

Screening of key genes related to the prognosis of mouse sepsis

Muhu Chen ^{1#}, Xue Chen ^{1#}, Yingchun Hu ^{1*}, Xianfu Cai ^{1*}

¹Department of emergency medicine, Affiliated Hospital of Southwest Medical University, Luzhou, China.

[#]They are co-first author and contributed equally to this study.

*Corresponding author

Yingchun Hu

No. 25, Taiping Street, Jiangyang District, Luzhou City, 646000, China

Tel: 86- 08303165120

Email: huyingchun913@swmu.edu.cn

Xianfu Cai

Email: caixianfu@126.com

Abstract

Sepsis is a common clinical disease with high mortality, and patients with sepsis have varied prognoses. Researchers need to explore the underlying mechanisms that determine the prognosis of sepsis. Hence, a mouse model was used to evaluate new potential prognostic markers of sepsis. Mice were randomly divided into low-dose group (n = 3, lipopolysaccharides [LPS], 20mg/kg) and high-dose group (n = 3; LPS, 40mg/kg). Total RNA was extracted from the peripheral blood of mice, and samples were then subjected to RNA sequencing. When complete data were normalized, the high-dose group and low-dose group

were screened for differentially expressed genes (DEGs, $\log_2FC \geq 1$ and $q \text{ value} \leq 0.05$). DEGs were analyzed by gene ontology enrichment, and potential core genes were screened using protein–protein interaction network (PPI) and weighted gene co-expression network analysis (WGCNA). Moreover, the survival data in GSE65682 was used to observe the correlation between core genes and prognosis. A total of 967 DEGs were identified in the low-dose group, of which 390 were upregulated and 577 were downregulated. These genes were mainly enriched in white blood cell activation, lymphocyte activation, immune system response, etc. LCK, ZAP70, ITK, CD247, and DOCK2 were found at the core of PPI network, while WGCNA found that IFI35, ITGB3, and MED25 may be potential core genes. It was demonstrated that CD247, DOCK2, IFI35, ITK, and LCK core genes were positively correlated with prognosis based on GSE65682. CD247, DOCK2, IFI35, ITK, LCK, and MED25 might be important targets affecting the prognosis of sepsis.

Keywords: Sepsis; Prognosis; PPI; co-expression; RNA-seq

Introduction

Sepsis is a common critical clinical illness with high mortality, and data shows that its incidence is about 19 million patients per year in the United States [1]. The latest guidelines defined sepsis as an infection that causes dysfunctions of more than two organs [2, 3]. The current clinical treatment for sepsis consists of anti-infection, fluid replacement, and organ function maintenance. With regard to its pathogenesis, current researchers have discovered that sepsis is related to inflammatory storms [4, 5]. Patients with sepsis may survive or die even when receiving the same treatment. However, the key factors and mechanisms that

determine its prognosis remains unclear. This study intends to simulate the different prognoses of sepsis through intraperitoneal injection of different concentrations of lipopolysaccharides (LPS) in mice and sequence its peripheral blood cells to explore potential core factors in order to screen out the key targets that determine its prognosis.

Materials and Methods

The experimental mice were obtained from the Experimental Animal Center of Southwest Medical University. Male Kunming mice aged 8–12 weeks weighing 25–30 g were selected in this study. The mice were housed in the animal laboratory of the Central Laboratory of Southwest Medical University, with a 12-hour day-night cycle and an ambient temperature of $24^{\circ}\text{C} \pm 2^{\circ}\text{C}$. Moreover, mice were raised in an ordinary rearing environment with free access to water and food. The experiment on animals began after five days of acclimatization to the environment. All animal experiments were approved by the Animal Ethics Association of Southwest Medical University. LPS was purchased from the American Sigma Inc (Fungus: O55: B5).

Grouping and modeling

The pre-experimental mice were randomly divided into 4 groups, each with 10 mice, which summed up to a total of 40 mice. The mice were fasted for 10 hours prior to the experiment and resumed eating 1 hour after modeling. The mice from each group were administered with intraperitoneal injection of LPS and equal dose of PBS solution with drug concentrations of 10 mg/kg, 20 mg/kg, 40 mg/kg, and 50 mg/kg. Thereafter, their body temperature and activity status were measured 0h, 2h, 4h, 8h, 12h, 24h, 2d, 3d, 4d, 5d, and 7d after injection, and the

survival curve was plotted to determine the best modeling conditions. In the observation study, some mice died within 7 days due to the LPS intraperitoneal injection; those surviving over 7 days were sacrificed by cervical spondylectomy. In the formal experiment, the mice were randomly divided into two groups: the survival group (n = 3) and the death group (n = 3). The survival group (low-dose group) received intraperitoneal injection of an equal volume of PBS solution and 20 mg/kg LPS, while the death group (high-dose group) received intraperitoneal injection of 40mg/kg LPS. The body temperature and activity status of each group of mice were measured during the above time points. When the body temperature of the mouse in the death group reached 23°C (dying state), the blood from the eyeball was taken. At the same time point, one mouse from the survival group was randomly selected to have its eyeball blood taken, and the sample was treated with EDTA-2K for anticoagulation. During collection of blood samples, the survival group mice died of massive hemorrhage due to eyeball enucleation.

Sample sequencing

In accordance with the instructions of Invitrogen, the TRizol method was used to extract the total RNA from whole blood, and DNase I (TaKara) was used to eliminate the genomic DNA. The RNA quality was then measured by 2100 Bioanalyzer (Agilent) and then quantified using ND-2000 (NanoDrop Technologies). Only high-quality RNA samples (OD260/280, 1.8–2.2; OD260/230 \geq 2.0; RIN, \geq 6.5; 28S, 18S \geq 1.0, >2 μ g) were used to construct sequencing library. 1 μ g of the qualified test sample was taken. The test samples were prepared, the library was built, and the RNA-seq was sequenced according to the Illumina TruSeq™ RNA sample manual.

Differential gene screening

All clean sequencing data were compared using the Genbank mouse genome database, and subsequently, gene annotation was performed. To ensure the accuracy of gene expression levels, the fragments per kilobase of exon per million (FPKM) was employed to normalize gene expression levels. By performing log-uniform analysis on differentially expressed genes (DEGs), the differential genes in the survival and death groups caused by LPS were screened (screening conditions: $\log_2FC \geq 1$ and $q \text{ value} \leq 0.05$). Moreover, data visualization was carried out through integrated Differential Expression and Pathway analysis (iDEP) (<http://bioinformatics.sdstate.edu/idep/>) network tool based on R language [6].

GO analysis

GO analysis is a typical analysis method for big gene data. It can perform global analysis from three levels, BP, MF, or CC, which is beneficial to the classification and screening of batch genes in the later stage. In this study, GO enrichment was performed on upregulated and downregulated genes according to different gene expression trends.

PPI analysis

The protein-protein interaction (PPI) network is a common method used in screening potential core genes, and the STRING (<https://string-db.org/cgi/input.pl?>) online network tool is a popular tool for PPI analysis [7]. The basic principle is to connect proteins that directly and physically interact with each other. In theory, the more core factor functions, the more connections and the closer to the core area. Considering the reliability of the data, this study screened all experimentally verified results. In view of the integrity of the network, the correlation strength coefficient was set to 0.3. Subsequently, through the online analysis tool,

the functions of interest in this study were screened out, and core genes functions were reviewed.

Co-expression analysis

In order to prevent the PPI method from being missed, the WGCNA co-expression method was adopted for core gene screening [8]. The purpose of co-expression analysis is to screen out possible key factors based on the common trend of expression values [9]. WGCNA co-expression analysis is divided into different modules according to different expression situations. This method can reasonably speculate the possible functions of genes that are less studied. For core genes screened by PPI and WGCNA, heat maps were used to establish their expression in different groups.

Survival curve of core genes

Clinical data is the foundation of scientific research. In order to better screen genes related to prognosis and verify the relationship between core genes and prognosis, human sepsis data GSE65682 [10] was used. GSE65682 was a big data study on genes in the peripheral blood of patients with sepsis to identify the biomarkers of sepsis and other diseases. It recorded the peripheral blood cell gene data and clinical data of 479 patients with sepsis as well as the 28-day survival data. The specific genes of all 479 patients were sequenced from the low expression group to the high expression group. The first half ($n = 239$) was the low expression group, while the last half ($n = 240$) was the high expression group. As regards GSE65682 data, GraphPad prism 7.0 was adopted to plot the survival curve of the above core genes. Log-rank (Mantel-Cox) test was used to identify statistical differences between the two groups of data.

Results

Establishment of sepsis survival and death model

After an intraperitoneal injection of different concentrations of LPS, the mortality of mice has shown to be dose-dependent. In general, the higher the concentration, the higher the mortality rate (Figure 1A). Furthermore, the overall survival rate of mice in the 20 mg/kg group was 80%, while the mice in the 50 mg/kg group almost died in 48 hours. In order to better simulate the clinical situation of sepsis, we chose 20 mg/kg as the survival group and 40 mg/kg as the death group for follow-up research. All survival mice in the low-dose group showed hypomotility, appetite loss, yellowing hair, and other inflammatory manifestations similar to sepsis.

Identification of differential genes

After sequencing and gene annotation of peripheral blood cells of mice, the gene expression data was logarithmically processed (Figure 1B). As a result, the data was uniform, and the two groups were comparable. In comparison to the high-dose group, a total of 967 DEGs were found in the low-dose group, of which 390 gene expressions were upregulated, 577 gene expressions were reduced, and differentially expressed gene distributions were even (Figure 1C).

GO enrichment analysis

On the one hand, GO analysis revealed that the highly expressed genes in the survival group were mainly enriched in leukocytes, T cells, lymphocyte activation and differentiation, immune system processes, positive regulation of leukocyte adhesion, and peptide metabolic

processes. On the other hand, genes highly expressed in the death group were mainly enriched in immune response, endogenous immune response, defense response, and response to external stimuli. The enrichment of these functions basically reflected the changes of biological processes in different states, as illustrated in Figure 1D.

PPI screening of core factors

In this study, the PPI network analyzed by the STRING online tool included 67 core proteins, which were closely linked together. The proteins of LCK, ZAP70, JUN, ITK, CD247, DOCK2, and HCK were located at the core of the network, which might be involved in many different functions. In this study, proteins involved in the immune system, cell surface receptor signaling pathways, cell proliferation regulation, cell differentiation, and cell death regulation were paid more attention, as illustrated in Figure 2.

Co-expression screening of core genes

Co-expression analysis is a common method of bioinformatics, which supplements the PPI analysis deficiency. We employed the WGCNA analysis method to construct a co-expression module for genes. A total of five modules were screened out in this study, and two of them were mainly distributed. We visualized the first 25 genes of these two modules and discovered that IFI35, ITGB3, ZHX2, USP25, FECH, KDM6B, and MED25 might be potential core genes, as shown in Figures 3A–3C. In Figure 4D, we further observed the expression of 14 potential core genes screened by two of the abovementioned methods of PPI and WGCNA. CD247, ZHX2, KDM6B, DOCK2, ZAP70, LCK, and ITK were highly expressed in the survival group, while IFI35, ITGB3, JUN, MED25, USP25, and FECH were highly expressed in the death group.

Survival curve of core genes

To further explore the correlation between genes, we screened the prognosis of clinical patients, and the genetic data of sepsis peripheral blood cells in GSE65682 was used for survival analysis. Finally, we found that the expression values of CD247, DOCK2, IFI35, and LCK manifested a positive correlation with the prognosis of patients ($P < 0.05$), and the difference in the statistics of ITK survival curve was $P = 0.06$. The expression value of MED25 demonstrated a negative correlation with the prognosis of patients ($P < 0.05$), as shown in Figure 4.

Discussion

Sepsis is a common critical clinical condition, especially in patients in the ICU. Sepsis involves multidisciplinary and multisystem changes [2]. Clinically, some patients with sepsis died despite strong anti-infection treatment, especially when sepsis is combined with septic shock. In clinical practice, the key factors that determine the prognosis of patients and their underlying mechanisms deserve our in-depth study [11]. In this study, different concentrations of LPS were used to construct the survival and death model of sepsis. Moreover, the mortality was used as the main standard to screen and simulate different prognoses of sepsis. Furthermore, the cecal ligation and perforation (CLP) method is another common model construction method [12]. However, achieving the expected results of CLP during the construction process is challenging. This may be related to the large differences in individual defense responses in the abdominal cavity of mice, especially the omentum.

In this study, peripheral blood cells were sequenced and screened under bioinformatics

technology. A total of 967 DEGs were screened out. Among them, 390 genes were upregulated and 577 genes were downregulated in the survival group. Through GO analysis, it was found that gene function enrichment in the survival group was mainly related to the activation and differentiation of white blood cells, lymphocytes, and T cells, which suggested that the body is establishing a second line of defense, or may participate in the establishment of endotoxin tolerance mechanisms. The gene function enrichment in the death group was mainly related to immune response and endogenous defense response, which indicated that the body produces a large number of inflammatory mediators and induces an inflammatory cascade amplification effect. This was consistent with our “inflammatory storm” theory.

In the method of screening core genes, we used PPI and WGCNA to screen the genes together [7-9], as far as possible, to prevent possible missed selection of a single method. From a methodological point of view, the PPI method is more likely to be verified later due to the multiple functions of these core proteins confirmed by a large amount of data in the early stage. Additionally, co-expression will likely select core genes that are less studied, so it is with higher innovation. However, all core genes theoretically screened need to be verified by experiment.

To further increase the positive rate of bioinformatics screening, we included the clinical prognostic data of patients. Verification in the human body will make the conclusion more convincing. This study screened six potential core genes: CD247, DOCK2, IFI35, LCK, MED25, and ITK. Among them, CD247, DOCK2, IFI35, LCK, and ITK showed positive correlation with prognosis, suggesting that they might be beneficial to patient survival. The MED25 gene was negatively correlated with the prognosis, which indicated that it plays a

role in the promotion of death process of patients.

CD247 is a characteristic protein located in the T-cell membrane. Existing research revealed that it is mainly related to immune diseases [13-16], while there is no research on sepsis. Under this circumstance, it could be a potential drug treatment target. LCK is a non-receptor tyrosine protease, which plays a vital role in promoting the maturation and differentiation of T cells and their functions after maturation. Once T cells are stimulated, TCR activates the tyrosine kinase ZAP70, and ZAP70 is phosphorylated and activated by LCK. Thereafter, a large number of signal molecules are absorbed, which ultimately leads to the production of lymphokines [17]. Subsequently, studies have demonstrated that LCK phosphorylation inhibits inflammatory storm via TLR4 signaling pathway [18]. DOCK2 is an atypical guanine exchange factor specifically expressed in hematopoietic cells. It regulates the activation and migration of immune cells by activating Ras-related C3 botulinum toxin substrate (Rac). Moreover, Dock2 plays a significant role in the occurrence and development of various inflammatory diseases [19-22]. Interferon-inducible protein 35 (IFI35) is an IFN- γ -induced protein that plays an important role in anti-virus-related immune inflammation. Overexpression of IFI35 affects endothelial cell proliferation, migration, and re-endothelialization through the NF-KB pathway [23]. Furthermore, studies have also illustrated that IFI35 exacerbates H5N1 influenza by expressing IL-12p40 homodimer [24]. In addition, ITK and BTK regulate the thermal homeostasis in sepsis in mice through mast cell function. By regulating the activation of NF-KB, phosphatidylinositol-4,5-bisphosphate, 3-kinase/Akt, and p38 signaling pathway, they play a regulatory function in the downstream of mast cell toll-like receptor 4/LPS [25]. On the contrary, enterovirus 71 (EV71) is the main

pathogen of hand-foot-mouth disease, and its neurological complications are particularly serious, which is the main cause of death of the disease. EV71, VP1, and human mediator complex subunit 25 (MED25) have a highly expressed common epitope (co-epitope) in the brain stem, which indicates the involvement of MED25 in the occurrence and development of the disease [26].

This study confirmed the prognostic correlation of the six core genes screened. However, reports on the screened genes and sepsis are limited, which suggests that they may be potential research targets, while their specific functions and mechanisms need to be further studied.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Funding

This work was supported by the Southwest Medical University [grant number 02/00030695] and the Doctor foundation of Affiliated Hospital of Southwest Medical University [grant number 18094].

Author contribution

MHC performed experiments, analyzed experimental results, and wrote the paper; XC performed experiments; YCH designed study and wrote the paper; XFC designed the study and revised the paper. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Reference

1. Prescott HC, Angus DC. (2018) Enhancing recovery from sepsis: a review. *JAMA*. 319(1): 62-75
2. Shankar-Hari M, Phillips GS, Levy ML, Seymour CW, Liu VX, Deutschman CS, et al. (2016) Developing a new definition and assessing new clinical criteria for septic shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 315(8): 775-787
3. Klein Klouwenberg PM, Ong DS, Bonten MJ, Cremer OL. (2012) Classification of sepsis, severe sepsis and septic shock: the impact of minor variations in data capture and definition of SIRS criteria. *Intensive Care Med*. 38(5): 811-819
4. Demaret J, Venet F, Friggeri A, Cazalis MA, Plassais J, Jallades L, et al. (2015) Marked alterations of neutrophil functions during sepsis-induced immunosuppression. *J Leukoc Biol*. 98(6): 1081-1090
5. Severino P, Silva E, Baggio-Zappia GL, Brunialti MK, Nucci LA, Rigato O Jr, et al. (2014) Patterns of gene expression in peripheral blood mononuclear cells and outcomes from patients with sepsis secondary to community acquired pneumonia. *PLoS One*. 9(3): e91886
6. Ge SX, Son EW, Yao R. (2018) iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. 19(1): 534
7. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 45(D1): D362-d368
8. Langfelder P, Horvath S. (2008) WGCNA: an R package for weighted correlation

network analysis. *BMC Bioinformatics*. 9: 559

9. Pei G, Chen L, Zhang W. (2017) WGCNA application to proteomic and metabolomic data analysis. *Methods Enzymol*. 585: 135-158
10. Scicluna BP, Klein Klouwenberg PM, van Vught LA, Wiewel MA, Ong DS, Zwinderman AH, et al. (2015) A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *Am J Respir Crit Care Med*. 192(7): 826-835
11. Fernando SM, Rochwerf B, Seely AJE. (2018) Clinical implications of the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *CMAJ*. 190(36): E1058-e1059
12. Rittirsch D, Huber-Lang MS, Flierl MA, Ward PA. (2009) Immunodesign of experimental sepsis by cecal ligation and puncture. *Nat Protoc*. 4(1): 31-36
13. Lundholm M, Mayans S, Motta V, Löfgren-Burström A, Danska J, Holmberg D. (2010) Variation in the Cd3 zeta (Cd247) gene correlates with altered T cell activation and is associated with autoimmune diabetes. *J Immunol*. 184(10): 5537-5544
14. Martins M, Williams AH, Comeau M, Marion M, Ziegler JT, Freedman BI, et al. (2015) Genetic association of CD247 (CD3 ζ) with SLE in a large-scale multiethnic study. *Genes Immun*. 16(2): 142-150
15. Takeuchi T, Suzuki K. (2013) CD247 variants and single-nucleotide polymorphisms observed in systemic lupus erythematosus patients. *Rheumatology (Oxford)*. 52(9): 1551-1555
16. Teruel M, McKinney C, Balsa A, Pascual-Salcedo D, Rodriguez-Rodriguez L, Ortiz AM, et al. (2013) Association of CD247 polymorphisms with rheumatoid arthritis: a replication

study and a meta-analysis. *PLoS One*. 8(7): e68295

17. Jung SH, Yoo EH, Yu MJ, Song HM, Kang HY, Cho JY, et al. (2016) ARAP, a novel adaptor protein, is required for TCR signaling and integrin-mediated adhesion. *J Immunol*. 197(3): 942-952
18. Mitchell J, Kim SJ, Seelmann A, Veit B, Shepard B, Im E, et al. (2018) Src family kinase tyrosine phosphorylates Toll-like receptor 4 to dissociate MyD88 and Mal/Tirap, suppressing LPS-induced inflammatory responses. *Biochem Pharmacol*. 147: 119-127
19. Alosaimi MF, Shendi H, Beano A, Stafstrom K, El Hawary R, Meshaal S, et al. (2019) T-cell mitochondrial dysfunction and lymphopenia in DOCK2-deficient patients. *J Allergy Clin Immunol*. 144(1): 306-309.e302
20. Chen Y, Meng F, Wang B, He L, Liu Y, Liu Z. (2018) Dock2 in the development of inflammation and cancer. *Eur J Immunol*. 48(6): 915-922
21. Dobbs K, Domínguez Conde C, Zhang SY, Parolini S, Audry M, Chou J, et al. (2015) Inherited DOCK2 deficiency in patients with early-onset invasive infections. *N Engl J Med*. 372(25): 2409-2422
22. Hu N, Pang Y, Zhao H, Si C, Ding H, Chen L, et al. (2019) High expression of DOCK2 indicates good prognosis in acute myeloid leukemia. *J Cancer*. 10(24): 6088-6094
23. Jian D, Wang W, Zhou X, Jia Z, Wang J, Yang M, et al. (2018) Interferon-induced protein 35 inhibits endothelial cell proliferation, migration and re-endothelialization of injured arteries by inhibiting the nuclear factor-kappa B pathway. *Acta Physiol (Oxf)*. 223(3): e13037
24. Gounder AP, Yokoyama CC, Jarjour NN, Bricker TL, Edelson BT, Boon ACM. (2018) Interferon induced protein 35 exacerbates H5N1 influenza disease through the expression of

IL-12p40 homodimer. *PLoS Pathog.* 14(4): e1007001

25. Huang W, Morales JL, Gazivoda VP, August A. (2016) Nonreceptor tyrosine kinases ITK and BTK negatively regulate mast cell proinflammatory responses to lipopolysaccharide. *J Allergy Clin Immunol.* 137(4): 1197-1205
26. Fan P, Li X, Sun S, Su W, An D, Gao F, et al. (2015) Identification of a common epitope between enterovirus 71 and human MED25 proteins which may explain virus-associated neurological disease. *Viruses.* 7(4): 1558-1577

Figure Legends

Figure 1. Establishment of LPS model of sepsis in mice and homogenization of blood RNA sequencing data. A: Establishment of sepsis model in mice. Different concentrations of LPS were used for intraperitoneal injection. Within 7 days, no mice died in the 10 mg/kg group (simulated SIRS); 20% died in the 20 mg/kg group (survival group); 70% died in the 40 mg/kg group (death group); and 100% died in the 50 mg/kg group (super dose group) within 2 days. B: Box chart after the homogenization of sequencing data of the low-dose group (survival group) and the high-dose group (death group). C: Volcanic map results of the differential gene analysis between the two groups (survival group and death group). The vertical coordinate is the negative logarithm of FDR value (the base number is 10); the horizontal coordinate is fold change: the red on the right is the upregulated gene, and the blue on the left is the downregulated gene. D: Functional annotation of upregulated and downregulated differentially expressed genes.

Figure 2. The PPI network of DEGs was constructed using STRING tool; different colors

represent different functions: red is the immune system process; blue is the cell surface receptor signaling pathway; yellow is the regulation of cell population proliferation; green is cell differentiation; and violet is the regulation of cell death.

Figure 3. Construction of co-expression network based on WGCNA and screening of core genes. A: The module distribution of WGCNA co-expression analysis of DEGs. B–C: Module 1/2 shows the first 25 genes of degree construct network with a correlation coefficient of 0.4. D: heat-map of core genes screened using PPI and co-expression analysis; red for upregulated; blue for downregulated.

Figure 4. Survival curve of six potential cores in GSE65682. A–F: A, CD247; B, DOCK2; C, IFI35; D, ITK; E, LCK; F, MED. Red for high expression group, green for low expression group.







