

Article Bioinformatics Analysis of Gene Expression Profiles for Diagnosing Sepsis and Risk Prediction in Patients with Sepsis

Hayoung Choi^{1,†}, Jin Young Lee^{2,†}, Hongseok Yoo², and Kyeongman Jeon^{2,3,*}

- ¹ Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Internal Medicine, Hallym University Kangnam Sacred Heart Hospital, Hallym University College of Medicine, Seoul 07441, Republic of Korea; hychoimd@gmail.com
- ² Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Republic of Korea; yenayein@gmail.com (J.Y.L.); hongseok.yoo@samsung.com (H.Y.)
- ³ Department of Health Sciences and Technology, SAIHST, Sungkyunkawan University, Seoul 06351, Republic of Korea
- Correspondence: kjeon@skku.edu; Tel.: +82-2-3410-3429; Fax: +82-2-3410-6956
- + These authors contributed equally to this work.

Abstract: Although early recognition of sepsis is essential for timely treatment and can improve sepsis outcomes, no marker has demonstrated sufficient discriminatory power to diagnose sepsis. This study aimed to compare gene expression profiles between patients with sepsis and healthy volunteers to determine the accuracy of these profiles in diagnosing sepsis and to predict sepsis outcomes by combining bioinformatics data with molecular experiments and clinical information. We identified 422 differentially expressed genes (DEGs) between the sepsis and control groups, of which 93 immune-related DEGs were considered for further studies due to immune-related pathways being the most highly enriched. Key genes upregulated during sepsis, including S100A8, S100A9, and CR1, are responsible for cell cycle regulation and immune responses. Key downregulated genes, including CD79A, HLA-DQB2, PLD4, and CCR7, are responsible for immune responses. Furthermore, the key upregulated genes showed excellent to fair accuracy in diagnosing sepsis (area under the curve 0.747-0.931) and predicting in-hospital mortality (0.863-0.966) of patients with sepsis. In contrast, the key downregulated genes showed excellent accuracy in predicting mortality of patients with sepsis (0.918-0.961) but failed to effectively diagnosis sepsis. In conclusion, bioinformatics analysis identified key genes that may serve as biomarkers for diagnosing sepsis and predicting outcomes among patients with sepsis.

Keywords: sepsis; biomarkers; diagnosis; genes; bioinformatics

1. Introduction

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response [1]. Sepsis and septic shock are major healthcare problems that affect millions of patients worldwide each year and kill approximately 17–33% of those affected [2,3]. Unfortunately, the reported incidence of sepsis and its associated healthcare burden are both currently on the rise due to the global trend of population aging and patients having a greater number of comorbidities [4,5]. Biomarker development could serve as a cornerstone of sepsis management and may ameliorate the healthcare burden of sepsis because early recognition is essential for timely treatment and the improvement of sepsis outcomes [6].

Numerous biomarkers for sepsis, including C-reactive protein and procalcitonin, have been investigated previously [7,8]. However, to date no marker has demonstrated sufficient discriminatory power [9,10]. Bioinformatics approaches integrate computational and life sciences to screen molecular and clinical data via data mining, pathway analysis, statistical analysis, and visual processing. These methods can investigate disease on



Citation: Choi, H.; Lee, J.Y.; Yoo, H.; Jeon, K. Bioinformatics Analysis of Gene Expression Profiles for Diagnosing Sepsis and Risk Prediction in Patients with Sepsis. *Int. J. Mol. Sci.* 2023, 24, 9362. https:// doi.org/10.3390/ijms24119362

Academic Editors: Vasso Apostolopoulos, Jack Feehan and Vivek P. Chavda

Received: 10 March 2023 Revised: 19 May 2023 Accepted: 23 May 2023 Published: 27 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the molecular level and have been widely used to identify significant biomarkers for sepsis [11–15]. In addition to the previous studies using datasets downloaded from public repositories, some prospective cohort studies also performed bioinformatics analysis to explore potential biomarkers for sepsis diagnosis [16,17]. Although previous studies have provided important insight into biomarkers and the pathophysiology of sepsis, more bioinformatics studies are warranted to validate the study results in association with real-world clinical outcomes.

Thus, the present study aimed to (1) use bioinformatics analyses to explore key differentially expressed genes (DEGs) between patients with sepsis from a prospective cohort and healthy volunteers and (2) validate the accuracy of bioinformatics analyses for diagnosing sepsis and predicting sepsis outcomes by integrating molecular experiments with clinical information.

2. Results

2.1. Clinical Characteristics of Patients with Sepsis

During the study period, we enrolled 133 critically ill patients with sepsis after excluding 63 who were not diagnosed with sepsis and two who withdrew their consent. Of the 133 patients, 90 (67.7%) were male, and the median age of patients was 66 years (interquartile range (IQR), 58–73 years). When patients were admitted to intensive care units, 60 (45.1%) and 69 (51.9%) received mechanical ventilation therapy and vasopressors, respectively. With respect to the severity of illness, the median SAPS 3 score was 55 (IQR, 47–63), the APACHE II score was 24 (IQR, 20–30), and the initial Sequential Organ Failure Assessment score was 9 (IQR, 7–11). The rates of 28-day mortality and in-hospital mortality were 16.5% and 24.8%, respectively (Table 1).

Table 1. Characteristics of patients with sepsis.

Variable	Value (n = 133)
Age (years)	66 (58–73)
Male	90 (67.7)
Comorbidities	
All malignancies	51 (38.3)
Solid organ malignancies	35 (26.3)
Hematologic malignancies	16 (12.0)
Diabetes mellitus	41 (30.8)
Chronic obstructive pulmonary disease	16 (12.0)
Chronic kidney disease	10 (7.5)
Myocardial infarction	8 (6.0)
Congestive heart failure	7 (5.3)
Cerebrovascular disease	8 (6.0)
Chronic liver disease	11 (8.3)
Charlson Comorbidity Index	2 (1–3)
Clinical status on ICU admission	
Mechanical ventilation	60 (45.1)
Vasopressor support	69 (51.9)
Laboratory findings	
WBC (/ µL)	13,640 (5060–20,340)
Hemoglobin (g/dL)	10.1 (8.9–11.7)
Platelet (/ µL)	136,000 (59,000–214,000)
Albumin (g/dL)	2.9 (2.6–3.2)
CRP (mg/dL)	12.5 (5.8–24.5)
Lactate (mg/dL)	3.0 (2.0-4.4)
PaO_2/FiO_2 ratio	216 (135–323)
Severity of illness	
SAPS 3 score	55 (47–63)
APACHE II score	24 (20–30)
SOFA score, initial	9 (7–11)
Outcome	
28-day mortality	22 (16.5)
In-hospital mortality	33 (24.8)

Data are presented as count (percentage) or median (interquartile range). Abbreviations: APACHE II, Acute Physiology and Chronic Health Evaluation II; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; ICU, intensive care unit; PaO_2/FiO_2 ratio, ratio of arterial oxygen pressure to fractional inspired oxygen; SAPS 3, Simplified Acute Physiology Score 3; SOFA, Sequential Organ Failure Assessment.

This study initially identified 422 DEGs between 133 patients with sepsis and 12 healthy volunteers. A principal component analysis (PCA) of global gene expression profiles revealed that sepsis patients were clearly separate from healthy volunteers (Figure 1A). In addition, distinct patterns of gene expression existed in sepsis patients when transcriptomic profiles of sepsis patients were compared to those of healthy volunteers (Figure 1B). Further bioinformatics analyses were conducted to identify key genes related to sepsis. First, enriched gene ontology (GO) functional analysis revealed that the identified DEGs were mainly involved in the immune response (Figure 1C,D).



Figure 1. Cont.



Figure 1. Gene ontology (GO) enrichment analysis of 422 differentially expressed genes (DEGs) between patients with sepsis and healthy volunteers. (**A**) Principal component analysis of the RNA transcriptome from sepsis patients (green) and that from healthy volunteers (blue). (**B**) Volcano plot of DEGs showing upregulated genes in red and downregulated genes in green. (**C**) GO enrichment analysis visualizing main DEGs that are mainly involved in the response to sepsis. (**D**) Number of identified DEGs according to their biological process (red), cellular component (blue), or molecular function (green) categorization.

Second, a protein–protein interaction (PPI) network analysis of the 422 DEGs also showed that the most extensive module was composed of 78 seeds, 1381 nodes, and 1823 edges. Moreover, it appeared to be most strongly enriched in immune-related pathways (Figure 2).



Figure 2. Protein–protein interaction network analysis of all 422 differentially expressed genes between patients with sepsis and healthy volunteers. Immune-related pathways, presented as the largest module (78 seeds, 1381 nodes, and 1823 edges), show the greatest enrichment.

A following PPI network analysis of 93 immune-related DEGs revealed that these enriched immune-related pathways included adaptive immune response, positive regulation of immune response, positive regulation of leukocyte cell–cell adhesion, cell activation, and positive regulation of cytokine production (Figure 3).



Figure 3. Protein–protein interaction network analysis of 93 immune-related differentially expressed genes between patients with sepsis and healthy volunteers. Enriched pathways are also shown.

Results of a molecular complex detection (MCODE) analysis performed to screen the significant modules of the PPI network are shown in Figure 4.





GO	Description	Log10(P)	
GO:0050778	positive regulation of immune response	-36.9	
GO:0002250	adaptive immune response	-33.9	
GO:1903039	positive regulation of leukocyte cell-cell adhesion	-30.7	

Figure 4. Cont.



Color	MCODE	GO	Description	Log10(P)
	MCODE 1	R-HSA-202430	Translocation of ZAP-70 to Immunological synapse	-32.2
	MCODE 1	R-HSA-202427	Phosphorylation of CD3 and TCR zeta chains	-31.4
	MCODE 1	R-HSA-202433	Generation of second messenger molecules	-29.1
	MCODE 2	GO:0051238	sequestering of metal ion	-8.5
	MCODE 2	GO:0042063	gliogenesis	-7.7
	MCODE_2	GO:2001244	positive regulation of intrinsic apoptotic signaling pathway	-7.1
	MCODE 3	GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	-9.1
	MCODE 3	GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	-9
	MCODE_3	GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	-8

Figure 4. Molecular complex detection analysis results used to screen the significant modules identified in the protein–protein interaction network.

The 93 immune-related DEGs identified here included 51 upregulated and 42 downregulated genes in patients with sepsis compared to healthy volunteers. Table 2 summarizes the top 10 upregulated and downregulated genes in patients with sepsis. Significantly upregulated genes included S100A8, VNN1, HMGB2, and S100A9, whereas significantly downregulated genes included CD79A, HLA-DQB2, PLD4, and CCR7.

Table 2. Top 10 upregulated and downregulated immune-related genes in patients with sepsis relative to healthy volunteers.

Top 10 Upregulated Genes			Top 10 Downregulated Genes			
Entrez ID	Gene Symbol	Fold Change	Entrez ID	Gene Symbol	Fold Change	
6279	S100A8	16.17	973	CD79A	0.12	
8876	VNN1	7.83	3120	HLA-DQB2	0.13	
3148	HMGB2	7.23	122618	PLD4	0.14	
6280	S100A9	5.43	1236	CCR7	0.15	
301	ANXA1	5.42	259197	NCR3	0.17	
665	BNIP3L	4.67	29802	VPREB3	0.18	
200315	APOBEC3A	4.61	6932	TCF7	0.199	
353514	LILRA5	4.49	933	CD22	0.201	
1378	CR1	4.46	10578	GNLY	0.206	
1604	CD55	4.29	974	CD79B	0.213	

Abbreviations: S100A8, S100 calcium-binding protein A8; VNN1, vanin 1; HMGB2, high-mobility group box 2; S100A9, S100 calcium-binding protein A9; ANXA1, annexin A1; BNIP3L, BCL2 interacting protein 3-like; APOBEC3A, apolipoprotein B mRNA editing enzyme catalytic subunit 3A; LILRA5, leukocyte immunoglobulinlike receptor A5; CR1, complement C3b/C4b receptor 1; CD55, CD55 molecule; CD79A, CD79a molecule; HLA-DQB2, major histocompatibility complex, class II, DQ beta 2; PLD4, phospholipase D family member 4; CCR7, C-C motif chemokine receptor 7; NCR3, natural cytotoxicity triggering receptor 3; VPREB3, V-set pre-B cell surrogate chain 3; TCF7, transcription factor 7; CD22, CD22 molecule; GNLY, granulysin; CD79B, CD79b molecule.

2.3. Experimental and Clinical Validation of Potential Biomarkers from the mRNA Profile

Next, we used quantitative real-time PCR (qPCR) to measure the expression levels of identified DEGs and thereby validate our bioinformatics analyses using molecular data

in 133 patients with sepsis and 12 healthy volunteers. Among the 10 upregulated and 10 downregulated genes, 3 upregulated and 4 downregulated genes showed consistent results with bioinformatics analysis. The three upregulated genes, S100A8, S100A9, and CR1, showed significantly higher expression levels in patients with sepsis than in healthy volunteers (p < 0.001 for S100A8 and S100A9, and p = 0.005 for CR1), whereas the four downregulated genes, CD79A, HLA-DQB2, PLD4, and CCR7, showed significantly lower expression levels in patients with sepsis than in healthy volunteers (p < 0.001 for all four genes) (Table 3).

Table 3. Comparisons of the expression levels of differentially expressed genes between patients with sepsis and healthy volunteers, and between patients who died and survived.

Patients with Sepsis vs. Healthy Volunteers *						
		Patients with Sepsis	Healthy Volunteers	<i>p</i> -Value		
Upregulated genes	S100A8	19.8 (11.6–24.2)	7.6 (7.1–8.2)	< 0.001		
1 0 0	S100A9	7.6 (3.8–10.3)	3.8 (2.6–4.2)	< 0.001		
	CR1	5.5 (2.5-8.8)	2.7 (2.3–2.8)	0.005		
Downregulated genes	CD79A	0.3 (0.2–0.4)	0.5 (0.5–0.6)	< 0.001		
0 0	HLA-DQB2	0.2 (0.2–0.2)	0.3 (0.3–0.4)	< 0.001		
	PLD4	0.5 (0.2–0.7)	0.9 (0.9–1.0)	< 0.001		
	CCR7	0.3 (0.1–0.4)	0.5 (0.5–0.6)	< 0.001		
The Dead vs. Surviving among Patients with Sepsis *						
		Patients (Dead)	Patients (Surviving)	<i>p</i> -Value		
Upregulated genes	S100A8	25.9 (24.1–27.8)	17.1 (11.4–21.2)	< 0.001		
1 0 0	S100A9	10.7 (9.7–11.8)	6.3 (3.2–8.5)	< 0.001		
	CR1	10.3 (9.3–11.9)	4.0 (2.0-6.6)	< 0.001		
Downregulated genes	CD79A	0.5 (0.4–0.6)	0.2 (0.1–0.3)	< 0.001		
0 0	HLA-DQB2	0.3 (0.2–0.3)	0.2 (0.1–0.2)	< 0.001		
	PLD4	0.7 (0.6–0.8)	0.4 (0.2–0.6)	< 0.001		
	CCR7	0.4 (0.4–0.5)	0.2 (0.1–0.3)	< 0.001		

Data are presented as median (interquartile range). * Values denote fold changes of genes.

A receiver operating characteristic (ROC) curve analysis was then conducted to determine the accuracy of these seven genes in diagnosing sepsis: S100A8 showed very high accuracy (i.e., area under the curve (AUC): 0.931, 95% confidence interval (CI): 0.880–0.982), while both S100A9 (AUC: 0.791, 95% CI: 0.711–0.871) and CR1 (AUC: 0.747, 95% CI: 0.647–0.820) showed fair accuracy. In contrast to the three upregulated genes, all four downregulated genes failed to effectively discriminate between patients with sepsis and healthy volunteers (Table 4).

Further analyses were then performed to compare gene expression in patients with sepsis who died and survived among the 133 with sepsis. All three upregulated genes (S100A8, S100A9, and CR1) and all four downregulated genes (CD79A, HLA-DQB2, PLD4, and CCR7) showed significantly higher expression levels in the dead than in the survivors among patients with sepsis (p < 0.001 for all) (Table 3). The ROC curve analysis was then used to assess the accuracy of the seven genes in predicting in-hospital mortality among patients with sepsis. All upregulated and downregulated genes demonstrated excellent accuracy, with CR1 showing the highest accuracy (AUC: 0.966, 95% CI: 0.939–0.993), followed by CD79A (AUC: 0.961, 95% CI: 0.930–0.992) and HLA-DQB2 (AUC: 0.936, 95% CI: 0.895–0.978) (Table 4).

	Sepsis Diagnosis					
mRNAs	AUC	SE	95% CI	Optimal Cut-Off	Sensitivity	Specificity
S100A8	0.931	0.026	0.880-0.982	≥9.0	94.0%	83.3%
S100A9	0.791	0.041	0.711-0.871	\geq 5.0	69.2%	100%
CR1	0.747	0.037	0.674-0.820	\geq 3.0	72.9%	100%
CD79A	0.113	0.028	0.059-0.168	NA	NA	NA
HLA-DQB2	0.003	0.003	0-0.009	NA	NA	NA
PLD4	0.013	0.013	0-0.038	NA	NA	NA
CCR7	0.016	0.010	0-0.036	NA	NA	NA
	Mortality Prediction in Sepsis					
	AUC	SE	95% CI	Optimal Cut-Off	Sensitivity	Specificity
S100A8	0.919	0.026	0.868-0.970	≥23.6	84.9%	88.0%
S100A9	0.863	0.033	0.797-0.928	≥ 9.6	78.8%	85.0%
CR1	0.966	0.014	0.939-0.993	≥ 9.0	84.9%	95.0%
CD79A	0.961	0.016	0.930-0.992	≥ 0.42	93.9%	92.0%
HLA-DQB2	0.936	0.021	0.895-0.978	≥ 0.23	84.9%	88.0%
PLD4	0.918	0.026	0.866-0.969	≥ 0.64	84.9%	88.0%
CCR7	0.918	0.026	0.866-0.969	≥ 0.39	78.8%	90.0%

Table 4. Receiver operating characteristic curve analysis and suggested optimal cut-off values of seven mRNAs in diagnosing sepsis and predicting in-hospital mortality among patients with sepsis.

Abbreviations: AUC, area under curve; SE, standard error; CI, confidence interval; NA, not applicable.

3. Discussion

In the present study, bioinformatics analysis revealed that immune-related pathways were strongly enriched in patients with sepsis relative to healthy volunteers. In addition, we identified three key genes that were upregulated in patients with sepsis (S100A8, S100A9, and CR1) as well as four key genes that were downregulated (CD79A, HLA-DQB2, PLD4, and CCR7). Furthermore, validation of these findings by molecular experiments and clinical outcomes determined that the key upregulated genes showed excellent to fair accuracy for both diagnosing sepsis and predicting in-hospital mortality of patients with sepsis. In contrast, the key downregulated genes showed excellent accuracy in predicting in-hospital mortality of patients with sepsis but failed to effectively diagnose sepsis.

Immune-related pathways were the most highly enriched biological pathways in patients with sepsis relative to healthy volunteers; this finding is consistent with previous bioinformatics studies, which also revealed DEGs were significantly enriched in pathways related to neutrophil activation, the TNF signaling pathway, and cytokine secretion [11,13–16,18]. These results suggest that the pathophysiology of sepsis is driven by "an aberrant or dysregulated host response to infection", which is also reflected in the current definition of sepsis [1]. Furthermore, a recent RNA sequencing study provided more insights into the pathophysiology of sepsis. The study compared RNA sequencing results between 18 immunocompromised patients with sepsis and 18 Sequential Organ Failure Assessment score-matched immunocompetent controls, and it demonstrated that patients with sepsis were more likely to show compromised T cell function, decreased T cell diversity, and altered metabolic signaling than controls [19]. In this way, bioinformatics studies will increasingly contribute to unveiling the pathophysiology of sepsis.

Notably, the key genes upregulated during sepsis are responsible for regulating cell cycle progression and differentiation (i.e., S100A8 and S100A9) and immune responses (i.e., ANXA1, APOBEC3A, LILRA5, CR1, and CD55) [20]. In agreement with our findings, one Chinese prospective cohort study also performed a comprehensive transcriptome profile analysis and qPCR validation, and suggested S100A8, S100A9, and ANXA3 as key genes differentially expressed between sepsis patients and healthy controls [16]. Potential underlying mechanisms of upregulated S100A8 and S100A9 may also include

altering MyD88-dependent gene programs, which consequently prevent hyperinflammatory responses without impairing pathogen defense [21], and mediating endotoxin-induced cardiomyocyte dysfunction [16,22]. Our qPCR results and ROC curve analyses confirmed the accuracy of S100A8, S100A9, and CR1 in diagnosing sepsis as well as in predicting in-hospital mortality among patients with sepsis. Thus, our results suggest that the expression levels of the key upregulated genes identified here may serve as potential biomarkers of sepsis.

In this study, the key genes downregulated during sepsis also encoded proteins responsible for the immune response, including the expression of antigen-presenting cells, the regulation of cytokine production, and the activation of B and T lymphocytes (i.e., CD79A, HLA-DQB2, PLD4, and CCR7) [20]. This is in line with prior studies showing that genes involving Jak-STAT signaling, T cell receptor signaling, and natural killer cell-mediated pathways were downregulated [23–25]. The results showed that genes participating in the immune response have mixed differential expression patterns of both up- and downregulation. Considering sepsis has been found to manifest a balance between competing proand anti-inflammatory pathways [26], the downregulation of immune response genes in patients with sepsis implies the homeostatic regulation of immunity during sepsis. Thus, the failed homeostatic regulation of immunity may have consequently resulted in the development and progression of sepsis. Interestingly, the key downregulated genes revealed an excellent accuracy in predicting in-hospital mortality of patients with sepsis but failed to diagnose sepsis in the ROC curve analyses. Taken together, key downregulated genes helped us understand more about sepsis pathophysiology; however, they may not be useful as sepsis biomarkers compared with key upregulated genes.

This study's most important strength is the validation of key DEGs between patients with sepsis and healthy volunteers via molecular experiments and clinical information from a prospective cohort. These validation methods can elucidate which genes may act as biomarkers of the diagnosis and mortality prediction of sepsis. Nonetheless, two limitations to this study should also be acknowledged. First, the study population was relatively small; thus, future bioinformatics studies, including a larger sample size, are necessary to confirm our findings. Second, this study was conducted in Korea, which might limit the generalizability of our results to other countries or ethnic groups.

4. Materials and Methods

4.1. Study Population

This study included patients with sepsis from the Samsung Medical Center Registry of Critical Illness (SMC-RoCI), a prospective observational study conducted at the Samsung Medical Center (i.e., a 1989-bed, university-affiliated, tertiary referral hospital in Seoul, Republic of Korea) between October 2015 and January 2020 as previously described [27]. Sepsis was defined according to the third edition of the International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [1]. Consequently, patients enrolled before the release of this new definition were reclassified according to the Sepsis-3 scheme.

In addition to patients with sepsis, we used a control consisting of 12 healthy volunteers (\geq 19 years of age) who donated blood specimens for research purposes. Written informed consent was obtained from all participants or their legally authorized representatives before enrollment. This study was conducted according to the Declaration of Helsinki, and all experimental procedures were approved by the institutional review board of the Samsung Medical Center (Application No. 2013-12-033).

4.2. Sample Collection

Blood samples consisted of 19 mL of whole blood collected into ethylenediaminetetraacetic acid tubes within 48 h of enrollment in the SMC-ROCI. Samples were centrifuged at $480 \times g$ (Eppendorf Centrifuge 5810 No. 0012529-rotor A-4-81) for 10 min at 4 °C within 4 h of collection. Several plasma aliquots from each study participant were then isolated and stored at -80 °C for further analysis.

4.3. Total RNA Isolation and Quality Analysis

For RNA isolation, whole blood (2 mL) was also collected in PAXgene tubes, using BD PAXgene blood RNA tubes (BD, cat. no. 762165). Total RNA was isolated from whole blood using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol [28]. RNA quantity and purity were measured using a NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA). RNA quality, yield, and distribution were determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) [29]. Blood samples (5 mL) were also collected from healthy volunteers; these samples were also prepared using the method described above.

4.4. Library Preparation and Sequencing

Libraries were prepared from total RNA using a NEBNext Ultra II Directional RNA-Seq Kit (New England BioLabs, UK). Messenger RNA (mRNA) was isolated using a Poly(A) RNA Selection Kit (Lexogen, Inc., Vienna, Austria). Isolated mRNA was then used for cDNA synthesis and shearing by following the manufacturer's instructions. Indexing was performed using Illumina indices 1–12. Enrichment was performed by polymerase chain reaction (PCR). Subsequently, libraries were checked using an Agilent 2100 Bioanalyzer (DNA High Sensitivity Kit) to evaluate the mean fragment size. Quantification was performed using a library quantification kit on a StepOne Real-Time PCR System (Applied Biosystems Life Technologies, Carlsbad, CA, USA). High-throughput sequencing was performed as paired-end 100 bp sequencing on a NovaSeq 6000 sequencing platform (Illumina, Inc., San Diego, CA, USA) [29,30].

4.5. Data Analysis

Quality control of the raw sequencing data was performed using FastQC [31]. Adapter sequences and low-quality reads (<Q20) were removed using the fastx_clipper function implemented in the FASTX_Toolkit and by BBMap [32]. Trimmed reads were then mapped to the reference genome using TopHat [33]. Gene expression levels were estimated as fragments per kilobase of transcript per million (FPKM) mapped reads values as determined by Cufflinks [34]. All FPKM values were normalized based on the quantile normalization method implemented by the EdgeR package for R [35].

4.6. Identification of DEGs

GEO2R was used to screen for DEGs between the sepsis and control groups. GEO2R is an R-based interactive web tool that helps to identify and visualize differential gene expression [36]. PCA of the different groups' samples was performed on the gene expression matrix. We set the threshold of differential expression to the default standard (i.e., $|\log 2$ (fold change [FC])| > 1 and adjusted p < 0.05) to identify significant DEGs between the two groups. Thus, significantly upregulated DEGs showed log2 FC > 1, and significantly downregulated DEGs showed log2 FC < 1 [37]. Significance was defined as an adjusted p value < 0.05 to control for type I errors in multiple tests.

4.7. Functional and Pathway Enrichment Analyses

To further recognize the underlying biological functions of the DEGs identified in the previous step, we performed GO functional analysis to annotate all DEGs according to the three main GO categories: molecular function, cellular component, and biological process [38]. Furthermore, to further elucidate the DEG pathways, we performed a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis [39]. For GO functional annotation and KEGG pathway enrichment analyses, we used the WEB-based Gene SeT AnaLysis Toolkit (WebGestalt), the web-based Database for Annotation, Visualization, and Integrated Discovery (DAVID) tool version 6.8 [37,40], and Metascape [41].

4.8. Protein–Protein Interaction Network Analysis

Because proteins function in a coordinated manner within a complicated and dynamic network, we constructed a PPI network of the target genes using the Search Tool for the Retrieval of Interacting Genes (STRING) database, version 11.0 [42]. In addition, the Cytoscape plug-in Network Analyzer was also used for further analyses. Furthermore, the topological properties of the PPI network, including node degree, were calculated by searching for hub genes using the PPI network [43]. MCODE analysis implemented in Cytoscape was then performed to screen for significant modules of the PPI network using the following cut-off parameters: node score cut-off = 0.2, K-core = 2, and degree cut-off = 2.

4.9. Quantitative Real-Time PCR

To evaluate the expression levels of key genes, we performed qPCR analyses in duplicate. The reaction conditions were as follows: an initial step of 50 °C for 2 min, denaturing at 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s and 58.5 °C for 1 min. qPCR was performed on an ABI ViiA 7 Real-Time PCR System (Applied Biosystems) and was followed by a melting curve analysis. Glyceraldehyde-3-phosphate dehydrogenase was selected as an internal control. The $2^{-\Delta\Delta CT}$ algorithm (Δ CT = Ct. target – Ct. reference) was employed for downstream data analysis [44].

4.10. Statistical Analysis

Categorical variables were compared using the chi-square or Fisher's exact tests. Continuous variables were compared using Mann–Whitney *U* tests. For clinical validation of bioinformatics analysis results, an ROC curve was used to analyze the diagnostic accuracy of mRNA expression for (1) discriminating between patients with sepsis and healthy volunteers and (2) predicting in-hospital mortality among patients with sepsis. The sensitivity and specificity were also calculated to suggest the optimal cut-off value of each gene. All tests were two-tailed, and p < 0.05 was used as the threshold of statistical significance. Data were analyzed using STATA version 16 (Stata Corp., College Station, TX, USA).

5. Conclusions

Bioinformatics analysis revealed that immune-related pathways were the most enriched in patients with sepsis relative to healthy volunteers. In addition, we identified key genes that were upregulated in sepsis, namely, S100A8, S100A9, and CR1, as well as those that were downregulated, namely, CD79A, HLA-DQB2, PLD4, and CCR7. The key upregulated genes showed excellent to fair accuracy in diagnosing sepsis and predicting in-hospital sepsis mortality; however, the key downregulated genes showed excellent accuracy in predicting in-hospital sepsis mortality but failed to effectively diagnose sepsis.

Author Contributions: Conceptualization, K.J.; methodology, H.C., J.Y.L., H.Y. and K.J.; software, J.Y.L.; validation, H.C. and H.Y.; formal analysis, H.C., J.Y.L. and K.J.; investigation, H.C. and J.Y.L.; resources, J.Y.L. and K.J.; data curation, H.C., J.Y.L., H.Y. and K.J.; writing—original draft preparation, H.C., J.Y.L. and K.J.; writing—review and editing, H.C., J.Y.L., H.Y. and K.J.; visualization, H.C. and J.Y.L.; supervision, K.J.; project administration, K.J.; funding acquisition, K.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Future Medicine 20*30 Project of the Samsung Medical Center (SMX1230441), a National Research Foundation of Korea grant funded by the Korean Government (MSIT) (NRF-2022R1F1A1074855), and a grant from the Korean Academy of Tuberculosis and Respiratory Disease (2022).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and was approved by the Institutional Review Board of the Samsung Medical Center (IRB no. 2013-12-033).

Informed Consent Statement: Written informed consent was obtained from all patients or their legally authorized representatives prior to enrollment.

Data Availability Statement: The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE232753 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE232753).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Singer, M.; Deutschman, C.S.; Seymour, C.W.; Shankar-Hari, M.; Annane, D.; Bauer, M.; Bellomo, R.; Bernard, G.R.; Chiche, J.D.; Coopersmith, C.M.; et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016, 315, 801–810. [CrossRef]
- Fleischmann, C.; Scherag, A.; Adhikari, N.K.; Hartog, C.S.; Tsaganos, T.; Schlattmann, P.; Angus, D.C.; Reinhart, K.; International Forum of Acute Care Trialists. Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. Am. J. Respir. Crit. Care Med. 2016, 193, 259–272. [CrossRef] [PubMed]
- Fleischmann-Struzek, C.; Mellhammar, L.; Rose, N.; Cassini, A.; Rudd, K.E.; Schlattmann, P.; Allegranzi, B.; Reinhart, K. Incidence and mortality of hospital- and ICU-treated sepsis: Results from an updated and expanded systematic review and meta-analysis. *Intensive Care Med.* 2020, 46, 1552–1562. [CrossRef] [PubMed]
- 4. Gaieski, D.F.; Edwards, J.M.; Kallan, M.J.; Carr, B.G. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit. Care Med.* **2013**, *41*, 1167–1174. [CrossRef]
- Iwashyna, T.J.; Cooke, C.R.; Wunsch, H.; Kahn, J.M. Population burden of long-term survivorship after severe sepsis in older Americans. J. Am. Geriatr. Soc. 2012, 60, 1070–1077. [CrossRef] [PubMed]
- Seymour, C.W.; Gesten, F.; Prescott, H.C.; Friedrich, M.E.; Iwashyna, T.J.; Phillips, G.S.; Lemeshow, S.; Osborn, T.; Terry, K.M.; Levy, M.M. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N. Engl. J. Med. 2017, 376, 2235–2244. [CrossRef]
- 7. Simon, L.; Saint-Louis, P.; Amre, D.K.; Lacroix, J.; Gauvin, F. Procalcitonin and C-reactive protein as markers of bacterial infection in critically ill children at onset of systemic inflammatory response syndrome. *Pediatr. Crit. Care Med.* 2008, *9*, 407–413. [CrossRef]
- 8. Yang, Y.; Xie, J.; Guo, F.; Longhini, F.; Gao, Z.; Huang, Y.; Qiu, H. Combination of C-reactive protein, procalcitonin and sepsis-related organ failure score for the diagnosis of sepsis in critical patients. *Ann. Intensive Care* **2016**, *6*, 51. [CrossRef]
- 9. Pierrakos, C.; Vincent, J.L. Sepsis biomarkers: A review. Crit. Care 2010, 14, R15. [CrossRef]
- 10. Yoo, H.; Lee, J.Y.; Park, J.; Yang, J.H.; Suh, G.Y.; Jeon, K. Association of Plasma Level of TNF-Related Apoptosis-Inducing Ligand with Severity and Outcome of Sepsis. *J. Clin. Med.* **2020**, *9*, 1661. [CrossRef]
- 11. Hu, Y.; Cheng, L.; Zhong, W.; Chen, M.; Zhang, Q. Bioinformatics Analysis of Gene Expression Profiles for Risk Prediction in Patients with Septic Shock. *Med. Sci. Monit.* 2019, 25, 9563–9571. [CrossRef] [PubMed]
- 12. Li, Z.; Huang, B.; Yi, W.; Wang, F.; Wei, S.; Yan, H.; Qin, P.; Zou, D.; Wei, R.; Chen, N. Identification of Potential Early Diagnostic Biomarkers of Sepsis. J. Inflamm. Res. 2021, 14, 621–631. [CrossRef]
- 13. Niu, J.; Qin, B.; Wang, C.; Chen, C.; Yang, J.; Shao, H. Identification of Key Immune-Related Genes in the Progression of Septic Shock. *Front. Genet.* **2021**, *12*, 668527. [CrossRef] [PubMed]
- 14. She, H.; Tan, L.; Zhou, Y.; Zhu, Y.; Ma, C.; Wu, Y.; Du, Y.; Liu, L.; Hu, Y.; Mao, Q.; et al. The Landscape of Featured Metabolism-Related Genes and Imbalanced Immune Cell Subsets in Sepsis. *Front. Genet.* **2022**, *13*, 821275. [CrossRef] [PubMed]
- 15. Zeng, X.; Feng, J.; Yang, Y.; Zhao, R.; Yu, Q.; Qin, H.; Wei, L.; Ji, P.; Li, H.; Wu, Z.; et al. Screening of Key Genes of Sepsis and Septic Shock Using Bioinformatics Analysis. *J. Inflamm. Res.* **2021**, *14*, 829–841. [CrossRef]
- Wu, T.; Liang, X.; Jiang, Y.; Chen, Q.; Zhang, H.; Zhang, S.; Zhang, C.; Lv, Y.; Xin, J.; Jiang, J.; et al. Comprehensive Transcriptome Profiling of Peripheral Blood Mononuclear Cells from Patients with Sepsis. *Int. J. Med. Sci.* 2020, 17, 2077–2086. [CrossRef] [PubMed]
- 17. Herwanto, V.; Tang, B.; Wang, Y.; Shojaei, M.; Nalos, M.; Shetty, A.; Lai, K.; McLean, A.S.; Schughart, K. Blood transcriptome analysis of patients with uncomplicated bacterial infection and sepsis. *BMC Res. Notes* **2021**, *14*, 76. [CrossRef] [PubMed]
- 18. Zhai, J.; Qi, A.; Zhang, Y.; Jiao, L.; Liu, Y.; Shou, S. Bioinformatics Analysis for Multiple Gene Expression Profiles in Sepsis. *Med. Sci. Monit.* 2020, *26*, e920818. [CrossRef] [PubMed]
- 19. Cheng, P.L.; Chen, H.H.; Jiang, Y.H.; Hsiao, T.H.; Wang, C.Y.; Wu, C.L.; Ko, T.M.; Chao, W.C. Using RNA-Seq to Investigate Immune-Metabolism Features in Immunocompromised Patients With Sepsis. *Front. Med.* **2021**, *8*, 747263. [CrossRef]
- National Library of Medicine (US), N.C.f.B.I. Gene. Available online: https://www.ncbi.nlm.nih.gov/gene/ (accessed on 31 January 2023).
- Ulas, T.; Pirr, S.; Fehlhaber, B.; Bickes, M.S.; Loof, T.G.; Vogl, T.; Mellinger, L.; Heinemann, A.S.; Burgmann, J.; Schöning, J.; et al. S100-alarmin-induced innate immune programming protects newborn infants from sepsis. *Nat. Immunol.* 2017, 18, 622–632. [CrossRef]
- 22. Boyd, J.H.; Kan, B.; Roberts, H.; Wang, Y.; Walley, K.R. S100A8 and S100A9 mediate endotoxin-induced cardiomyocyte dysfunction via the receptor for advanced glycation end products. *Circ. Res.* **2008**, *102*, 1239–1246. [CrossRef] [PubMed]
- 23. Forel, J.M.; Chiche, L.; Thomas, G.; Mancini, J.; Farnarier, C.; Cognet, C.; Guervilly, C.; Daumas, A.; Vély, F.; Xéridat, F.; et al. Phenotype and functions of natural killer cells in critically-ill septic patients. *PLoS ONE* **2012**, *7*, e50446. [CrossRef]

- 24. Lv, X.; Zhang, Y.; Cui, Y.; Ren, Y.; Li, R.; Rong, Q. Inhibition of microRNA-155 relieves sepsis-induced liver injury through inactivating the JAK/STAT pathway. *Mol. Med. Rep.* **2015**, *12*, 6013–6018. [CrossRef]
- 25. Winkler, M.S.; Rissiek, A.; Priefler, M.; Schwedhelm, E.; Robbe, L.; Bauer, A.; Zahrte, C.; Zoellner, C.; Kluge, S.; Nierhaus, A. Human leucocyte antigen (HLA-DR) gene expression is reduced in sepsis and correlates with impaired TNFα response: A diagnostic tool for immunosuppression? *PLoS ONE* 2017, *12*, e0182427. [CrossRef]
- Nedeva, C.; Menassa, J.; Puthalakath, H. Sepsis: Inflammation Is a Necessary Evil. Front. Cell Dev. Biol. 2019, 7, 108. [CrossRef] [PubMed]
- Choi, H.; Yoo, H.; Lee, J.Y.; Park, J.; Jeon, K. Plasma Mitochondrial DNA and Necroptosis as Prognostic Indicators in Critically III Patients with Sepsis. *Biomedicines* 2022, 10, 2386. [CrossRef]
- 28. Li, D.; Ren, W.; Wang, X.; Wang, F.; Gao, Y.; Ning, Q.; Han, Y.; Song, T.; Lu, S. A modified method using TRIzol reagent and liquid nitrogen produces high-quality RNA from rat pancreas. *Appl. Biochem. Biotechnol.* **2009**, *158*, 253–261. [CrossRef]
- Sp, N.; Kang, D.Y.; Jo, E.S.; Rugamba, A.; Kim, W.S.; Park, Y.M.; Hwang, D.Y.; Yoo, J.S.; Liu, Q.; Jang, K.J.; et al. Tannic Acid Promotes TRAIL-Induced Extrinsic Apoptosis by Regulating Mitochondrial ROS in Human Embryonic Carcinoma Cells. *Cells* 2020, 9, 282. [CrossRef] [PubMed]
- Kukurba, K.R.; Montgomery, S.B. RNA Sequencing and Analysis. Cold Spring Harb. Protoc 2015, 2015, pdb-top084970. [CrossRef] [PubMed]
- 31. Wingett, S.W.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **2018**, *7*, 1338. [CrossRef] [PubMed]
- 32. Bushnell, B.; Rood, J.; Singer, E. BBMerge–Accurate paired shotgun read merging via overlap. *PLoS ONE* **2017**, *12*, e0185056. [CrossRef]
- Trapnell, C.; Pachter, L.; Salzberg, S.L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 2009, 25, 1105–1111. [CrossRef] [PubMed]
- Roberts, A.; Trapnell, C.; Donaghey, J.; Rinn, J.L.; Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* 2011, 12, R22. [CrossRef] [PubMed]
- 35. R Core Team. *R: A Language and Environment for Statistical Computing;* R Foundation for Statistical Computing: Vienna, Austria, 2013. Available online: http://www.R-project.org (accessed on 5 January 2023).
- 36. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2012**, *41*, D991–D995. [CrossRef]
- Udhaya Kumar, S.; Thirumal Kumar, D.; Bithia, R.; Sankar, S.; Magesh, R.; Sidenna, M.; George Priya Doss, C.; Zayed, H. Analysis of Differentially Expressed Genes and Molecular Pathways in Familial Hypercholesterolemia Involved in Atherosclerosis: A Systematic and Bioinformatics Approach. *Front. Genet.* 2020, *11*, 734. [CrossRef] [PubMed]
- Harris, M.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004, *32*, D258–D261.
- Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009, 37, 1–13. [CrossRef]
- 40. Liao, Y.; Wang, J.; Jaehnig, E.J.; Shi, Z.; Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 2019, 47, W199–W205. [CrossRef]
- 41. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523. [CrossRef]
- Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019, 47, D607–D613. [CrossRef]
- Wang, J.; Zhong, J.; Chen, G.; Li, M.; Wu, F.X.; Pan, Y. ClusterViz: A Cytoscape APP for Cluster Analysis of Biological Network. IEEE/ACM Trans. Comput. Biol. Bioinform. 2015, 12, 815–822. [CrossRef] [PubMed]
- Maru, D.M.; Singh, R.R.; Hannah, C.; Albarracin, C.T.; Li, Y.X.; Abraham, R.; Romans, A.M.; Yao, H.; Luthra, M.G.; Anandasabapathy, S.; et al. MicroRNA-196a is a potential marker of progression during Barrett's metaplasia-dysplasia-invasive adenocarcinoma sequence in esophagus. *Am. J. Pathol.* 2009, 174, 1940–1948. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.