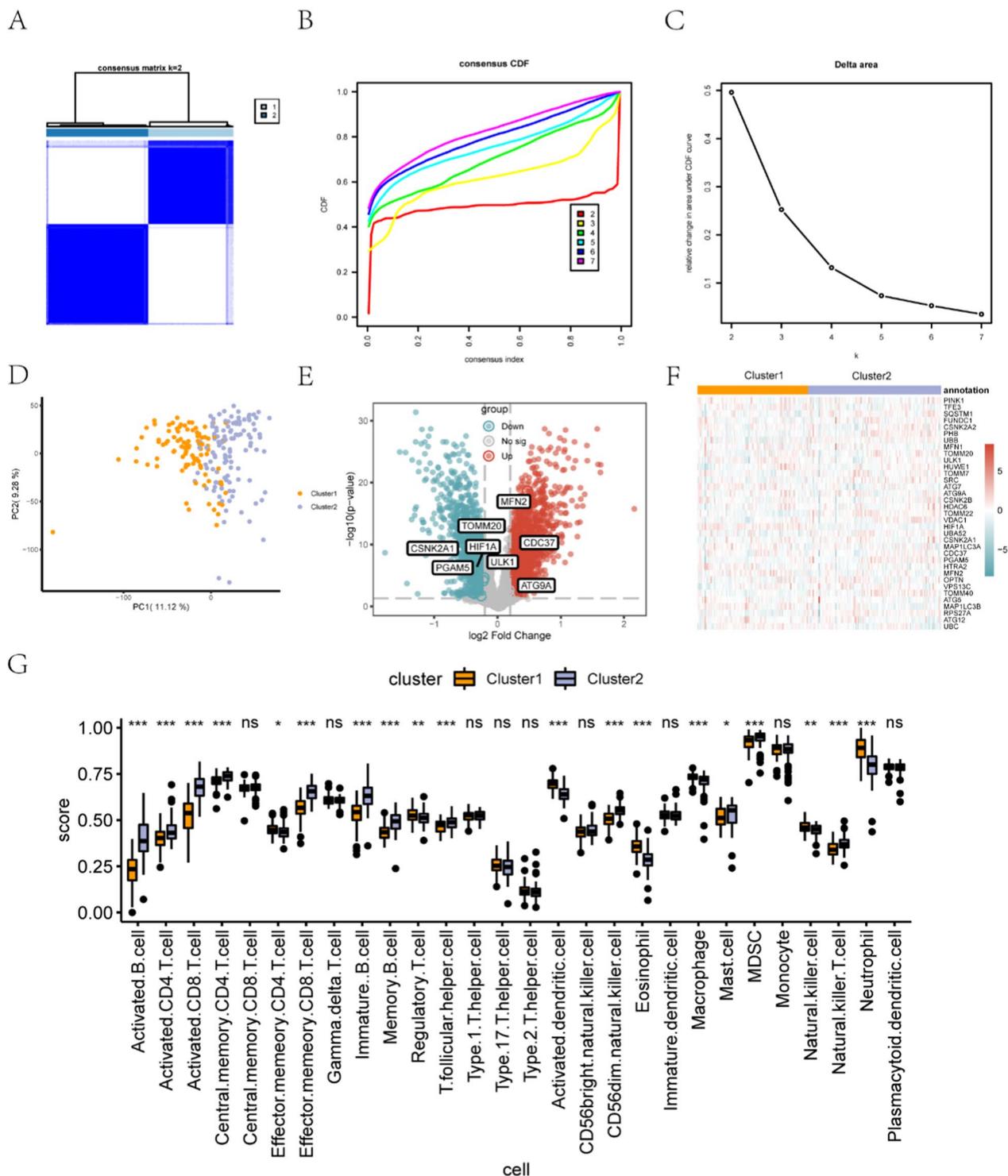
and 3). These genes were primarily correlated to entries for molecular functions such as macroautophagy, mitochondrial autophagy, mitochondrial disassembly, binding of ubiquitin protein ligases, binding of ubiquitin-like protein linkers, and protein transmembrane transporter activity, as well as entries for cellular components such as the site of phagosome assembly, autophagosome, and outer mitochondrial membrane (Fig. 5A, B). KEGG analysis revealed that differentially expressed mitochondrial autophagy-related genes were correlated with mitophagy (Fig. 5A, C), ferroptosis (Fig. 5A, D), the NOD-like

receptor signaling pathway (Fig. 5A, E), and other biological pathways.

**GSEA and GSVA**

To identify the impact of gene expression in sepsis, we analyzed the biological pathways impacted by the studied genes in the healthy controls and sepsis groups (Fig. 6A). Results showed that the toll-like receptor signaling pathway (Fig. 6B), the RIG-I-like receptor signaling pathway (Fig. 6D), and the NOD-like receptor signaling pathway (Fig. 6E) were upregulated in the sepsis group. In contrast, biological pathways





**Fig. 4** Immune signature subtyping. **A** Consistent clustering plot (k=2) results for the immune infiltration matrix. **(B, C)** CDF of consistent clustering plot for different numbers of clusters **(B)**, and area under the CDF curve **(C)**. **D** PCA of the two clusters, cluster1 colored in yellow and cluster2 in purple. **E** Key genes expression profiles in different immune signature subtypes. Red indicates up-regulated genes, cyan shows down-regulated genes. **F**: Heatmap of key gene expression, horizontal coordinates are mitochondrial autophagy-related genes, vertical coordinates are different immune signature subtypes. Red represents high gene expression, cyan represents low gene expression. **G** Comparison of gene expression in different immune cells in the two clusters; horizontal coordinates indicate cells, vertical coordinates indicate infiltration levels. Yellow color represents cluster1, purple represents cluster2. ns:  $P > 0.05$ , \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , \*\*\*\*  $P < 0.0001$ . PCA, principal component analysis; CDF, cumulative distribution function

**Table 1** List of sepsis information

	GSE154918	GSE32707	GSE54514
Species	Homo sapiens	Homo sapiens	Homo sapiens
Tissue	whole blood	whole blood	whole blood
Samples in Sepsis group	20	58	127
Samples in Normal group	40	34	36
Platform	GPL20301	GPL10558	GPL6947
References	Herwanto, V. et al. [24]	Dolinay, T. et al. [25]	Parnell, G. P. et al. [26]

**Table 2** Hub genes GO analysis: Based on GO database

ONTOLOGY	ID	Description	p.adjust
BP	GO:0016236	macroautophagy	7.15E-23
BP	GO:0000422	autophagy of mitochondrion	5.23E-22
BP	GO:0061726	mitochondrion disassembly	5.23E-22
BP	GO:1903008	organelle disassembly	2.20E-20
BP	GO:0000423	mitophagy	4.25E-12
CC	GO:0000407	phagophore assembly site	5.15E-09
CC	GO:0005776	autophagosome	5.56E-09
CC	GO:0005741	mitochondrial outer membrane	1.76E-07
CC	GO:0031968	organelle outer membrane	3.20E-07
CC	GO:0019867	outer membrane	3.35E-07
MF	GO:0031625	ubiquitin protein ligase binding	1.17E-06
MF	GO:0044389	ubiquitin-like protein ligase binding	1.59E-06
MF	GO:0008320	protein transmembrane transporter activity	9.67E-05
MF	GO:0022884	macromolecule transmembrane transporter activity	0.000137917
MF	GO:0140318	protein transporter activity	0.000186349

**Table 3** Hub genes KEGG analysis: Based on KEGG database

ID	Description	p.adjust	q value
hsa04137	Mitophagy-animal	2.31E-12	4.85E-11
hsa04140	Autophagy-animal	2.85E-10	3.00E-09
hsa04136	Autophagy-other	1.16E-05	8.17E-05
hsa05131	Shigellosis	1.60E-05	8.40E-05
hsa04216	Ferroptosis	2.48E-05	8.97E-05
hsa05022	Pathways of neurodegeneration-multiple diseases	2.56E-05	8.97E-05
hsa05167	Kaposi sarcoma-associated herpesvirus infection	0.00013	0.000392
hsa05014	Amyotrophic lateral sclerosis	0.00143	0.003764
hsa04621	NOD-like receptor signaling pathway	0.002114	0.004945
hsa04371	Apelin signaling pathway	0.016889	0.033656
hsa04120	Ubiquitin mediated proteolysis	0.017585	0.033656
hsa04217	Necroptosis	0.021762	0.038179
hsa04215	Apoptosis-multiple species	0.045928	0.074377

NAL\_ABSORPTION, BIOCARTA\_PEP1\_PATHWAY and other biologically relevant pathways (Fig. 6F, Table 5).

#### Clinical diagnostic model

To construct a diagnostic model for sepsis, we entered 14 differentially expressed mitochondrial autophagy-associated genes into a random forest algorithm to measure

their importance (Fig. 7A). The Gini index is commonly used as a gene importance measure, so we used a stochastic algorithm to rank the differentially expressed mitochondrial autophagy-related genes. We selected genes with importance greater than 0 for further neural network construction (Fig. 7B), with a total of 14 genes (HIF1A, HTRA2, ATG7, ATG9A, ATG12, MAP1LC3A, MAP1LC3B, PGAM5, SQSTM1, TOMM20, TOMM40, UBA52, HUWE1, and VPS13C). The final model consisted of an input level of 14 variables, a concealed level of 5 variables, and an outcome level for binary classification (Fig. 7D). Using this ANN model, the AUC value was 0.900 (Fig. 7C), indicating that the model was suitable for sepsis diagnose.

### Immune cells infiltration correlations

To analyze the biological link between key mitochondrial autophagy-related genes and the immune micro-environment in sepsis, we used the ssGSEA algorithm. Results showed a significant correlation between key mitochondrial autophagy-related genes and a variety of immune-infiltrating cells in sepsis (Fig. 8A). Of these, activated dendritic cells were significantly correlated with all key genes and were more strongly correlated with, HIF1A (Fig. 8B,  $R=0.3$ ), HTRA2 (Fig. 8C,  $R=0.48$ ), ATG7 (Fig. 8D,  $R=0.58$ ), MAP1LC3A (Fig. 8E,  $R=0.41$ ), PGAM5 (Fig. 8F,  $R=-0.46$ ), TOMM20 (Fig. 8G,  $R=-0.45$ ), UBA5290 (Fig. 8H,  $R=-0.4$ ), and HUWE1 (Fig. 8I,  $R=-0.5$ ). These observations suggest that immune cell infiltration by activated dendritic cells might play an instrumental role in the progression of sepsis.

### Molecular typing

To explore the subtypes associated with mitochondrial autophagy in sepsis, we used expression profile data of key genes associated with mitochondrial autophagy in sepsis to subtype a comprehensive dataset of sepsis patients by unsupervised consistent clustering. We identified two distinct subtypes (Fig. 9A), cluster1, comprising 123 sepsis samples, and cluster2, comprising 82 sepsis samples. Figure 9B and C show the CDF plots and the area under the CDF curve delta plots for the different numbers of clusters. PCA of the samples

revealed that cluster1 and cluster2 could be distinguished from each other (Fig. 9D). Moreover, the 14 key genes (HIF1A, HTRA2, ATG7, ATG9A, ATG12, MAP1LC3A, MAP1LC3B, PGAM5, SQSTM1, TOMM20, TOMM40, UBA52, HUWE1, and VPS13C), with the exception of SQSTM1, had 13 genes in both isoforms with significant differences (Fig. 9E and F  $P<0.05$ ). Of these, HIF1A, HTRA2, ATG7, ATG9A, ATG12, MAP1LC3A, and MAP1LC3B were significantly down-regulated in cluster2, and PGAM5, TOMM20, TOMM40, UBA52, HUWE1, and VPS13C were significantly up-regulated in cluster2.

### Analysis of immune infiltration in sepsis subtypes

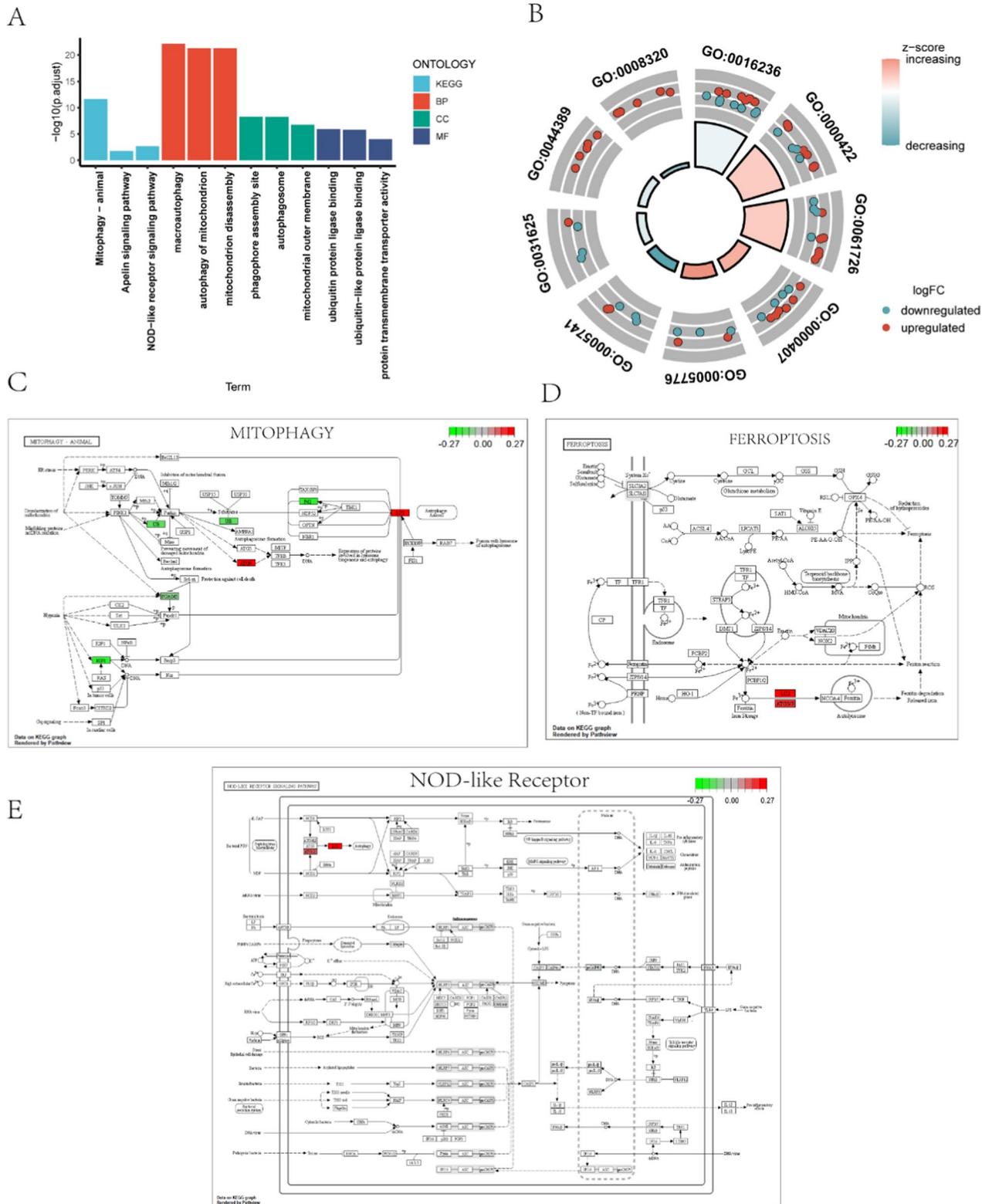
In order to investigate differences in the immune cell infiltration in the different subtypes of sepsis, we calculated the abundance of immune cells for 22 immune infiltrates in disease and control sample using the CIBERSORT algorithm (Fig. 10A) and then compared the differences in immune cell abundance in disease subtypes (Fig. 10B). We found significant differences in immune cell types infiltration between the different disease subtypes, the more significant being plasma cells, CD8<sup>+</sup> T cells, regulatory T cells (Tregs), gamma delta T cells, resting NK cells, M0 macrophages, M2 macrophages, activated mast cells, eosinophils and neutrophils ( $P<0.05$ ).

### Analysis of key genes associated with mitochondrial autophagy in sepsis

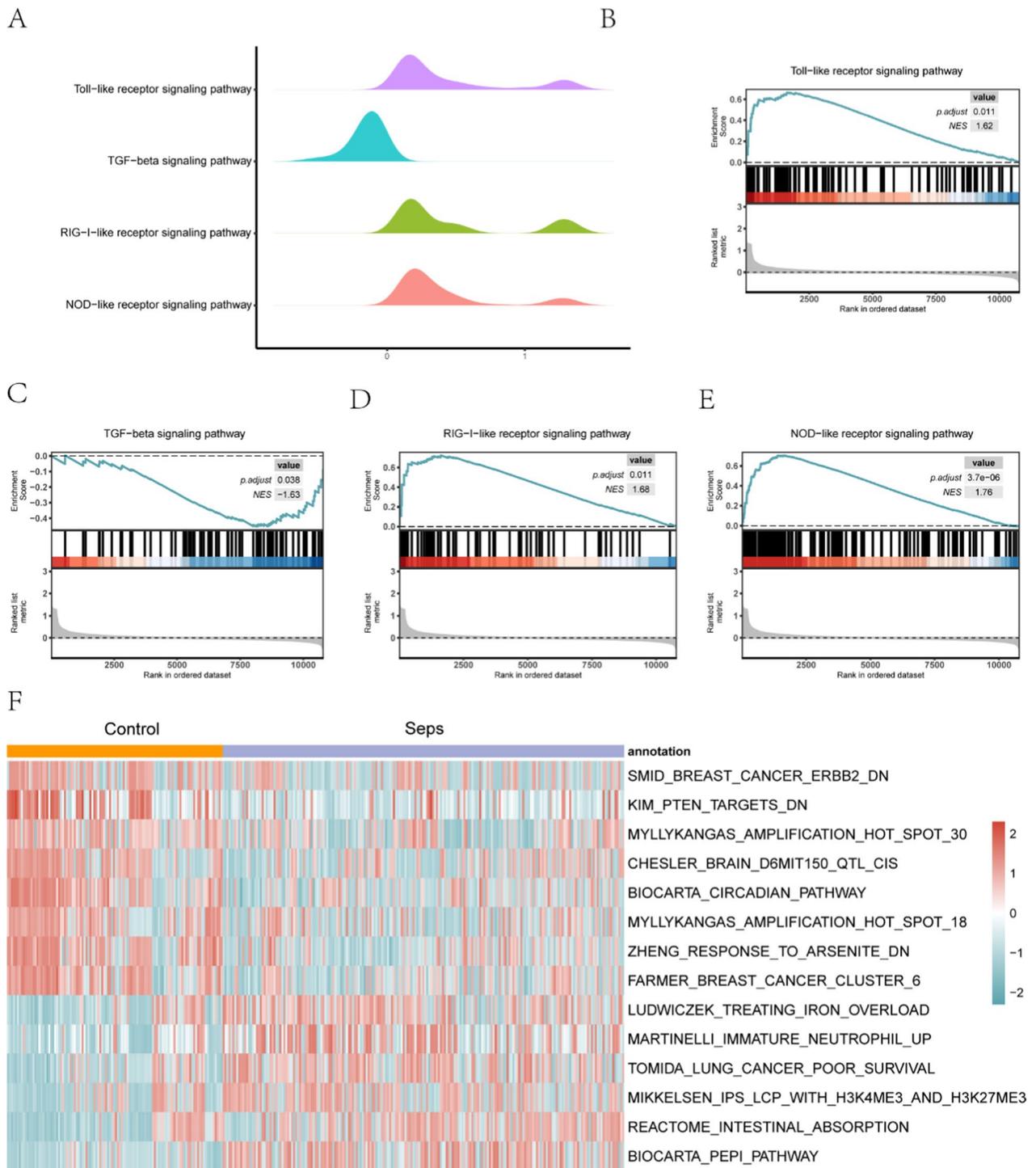
Further analysis gene expression of key genes involved in mitochondrial autophagy in sepsis revealed clear differences in both sepsis and control samples (Fig. 11A,  $P$  value  $<0.05$ ). With logistic regression profiling of the 14 genes and the findings revealed by forest plots, SQSTM1, TOMM20, HUWE1, HIF1A, UBA52, PGAM5, VPS13C, and TOMM40 were located on the left-hand side of the null hypothesis line and considered as protective genes for sepsis. On the opposite, HTRA2 was located on the right-hand side of the null hypothesis line and was found to be an unfavorable factor for sepsis (Fig. 11B). These 14 genes were also evaluated by Calibration and decision curve (DCA) analyses. The x-axis in the DCA plot (Fig. 11C) represents the probability throw, and the y-axis

(See figure on next page.)

**Fig. 5** Functional enrichment analysis. **A** GO and KEGG functional enrichment analysis: horizontal coordinates are different GO entries enriched with the differentially expressed mitochondrial autophagy-associated genes; vertical coordinates are  $-\log_{10}(P$  value). Red color indicates biological processes, green indicates cellular components, dark blue indicates molecular functions, and light blue indicates enriched biological pathways. **B** Circle diagram of GO enrichment pathways, divided in an inner and outer part. Each bar in the inner circle corresponds to an entry and the height is the relative size of the adjusted  $p$  value, the higher it is the smaller the adjusted  $p$  value for that ID. **C** Mitophagy—animal pathway, **D** Ferroptosis pathway, and **E** NOD-like receptor signaling pathway. Each node indicates a key gene in the pathway. The color of the node is determined by  $\log_2FC$ , with green indicating differentially down-regulated genes and red indicating differentially up-regulated genes. GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; BP, biological process; CC, cellular component; MF, molecular function



**Fig. 5** (See legend on previous page.)



**Fig. 6** GSEA and GSVA (Samples in Sepsis group; N = 205, Samples in Normal group; N = 110). **A** Mountain range plot of GSEA results for the healthy versus sepsis groups. **B–E** GSEA visualization; x-axis represents the rank of DEGs up-regulated (greater than zero) and down-regulated (lower than zero), the upper y-axis refers to the enrichment score and the lower y-axis is the logFC value. Each color represents a pathway: Toll-like receptor signaling pathway (**B**), RIG-I-like receptor signaling pathway (**D**), and NOD-like receptor signaling pathway (**E**), up-regulated in the sepsis group; and TGF- $\beta$  signaling pathway (**C**), down-regulated in the sepsis group. (**F**) GSVA enrichment results; vertical coordinates represent different gene set entries and horizontal coordinates represent different samples. Yellow annotation bars indicate the control group and purple bars indicate the sepsis group. GSEA, gene set enrichment analysis; GSVA: gene set variation analysis

**Table 4** Sepsis vs Control GSEA: Based on c2.cp.kegg.v7.4.entrez.gmt database

ID	Description	NES	p.adjust
hsa04740	Olfactory transduction	2.258522	3.28E-08
hsa03010	Ribosome	-2.09257	2.97E-06
hsa04621	NOD-like receptor signaling pathway	1.75864	3.69E-06
hsa04080	Neuroactive ligand-receptor interaction	1.695193	4.08E-05
hsa05150	Staphylococcus aureus infection	1.808086	0.000292
hsa03013	Nucleocytoplasmic transport	-1.84597	0.000958
hsa03008	Ribosome biogenesis in eukaryotes	-2.01971	0.001244
hsa04060	Cytokine-cytokine receptor interaction	1.512297	0.004564
hsa04622	RIG-I-like receptor signaling pathway	1.678853	0.010621
hsa04620	Toll-like receptor signaling pathway	1.619891	0.011103
hsa00982	Drug metabolism-cytochrome P450	1.705581	0.011112
hsa04623	Cytosolic DNA-sensing pathway	1.72091	0.013404
hsa05152	Tuberculosis	1.474506	0.019159
hsa03040	Spliceosome	-1.56296	0.037096
hsa05164	Influenza A	1.475047	0.037096
hsa04350	TGF-beta signaling pathway	-1.6305	0.038166
hsa03018	RNA degradation	-1.65637	0.042924

**Table 5** Sepsis vs Control GSVA: Based on c2.all.v7.5.1.symbols.gmt database

Term	logFC	adj.P.Val
LUDWICZEK_TREATING_IRON_OVERLOAD	0.404753	2.25E-10
MARTINELLI_IMMATURE_NEUTROPHIL_UP	0.414062	2.27E-11
TOMIDA_LUNG_CANCER_POOR_SURVIVAL	0.431438	8.10E-11
MIKKELSEN_IPS_LCP_WITH_H3K4ME3_AND_H3K27ME3	0.483689	1.17E-06
REACTOME_INTESTINAL_ABSORPTION	0.509022	8.56E-10
BIOCARTA_PEPI_PATHWAY	0.627122	3.87E-15
SMID_BREAST_CANCER_ERBB2_DN	-0.48367	1.15E-06
KIM_PTEN_TARGETS_DN	-0.43807	5.57E-13
MYLLYKANGAS_AMPLIFICATION_HOT_SPOT_30	-0.43658	7.83E-09
CHESLER_BRAIN_D6MIT150_QTL_CIS	-0.42936	8.76E-07
BIOCARTA_CIRCADIAN_PATHWAY	-0.4229	7.38E-13
MYLLYKANGAS_AMPLIFICATION_HOT_SPOT_18	-0.41775	2.01E-07
ZHENG_RESPONSE_TO_ARSENITE_DN	-0.41705	5.66E-14
FARMER_BREAST_CANCER_CLUSTER_6	-0.40038	8.67E-12

represents the net gain. The prediction line (dashed line) of the model constructed based on the 14 genes was steadily higher than the All and None lines based on the range of x-values to determine the clinical effect, with a larger range of x-values being more effective. The clinical effect was better. The main function of the calibration curve graph (Fig. 11D) was to assess the diagnostic accuracy of the disease. The fitted curve in Fig. 11D shows

a high overlap with the prediction curve (dashed line), indicating that the 14 genes are good disease predictors and may be potential therapeutic markers for sepsis diagnosis.

### WGCNA

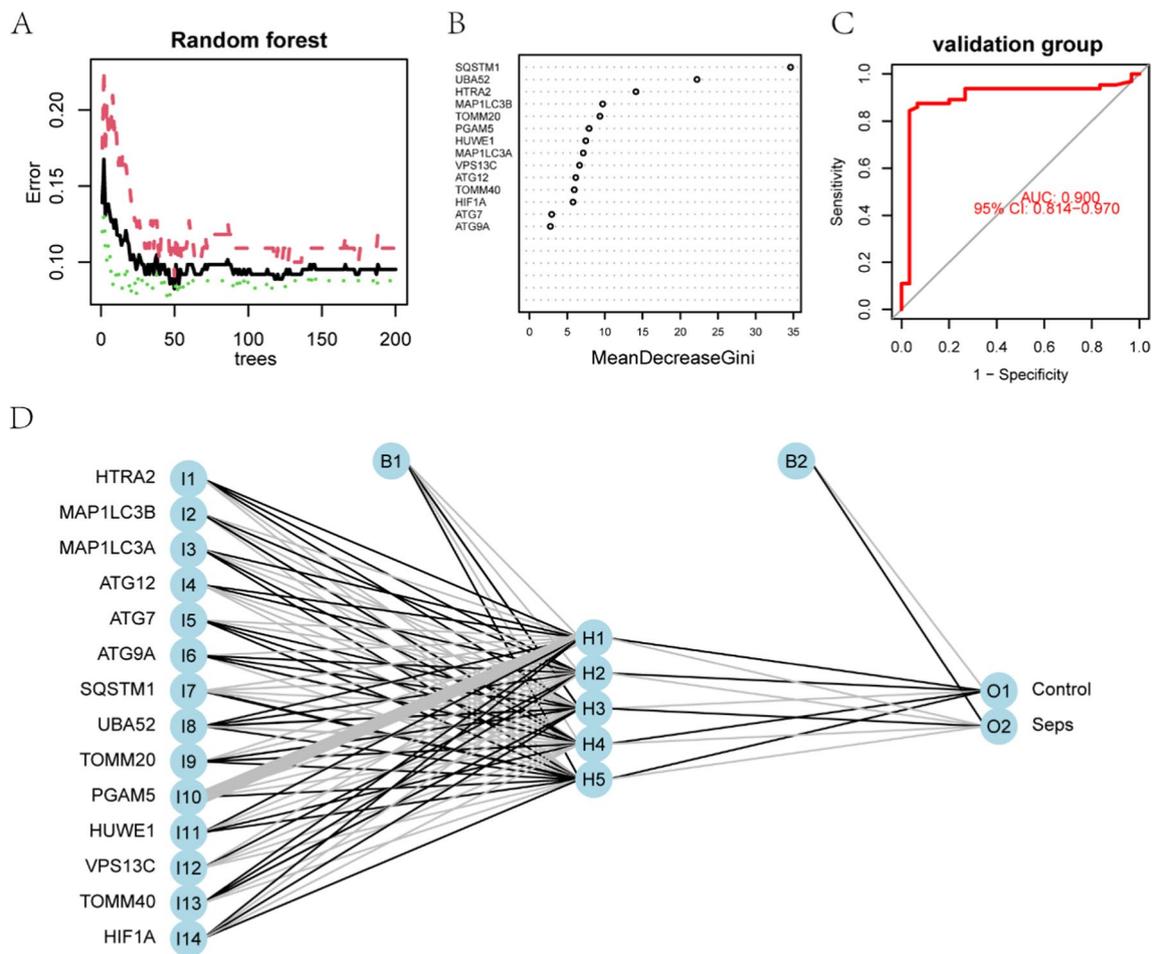
By exploring the co-expression of genes in profiles of the sepsis group, using the R package “WGCNA”, we constructed co-expression modules using the top 5,000 genes in this dataset. The samples were clustered using Pearson’s correlation coefficients in combination with the clinical information of the patients (Fig. 12A). Next, we set a soft threshold of 9 to construct a scale-free network (Fig. 12B), where  $R^2 > 0.85$ , and network connectivity was high. Finally, we set the smallest module number of genes to 50 and the combined module cut height to 0.2, to calculate the eigengenes. Using the dynamic cut tree algorithm, we obtained a total of 17 gene co-expression blocks, divided into different colors (Fig. 12C). We then analyzed the correlation between the different modules and the sepsis group (Fig. 12D). The light cyan and turquoise modules were positively correlated with the sepsis group ( $P < 0.05$ ), with the turquoise module having the strongest correlation ( $r = 0.34$ ,  $P = 6e-10$ ). The red module was negatively associated with sepsis ( $r = -0.28$ ,  $P = 3e-07$ ). We chose the turquoise module, to intersect with the differentially expressed mitochondrial autophagy genes, and obtain 2 interest genes (ATG7 and HTRA2) (Fig. 12E).

### Construction of ceRNA networks for key genes

To explore the potential ceRNA regulatory network of key genes, we first entered the previously selected key genes into the miRcode database to obtain miRNAs that might target the key genes associated with mitochondrial autophagy in sepsis (total score  $> 50$ ). Next, we predicted the lncRNAs to which the selected miRNAs might bind, selecting only miRNA-lncRNA relationships with high scores (total score  $> 70$ ). Finally, by integrating mRNA-miRNA and miRNA-lncRNA interaction pairs, we obtained a hub gene-associated ceRNA regulatory network consisting of 23 miRNAs, 2 mRNAs, and 95 lncRNAs, visualized using the Cytoscape software (Fig. 13A). Our “cytoHubba” plug-in clustered the ceRNA interaction network and identified key modules within it. We identified a key ceRNA module (Fig. 13B) that suggested ATG7 as an important player in sepsis.

### Validation of autophagy-related genes in sepsis

The 14 hub genes were further validated by quantitative PCR (Fig. 14), and the two groups were statistically significant in ATG9A (Fig. 14C), HTRA2 (Fig. 14E),



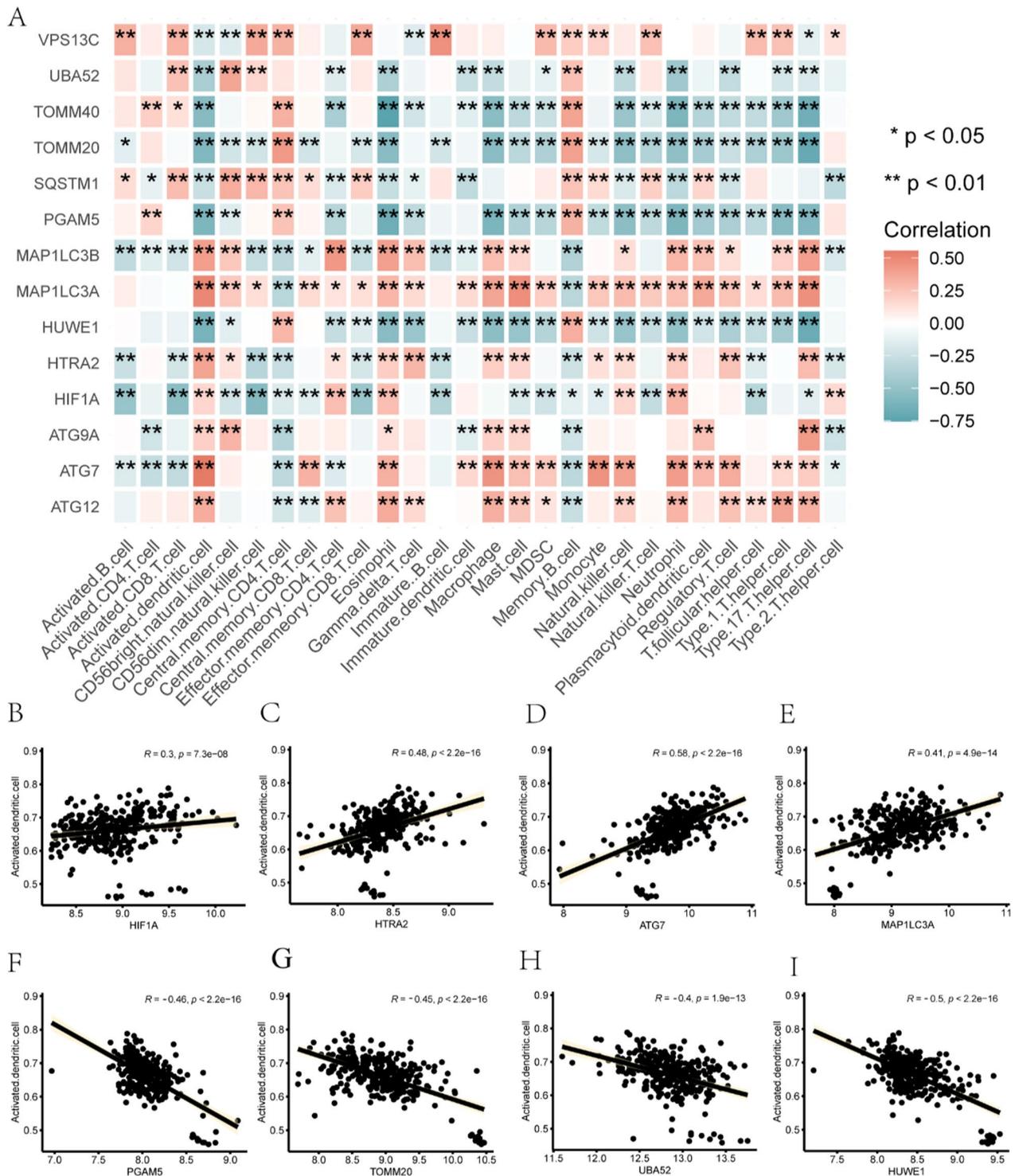
**Fig. 7** Construction of the clinical prediction model (Samples in Sepsis group; N=205, Samples in Normal group; N=110). **A** Relationship between the number of decision trees and the error of cross-validation; vertical coordinates indicate the error, and horizontal coordinates indicate the number of decision trees. The green line shows the error for the control group, the red line shows the error for the sepsis group, and the black line shows the error for all samples. **B** Scatter plot of gene importance in random forest variables; the mean falling Gini index at the horizontal coordinate and the different genes at the vertical coordinate. **C** ROC curves; horizontal coordinates define specificity, vertical coordinates define sensitivity, area under the curve and 95% confidence intervals are annotated in the lower right corner of the Fig **D**: From left to right, the nodes indicate the input layer, hidden layer, and different groupings of the neural network. The grey lines represent negative connection weights and the black lines represent positive connection weights. ROC, receiver operating characteristic

HUWE1 (Fig. 14F), MAP1LC3A (Fig. 14G), MAP1LC3B (Fig. 14H), SQSTM1 (Fig. 14J), TOMM40 (Fig. 14L), UBA52 (Fig. 14M).

## Discussion

In sepsis, organ malfunction results from a dysfunctional response to injury. This disease affects more than 30 million individuals every year, causing more than six million deaths worldwide [54, 55]. Despite the immense efforts to find effective treatments, the mortality rate associated with sepsis remains alarmingly high (15–50%). Reducing mortality has become the goal of treatment approaches [56, 57].

The analysis of gene expression profiles has greatly influenced the medical field. In oncology, molecular isoforms of the disease have been determined [58], transcriptional features that predict clinical outcomes [59] and response to particular treatments [60]. However, the impact of genomic science is far from universal in critical care medicine. Sepsis progresses rapidly, leading to systemic organ failure and consequent mortality. Although current research has identified genetic factors involved in biological functions that might contribute to individual differences in sepsis and its complex pathophysiological mechanisms [61], the clinical diagnostic value of this single biomarker is suboptimal.



**Fig. 8** Immune cells infiltration correlation. **A** Heatmap showing the correlation analysis between mitochondrial autophagy-related key genes and immune cells in sepsis; the horizontal axis indicates different immune cells and the vertical axis indicates mitochondrial autophagy-related key genes in sepsis. The more intense the red color, the stronger the correlation between the genes and immune cells. **B** HIF1A, **C** HTRA2, and **D** ATG7 correlation with activated dendritic cells; horizontal coordinates are the expression level of each gene and vertical coordinates are the ssGSEA enrichment score of activated dendritic cells. Scatterplot of the correlation between genes MAP1LC3A **E**, PGAM5 **F**, TOMM20 **G**, UBA5290 **H**, HUWE1 **I** and activated dendritic cells. ns:  $P > 0.05$ , \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , \*\*\*\*  $P < 0.0001$ . Correlation coefficients ( $r$ )  $> 0.8$  are strongly correlated;  $r: 0.5-0.8$  are moderately correlated;  $r: 0.3-0.5$  are weakly correlated;  $r < 0.3$  are weakly or not correlated

Accumulating evidence suggests that autophagy plays an essential role in sepsis pathogenesis. Activation of various autophagy-associated proteins in early stage of sepsis can reduce cytokine release and attenuate inflammatory responses [62]. Others have argued that autophagy may accelerate the disease in advanced stages, through the excessive accumulation of autophagosomes [63, 64]. Therefore, identifying the different players and pathways involved in autophagy in sepsis, through bioinformatics, will help us understand the real impact of autophagy in the development of this disease.

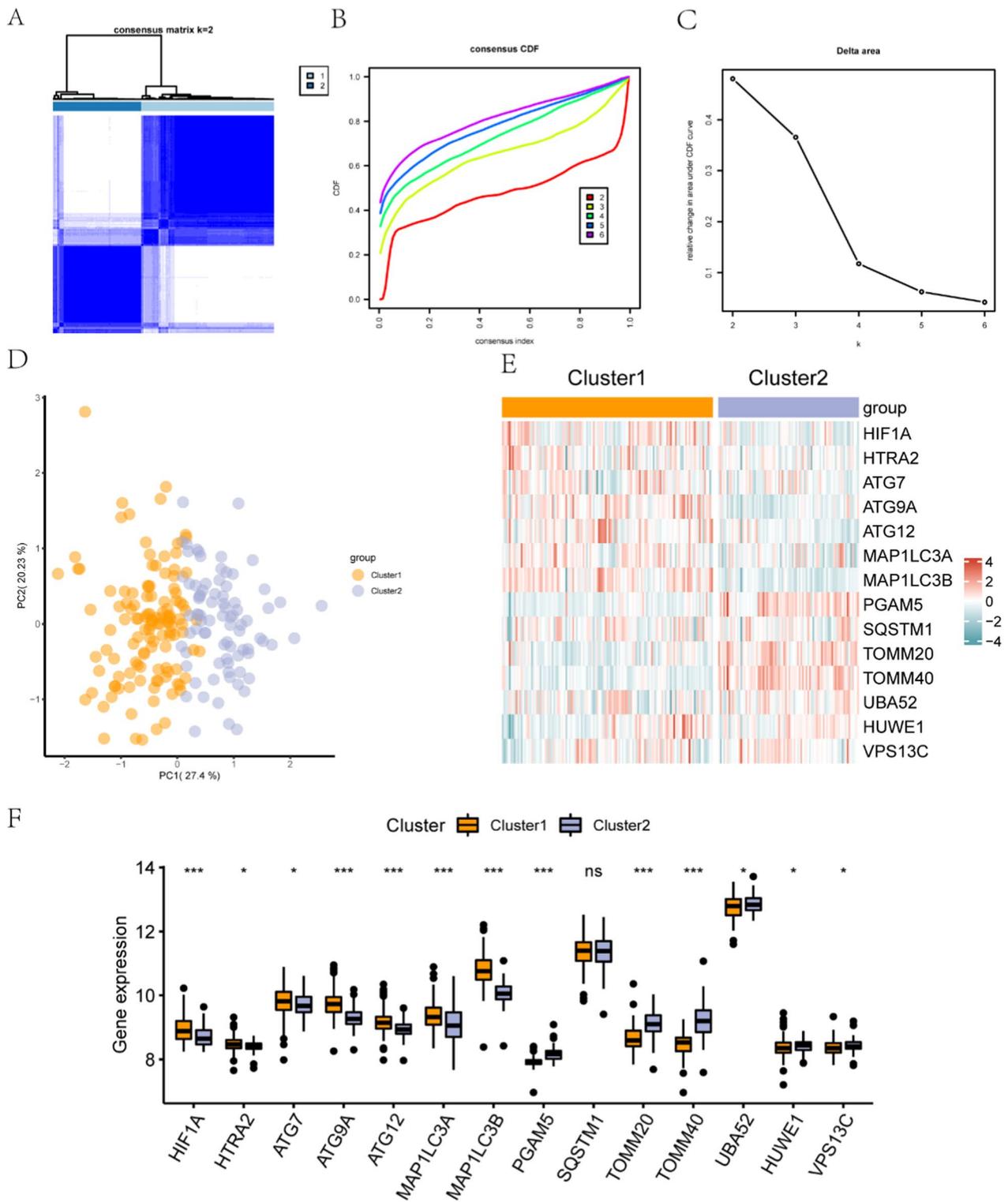
In our study, DEGs and gene clusters related to sepsis were identified in the GSE154918, GSE32707, and GSE54514 datasets, through the DEGs and WGCNA analyses. In addition, patients with sepsis from the three datasets were grouped into an unsupervised cluster using the “Consensus Cluster Plus” package. A grand total of 14 co-DEGs were selected from the crossover of DEGs, specific gene clusters, and mitochondrial autophagy-associated genes. We employed RF and ANN to construct a sepsis prediction model with DCA, determining the 14 DEARs (HIF1A, HTRA2, ATG7, ATG9A, ATG12, MAP1LC3A, MAP1LC3B, PGAM5, SQSTM1, TOMM20, TOMM40, UBA52, HUWE1, and VPS13C) as specific and sensitive for sepsis diagnosis. Immunoassays using CIBERSORT revealed 11 immune cell infiltrations at significantly different levels. We also constructed a ceRNA network of key genes, and identified a key ceRNA module, which suggested that ATG7 might play an important role in sepsis. Frame-shifting mutations in the ATG gene cause premature cessation of amino acid synthesis of the affected protein and might disrupt autophagy [65]. Others have supported the treatment of acute myeloid leukemia (AML) through targeting ATG7, an autophagy regulator [66]. ATG9 is the only evolutionarily conserved membrane-spanning protein known among the ATG gene products and may be involved in membrane trafficking and cycling during autophagosome biogenesis. In addition, ATG9 enhances filamentous cell outgrowth in elementary fetal neurons, indicating a regulatory function in the general motility of the actin cytoskeleton [67]. Recently, ATG12 silencing

was shown to significantly decrease the growth of breast cancer cells in mice, suggesting an oncogenic activity of this ATG [68]. HTRA2 is a serine protease localized in the intermembrane space, which is 4,180 base pairs long and is commonly expressed in human tissues [69]. Mutations that inactivate HTRA2 have been reported in the literature to be associated with Parkinson’s disease (PD) and idiopathic tremor, neurodegenerative disorders [70, 71]. It has also been proposed that HTRA2 is also linked to other neurodegenerative diseases such as Alzheimer’s disease (AD) [72]. Many reports have linked HTRA2 dysregulation to cancer. In non-small cell lung cancer (NSCLC) tissues, reduced HTRA2 levels correlate with clinical stage and histological differentiation [73]. Furthermore, p53 activates HTRA2, which regulates cell invasion by inducing alterations in actin polymerisation and subsequent phosphorylation of p130Cas [74]. At that time the role of the HTRA2 gene in the pathogenesis of sepsis was currently unreported. HUWE1 is an E3 ubiquitin-conjugating enzyme with a HECT structural region [75], playing an essential role in the regulation of apoptosis through the polyubiquitination of mcl-1 [76]. Furthermore, HUWE1 protein levels increase during differentiation, suggesting a role in differentiation [75]. Additionally, HUWE1 regulates mitochondrial morphology and function [76]. Although few studies have investigated the association of these 14 key genes with sepsis, our findings suggest their potential as biomarkers for the diagnosis of sepsis at an early stage.

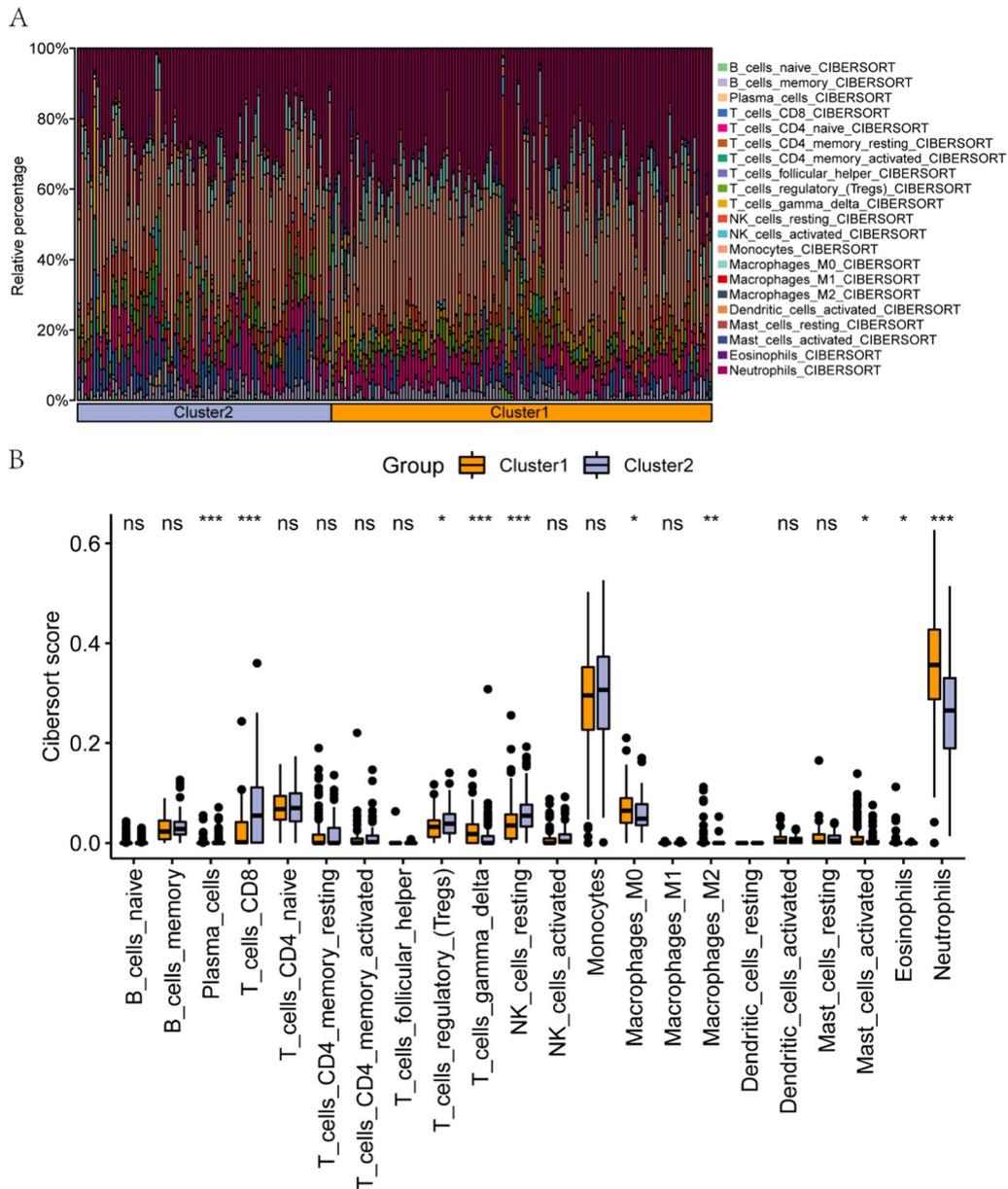
Different patients with sepsis display different gene expression profiles, which may be due to the variety of routes involving various key genes. In this study, to analyze the divergent expression profiles of the control and sepsis groups, we performed a differential analysis of the combined dataset, using genes with  $\text{adj}P$  value  $< 0.05$  as a threshold, and taking the intersection of mitochondrial autophagy-related genes with sepsis differentially expressed genes, obtaining these 14 key genes. We further analyzed the biological processes, molecular functions, cellular components, and biological pathways of these 14 genes and revealed that these were mostly related to mitochondrial autophagy, mitochondrial disassembly, binding of ubiquitin protein ligases, binding

(See figure on next page.)

**Fig. 9** Molecular typing and immune infiltration. **A** Plot of consistent clustering ( $k=2$ ) results for the immune infiltration matrix. **B, C** Plot of the cumulative distribution function of consistent clustering for different numbers of clusters (**B**), and area under the CDF curve delta plot (**C**). **D** PCA of the two clusters, cluster1 is represented in yellow and cluster2 in purple. **E** Heatmap of key gene expression; horizontal coordinates are key genes associated with sepsis mitochondrial autophagy and vertical coordinates are different subtypes of sepsis. Red represents high gene expression, cyan represents low gene expression. **F** Comparison of key genes expression in the two clusters; horizontal coordinates represent genes and vertical coordinates represent gene expression. Cluster1 is represented in yellow and cluster2 in purple. ns:  $P > 0.05$ , \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , \*\*\*\*  $P < 0.0001$



**Fig. 9** (See legend on previous page.)

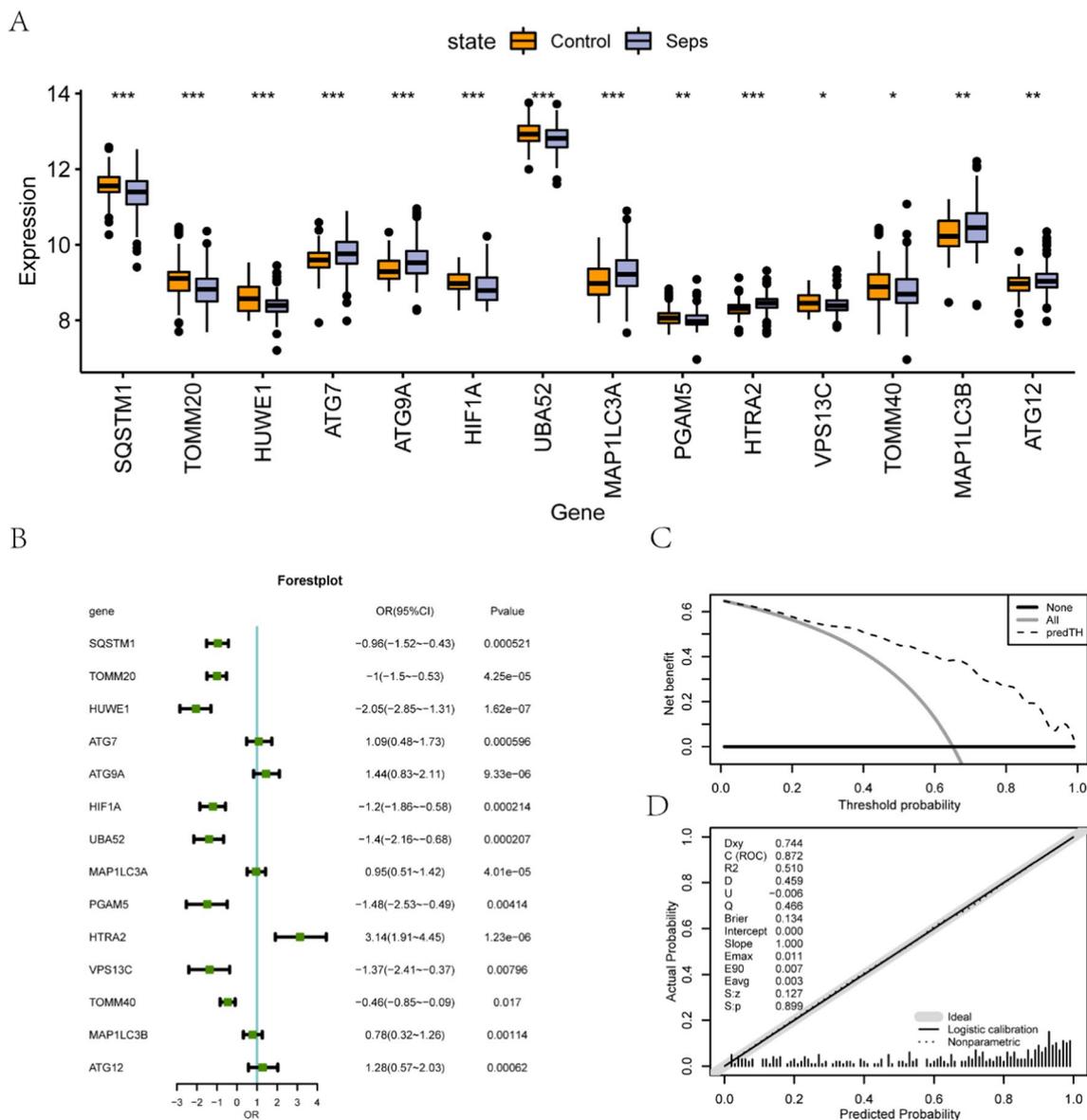


**Fig. 10** Immune infiltration analysis. **A** Bar plot displaying the immune infiltration analysis in different subtypes of sepsis, in the combined dataset. **B** Comparison of the 22 immune cell infiltrates in sepsis cluster1 and cluster2. ns:  $P > 0.05$ , \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ , \*\*\*\*  $P < 0.0001$

of ubiquitin-like protein ligases, and protein transmembrane transporter activity.

KEGG analysis revealed that the differentially expressed mitochondrial autophagy-related genes were related to biological pathways, such as mitophagy, ferroptosis, and NOD-like receptor signaling pathways. Ferroptosis is a specialized form of iron-dependent cell death related to fatal lipid peroxidation, and indeed iron toxicity has been related to a wide range of pathophysiological conditions and illnesses [77]. Autophagy is being extensively studied

for the treatment of several diseases [78, 79]. Autophagy acts as an immunomodulator and inhibits tissue injury following sepsis by modulating the export of a wide range of immune stores, and has been shown to be able to mitigate post-sepsis organ failure caused by oxidative stress [80]. We cross-tabulated the identified DEGs with disease subtype-associated specific genes and found significant correlations between most of them. GO, KEGG, and ssGSEA analyses suggested that the immune reaction might be strongly involved in the severity of sepsis.

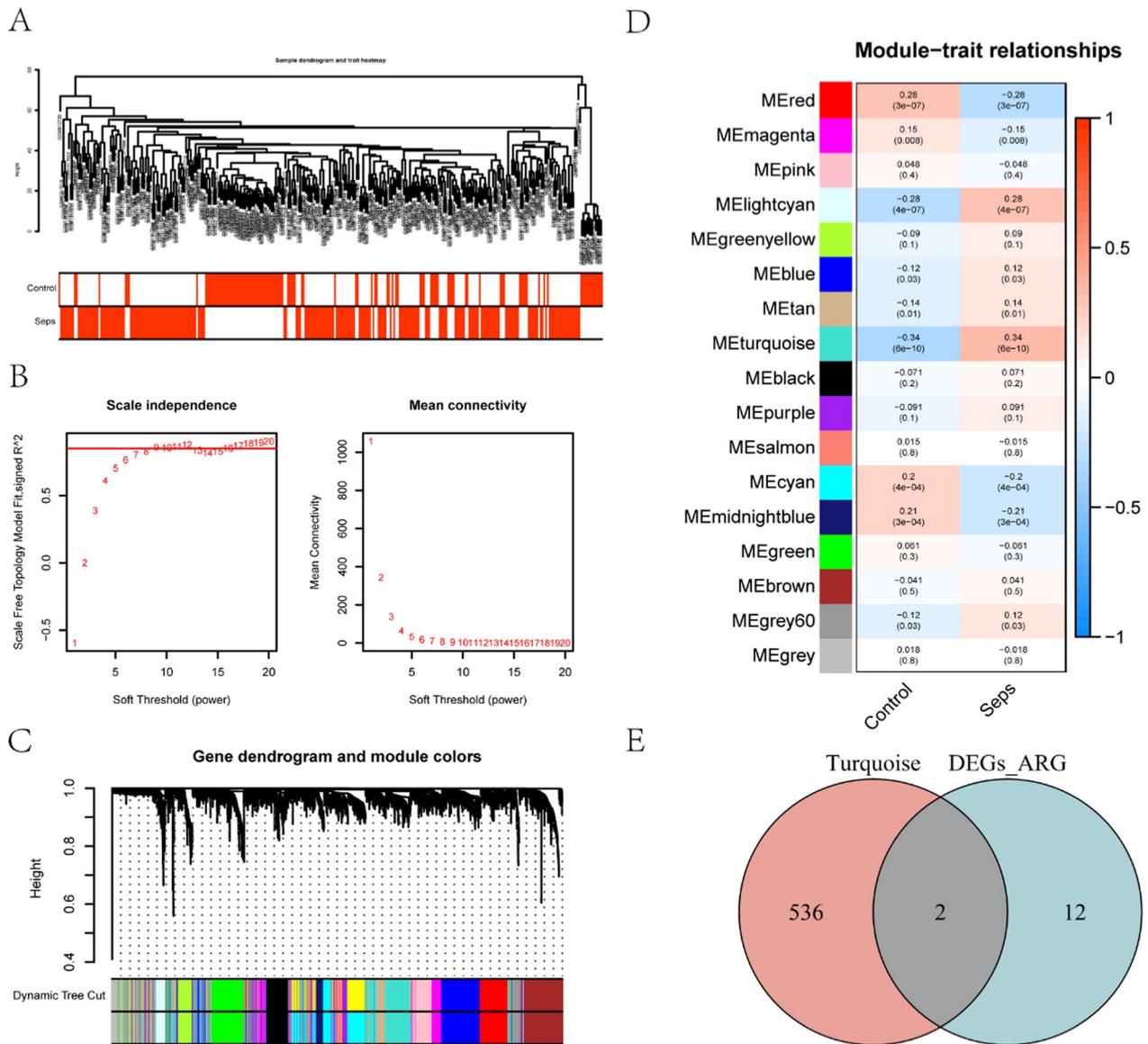


**Fig. 11** Analysis of key genes associated with mitochondrial autophagy in sepsis (Samples in Sepsis group; N = 205, Samples in Normal group; N = 110). **A** Grouped comparison plot of the selected 14 genes in the sepsis and control groups. **B** Forest plot of 14 genes by logistic regression analysis to calculate clinical and disease predictive effects, demonstrated by DCA plots (**C**) and calibration curves (**D**). ns:  $P > 0.05$ ,  $** P < 0.01$ ,  $*** P < 0.001$ ,  $**** P < 0.0001$ . DCA: decision curve analysis

Thus, we postulate that the immune response in interplay with autophagy and ferroptosis routes may be key factors in the evolution of sepsis. These distinct pathways could uncover new potential therapeutic targets for patients with sepsis, according with different disease subtypes and genetic profile.

Sepsis may damage the innate and adaptive immune responses of the host, predisposing to primary and secondary infections [81]. Several immune cell types are involved in this process, such as neutrophils,

lymphocytes, and monocytes/macrophages, which recognize and engulf pathogens triggering the release of cytokines that, in turn, activate other cells [82]. We computed the abundance of immune cells for 22 immune infiltrates in disease and control sample species using the CIBERSORT algorithm on the combined dataset; among them, plasma cells, CD8<sup>+</sup> T cells, regulatory T cells (Tregs), gamma delta T cells, resting NK cells, M0 macrophages, M2 macrophages, activated mast cells, eosinophils, and neutrophils were significantly different



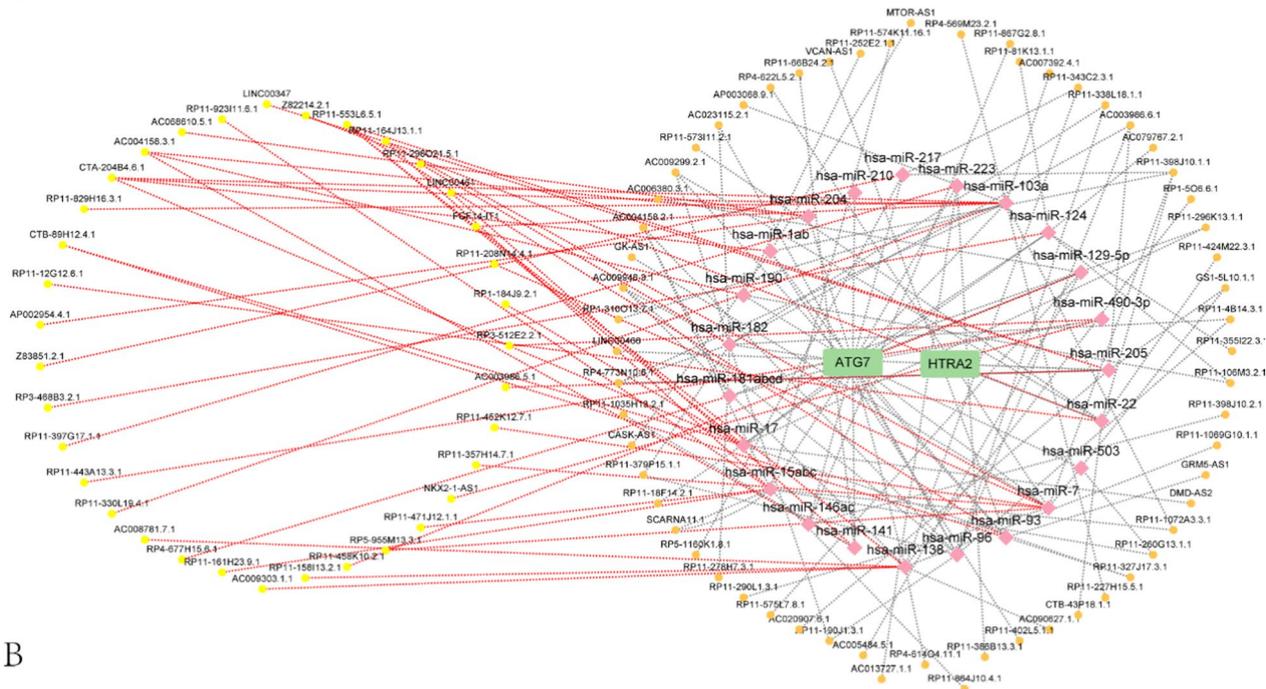
**Fig. 12** WGCNA based on the gene expression spectrum of the dataset. **A** Sample clustering was used to detect anomalous samples; each branch represents 1 sample and the red box in the sample information indicates the category of each sample. **B** Scale-free topological model fit (left) and mean connectivity (right), used to determine the best soft threshold. The horizontal coordinate of the left panel indicates the soft threshold of fit and the vertical coordinate indicates the level of fit (R-squared) in the scale-free topological model. The right panel indicates the soft threshold of fit in horizontal coordinates and the average connectivity between modules in vertical coordinates. **C** Dynamic shear clustering tree for different genes. The upper tree shows gene co-expression, with each gene represented by a leaf in the tree and each module represented by a main trunk branch. The lower colored bars indicate the corresponding 17 modules, labelled in the indicated colors. **D** Heatmap of the correlation between different modules and the sepsis group; x-axis refers to sample classification and y-axis to the module color. Red in the correlation heatmap indicates positive correlation and dark blue indicates negative correlation. Textual meanings of the color blocks in the heatmap: Pearson correlation coefficient in the upper part; statistical p-value in the lower brackets. **E** Turquoise module genes and mitochondrial autophagy differential genes intersection

among the different sepsis subtypes ( $p < 0.05$ ). Macrophages exhibit an immunosuppressive role and have been shown to play a deleterious role in sepsis, through CASP11-dependent pyrogenesis [82, 83]. Our study

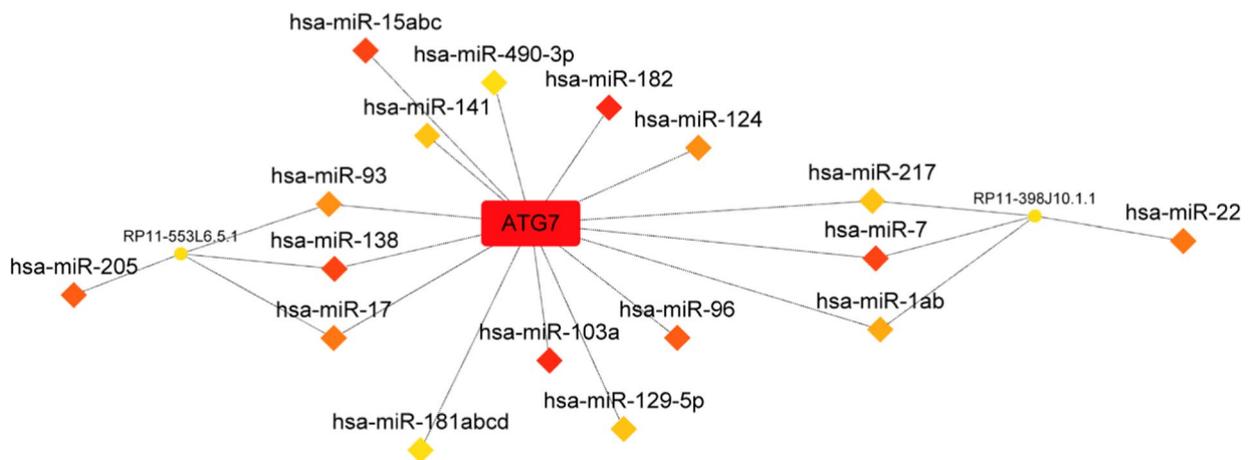
confirmed that immune cells play a determinant role in the pathogenesis of sepsis.

This study has some limitations. The expression levels of biomarkers, which were not validated by wet experiments, as well as the lack of corresponding clinical

A



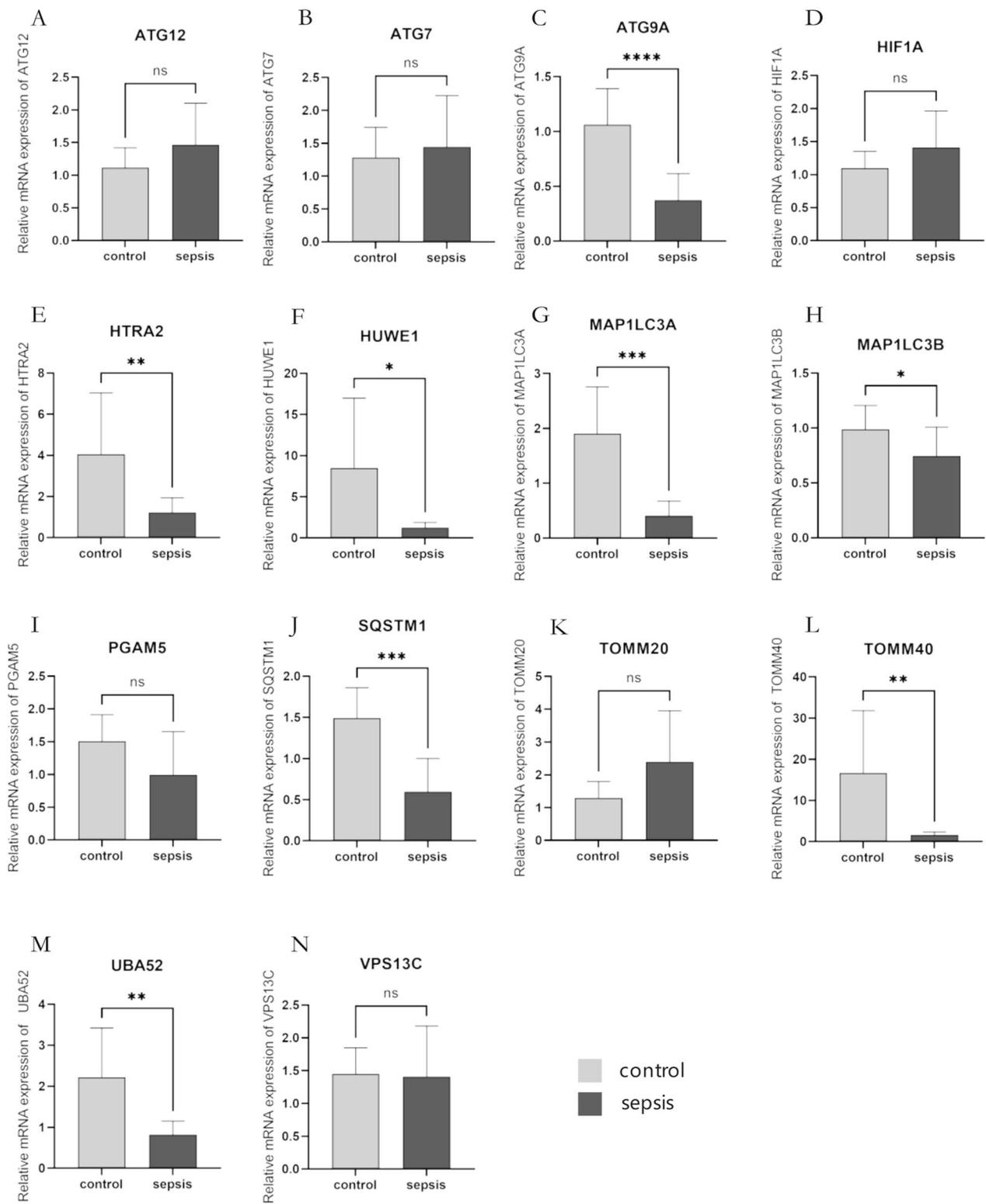
B



**Fig. 13** ceRNA network construction for hub genes. **A** Key genes-associated ceRNA network construction based on the miRcode database; green nodes indicate mRNAs, pink nodes indicate miRNAs, and yellow nodes indicate lncRNAs. **B** Clusters of key genes-associated ceRNA networks, with square ones for mRNAs, prismatic ones for miRNAs, and round ones for lncRNAs

(See figure on next page.)

**Fig. 14** PCR of hub genes. **A** PCR of ATG12 gene in control and sepsis groups. **B** PCR of ATG7 gene in control and sepsis groups. **C** PCR of ATG9A gene in control and sepsis groups. **D** PCR of HIF1A gene in control and sepsis groups. **E** PCR of HTRA2 gene in control and sepsis groups. **F** PCR of HUWE1 gene in control and sepsis groups. **G** PCR of MAP1LC3A gene in control and sepsis groups. **H** PCR of MAP1LC3B gene in control and sepsis groups. **I** PCR of PGAM5 gene in control and sepsis groups. **J** PCR of SQSTM1 gene in control and sepsis groups. **K** PCR of TOMM20 gene in control and sepsis groups. **L** PCR of TOMM40 gene in control and sepsis groups. **M** PCR of UBA52 gene in control and sepsis groups. **N** PCR of VPS13C gene in control and sepsis groups



**Fig. 14** (See legend on previous page.)

correlation between these genes and clinical features, could potentially and unexpectedly adversely affect the precision of the prediction model and the results obtained in this study. Therefore, these results should be interpreted with caution.

## Conclusions

In this study, we explored the pathogenesis of sepsis using bioinformatics tools, screened for potential key biomarkers and validated them by PCR and constructed a diagnostic model for sepsis. In addition, we found a strong association between DEPRGs and the immune microenvironment during sepsis. These findings provide new insights into the mechanisms underlying sepsis and highlight potential new molecular therapeutic targets.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12985-025-02683-0>.

Additional file 1 (DOCX 13 KB)

## Acknowledgements

Not applicable.

## Author contributions

Jian Lu and Yu Zhang designed this study. Dong-po Wei and Wei-wei Jiang performed the data analysis and drew the pictures. Chang-xing Chen, Zi-yang Chen and Fang-qing Zhou contributed to the writing of the protocol and setting of figures. All authors read and approved the final manuscript.

## Funding

Not applicable.

## Availability of data and materials

Research data available in the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/gds>) database.

## Declarations

### Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of Shanghai First People's Hospital (2024112).

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no conflicts of interest.

Received: 31 March 2024 Accepted: 25 February 2025

Published online: 20 March 2025

## References

- Keeley A, Hine P, Nsutebu E. The recognition and management of sepsis and septic shock: a guide for non-intensivists. *Postgrad Med J*. 2017;93(1104):626–34.
- Weis S, et al. Metabolic adaptation establishes disease tolerance to sepsis. *Cell*. 2017;169(7):1263–75.
- Pan B, et al. The fifth epidermal growth factor like region of thrombomodulin alleviates LPS-induced sepsis through interacting with GPR15. *Thromb Haemost*. 2017;117(3):570–9.
- Uchimido R, Schmidt EP, Shapiro NI. The glycocalyx: a novel diagnostic and therapeutic target in sepsis. *Crit Care*. 2019;23(1):16.
- Beurskens DM, et al. Decreased endothelial glycocalyx thickness is an early predictor of mortality in sepsis. *Anaesth Intensive Care*. 2020;48(3):221–8.
- Zhang Q, Li CS. Risk stratification and prognostic evaluation of endothelial cell-specific molecule1, von Willebrand factor, and a disintegrin-like and metalloprotease with thrombospondin type 1 motif for sepsis in the emergency department: an observational study. *Exp Ther Med*. 2019;17(6):4527–35.
- Zhao Y, Vanhoutte PM, Leung SW. Vascular nitric oxide: beyond eNOS. *J Pharmacol Sci*. 2015;129(2):83–94.
- Kowalczyk A, et al. The role of endothelin-1 and endothelin receptor antagonists in inflammatory response and sepsis. *Arch Immunol Ther Exp (Warsz)*. 2015;63(1):41–52.
- Zeh HJ, et al. A randomized phase II preoperative study of autophagy inhibition with high-dose hydroxychloroquine and gemcitabine/nab-paclitaxel in pancreatic cancer patients. *Clin Cancer Res*. 2020;26(13):3126–34.
- Reyes M, et al. An immune-cell signature of bacterial sepsis. *Nat Med*. 2020;26(3):333–40.
- Sun Y, et al. Beclin-1-dependent autophagy protects the heart during sepsis. *Circulation*. 2018;138(20):2247–62.
- Zhao H, et al. Autophagy activation improves lung injury and inflammation in sepsis. *Inflammation*. 2019;42(2):426–39.
- Vishnupriya S, et al. Autophagy markers as mediators of lung injury-implication for therapeutic intervention. *Life Sci*. 2020;260: 118308.
- Yang L, et al. EGFR TKIs impair lysosome-dependent degradation of SQSTM1 to compromise the effectiveness in lung cancer. *Signal Transduct Target Ther*. 2019;4:25.
- Vo MT, et al. Activation of NIX-mediated mitophagy by an interferon regulatory factor homologue of human herpesvirus. *Nat Commun*. 2019;10(1):3203.
- Ilyas G, et al. Macrophage autophagy limits acute toxic liver injury in mice through down regulation of interleukin-1beta. *J Hepatol*. 2016;64(1):118–27.
- Neumann Y, et al. Intracellular *Staphylococcus aureus* eludes selective autophagy by activating a host cell kinase. *Autophagy*. 2016;12(11):2069–84.
- Oami T, et al. Suppression of T cell autophagy results in decreased viability and function of T cells through accelerated apoptosis in a murine sepsis model. *Crit Care Med*. 2017;45(1):e77–85.
- Pena OM, et al. An endotoxin tolerance signature predicts sepsis and organ dysfunction at initial clinical presentation. *EBioMedicine*. 2014;1(1):64–71.
- Sweeney TE, et al. Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Crit Care Med*. 2018;46(6):915–25.
- Lukaszewski RA, et al. Presymptomatic diagnosis of postoperative infection and sepsis using gene expression signatures. *Intensive Care Med*. 2022;48(9):1133–43.
- Barrett T, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acids Res*. 2007;35:D760–5.
- Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846–7.
- Herwanto V, et al. Blood transcriptome analysis of patients with uncomplicated bacterial infection and sepsis. *BMC Res Notes*. 2021;14(1):76.
- Dolinay T, et al. Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am J Respir Crit Care Med*. 2012;185(11):1225–34.
- Parnell GP, et al. Identifying key regulatory genes in the whole blood of septic patients to monitor underlying immune dysfunctions. *Shock*. 2013;40(3):166–74.
- Liberzon A, et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–25.
- Jassal B, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48(D1):D498–503.

29. Chen H, et al. Development and validation of a novel mitophagy-related gene prognostic signature for hepatocellular carcinoma based on immunoscore classification of tumor. *J Oncol*. 2021;2021:5070099.
30. Newman AM, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773–82.
31. Zeng D, et al. IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front Immunol*. 2021;12: 687975.
32. Bindea G, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39(4):782–95.
33. Zhang G, et al. m6A regulatory gene-mediated methylation modification in glioma survival prediction. *Front Genet*. 2022;13: 873764.
34. The Gene Ontology C. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330–8.
35. Ashburner M, et al. Gene ontology: tool for the unification of biology the gene ontology consortium. *Nat Genet*. 2000;25(1):25–9.
36. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
37. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
38. Wu T, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*. 2021;2(3): 100141.
39. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*. 2013;29(14):1830–1.
40. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545–50.
41. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
42. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14(1):1–15.
43. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7): e47.
44. Liaw, A. and M. Wiener, *Classification and Regression by Random Forest*. 2001.
45. Beck MW. NeuralNetTools: visualization and analysis tools for neural networks. *J Stat Softw*. 2018;85(11):1–20.
46. Robin X, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. 2011;12:77.
47. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
48. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell*. 2016;29(4):452–63.
49. Lock EF, Dunson DB. Bayesian consensus clustering. *Bioinformatics*. 2013;29(20):2610–6.
50. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–3.
51. Zhang B, et al. m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol Cancer*. 2020;19(1):53.
52. Rezk M, et al. Biomarker screening in preeclampsia: an RNA-sequencing approach based on data from multiple studies. *J Hypertens*. 2022;40(10):2022–36.
53. Leek JT, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3.
54. Yan J, et al. The BH3-only protein BAD mediates TNF $\alpha$  cytotoxicity despite concurrent activation of IKK and NF- $\kappa$ B in septic shock. *Cell Res*. 2018;28(7):701–18.
55. Fleischmann C, 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):259–72.
56. Jensen IJ, et al. Sepsis-induced T cell immunoparalysis: the ins and outs of impaired T cell immunity. *J Immunol*. 2018;200(5):1543–53.
57. Nguyen HB, et al. Early goal-directed therapy in severe sepsis and septic shock: insights and comparisons to ProCESS, ProMISe, and ARISE. *Crit Care*. 2016;20(1):160.
58. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503–11.
59. van 't Veer LJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530–6.
60. Paez JG, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*. 2004;304(5676):1497–500.
61. Maslove DM, Wong HR. Gene expression profiling in sepsis: timing, tissue, and translational considerations. *Trends Mol Med*. 2014;20(4):204–13.
62. Xu Y, et al. SPIONs enhances IL-10-producing macrophages to relieve sepsis via Cav1-Notch1/HES1-mediated autophagy. *Int J Nanomed*. 2019;14:6779–97.
63. Lo S, et al. Lc3 over-expression improves survival and attenuates lung injury through increasing autophagosomal clearance in septic mice. *Ann Surg*. 2013;257(2):352–63.
64. Guo L, et al. CaMKII $\alpha$  regulates AMP kinase-dependent, TORC1-independent autophagy during lipopolysaccharide-induced acute lung neutrophilic inflammation. *J Immunol*. 2013;190(7):3620–8.
65. Kang MR, et al. Frameshift mutations of autophagy-related genes ATG2B, ATG5, ATG9B and ATG12 in gastric and colorectal cancers with microsatellite instability. *J Pathol*. 2009;217(5):702–6.
66. Piya S, et al. Atg7 suppression enhances chemotherapeutic agent sensitivity and overcomes stroma-mediated chemoresistance in acute myeloid leukemia. *Blood*. 2016;128(9):1260–9.
67. Kiss V, et al. Drosophila Atg9 regulates the actin cytoskeleton via interactions with profilin and Ena. *Cell Death Differ*. 2020;27(5):1677–92.
68. Cufi S, et al. Autophagy-related gene 12 (ATG12) is a novel determinant of primary resistance to HER2-targeted therapies: utility of transcriptome analysis of the autophagy interactome to guide breast cancer treatment. *Oncotarget*. 2012;3(12):1600–14.
69. Gibellini L, et al. Mitochondrial proteases as emerging pharmacological targets. *Curr Pharm Des*. 2016;22(18):2679–88.
70. Strauss KM, et al. Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Hum Mol Genet*. 2005;14(15):2099–111.
71. Unal Gulsuner H, et al. Mitochondrial serine protease HTRA2 p.G3995 in a kindred with essential tremor and Parkinson disease. *Proc Natl Acad Sci USA*. 2014;111(51):18285–90.
72. Behbahani H, et al. Association of Omi/HtrA2 with gamma-secretase in mitochondria. *Neurochem Int*. 2010;57(6):668–75.
73. Mao G, et al. The expression levels and prognostic value of high temperature required A2 (HtrA2) in NSCLC. *Pathol Res Pract*. 2014;210(12):939–43.
74. Yamauchi S, et al. p53-mediated activation of the mitochondrial protease HtrA2/Omi prevents cell invasion. *J Cell Biol*. 2014;204(7):1191–207.
75. Cho HI, et al. BRPF3-HUWE1-mediated regulation of MYST2 is required for differentiation and cell-cycle progression in embryonic stem cells. *Cell Death Differ*. 2020;27(12):3273–88.
76. Zhao Z, et al. Hepatic PPAR $\alpha$  function is controlled by polyubiquitination and proteasome-mediated degradation through the coordinated actions of PAQR3 and HUWE1. *Hepatology*. 2018;68(1):289–303.
77. Chen PH, et al. Kinome screen of ferroptosis reveals a novel role of ATM in regulating iron metabolism. *Cell Death Differ*. 2020;27(3):1008–22.
78. Levy RJ. Mitochondrial dysfunction, bioenergetic impairment, and metabolic down-regulation in sepsis. *Shock*. 2007;28(1):24–8.
79. Su LJ, et al. Reactive oxygen species-induced lipid peroxidation in apoptosis, autophagy, and ferroptosis. *Oxid Med Cell Longev*. 2019;2019:5080843.
80. Thiessen SE, et al. The role of autophagy in critical illness-induced liver damage. *Sci Rep*. 2017;7(1):14150.
81. McBride MA, et al. The metabolic basis of immune dysfunction following sepsis and trauma. *Front Immunol*. 2020;11:1043.
82. Wen X, et al. The “self-sacrifice” of immunecells in sepsis. *Front Immunol*. 2022;13: 833479.
83. Dash SP, Chakraborty P, Sarangi PP. Inflammatory monocytes and subsets of macrophages with distinct surface phenotype correlate with specific integrin expression profile during murine sepsis. *J Immunol*. 2021;207(11):2841–55.

## Publisher's Note

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