Differential gene expression analysis reveals novel genes and pathways in pediatric septic shock patients

Akram Mohammed[1], Yan Cui[1], Valeria R. Mas[1], Rishikesan Kamaleswaran[1*]

[1] University of Tennessee Health Science Center, Memphis, TN, USA

[*] Corresponding Author
Rishikesan Kamaleswaran, Ph.D.
50 N. Dunlap St, 4th Floor 491R,
Memphis, 38103, TN, USA
E: rkamales@uthsc.edu

Abstract

Septic shock is a severe health condition caused by uncontrolled sepsis. Advancements in the high-throughput sequencing techniques have risen the number of potential genetic biomarkers under review. Multiple genetic markers and functional pathways play a part in the development and progression of pediatric septic shock. Fifty-four differentially expressed pediatric septic shock gene biomarkers were identified using gene expression data from 181 pediatric intensive care unit (PICU) patients within the first 24 hours of admission. The gene expression signatures discovered showed discriminatory power between pediatric septic shock survivors and nonsurvivors types. Using functional enrichment analysis of differentially expressed genes (DEGs), the known genes and pathways in septic shock were validated, and unexplored septic shock-related genes and functional groups were identified. Septic shock survivors were distinguished from septic shock non-survivors by differential expression of genes involved in the immune response, chemokine-mediated signaling, neutrophil chemotaxis, and chemokine activity. The identification of the septic shock gene biomarkers may facilitate in septic shock diagnosis, treatment, and prognosis.

Introduction

Septic shock is a life-threatening organ dysfunction caused by imbalanced host response to infection [1]. Multi-omics sequencing technologies have increased the number of genetic biomarkers [2]. Single or combination biomarkers are increasingly being analyzed and tested in the context of genes, RNA, or proteins [3–6]. Many strategies for uncovering biomarkers exist, such as mass-spectrometry-based, protein arrays and gene-expression profiling. Furthermore, it has been demonstrated that multiple genes and immune system-related pathways participate in the development of pediatric septic shock [7].

High-throughput technologies have enabled analysis of the expression of a number of genes and determine the activity of these genes in different conditions [8]. Statistical testing and machine learning methods have been developed to successfully utilize the omics data for biomarker discovery [2,9–18].

The purpose of this study is to identify differentially expressed pediatric septic shock biomarkers using gene expression data. To this end, gene expression data from 181 samples from PICU within the first 24 hours were analyzed using multiple statistical testing methods to identify gene biomarkers. The gene expression profiles discovered by this statistical approach may lead to new insights into genetic biomarkers for successful septic shock diagnosis [19]. Using functional gene-set enrichment analysis, we validated the known septic shock-related genes, pathways and functional groups, and identified the unexplored septic shock-related genes, and functional groups. The discovery of the potential gene biomarkers may provide effective septic shock diagnosis, treatment, and prognosis.

Results

Identification of Differentially Expressed Upregulated and Down-regulated Genes

Based on the preset criteria of an adjusted p-value < 0.05, a total of 54 genes from 21,731 were shown to be differentially expressed between the Septic Shock Survivor and Non-survivor

samples, including 47 genes that were up-regulated and 7 genes that were down-regulated. Sixteen DEGs with a fold change of at least 1.5 is shown in Table 1 (For the complete list, refer to Supplementary File 1).

Table 1: List of most significant up-regulated and down-regulated genes in septic shock

Gene	Fold Change	Average Expression	t-statistics	p-value	adj p-value	Reference
DDIT4	2.12	8.441545	5.490723	1.33E-07	0.000592	[20,21]
CCL3	2.12	6.208106	4.627029	7.02E-06	0.012717	[22]
PRG2	2.11	5.48492	5.611777	7.35E-08	0.000533	-
MT1M	1.78	3.72997	5.641576	6.35E-08	0.000533	-
CDC20	1.68	6.18643	4.059008	7.31E-05	0.040893	[23]
KIF20A	1.66	4.940242	4.673723	5.74E-06	0.012717	-
MAFF	1.64	6.815799	4.437113	1.58E-05	0.015559	[24]
EBI3	1.64	5.41593	4.517058	1.12E-05	0.013575	[25]
MELK	1.63	6.413027	4.141523	5.27E-05	0.034718	-
TOP2A	1.58	4.997892	4.045113	7.72E-05	0.040893	[26]
NUSAP1	1.54	6.627514	3.928091	0.000121	0.049772	[27]
RGL1	1.52	7.288761	4.447937	1.51E-05	0.015559	[28]
ARHGEF40	-1.66	6.880613	-3.994983	9.38E-05	0.043368	-
LOC254896	-1.65	8.498379	-4.452352	1.48E-05	0.015558	-
SLC46A2	-1.61	5.880150	-4.084894	6.60E-05	0.039833	-
TNFRSF10C	-1.54	8.678817	-4.643272	6.55E-06	0.012717	[29]

Functional Enrichment Analysis of Differentially Expressed Genes

54 DEGs were analyzed by KEGG pathway and Gene Ontology (GO) term enrichment. A total of 52 genes were recognized in the DAVID database. KEGG pathway analysis revealed rheumatoid arthritis (RA) (hsa: 05323) and cell cycle (has: 04110) pathways as the most significant pathways (Table 2). GO analyses of the DEGs demonstrated that mitotic sister chromatid segregation (GO: 0000070), immune response (GO: 0006955), cell division (GO: 0051301), and chemokine-mediated signaling pathway (GO: 0070098) were the most enriched biological process (BP) terms (Table 2). ‘Chemokine activity (GO: 0008009) was the most enriched term under molecular function (Table 2). Chemokine interleukin-8-like domain (IPR001811), CC chemokine, conserved site (IPR000827) InterPro protein functional groups

were among the significantly enriched functional classes associated with septic shock development (Table 2).

Table 2: Functional enrichment of differentially expressed genes

Functional Category	ID	Functional Term	Gene Count	adjusted p-value	Fold Change	Benjamini Score
BP	GO:0000070	Mitotic sister chromatid segregation	4	4.82E-05	54.83	0.0246
BP	GO:0006955	Immune response	8	1.80E-04	6.51	0.0453
BP	GO:0051301	Cell division	7	4.61E-04	6.85	0.0763
BP	GO:0070098	Chemokine-mediated signaling	4	0.001093	19.30	0.0776
BP	GO:0007059	Chromosome segregation	4	9.64E-04	20.15	0.0797
BP	GO:0007067	Mitotic nuclear division	6	6.88E-04	8.29	0.0851
BP	GO:0030593	Neutrophil chemotaxis	4	8.84E-04	20.76	0.0873
MF	GO:0008009	Chemokine activity	4	3.63E-04	28.12	0.0467
KEGG pathway	hsa05323	Rheumatoid arthritis	5	2.58E-04	15.03	0.0226
KEGG pathway	hsa04110	Cell cycle	5	9.49E-04	10.66	0.0413
InterPro	IPR001811	Chemokine interleukin-8-like domain	4	2.01E-04	34.33	0.0247
InterPro	IPR000827	CC chemokine, conserved site	3	0.001602	49.35	0.0953

BP: Biological Process; MF: Molecular Function; CC: Cellular Component

Discussion

This study of peripheral blood mRNA sequences revealed key genes and functional characterization associated with septic shock survivor and septic shock non-survival[30]. From the differential gene expression analysis, we identified the potential septic shock biomarkers that may help in an unbiased sepsis diagnosis, effective treatment, and ultimately improving prognoses. DEGs analysis using septic shock samples provides insights into the functional characterization of the genes between groups of septic shock survivor and non-survivor samples. However, like any other microarray data analysis is incomplete without performing adjustment for multiple testing. Due to approximately 20,000 (the approximate number of genes on a standard microarray chip) independent tests, it is expected to get at least 20 test scores by random chance when we allow a stricter p-value threshold of say 0.001. To avoid this situation, adjustment for multiple testing was utilized, and we have used the Benjamini Hochberg method.

For a false-discovery rate (FDR) controlling procedure, the adjusted p-value of an individual hypothesis is the minimum value of FDR for which the hypothesis is first included in the set of rejected hypotheses, and we used an adjusted p-value cut-off of 0.05[31].

We identified *CDC20* as one of the top up-regulated genes, along with *LCN2*, and *CD24*, similar to the findings of Dong et al., 2018 [23], which studied the development of trauma-induced sepsis in patients. However, our study population was much more diverse (Table 1). The most significantly up-regulated gene identified was *DDIT4* (DNA damage-inducible transcript 4-like). PERSEVERE-XP study had also identified *DDIT4* gene directly related to *TP53* [7]. *DDIT4* (*REDD1)* is increased in the septic shock and can negatively regulate mTORC1 activity and plays an important role in energy homeostasis [21]. We found *CCL3* as the second-most significantly up-regulated chemokine, a fundamental component of the acute-phase response to endotoxin in humans and regulates the leukocyte activation and trafficking [32]. Elevated levels of *CCL3* has been detected within the first 24 hours of sepsis, suggesting its unique role in innate immune function [22,33]. Further studies are needed to understand the mechanisms of these identified genes in the septic shock development. On the other hand, *TNFRSF10C*, a down-regulated gene has been shown to play an essential role in the sepsis immune response [34].

The set of genes identified is then examined for over-representation of specific functions or pathways. Septic shock survivors were distinguished from septic shock non-survivor by differential expression of genes involved in the immune response, chemokine-mediated signaling, neutrophil chemotaxis, and chemokine activity. Sepsis impacts the immune responses by directly altering the life span, production, and function of effector cells responsible for homeostasis [35]. We identified the immune response (BP GO:0006955) term from DAVID

analysis, and there has also been evidence that understanding the immune response to sepsis provides opportunities to develop effective treatment strategies [36].

Chemokines play a critical role in the sepsis and septic shock development, and molecules that block chemokine and chemokine receptor activity may prove to be useful in the identification of sepsis. [37]. Our differentially expressed genes mapped to chemokine-mediated signaling (GO:0070098), chemokine activity (GO:0008009) molecular function, chemokine interleukin-8 like domain (IPR001811) and chemokine conserved site (IPR000827).

Sepsis and Rheumatoid Arthritis (RA) have been known to be associated for over 50 years [38]. RA is shown to be a risk factor in sepsis patients, and sepsis infection could trigger the RA [39]. We identified RA KEGG pathway (hsa:05323) using our differentially expressed gene set to be statistically significant (Table 2).

The gene expression changes shown in our results are based on the peripheral blood cells and may not be extrapolated as occurring at the organ or tissue level [30,40]. Therefore, extra care must be taken while generalizing host immune responses or chemokine activities in septic shock patients. Besides, variations in the gene expression profiles of survivors and non-survivors of septic shock patients could be due to other unexplored confounding factors (such as patients demographics) rather than sepsis-related biology [41]. On the other hand, the blood-based biomarkers have the advantage to be minimally-invasive. Large cohorts replication studies and network analysis studies are needed to gain insights into the relationships between these biomarkers and the survival/non-survival of cohorts [42]. To avoid the possibility of selection bias the analysis must be expanded to the other independent data sets.

This work can be expanded by experimentally validating the identified blood-based biomarkers, developing robust machine learning methods to build septic shock prediction model using different omics data from diversified patient cohorts.

Materials and Methods

Data collection

Expression microarray data was collected from the NCBI Gene Expression Omnibus repository [43]. The dataset contains the gene expression profiles of the peripheral blood samples from 181 septic shock patients including 154 survivors and 27 non-survivors, who were admitted to the pediatric intensive care unit within the first 24 hours [44]. The GEO accession number for the data used in the study is GSE66099. The data was collected from the Affymetrix Human Genome HG-U133_Plus_2 (GPL570 platform).

Normalization and Background Correction

The R Affy module [45] was used to remove the technical variations and background noise. The Quantile Normalization Method [46] was used to normalize the data, and the background correction was performed using the Robust Multi-Average [47] parameter method[48].

Probe to Gene Mapping

Affymetrix probes were mapped to the genes using the information provided in the Affymetrix database (hgu133plus2.db). We used average expression values when multiple probes mapped to the same gene [19].

Identification of Differentially Expressed Genes

Differentially Expressed Genes (up-regulated and down-regulated genes) were identified using R limma package with a Benjamini-Hochberg (BH) correction method and the adjusted p-value of < 0.05 was used.

Functional Analysis

We used DAVID [49] for functional enrichment analysis of the DEGs from samples of septic shock survivor or non-survivor. The biological process (BP), cellular component (CC), and molecular function (MF) were identified from the Gene Ontology database. For the GO functional groups, KEGG pathways, and InterPro functional terms returned from DAVID functional analysis; we considered an adjusted p-value threshold of ≤ 0.05 and gene count of 3 or more from this study.

Statistical analysis

R programming language [50] is used for downloading the Affymetrix data and gene mapping using R Affy, and Bioconductor package. A Fisher-exact test was performed for determining statistical significance among the gene ontology terms and functional classes. Benjamini Hochberg multiple test correction method was used for calculating the differentially expressed genes.

Data availability

The R scripts other related files used for data preprocessing, normalization and differential gene expression analysis are available from https://github.com/akram-mohammed/septic_shock_degs. The datasets generated and analyzed during the study are available upon request.

Acknowledgments

We would like to acknowledge Timothy E. Sweeney and Hector R. Wang for providing the data.

Author contributions

RK and AM conceived and designed the study, developed the method and performed the analysis. YC, VRM contributed to the analysis. All authors wrote and proofread the manuscript.

Competing interests

The authors declare no competing interest.

References

1. Delano, M. J. & Ward, P. A. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol. Rev.* **274,** 330–353 (2016).
2. Zuo, Y. *et al.* INDEED: Integrated differential expression and differential network analysis of omic data for biomarker discovery. *Methods* **111,** 12–20 (2016).
3. Fung, K. Y. C. *et al.* Blood-Based Protein Biomarker Panel for the Detection of Colorectal Cancer. *PLoS One* **10,** e0120425 (2015).
4. Tang, Q., Cheng, J., Cao, X., Surowy, H. & Burwinkel, B. Blood-based DNA methylation as biomarker for breast cancer: a systematic review. *Clin. Epigenetics* **8,** 115 (2016).
5. Birse, C. E. *et al.* Blood-based lung cancer biomarkers identified through proteomic discovery in cancer tissues, cell lines and conditioned medium. *Clin. Proteomics* **12,** 18 (2015).
6. Yörüker, E. E., Holdenrieder, S. & Gezer, U. Blood-based biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *Clin. Chim. Acta* **455,** 26–32 (2016).
7. Wong, H. R. *et al.* Improved risk stratification in pediatric septic shock using both protein and mRNA Biomarkers: Persevere-XP. *Am. J. Respir. Crit. Care Med.* **196,** 494–501 (2017).
8. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46,** 389–422 (2002).
9. Akay, M. F. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Syst. Appl.* **36,** 3240–3247 (2009).
10. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6,** pl1-pl1 (2013).
11. Aliferis, C. F., Tsamardinos, I., Mansion, P., Statnikov, A. & Hardin, D. Machine learning models for lung cancer classification using array comparative genomic hybridization. *16th Int. FLAIRS Conf.* 67–71 (2002).
12. Liu, J. J. *et al.* Multiclass cancer classification and biomarker discovery using GA-based algorithms. *Bioinformatics* **21,** 2691–2697 (2005).
13. Pirooznia, M., Yang, J. Y., Yang, M. Q. & Deng, Y. A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics* **9,** S13 (2008).
14. Mao, Y., Zhou, X., Pi, D., Sun, Y. & Wong, S. T. C. Multiclass cancer classification by using fuzzy support vector machine and binary decision tree with gene selection. *J. Biomed. Biotechnol.* **2005,** 160–171 (2005).
15. Peng, Y. A novel ensemble machine learning for robust microarray data classification. *Comput. Biol. Med.* **36,** 553–573 (2006).
16. Duan, K. B., Rajapakse, J. C., Wang, H. & Azuaje, F. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans. Nanobioscience* **4,** 228–233 (2005).
17. Kallio, M. A. *et al.* Chipster: User-friendly analysis software for microarray and other high-throughput data. *BMC Genomics* **12,** 507 (2011).
18. Kolesnikov, N. *et al.* ArrayExpress update-simplifying data submissions. *Nucleic Acids Res.* **43,** D1113–D1116 (2015).
19. Mohammed, A., Biegert, G., Adamec, J. & Helikar, T. Identification of potential tissue-specific cancer biomarkers and development of cancer versus normal genomic classifiers. *Oncotarget* **8,** 85692–85715 (2017).
20. Sweeney, T. E. *et al.* A community approach to mortality prediction in sepsis via gene

expression analysis. *Nat. Commun.* **9,** 694 (2018).
21. Gordon, B. S., Steiner, J. L., Williamson, D. L., Lang, C. H. & Kimball, S. R. Emerging role for regulated in development and DNA damage 1 (REDD1) in the regulation of skeletal muscle metabolism. *Am. J. Physiol. - Endocrinol. Metab.* **311,** E157–E174 (2016).
22. Dapunt, U., Maurer, S., Giese, T., Gaida, M. M. & Hänsch, G. M. The macrophage inflammatory proteins MIP1 (CCL3) and MIP2 (CXCL2) in implant-associated osteomyelitis: Linking inflammation to bone degradation. *Mediators Inflamm.* **2014,** (2014).
23. Dong, L., Li, H., Zhang, S. & Su, L. Identification of genes related to consecutive trauma-induced sepsis via gene expression profiling analysis. *Med. (United States)* **97,** (2018).
24. Wong, H. R. *et al.* Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. *Physiol. Genomics* **30,** 146–155 (2007).
25. Wirtz, S. *et al.* Protection from lethal septic peritonitis by neutralizing the biological function of interleukin 27. *J. Exp. Med.* **203,** 1875–1881 (2006).
26. Wang, M. *et al.* Candidate genes and pathogenesis investigation for sepsis-related acute respiratory distress syndrome based on gene expression profile. *Biol. Res.* **49,** 25 (2016).
27. Yang, J., Zhang, P. & Wang, L. Gene Network for Identifying the Entropy Changes of Different Modules in Pediatric Sepsis. *Cell. Physiol. Biochem.* **40,** 1153–1162 (2016).
28. Zaba, L. C. *et al.* Effective treatment of psoriasis with etanercept is linked to suppression of IL-17 signaling, not immediate response TNF genes. *J. Allergy Clin. Immunol.* **124,** (2009).
29. Dickinson, P. *et al.* Whole blood gene expression profiling of neonates with confirmed bacterial sepsis. *Genomics Data* **3,** 41–48 (2015).
30. Tsalik, E. L. *et al.* An integrated transcriptome and expressed variant analysis of sepsis survival and death. *Genome Med.* **6,** 111 (2014).
31. Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19,** 368–375 (2003).
32. O'Grady, N. P. *et al.* Detection of Macrophage Inflammatory Protein (MIP)□1α and MIP□1β during Experimental Endotoxemia and Human Sepsis. *J. Infect. Dis.* **179,** 136–141 (1999).
33. Tsujimoto, Y. & Shimizu, S. Another way to die: Autophagic programmed cell death. *Cell Death and Differentiation* **12,** 1528–1534 (2005).
34. Smith, C. L. *et al.* Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat. Commun.* **5,** (2014).
35. Delano, M. J. & Ward, P. A. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol. Rev.* **274,** 330–353 (2016).
36. Luan, Y., Yao, Y., Xiao, X. & Sheng, Z. Insights into the Apoptotic Death of Immune Cells in Sepsis. *J. Interf. Cytokine Res.* **35,** 17–22 (2015).
37. Murdoch, C. & Finn, A. The role of chemokines in sepsis and septic shock. *Contrib. Microbiol.* **10,** 38–57 (2003).
38. Baghai, M. *et al.* Fatal Sepsis in a Patient With Rheumatoid Arthritis Treated With Etanercept. *Mayo Clin. Proc.* **76,** 653–656 (2001).
39. Barrett, O., Abramovich, E., Dreiher, J., Novack, V. & Abu-Shakra, M. Short- and long-term mortality due to sepsis in patients with rheumatoid arthritis. *Rheumatol. Int.* **37,** 1021–1026 (2017).

40. Langley, R. J. *et al.* Sepsis: An integrated clinico-metabolomic model improves prediction of death in sepsis. *Sci. Transl. Med.* **5,** 195ra95-195ra95 (2013).
41. Iwashyna, T. J., Netzer, G., Langa, K. M. & Cigolle, C. Spurious inferences about long-term outcomes: The case of severe sepsis and geriatric conditions. *Am. J. Respir. Crit. Care Med.* **185,** 835–841 (2012).
42. Tsalik, E. L. *et al.* An integrated transcriptome and expressed variant analysis of sepsis survival and death. *Genome Med.* **6,** 111 (2014).
43. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–210 (2002).
44. St. John, M. A. R. *et al.* Interleukin 6 and Interleukin 8 as Potential Biomarkers for Oral Cavity and Oropharyngeal Squamous Cell Carcinoma. *Arch. Otolaryngol. Neck Surg.* **130,** 929 (2004).
45. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20,** 307–315 (2004).
46. Bolstad, B. M., Irizarry, R. A., Åstrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19,** 185–193 (2003).
47. Tai, Y. C. & Speed, T. P. A multivariate empirical Bayes statistic for replicated microarray time course data. *Sel. Work. Terry Speed* **4,** 617–642 (2012).
48. Mohammed, A., Biegert, G., Adamec, J. & Helikar, T. CancerDiscover: an integrative pipeline for cancer biomarker and cancer class prediction from high-throughput sequencing data. *Oncotarget* **9,** 2565–2573 (2018).
49. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4,** R60 (2003).
50. Snyder, R. G. Vibrational spectra of crystalline n-paraffins. II. Intermolecular effects. *J. Mol. Spectrosc.* **7,** 116–144 (1961).