Transcriptomic markers in pediatric septic shock prognosis: an integrative analysis of gene expression profiles

Qian Wang1, Jie Huang2, Xia Chen1, Jian Wang3 and Fang Fang3

1Anesthesiology Department, Children’s Hospital of Soochow University, Suzhou, China
2Department of Cardiology, Children’s Hospital of Soochow University, Suzhou, China
3Institute of Pediatric Research, Children’s Hospital of Soochow University, Suzhou, China

Abstract

The goal of this study was to identify potential transcriptomic markers in pediatric septic shock prognosis by an integrative analysis of multiple public microarray datasets. Using the R software and bioconductor packages, we performed a statistical analysis to identify differentially expressed (DE) genes in pediatric septic shock non-survivors, and further performed functional interpretation (enrichment analysis and co-expression network construction) and classification quality evaluation of the DE genes identified. Four microarray datasets (3 training datasets and 1 testing dataset, 252 pediatric patients with septic shock in total) were collected for the integrative analysis. A total of 32 DE genes (18 upregulated genes; 14 downregulated genes) were identified in pediatric septic shock non-survivors. Enrichment analysis revealed that those DE genes were strongly associated with acute inflammatory response to antigenic stimulus, response to yeast, and defense response to bacterium. A support vector machine classifier (non-survivors vs survivors) was also trained based on DE genes. In conclusion, the DE genes identified in this study are suggested as candidate transcriptomic markers for pediatric septic shock prognosis and provide novel insights into the progression of pediatric septic shock.

Key words: Pediatric; Septic shock; Transcriptomic markers; Co-expression network; Prognosis

Introduction

Sepsis is the world’s leading cause of death of children, and it represents a complex disease with dysregulated inflammatory responses and a high mortality rate (1). An immune response initiated through an invading pathogen cannot be controlled to restore homeostasis in sepsis, which generates a pathological syndrome with sustained excessive inflammation as well as immune suppression (2). Toll-like receptors, nuclear factor κB, and cytokines including tumor necrosis factor α, interleukins 1, 2, 6, 8, and neutrophils all take part in the generation of sepsis (3). Although there are common points between adult and pediatric sepsis, crucial differences do exist in pathophysiology, clinical presentation, as well as therapeutic approaches (4).

High-throughput transcriptomic data grow rapidly, enabling gene expression profiling and identification of prognostic targets in disease. Over the last decade, several studies have explored the prognosis of pediatric septic shock by transcriptional profiling using microarrays (5–8). Integrative analysis based on multiple transcriptomic datasets could help to discover robust prognostic candidates. Therefore, in this study, the gene expression patterns of pediatric septic shock patients (of survivors and non-survivors) were investigated using public microarray datasets. The differently expressed (DE) genes identified were then further interpreted by enrichment analysis, construction of co-expression networks, and receiver operating characteristic (ROC) curve analysis.

In this study, data pre-processing, identification of DE genes, ROC analysis, and support vector machine (SVM) model training were performed using R software (http://www.r-project.org/) and bioconductor packages (9). Enrichment analysis and construction of co-expression networks were also performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) (10,11) and Cytoscape (12) software, respectively.
Material and Methods

Microarray datasets search and selection

In this study, public microarray datasets were searched from the earliest dataset until Dec 3, 2018 using the keywords “sepsis”/“septic shock” in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) (13). Further selection of the datasets was performed for subsequent analyses. The selection criteria were: i) dataset using whole blood for gene expression analysis; ii) dataset with detailed gene expression data for both survivors and non-survivors; and iii) dataset with sample size not fewer than 30. Animal studies and adult studies were excluded.

Data collection from each eligible dataset was performed by two investigators independently. The data consisted of the sample size, GEO accession, sample source, raw gene expression data, and platform used. Final data collection was determined by checking between the two investigators.

Data analysis methods

Four Affymetrix datasets were included in this study (3 datasets for training and 1 dataset for testing). Raw data (CEL files) of the 3 training datasets were downloaded from the GEO database, merged, and pre-processed using the R package “affy”, based on the Robust Multichip Average (RMA) method (14). Hybridization probes were then mapped to genes (Entrez IDs) based on the platform table. Probes not mapping to genes and probes mapping to multiple genes were excluded. If two or more probes mapped to the same gene, we calculated the arithmetic mean of the probe values to represent the gene expression. Using the R package “limma” (15,16), DE genes in septic shock non-survivors compared with survivors were identified based on the following criteria: i) absolute log2 fold change (LFC) greater than 0.8; and ii) false discovery rate (FDR)-adjusted P-value less than 0.05.

Results

DE genes in pediatric septic shock non-survivors

A total of 8 studies were obtained through the original search, among which 4 adult studies were excluded. The remaining 4 microarray datasets (3 training datasets and 1 testing dataset all from the Affymetrix microarrays) (5–8) were used for subsequent analysis. Table 1 presents detailed information about the 4 datasets.

Table 1. Microarray datasets used in this analysis.

| Dataset | GEO accession | Sample size | Sample source | Platform | Dataset usage | Submission date |
|---------|---------------|-------------|---------------|----------|--------------|----------------|
| 1       | GSE9692 (5)   | 6           | 24            | Whole blood | Affymetrix Human Genome U133 Plus 2.0 Array | Training | 2007     |
| 2       | GSE26378 (6)  | 12          | 70            | Whole blood | Affymetrix Human Genome U133 Plus 2.0 Array | Training | 2010     |
| 3       | GSE26440 (7)  | 17          | 81            | Whole blood | Affymetrix Human Genome U133 Plus 2.0 Array | Training | 2011     |
| 4       | GSE4607 (8)   | 9           | 33            | Whole blood | Affymetrix Human Genome U133 Plus 2.0 Array | Testing | 2006     |

GEO: Gene Expression Omnibus.
Figure 1 shows the heat map of expression data for the 32 DE genes in the collection of the 3 training datasets. Among the 32 DE genes, 18 genes were up-regulated and 14 genes were downregulated. Table 2 presents the top 10 most significantly upregulated and top 10 most significantly downregulated genes. The top upregulated gene was olfactomedin 4 (OLFM4) (LFC=1.72). OLFM4 is a glycoprotein that functions in innate immunity, inflammation, and cancer (22). It was suggested to be a neutrophil subset marker in patients suffering from septic shock (23). The most downregulated gene was membrane metalloendopeptidase (MME) (LFC=−1.11). MME is widely present in organs and tissues, contributing to the inactivation of many biological peptides (24,25). Decreased MME expression capacity was detected in the neutrophils from patients with septic shock (25).

Figure 1. Heatmap of expression data for 32 differentially expressed genes in the collection of 3 training datasets.

Table 2. The 10 most significantly upregulated and most significantly downregulated genes in pediatric septic shock non-survivors.

| Entrez ID | Gene symbol | Gene full name | Log2 fold-change | False discovery rate-adjusted P-value |
|-----------|-------------|----------------|------------------|---------------------------------------|
| 10562     | OLFM4       | olfactomedin 4  | 1.717146732      | 0.029147425                          |
| 1088      | CEACAM8     | carcinoembryonic antigen-related cell adhesion molecule 8 | 1.358191894 | 0.038663019 |
| 84419     | C15orf48    | chromosome 15 open reading frame 48 | 1.283283183 | 0.002649534 |
| 1991      | ELANE       | elastase, neutrophil-expressed | 1.162470483 | 0.022391364 |
| 5657      | PRTN3       | proteinase 3    | 1.104928095      | 0.005347329                          |
| 54541     | DDIT4       | DNA damage inducible transcript 4 | 1.069842902 | 0.000011715 |
| 5553      | PRG2        | proteoglycan 2, pro eosinophil major basic protein | 1.01469606 | 0.000011715 |
| 100133941 | CD24        | CD24 molecule   | 1.014691224      | 0.001427259                          |
| 6241      | RRM2        | ribonucleotide reductase regulatory subunit M2 | 1.001189643 | 0.014703381 |
| 55165     | CEP55       | centrosomal protein 55 | 1.001180377 | 0.001823254 |
| 4311      | MME         | membrane metalloendopeptidase | −1.111678792 | 0.005267437 |
| 10875     | FGL2        | fibrinogen-like 2 | −1.099276289 | 0.000425566 |
| 8843      | HCAR3       | hydroxycarboxylic acid receptor 3 | −1.083432919 | 0.027926014 |
| 100134822 | LOC100134822| LOC100134822 | −1.065641306 | 0.003118448 |
| 3579      | CXCR2       | C-X-C motif chemokine receptor 2 | −1.028150049 | 0.001756143 |
| 5997      | RGS2        | regulator of G protein signalling 2 | −0.962605787 | 0.008475327 |
| 2215      | FCGR3B      | Fc fragment of IgG receptor IIb | −0.923733902 | 0.001756143 |
| 54504     | CPVL        | carboxypeptidase, vitellogenic-like | −0.922107067 | 0.017614822 |
| 2867      | FFAR2       | free fatty acid receptor 2 | −0.84291548 | 0.015277801 |
| 401233    | HTATSF1P2   | HIV-1 Tat specific factor 1 pseudogene 2 | −0.837903188 | 0.022307329 |
Enrichment analysis and co-expression network construction

For further functional investigation, we performed advanced analyses (GO and pathway analysis) of the DE genes. Figure 2A and Supplementary Table S2 show the results of GO biological process analysis, which were similar to those reported in the original microarray studies (5–8). The DE genes were significantly enriched in 6 GO terms, and the top 3 were “acute inflammatory response to antigenic stimulus” (P=0.013), “response to yeast” (P=0.021), and “defense response to bacterium” (P=0.024). In the pathway analysis, when DE genes were mapped to the KEGG database, no significant pathway was identified.

We also constructed a co-expression network using significantly correlated DE gene pairs (correlation coefficients greater than 0.8). The co-expression network contained 16 nodes and 23 edges (Figure 2B). We further extracted sub-networks for the GO enrichment analysis. One GO term (“chemotaxis”) was significantly enriched in the sub-network composed of ENPP2, LOC100134822, CREB5, CXCR2, and FCGR3B. In the sub-network containing ELANE, MPO, and PRTN34, 4 significant GO terms (“response to yeast”, “negative regulation of growth of symbiont in host”, “defense response to bacterium”, and “response to lipopolysaccharide”) were identified.

Classification performance evaluation and validation

ROC curves and AUC values suggested that the top performing DE genes were DDIT4, PRG2, ARHGEF40, MME, CD24, CEP55, CD69, UHRF1, and KIF11 (Figure 3), with the best single gene’s AUC as great as 0.829. The results of recursive feature selection showed that 11 genes (Supplementary Table S3) were sufficient to achieve close to 91% prediction accuracy. Based on the selected features, we trained an SVM classifier (pediatric septic shock non-survivors vs survivors) and evaluated its performance by 10-fold cross-validation (cross-validation error: 0.133). The SVM classifier performed well in the training dataset (AUC=0.986, Figure 4A). The pre-processed testing dataset was also used for independent validation, and the SVM classifier was...
constructed according to the training datasets, showing modest performance in the testing dataset (AUC=0.722, Figure 4B).

**Discussion**

Many genes are reported to be associated with septic shock mortality (23,26,27). It is a challenge currently to identify the most important candidate genes and pathways in septic shock prognosis. Growth of high-throughput transcriptomic data has enabled the integrative analysis of multiple datasets to discover robust candidates for prognosis and treatment. Therefore, to identify potential transcriptomic markers in pediatric septic shock prognosis, we performed an integrative analysis of multiple public microarray datasets.

In this study, 32 DE genes were identified in pediatric septic shock non-survivors, among which 18 genes were upregulated and 14 were downregulated. The most upregulated gene was OLFM4. The most downregulated gene was MME. Among the 32 DE genes identified, OLFM4 was reported to be present in only 20–25% of peripheral blood neutrophils and identified a subset of human neutrophils (28). MME is widely present in organs and tissues, contributing to the inactivation of many biological peptides (24,25). Decreased MME expression capacity was detected in the neutrophils from patients with septic shock (25). CXCR2 is a key stimulant to immune cell migration as well as recruitment, especially for neutrophils. The CXCR2 signaling pathway is suggested to be a potential target for modification of neutrophil dynamics in inflammatory disorders (29). CEACAM8 belongs to CEA family, which is known to function as intercellular adhesion molecules. CEACAM8 is suggested to be involved in the regulation of the neutrophil adhesion (30). ELANE encodes neutrophil elastase, which is a
cytotoxic serine protease stored in azurophil granules and released after neutrophil activation. ELANE also functions in neutrophil extracellular traps (31). Neutrophil dysfunction is known to promote sepsis, and the alterations of chemokines, cytokines, as well as other mediators lead to neutrophil dysfunction in sepsis. Thus, the DE genes identified in this study, such as OLFM4, MME, CXCR2, CEACAM8, and ELANE, probably take part in pediatric sepsis development by regulation of neutrophil function (32).

Furthermore, a series of DE genes identified in this study have not yet been investigated in pediatric septic shock prognosis, and the exact contributions of these genes to pediatric septic shock are not yet clear. Since these DE genes could serve as potential transcriptomic markers for pediatric septic shock, further research is required.

In the enrichment analysis of the 32 DE genes, “response to yeast” (P=0.021) was the second most enriched GO term. This is interesting because yeast infection has been reported to cause sepsis (33).

According to the literature, machine learning methods, including SVM, usually have high prediction accuracy (34). Hence, in this study, we trained an SVM classifier (pediatric septic shock in non-survivors vs survivors) using gene expression data to improve the performance of prognostic prediction. The SVM classifier performed well in the training dataset with an AUC value of 0.986 (Figure 4A), compared with Pediatric Sepsis Biomarker Risk Model-II (35). However, it showed modest performance in the testing dataset (AUC=0.722, Figure 4B), suggesting that there could be an over-fitting problem. To the best of our knowledge, no gene expression-based SVM model has yet been reported for pediatric septic shock prognostic prediction. We hope the predictive model trained in this study facilitates better prognosis and treatment of pediatric septic shock.

Because of the existence of several limitations, the results of our study should be considered cautiously. First, the sample size used in our study was insufficient. Second, stratified analyses according to potential influential factors (age, sex, and platform usage) could not be performed in this study due to the lack of corresponding data. Third, both the biological knowledge base and the pathway data are currently incomplete. In consideration of the reported impact of age and sex on pediatric sepsis patients (36), stratified analyses, based on factors including age, sex, and platform usage, are required in the future. Further analysis based on larger sample sizes is also necessary. Moreover, to explore the exact roles of candidate DE genes in pediatric septic shock, experimental verification and functional studies of these DE genes must also be performed in the future.

In conclusion, consistently DE genes in pediatric septic shock non-survivors were identified in this study, and they could be potential transcriptomic markers. GO analysis suggested that these candidates were strongly correlated with acute inflammatory response to antigenic stimuli, response to yeast, and defense responses to bacteria. We also trained an SVM classifier (non-survivors vs survivors) using candidate transcriptomic markers. These results provided novel insights into the prognosis of pediatric septic shock and promoted the generation of prognostic gene sets.

**Supplementary Material**

Click to view [pdf].

**Acknowledgments**

This work was supported by grants from National Natural Science Foundation (grant numbers 81971867,
References

1. Kissous N, Carapetis J. Pediatric sepsis in the developing world. J Infect 2015; 711: S21–S26, doi: 10.1016/j.jinf.2015.04.016.
2. van der Poll T, van de Veerdonk FL, Sestan AB, Netea MG. The immunopathology of sepsis and potential therapeutic targets. Nat Rev Immunol 2017; 17: 407–420, doi: 10.1038/nri.2017.36.
3. Cecconi M, Evans L, Levy M, Rhodes A. Sepsis and septic shock. Lancet 2018; 392: 75–87, doi: 10.1016/S0140-6736(18)30696-2.
4. Emr BM, Alcamo AM, Cardillo JA, Aneja RK, Mollen KP. Pediatric sepsis update: how are children different? Surg Infect (Larchmt) 2018; 19: 176–183, doi: 10.1089/sur.2017.316.
5. Cvijanovich N, Stanley TP, Lin R, Allen GL, Thomas NJ, Checchia P, et al. Validating the genomic signature of pediatric septic shock. Physiol Genomics 2008; 34: 127–134, doi: 10.1152/physiolgenomics.00025.2008.
6. Wynn JL, Cvijanovich NZ, Allen GL, Thomas NJ, Freishtat RJ, Anas N, et al. The influence of developmental age on the early transcriptomic response of children with septic shock. Mol Med 2011; 17: 1146–1156, doi: 10.2119/molmed.2011.00169.
7. Wong HR, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Wilsson DF, et al. Identification of pediatric septic shock subclasses based on genome-wide expression profiling. BMC Med 2009; 7: 34, doi: 10.1186/1741-7015-7-34.
8. Wong HR, Stanley TP, Sathivel B, Cvijanovich N, Lin R, Allen GL, et al. Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. Physiol Genomics 2007; 30: 146–155, doi: 10.1152/physiolgenomics.00024.2007.
9. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dubot D, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004; 5: R80, doi: 10.1186/gb-2004-5-10-r80.
10. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44–57, doi: 10.1038/nprot.2008.211.
11. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37: 1–13, doi: 10.1093/nar/gkn923.
12. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498–2504, doi: 10.1101/gr.1239303.
13. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013; 41: D991–D995, doi: 10.1093/nar/gks1193.
a preliminary study. *PLoS One* 2017; 12: e0169653, doi: 10.1371/journal.pone.0169653.

28. Clemmensen SN, Bohr CT, Rørvig S, Glenthøj A, Mora-Jensen H, Cramer EP, et al. Olfactomedin 4 defines a subset of human neutrophils. *J Leukoc Biol* 2012; 91: 495–500, doi: 10.1189/jlb.0811417.

29. Zhang X, Guo R, Kambara H, Ma F, Luo HR. The role of CXCR2 in acute inflammatory responses and its antagonists as anti-inflammatory therapeutics. *Curr Opin Hematol* 2019; 26: 28–33, doi: 10.1097/MOH.0000000000000476.

30. Zhao L, Furebring M, Xu S, Venge P. Subcellular localization and mobilization of carcinoembryonic antigen-related cell adhesion molecule 8 in human neutrophils. *Br J Haematol* 2004; 125: 666–673, doi: 10.1111/j.1365-2141.2004.04963.x.

31. Skokowa J, Dale DC, Touw IP, Zeidler C, Welte K. Severe congenital neutropenias. *Nat Rev Dis Primers* 2017; 3: 17032, doi: 10.1038/nrdp.2017.32.

32. Zhang F, Liu AL, Gao S, Ma S, Guo SB. Neutrophil dysfunction in sepsis. *Chin Med J (Engl)* 2016; 129: 2741–2744.

33. Toh J, Snowball V. Postpartum sepsis caused by an intra-abdominal yeast infection. *J Obstet Gynaecol* 2020; 40: 432–433, doi: 10.1080/01443615.2019.1645821.

34. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. NY: Springer; 2009.

35. Wong HR, Cvijanovich NZ, Anas N, Allen GL, Thomas NJ, Bigham MT, et al. Pediatric sepsis biomarker risk model-II: redefining the pediatric sepsis biomarker risk model with septic shock phenotype. *Crit Care Med* 2016; 44: 2010–2017, doi: 10.1097/CCM.0000000000001852.

36. Maat M, Buysse CM, Emonts M, Spanjaard L, Joosten KF, de Groot R, et al. Improved survival of children with sepsis and purpura: effects of age, gender, and era. *Crit Care* 2007; 11: R112, doi: 10.1186/cc6161.