ISSN 2706-6819 Vol.6, Issue 9: 7-13, DOI: 10.25236/IJFM.2024.060902

WGCNA algorithm and Mendelian randomization analysis identified immune checkpoint CD96 as a biomarker for sepsis

Bo Chen[1,2,a], Quan Gong[1,b,*]

[1]*Yangtze University Health Science Center, Yangtze University, Jingzhou, Hubei, 434023, China*
[2]*Dazhou Vocational and Technical College, Dazhou, Sichuan, 635000, China*
[a]*584413767@qq.com,* [b]*gongquan1998@163.com*
[*]*Corresponding author*

Abstract: *Immune checkpoints are a general term for a class of molecules that can regulate the body's immune function, and have a certain correlation with many diseases. However, immune checkpoints in sepsis are not well understood and potential immune checkpoint biomarkers need to be further developed. This study aims to study the effects of immune checkpoints on immune cells in patients with sepsis and identify biomarkers. We drew on the work of other scholars to get immune checkpoint genes from the published literature. Based on the GEO database, we identified differentially expressed immune checkpoint genes (DEICGs) in sepsis by differentially expressed gene (DEG) analysis. Then, CIBERSORT was applied to investigate the Infiltration level of 22 immune cells in the peripheral blood of sepsis, and we analyzed the correlation between DEICGs and the immune cells. After that, the weighted gene co-expression network (WGCNA) algorithm was used to explore new sepsis susceptibility modules, and the hub genes were extracted from the intersection of the results of WGCNA analysis and DEG. Subsequently, we performed Mendelian randomization (MR) analysis to explore the casual association of hub genes on sepsis. Positive results from MR analysis are considered the most reliable sepsis biomarkers and have been validated in external datasets by ROC analysis. A total of 36 DEICGs were identified by DEG. Through CIBERSORT, the correlation between the DEICGs and immune cells was investigated. Integrating the results of DEG and WGCNA, CD96, CD27, CD28, CD160, and TIGIT were identified as hub genes. In inverse variance weighting, we found that CD96 is a protective factor for sepsis with an OR of 0.952 (95%CI = 0.913 - 0.994, p = 0.027). According to ROC analysis, CD96 demonstrated excellent diagnostic efficacy, with an AUC of 0.945 (cutoff value = 4.074, 95%CI = 0.841 - 0.960). Generally, this research identified CD96 as a biomarker that could serve as a novel target for the diagnosis and therapy of sepsis.*

Keywords: *Sepsis, immune checkpoints, WGCNA, Mendelian randomization, biomarker*

1. Introduction

With approximately 14,000 deaths per day, sepsis is one of the main motives of dying in intensive care unit sufferers[1]. Sepsis is a syndrome in which the host responds poorly to infection and progresses rapidly to lead to multi-organ failure and life-threatening patients and accurate diagnosis and effective intervention at the beginning of the disease course are crucial for patient prognosis. Immune disorders, excessive inflammatory responses, and immunosuppression are common in patients with sepsis[2], while the mechanisms of the disease have not been fully elucidated.

Immune checkpoints (ICs) are a group of proteins on the surface of immune cells that can regulate immune cell activity to maintain immune system balance[3]. Some scholars believe that there are at least three ways for immune checkpoints to regulate the body's immune status[4]: (1) The combination of immune checkpoints and their ligands directly act to form synergistic stimulation signals and mediate the regulation of the body's immune function; (2) Immune checkpoints competitively bind to costimulatory signal receptors on immune cells to costimulatory signal ligands and interfere with cell activation; (3) By expressing certain metabolism-related genes to exert immune regulation. Huang X demonstrated that high levels of PD-1 on neutrophils can cause inflammation-regulating cytokine secretion disorders, thereby mediating inflammatory damage and ultimately leading to increased mortality in LPS-treated mice[5]. Studies by Ligers A have shown that the level of CTLA-4 on T cells increases as the disease progresses and that the prognosis is related to nucleotide polymorphisms (SNPs)

ISSN 2706-6819 Vol.6, Issue 9: 7-13, DOI: 10.25236/IJFM.2024.060902

of CTLA-4[6]. Other scholars have demonstrated that compared with the control mice, the $CD4^+$T cells, the $CD8^+$T cells, the B cells, the Treg cells, and the dendritic cells of the sepsis mice expressed higher levels of LAG-3, and anti-LAG-3 treatment can significantly improve the survival rate of sepsis mice[7]. In conclusion, existing research suggests that immune checkpoints may have a crucial function in sepsis, affecting the onset, development, and prognosis of the disease.

This study explored the correlation between immune checkpoint genes and infiltrating immune cells in sepsis, and based on the results of a combination of differentially expressed genes analysis, WGCNA algorithm, and two-sample Mendelian randomization analysis, we identified CD96 as a reliable biomarker of sepsis which can be used for therapy and diagnosis in sepsis.

2. Manuscript Preparation

2.1. Data

2.1.1. GEO data

The mRNA profile datasets named GSE65682 and GSE57065 were downloaded from the GEO database (www.ncbi.nlm.nih.gov/geo). The details of GES65682 and GSE57065 are in table 1.

Table 1: Details of GEO data

Name	Type	Source	Description	Platform
GSE65682	Microarray	GEO	760 cases of sepsis and 42 healthy control	GPL 13667
GSE57065	Microarray	GEO	83 cases of sepsis and 26 healthy control	GPL 570

2.1.2. Immune checkpoint genes

The immune checkpoint genes were derived from published studies[8] and there were 60 immune checkpoint genes involved in our study (table 2).

Table 2: Immune checkpoint genes

NO.	Gene	NO.	Gene	NO.	gene	NO.	gene	NO.	gene
1	ADORA2A	13	CD274	25	HAVCR2	37	KIR2DS3	49	SIRPA
2	BTLA	14	CD276	26	ICOS	38	KIR2DS4	50	TDO2
3	BTN2A1	15	CD28	27	ICOSLG	39	KIR2DS5	51	TIGIT
4	BTN2A2	16	CD40	28	IDO1	40	KIR3DL1	52	TNFRSF14
5	BTN3A1	17	CD40LG	29	KIR2DL1	41	KIR3DL2	53	TNFRSF18
6	BTNL3	18	CD47	30	KIR2DL2	42	KIR3DL3	54	TNFRSF4
7	BTNL9	19	CD70	31	KIR2DL3	43	KIR3DS1	55	TNFRSF9
8	C10orf54	20	CD80	32	KIR2DL4	44	LAG3	56	TNFSF14
9	CD160	21	CD86	33	KIR2DL5A	45	LGALS9	57	TNFSF18
10	CD209	22	CD96	34	KIR2DL5B	46	PDCD1	58	TNFSF4
11	CD226	23	CEACAM1	35	KIR2DS1	47	PDCD1LG2	59	TNFSF9
12	CD27	24	CTLA4	36	KIR2DS2	48	PVR	60	VTCN1

2.2. Method

2.2.1. Identification of DEICGs

The GSE65682 and GSE57065 were cleaned and normalized by using the “limma” package. The P value was corrected by t-test, and the genes with $|logFC|> 0.5$ and $P < 0.05$ in GSE65682 were DEGs. By cross-referencing these DEGs with ICGs, differentially expressed immune checkpoint genes (DEICGsz) were identified. Then the “ggplot2” package was used to draw the expression box plot of ICGs in sepsis and healthy and the “circlize” package was used to draw the correlation circle plot of each gene in DEICGs.

2.2.2. Immune infiltration analysis

CIBERSORT[9] analysis was conducted to investigate the degree of immune cell infiltration in peripheral blood between sepsis and healthy, and “ggplot2” was conducted to draw the box map for the

ISSN 2706-6819 Vol.6, Issue 9: 7-13, DOI: 10.25236/IJFM.2024.060902

immune cells. After that, we analyzed the connections between DEICGs with immune cells by Spearman's correlation analysis and plotted the correlation map.

2.2.3. WGCNA and extracting the hub genes

WGCNA was performed on the gene expression matrix of the GSE65682 to find the gene modules co-expressed in sepsis and to identify the relationship between the gene network and sepsis, as well as the key genes in the network[10]. The "WCGNA" package and the "limma" package were used to eliminate outliers in the dataset, cluster samples, construct topological overlap and scaleless networks, analyze and merge dynamic modules, and find the core gene module with the strongest correlation with clinical information as the gene module most relevant to sepsis. The hub genes were extracted from the intersection of genes within the DEICGs and core gene modules of WGCNA, which were plotted by the "VennDiagram" package.

2.2.4. Mendelian randomization (MR)analysis

Two-sample MR was applied to evaluate the causality between the hub genes and the risk of sepsis with the definition of SNPs as instrumental variables(IVs)[11]. By searching the Genome-Wide Association Study (GWAS) database for data of hub genes and sepsis, we found that data of CD96 and sepsis were available. The data of CD96 (GWAS ID: PROT-A-470) contains 3301 samples, and the data of sepsis (GWAS ID: ieu−b−5088) contains 462986 samples. P less than 5×10^{-5} was used as the significance threshold and SNPs that conform to this criterion are used as IVs.To address linkage disequilibrium, we only included SNPs with an r^2 less than 0.001 within a 10,000 kb radius of the most significantly associated SNP. By using the "TwoSampleMR" package for analysis, considering the stability of the inverse weighting of variance, this method was selected as the main analysis method while MR-Egger regression, weighted median, weighted model, and simple mode are complementary. Heterogeneity analysis was performed by Cochran Q test, while pleiotropy analysis was judged by MR-Egger regression intercept, and if $P > 0.05$, there was no horizontal pleiotropy. The leave-one-out test was applied to see if a single SNP affected the outcome.

2.2.5. Validation of the biomarker

The potency of the screened biomarkers was verified by plotting the receiver operating characteristic (ROC) curve on the GSE57065 dataset through the "pROC" package, and the area under the curve can represent the validity of the diagnostic test, and the closer the area under the curve is to 1, the better the potency.

3. Results

3.1. Identification and correlation of DEICGs

DEG analysis screened out 36 DEICGs, of which 16 were up-regulated and 20 were down-regulated (figure 1A). The correlation between each of these genes is shown in the figure 1B.

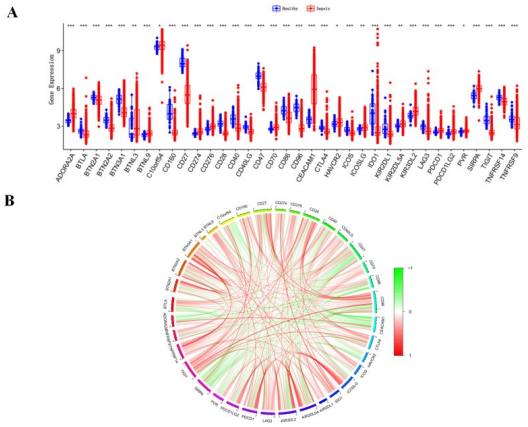

*Figure 1: Identification and correlation of DEICGs. (A) Box plot displaying the expression of DEICGs in the healthy group and the sepsis group. (B)Circle plot showing the correlation between each of DEICGs. *P <0.05,**P <0.01,***P <0.001.*

ISSN 2706-6819 Vol.6, Issue 9: 7-13, DOI: 10.25236/IJFM.2024.060902

3.2. Analysis of Immune infiltration

There are a total of 17 types of immune cells were differentially distributed between the sepsis patients and healthy people (figure 2A). The correlation between DEICGs and immune cells is shown in figure 2B.

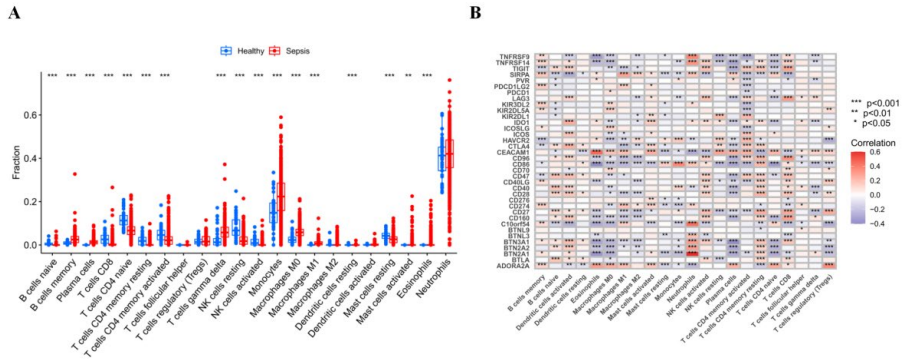

Figure 2: Immune infiltration Analysis and correlation between DEICGs and immune cells. (A) Box plot showing the different distribution of immune cells between health control and sepsis. (B) Diagram of the correlation of DEICGs and immune cells.

3.3. WGCNA

The gene expression matrix of GSE65682 was sorted by expression standard deviation size, and the genes with the top 25% of the median absolute deviation were screened as the genes with the largest fluctuation for WGCNA analysis. There were no outliers in the dataset, and with the soft threshold set to 15, a scale-free network was constructed. The genes were classified into 10 different modules according to the expression profile, and the results of analysis of the correlation coefficients of the modules showed that all modules were $P<0.05$, and the pink module ($r=\pm0.58$, $P=4\times10^{-73}$) contained 740 genes, which were considered to be the genes most associated with the sepsis(figure 3A). The results of WGCNA and DEG were intersected, and 5 hub genes were obtained, which were CD27, CD96, CD28, CD160, and TIGIT(figure 3B).

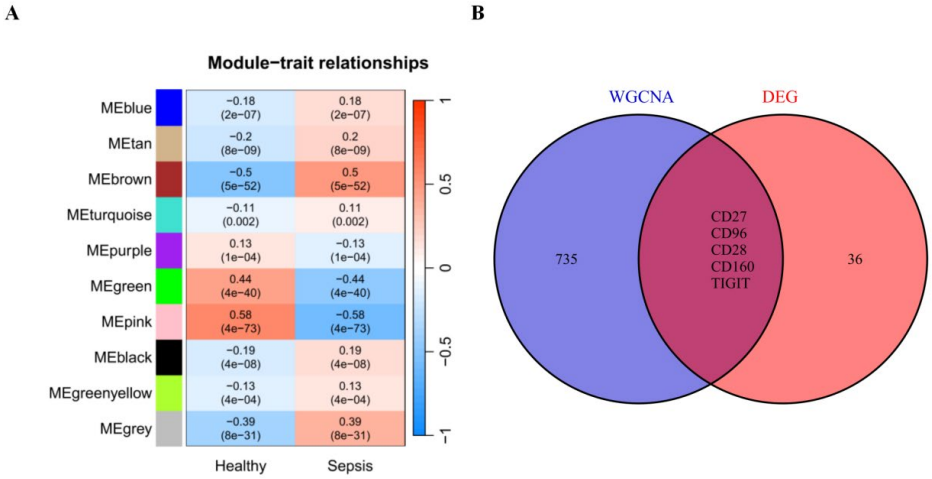

Figure 3: Identification of 5 hub genes using WGCNA. (A) Module-trait heatmap showing the correlation between the clustering gene module and sepsis. (B)Venn plot of the the 5 hub genes.

3.4. CD96 was causally associated with the risk of sepsis

Following the IVS filter criteria set before, a total of 20 SNPs were obtained for Mendelian randomization analysis (Table 3). Among them, rs116852869, rs11979013, rs11981103, rs12380605, rs17438332, rs181355252, rs182518981, rs2583450, rs3132451, rs6802119, rs71602118, rs75957873, and rs948070 showed a protective effect on outcome factors, while rs117023749, rs28361034, rs5746459, rs57741805, rs7021048, rs72655120, and rs78298585 showed a promoting effect on the outcome factors,

ISSN 2706-6819 Vol.6, Issue 9: 7-13, DOI: 10.25236/IJFM.2024.060902

and finally the algorithm comprehensively determined that all the 20 SNPs used as IVs had a protective effect on the outcome factors. The effect of each instrumental variable on outcome factors is shown in Figure 4A.

The results of IVW (*OR* = 0.952, 95%*CI*: 0.91-0.99, *P* = 0.026) indicated that there was a causal relationship between CD96 and sepsis, and the *OR* of CD96 less than 1 suggested CD96 has a protective effect on sepsis, which could intervene in the occurrence of sepsis. With the exception of simple model, all other complementary methods were consistent with the direction of IVW (*OR*＜1). (Table 4, Fig 4B). The Cochran Q test showed that there was no heterogeneity between IVW(*P*=0.951) and MR-Egger regression (*P*=0.930). The intercept *P* value of MR-Egger regression was 0.938, indicating the absence of underlying horizontal pleiotropy.

Table 3: The details of the 20 SNPs used as IVs

SNP	Effector alleles	Non-effector alleles	*β*	*se*	*F*
rs116852869	C	T	-0.5882	0.0744786	62.37162102
rs117023749	T	C	0.4279	0.0530043	65.17219769
rs11979013	C	T	-0.4196	0.0480335	76.31018264
rs11981103	T	C	-0.1143	0.0134972	71.71418947
rs12380605	T	A	-0.4821	0.0503018	91.85593257
rs17438332	G	A	-0.7391	0.095384	60.04202661
rs181355252	T	C	-0.1394	0.0155643	80.21698988
rs182518981	C	T	-0.1242	0.0149923	68.62884086
rs2583450	G	C	-0.486	0.0552371	77.41244764
rs28361034	G	A	0.2666	0.0170667	244.0176168
rs3132451	C	G	-0.3605	0.0163862	484.0096667
rs5746459	T	C	0.1141	0.0133739	72.78722223
rs57741805	A	G	0.1186	0.0138964	72.83913253
rs6802119	T	C	-0.1142	0.0135388	71.14950162
rs7021048	T	C	0.1203	0.0136639	77.51431345
rs71602118	G	T	-0.1243	0.0144766	73.72403692
rs72655120	A	G	0.4588	0.0472297	94.36619512
rs75957873	T	C	-0.2096	0.026002	64.9784053
rs78298585	T	C	0.2353	0.0257939	83.21657375
rs948070	A	G	-0.1173	0.0139368	70.83858682

Table 4: Results of MR

Outcome	Method	SNP num	*OR*	*P*
sepsis	inverse variance weighting	20	0.952	0.026
sepsis	MR Egger	20	0.949	0.254
sepsis	Weighted median	20	0.963	0.229
sepsis	Simple mode	20	1.018	0.725
sepsis	Weighted model	20	0.974	0.518

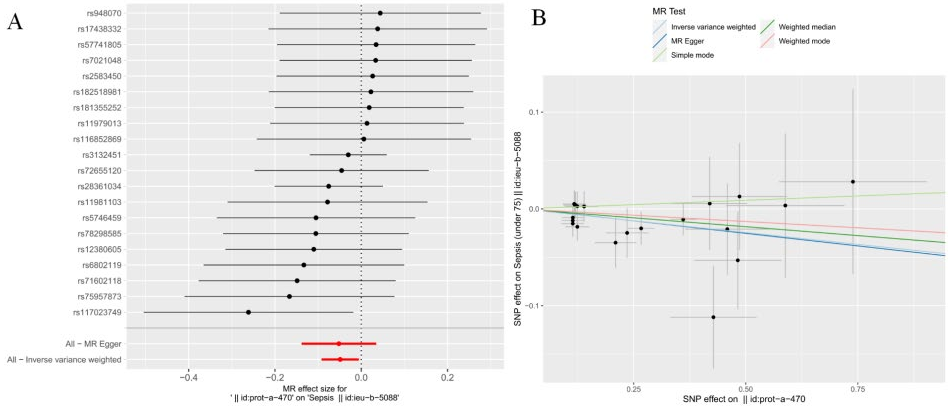

Figure 4: The results of the Mendelian randomization study. (A) Forest plot displaying the causal effect of each SNP on the risk of sepsis. (B) The comprehensive causal effect of CD96 on the risk of sepsis.

ISSN 2706-6819 Vol.6, Issue 9: 7-13, DOI: 10.25236/IJFM.2024.060902

3.5. Analysis of ROC

We calculated ROC curves for CD96 in GSE57065 datasets to determine the diagnostic effect of CD96(figure 5). The AUC values of CD96 was 0.945(cut-off value = 4.074, 95%CI = 0.841-0.960), suggesting CD96 is a good method to determine whether a person is suffering from sepsis.

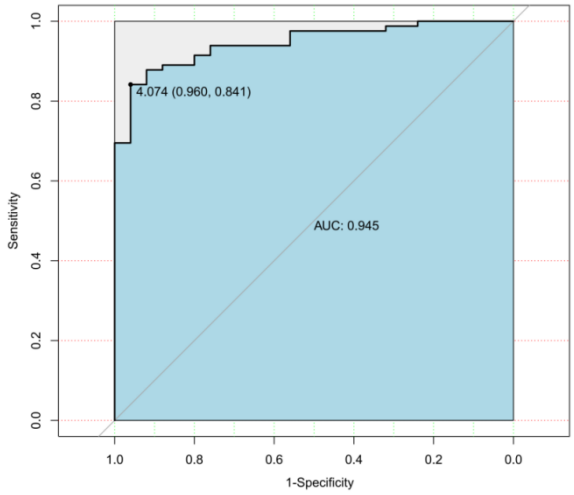

Figure 5: ROC curve of CD96.

4. Discussions

Sepsis progresses rapidly and can easily develop into septic shock and endanger the lives of patients. The SOFA score based on the patient's symptoms and signs is the main basis for clinicians to judge sepsis in the early stage, but the SOFA index lacks specificity and is greatly affected by individual patient differences, which is prone to misdiagnosis[12]. As the gold standard for the laboratory diagnosis of sepsis, blood culture has the shortcomings of long culture time, high false negative rate, and many factors that affect the results of blood culture experiments, which limit the clinical reference value of blood culture results[13]. Laboratory indicators such as absolute white blood cell value, neutrophil percentage, PCT, and CRP have low specificity, which can only be used as a reference diagnosis of sepsis and cannot accurately predict the disease. As for the treatment of sepsis, conventional treatment has a limited impact on the prognosis of sepsis due to the poor immune status. Several studies suggested therapies targeting immune checkpoints can improve symptoms and reduce mortality on patients with sepsis[14-16]. However, immune checkpoint therapy is not effective in all patients due to individual patient differences and different pathogens of infection[17]. Therefore, there is a need to develop a reliable diagnostic method of sepsis for early intervention of the disease and more immune checkpoint targeted therapies.

Our study explored the association between immune immune checkpoints and immune-infiltrating cells. Furthermore, we used a combination of methods to identify CD96 as a potential biomarker that could contribute to the diagnosis and treatment of sepsis. Existing research suggests that CD96 may has a vital function in immune regulation. A study confirmed that $CD8^+$T cells with low CD96 expression had worse proliferative potential than those with high CD96 expression, and the loss of CD96 expression in HIV-specific $CD8^+$T cells was once additionally related with bad response to HIV antigens[18]. Scholars conducted experiments on CD96-deficient mice and found that CD96 can indirectly or directly inhibit NK cell effector function and CD96-deficient mice are able to secrete more IFN-γin response to LPS stimulation, resulting in a stronger inflammatory response[19]. Another study demonstrated that deficiencies in CD96 and its ligand CD155 lead to increased levels of IL-9, which can promote tissue damage in autoimmune vasculitis[20]. The above studies suggest that CD96 was a protective factor to intervene in the development of infectious diseases, which was consistent with the results of our MR analysis in sepsis. However this study did not explore how immune cells are specifically affected by CD96, nor did we investigate the relationship between CD96 and various cytokines that exert inflammatory effects. The role of CD96 in sepsis requires a large number of clinical and experimental studies to explore.

ISSN 2706-6819 Vol.6, Issue 9: 7-13, DOI: 10.25236/IJFM.2024.060902

5. Conclusions

In this study, CD96 was identified as the biomarker of sepsis, which has a causal relationship with sepsis and might be a protective factor for sepsis. However, follow-up experimental studies are needed to elucidate the mechanism of CD96 in sepsis.

References

[1] Kellum J. A., Formeck C. L., Kernan K. F., Gómez H., Carcillo J. A. (2022). Subtypes and mimics of sepsis [J]. Crit. Care Clin. 2022; 38(2), 195–211.
[2] Liu D., Huang S. Y., Sun J. H., Zhang H. C., Cai Q. L., Gao C., et al. (2022). Sepsis-induced immunosuppression: mechanisms, diagnosis and current treatment options[J]. Mil Med. Res. 2022 Oct 9; 9(1):56.
[3] Morad G., Helmink B. A., Sharma P., Wargo J. A. Hallmarks of response, resistance, and toxicity to immune checkpoint blockade[J]. Cel. 2021 Oct 14;184(21):5309-5337.
[4] Topalian S. L., Sznol M., McDermott D. F., et al. Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab[J]. J Clin Oncol. 2023;41(5):943-954.
[5] Huang X., Chen Y. P., Chung C. S., et al. Identification of B7-H1 as a novel mediator of the innate immune/proinflammatory response as well as a possible myeloid cell prognostic biomarker in sepsis[J]. Immunol (2014) 192(3):1091-1112.
[6] Ligers A., Teleshova N., Masterman T., Huang W. X., Hillert J. CTLA-4 gene expression is influenced by promoter and exon 1 polymorphisms[J]. Genes Immun (2001) 2(3):145–152.
[7] Lou J., Wang J., Fei M., Zhang Y., Wang J., Guo Y., Bian J., Deng X. Targeting Lymphocyte Activation Gene 3 to Reverse T-Lymphocyte Dysfunction and Improve Survival in Murine Polymicrobial Sepsis[J]. J Infect Dis. (2020)222(6):1051-1061.
[8] Hu F. F., Liu C. J., Liu L. L., Zhang Q., Guo A. Y. Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. Brief. Bioinform. 2021;22:bbaa176.
[9] Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., & Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT[J]. Methods Mol Biol. 2018:1711:243-259.
[10] Tian Z. L., He W. X., Tang J. N., et al. Identification of important modules and biomarkers in breast cancer based on WGCNA[J]. OncoTargets and Therapy, 2020, 13: 6805-6817.
[11] Gkatzionis A, Burgess S, Newcombe PJ (2023) Statistical methods for cis-Mendelian randomization with two-sample summary-level data. Genet Epidemiol 47(1):3–25.
[12] Lambden S, Laterr P. F., Levy M. M., et al. The SOFA score-development, utility and challenges of accurate assessment in clinical trials[J]. Crit Care, 2019, 23(1):374.
[13] Ali MA, Khamis Hussein S, Ali Mohamed E, et al. Diagnostic and prognostic values of miR181b-5p and miR21-5p for neonatal sepsis risk and their link to SNAP II score and disease mortality[J]. Noncoding RNA Res, 2022, 8(1):115-125.
[14] Huang S. Y., Liu D., Sun J. H., et al. Tim-3 regulates sepsis-induced immunosuppression by inhibiting the NF-kB signaling pathway in CD4 T cells[J]. Mol Ther. 2022; 30(3):1227–38.
[15] Chen R., Zhou L. PD-1 signaling pathway in sepsis: does it have a future?[J]. Clin Immunol. 2021;229:108742.
[16] Shindo Y., McDonough J. S., Chang K. C., et al. Anti-PD-L1 peptide improves survival in sepsis[J]. J Surg Res. 2017; 208: 33–39.
[17] Li W., Wang M., Zhu B., Zhu Y., Xi X. Prediction of median survival time in sepsis patients by the SOFA score combined with different predictors[J]. Burns Trauma. 2020; 8: tkz006.
[18] Bunet R., Nayrac M., Ramani H., et al. Loss of CD96 Expression as a Marker of HIV-Specific CD8+ T-Cell Differentiation and Dysfunction. Front. Immunol. 2021; 12: 673061. .
[19] Chiang E. Y., de Almeida P. E., de Almeida Nagata D. E., et al. CD96 functions as a co-stimulatory receptor to enhance CD8+ T cell activation and effector responses. Eur J Immunol. 2020; 50(6):891–902.
[20] Ohtsuki S., Wang C., Watanabe R., et al. Deficiency of the CD155-CD96 immune checkpoint controls IL-9 production in giant cell arteritis. Cell Rep. Med. 2023;4:101012.