

## Research Paper

# Comprehensive Transcriptome Profiling of Peripheral Blood Mononuclear Cells from Patients with Sepsis

Tianzhou Wu<sup>1\*</sup>, Xi Liang<sup>1\*</sup>, Yongpo Jiang<sup>2\*</sup>, Qi Chen<sup>1</sup>, Huaping Zhang<sup>6</sup>, Sheng Zhang<sup>2</sup>, Chao Zhang<sup>3</sup>, Yuhang Lv<sup>6</sup>, Jiaojiao Xin<sup>1,4</sup>, Jing Jiang<sup>1,4</sup>, Dongyan Shi<sup>1,4</sup>, Xin Chen<sup>1,5</sup>✉, Jun Li<sup>1,4</sup>✉ and Yinghe Xu<sup>6</sup>✉

1. Precision Medicine Center, Taizhou Central Hospital, Taizhou University Medical School, Taizhou, China.
2. Department of intensive care unit, Taizhou Hospital of Zhejiang Province, Wenzhou Medical University, Taizhou, China.
3. Department of intensive care unit, Taizhou Enze Medical Center (Group) Enze Hospital, Taizhou, China.
4. State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative National Clinical Research Center for Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine.
5. Institute of Pharmaceutical Biotechnology, Zhejiang University School of Medicine, Hangzhou, China.
6. Department of intensive care unit, Taizhou Central Hospital, Taizhou University Medical School, Taizhou, China.

\*These authors contributed equally to this work.

✉ Corresponding authors: Yinghe Xu. M.D. Professor, Department of intensive care unit, Taizhou Central Hospital, Taizhou University Medical School, Taizhou, China, E-mail: xuyh@tzc.edu.cn; Co-corresponding: Jun Li. M.D. & Ph.D. Professor, E-mail: lijun2009@zju.edu.cn; Xin Chen. Ph.D. Professor. E-mail: xinchen@zju.edu.cn.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2020.04.11; Accepted: 2020.07.14; Published: 2020.07.25

## Abstract

**Background:** Sepsis, as a clinical emergency, usually causes multiorgan dysfunction and can lead to high mortality. Establishment of specific and sensitive biomarkers for early diagnosis is critical to identify patients who would benefit from targeted therapy. In this study, we investigated this syndrome by analyzing the transcriptome of peripheral blood mononuclear cells (PBMCs) from patients with sepsis and identified sepsis-specific biomarkers.

**Methods:** In this study, a total of 87 patients with sepsis and 40 healthy controls from a prospective multicenter cohort were enrolled. Samples from 44 subjects (24 patients with sepsis and 20 healthy controls) were sequenced and the remaining patients were included in the validation group. Using high-throughput sequencing, a gene expression profile of PBMCs from patients with sepsis was generated to elucidate the pathophysiology of sepsis and identify sepsis-specific biomarkers.

**Results:** Principal component analysis (PCA) and unsupervised hierarchical cluster analysis showed that patients with sepsis separated from healthy controls. A total of 1639 differentially expressed genes (DEGs) were identified ( $|\log_2 \text{fold change}| > 2$ , adjusted P value  $< 0.05$ ) between these two groups, with 1278 (78.0%) upregulated and 361 (22.0%) downregulated in patients with sepsis. Gene Ontology (GO) analysis of the upregulated DEGs identified 194 GO terms that were clustered into 27 groups, and analysis of the downregulated DEGs identified 20 GO terms that were clustered into 4 groups. Four unique genes were identified that could be predictive of patients with sepsis. External validation of the four genes using quantitative real-time polymerase chain reaction (qRT-PCR) was consistent with the results of mRNA sequencing, revealing their potential in sepsis diagnosis.

**Conclusions:** The transcriptome characteristics of PBMCs, which were significantly altered in sepsis patients, provide new insights into sepsis pathogenesis. The four identified gene expression changes differentiated patients with sepsis from healthy subjects, which could serve as a convenient tool contributing to sepsis diagnosis.

Key words: sepsis; transcriptomics; peripheral blood mononuclear cells; biomarker

## Introduction

Sepsis is a potentially life-threatening condition in the intensive care unit (ICU) caused by the body's response to an infection, that can lead to tissue

damage, organ failure, and death. The incidence of sepsis is very high (2.8 million deaths per year worldwide) in adults, with high mortality rates of 10-

50%, depending on age and disease severity [1, 2]. Despite the burden on patients and the healthcare service system, treatment remains mainly futile. Thus, precise diagnosis and targeted therapy for sepsis is crucial to reducing mortality, with every hour of delay increasing mortality risk [3].

Recent findings in transcriptome analysis have given us a greater understanding of the mechanisms behind diseases. Genome-wide gene expression profiling is extensively used to discover new potential biomarkers for the diagnosis or prediction of disease severity and identification of novel drug targets. However, tissue sampling has limited the accurate investigation of sepsis pathogenesis. Therefore, blood can be used as a surrogate and can be obtained with a minimally invasive procedure. Many studies have indicated that the messenger ribonucleic acid (mRNA) expression levels of specific genes in peripheral blood mononuclear cells (PBMCs) can serve as a signature of specific diseases, such as active pulmonary tuberculosis [4] and enterocolitis syndrome [5]. Recently, many genes have also been identified for sepsis diagnosis and prognosis; for example, SeptiCytte Lab combined with CEACAM4, LAMP1, PLA2G7, and PLAC8 genes was identified to determine which patients had sepsis [6]. Meanwhile, seven genes (DYRK2, CCNB1IP1, TDRD9, ZAP70, ARL14EP, MDC1, and ADGRE3) had been identified as predictors of septic patients and their relationships with clinical outcomes were observed [7]. However, these biomarkers lack sensitivity and specificity and have been analyzed by microarray data [8, 9].

We hypothesized that the pathological process of sepsis changes the transcriptome profile of PBMCs. In this study, we characterized the transcriptome profile of PBMCs in patients with sepsis using high-throughput sequencing to establish alterations present in gene expression patterns and identify potential biomarkers that could serve as a convenient tool to improve sepsis diagnosis.

## Methods

The study was approved by the Clinical Research Ethics Committee of Taizhou Central Hospital (Taizhou University Medical School) (Registration number: 2019-016, Principal investigator: Yinghe Xu, Date of registration: February 26, 2019). This study was registered in the Chinese Clinical Trial Registry (ChiCTR1900022081, Principal investigator: Yinghe Xu, Date of registration: March 21, 2019). The trial was registered prior to patient enrollment. Written informed consent was obtained from all participants or their legal representative.

## Study design

Patients were prospectively screened at 3 participating ICUs and enrolled between April 1, 2019, and November 20, 2019. The last follow-up was completed on February 18, 2020. Adult patients (aged >18 years) who were admitted to the ICU with sepsis were enrolled in the study. Healthy subjects (with 0 in SOFA and without infection) were also recruited as the control group from Physical Examination Center during the same time. Sepsis was defined as the presence of an infection combined with an acute change in SOFA score of 2 or more points [10]. Detailed clinical data and outcomes for all enrolled patients were collected and recorded in case report forms at admission and during the 28/90-day follow-up.

## Methods of mRNA sequencing/mRNA sequencing sample preparation

Samples were obtained from 105 patients with an acute change in SOFA score of 2 or more points following admission to ICU, and 10 mL whole blood sample were taken. PBMCs were immediately isolated by Ficoll-Paque™ PLUS medium (GE Healthcare, Uppsala, Sweden). Eighty-seven patients diagnosed with infection were finally defined as sepsis, and their samples were used for mRNA sequencing or validation, as well as, samples of 40 healthy controls. Total RNA, which was freshly extracted, was stored at -80°C for subsequent testing. A sequencing library was then prepared following the manufacturer's instructions (TruSeq™ Small RNA Sample Preparation Kit, Illumina, San Diego, CA, USA), consisting of purification of poly-A-containing mRNA molecules, fragmentation of mRNA, end repair, addition of a single 'A' base, ligation of adapters, reverse transcription, PCR amplification and pooled gel purification steps. The pooled library consisted of sequences with lengths of approximately 250 nucleotides. The library was sequenced using the HiSeq 2500 sequencing system (Illumina, San Diego, CA, USA).

## Quantitative real-time polymerase chain reaction

Quantitative real-time polymerase chain reaction (qRT-PCR) was used to validate the results of the transcriptome analysis. Total RNA was isolated from PBMCs and reverse transcribed. qRT-PCR was performed using a two-step protocol. Complementary DNA (cDNA) was synthesized using a RevertAid First Strand cDNA Synthesis Kit (Thermo, IL, USA). The amount of cDNA was optimized so that amplification of both control gene cDNA and the cDNAs of interest were in the exponential phase.

qRT-PCRs were performed with PowerUp SYBR Green Master Mix (Thermo, IL, USA) using specific gene primers (BioSune, Shanghai, China) following the manufacturer's instructions. The specific quantitative primers for 5 mRNAs are provided in Supplementary Table 1. Thermal cycling was performed using a 7500 Real-Time PCR instrument (ABI, CA, USA). The target genes were assayed in triplicate on each plate. Target mRNA levels were quantified using the comparative Ct method and normalized using glyceraldehyde phosphate dehydrogenase (GAPDH) as a housekeeping gene.

### Enzyme-linked immunosorbent assay

Flow cytometric bead array (CBA) was used as a rapid determination tool for Th1/Th2 cytokines, including IL-2, IL-4, IL-6, IL-10, TNF and IFN- $\gamma$ . The quantification of the 6 cytokines was evaluated by a FACSria II Cell Sorter (BD, CA, USA) with a human Th1/Th2 subpopulation detection kit (CEGER, Zhejiang, China). The data of the samples acquired from the FACSria II were analyzed by BD FCAP Array v3.0.1 software. Six standard curves were obtained from one set of calibrators, and six results were obtained from one test sample. The maximum and minimum limits of the six detected cytokines were 1.0 and 5000 pg/ml, respectively. Biochemical measurement was performed by a Siemens advia2400 (Siemens, BER, GER).

### Bioinformatics

The bioinformatics pipeline used to process RNA-Seq data is presented in Supplementary Figure 1. The read quality of the sequencing data was evaluated by using Fastqc [11]. Adapter contamination and low-quality read filtering were performed by using Trimmomatic [12] v0.36 with default parameters. Paired-end reads were mapped to the human genome reference GRCh38.87 by using HISAT [13] v2.0.5. All parameters were set to the default values, including mismatch tolerance. The output SAM files were sorted and converted to BAM files by samtools v1.3.1. Raw read counts of each gene were computed by using HTSeq [14]. Differential expression analysis was performed by using DESeq2 [15]. Genes with low abundance were excluded from downstream analysis if their total expression was <10 and if their expression variance was <10 across samples. Significance was defined as an adjusted  $P$  value <0.05, which was determined by using the Benjamini-Hochberg procedure (as implemented in the R function `p.adjust`) to control type I error in multiple tests. The significantly upregulated or downregulated genes in patients with sepsis were then submitted to ClueGO [16] for annotation

enrichment analysis. The enriched Gene Ontology (GO) terms with a false discovery rate of <0.05 were considered significantly enriched.

### Statistical analysis

Data are expressed as the mean  $\pm$  standard deviation, median with interquartile range or percent with number of patients. Normality was evaluated by using the Shapiro-Wilk test ( $P$  value >0.05). Comparisons between groups of continuous variables, which were normally distributed, were carried out using Student's  $t$ -test. Comparisons of other scenarios were performed using the Mann-Whitney U test. Comparisons of categorical variables were performed using the  $\chi^2$  test.  $P$  values <0.05 were considered significant.

All statistical analyses were performed by using the R software package. Principal component analysis (PCA) of the different groups' samples was performed on the gene expression matrix by using the 'sva' function [17]. Unsupervised hierarchical clustering was performed with all protein-coding genes except low-abundance genes by using the 'pheatmap' function [18], based on Ward's cluster method and the Euclidean distance matrix. The significantly differentially expressed genes (DEGs) were analyzed by using the DESeq2 package. An adjusted  $P$  value <0.05 was considered to indicate a significant difference, which employed the Benjamini-Hochberg method for multiple-test correction. The DEGs validated via qRT-PCR were identified using Student's  $t$ -test. A  $P$  value <0.05 was considered significant.

## Results

### Patient and clinical characteristics

The demographic and clinical characteristics at admission of patients with sepsis and of the healthy controls whose PBMCs were sequenced are summarized in Table 1. A total of 24 patients with sepsis and 20 healthy controls were enrolled in this study. The SOFA score was  $7.6 \pm 3.7$ , and the APACHE II score was  $22.3 \pm 6.8$ . Among these patients, one patient had a gram-positive bacterial infection, more than 30% had a gram-negative bacterial infection, and one patient had a viral infection. The short-term (28/90 day) mortality of the patients with sepsis was 29.2%/29.2%. The levels of laboratory indices, including white blood cell count, hemoglobin, hematocrit, platelet count, albumin, aspartate aminotransferase, and creatinine, were significantly worse in the sepsis patients compared to those seen in healthy controls.

**Table 1.** Characteristics of enrolled patients and healthy subjects included in the sequencing group

	Sepsis	Healthy	P value
<b>n</b>	24	20	
Male (%)	20 (83.3%)	11 (55.0%)	0.086
Age (years)	76.5 [69.3, 80.3]	69.0 [67.2, 74.0]	0.067
SOFA	7.6 ± 3.7		
APACHE II	22.3 ± 6.8		
<b>Infection</b>			
Gram-positive bacteria (%)	1 (4.2%)		
Gram-negative bacteria (%)	9 (37.5%)		
Viral (%)	1 (4.2%)		
Other (%)	13 (54.1%)		
<b>CRRT</b>	2 (8.3%)		
<b>Vasopressors</b>	16 (66.7%)		
<b>Mechanical ventilation</b>	13 (54.2%)		
<b>Mortality</b>			
28-day	7 (29.2%)		
90-day	7 (29.2%)		
<b>Laboratory Data</b>			
MAP (mmHg)	73.1 ± 16.6		
White blood cell count	12.9 [10.4, 18.2]	5.8 [5.2, 6.6]	<0.001
Hemoglobin (g/L)	114.5 ± 29.4	142.8 ± 20.6	0.001
Hematocrit (%)	35.6 [27.3, 40.9]	44.5 [38.0, 47.2]	0.007
Platelet count	110.3 ± 79.0	212.2 ± 73.1	<0.001
Albumin (g/dL)	25.81 ± 5.09	44.3 ± 2.1	<0.001
Aspartate aminotransferase	44.0 [29.5, 84.5]	16.0 [14.0, 27.0]	<0.001
Alanine aminotransferase	27.0 [17.0, 54.5]	20.0 [17.0, 25.0]	0.150
Total bilirubin (μmol/L)	13.8 [8.0, 18.9]	16.5 [12.0, 20.5]	0.273
Creatinine (μmol/L)	225.5 ± 129.2	69.8 ± 15.6	<0.001
INR	1.2 [1.1, 1.4]		

APACHE II = Acute Physiology and Chronic Health Evaluation II;

INR = International Normalized Ratio; MAP = mean arterial pressure;

SOFA = Sequential Organ Failure Assessment on day of sampling;

Data are expressed as the mean ± standard deviation (SD), median (interquartile range) or number of patients (percentages).

The continuous variables were compared by using Student's t-test and the Mann-Whitney U test, and the categorical variables were compared by using the  $\chi^2$  or Fisher's exact test between the discovery and validation groups.

### Qualitative analysis of the PBMC transcriptomes of patients with sepsis and healthy controls

To identify changes in the transcriptome profile associated with sepsis, the transcriptomes of PBMCs from 44 subjects (24 patients with sepsis, 20 healthy controls) were characterized. The number of raw paired-end sequencing reads for each sample ranged from 20.8 million to 58.2 million, and the number of clean reads for each sample ranged from 15.1 million to 55.8 million. The average amount of clean data was approximately 36.5 million reads. An average of 97.4% of reads from each sample was uniquely mapped to the human genome. Overall, 58056 genes/19960 protein-coding genes were identified in the patients with sepsis and healthy controls. Detailed information about the qualitative analysis of each sample is recorded in Supplementary Table 2.

### Detection of DEGs in sepsis patients

A PCA of global gene expression profiles of PBMCs revealed that sepsis patients were clearly separate from healthy individuals (Figure 1A). Additionally, unsupervised hierarchical cluster

analysis based on the normalized expression of protein-coding genes except the low-abundance genes showed that all of the patients with sepsis clustered together (Figure 1B). To determine whether there were distinct patterns of gene expression in sepsis patients, transcriptomic profiles of sepsis patients were compared to those of healthy controls. Differential gene expression analysis showed that sepsis patients had obvious changes in their PBMCs compared to those of healthy subjects (Figure 1C). The expression of 1639 genes was significantly different (using a log<sub>2</sub> fold change (log<sub>2</sub> fold change) >2 and a Benjamini-Hochberg adjusted P value <0.05) in sepsis patients. Of these genes, 1278 (78.0%) were up-regulated and 361 (22.0%) were downregulated (Figure 1D, Supplementary Table 3 and Supplementary Table 4). The expression pattern of DEGs was shown by using unsupervised hierarchical cluster analysis (Supplementary Figure 2).

Significant differences were observed in the expression of the pro-inflammatory cytokine genes IL-2, IL-6, TNF, and IFN- $\gamma$  and the anti-inflammatory cytokine genes IL-4 and IL-10 between sepsis patients and healthy controls (Figure 1E).

### Annotation enrichment analysis

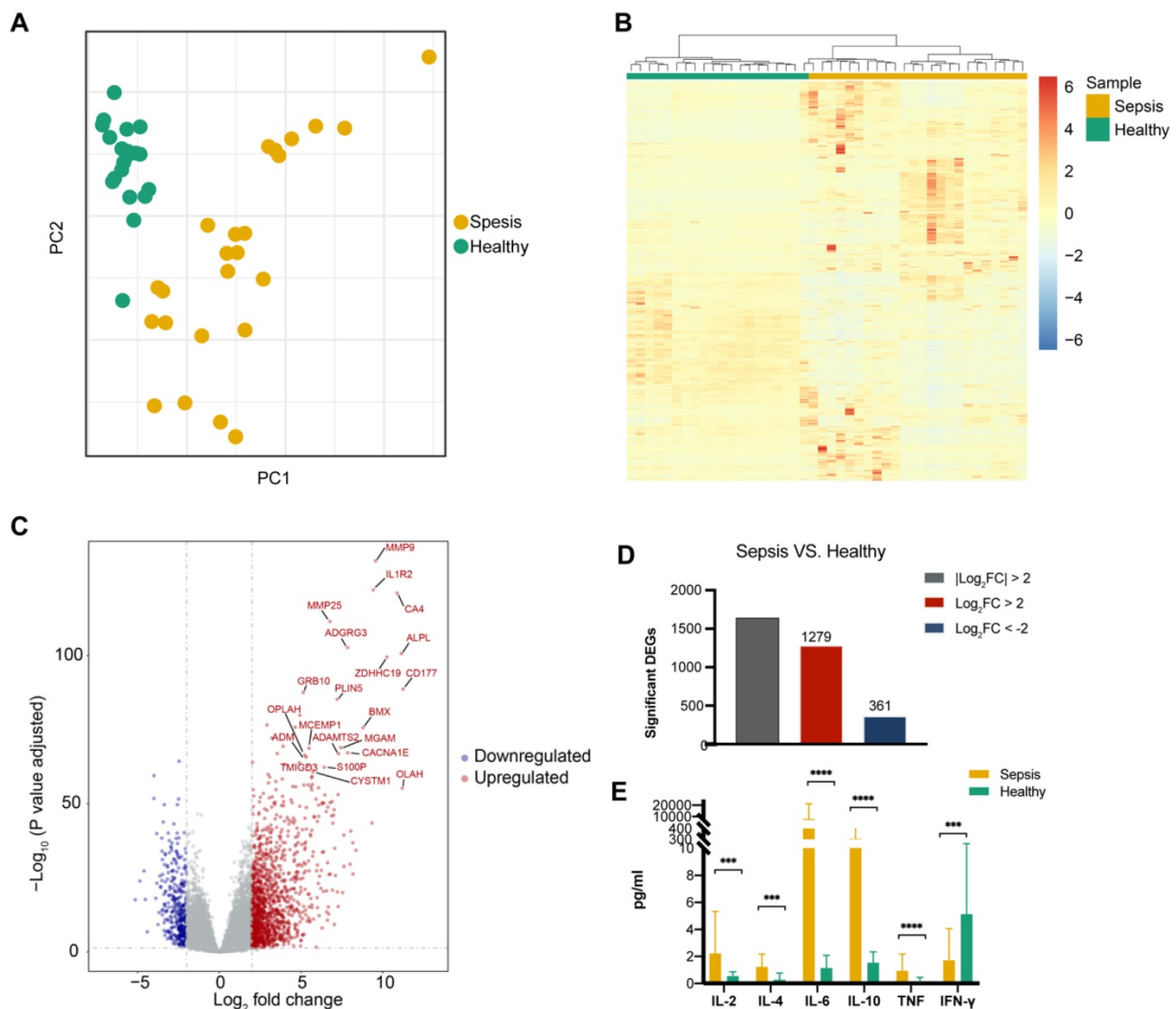
To elucidate the pathogenesis of sepsis and study the possible functions of these observed changes in gene expression, the biological processes that these DEGs participate in were identified by annotation enrichment analysis using ClueGO. In total, 214 GO terms were identified. The over-expressed genes showed functional significant differences (Benjamini-Hochberg adjusted probability <0.05) related to 194 GO terms, as shown in Figure 2A and detailed in Supplementary Table 5. These GO terms are related to the immune response; cytokine secretion, such as "interleukin-6 production" and "regulation of tumor necrosis factor production"; and the inflammatory response. In addition, biological processes such as the defense response, cell chemotaxis and migration, angiogenesis, hemopoiesis, and metalloendopeptidase activity were also identified among these upregulated genes in sepsis patients. Immune response pathways (Figure 2B, Supplementary Table 6) such as lymphocyte/(alpha-beta) T cell activation, leukocyte/lymphocyte/natural killer cell-mediated immunity, and the cellular defense response were enriched among downregulated genes. The top eight pathways enriched according to ClueGO are depicted in Figure 2C. These results indicated that excessive inflammation and immune suppression are concurrent in the host response, which is the principal pathophysiology in sepsis. Additionally, to identify



whether there were some specific pathways related to changes in the sepsis transcriptome profile, an annotation enrichment analysis was also performed based on the canonical pathways described in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. The significantly enriched pathways of upregulated DEGs in septic patients compared to those in healthy subjects included “complement and coagulation cascades” (Figure 2D). “T cell receptor signaling pathway” and “natural killer cell-mediated cytotoxicity” were enriched among downregulated DEGs (Figure 2E). Moreover, the TNF signaling pathway, Toll-like receptor signaling pathway and Jak-STAT signaling pathway were also altered in sepsis patient PBMCs.

### Identification of representative and specific genes in sepsis patients

To identify the key molecules that were differentially expressed and associated with sepsis pathophysiology, the differential gene expression profiles between sepsis patients and healthy subjects and their associated functions were used as the foundation for derivation of a candidate sepsis signature and biomarker. The frequencies of all identified DEGs participating in each GO term cluster were calculated. An importance score was defined based on the gene expression, significance of the difference and function participation frequency. Four unique representative genes (matrix metalloendopeptidase 9, MMP9; S100 calcium-binding protein A8,



**Figure 1.** Transcriptional landscape in peripheral blood mononuclear cells (PBMCs) from patients with sepsis and healthy controls. **(A)** Principal component analysis (PCA) of the RNA transcriptome of PBMCs from sepsis patients and healthy subjects. **(B)** Unsupervised hierarchical cluster analysis of sepsis patients and healthy subjects based on the normalized expression of protein-coding genes except the low-abundance genes, the sum and variance of which are lower than 10. Each column represents one sample, and each row represents a gene. The expression level of each gene in a single sample is depicted according to the color scale. **(C)** Volcano plot of differentially expressed genes (DEGs) from the comparison between sepsis patients and healthy subjects. Red coloring shows upregulated genes, and blue coloring shows downregulated genes ( $|\log_2$  fold change  $> 2$ , adjusted P value  $< 0.05$ ). **(D)** Number of up- and downregulated genes, which are colored in red and blue, respectively. **(E)** Enzyme-linked immunosorbent assay (ELISA) measurement of serum IL-2, IL-4, IL-6, IL-10, TNF and IFN- $\gamma$  levels. \*\*\* P value  $< 0.001$ , \*\*\*\* P value  $< 0.0001$ ; Mann-Whitney U test.

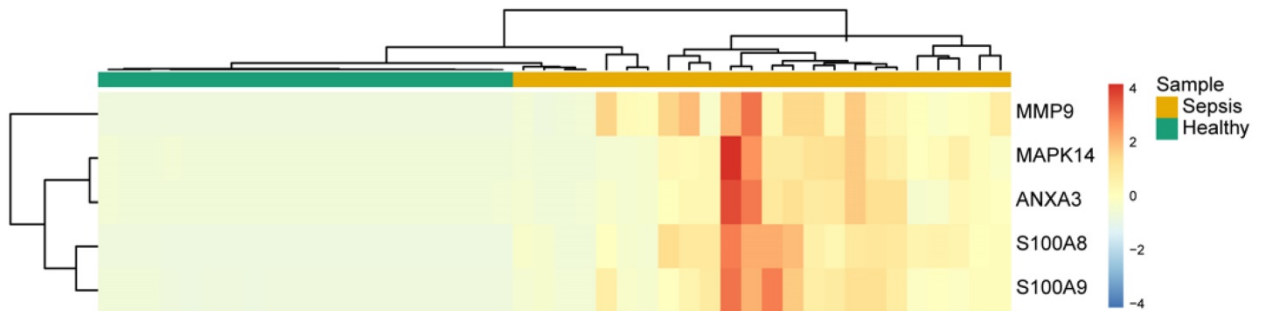


**A**

Gene Symbol	Description	Ensemble ID	Log <sub>2</sub> FoldChange	P-value adjusted <sup>a</sup>
S100A8	S100 calcium binding protein A8	ENSG00000143546	5.16	5.88E-15
S100A9	S100 calcium binding protein A9	ENSG00000163220	4.48	2.95E-31
MMP9	matrix metalloproteinase 9	ENSG00000100985	9.56	1.13E-132
ANXA3	annexin A3	ENSG00000138772	5.75	1.96E-34

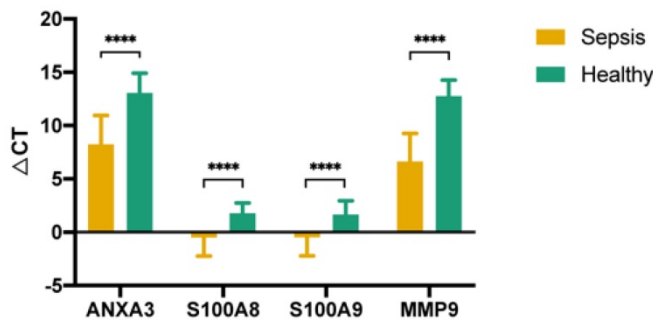
<sup>a</sup> Adjusted P-value by the Benjamini and Hochberg procedure

**B**

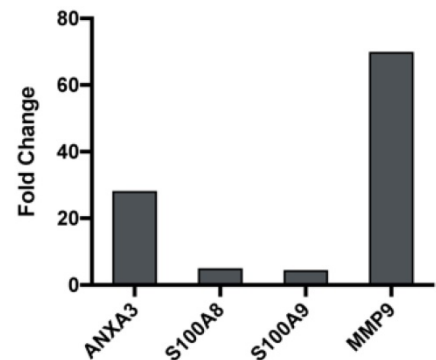


**Figure 3.** Four pathophysiology-based biomarkers of sepsis. **(A)** Top 4 differentially expressed genes (DEGs) between sepsis patients and healthy subjects, as potential disease biomarkers. The P value was adjusted by the Benjamini-Hochberg procedure. **(B)** Unsupervised hierarchical cluster analysis of sepsis patients and healthy subjects based on the normalized expression of the four biomarkers. Each column represents one sample, and each row represents a gene. The expression level of each gene in a single sample is depicted according to the color scale.

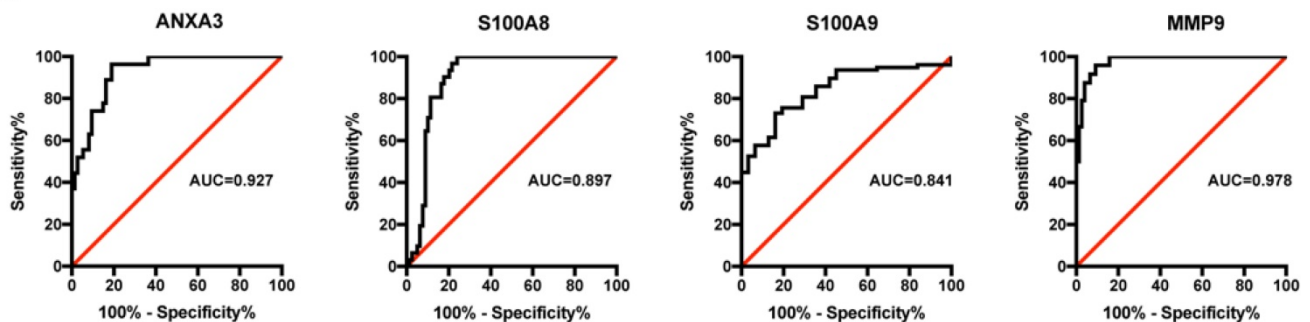
**A**



**B**



**C**



**Figure 4.** Validation of four potential biomarkers of sepsis. **(A)** The  $\Delta$ CT results of the top 4 differentially expressed genes (DEGs) identified using quantitative real-time polymerase chain reaction (qRT-PCR). \*\*\*\* P value <0.0001; Mann-Whitney U test. **(B)** Comparison of the fold changes in the top 4 DEGs identified using qRT-PCR between sepsis patients and healthy control subjects. **(C)** ROC curves for the top 4 DEGs: ANXA3, S100A8, S100A9 and MMP9.

**Table 2.** Characteristics of enrolled patients and healthy subjects included in the validation group

	Sepsis	Healthy	P value
<b>n</b>	63	20	
Male (%)	29 (46.0%)	12 (60.0%)	0.405
Age (years)	65.5 [55.2, 76.5]	63.5 [57.0, 66.0]	0.142
SOFA	6.0 [4.0, 8.0]		
APACHE II	15.0 [10.7, 22.0]		
<b>Infection</b>			
Gram-positive bacteria (%)	9 (14.3%)		
Gram-negative bacteria (%)	25 (39.7%)		
Viral (%)	0		
Other (%)	29 (46.0%)		
CRRT	6 (9.5%)		
Vasopressors	23 (36.5%)		
Mechanical ventilation	22 (34.9%)		
<b>Mortality</b>			
28-day	6 (9.5%)		
90-day	7 (11.1%)		
<b>Laboratory Data</b>			
MAP (mmHg)	81.2 ± 13.5		
White blood cell count	10.40 [6.30, 17.50]	5.40 [4.45, 6.30]	<0.001
Hemoglobin (g/L)	109.2 ± 30.9	149.9 ± 13.3	<0.001
Hematocrit (%)	32.9 ± 8.6	44.4 ± 3.7	<0.001
Platelet count	134.0 [79.5, 187.0]	228.0 [206.0, 246.0]	<0.001
Albumin (g/dL)	28.0 [25.7, 31.6]	45.7 [45.1, 46.8]	<0.001
Aspartate aminotransferase	47.0 [26.0, 111.5]	20.0 [15.0, 27.5]	<0.001
Alanine aminotransferase	30.0 [16.5, 61.0]	20.0 [18.0, 22.2]	0.052
Total bilirubin (μmol/L)	15.4 [8.7, 25.0]	14.5 [12.0, 18.9]	0.966
Creatinine (μmol/L)	111.0 [76.5, 214.5]	68.5 [57.2, 79.5]	<0.001
INR	1.2 [1.1, 1.3]		

APACHE II = Acute Physiology and Chronic Health Evaluation II;

INR = International Normalized Ratio; MAP = mean arterial pressure;

SOFA = Sequential Organ Failure Assessment on day of sampling.

Data are expressed as the mean ± standard deviation (SD), median (interquartile range) or number of patients (percentages).

The continuous variables were compared by using Student's t-test and the Mann-Whitney U test, and the categorical variables were compared by using the  $\chi^2$  or Fisher's exact test between the discovery and validation groups.

## Discussion

In this study, mRNA sequencing was implemented to identify four potential diagnostic biomarkers for sepsis, which were involved in anti-coagulation, endothelial barrier dysfunction and inflammatory response. The diagnostic performance of these markers has been validated on independent patients through qRT-PCR-based ROC analysis. Conversion of the classifier from a sequencing format to qRT-PCR format is convenient for clinical utility. These candidates' blood biomarkers assist in the rapid diagnosis of septic patients at ICU admission.

To determine the basic mRNA expression status, this study identified entire mRNAs that changed significantly in the PBMCs of patients with sepsis compared with those in the PBMCs of healthy subjects. According to quality control and qRT-PCR validation, our sequencing results had high reliability and quality. Through bioinformatics analysis, potential functions of significantly DEGs were predicted by GO term enrichment and KEGG pathway enrichment. Identifying important biological function changes could help us to more clearly

understand the molecular mechanisms of sepsis.

Based on differential expression analysis, 1639 DEGs were found between the sepsis and healthy groups. GO enrichment analysis was performed and revealed that the biological process terms that were significantly enriched included the immune response, cytokine secretion-related processes, the inflammatory response, and neutrophil/leukocyte-related processes, which are the main pathophysiological conditions of sepsis [19-21]. According to the KEGG enrichment results, the Toll-like receptor signaling pathway and TNF signaling pathway were identified in the upregulated gene set. The Jak-STAT signaling pathway, T cell receptor signaling pathway, and natural killer cell-mediated pathway appeared in the downregulated gene set. The potential impact of these pathways in sepsis has already been stated in published studies [22-26]. The results showed that genes participating in immune and inflammatory responses have mixed differential expression patterns of both up- and downregulation. These gene expression and functional dysregulation results are consistent with previous studies showing that the host response in sepsis contains both sustained excessive inflammation and immune suppression before returning to normal homeostasis. Apoptosis is another manifest pathological process of sepsis. However, in this study, dysregulation of the genes related to apoptosis was not detected.

The identification of potential biomarkers for sepsis not only improves clinical practice but also provides molecular insights into the pathophysiological basis of this disease. We used a clinical manifestation-directed analysis of the functional synergy among DEGs and discovered four potential biomarkers, which were all validated with external samples. MMP9 has been reported to be involved in extracellular matrix degradation and leukocyte migration, which are essential components of an effective host response to *Escherichia coli* peritonitis [27]. In addition, MMP9 has been shown to regulate platelet-dependent infiltration of neutrophils and tissue damage in septic lung injury by controlling CD40L shedding from platelets, which indicates that targeting MMP9 may be a useful strategy to limit acute lung injury in abdominal sepsis [28]. S100A8 and S100A9 have been reported to alter MyD88-dependent proinflammatory gene programs, which prevent hyperinflammatory responses without impairing pathogen defense [29]. In addition, S100A8 and S100A9 have been demonstrated to mediate endotoxin-induced cardiomyocyte dysfunction via the receptor for advanced glycation end products [30], which may change the condition of cardiomyocyte dysfunction as a result of sepsis, which the leading



cause of death in the critically ill. ANXA3 has been reported to play a role in anticoagulation [31]. The augmented tendency for thrombosis during sepsis is caused mainly by compromised activity of the three main anticoagulant pathways [21]. The level of ANXA3 expression was significantly increased in patients with sepsis, indicating its function in limiting the formation and expansion of thrombosis.

In summary, we performed a comprehensive transcriptome profile analysis of PBMCs to identify the manifest pathological processes of sepsis. Further qRT-PCR validation indicated that four genes are differentially expressed between patients with sepsis and healthy subjects (MMP9, S100A8, S100A9, and ANXA3). These genes have the potential to be treatment targets or biomarkers for sepsis diagnosis and could provide a new direction toward understanding pathological processes. To permit the clinical application of these biomarkers, a larger cohort of sepsis patients' needs to be further validated using high-throughput sequencing. The relationship between infecting microorganisms and the transcriptome profile also needs to be clarified in a further study.

## Abbreviations

PBMCs: peripheral blood mononuclear cells; PCA: Principal component analysis; DEGs: differentially expressed genes; GO: Gene Ontology; qRT-PCR: quantitative real-time polymerase chain reaction; ICU: intensive care unit; mRNA: messenger ribonucleic acid; cDNA: complementary DNA; GAPDH: glyceraldehyde phosphate dehydrogenase; CBA: cytometric bead array; KEGG: Kyoto Encyclopedia of Genes and Genomes; MMP9: matrix metalloendopeptidase 9; S100A8: S100 calcium-binding protein A8; S100A9: S100 calcium-binding protein A9; ANXA3: annexin A3.

## Supplementary Material

Supplementary figures and tables.

<http://www.medsci.org/v17p2077s1.pdf>

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (81830073), The Science and Technology Project of Taizhou (1801KY70, 1902KY02).

## Author Contributions

The study conception and design, the manuscript revision were performed by Xin Chen, Jun Li, Yinghe Xu. The data analysis and manuscript draft were performed by Tianzhou Wu and Xi Liang. The experiments were performed by Qi Chen, Jiaojiao Xin,

Jing Jiang and Dongyan Shi. The study implementation and manuscript revision were performed by Yongpo Jiang, Huaping Zhang, Sheng Zhang, Yuhang Lv. All authors read and approved the final manuscript.

## Ethics approval

Ethics approval was granted by the Clinical Research Ethics Committee of Taizhou Central Hospital (Taizhou University Medical School) (Registration number: 2019-016, Principal investigator: Yinghe Xu, Date of registration: February 26, 2019).

## Clinical trial

This study was registered in the Chinese Clinical Trial Registry (ChiCTR1900022081, Principal investigator: Yinghe Xu, Date of registration: March 21, 2019).

## Competing Interests

The authors have declared that no competing interest exists.

## References

- Dombrovskiy VY, Martin AA, Sunderram J, Paz HL. Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: a trend analysis from 1993 to 2003. *Crit Care Med.* 2007; 35: 1244-50.
- Adhikari NKJ, Fowler RA, Bhagwanjee S, Rubenfeld GD. Critical Care 1 Critical care and the global burden of critical illness in adults. *Lancet.* 2010; 376: 1339-46.
- Ferrer R, Martin-Loeches I, Phillips G, Osborn TM, Townsend S, Dellinger RP, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Crit Care Med.* 2014; 42: 1749-55.
- Zhuang ZG, Zhang JA, Luo HL, Liu GB, Lu YB, Ge NH, et al. The circular RNA of peripheral blood mononuclear cells: Hsa\_circ\_0005836 as a new diagnostic biomarker and therapeutic target of active pulmonary tuberculosis. *Mol Immunol.* 2017; 90: 264-72.
- Goswami R, Blazquez AB, Kosoy R, Rahman A, Nowak-Wegrzyn A, Berin MC. Systemic innate immune activation in food protein-induced enterocolitis syndrome. *J Allergy Clin Immunol.* 2017; 139: 1885-96 e9.
- McHugh L, Seldon TA, Brandon RA, Kirk JT, Rapisarda A, Sutherland AJ, et al. A Molecular Host Response Assay to Discriminate Between Sepsis and Infection-Negative Systemic Inflammation in Critically Ill Patients: Discovery and Validation in Independent Cohorts. *PLoS Med.* 2015; 12: e1001916.
- Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, et al. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. *Lancet Respir Med.* 2016; 4: 259-71.
- Hu Y, Zhong W, Chen M, Zhang Q. Identifying crucial genes for prognosis in septic patients: Gene integration study based on PRISMA guidelines. *Medicine (Baltimore).* 2019; 98: e16807.
- Yang J, Zhang S, Zhang J, Dong J, Wu J, Zhang L, et al. Identification of key genes and pathways using bioinformatics analysis in septic shock children. *Infect Drug Resist.* 2018; 11: 1163-74.
- Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016; 315: 801-10.
- Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics.* 2017; 33: 3137-9.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30: 2114-20.
- Kim D, Landmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015; 12: 357-U121.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015; 31: 166-9.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550.
- Bindea G, Mecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009; 25: 1091-3.

17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012; 28: 882-3.
18. Lin L, Finak G, Ushey K, Seshadri C, Hawn TR, Frahm N, et al. COMPASS identifies T-cell subsets correlated with clinical outcomes. *Nat Biotechnol*. 2015; 33: 610-6.
19. Delano MJ, Ward PA. Sepsis-induced immune dysfunction: can immune therapies reduce mortality? *J Clin Invest*. 2016; 126: 23-31.
20. Delano MJ, Ward PA. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol Rev*. 2016; 274: 330-53.
21. van der Poll T, van de Veerdonk FL, Scicluna BP, Netea MG. The immunopathology of sepsis and potential therapeutic targets. *Nat Rev Immunol*. 2017; 17: 407-20.
22. Forel JM, Chiche L, Thomas G, Mancini J, Farnarier C, Cognet C, et al. Phenotype and functions of natural killer cells in critically-ill septic patients. *PLoS One*. 2012; 7: e50446.
23. Lv X, Zhang Y, Cui Y, Ren Y, Li R, Rong Q. Inhibition of microRNA155 relieves sepsis-induced liver injury through inactivating the JAK/STAT pathway. *Mol Med Rep*. 2015; 12: 6013-8.
24. Kuzmich NN, Sivak KV, Chubarev VN, Porozov YB, Savateeva-Lyubimova TN, Peri F. TLR4 Signaling Pathway Modulators as Potential Therapeutics in Inflammation and Sepsis. *Vaccines (Basel)*. 2017; 5.
25. Rosadini CV, Kagan JC. Early innate immune responses to bacterial LPS. *Curr Opin Immunol*. 2017; 44: 14-9.
26. Winkler MS, Rissiek A, Priefler M, Schwedhelm E, Robbe L, Bauer A, et al. Human leucocyte antigen (HLA-DR) gene expression is reduced in sepsis and correlates with impaired TNFalpha response: A diagnostic tool for immunosuppression? *PLoS One*. 2017; 12: e0182427.
27. Renckens R, Roelofs JJ, Florquin S, de Vos AF, Lijnen HR, van't Veer C, et al. Matrix metalloproteinase-9 deficiency impairs host defense against abdominal sepsis. *J Immunol*. 2006; 176: 3735-41.
28. Rahman M, Zhang S, Chew M, Syk I, Jeppsson B, Thorlacius H. Platelet shedding of CD40L is regulated by matrix metalloproteinase-9 in abdominal sepsis. *J Thromb Haemost*. 2013; 11: 1385-98.
29. Ulas T, Pirr S, Fehlhaber B, Bickes MS, Loof TG, Vogl T, et al. S100-alarmin-induced innate immune programming protects newborn infants from sepsis. *Nat Immunol*. 2017; 18: 622-32.
30. Boyd JH, Kan B, Roberts H, Wang Y, Walley KR. S100A8 and S100A9 mediate endotoxin-induced cardiomyocyte dysfunction via the receptor for advanced glycation end products. *Circ Res*. 2008; 102: 1239-46.
31. Kubisch R, Meissner L, Krebs S, Blum H, Gunther M, Roidl A, et al. A Comprehensive Gene Expression Analysis of Resistance Formation upon Metronomic Cyclophosphamide Therapy. *Transl Oncol*. 2013; 6: 1-9.