Author response to reviewer comments v.1

Reviewer: 1

Comments to the Author

1.Proof reading is needed to correct grammar and scientific writing, such as the department of the Corresponding authors and so on.

Respond: We carefully checked our manuscript and fixed these errors.

2.This paper focused on sepsis-induced ARDS, what about other ARDS?

Respond: Thanks for this comment to help us with further research. Sepsis is one of the causes of ARDS. The focus of the present study is to explore the underlying molecular mechanisms of ARDS caused by sepsis. Therefore, whether the ARDS caused by other reasons has the same mechanism still needs further verification.

Reviewer: 2

Comments to the Author

Therapeutic Advances in Respiratory Disease Review

“NF-kappa B signal transduction through the NFKB1-GSK3B

bridge and a series of inflammatory pathways that mediate sepsis-associated ARDS”

Zhang et al have studied sepsis-induced ARDS and lung injury using publicly-available microarray datasets from the GEO repository. To minimize batch effects, they used batch correction and used PCA analysis to evaluate data quality. The authors proceeded to perform differential expression and GSEA enrichment and constructed a weighting model using forest tree, STRING and functional enrichment methods to identify pathways. Finally, they used the "TRRUST" database to identify regulatory interactions. The authors found enrichment of immunological genes. The genes enriched for cell-substrate adherens junction and genes involved in the NF-kappa B signal transduction. Overall, this is an interesting paper that adds new knowledge to the field and to understanding genomic basis for ARDS due to sepsis. Issues that limit enthusiasm and need to be addressed are listed below.

1. The concept is appealing, however, the strategy and the methods are confusing without the clear rationale for the different technologies that were applied.

Respond: Thanks for this comment to help our manuscript writing. We almost rewrote the method part. The statement of the method is detailed. A flow chart and detailed description were provided. In the present study, the further dysregulated genes (FDGs) were identified in the process evolving from healthy control through sepsis to sepsis-induced ARDS. The protein-protein interaction networks of FDGs was constructed and used to modular analysis. Functional enrichment analysis was performed for the functional modules. Hypergeometry test was proceeded to identify potential transcription factors of modules. A TF-module-pathway global transcriptional regulatory network was constructed. ROC analysis was used to access the potential diagnostic markers

2. The authors used five different expression arrays, four from Affymetrix (GSE76293 HG-U133_Plus_2, GSE66890 HuGene-1_0-st, GSE10474 HG-U133A_2, GSE10361 HG-U133A) and one from Illumina (GSE32707 GPL10558 Illumina HumanHT-12). It is unclear if the arrays were combined or analyzed independently. This needs to be clarified. The section of the batch effect describes using "combit package" a package that does not exist in any of the R repositories. Furthermore, the details on how the data was analyzed were not described in detail or missing, like the case of the GSE32707 arrays.

Respond: Thank you very much for this comment, which led us to reconsider the study and almost revised the manuscript entirely. The data set in the original analysis were not really from the same platform, so it is unreasonable to merge them and remove batch effects. Consequently, we redesigned the study and filtered most of the samples, and only GSE32707 was retained.

In addition, "combat" is a function in "sva" package (https://www.bioconductor.org/packages/release/bioc/html/sva.html), not a package. In our revised manuscript, we did not use it because there were no data sets based on the same platform.

3. There are variabilities between the datasets that can't be overlooked for example, GSE76293 originated from polymorphonuclear neutrophils (PMNs) from bronchoalveolar lavage and blood, while the rest of the samples are from whole blood. The samples were from 2007 to 2016 at multiple centers, and the potential differences in diagnosis between centers need to be discussed.

Respond:

1) We redesigned the study and filtered most of the samples, and only GSE32707 was retained.

2) Only GSE32707 was retained in the revised manuscript, so we may not need to discuss diagnostic differences.

4. What phenotypes were compared for each dataset and analysis? What phenotypes were used for the GSEA and differential expression? This requires clarification.

Respond:

1) Compared to the control samples, the DEGs in sepsis samples and sepsis-induced samples were respectively screened using the limma package in R.

2) "Sepsis vs control" and "Sepsis-induced ARDS" as phenotypes for GSEA analysis separately.

5. There is inconsistency with p-values described as P-value < 0.05, in others as p-value < 5% and in other instances they used FDR, but the FDR was not present in the tables (adj.P.Val is not necessarily FDR, need to see the commands used in limma). During differential expression analysis, a threshold that combines FDR and fold change is usually utilized. For example, an absolute value of logFC of > 1 and FDR of < 0.01 is commonly used for gene expression. The tables values of logFC < 1.0 are listed as differentially expressed whereas they are not.

Respond: We redesigned the study. Genes with |log2FC|>1 and P adjusted by the false discovery rate (FDR) < 0.05 were considered as differentially expressed genes.

6. The random forest analysis general information about the method, but it is not clear what was the input for the algorithm. An input table is necessary.

Respond: In our redesigned study, we did not use random forest analysis in our present study.

7. The authors need to provide the software used. They indicate using the method “train_test_split” for the training and testing, did they used the python scikit-learn package or some other software? The same critique is for the protein-based network identification and the pathway-enriched method. Figure 1 helps to understand the general approach, but the method section should be rewritten to provide all the scripts and tables used at each step in a GitHub repository.

Respond: In our redesigned study, we rewrote the method section and explained the software and parameters used in the analysis in detail. So the scripts and tables may be not necessary to upload to GitHub.