

OPEN

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

Akram Mohammed , Yan Cui, Valeria R. Mas & Rishikesan Kamaleswaran 

Septic shock is a devastating health condition caused by uncontrolled sepsis. Advancements in high-throughput sequencing techniques have increased the number of potential genetic biomarkers under review. Multiple genetic markers and functional pathways play a part in development and progression of pediatric septic shock. We identified 53 differentially expressed pediatric septic shock biomarkers using gene expression data sampled from 181 patients admitted to the pediatric intensive care unit within the first 24 hours of their admission. The gene expression signatures showed discriminatory power between pediatric septic shock survivors and nonsurvivor types. Using functional enrichment analysis of differentially expressed genes, we validated the known genes and pathways in septic shock and identified the unexplored septic shock-related genes and functional groups. Differential gene expression analysis revealed the genes involved in the immune response, chemokine-mediated signaling, neutrophil chemotaxis, and chemokine activity and distinguished the septic shock survivor from non-survivor. The identification of the septic shock gene biomarkers may facilitate in septic shock diagnosis, treatment, and prognosis.

Septic shock is a life-threatening organ dysfunction caused by an imbalanced host response to infection¹. Multi-omics sequencing technologies have increased the number of genetic biomarkers². 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, protein arrays, and gene-expression profiling. Furthermore, it has been demonstrated that multiple genes and immune system-related pathways participate in development of pediatric septic shock⁷.

High-throughput technologies support the analysis of gene expressions and also enable the ability to determine activity of these genes in different conditions⁸. Statistical testing and machine learning methods have been frequently used to successfully utilize 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 predict long-term outcomes. To that end, gene expression data from 181 samples collected from critically ill patients admitted to the pediatric intensive care unit (PICU) within the first 24 hours of their admission, 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 that support successful prognosis, especially among patients with poor long term outcomes¹⁹. Furthermore, these approaches can augment parallel efforts using clinical markers for earlier identification of sepsis and septic shock, providing meaningful input at the point of care^{20–22}. 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. Discovery of potential gene biomarkers may provide effective septic shock diagnosis, treatment, and prognosis.

Results

The pediatric septic shock dataset contains gene expression profiles of peripheral blood samples collected from 181 patients including 154 survivors and 27 non-survivors within the first 24 hours of admission. Overall mean age of the patient cohort was 3.69 years with a standard deviation of 3.28. Mean age of the survivors was 3.78 years (standard deviation of 3.15), compared to 3.19 (standard deviation of 3.93) for the non-survivors. Males (109;

University of Tennessee Health Science Center, Memphis, TN, USA. Correspondence and requests for materials should be addressed to R.K. (email: rkamales@uthsc.edu)

Etiology	Number of patients	Total (%)
<i>Streptococcus pneumoniae</i>	17	9.39
<i>Staphylococcus aureus</i>	17	9.39
<i>Streptococcus pyogenes</i>	12	6.63
<i>Klebsiella</i>	9	4.97
<i>Streptococcus agalactiae</i>	7	3.87
<i>Neisseria meningitidis</i>	7	3.87
<i>Enterococcus</i>	5	2.76
Gram-negative rods	4	2.21
Simplexvirus	3	1.66
influenza A virus	3	1.66
Mixed	3	1.66
<i>Enterobacter</i>	2	1.11
<i>Hemophilus</i>	2	1.11
<i>Pseudomonas</i>	2	1.11
<i>Serratia</i>	2	1.11
<i>Acinetobacter</i>	1	0.55
<i>Adenovirus</i>	1	0.55
BK virus	1	0.55
<i>Candida</i>	1	0.55
<i>Clostridium</i>	1	0.55
<i>Cytomegalovirus</i>	1	0.55
<i>Escherichia coli</i>	1	0.55
Group F <i>Streptococcus</i>	1	0.55
<i>Human metapneumovirus</i>	1	0.55
<i>Moraxella</i>	1	0.55
<i>Parainfluenza</i>	1	0.55
<i>Rickettsia</i>	1	0.55
Unknown	74	40.88

Table 1. Etiology of septic shock in 181 patients.

60.2%) had higher representation than females (72; 39.8%). Of the 27 non-survivors, 19 were male and only 8 were female; whereas, in the non-survivor group the male and females were 90 and 64, respectively. Clinical characteristics (such as mortality, organism, source of infection, malignancies, immunosuppression, gram-positive bacteria, and acute kidney injury) of the 181 patients are included in Supplementary Dataset 1.

Etiologic diagnosis was identified for 107 (59.12%) patients. As shown in Table 1, the most frequently identified pathogens were *Streptococcus pneumoniae* (in 17 cases), *Staphylococcus aureus* (in 17 cases), and *Streptococcus pyogenes* (in 12 cases). Detailed clinical characteristics including source of infection and organism are given in Supplementary Dataset 1.

Identification of differentially expressed upregulated and downregulated genes. Based on pre-set criteria of an adjusted p-value < 0.05, a total of 53 genes from 21,731 were shown to be differentially expressed genes (DEG) between Survivor and Non-survivor types, including 47 genes that were upregulated and 7 genes that were downregulated. Those DEGs with a fold change of at least 1.5 (n = 16) is shown in Table 2 (For complete list, refer to Supplementary Dataset 1).

Functional enrichment analysis of differentially expressed genes. 53 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 3). 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 3). Chemokine activity (GO: 0008009) was the most enriched term under molecular function (MF) (Table 3). 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 3).

Figure 1 illustrates a heatmap of 53 selected DEGs in survivor and non-survivor groups. The vertical axis represents patient samples (see Supplementary Dataset 1 for more details), and the horizontal axis represents DEGs with the log₂ gene expression intensity values. Red and green represent the upregulated, and downregulated genes, respectively. The heatmap is clustered using both samples and genes categories. Right vertical bars are used for annotation of mortality, pathogen, gram-positive, immunosuppression, gender status, and infection source.

Gene	Fold Change	Average Expression	t-statistics	p-value	adj. p-value	Reference
<i>DDIT4</i>	2.12	8.441545	5.490723	1.33E-07	0.000592	^{26,49}
<i>CCL3</i>	2.12	6.208106	4.627029	7.02E-06	0.012717	²⁹
<i>PRG2</i>	2.11	5.48492	5.611777	7.35E-08	0.000533	—
<i>MT1M</i>	1.78	3.72997	5.641576	6.35E-08	0.000533	—
<i>CDC20</i>	1.68	6.18643	4.059008	7.31E-05	0.040893	²⁵
<i>KIF20A</i>	1.66	4.940242	4.673723	5.74E-06	0.012717	—
<i>MAFF</i>	1.64	6.815799	4.437113	1.58E-05	0.015559	⁵⁰
<i>EBI3</i>	1.64	5.41593	4.517058	1.12E-05	0.013575	⁵¹
<i>MELK</i>	1.63	6.413027	4.141523	5.27E-05	0.034718	—
<i>TOP2A</i>	1.58	4.997892	4.045113	7.72E-05	0.040893	⁵²
<i>NUSAP1</i>	1.54	6.627514	3.928091	0.000121	0.049772	⁵³
<i>RGL1</i>	1.52	7.288761	4.447937	1.51E-05	0.015559	⁵⁴
<i>ARHGEF40</i>	-1.66	6.880613	-3.994983	9.38E-05	0.043368	—
<i>LOC254896</i>	-1.65	8.498379	-4.452352	1.48E-05	0.015558	—
<i>SLC46A2</i>	-1.61	5.880150	-4.084894	6.60E-05	0.039833	—
<i>TNFRSF10C</i>	-1.54	8.678817	-4.643272	6.55E-06	0.012717	⁵⁵

Table 2. List of most significant upregulated and downregulated genes in septic shock.

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

Table 3. Functional enrichment of differentially expressed genes. BP: Biological Process; MF: Molecular Function; CC: Cellular Component.

There are two major sample clusters at the top level of the dendrogram. Only 4 (15%) out of 27 non-survivors were in the first cluster, while, the remaining 85% formed the second cluster. Similarly, for the gram-positive category, samples formed two clusters. We observed no clear trends or cluster formation for the other annotation groups. Genes were clustered into two major groups of downregulated genes (*MT1M* to *CFH* except for *DDIT4* and *RGL1*), while the remaining genes clustered in the upregulated category. Overall, *DDIT4*, *NDUFV2*, and *TNFRSF10C* were the most highly expressed genes, whereas *AREG*, *CCL20*, and *CFH* were the least expressed genes.

Discussion

This study of peripheral blood mRNA sequences revealed key genes and functional characterization associated with septic shock survival and non-survival²³. From the differential gene expression analysis, we identified potential septic shock biomarkers that may help in an unbiased septic shock diagnosis and effective treatment, and ultimately improving prognoses.

Differentially expressed gene analysis using septic shock samples provides insights into the functional characterization of 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 (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, for example 0.001. To avoid this situation, adjustment for multiple testing was utilized, and we used the *Benjamini Hochberg* method. For a false-discovery rate (FDR) controlling procedure, 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 0.05²⁴.

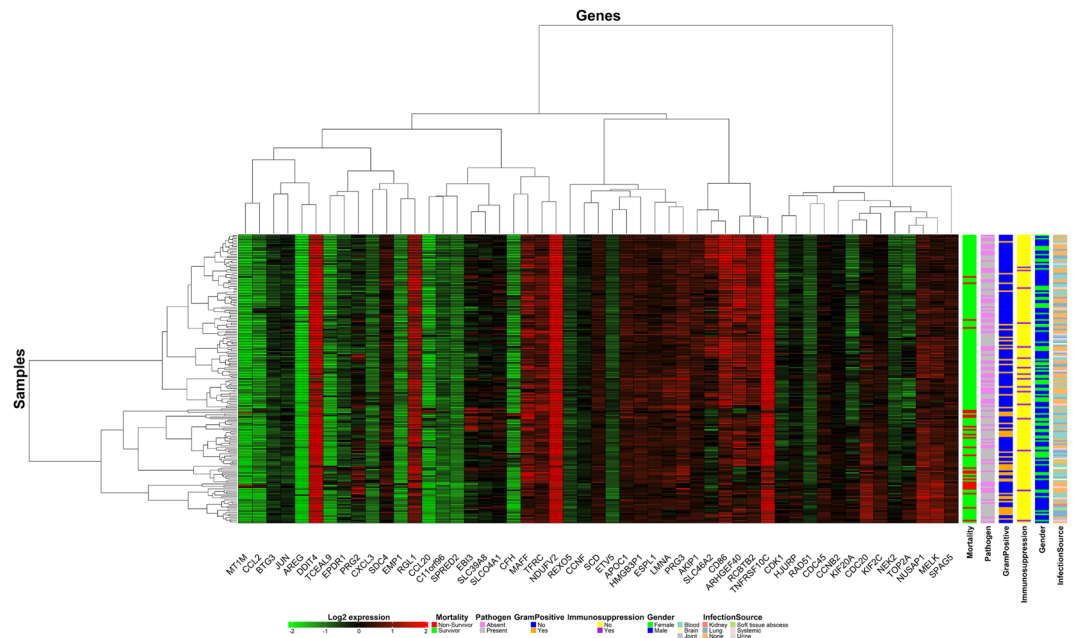


Figure 1. Heatmap representing differentially expressed genes between survivor and non-survivor groups, with pathogen, gram-positive, immunosuppression, gender status, and infection source as annotation.

We identified *CDC20* as one of the top upregulated genes, along with *LCN2*, and *CD24*, similar to the findings of Dong *et al.*²⁵, which studied development of trauma-induced sepsis in patients. However, our study population was much more diverse (Table 2 and Supplementary Dataset 1). The most significant upregulated 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 septic shock, can negatively regulate mTORC1 activity and plays an important role in energy homeostasis²⁶. We found *CCL3* as the second-most significant upregulated chemokine, a fundamental component of acute-phase response to endotoxin in humans and regulation of leukocyte activation and trafficking²⁷. Elevated levels of *CCL3* have been detected within the first 24 hours of sepsis, suggesting its unique role in innate immune function^{28,29}. Further studies are needed to understand the mechanisms of these identified genes in septic shock development. On the other hand, *TNFRSF10C*, a downregulated gene has been shown to play an essential role in sepsis immune response³⁰.

The set of genes identified is then examined for over-representation of specific functions or pathways. Septic shock survivors were distinguished from non-survivors 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 life span, production, and function of effector cells responsible for homeostasis³¹. We identified the immune response (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 strategies³².

Chemokines play a critical role in sepsis and septic shock development, and molecules that block chemokine and chemokine receptor activity may prove to be useful in the identification of sepsis³³. 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 years³⁴. RA is shown to be a risk factor in sepsis patients and sepsis infection could trigger RA³⁵. We identified RA KEGG pathway (hsa:05323) using our differentially expressed gene set to be statistically significant (Table 3). We analyzed the literature to look for other underlying genetic factors responsible for septic shock. Five (*CCL3*, *CDC20*, *TNFRSF10C*, *EBI3*, *TOP2A*) of the 53 identified genes were already shown to be involved in sepsis, and the remaining 48 genes from our list were not shown to be involved in sepsis. These genes may contribute to the septic shock response. Table 1 and the Discussion section above explains it in further detail. The *SLC46A2* is downregulated in the septic shock non-survivor group (Fig. 1) and have been shown to be involved in T-cell homeostasis (GO:0043029). On the other hand, the *LCN2* was identified as the upregulated gene and recently proposed as a biomarker involved in acute kidney injury (AKI)³⁶. Overall, 22.6% (41/181) patients had AKI, whereas, in the non-survivors, there were 70% (19/27) patients who had AKI (see Supplementary Dataset 1 for more details).

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^{23,37}. Therefore, extra care must be taken while generalizing host immune responses or chemokine activities in septic shock patients. Besides, variations in gene expression profiles of survivors and non-survivors of septic shock patients could be due to other unexplored confounding factors (such as patient demographics) rather than sepsis-related biology³⁸. On the other hand, the blood-based

biomarkers have the advantage of being minimally-invasive. Therefore, in combination with other clinical phenotypic data, these minimally-invasive biomarkers may enable rapid recognition of poor long term outcomes³⁹. Large cohorts replication studies and network analysis studies are needed to gain insights into relationships between these biomarkers and the survival/non-survival of cohorts⁴⁰. To avoid the selection bias, analysis must be expanded to other independent data sets.

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

Materials and Methods

Sepsis and septic shock. Bacteremia or another pathogenic infection triggers a serious inflammatory response called sepsis, which typically includes an increased number of white blood cells, rapid heart rate and breathing rate, and fever¹. The response also affects many major organ failures. Sepsis that results in poor perfusion to end organs resulting in multiple organ failure is defined as a septic shock⁷.

Data collection. Expression microarray data was collected from the NCBI Gene Expression Omnibus repository⁴¹. The dataset contains gene expression profiles of peripheral blood samples from 181 septic shock patients including 154 survivors and 27 non-survivors who were admitted to the pediatric intensive care unit (PICU) within the first 24 hours⁴². For all 181 samples, blood was drawn on day 1 of the PICU stay. GEO accession number for data used in the study is GSE66099. Data was collected from the Affymetrix Human Genome HG-U133_Plus_2 (GPL570 platform).

Normalization and background correction. The R Affy module⁴³ was used to remove technical variations and background noise. The Quantile Normalization Method⁴⁴ was used to normalize data and background correction was performed using the Robust Multi-Average⁴⁵ parameter method⁴⁶.

Probe to gene mapping. Affymetrix probes were mapped to the genes using information provided in the Affymetrix database (hgu133plus2.db). We used average expression values when multiple probes mapped to the same gene⁴⁹.

Identification of differentially expressed genes. Differentially Expressed Genes (upregulated and downregulated genes) were identified using R limma package with a Benjamini-Hochberg (BH) correction method and adjusted p-value of <0.05.

Functional analysis. We used DAVID⁴⁷ for functional enrichment analysis of the DEGs from samples of septic shock survivors or non-survivors. 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 language⁴⁸ was 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 gene ontology terms and functional classes. Benjamini Hochberg multiple test correction method was used for calculating the differentially expressed genes. Heatmap of differentially expressed genes between survivor and non-survivor groups for top selected genes were plotted using the complexHeatmap function in R programming. Other variables (pathogen, gram-positive, immunosuppression, gender status, and infection source) were added to the heatmap for annotation.

Data Availability

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. Datasets generated and analyzed during the study are available upon request.

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. 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).
11. Liu, J. J. *et al.* Multiclass cancer classification and biomarker discovery using GA-based algorithms. *Bioinformatics* **21**, 2691–2697 (2005).
12. 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).
13. 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).
14. Peng, Y. A novel ensemble machine learning for robust microarray data classification. *Comput. Biol. Med.* **36**, 553–573 (2006).
15. 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).
16. Kallio, M. A. *et al.* Chipster: User-friendly analysis software for microarray and other high-throughput data. *BMC Genomics* **12**, 507 (2011).
17. Kolesnikov, N. *et al.* ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.* **43**, D1113–D1116 (2015).
18. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1–pl11 (2013).
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. Kamaleswaran, R. *et al.* Applying Artificial Intelligence to Identify Physiobiomarkers Predicting Severe Sepsis in the PICU. *Pediatr. Crit. Care Med.* **19**, e495–e503 (2018).
21. Van Wyk, F. *et al.* How much data should we collect? A case study in sepsis detection using deep learning. In *2017 IEEE Healthcare Innovations and Point of Care Technologies, HI-POCT 2017 2017–Decem.*, 109–112 (2017).
22. Sutton, J., Mahajan, R., Akbilgic, O. & Kamaleswaran, R. PhysOnline: An Online Feature Extraction and Machine Learning Pipeline for Real-Time Analysis of Streaming Physiological Data. *IEEE J. Biomed. Heal. Informatics*, <https://doi.org/10.1109/JBHI.2018.2832610> (2018).
23. Tsalik, E. L. *et al.* An integrated transcriptome and expressed variant analysis of sepsis survival and death. *Genome Med.* **6**, 111 (2014).
24. Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368–375 (2003).
25. 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).
26. 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).
27. O’Grady, N. P. *et al.* Detection of Macrophage Inflammatory Protein (MIP)-1 α and MIP-1 β during Experimental Endotoxemia and Human Sepsis. *J. Infect. Dis.* **179**, 136–141 (1999).
28. Tsujimoto, Y. & Shimizu, S. Another way to die: Autophagic programmed cell death. *Cell Death and Differentiation* **12**, 1528–1534 (2005).
29. 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).
30. Smith, C. L. *et al.* Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat. Commun.* **5** (2014).
31. 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).
32. 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).
33. Murdoch, C. & Finn, A. The role of chemokines in sepsis and septic shock. *Contrib. Microbiol.* **10**, 38–57 (2003).
34. Baghai, M. *et al.* Fatal Sepsis in a Patient With Rheumatoid Arthritis Treated With Etanercept. *Mayo Clin. Proc.* **76**, 653–656 (2001).
35. Barrett, O., Abramovich, E., Dreiherr, 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).
36. Kanda, J. *et al.* An AKI biomarker lipocalin 2 in the blood derives from the kidney in renal injury but from neutrophils in normal and infected conditions. *Clin. Exp. Nephrol.* **19**, 99–106 (2015).
37. Langley, R. J. *et al.* Sepsis: An integrated clinico-metabolomic model improves prediction of death in sepsis. *Sci. Transl. Med.* **5**, 195ra95–195ra95 (2013).
38. 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).
39. van Wyk, F. *et al.* A minimal set of physiobiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. *Int. J. Med. Inform.* **122**, 55–62 (2019).
40. Tsalik, E. L. *et al.* An integrated transcriptome and expressed variant analysis of sepsis survival and death. *Genome Med.* **6**, 111 (2014).
41. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
42. 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).
43. 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).
44. 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).
45. 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).
46. 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).
47. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, (R60) (2003).
48. Snyder, R. G. Vibrational spectra of crystalline n-paraffins. II. Intermolecular effects. *J. Mol. Spectrosc.* **7**, 116–144 (1961).
49. Sweeney, T. E. *et al.* A community approach to mortality prediction in sepsis via gene expression analysis. *Nat. Commun.* **9**, 694 (2018).
50. 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).
51. Wirtz, S. *et al.* Protection from lethal septic peritonitis by neutralizing the biological function of interleukin 27. *J. Exp. Med.* **203**, 1875–1881 (2006).
52. 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).

53. 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).
54. 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).
55. Dickinson, P. *et al.* Whole blood gene expression profiling of neonates with confirmed bacterial sepsis. *Genomics Data* **3**, 41–48 (2015).

Acknowledgements

We would like to acknowledge Timothy E. Sweeney and Hector R. Wang for providing the data. We would also like to thank Beth Hines for copy editing and proofreading the manuscript.

Author Contributions

R.K. and A.M. conceived and designed the study, developed the method and performed the analysis. Y.C. and V.R.M. contributed to the analysis. All authors wrote and proofread the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-47703-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019