Am J Transl Res 2022;14(3):1807-1817
www.ajtr.org /ISSN:1943-8141/AJTR0138818

Original Article

SOX18-associated gene signature predicts sepsis outcome

Anlin Feng[1,3,4], Wenli Ma[1,3,4], Reem Faraj[1,2], Stephen M Black[1,3,4], Ting Wang[1,3,4]

[1]Center of Translational Science, Florida International University, Port St. Lucie, FL 34987, USA; [2]Graduate Program of Clinical and Translational Sciences, University of Arizona, Phoenix, AZ 85004, USA; Departments of [3]Environmental Health Sciences, Robert Stempel College of Public Health and Social Work, [4]Cellular Biology & Pharmacology, Howard Wertheim College of Medicine, Florida International University, Miami, FL 33199, USA

Received September 2, 2021; Accepted November 10, 2021; Epub March 15, 2022; Published March 30, 2022

Abstract: Objectives: Sepsis is a critical medical condition associated with an high mortality. Currently, there are no reliable diagnostic or prognostic biomarkers to evaluate sepsis outcomes. SRY (sex-determining region on the Y chromosome)-box transcription factor 18 (SOX18) is an endothelial barrier protective protein, and a decreased level of SOX18 expression is involved in disruption of human endothelial cell barrier integrity. Over-expression of SOX18 attenuates the bacterial lipopolysaccharide (LPS)-mediated disruption of the vascular barrier and is associated with favorable prognosis. The utility of SOX18-related genes as biomarkers in sepsis is uncertain. Methods: Transcriptomic analysis was used to profile the PBMC samples of patients with sepsis across two Gene Expression Omnibus (GEO) datasets with survival data. An 84-gene signature was derived from discovery datasets that correlated with SOX18 gene expression and sepsis survival. Results: Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis revealed Th1 and Th2 cell differentiation, Cytokine-cytokine receptor interaction, and T cell receptor signaling pathways as the most significantly enriched KEGG pathways among 84 genes. A severity score based on the gene expression of 84 genes was allocated to each patient. A notable increase was detected in sepsis patients compared to healthy controls in both discovery and validation cohorts. SOX18-associated gene signature discriminated severe cases from mild cases and performed significantly better than both random 84-gene sets from whole genomes or sepsis survival-related genes. Furthermore, we obtained an 18-gene signature from screening these 84 genes in a LASSO model, which performed better in both discovery and validation cohorts. Conclusions: Data support SOX18-associated gene signatures as a prognostic biomarker for sepsis.

Keywords: SOX18, sepsis survival, gene signature

Introduction

Sepsis is still a poorly understood and diverse clinical condition resulting in shock, multi-organ failure, and death [1]. Sepsis often occurs secondary to bacterial infections in the lungs, abdomen, and urinary tract and can ultimately result in systemic inflammation and organ failure [2]. In the last few decades, improving sepsis outcomes has focused primarily on prevention and standardizing treatment protocols developed by individual respective clinical centers or hospitals. However, the complex and diverse clinical nature of sepsis presents a challenge in the development of effective, accurate, and selective biomarkers to predict the outcome to determine an appropriate medical response. Therefore, there is a significant need for these biomarkers.

Theoretically, a sepsis biomarker can measure biologic or pathogenic processes which can then be utilized for diagnostic or prognostic purposes as well as to stage sepsis severity. Sepsis-associated infections and systematic inflammation exert significant impacts on gene expression, which subsequently impacts the pathogenesis of sepsis. Of note, sepsis is associated with more than 3,700 gene expression alternations in the innate immune response [3], illustrating that expression analysis is effective for sepsis biomarker discovery. Microarray-based gene expression profiling has played an important role in clinical research [4], and

recent gene expression profiling has identified gene signatures [3-8] that may have utility as effective biomarkers in sepsis. However, more work is needed to validate them.

Transcription factors (TFs) are proteins which contain a DNA-binding domain and multiple protein partners. The (SRY)-box transcription factor 18 (SOX18), is a member of the SOXF transcription factor family which possess critical and diverse roles as master regulators of cellular reprogramming and cell fate decisions [9]. In humans, SOX18 modulates lymphangiogenesis and angiogenesis [10, 11]. Prior studies have utilized SOX18 as a prognostic biomarker in hepatocellular carcinoma, lung cancer, and esophageal cancer [12, 13]. Further, our recent work has shown that SOX18 sustains the integrity of endothelial barrier by increasing the expression of tight junction protein Claudin-5 (CLDN5) under conditions of laminar shear stress [14]. We also illustrated that the suppression of SOX18-CLDN5 cascade by NF-κB is involved in disruption of endothelial cell barrier integrity, implying that SOX18 plays a role in abrogating the endothelial barrier dysfunction associated with LPS-mediated sepsis {Gross, 2017 #3931}. If verified, SOX18 may be a novel target for assessment of endothelial barrier integrity in sepsis resulting from gram-negative bacteria. To begin to evaluate this possibility we investigated SOX18 co-expressed genes with gene expression data in datasets from septic patients to determine the SOX18-associated gene signature as a biomarker for sepsis. First, we discovered a set of SOX18 co-expressed genes from the discovery cohort. Additionally, we identified a set of differentially-expressed genes as sepsis survival-related genes in the discovery cohort. By intersecting these two gene sets, we prioritized our SOX18-associated gene signature which has been demonstrated to successfully discriminate severe cases from mild cases. Our hypothesis was that SOX18-based gene signature can assess prognosis following a diagnosis of sepsis.

Methods

Microarray datasets

Two ArrayExpress datasets (E-MTAB-4421 and E-MTAB-4451) with a total of 371 sepsis patients which meet the following criteria were included in our study: PBMCs samples from sepsis patients, microarray datasets with survival data, and sample counts large than 100. E-MTAB-4421 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4421/), with 78 sepsis non-survivors and 187 sepsis survivors was assigned as the discovery cohort. Samples were obtained for gene expression profiling by quickly isolating peripheral blood mononuclear cells (PBMCs). Samples from survivors were assessed 4 weeks from intensive care unit (ICU) admission. E-MTAB-4451 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4451/) included 56 sepsis non-survivors and 50 sepsis survivors (Supplementary Table 1).

Sepsis survival-related genes and SOX18-related genes

We utilized the R "limma" (version 3.13) and "gcrma" (version 3.13) package to find differentially expressed genes (DEGs) between survivors and non-survivors in the discovery cohort which was defined as "sepsis survival-related genes" in this study. The discovery cohort was chosen to perform the SOX18 correlation test. We ran a Pearson correlation test for all protein-coding genes with SOX18 in the discovery cohort to determine all the SOX18 co-expressed genes which were deemed as "SOX18-related genes".

Survival score

The survival score was assigned for each patient using a linear combination of gene expression values and weight value of genes in the SOX18-GS and SOX18-GS2. The formula for survival score is:

$$risk\ \ score = \sum_{i=1}^{n} W_i \left(\frac{e_i - \mu_i}{s_i} \right)$$

The sign n represents the count of genes included in SOX18-GS and SOX18-GS2 in each dataset, W_i shows the weight value of each gene (see in **Tables 1** and **2**). $e_{a.}$ shows the expression level of each gene. q_i and s_{ix} are the corresponding mean and standard deviation value for the *i*t gene among whole samples.

LASSO regression

Gene expression of SOX18-GS (84 genes) in the discovery cohort were used to build a binomial LASSO model by R package glmnet [23]

Table 1. 84-gene signature

Gene Name	Weight
GGA2	1.00
FAM38A	1.00
CX3CR1	1.00
CD3D	1.00
PRF1	-1.00
NAPSB	1.00
PEA15	-1.00
CD247	1.00
TMEM204	-1.00
ACP5	1.00
ESYT1	1.00
RGL4	-1.00
HDC	1.00
KCNMB1	1.00
CST7	1.00
LOC648470	1.00
PIK3IP1	-1.00
RASGRP3	-1.00
PVRIG	-1.00
NCR3	1.00
FCRLA	1.00
GIMAP7	1.00
CD27	1.00
SH3TC1	-1.00
RETN	-1.00
TNS3	-1.00
CD5	1.00
CCR7	1.00
ABI3	1.00
CD79B	1.00
PRKCQ	-1.00
SPOCK2	-1.00
CYYR1	1.00
GZMA	1.00
SLC27A1	-1.00
MATK	1.00
CD3E	1.00
CECR1	1.00
TLR7	-1.00
ENG	1.00
GZMK	1.00
P2RY5	-1.00
MCOLN2	1.00
IL18R1	1.00
PPP3CC	-1.00
SLAMF6	-1.00
CXCR5	1.00
BTN3A3	1.00
FLT3LG	1.00
GZMH	1.00
GPBAR1	1.00
CD40LG	1.00
GPR114	1.00
EOMES	1.00
IL7R	1.00
TGFBR3	-1.00
CA4	1.00
ENOPH1	1.00
PTGDR	-1.00
LBH	1.00
P2RX7	-1.00
PTPRCAP	-1.00
CD4	1.00
OAF	1.00
RUNX3	-1.00
IL2RB	1.00
RRP15	-1.00
LY86	1.00
NOP58	1.00
DOK3	1.00
PPP1R16B	-1.00
LRP1	1.00
CSF1R	1.00
FAIM3	1.00
UBASH3A	-1.00
CD2	1.00
GIMAP5	1.00
CD52	1.00
ADARB1	1.00
POU2F2	-1.00
PYHIN1	-1.00
KLRB1	1.00
CD6	1.00
MPRIP	1.00

(version 4.1-2). The k-fold cross-validation was performed to determine the best lambda value that obtains the lowest test mean squared error (MSE). Lastly, the final LASSO model was created by this optimal lambda value. The LASSO regression model was used to calculate the survival score. Genes with zero coefficient had been removed from the 84 gene list, while other genes' coefficients were utilized as weight values in the survival score.

Enrichment analysis

To analyze the gene list at the functional and biologic level, the online Database for Annotation, Visualization and Integrated Discovery

Table 2. 18-gene signature

Gene Name	Weight
ACP5	0.0096
CA4	-0.0203
CD6	0.0376
DOK3	-0.0450
ESYT1	0.0346
FAM38A	0.1171
GIMAP7	0.0114
GPR114	0.0011
IL18R1	-0.0188
LBH	0.0742
LOC648470	0.0271
NOP58	0.0611
OAF	0.0580
POU2F2	0.0041
RGL4	-0.0057
RRP15	0.0200
RUNX3	0.0015
TGFBR3	0.0527

(DAVID) v6.8 database [24, 25] (https://david.ncifcrf.gov/tools.jsp) which is a tools for the analysis of the relevant biologic and functional annotation of gene lists, was utilized to provide signaling pathway interpretation of the KEGG pathway analysis for our gene lists.

ChEA3 (https://maayanlab.cloud/chea3/) [15], a web-based TFEA tool, was used for transcription factor enrichment analysis. This web application output contained transcription factor enrichment results in the form of sortable tables for every library, and the integration result. The average integrated ranks across the library method were chosen for enrichment analysis of our gene lists.

Protein-protein interaction network

The STRING database (https://string-db.org) is an online tool [26] (version 11.0b) to assess the protein-protein interaction (PPI) information for gene list. To estimate the PPI relationships, the identified 84 genes were mapped by STRING, and genes with a combined score >0.4 were selected to generate a PPI network. The PPI networks were visualized by Cytoscape software (version 3.8.2).

Statistical analyses

Statistical calculations were performed in the R language. The R packages (ade4 and price) were used for principal component analysis and ROC curves. The cutoffs for the DEGs were: |fold change| >2 and false discovery rate (FDR) <0.05. Cutoffs for SOX18 co-expression analysis were: Pearson r value >0.4, adjusted-*P* value <0.05. ROC curves were used to evaluate SOX18 gene signatures classification by identifying the dataset's ability to discriminate between classes, and AUC values measured the performance of our gene signatures. The dimensionality reduction method PCA was used to decrease the dimensionality of large data sets.

Results

Discovery of SOX18 co-expressed genes and sepsis survival-related genes

We chose E-MTAB-4421 with a larger sample count (265 patients) [6] as the discovery cohort to identify sepsis survival-related genes, which were determined by differentially expressed genes (DEGs) between the 108 sepsis high risk (SHR) group and 157 sepsis low risk (SLR) group. SHR and SLR grouping were based on the clustering of gene expression data in E-MTAB-4421 (Supplementary Figure 1 and Supplementary Table 1). The clustering was formed by Ward's method, Euclidean distance, and k-means to establish the proper number of groups and consolidation of group members. Compared to SLR, SHR possessed significantly higher early mortality that was associated with immunosuppression and T-cell exhaustion [6]. We identified 739 genes (false discovery rate [FDR] <5% and fold change [FC] >2) differentially expressed between SHR and SLR in the discovery cohort and represented them as sepsis survival-associated genes in this study (**Figure 1A** and Supplementary Table 2). The KEGG analysis of these 739 genes illustrated that T cell receptor signaling pathways, Th1 and Th2 differentiation, and immunodeficiency (**Figure 1B**) were enriched among sepsis survival-associated genes. These KEGG pathways were consistent with biologic functions from SHR and SLR grouping [6]. We next utilized gene co-expression analysis to obtain the entire SOX18 co-expressed genes from the discovery cohort. 813 genes were identified as SOX18 co-expressed genes (Pearson r>0.4, adjusted-*P* value <0.05; Supplementary Table 3). KEGG analysis of these 813 genes showed that SOX18-related genes correlated with cell adhesion molecules, Th1 and Th2 differentiation

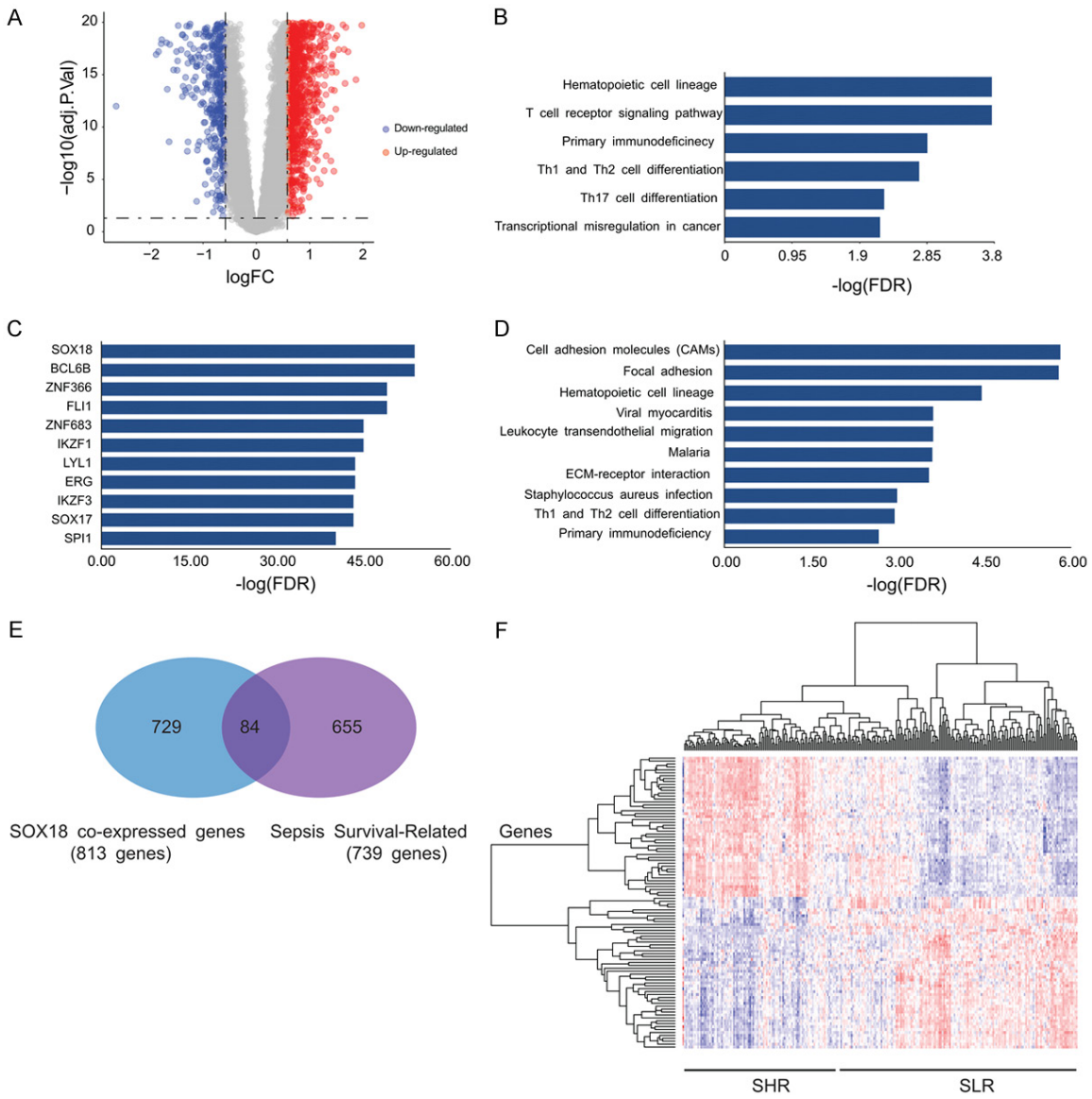


Figure 1. Discovery of SOX18 co-expressed genes and sepsis survival-related genes. A. Volcano plot showing DEGs among sepsis survivors and non-survivors. B. Enriched KEGG pathways among sepsis survival-related genes. C. Enriched TF target genes among SOX18 co-expressed genes; D. Enriched KEGG pathways among SOX18 co-expressed genes. E. Venn plot depicting sepsis survival and SOX18 co-expressed overlapping genes; F. Heat map revealing the 84-gene signature gene expression in the discovery cohort. Red represents increased gene expression, while blue represents decreased gene expression.

and immunodeficiency (**Figure 1D**). Transcription Factor Enrichment Analysis (TFEA) is a method based on the overlap between input given lists and annotated transcription factor targets collected from multiple databases [15]. We used TFEA to confirm SOX18 as the central regulator of these 813 genes (**Figure 1C**). The transcription factor targets over-expression analysis showed significant overlapping genes between SOX18 co-expressed genes and SOX-18 target genes (**Figure 1C**) with a significant enriched TF score, indicating that our SOX18 co-expressed genes are not random co-expression genes and are capable of being considered SOX18-associated genes. By intersecting the 813 SOX18-related genes and the 739-sepsis survival-associated genes, we identified 84 overlapping genes (**Figure 1E**). These 84 genes were assigned as our SOX18-related gene signature for sepsis (SOX18-SG) (**Table 1**). This

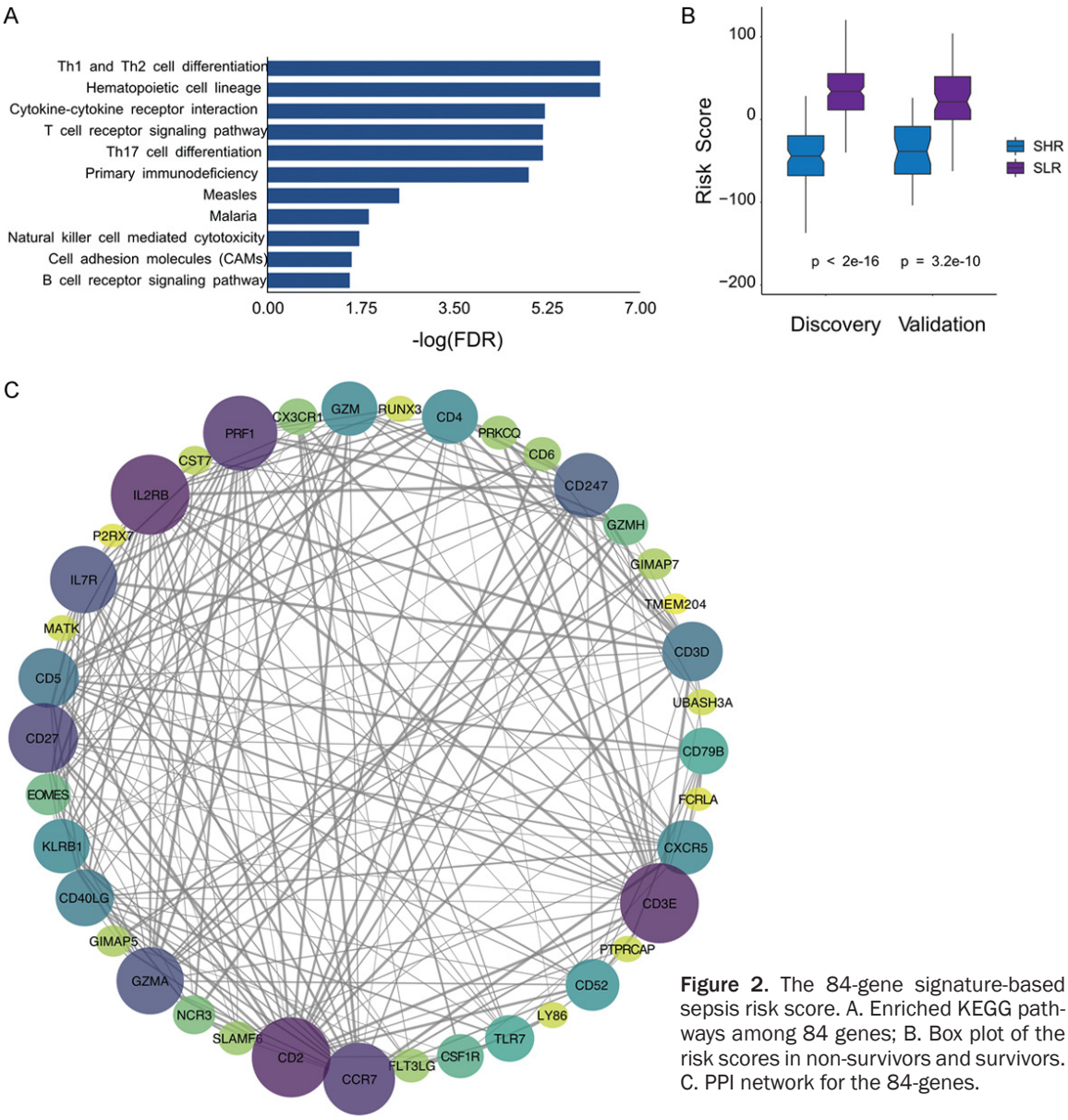


Figure 2. The 84-gene signature-based sepsis risk score. A. Enriched KEGG pathways among 84 genes; B. Box plot of the risk scores in non-survivors and survivors. C. PPI network for the 84-genes.

overlap is statistically significant (hypergeometric *P* value <0.05), illustrating that SOX18-related genes are enriched in sepsis survival-associated genes. The heatmap we produced demonstrates that the 84 genes (SOX18-SG) can differentiate SHR from SLR in both the discovery and a validation cohort (**Figure 1F**).

84-gene signature-based sepsis risk score

We next identified the signaling pathways and functions of the 84 genes in the SOX18-SG. KEGG pathways such as Th1 and Th2 differentiation, immunodeficiency, and T cell receptor signaling pathway, among others (**Figure 2A**) were significantly enriched among these 84 genes. A protein and protein interaction (PPI) network was then utilized to verify whether an interaction or relationship exists between these genes. The PPI network demonstrated that 39 genes establish a strong interaction network with each other (**Figure 2C**). Thus, our results confirm that the 84 genes identified in our SOX18-SG possess a vigorous connection and network. Further, these results strongly suggest that SOX18-SG builds a bridge between SOX18-associated proteins and sepsis prognosis. To estimate the probability of a specific clinical outcome in a sepsis patient, a survival rating score was employed using a formula consisting of linear combinations of the 84 genes

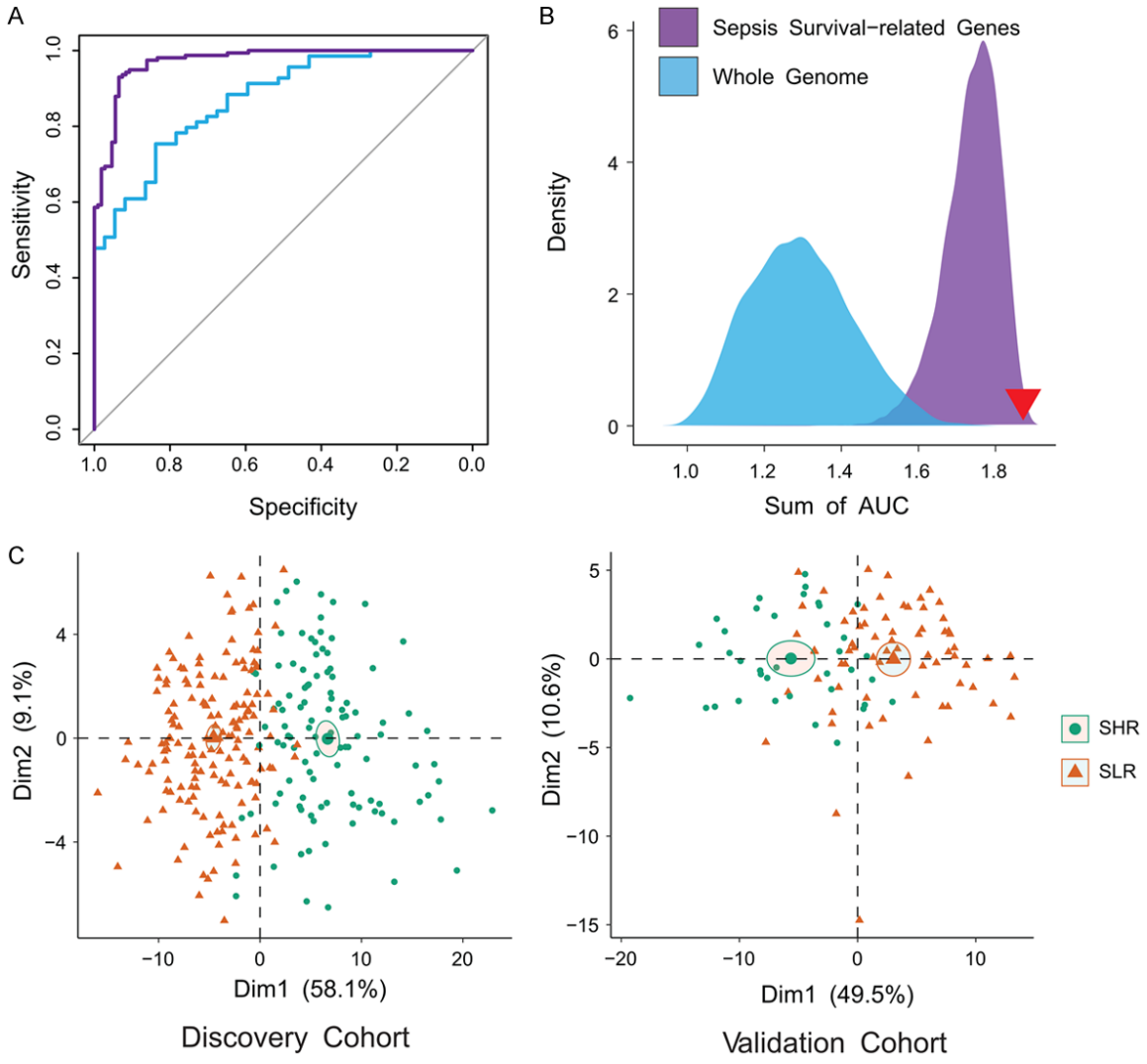


Figure 3. 84-gene signature-based sepsis risk score differentiates non-survivors from survivors in both the discovery and validation cohort. A. ROC curves of the 84-gene signature in the discovery and validation cohorts. B. Density distribution plot of random gene signatures from the whole genome genes and only sepsis survival-related genes, the 84-gene signature-based AUC value was marked as red inverted triangle; C. PCA plot of the 84-gene signature for the discovery cohort and validation cohorts.

identified in our SOX18-SG (**Table 1**). A higher survival score correlates with a better prognosis and survival rate. Survival scores from SLR were significantly higher than those of SHR in both the discovery (*P*-value <2e-16) and validation cohorts (*P*-value =3.2e-10) (**Figure 2B**), confirming that our SOX18-SG can estimate clinical outcome in sepsis.

84-gene signature-based sepsis risk score differentiates non-survivors from survivors

We next performed a validation study. A receiver operating characteristic (ROC) curve is a graphical plot showing the discrimination capacity of a binary classifier structure. The area under the curve (AUC) illustrates the degree of separability and capability of discriminating between classes. We obtained excellent AUC value under the ROC curve for both the discovery (AUC: 0.97) and validation cohorts (AUC: 0.87) respectively (**Figure 3A**). Additionally, a principal component analysis (PCA) was performed for our 84-gene expression model to simplify its complexity. PCA is a dimensionality reduction method that is utilized to decrease the dimensionality of large data sets. PCA revealed that the 84 genes entirely discriminated SHR patients from SLR patients (**Figure**

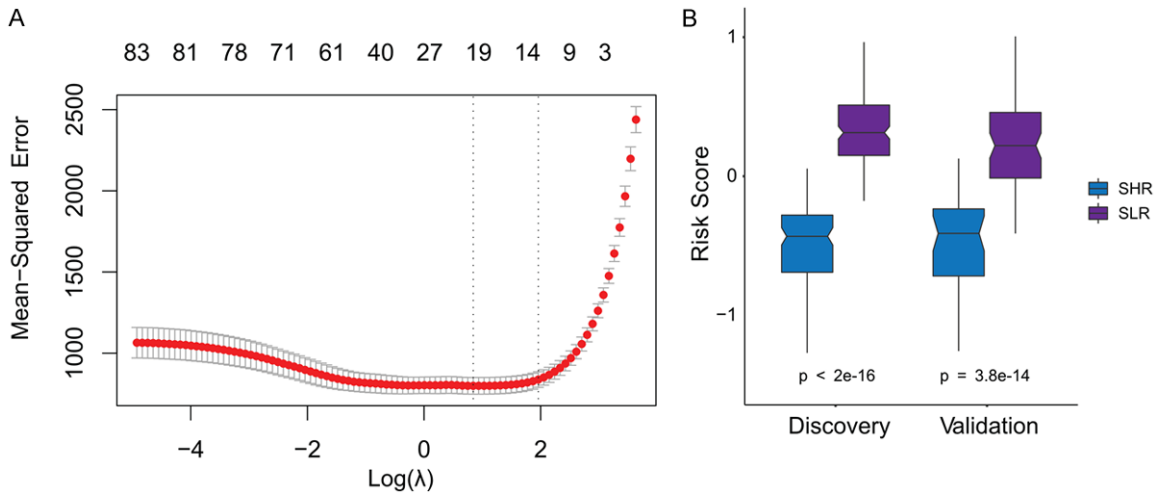


Figure 4. The 18-gene signature-based sepsis risk score. A. The gene signature screening of optical parameters in LASSO model; B. Box plot of the risk scores in non-survivors and survivors.

3C) in both the discovery and validation cohorts. Additionally, the SOX18-GS demonstrated superiority in stratifying sepsis patients based on likelihood of survival relative to random gene sets with an identical gene count from the whole genome or survival-related genes (**Figure 3B**).

18-gene signature-based sepsis risk score

Next, to strengthen the ability of SOX18-GS in stratifying sepsis patients according to likelihood of survival, a LASSO regression model was conducted to screen the 84 genes in the discovery cohort. A λ value of 0.76 was chosen to determine genes that could predict sepsis survival most accurately (**Figure 4A**). From these analyses, we obtained an 18-gene signature (**Table 2**) with a non-zero regression coefficient. This was utilized as the second SOX18 gene signature (SOX18-SG2) in this study. SOX18-SG2's survival scores from SLR were also higher than those of SHR in both the discovery (*P*-value <2e-16) and validation cohorts (*P*-value =3.8e-14) (**Figure 4B**). SOX18-SG2 also had better AUC values under the ROC curve for both the discovery (AUC: 0.99) and validation cohorts (AUC: 0.94) (**Figure 5A**). Further, SOX18-GS2 was superior in grouping sepsis patients according to survival status relative to random gene sets containing 18 genes from either the whole genome or survival-related genes (**Figure 5B**). A PCA plot for our 18-gene expression revealed that these 18 genes were entirely able to discriminate SHR patients from SLR patients (**Figure 5C**) in both the discovery and validation cohorts.

Discussion

Despite the publication of a few sepsis gene signatures, evaluating a prognosis following a sepsis diagnosis is still very difficult and overly reliant on the diagnosis of the attending physician. SOX18 mRNA and protein expression levels are decreased in mice exposed to LPS, and increasing expression of SOX18 has demonstrated protection against LPS exposure likely through the regulation of genes that protect endothelial cell barrier integrity [14]. Therefore, SOX18 and SOX18-related proteins may be novel targets for sepsis therapy. Further, a SOX18-related gene signature may also be a prognostic indicator for sepsis progression and survival. In this study, utilizing two GEO datasets of sepsis, including gene expression data from PBMCs samples, we made two important contributions to the field. First, the correlation of a SOX18 target and associated gene in sepsis survival could be verified with gene expression analysis; and second, SOX18-GS and SOX18-GS2 are "independent" prognostic markers for sepsis survival. The categorization of diverse sepsis populations using molecular markers may guide the management and treatment of sepsis, ultimately allowing for targeted therapies and preventing unnecessary and harmful interventions. Currently, there are no gene signatures available for use in the evaluation of sepsis severity [16-18]. It is important to

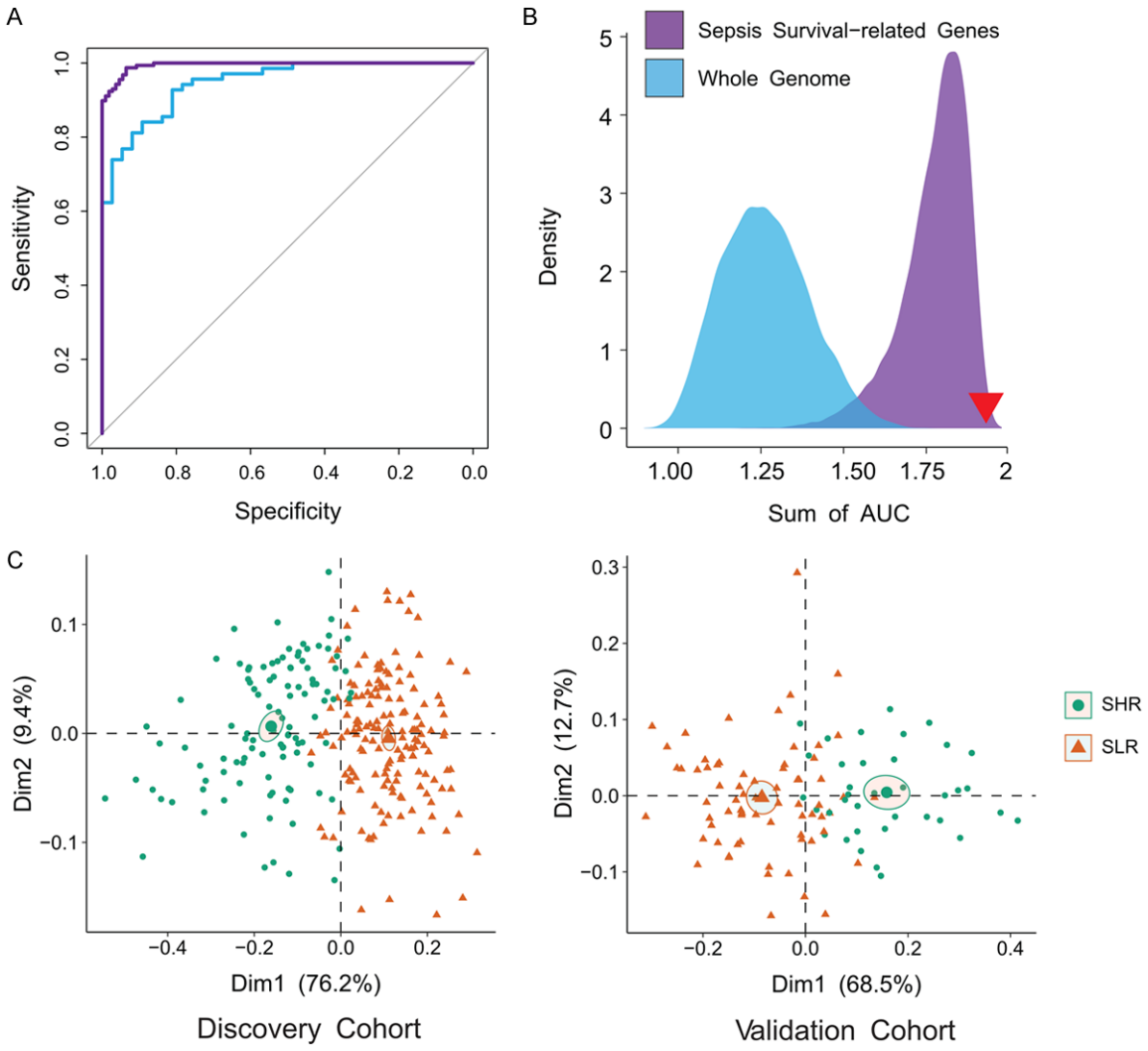


Figure 5. 18-gene signature-based sepsis risk score differentiated non-survivors from survivors in both the discovery and validation cohort. A. ROC curves of the 18-gene signature in the discovery and validation cohorts. B. Density distribution plot of random gene signatures from the whole genome genes and only sepsis survival-related genes, the 18-gene signature-based AUC value was marked as red inverted triangle; C. PCA plot of the 18-gene signature for the discovery cohort and validation cohorts.

note that biomarkers utilizing a single gene are not sufficient for the accurate diagnosis of a sepsis patient and cannot accurately determine disease severity. Prior studies have attempted to randomly combine pro- and anti-inflammatory molecular markers but without success [19] but few associations, or biologic processes, identifying corresponding molecular mechanisms were found among these biomarkers. However, by evaluating gene signatures amongst overlapping genes between SOX18 co-expressed genes and sepsis survival-related genes we were able to identify two gene signature sets: SOX18-GS and SOX18-GS2 that are very promising molecular biomarkers for assessing clinical outcomes of sepsis based on the survival assessment score in our included datasets.

The restoration of immune function is vital for the survival of patients with sepsis. Therapies that enhance host immunity might improve sepsis survival. Our KEGG signaling pathway analysis identified several immunity pathways as being enriched in our SOX18-GS. These included Th1 and Th2 differentiation, and B and T cell receptor signaling pathways. This also activates transcription of PROX1 and other

genes coding for lymphatic endothelial markers. CD4+ T cells play a major role in orchestrating cellular and humoral immune responses in sepsis. IFN-I activated B cells increase as an early protective innate immune response during bacterial sepsis [20]. Further, the gene expression level of these immune pathways is higher in patients with mild sepsis, suggesting they support the immune response [21]. The 84 GS we identified had significant correlations and interactions with each other. Consequently, the 84-gene signature we identified is not only an effective prognostic biomarker for sepsis patients but also possibly indicates a molecular mechanism in the pathogenesis of sepsis.

Gao et al. [22] found that exosomes derived from non-immune cells in sepsis could promote the differentiation of Th1/Th2 cells in the middle and late phase of sepsis. The growth factor GM-CSF increases the proliferation of T cells by the Toll-Like Receptor 4 pathway. Various studies have clarified many different pathophysiologic processes involved in sepsis and have revealed an important regulatory role of pro- and anti-inflammatory cytokines in disease progression. These findings have led to the development of promising anti-cytokine and immunomodulating treatment strategies.

SOX18-GS and SOX18-GS2 have two advantages relative to other sepsis biomarkers. First, we have obtained two gene signatures in this study, including the 84 gene signature, that have a robust association with immune signaling pathways and a PPI network. Our SOX18 gene signatures suggest that target genes of SOX18 may reflect immune system status. Thus, this study revealed that immunity pathways regulated by SOX18 are central factors for assessing the risk of sepsis patients. Second, our study was validated using a number of quality control methods. PCA plots demonstrate that our SOX18 gene signatures are robust tools to differentiate high-risk patients from low-risk patients in all included datasets. In contrast to most AUC values from random 84 or 18 gene signatures, our two gene signatures showed significantly higher prognostic power in both the discovery and validation cohorts.

In conclusion, we have identified two SOX18-associated gene signatures containing 84 and 18 genes, respectively, that can predict clinical outcome in patients with sepsis. Thus, SOX18-GS and SOX18-GS2 have the potential to be second-generation biomarkers for identifying sepsis progression in patients. This may guide the management and treatment of this serious clinical condition.

Acknowledgements

This study is supported by National Institutes of Health research grants P01HL134610, P01-HL146369, and T32HL007249.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Ting Wang, Center of Translational Science, Florida International University, 11350 SW Village Parkway, Port St. Lucie, FL 34987, USA. Tel: 772-345-4716; E-mail: tinwang@fiu.edu

References

[1] Cohen J, Vincent JL, Adhikari NKJ, Machado FR, Angus DC, Calandra T, Jaton K, Giulieri S, Delaloye J, Opal S, Tracey K, van der Poll T and Pelfrene E. Sepsis: a roadmap for future research. Lancet Infect Dis 2015; 15: 581-614.

[2] Annane D. Sepsis clinical knowledge: a role of steroid treatment. Minerva Anestesiol 2003; 69: 254-257.

[3] Deigner HP and Kohl M. The molecular sepsis signature. Crit Care Med 2009; 37: 1137-1138.

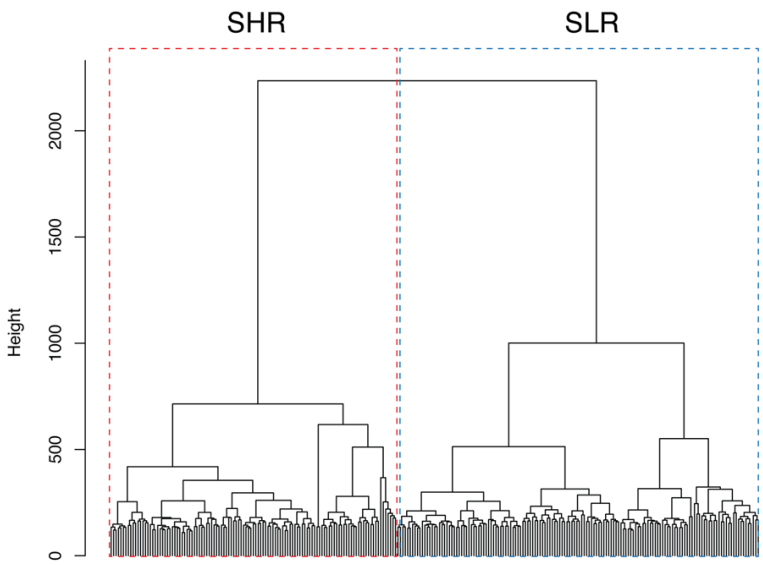
[4] Howrylak JA, Dolinay T, Lucht L, Wang Z, Christiani DC, Sethi JM, Xing EP, Donahoe MP and Choi AM. Discovery of the gene signature for acute lung injury in patients with sepsis. Physiol Genomics 2009; 37: 133-139.

[5] Reyes M, Filbin MR, Bhattacharyya RP, Billman K, Eisenhaure T, Hung DT, Levy BD, Baron RM, Blainey PC, Goldberg MB and Hacohen N. An immune-cell signature of bacterial sepsis. Nat Med 2020; 26: 333-340.

[6] Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nürnberg P, Schultz MJ, Horn J, Cremer OL, Bonten MJ, Hinds CJ, Wong HR, Knight JC and van der Poll T; MARS consortium. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. Lancet Respir Med 2017; 5: 816-826.

[7] Feng A, Rice AD, Zhang Y, Kelly GT, Zhou T and Wang T. S1PR1-associated molecular signature predicts survival in patients with sepsis. Shock 2020; 53: 284-292.

[8] Feng A, Ma W, Faraj R, Kelly GT, Black SM, Fallon MB and Wang T. Identification of S1PR3 gene signature involved in survival of sepsis patients. BMC Med Genomics 2021; 14: 43.
[9] Grimm D, Bauer J, Wise P, Krüger M, Simonsen U, Wehland M, Infanger M and Corydon TJ. The role of SOX family members in solid tumours and metastasis. Semin Cancer Biol 2020; 67: 122-153.
[10] Duong T, Koltowska K, Pichol-Thievend C, Le Guen L, Fontaine F, Smith KA, Truong V, Skoczylas R, Stacker SA, Achen MG, Koopman P, Hogan BM and Francois M. VEGFD regulates blood vascular development by modulating SOX18 activity. Blood 2014; 123: 1102-1112.
[11] François M, Caprini A, Hosking B, Orsenigo F, Wilhelm D, Browne C, Paavonen K, Karnezis T, Shayan R, Downes M, Davidson T, Tutt D, Cheah KSE, Stacker SA, Muscat GEO, Achen MG, Dejana E and Koopman P. Sox18 induces development of the lymphatic vasculature in mice. Nature 2008; 456: 643-647.
[12] Jethon A, Pula B, Olbromski M, Werynska B, Muszczynska-Bernhard B, Witkiewicz W, Dziegiel P and Podhorska-Okolow M. Prognostic significance of SOX18 expression in non-small cell lung cancer. Int J Oncol 2015; 46: 123-132.
[13] Zhu YH, Yang F, Zhang SS, Zeng TT, Xie X and Guan XY. High expression of biglycan is associated with poor prognosis in patients with esophageal squamous cell carcinoma. Int J Clin Exp Pathol 2013; 6: 2497-2505.
[14] Gross CM, Kellner M, Wang T, Lu Q, Sun X, Zemskov EA, Noonepalle S, Kangath A, Kumar S, Gonzalez-Garay M, Desai AA, Aggarwal S, Gorshkov B, Klinger C, Verin AD, Catravas JD, Jacobson JR, Yuan JX, Rafikov R, Garcia JGN and Black SM. LPS-induced acute lung injury involves NF-κB-mediated downregulation of SOX18. Am J Respir Cell Mol Biol 2018; 58: 614-624.
[15] Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, Jagodnik KM, Kropiwnicki E, Wang Z and Ma'ayan A. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res 2019; 47: W212-W224.
[16] Faix JD. Biomarkers of sepsis. Crit Rev Clin Lab Sci 2013; 50: 23-36.
[17] Huang M, Cai S and Su J. The pathogenesis of sepsis and potential therapeutic targets. Int J Mol Sci 2019; 20: 5376.
[18] Pierrakos C and Vincent JL. Sepsis biomarkers: a review. Crit Care 2010; 14: R15.
[19] Wang L, Ko ER, Gilchrist JJ, Pittman KJ, Rautanen A, Pirinen M, Thompson JW, Dubois LG, Langley RJ, Jaslow SL, Salinas RE, Rouse DC, Moseley MA, Mwarumba S, Njuguna P and Mturi N; Wellcome Trust Case Control Consortium 2, Kenyan Bacteraemia Study Group; Williams TN, Scott JAG, Hill AV, Woods CW, Ginsburg GS, Tsalik EL and Ko DC. Human genetic and metabolite variation reveals that methylthioadenosine is a prognostic biomarker and an inflammatory regulator in sepsis. Sci Adv 2017; 3: e1602096.
[20] Kelly-Scumpia KM, Scumpia PO, Weinstein JS, Delano MJ, Cuenca AG, Nacionales DC, Wynn JL, Lee PY, Kumagai Y, Efron PA, Akira S, Wasserfall C, Atkinson MA and Moldawer LL. B cells enhance early innate immune responses during bacterial sepsis. J Exp Med 2011; 208: 1673-1682.
[21] Delano MJ and Ward PA. Sepsis-induced immune dysfunction: can immune therapies reduce mortality? J Clin Invest 2016; 126: 23-31.
[22] Gao K, Jin J, Huang C, Li J, Luo H, Li L, Huang Y and Jiang Y. Exosomes derived from septic mouse serum modulate immune responses via exosome-associated cytokines. Front Immunol 2019; 10: 1560.
[23] Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33: 1-22.
[24] Huang DW, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.
[25] Huang DW, Sherman BT and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37: 1-13.
[26] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.



Supplementary Figure 1. Hierarchical clustering of sepsis patients in discovery cohort.