Whole blood transcriptomic investigation identifies long non-coding RNAs as regulators in sepsis

Lixin Cheng (✉ easonlcheng@gmail.com)

Shenzhen People's Hospital <https://orcid.org/0000-0002-9427-383X>

Chuanchuan Nan

Department of Critical Care Medicine, Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Lin Kang

Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Ning Zhang

Department of Critical Care Medicine, Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Sheng Liu

Department of Critical Care Medicine, Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Huaisheng Chen

Department of Critical Care Medicine, Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Chengying Hong

Department of Critical Care Medicine, Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Youlian Chen

Department of Critical Care Medicine, Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Zhen Liang

Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Xueyan Liu

Department of Critical Care Medicine, Shenzhen People's Hospital, The Second Clinical Medicine College of Jinan University, Shenzhen, China

Research

Keywords: sepsis, lncRNA, functional module, gene coexpression, survival analysis, differential analysis

Posted Date: February 27th, 2020

DOI: <https://doi.org/10.21203/rs.2.24685/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Translational Medicine on May 29th, 2020. See the published version at <https://doi.org/10.1186/s12967-020-02372-2>.

Abstract

Background: Sepsis is a fatal disease referring to the presence of a known or strongly suspected infection coupled with systemic and uncontrolled immune activation causing multiple organ failure. However, neither pathogenic long non-coding RNAs (lncRNAs) nor biological network analysis in sepsis draws enough attention to the society of sepsis studies.

Methods: We performed an in-silico investigation of the gene coexpression pattern for the patients response to all-cause sepsis in consecutive intensive care unit (ICU) admissions. Sepsis coexpression gene modules were identified using WGCNA and enrichment analysis. lncRNAs were determined as sepsis biomarkers based on the interactions among lncRNAs and the identified modules.

Results: Twenty-three sepsis modules, including both differentially expressed modules and prognostic modules, were identified from the whole blood RNA expression profilings of sepsis patients. Five lncRNAs, FENDRR, MALAT1, TUG1, CRNDE, and ANCR, were detected as sepsis regulators based on the interactions among lncRNAs and the identified coexpression modules. Furthermore, we found that CRNDE and MALAT1 may act as miRNA sponges of sepsis related miRNAs to regulate the expression of sepsis modules. Ultimately, FENDRR, MALAT1, TUG1, and CRNDE were reannotated using three independent lncRNA expression datasets and validated as differentially expressed lncRNAs.

Conclusion: The procedure facilitates the identification of prognostic biomarkers and novel therapeutic strategies of sepsis. Our findings highlight the importance of transcriptome modularity and regulatory lncRNAs in the progress of sepsis.

Background

Sepsis refers to the presence of a known or strongly suspected infection coupled with systemic and uncontrolled immune activation causes multiple organ dysfunction with worldwide mortality among 17%-26% [1, 2]. Common symptoms of sepsis contain fever, increased heart rate, increased breathing rate, and confusion, while specific symptoms include a cough with pneumonia or painful urination with a kidney infection. Sepsis can progress to septic shock with dramatically dropped blood pressure leading to a much higher mortality of 40%. However, sepsis is a complex heterogeneous disease implicating a variety of cellular processes and we can hardly identify reliable diagnostic and prognostic biomarkers for sepsis in clinical [3].

High-throughput gene expression analysis can detect tens of thousands of genes simultaneously, which provides vast opportunities to improve prognostic accuracy and address clinical questions that otherwise cannot be answered. Transcriptomic strategies have been adopted among numerous diseases to investigate differential expression analysis, coexpression pattern, survival analysis, prediction modeling, etc., leading to substantial advances in the identification of promising diagnostic biomarkers as well as clinical use and disease treatment [3–5]. Poll and his colleagues have utilized the high-throughput blood gene expression profiling to carry out the comparative analysis of the systemic response for sepsis

patients diagnosed in distinct subgroups and endotypes, such as community-acquired and hospital-acquired pneumonia, bacterial sepsis and fungal sepsis, hyper-inflammatory and hypo-inflammatory, and critically ill patients in different platelet counts [6–9]. Gene expression signatures and candidate plasma proteins have been identified and characterized from critically ill patients with different subtypes of sepsis [10–12].

Also, biological network analysis coupled with functional module analysis have been commonly deployed in the domain of cancer study, to probe the tumor biogenesis and dysfunction in patients with cancer, which facilitated the pathway and mechanism studies that otherwise would be hardly discovered [13]. We previously designed a procedure SMILE for the identification of protein modules taking account into the subcellular localization of proteins [14, 15]. The resulting modules showed high correspondence with known modules and canonical pathways. Moreover, a computational framework was proposed to predict moonlighting lncRNAs by clustering the protein interaction network to determine modules with independent functions [16].

Long non-coding RNAs (lncRNAs) are a type of transcripts with more than 200 nucleotides that have low protein-coding potential, which function in a variety of cellular processes and usually serve as disease diagnostic and prognostic markers [16, 17]. Numerous studies have implicated the mutations and dysregulations of lncRNAs contribute to the development of immunity diseases and cancers [18–20]. Accumulating evidence has demonstrated lncRNAs playing roles as competing endogenous RNAs (ceRNAs) to determine the fate of gene transcripts in a variety of diseases [21, 22]. However, the role of lncRNAs in sepsis remains largely unknown, although sporadic works reported that organ failure in sepsis is associated with the expression change of lncRNAs in some tissues, i.e., liver, kidney, and skeletal muscle [23]. Thus, in this context, we need to find new lncRNA therapeutic targets and investigate their regulatory mechanisms in sepsis for the severely ill sepsis patients.

We comprehensively performed an in-silico investigation of the gene coexpression pattern for the patients response to all-cause sepsis in consecutive intensive care unit (ICU) admissions. We identified diagnostic modules based on the whole blood RNA expression profiles of sepsis patients, and subsequently predicted sepsis associated lncRNAs on the basis of the interactions among lncRNAs and the identified coexpression modules. Afterthat, we established five candidate lncRNA regulators of sepsis and investigated their regulatory mechanism through miRNAs playing in a competing endogenous RNA fashion. Ultimately, FENDRR, MALAT1, TUG1, and CRNDE were reannotated using three independent expression cohorts and validated as differentially expressed lncRNAs.

Materials And Methods

1. Gene expression datasets and data preprocessing

Five microarray datasets GSE65682, GSE69528, GSE95233, GSE57065, and GSE28750 from the NCBI GEO database were used in this study (Table 1) [24]. Raw array data preprocessing was performed using

the affy package in the R environment [25]. The raw gene expression matrixes were normalized by the RMA method [26–28]. We correct for batch effects of the datasets by ComBat [29]. Only the common genes detected in both datasets remained for analysis. Average expression intensities were used when multiple probe sets mapped an individual gene symbol. The genefilter algorithm was used to filter genes with interindividual variability over 0.5 [30], resulting in 11,222 most variable genes to construct the sepsis coexpression network. To overcome multiple comparison, Benjamini-Hochberg adjusted probabilities were used to define significance throughout the paper [31].

Table 1
Whole Blood Expression Datasets

GSE Number	Tissue	Control	Sepsis	Platform
mRNA expression:				
GSE65682[12]	Whole blood	42	522	Affymetrix Human Genome U219 Array
GSE69528[46]	Whole blood	28	83	Illumina HumanHT-12 V4.0 expression beadchip
lncRNA expression:				
GSE95233[47]	Whole blood	22	51	Affymetrix Human Genome U133 Plus 2.0 Array
GSE57065[48]	Whole blood	25	28	Affymetrix Human Genome U133 Plus 2.0 Array
GSE28750[49]	Whole blood	20	10	Affymetrix Human Genome U133 Plus 2.0 Array

Series GSE65682 was analyzed using the Affymetrix HG-U129 platform, including 42 healthy samples and 760 patients admitted to the ICU with sepsis. 522 patients with sepsis among them were picked up for further analysis. We used the dataset GSE65682 as the core discovery dataset and the primary results were based on the dataset, because it has the largest size of whole blood sepsis samples and most of the samples have clinical information. Series GSE69528 contains 83 sepsis and 28 healthy whole blood samples analyzed using Illumina Human HT-12 V4.0 expression beadchip. This dataset was used for the validation of module identification, as it has the second largest size of adult whole blood sepsis samples.

2. Reannotation of Gene expression datasets

To explore how the lncRNAs are expressed in sepsis, we reannotated lncRNAs based on three sepsis adult whole-blood gene expression datasets, GSE95233, GSE57065, and GSE28750. All of them are on the same platform of Affymetrix Human Genome U133 Plus 2.0 that were designed for detecting the expression intensity of coding genes. The platform of Affymetrix Human Genome U133 Plus 2.0 Array has been widely used for gene expression profiling of patient with sepsis. On top of this, it has the most comprehensive coverage of the annotated human lncRNAs. Using the latest NetAffx Annotation File, HG-

U133_Plus_2 Annotations (Release 35, 04/16/15), we reannotated the lncRNAs of the three datasets as follows: 1) The Refseq ID labeled with NR_ or XR_, indicative of non-coding RNAs, are retained; 2) the Ensemble gene IDs annotated with antisense, processed transcripts, sense overlapping, non-sense mediated decay, sense intronic or lincRNA are retained; and 3) pseudogenes, rRNAs, microRNAs, and other small RNAs including tRNAs, snRNAs and snoRNAs are filtered out. Finally, 5,016 probesets were detected as lncRNAs representing 3,640 unique lncRNAs. Probesets encoding more than one lncRNA were averaged.

3. Coexpression network construction

The sepsis expression cohort was independently processed using the weighted gene coexpression network analysis (WGCNA) for both datasets[32, 33]. A coexpression matrix is build up firstly, which is an adjacent matrix measuring the Pearson Correlation Coefficient (PCC) of all gene pairs. Then, a power function $f(x) = x^b$ is used to tune the weighted matrix or network to be scale-free. A common linear model that regressed the connectivity frequency on gene connectivity is used to assess the network scale-free degree, with the fitting index R^2 close to 1 indicates perfect organized. b was set as 6 for both datasets to construct the scale-free networks (Supplementary Figs. 1 and 2). Afterward, the weighted coexpression matrix is transformed into a topological overlap matrix (TOM), which is a classical algorithm considering both direct and indirect interactions of all the gene members in the network, resulting in biologically more meaningful modules. Modules with gene number over 20 were determined for further analysis.

4. Differentially expressed genes and modules

To identify differentially expressed genes (DEGs) between sepsis and normal samples, gene expression data were analyzed by the two-tailed t-test with a threshold of 0.01 and log2 transformed absolute Fold Change (FC) value of 1. A module is defined as Differentially Expressed Module (DEM) if the module significantly overrepresents the DEGs. Similarly, a module is defined up-regulated (or down-regulated) DEM if the module significantly overrepresents the up-regulated (or down-regulated) DEGs. The statistical significance is assessed by the hypergeometric test with p-value less than 0.01, which is defined as follows,

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{i}{x} \binom{m-i}{n-x}}{\binom{m}{n}}$$

where n is the network size or the total number of genes of the coexpression network, m is the module size, x is the number of DEGs, and i is the number of DEGs included in the module. The clusterprofiler package in R was adopted to perform the functional annotation of the identified DEGs and gene modules [34].

5. Survival associated modules

Principal component analysis (PCA) was used to evaluate whether gene modules are relevant to the clinical outcome of sepsis patients. For each module, the first principal component of its gene members is calculated as module eigengene (ME), which served as the most representative gene expression of all genes in a module [17]. It was used to risk-stratified the sepsis patients into two subgroups. Then, we examined the correlation between ME and patient overall outcome to compute module-trait relevance. A module is associated with a survival outcome if the correlation p-value is below 0.05. Kaplan–Meier survival curves were used for illustrating the result of survival analysis, in which ME is the risk score assessing the prognosis ability. For the 760 sepsis samples of the discovery dataset, only 479 of them having clinical information were utilized for survival analysis.

6. ncRNA-module interaction

The interactions between lncRNAs and gene products were obtained from two databases, LncRNA2Target v2.0 [35] and RAID v2.0 [36]. LncRNA2Target v2.0 is a high-confidence resource containing the relationships between lncRNAs and their target genes. We only adopted the literature mining low-throughput interactions. RAID v2.0 is an online repository of RNA-protein interactions, including interactions between proteins and lncRNA, circRNA, pseudogene, and miRNA, and only the experimental lncRNA-protein interactions were applied in this study. Together, 1,724 lncRNAs with 31,179 gene/protein targets were established for further analysis. We define a lncRNA regulates a module if the genes in the module significantly overrepresent the target genes of the lncRNA (p-value < 0.01, hypergeometric test). The same strategy was also adopted for the miRNA-module interaction, where the miRNA targets were obtained from mirCode [37], mirDB [38], and mirTarBase [39]. For competing endogenous RNA analysis, only the lncRNA-module pairs shared at least one miRNAs were determined as the lncRNA-miRNA-mRNA interaction. Additionally, we performed a literature search of the sepsis related miRNAs and collected 30 unique miRNAs as the sepsis diagnostic miRNAs (Supplementary Table 1).

7. Workflow of sepsis lncRNA identification

As shown in **Figure 1**, the main procedure consists of the 12 steps as follows:

- Preprocess the raw data.
- Establish the gene expression matrix for the genes with high variation.
- Construct the gene coexpression network.
- Identify gene coexpression modules using WGCNA.
- Eliminate unstable modules by another expression dataset.
- Screen differentially expressed genes (DEGs).
- Identify modules enriched with DEGs (DEMs).
- Calculate module eigengene and perform survival analysis.
- Identify survival associated modules (SAMs).
- Define sepsis modules by integrating DEMs and SAMs.
- Construct the lncRNA-module interaction network.

- Select candidate sepsis lncRNAs that are topologically critical.

We can obtain the sepsis lncRNA candidates using the procedure once the gene expression data, clinical data, and lncRNA-gene interaction data are imported.

Results

1. Overview of workflow

We aimed to construct a lncRNA-module network composed of modules associated with sepsis pathology and lncRNAs with prognostic potential. To construct the network, we started by collecting sepsis gene expression datasets. Two datasets GSE65682 and GSE69528 were used in this study and were served as the primary and validation datasets, respectively. Then the analysis was performed mainly on the primary dataset following the procedure in Fig. 1. (1) Preprocessing the raw data using RMA. (2) Establishing the gene expression matrix for the genes with high expression variance. (3) Constructing the gene coexpression network represented by the Pearson correlation coefficients of all gene pairs. (4) Identifying gene coexpression modules using WGCNA. (5) Filtering out unstable modules by another validation expression dataset. Only the modules detected in both datasets were retained for subsequent analysis. (6) Screening DEGs between the sepsis and normal samples for the primary dataset. (7) Identifying DEMs using hypergeometric test. (8) Calculating module eigengene (ME) and perform survival analysis. (9) Identifying survival associated modules (SAMs) by examining the correlation between ME and patient survival outcome. (10) Combining DEMs and SAMs and define them as sepsis modules. (11) Constructing the lncRNA-module interaction network. The interactions were established using hypergeometric test to assess whether a sepsis module significantly overrepresents the target genes of a lncRNA. (12) Select the hub lncRNAs connecting more than three sepsis modules as candidate sepsis lncRNAs. Five sepsis lncRNAs were ultimately identified, FENDRR, MALAT1, TUG1, CRNDE, and ANCR.

2. Coexpression network and modules

The primary results were based on the GSE65682 dataset as it has the largest sample size, 760 sepsis vs 42 normal samples. For this working dataset, we constructed a coexpression network consisting of 11,222 genes with expression variance over 0.5 across the 760 sepsis patient samples. The topological overlap matrix illustrates an apparent organizational structure of the sepsis gene coexpression network, demonstrating that sepsis configures an array of specific coexpression structure. In total 59 modules were detected with sizes ranging from 30 to 750 (Fig. 2A). Different coexpression modules are highlighted in distinct colors. The detailed procedure of module identification and the module dendrogram are shown in Material and Method section and Supplementary Fig. 1.

Using the same procedure, we also identified another set of gene module based on an independent microarray dataset GSE69528 for validation (Supplementary Figs. 2 and 3). Common genes detected in both datasets were used for the coexpression network construction. Only the reproducible modules were retained for the subsequent analysis to investigate the expression change of modules during disease progression. As shown in Fig. 2B, rows are modules identified from our primary dataset GSE65682, while

columns are modules determined from the validation dataset GSE69528. Significance of pairwise module overlap was measured by the $-\log_{10}$ transferred hypergeometric test p-values. It is clear that a high reproducibility was achieved for the two module lists. 52 out of 59 modules have at least one significant ($P < 0.01$, hypergeometric test) overlapping modules in the validation dataset (Fig. 2C).

3. Establishment of sepsis modules

Using the t-test p-value of 0.01 and absolute fold change of 2 as thresholds, we screened 750 down-regulated DEGs and 391 up-regulated DEGs from the primary dataset (Fig. 3A). The down-regulated DEGs are significantly involved in biological processes like neutrophil mediated immunity, defense response to bacterium, platelet degranulation, etc. (Fig. 3B), while the up-regulated DEGs are enriched in the functional categories of T cell activation, Lymphocyte activation, T cell receptor signaling pathway, etc. (Fig. 3C).

To determine the expression difference of modules between the sepsis and normal samples, we adopted the hypergeometric test to evaluate whether a module significantly overrepresents up-regulated or down-regulated DEGs. A module is referred to as Differentially Expressed Module (DEM) if a substantial large fraction of genes is differentially expressed, indicating distinct expression pattern between the sepsis patients and the healthy samples. Thus, some modules are over expressed in sepsis whereas some others are low expressed. In total ten up-regulated and 13 down-regulated DEMs were detected from the sepsis coexpression network (Fig. 2D).

Moreover, to identify the modules associated with clinical outcome in sepsis, we performed multivariate Cox regression analysis to assess the significance of the correlation between patient overall survival and the Eigengene (EG) values of each module. As shown in Fig. 3E, the risk scores of the EG values were sorted with corresponding survival information for module 32. The dotted line in the middle of the figures corresponds to the median of EG value, which stratifies the sepsis patients into two subgroups with high and low risk. Figure 3D illustrates the Kaplan–Meier curves for the patients with clinical information according to the EG of M32. Patients with high EG values show much poorer prognostic than those with low EG values, indicating that the dysfunction or oncogenic of M32 is close related to the prognosis of sepsis patients. The expression profilings of the DEGs in module 32 are illustrated as a heatmap in Fig. 3F. In total, we identified 14 modules from sepsis samples whose EGs are substantially correlated with patient overall survival and we defined them as survival-associated modules (SAM).

4. Characteristic of sepsis modules

31 sepsis modules, including both SAM and DEM, were screened from the primary dataset, implying novel gene signatures associated with sepsis pathology. We note an overlap of six modules (around 20%) between the two sets of SAM and DEM. Three of them are down-regulated, i.e., M22, M32, and M4, while the other three are up-regulated, i.e., M15, M23, and M47. The down-regulated module M22, for instance, consists of 22 genes closely co-expressed with each other; nine out of them are down-regulated DEGs playing as hub genes in the module (Fig. 4A). Kaplan–Meier curves were plotted for the rank-ordered Eigen Module values of M22 to carry out the 28-day survival analysis (Fig. 4B). It is apparent that patients with high EG value have substantial shorter survival time than those with low EG value. M22 are mainly

implicated in biological processes like T cell activation, regulation of lymphocyte activation, leukocyte cell-cell adhesion, etc. (Fig. 4C). For the up-regulated module M47, it has 26 gene members and nine of them are DEGs that up-regulated and more topologically important (Fig. 4D). The Kaplan–Meier curves show that patients with high EG value of M47 have a significantly worse prognosis than the low-EG ones (Fig. 4E). M47 are involved in function categories of neutrophil mediated immunity as well as neutrophil activation and degranulation (Fig. 4F). Some other sepsis modules and their corresponding Kaplan–Meier curves are shown in Supplementary Figs. 4 and 5.

Interestingly, we found that DEGs in the sepsis modules, either up-regulated or down-regulated, are prone to play a central role topologically in comparison to the non-DEGs. For instance, DEGs in M47 have an average correlation coefficient of 0.6 while the connectivity is merely 0.45 for the other genes ($p < 3.96 \times 10^{-5}$, Mann–Whitney U test). Similar results can be observed for the other four modules of both SAM and DEM, including M15, M22, M23, and M45 (Fig. 4G). The DEGs of M32 are overall more correlated with module gene members in expression, although not significantly. Generally, the average correlation coefficient of module DEGs is significantly higher than that of the other genes, suggesting that genes differentially expressed may drive the biogenesis or dysfunction of the coexpression gene modules.

5. Sepsis lncRNA candidates

We constructed a lncRNA-module interaction network including 251 interactions between 23 sepsis modules and 201 lncRNAs (Fig. 5A). Although most of the lncRNAs regulate none or merely a single sepsis module (Fig. 5B), FENDRR, MALAT1, TUG1, CRNDE, and ANCR connect multiple sepsis modules with the connectivity of 14, 10, 10, 8, and 5, respectively, which are expected to have high potentials to be involved in the sepsis progress.

A subnetwork concentrating on the five candidate sepsis lncRNAs and their regulated sepsis modules is shown on the bottom panel of Fig. 5A. FENDRR, the FOXF1 adjacent non-coding developmental regulatory RNA, plays as a hub regulator mediating 14 sepsis modules in the lncRNA-module interaction network. Both MALAT1 and CRNDE regulate ten sepsis modules and they share seven common modules, six out of them are down-regulated. TUG1 interacts with six down-regulated and two up-regulated DEMs, indicating that TUG1 tend to involve in under expression pathways. In contrast, ANCR (Angelman syndrome chromosome region) links four up-regulated DEMs and only one down-regulated DEMs, suggesting that ANCR may mediate some over expression pathways implicated in sepsis.

Furthermore, we investigated the regulatory mechanism of how the lncRNAs regulate the modules in sepsis from the perspective of competing endogenous RNAs (ceRNAs), which impact the translation rate of mRNAs by competing for shared miRNAs [21, 22]. lncRNAs are able to share the same miRNA response elements with mRNAs transferred by the sepsis modules, thereby sponging miRNAs intended to bind to these mRNAs and depressing the overall expression level of sepsis modules (Fig. 5C). Several miRNAs have been previously validated as potential regulators in sepsis, such as miR-34a, miR-206, and miR-199b-5p. By these miRNAs, we found that CRNDE regulates module 5 and module 20 through miR-199b-5p (CRNDE $\diamond$ miR-199b-5p $\diamond$ M5/M20), indicating that CRNDE acts as a miRNA sponge of miR-

199b-5p and thereby modulating the transcripts of genes in module 5 and module 20 (Fig. 5D). Similarly, MALAT1 regulates module 7 and module 20 through the miR-206-mediated lncRNA-mRNA interactions (MALAT1 $\diamond$ miR-206 $\diamond$ M7/M20). The in-detail information of these candidate lncRNAs including secondary structure and miRNA targets are provided in Fig. 5E and F.

6. Validation using independent lncRNA datasets

Following a computational pipeline (see Methods), we reannotated the probes from three array datasets to obtain the lncRNA expression profilings. We separately screened the differentially expressed lncRNAs (DELs) from the datasets of GSE95233, GSE57065, and GSE28750. Our finding shows that four out of the five sepsis lncRNA candidates are differentially expressed in at least one independent dataset except for ANCR, whose probes were not covered by the array platform (Fig. 6). Specifically, CRNDE is significantly up-regulated in the sepsis samples of all the three datasets, in which the (log2 transferred) fold changes are 0.43, 0.55, and 0.66, respectively. FENDRR and MALAT1 are significantly down-regulated in two datasets, while TUG1 is differentially expressed only in the dataset of GSE95233.

CRNDE, an oncogene that is usually overexpressed in tumor cells, contributes a lot to cellular proliferation, migration, invasion, and apoptosis [40]. More importantly, CRNDE can modulate the TLR3-NF- κ B cytokine signaling pathway to trigger inflammation [41, 42], suggesting that CRNDE may serve as a regulator in sepsis. In sepsis, genes or gene modules induced by MALAT1 may modulate their expression pattern in endothelial cells, which is critical as MALAT1 has been reported to mediate inflammation in traumatic brain injury [42]. Also, it was reported that TUG1 is able to affect the development of sepsis-associated acute kidney injury via modulating NF- κ B pathway [43]. FENDRR has never been mentioned in the induction or progress of sepsis before, so it can be considered as a novel lncRNA regulator for sepsis.

Discussion

We used a module-centric algorithm to identify sepsis lncRNAs via a network linking lncRNAs and coexpression modules. Twenty-three sepsis modules, including both differentially expressed modules and prognostic modules, were detected from the sepsis whole blood gene expression profilings. We identified five sepsis lncRNAs, FENDRR, MALAT1, TUG1, CRNDE, and ANCR, all of which are connecting five or more sepsis modules, indicating their functions are highly related with biological processes of sepsis. Further, we probed the regulatory mechanism of CRNDE and MALAT1 who are acting as competing endogenous RNAs (ceRNAs). CRNDE interacts module 5 and module 20 through miR-199b-5p, while MALAT1 sponges miR-206 to regulate the target module 7 and module 20. At last, the five sepsis lncRNAs were independently validated in three gene expression datasets of sepsis. Four out of them were reannotated and detected as differentially expressed lncRNAs in at least one dataset.

Genome-wide expression study of sepsis is relatively at its infancy and several technologies prevalently used in other diseases have not been widely adopted in sepsis. In order to detect the sepsis lncRNAs, we integrated the conventional approaches including gene coexpression, module identification, differential

analysis, survival analysis, and lncRNA-gene interaction, as well as mathematical and statistical algorithms. We comprehensively studied the gene coexpression pattern of patients with all-cause sepsis in ICU admissions in this study, although sepsis is a heterogeneous immunity disease and the mortalities of sepsis patients in distinct subtypes are substantially different [44]. In the future, we will investigate the coexpression pattern of patients with sepsis in specific subtypes, such as community-acquired and hospital-acquired pneumonia, bacterial sepsis and fungal sepsis, hyper-inflammatory and hypo-inflammatory, and endotypes classified by platelet counts [6, 8]. Since the interactions among lncRNAs and target genes are far from complete, the discovery of sepsis lncRNAs is limited by the interaction coverage [16, 17]. An alternative strategy is to produce the genome-wide RNA-seq data including both coding and non-coding genes, then a coding-non-coding network can be constructed and the association among coding and non-coding genes would be well established [45].

This is the first work computationally detecting the sepsis lncRNAs using coexpression and network analysis for application in the intensive care unit environment. Also, FENDRR is first proposed as a sepsis related lncRNA. The predicted sepsis lncRNAs is helpful for the diagnosis of sepsis and can improve our understanding of sepsis progress and development, although further experimental validation is required to elaborate how lncRNAs modulate the molecular signaling pathways of sepsis. The procedure will facilitate the identification of other types of sepsis-related molecules, such as circRNAs and pseudogenes, for the patients in critical care settings.

Conclusion

This study identified five lncRNAs as sepsis regulators based on the interactions among lncRNAs and the identified sepsis modules, four of which were differentially expressed in three independent datasets. The procedure facilitates the identification of prognostic biomarkers and novel therapeutic strategies of sepsis. Our findings highlight the importance of transcriptome modularity and regulatory lncRNAs in the progress of sepsis.

Additional Files

Additional file 1: Figure S1. Identification of co-expression modules for dataset GSE65682. A) Paramater setup. B, Gene dendrogram and module colors. C) Module dendrogram. **Figure S2.** Identification of co-expression modules for dataset GSE69528. A) Paramater setup. B, Gene dendrogram and module colors C) Module dendrogram. **Figure S3.** Identification of co-expression modules from the topological overlap matrix using WGCNA for dataset GSE69528. **Figure S4.** Kaplan–Meier curves of two patient groups with higher or lower EG value for module 15, 23, 45, and 36, respectively. **Figure S5.** Example of the coexpression moudules enriched of up (31 and 37) or down-regulated DEGs (45 and 39). Vertexes correspond to genes and edges correspond to expression correlation. Only the edges with the abslute value of PCC greater than 0.5 are shown. Up-regulated DEGs are colored in red while down-regulated DEGs are in blue.

Additional file 2: Supplementary table of the curated sepsis miRNA biomarkers. (xlsx 14 kb)

Declarations

Authors' contributions

LC and LK conceived the idea and drafted the manuscript. LC performed data analysis. ZL and XL supervised this project. SL, and NZ performed data management and analysis. CN, KC, HC, CH, and YC helped interpret the results and give suggestions. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Data are available on request.

Competing interests

none declared.

Funding

This work was supported by Health and Family Planning Commission of Shenzhen Municipality (SZXJ2017027 to X.L.)

Acknowledgements

None.

References

1. Angus DC and van der Poll T. Severe sepsis and septic shock. *The New England journal of medicine*. 2013; 369(9):840-851.
2. Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, Angus DC, Reinhart K and International Forum of Acute Care T. Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. *American journal of respiratory and critical care medicine*. 2016; 193(3):259-272.
3. Parnell GP, Tang BM, Nalos M, Armstrong NJ, Huang SJ, Booth DR and McLean AS. Identifying key regulatory genes in the whole blood of septic patients to monitor underlying immune dysfunctions. *Shock*. 2013; 40(3):166-174.
4. Cheng L, Lo LY, Tang NL, Wang D and Leung KS. CrossNorm: a novel normalization strategy for microarray data in cancers. *Scientific reports*. 2016; 6:18898.
5. Cheng L, Wang X, Wong PK, Lee KY, Li L, Xu B, Wang D and Leung KS. ICN: a normalization method for gene expression data considering the over-expression of informative genes. *Molecular bioSystems*. 2016; 12(10):3057-3066.
6. Bos LD, Scicluna BP, Ong DY, Cremer O, van der Poll T, Schultz MJ and consortium M. Understanding Heterogeneity in Biological Phenotypes of ARDS by Leukocyte Expression Profiles. *American journal of respiratory and critical care medicine*. 2019.
7. Scicluna BP, Klein Klouwenberg PM, van Vught LA, Wiewel MA, Ong DS, Zwinderman AH, Franitza M, Toliat MR, Nurnberg P, Hoogendijk AJ, Horn J, Cremer OL, Schultz MJ, et al. A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *American journal of respiratory and critical care medicine*. 2015; 192(7):826-835.
8. van Vught LA, Scicluna BP, Wiewel MA, Hoogendijk AJ, Klein Klouwenberg PM, Franitza M, Toliat MR, Nurnberg P, Cremer OL, Horn J, Schultz MJ, Bonten MM and van der Poll T. Comparative Analysis of the Host Response to Community-acquired and Hospital-acquired Pneumonia in Critically Ill Patients. *American journal of respiratory and critical care medicine*. 2016; 194(11):1366-1374.
9. van Vught LA, Scicluna BP, Wiewel MA, Hoogendijk AJ, Klein Klouwenberg PMC, Ong DSY, Cremer OL, Horn J, Franitza M, Toliat MR, Nurnberg P, Bonten MMJ, Schultz MJ, et al. Association of Gender With Outcome and Host Response in Critically Ill Sepsis Patients. *Critical care medicine*. 2017; 45(11):1854-1862.
10. Bos LD, Schouten LR, van Vught LA, Wiewel MA, Ong DSY, Cremer O, Artigas A, Martin-Loeches I, Hoogendijk AJ, van der Poll T, Horn J, Juffermans N, Calfee CS, et al. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. *Thorax*. 2017; 72(10):876-883.
11. Cheng SC, Scicluna BP, Arts RJ, Gresnigt MS, Lachmandas E, Giamarellos-Bourboulis EJ, Kox M, Manjeri GR, Wagenaars JA, Cremer OL, Leentjens J, van der Meer AJ, van de Veerdonk FL, et al.

- Broad defects in the energy metabolism of leukocytes underlie immunoparalysis in sepsis. *Nature immunology*. 2016; 17(4):406-413.
12. Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nurnberg P, Schultz MJ, Horn J, Cremer OL, Bonten MJ, Hinds CJ, Wong HR, et al. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *The Lancet Respiratory medicine*. 2017; 5(10):816-826.
 13. Cheng L, Fan K, Huang Y, Wang D and Leung KS. Full Characterization of Localization Diversity in the Human Protein Interactome. *Journal of proteome research*. 2017; 16(8):3019-3029.
 14. Cheng L, Liu P and Leung K-S. (2017). SMILE: A Novel Procedure for Subcellular Module Identification with Localization Expansion. *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics: ACM*, pp. 754-755.
 15. Cheng L, Liu P and Leung KS. SMILE: a novel procedure for subcellular module identification with localisation expansion. *IET systems biology*. 2018; 12(2):55-61.
 16. Cheng L and Leung K-S. Identification and characterization of moonlighting long non-coding RNAs based on RNA and protein interactome. *Bioinformatics*. 2018; 1:10.
 17. Cheng L and Leung K-S. Quantification of non-coding RNA target localization diversity and its application in cancers. *Journal of molecular cell biology*. 2018; 10(2):130-138.
 18. Ning S, Zhang J, Wang P, Zhi H, Wang J, Liu Y, Gao Y, Guo M, Yue M, Wang L and Li X. Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic acids research*. 2016; 44(D1):D980-985.
 19. Gao Y, Wang P, Wang Y, Ma X, Zhi H, Zhou D, Li X, Fang Y, Shen W, Xu Y, Shang S, Wang L, Wang L, et al. Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic acids research*. 2019; 47(D1):D1028-D1033.
 20. Zhou M, Zhao H, Wang X, Sun J and Su J. Analysis of long noncoding RNAs highlights region-specific altered expression patterns and diagnostic roles in Alzheimer's disease. *Briefings in bioinformatics*. 2019; 20(2):598-608.
 21. Song YX, Sun JX, Zhao JH, Yang YC, Shi JX, Wu ZH, Chen XW, Gao P, Miao ZF and Wang ZN. Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. *Nature communications*. 2017; 8(1):289.
 22. Denzler R, Agarwal V, Stefano J, Bartel DP and Stoffel M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Molecular cell*. 2014; 54(5):766-776.
 23. Ho J, Chan H, Wong SH, Wang MH, Yu J, Xiao Z, Liu X, Choi G, Leung CC, Wong WT, Li Z, Gin T, Chan MT, et al. The involvement of regulatory non-coding RNAs in sepsis: a systematic review. *Critical care*. 2016; 20(1):383.
 24. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic acids research*. 2013; 41(Database issue):D991-995.

25. Gautier L, Cope L, Bolstad BM and Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004; 20(3):307-315.
26. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4(2):249-264.
27. Wang D, Cheng L, Zhang Y, Wu R, Wang M, Gu Y, Zhao W, Li P, Li B, Zhang Y, Wang H, Huang Y, Wang C, et al. Extensive up-regulation of gene expression in cancer: the normalised use of microarray data. *Molecular bioSystems*. 2012; 8(3):818-827.
28. Liu X, Li N, Liu S, Wang J, Zhang N, Zheng X, Leung K-S and Cheng L. Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review. *Frontiers in Bioengineering and Biotechnology*. 2019; 7(358).
29. Johnson WE, Li C and Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8(1):118-127.
30. Bourgon R, Gentleman R and Huber W. Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(21):9546-9551.
31. Benjamini Y, Drai D, Elmer G, Kafkafi N and Golani I. Controlling the false discovery rate in behavior genetics research. *Behavioural brain research*. 2001; 125(1-2):279-284.
32. Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008; 9:559.
33. Cheng L, Liu P, Wang D and Leung KS. Exploiting locational and topological overlap model to identify modules in protein interaction networks. *BMC bioinformatics*. 2019; 20(1):23.
34. Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*. 2012; 16(5):284-287.
35. Cheng L, Wang P, Tian R, Wang S, Guo Q, Luo M, Zhou W, Liu G, Jiang H and Jiang Q. LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic acids research*. 2019; 47(D1):D140-D144.
36. Yi Y, Zhao Y, Li C, Zhang L, Huang H, Li Y, Liu L, Hou P, Cui T, Tan P, Hu Y, Zhang T, Huang Y, et al. RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res*. 2017; 45(D1):D115-D118.
37. Jeggari A, Marks DS and Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012; 28(15):2062-2063.
38. Wong N and Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic acids research*. 2015; 43(Database issue):D146-152.
39. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic acids research*. 2018; 46(D1):D296-D302.

40. Wang H, Ke J, Guo Q, Barnabo Nampoukime KP, Yang P and Ma K. Long non-coding RNA CRNDE promotes the proliferation, migration and invasion of hepatocellular carcinoma cells through miR-217/MAPK1 axis. *Journal of cellular and molecular medicine*. 2018; 22(12):5862-5876.
41. Li H, Li Q, Guo T, He W, Dong C and Wang Y. LncRNA CRNDE triggers inflammation through the TLR3-NF-kappaB-Cytokine signaling pathway. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2017; 39(6):1010428317703821.
42. Chen Y, Fu Y, Song YF and Li N. Increased Expression of lncRNA UCA1 and HULC Is Required for Pro-inflammatory Response During LPS Induced Sepsis in Endothelial Cells. *Frontiers in physiology*. 2019; 10:608.
43. Liu X, Hong C, Wu S, Song S, Yang Z, Cao L, Song T and Yang Y. Downregulation of lncRNA TUG1 contributes to the development of sepsis-associated acute kidney injury via regulating miR-142-3p/sirtuin 1 axis and modulating NF-kappaB pathway. *Journal of cellular biochemistry*. 2019.
44. van der Poll T. Future of sepsis therapies. *Critical care*. 2016; 20(1):106.
45. Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, Skogerbo G, Wu Z and Zhao Y. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic acids research*. 2011; 39(9):3864-3878.
46. Pankla R, Buddhisa S, Berry M, Blankenship DM, Bancroft GJ, Banchereau J, Lertmemongkolchai G and Chaussabel D. Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome biology*. 2009; 10(11):R127.
47. Venet F, Schilling J, Cazalis MA, Demaret J, Poujol F, Girardot T, Rouget C, Pachot A, Lepape A, Friggeri A, Rimmelé T, Monneret G and Textoris J. Modulation of LILRB2 protein and mRNA expressions in septic shock patients and after ex vivo lipopolysaccharide stimulation. *Human immunology*. 2017; 78(5-6):441-450.
48. Tabone O, Mommert M, Jourdan C, Cerrato E, Legrand M, Lepape A, Allaouchiche B, Rimmelé T, Pachot A, Monneret G, Venet F, Mallet F and Textoris J. Endogenous Retroviruses Transcriptional Modulation After Severe Infection, Trauma and Burn. *Frontiers in immunology*. 2018; 9:3091.
49. Sutherland A, Thomas M, Brandon RA, Brandon RB, Lipman J, Tang B, McLean A, Pascoe R, Price G, Nguyen T, Stone G and Venter D. Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis. *Critical care*. 2011; 15(3):R149.

Figures

Figure 1

Workflow for the identification of sepsis lncRNA. DEGs, differentially expressed genes.

Figure 2

Module identification. (A) Identification of co-expression modules from the topological overlap matrix of GSE65682 using WGCNA. (B) Heatmap shows the reproducibility of two module lists. Rows represent modules identified from GSE65682 while columns represent modules from GSE69528. Grid color corresponds to $-\log_{10}$ transferred hypergeometric test P-values of the overlap of two modules. (C) Venn diagram of the identified modules from the two datasets. (D) Overview of the sepsis modules, including up-regulated DEMs, down-regulated DEMs, and SAMs. DEM, differentially expressed modules; SAMs, survival associated modules.

Figure 3

Sepsis module identification. (A) Volcano plot shows the differentially expressed genes. (B and C) Functional analysis of the up- and down-regulated DEGs, respectively. (D) Kaplan–Meier curves of two patient groups with higher or lower EG value for module 32. (E) EG score distribution and the survival time and vital status of detected patients. The dotted line represents the median EG score dividing patients into two groups of high (red) and low risk (light green). (F) Heatmap of the expression profiles for the genes in module 32. Rows represent genes while columns represent patients.

Figure 4

Identification of sepsis modules. A) A co-expression module enriched of up-regulated DEGs. Vertexes correspond to genes and edges correspond to expression correlation. Only the edges with the absolute value of PCC greater than 0.5 are shown. Up-regulated DEGs are colored in red while down-regulated DEGs are colored in blue. (B) Kaplan–Meier curves of module 22. (C) Enriched GO biological processes of Module 22. (D) A co-expression module enriched of down-regulated DEGs. (E) Kaplan–Meier curves of module 47. (F) Enriched GO biological processes of Module 47. (G) The difference of mean correlation coefficient between DEGs and the other genes in different modules.

Figure 5

Candidate sepsis lncRNAs. (A) Overview of the lncRNA-module network. The bottom panel illustrate the subnetwork of candidate sepsis lncRNAs and interacting modules. Up-regulated DEMs are colored in red, down-regulated DEMs are colored in blue, and SAMs are framed in red circle. The node size of modules correspond to module size. (B) Module connectivity of lncRNAs. The pie plot indicates the proportion of lncRNAs with different connectivity. The bar plot shows the lncRNAs linking more than one module. (C) ceRNA regulatory mechanism in sepsis. (D) ceRNA interactions. Red and blue lines indicate lncRNA-miRNA and miRNA-module interactions, respectively. Bold gray line represents the competing endogenous relationship. (E) Secondary structure of FENDRR, MALAT1, TUG1, and CRNDE. (F) Description of the sepsis lncRNA candidates. ceRNA, competing endogenous RNA. MREs, miRNA response elements. DEMs, differentially expressed modules. SAMs, survival associated modules.

Figure 6

Validation of the candidate sepsis lncRNAs in three independent datasets. The four candidate sepsis lncRNAs, FENDRR, MALAT1, TUG1, and CRNDE, were detected as differentially expressed in at least one dataset.

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.xlsx](#)