Transcriptomic study of sex differences in sepsis

Huachao Xu  
Cheeloo College of Medicine, Shandong University

Qi Chen  
The First Affiliated Hospital of USTC, University of Science and Technology of China

Yuanhao Shen  
Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine

Jun Lu (✉ lujun_sjtu@163.com)  
Cheeloo College of Medicine, Shandong University

Research Article

Keywords: sex difference, transcriptomic analysis, sepsis

Posted Date: November 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1994561/v1

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

BACKGROUND AND OBJECTIVES: There are significant differences in sepsis incidence and mortality between men and women, but the pathogenesis has remained unclear. The purpose of this study is to compare sex related expression at the transcriptional level in sepsis, and to explore potential target genes for further research.

METHODS: The train dataset GSE65682 from the Gene Expression Omnibus (GEO) database was utilized for bioinformatic analysis. Four subgroups were divided according to different infection types such as CAP, HAP, AS and HC. By analyzing the differentially expressed genes (DEGs) in each subgroup between men and women, we got a list of sex related DEGs list. We further analyzed the expression levels of these genes between sepsis and healthy control groups to find the genes of interest (GOIs) with significant changes. Finally the validation dataset GSE134364 was used to verify the changes of GOIs.

RESULTS: 9 genes were found by comparing the 4 subgroups DEGs of train dataset under the standard of $|\log_2 (FC)| > 0.5$ and adj P.value$<0.05$. In the following expression analysis we got 3 GOIs, which were distributed on the Y chromosome and their expression was significantly down regulated in sepsis compared with healthy control group. These changes of GOIs were further verified by the validation set.

CONCLUSIONS: The sex-related differences of sepsis in large samples were mainly manifested in sex chromosome at the transcriptomic level. The expressions of TXLNGY, USP9Y and PRKY located in the male Y chromosome were down-regulated significantly in different sepsis types, which might affect their prognosis in sepsis.

1. Introduction

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection[1]. Due to different hosts may present different responses to the same infection, sepsis is a heterogeneous syndrome, and it is essential to identify clinically relevant subphenotypes. Many studies have found that sexual dimorphism, which influences the host immune response, is an important characteristic of sepsis[2-4]. Sexual dimorphism could be approached as a first step in the personalized management of septic patients[5].

However, a large number of observational clinical studies investigating sex dependent differences in sepsis have reported conflicting evidence. McGowan et al[6] first reported a higher incidence of sepsis in men than women admitted to Boston City Hospital between 1935 and 1972. Subsequently some large epidemiologic studies of sepsis[7-10] corroborated this discrepancy. In terms of sex-dependent mortality from sepsis and septic shock, conflicting evidence from numerous individual studies has been reported. Some have reported that males experienced a higher mortality rate than females[11, 12]. Others report the opposite—that females experienced higher mortality[8, 13-15]. However in a recent study[16] of global analysis of sepsis incidence and mortality for the years 1990 to 2017 shows that incidence of women is
higher than that of men about 15.2 percent as a whole, however the total mortality of men is higher than women about 12.2 percent. The reasons for this difference are worthy of further study.

In order to explore the difference of sex dependence in sepsis, most of the studies are related to sex and gender. The sex of an individual is defined by the differential organization of chromosomes, reproductive organs, and sex steroid levels; it is distinct from gender, which includes behaviors and activities that are determined by society or culture in humans\textsuperscript{[17]}. These studies also provide some evidence for explaining the heterogeneity of sepsis. With the development of sequencing technology, more and more patients with sepsis have genetic testing during treatment. So the availability of numerous transcriptomic profiling databases provides unprecedented opportunity to use different types of statistical methods to discover the secret behind the sexual veil. These datasets also allow extensive external validations. Because these two datasets are the larger databases for sepsis research at present, and they are also widely used and cited, the research results have obvious significance. The aims of this study is to compare sex related expression at the transcriptional level in sepsis, and to explore potential target genes for further research.

2. Methods

2.1. Data selection and processing

The sepsis RNA expression data was downloaded from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). Data from the GSE65682\textsuperscript{[18]} (Platforms: GPL13667, Affymetrix Human Genome U219 Array) was used as a train set for differential expression analysis. This dataset was part of the Molecular Diagnosis and Risk Stratification of Sepsis (MARS) project, conducted between January 2011 and December 2013 in the mixed ICUs of two tertiary teaching hospitals in the Netherlands.\textsuperscript{[19-22]} Total 802 samples were collected, including 760 intensive care unit (ICU) patient samples with sepsis and 42 healthy controls. We extracted four subgroups such as community acquired pneumonia (CAP), hospital acquired pneumonia (HAP), abdominal sepsis (AS), and healthy control (HC) for differentially expressed genes (DEGs) analysis (Table 1). An advantage of this method is that normalization occurs at the probe level (rather than at the probeset level) across all of the selected hybridizations\textsuperscript{[23]}.

GSE134364\textsuperscript{[24]} (Platforms: GPL17586, [HTA-2.0] Affymetrix Human Transcriptome Array 2.0) was also a big sepsis dataset downloaded from GEO which contained total 334 samples, including 215 samples with sepsis and 83 healthy controls. This dataset was used as a validation set for genes of interest (GOIs) expression analysis.

Table 1. Characteristics of each subgroups in GSE65682 dataset.
<table>
<thead>
<tr>
<th>Group</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n.</td>
<td>Age(y)</td>
</tr>
<tr>
<td>CAP</td>
<td>63</td>
<td>60.8±14.9</td>
</tr>
<tr>
<td>HAP</td>
<td>54</td>
<td>61.6±15.2</td>
</tr>
<tr>
<td>AS</td>
<td>27</td>
<td>63.2±9.3</td>
</tr>
<tr>
<td>HC</td>
<td>18</td>
<td>48±22.3</td>
</tr>
</tbody>
</table>

These datasets were selected from other datasets with reference to the following criteria: i) dataset providing data information about gender and age; ii) the presence of corresponding controls in the same dataset (healthy individuals or individuals scheduled for elective surgery); iii) available processed data and iv) samples collected within 24 h after ICU admission. The exclusion criteria include: i) animal and pediatric studies and ii) datasets with small number of samples.

### 2.2. Identification of sex related DEGs

R version 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria) was used for data extraction and sorting to obtain the gene expression matrices. To obtain differentially expressed genes (DEGs), the limma package (3.52.1) in R was used to identify DEGs of man samples versus woman samples in the four subgroups of the GSE65682 transcriptome data. The adjusted P-values (adj.P.value) and Benjamini and Hochberg false discovery rate were applied to provide a balance between discovery of statistically significant genes. |log2-fold change (FC)| >0.5 and adj.P.values <0.05 were considered statistically significant. The Venn App (Origin 2022, OriginLab Corporation, Northampton, USA) was used to overlap the lists of sex related differentially expressed genes (DEGs) from the four subgroups. Then, a set of sex related DEGs was selected based on those overlapping. In addition, we selected only genes classified as “protein-coding” for further analysis.

### 2.3. Correlation enrichment analyses of DEGs

All sex related DEGs were selected for enrichment analysis to reveal their biological function and signaling pathways. Gene Ontology (GO) term enrichment analysis was acquired from the the PANTHER classification system[25] (http://www.pantherdb.org), searching the following categories: GO Biological Process(BP), GO Molecular Function(MF), GO Cellular Component(CC), and PANTHER pathway[26]. FDR<0.05 was considered statistically significant. GO annotation is a main bioinformatics tool to annotate genes and analyze biological process of DEGS. The PANTHER is a unique resource that classifies genes into canonical pathways to predict function[27].
2.4. Statistical Analysis of Gene expression in subgroups

In order to further study the expression change of sex related DEGs in subgroups, we extracted the mean expression values of males and females in CAP, HAP, AS and HC individually and form a distribution heat map. The mean expressions were extracted by the R package “tidyverse”. The heat map and cluster analysis of differential gene distribution were performed by the R package “pheatmap”. One-way ANOVA was used for inter group test, and Turkey test was used for mean comparison. Log2FC is obtained by calculating the average expression ratio of genes in sepsis group and healthy control group and taking the logarithm based on 2. P value < 0.05 were considered statistically significant. Sex related DEGs of $|\text{log}_2 (\text{FC})| > 0.4$ were our genes of interest (GOIs). Statistical analysis was performed with Origin 2022 (OriginLab Corporation, Northampton, USA).

2.5. Gene Set Enrichment Analysis

GSEA was performed using the R package clusterProfiler in the four groups of AS, CAP, HAP and HS to discover the significant functional difference between male and female. Significant pathway enrichment was identified by the normalized enrichment score ($|\text{NES}| > 1$), P value <0.05, and FDR q value <0.05[28]. The top five terms of HALLMARK analyses from MSigDB were exhibited respectively. Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used for graphical depiction of the unions, intersections and distinctions among different group HALLMARK gene sets, and a list of those intersections in all groups and sepsis groups was generated.

2.6. Validation GOIs

The dataset GSE134364 was used for validation analysis. First, the expression quantities related to the list of sex related DEGs were selected, and then the dataset was divided into four groups according to sex difference in sepsis and healthy group. One-way ANOVA is performed for each group, P values < 0.05 were considered statistically significant.

2.7. Expression of GOIs in normal tissues

RNA-seqs of GOIs were download from HPA RNA-seq normal tissues project in NIH (National Library of Medicine, USA) (https://www.ncbi.nlm.nih.gov/gene). RNA-seq was performed of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes.

3. Results

3.1. Identification of sex related DEGs in sepsis
After standardization of the microarray results, sex related DEGs in GSE65682 were identified. The DEGs of these 4 subgroups showed in Volcano diagram (Fig. 1a). The overlap among the 4 datasets contained 7 genes as shown in the Venn diagram (Fig. 1b), Groups of CAP, HAP and AS are the same 7 genes (6 up-regulated and 1 down-regulated), which are included in the 98 up-regulated and 1 down-regulated genes of HC group. We take their sum aggregate for further study.

3.2. Cluster and Enrichment analyses of DEGs

The mean expressions of DEGs in each subgroup were showed in the heat map (Fig. 2a). Gene ontology functional enrichment analysis results showed that changes in biological processes (BP) of DEGs were significantly enriched in cellular process (GO:0009987), reproductive process (GO:0022414), localization (GO:0051179), etc. Changes in molecular function (MF) were mainly enriched in translation regulator activity (GO:0045182), binding (GO:0005488), structural molecule activity (GO:0005198), ATP-dependent activity (GO:0140657) and catalytic activity (GO:0003824). Changes in Cellular Component (CC) were mainly enriched in cellular anatomical entity (GO:0110165) and protein-containing complex (GO:0032991) (Fig. 2b). There were 11 pathways we got by PANTHER pathways, such as muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043), metabotropic glutamate receptor group III pathway (P00039), histamine H2 receptor mediated signaling pathway (P04386), etc. (Fig. 2c).

3.3. Expression Analysis of DEGs

The expressions of the sex related DEGs in the 4 groups were displayed by male and female separately. (Fig.3) Further analysis of the inter subgroups revealed the expressions of some genes were significantly changed in sepsis groups (CAP, HAP and AS) compared with healthy control in male groups. (Table.2) The GOIs those we got by the standard of P value < 0.05 and |log2 (FC)| > 0.4 were TXLNGY, USP9Y and PRKY.

The arrow direction indicates the up or down-regulated change. The number of arrows indicates one or more groups in CAP, HAP and AS are significantly different from the HC group in the same gender. ns: not significant.

3.4. Identification of Hallmark Gene Sets by GSEA

GSEA was performed to understand the functional differences between male and female in each group. All differentially expressed genes were included in the GSEA. We identified many significant hallmark gene sets in the enrichment of MSigDB Collection (V7.5.0)[29], including inflammatory response, oxidative phosphorylation, interferon gamma response etc. (Fig.4. a). The Venn diagram approach revealed there were 6 intersections gene sets in all four groups, and 2 in sepsis groups (AS, CAP and HAP). (Fig.4. b)
3.5. Validation of Gene Expression Levels

GSE134364 was used to validate the afore mentioned changes in DEGs expression. One-way ANOVA was performed to compare gene expression level in each group. We also found that TXLNGY, USP9Y and PRKY three genes have significant differences compared with other 6 genes.(Fig.5)

3.6. Expression of GOIs in normal tissues.

By retrieving the HPA RNA-seq normal tissues project in NIH, we obtained the expression data of GOIs, which were all distributed on the Y chromosome. In human 27 different tissues, TXLNGY are highly expressed in prostate and small intestine, USP9 is highly expressed in prostate and testis, and PRKY is highly expressed in skin and prostate.(Fig.6)

4. Discussion

The sex disparity of sepsis is likely explained by a combination of biological sex differences and gender-specific factors. Many studies\cite{30-33} focus on sex hormones, such as estrogen, progesterone, androgen, etc. These hormone levels are significantly different between men and women, especially when women have menstruation. Different levels of these hormones have different effects on the outcome of sepsis\cite{34}. Our study find almost no difference in hormone levels at the transcriptome level. The analysis result might due to sample selection bias. We can find that the average age of both men and women is over 60 years old as shown in table 1. Although this age characteristic of patients in ICU (mainly the elderly) is similar to many statistical samples\cite{35, 36}, it still causes some statistical bias. The stratification of age groups may be an exceptionally effective strategy for solving this problem.

The strengths of our study include the relatively large sample size and generalizability. The train dataset covers most sepsis such as pneumonia and abdominal infection, and has complete gender and other related data\cite{22}. The detection platform is unified, which is convenient for statistics and reduces interference factors. The validation set is also a large sample. The data collected at one level avoids the batch effect and is more conducive to draw reliable conclusions from the analysis results.

To obtain valuable genes which could better explain the sex disparity of sepsis, we had made detailed grouping according to the disease characteristics, and subsequently got sex related DEGs. The heatmap helped us know these differential genes have different expression in the subgroups. After further quantitative analysis, we found three genes with significant differences: TXLNGY, USP9Y and PRKY, and these three genes are all located on the Y chromosome. GSEA results showed that even in all groups, including healthy controls, sex related DEGs were significantly enriched in inflammatory response, INF-α response and INF-γ response. In sepsis groups (CAP, HAP and AS), oxidative phosphorylation and MYC targets related hallmark were more obvious. From this perspective GSEA confirmed that the differential
genes between males and females had primitive different responses to infection. The following validation set also corroborated that the expression of these genes in sepsis was significant.

It seems that the gene changes in sepsis are mainly concentrated on the Y chromosome and the GOIs are most express in male organs in our study. In fact, the X chromosome is more enriched with the changes of the body's natural immunity, Many genes on the X chromosome code for proteins ranging from PRRs to cytokine receptors and transcriptional factors. The Y chromosome is also found that its polymorphism changes will be more susceptibility to some infections. It is unknown whether the changes of these genes affect the progress and prognosis of sepsis or whether sepsis affects the changes of these sex chromosome genes. The pathogenesis of this difference and the effect on mortality are less clear. Despite the long-held belief that this finding is related to sex hormones, the current study and a preponderance of other evidence suggest otherwise. When it comes to sex and sepsis, it's time we look beyond our hormonal differences.

There are still some limitations and unanswered questions from this study. The number of samples is still limited, failing to include more countries, nationalities and races. The sex related DEGs genes are worth of further verification, and the mechanism of fold change is still unclear. LncRNAs or miRNAs may also participate in the gender differential expression of sepsis, which still needs further verification.

5. Conclusion

In the large sample study, it was found that the sex bias of sepsis was mainly reflected in the difference of gene expression in sex chromosomes. The expressions of TXLNG, USP9Y and PRKY located in the male Y chromosome were down-regulated significantly in different sepsis types, which might affect the sex difference in the prognosis of sepsis at the level of gene transcription.

Abbreviations

AS = abdominal sepsis
CAP = community acquired pneumonia
DEGs = differentially expressed genes
GEO = gene expression omnibus
GO = gene ontology
GOIs = genes of interest
GSEA = gene set enrichment analysis
HAP = hospital acquired pneumonia
HC = healthy control

ICU = intensive care units

KEGG = Kyoto encyclopedia of genes and genomes

MSY = male-specific region of Y chromosome

Declarations

Author contributions

Jun Lu contributed to the conception of the study; Huachao Xu contributed to analysis and manuscript preparation; Qi Chen performed the data analyses and wrote the manuscript; Yuanhao Shen helped perform the analysis with constructive discussions.

References


Tables
Table 2 and 3 are available in the Supplementary Files section.

**Figures**

**Figure 1**

Distribution of sex related DEGs in each subgroup of GSE65682.

a. Volcano map for the DEGs of CAP, HAP, AS and HC subgroup, |log2 (FC)| >0.5 and adj.P.values <0.05 were considered statistically significant. b. Venn diagram of the DEGs in 4 subgroups and the distribution details. (Red indicates up-regulated genes and blue indicates down-regulated genes)
Sex related DEGs cluster and enrichment analysis.

a. The heat map of sex related DEGs mean expression in male and female of the four groups. b. Gene ontology functional enrichment analysis in the PANTHER classification system. c. 11 pathways of DEGs in the PANTHER pathway. M refers to male and F refers to female in each group.
Figure 3

Expression analysis of sex related DEGs in each subgroup. M refers to male and F refers to female in each group.
Figure 4

Hallmark gene sets in each group.

a. Top 5 hallmark gene set in the group of AS, CAP, HAP and HC. b. Venn diagram shows the intersections and distinctions among different group.
Figure 5

Expression analysis of sex related DEGs in validation dataset GSE134364. M refers to male and F refers to female in each group.

Figure 6

RNA-seq in Normal Tissues

<table>
<thead>
<tr>
<th>TXLINGY</th>
<th>USP9Y</th>
<th>PRKY</th>
</tr>
</thead>
<tbody>
<tr>
<td>adrenal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>appendix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bone marrow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>brain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>colon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>duodenum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>esophagus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gall bladder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>heart</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kidney</td>
<td></td>
<td></td>
</tr>
<tr>
<td>liver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lymph node</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ovary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pancreas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>placenta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prostate gland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>salivary gland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>small intestine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spleen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stomach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>testis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>thyroid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>urinary bladder</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Expression of *TXLNGY, USP9Y* and *PRKY* in normal tissues. (data from HPA RNA-seq normal tissues project)

**Supplementary Files**

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

- Table2.jpg
- Table3.jpg