Int J Hum Genet, 18(2): 68-75 (2018)
DOI: 10.31901/24566330.2018/18.2.689

Identification of Hub Genes and Pathways in Acute Respiratory Distress Syndrome Based on Differential Expression Network

Xin Ni, Ren-Zhi Yu, Lei Han, Li-Li Zhao, Yu-Feng Lyu, Li-Yan Sun and Mei Yuan

Respiratory Medicine, Hongqi Hospital of Mudanjiang Medical College, Mudanjiang, 157011, Heilongjiang Province, China

KEYWORDS Differential Expression Network. Differential Interactions. Hub Genes. Non-differential Interactions

ABSTRACT This study aimed to investigate significant genes and pathways in acute respiratory distress syndrome (ARDS) based on differential expression network (DEN). Firstly, differentially expressed genes (DEGs) between ARDS samples and normal controls were identified using linear models for microarray data (Limma) package. Secondly, differential interactions and non-differential interactions were selected to construct a DEN. Thirdly, topological centrality analysis (degree) was performed for exploring hub genes. Finally, to investigate functional biological processes related to hub genes, pathway enrichment analysis was conducted. A total of 229 DEGs were detected for ARDS. There were 759 nodes and 576 edges which included 550 differential edges and 26 non-differential interactions. By accessing topological centrality analysis, the researchers obtained 6 hub genes, *UBC*, *CSNK2A2*, *CUL5*, *SOX2*, *PARK2* and *CHAF1A*. The most significant pathway enriched by hub genes was T cell receptor signaling pathway. The hub genes and pathways would be potential biomarkers for ARDS diagnosis and treatment.

INTRODUCTION

Acute respiratory distress syndrome (ARDS), which induced by alveolocapillary inflammation and increased permeability (Hoeboer et al. 2015), is a life-threatening common disorder of urgent onset pulmonary failure with high morbidity and mortality requiring critical care in clinically patients (Yadav et al. 2017). Its mortality remains high ranging between 27 percent and 45 percent (Villar et al. 2014), in which adult has mortality rates between 45 percent and 55 percent (Sud et al. 2010). Currently, there are some measures to deal with ARDS, for example, prone positioning (Guerin et al. 2013), neuromuscular blockade (Papazian et al. 2010) and extra-corporeal therapies (Fitzgerald et al. 2014). But few research dig the causes of ARDS at the molecular biology level and provide a more effective prevention and treatment for it on fundamentally. Therefore, the search for accurate biomarkers reflecting the severity and course of ARDS is crucial.

Address for correspondence:
Mei Yuan
Respiratory Medicine
Hongqi Hospital of Mudanjiang Medical College,
No.5 on Tongxiang Road, Aimin District,
Mudanjiang, 157011,
Heilongjiang Province, China
Telephone & Fax: 086-0453-6582800
E-mail: meiyuan201604@sina.com

A majority of the genome-related studies aim at the verification of differentially expressed genes (DEGs) that are dramatically different in different conditions, or the differential network (DN) that extract disease related edges in occurred interactions across different static networks. However, there is rarely research about the new molecular network type involved in characterization alterations on a specific condition, that is, the differential expression network (DEN). The DEN not only covers DEGs and DN, but also includes non-differential interactions which are missed in DN. It can better describe phenotype differences from the network viewpoint in contrast to the traditional DEGs and DN methods. Description of the alterations in DEN present in association to a certain phenotype at gene, interaction, pathway, and even global network structure level (Sun et al. 2013), and provide additional information to determine the genes that could have a central role in disease. The DEN fully explored all disease-related interactions, thus it would offer a novel way to predict disease genes and disease interactions in an accurate manner. Hence, the researchers investigated significant genes and pathways in ARDS based on DEN. Six hub genes and one pathway which possess the potential to be biomarkers for ARDS diagnosis and treatment were identified in this study.

Objectives

In order to gain a system-wide understanding for ARDS, in this paper, after screening the DEGs between ARDS samples and normal controls, the researchers developed a new frame work by constructing a DEN, whose key step was the identification of differential interactions and non-differential interactions by spearman correlation coefficients (SCC) algorithm based on DEGs and protein-protein interaction (PPI) network downloaded from Biological General Repository for Interaction Datasets (BioGrid), in two conditions. Subsequently, to investigate the functional significant genes and biological processes of ARDS, topological centrality analysis of DEN and pathway enrichment analysis were carried out. The study contributed to the validation of potential therapeutic targets among the selected genes and the relevant networks.

METHODOLOGY

Collecting Data

The researchers collected the gene expression profile under the access number E-MEXP-3621 (Vassiliou et al. 2013) from ArrayExpress database. E-MEXP-3621, which consisted of 4 ARDS samples and 6 normal controls, existed in Affymetrix GeneChip Human Genome U133A 2.0 Platform. Before analysis, the original expression information from all conditions was carried on data preprocessing. Background correction and normalization were carried out in order to eliminate the influence of nonspecific hybridization via robust multichip average (RMA) method (Ma et al. 2006) and quantile based algorithm (Rifai and Ridker 2001), respectively. Perfect match and mismatch value was revised by Micro Array Suite 5.0 (MAS 5.0) algorithm (Pepper et al. 2007), the expression value was selected using the medianpolish. The data were screened by feature filter function of genefilter package. Each probe is mapped to one gene by getSYMBOL. The researchers eliminated the probe sets without corresponding official symbol. Finally, a total of 12493 genes were screened for further consideration.

Identifying DEGs

It is well confirmed that the propensity of many diseases can be reflected in the difference of gene expression levels (Zhao et al. 2011). Hence genes showed a different expression levels between control crowds and case groups are likely related to the disease. To get a more powerful and less subject to bias, multiple testing was employed via linear models for microarray data (Limma) package in this work (Smyth 2004). Here the T-test and F-test were utilized to identify differentially expressed genes, and then the P-values were transformed to -log10. Empirical Bayes (Datta et al. 2004) (eBayes) statistics and a false discovery rate (FDR) (Reiner et al. 2003) calibration of P-values for the data were conducted by lmFit function. DEGs were identified for further research with the threshold of $P < 0.05$ and $|\log_2 FoldChange| > 2$.

Constructing DEN

Generally, genes can't work alone, thus the candidate interactions were detected from the network perspective. If some genes had a closely connections with ARDS, it was reasonable to assume that they had some closely interactions. The researchers constructed the DEN by extracting the interactions which were capable of characterizing the initiation and progression of ARDS based on differential interactions and non-differential interactions. A differential interaction was an edge which had significantly changed under different conditions, and a non-differential interaction was an edge which had no significant change but its linked two genes or corresponding coded proteins were both differentially expressed.

Calculating the SCC

SCC was a popular method to describe the interaction between genes. Thus, in this study, SCC was used to present the reweighted value of each interaction (between -1 and 1, negative value represents the negative correlation, positive value as on behalf of the positive correlation). First, all PPIs of humans were downloaded from BioGrid, (http://thebiogrid.org/) (Chatr-Aryamontri et al. 2015), which includes a total of 15750 genes and 248584 interactions, and the gene pairs from the gene expression data were extracted for further analysis. The genes expression value in the control sample and disease group were computed and extracted, respectively. The SCC of PPI pairs in two conditions (A1 meant

control sample and A2 represented disease group) and its absolute value of the differences (|A1-A2|) between two groups were calculated.

Determining the Threshold of Differential and Non-differential Interactions

In order to determine how to choose the gene relationships for further research, here the researchers built two models (one for case, the other for control) randomly, and each model contained 200,000 PPI pairs respectively. The SCC of PPI pairs and its absolute value of differences (|A1-A2|) in two groups were calculated. The value of |A1-A2| was ranked in descending order. When setting P = 0.01, the absolute value of SCC differences in two condition was 1.743. Thus, the researchers selected the differential interactions with |A1-A2|> 1.743 and one of A1, A2 showed strong correlation (<0.7). Meanwhile, the non-differential interactions were screened if an interaction had |A1-A2| < 1.743 and the both two corresponding nodes belonged to DEGs.

To sum up, the DEN was finally constructed by incorporating all of the differential interactions and non-differential interactions. The network was performed using the Cytoscape 2.1 software.

Centrality Analysis

In the present study, in order to obtain the hub genes of ARDS, the researchers utilized the centricity analysis (Scardoni and Laudanna 2012) which was useful to identify key players in biological processes for the DEN. Common centrality measures mainly contained degree centrality, closeness centrality and shortest path betweenness centrality, in which degree was the simplest topological index. The degree of a node (gene or protein) is the average number of edges (interactions) incident to this node. Nodes with high degree were called "hubs" which was related to several other genes, suggesting a central role in the interaction network. An obvious order of the vertices of a graph can be established by sorting them according to their degree (Koschützki and Schreiber 2008). The degree of the genes in the network > 9 were defined as hub genes in this study.

Pathways Enrichment Analysis

To further assess the signaling pathway of the DEGs and nodes in the DEN, the researchers performed a pathway analysis based on Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The KEGG database is a collection of manually drawn pathway maps for metabolism, genetic information processing, environmental information processing such as signal transduction, various other cellular processes and human diseases (Kanehisa and Goto 2000). The DEGs and nodes in the DEN were applied to this database in order to investigate the biochemistry pathways that might be involved in the occurrence and development of ARDS. The analysis was performed using the online tool of Database for Annotation, Visualization and Integrated Discovery (Huang et al. 2008) (DAVID). The researchers got the significantly enriched pathways according to the threshold of empirical p-value of 0.05.

RESULTS

Identifying DEGs

In this study, under the criteria of $|\log_2 \text{FoldChange}| > 2$ and $P < 0.05$, the researchers obtained 229 DEGs, of which 96 were up-regulated and 133 were down-regulated genes, between ARDS samples and normal subjects by Limma package.

Constructing DEN

A total of 248584 PPI pairs of humans including of 15750 genes were downloaded from BioGrid, and the study extracted 185,873 PPI pairs which the genes existed in the above gene expression spectrum. Then the researchers performed the SCC distribution for the 185,873 PPI pairs between disease group and the control group and its absolute difference result. The results were shown in Figure 1 and Figure 2. Based on |A1-A2|> 1.743 and at least one of A1, A2 > 0.7, the researchers selected 550 differential interactions. Meanwhile, 26 non-differential interactions whose two corresponding nodes belonged to DEGs were screened with |A1-A2| < 1.743. Finally, the DEN was constructed by incorporating 550 differential interactions and 26 non-differential interactions. The main DEN was

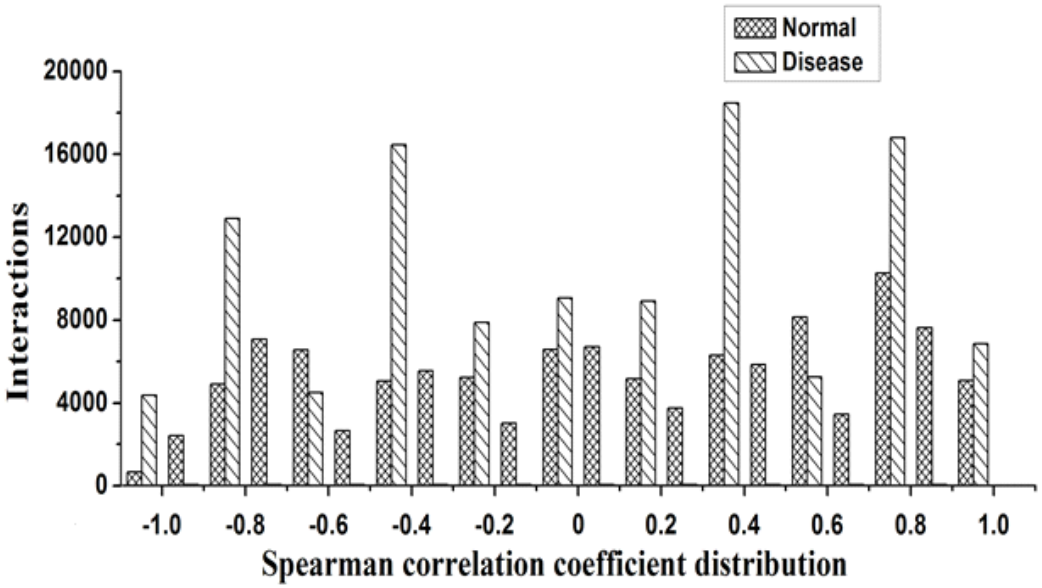

Fig. 1. The distribution of SCC in normal group and acute respiratory distress syndrome group between-1 and 1
Source: Author

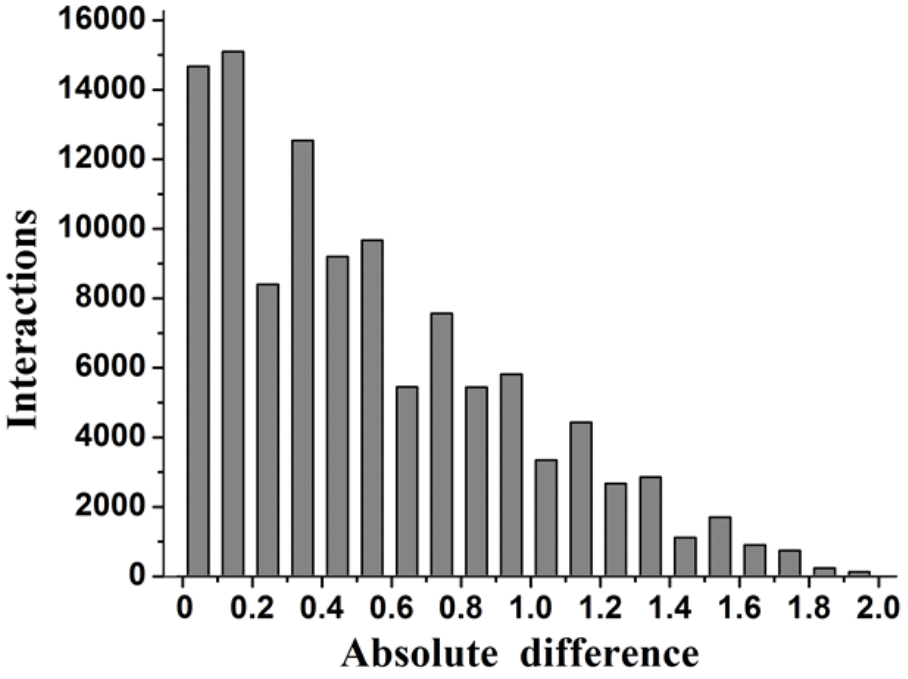

Fig. 2. The absolute difference of SCC between normal group and acute respiratory distress syndrome group
Source: Author

shown in Figure 3. The node stands for the protein, edge stands for the interaction of proteins.

Centrality Analysis

In this work, the researchers obtained a few hub genes based on the nodes degree centrality analysis of the main DEN. Then the six hub genes were picked out under the threshold value > 9 in descending order, which might play important roles in the biological processes of ARDS, as following: *UBC* (degree = 62), *CSNK2A2* (degree = 16), *CUL5* (degree = 14), *SOX2* (degree = 11), *PARK2* (degree = 11) and *CHAF1A* (degree = 9).

Pathways Enrichment Analysis

In this paper, for KEGG pathway enrichment analysis, on one hand, the enriched pathways of the DEGs showed that acute myeloid leuke-

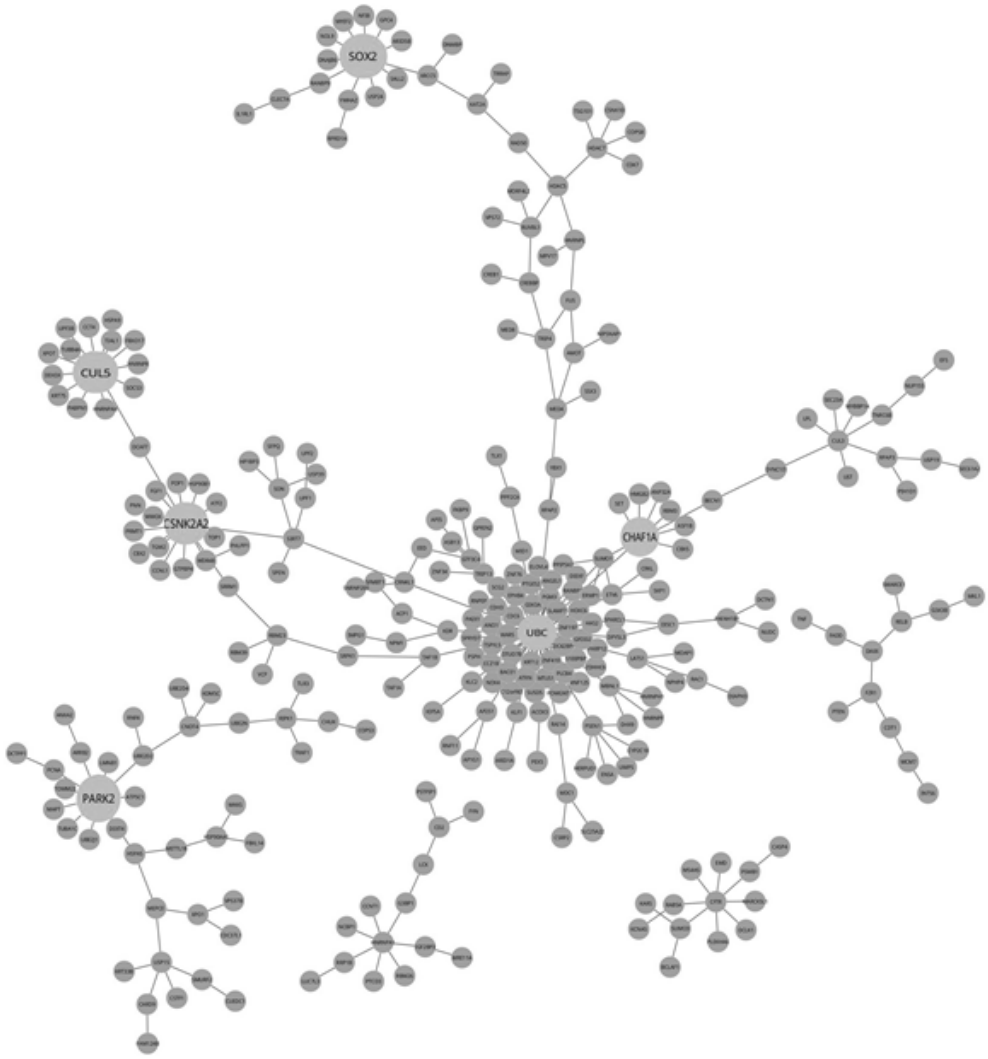

Fig. 3. The main DEN of 550 differential interactions and 26 non-differential interactions. The node stands for the protein (gene), edge stands for the interaction of proteins (genes). Six hub genes are filled with light grey
Source; Author

mia (P = 8.39E-03), T cell receptor signaling pathway (P = 1.69E-02) and endometrial cancer (P = 3.54E-02) were significant pathways under the condition of P < 0.05 (Table 1). On the other hand, the top three enriched pathways in DEN were pathways in cancer (P = 7.43E-10), T cell receptor signaling pathway (P = 8.14E-09), prostate cancer (P = 1.84E-08) at P < 0.05 (Table 1). At last, the researchers obtained T cell receptor signaling pathway that was the mutual pathway under the two conditions.

DISCUSSION

ARDS is a respiratory disease linked to numerous factors including cytokines, epithelial and endothelial damage, fibrogenesis and abnormal lung mechanics (Capelozzi et al. 2017). It was first described in 1967, and since then a majority of studies were raised to address its pathogenesis and therapies. Despite intense research efforts, few therapies for ARDS have been shown to be effective other than the use of lung protection strategies. Functional genomics approaches provide novel insights into understanding gene-environmental interactions controlling this complex process. In the present study, the researchers identified the key genes and the enriched pathways with ARDS using DEN method that maybe play critical roles in regulating the pathogenesis of ARDS.

In this study, the researchers obtained six hub genes based on the nodes degree centrality analysis of the main DEN. Among the six hub genes, *UBC* gene had the highest degree of 62 which maybe act as a genomic resource for functional studies of ARDS. *UBC* is one of two stress-inducible polyubiquitin genes in mammals, whose DNA sequence contains a first untranslated exon, an intron and 9 tandemly repeated ubiquitin moieties. *UBC*, together with *UBB*, was first described as a stress inducible gene, up-regulated upon different cell treatments as well as in some pathological conditions. Meanwhile *UBC* is thought to be essential for fetal liver development and cell-cycle progression (Ryu et al. 2007). It was also reported that the expression of *UBC* had a close relationship testicular gene expression (Sinnar et al. 2011). What was more, Earl et al. (2015) drew a conclusion that the *UBC*-40 was a predominance of more aggressive tumor subtypes among urothelial bladder cancer cell line in recently. Therefore *UBC* was an important gene of many diseases. This study found that it was a hub gene of ARDS based on centrality analysis of DEN rather than the DEG. If the researchers used the traditional DEG method to analyze the key gene of diseases, the *UBC* would be ignored. It was further showed the advanced nature of the DEN.

Moreover, the researchers explored the T cell receptor (TCR) signaling pathway that was the mutual pathway between the DEGs and nodes in the DEN. The self-recognition of the TCR signals and foreign antigens to control T cell homeostasis for immune tolerance were regulated by a variety of cytokines which determined T cell subset homeostasis and differentiation. It has been well documented that protein tyrosine phosphatases were involved in negative regulation of proximal TCR signaling (Ye et al. 2015). Several observations suggested that Treg cell homeostasis, cell-type-specific gene expression and suppressive function critically depended on continuous triggering of their TCR (Vahl et al. 2014). Signaling events triggered by TCR stimulation were importantly targets for the development of common therapeutics for various diseases. Such as Ye et al. (2015) found that the TCR pathway was associated with rheumatoid arthritis. Agostinelli et al. (2014) evaluated that intracellular TCR-signaling pathway played a novel marker for lymphoma diagnosis and potential therapeutic targets. Considering the sig-

Table 1: KEGG pathway enrichment analysis

Enriched pathways of the DEGs		*Enriched pathways in DEN*	
Pathway	*P value*	*Pathway*	*P value*
Acute myeloid leukemia	8.39E-03	Pathways in cancer	7.43E-10
T cell receptor signaling pathway	1.69E-02	T cell receptor signaling pathway	8.14E-09
Endometrial cancer	3.54E-02	prostate cancer	1.84E-08

KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes; DEN, differential expression network.

nificance of TCR signaling pathway in this study, it might also be a potential sign for the diagnosis and treatment of TCR.

CONCLUSION

In conclusion, the researchers utilized the DEN to fully explore ARDS related interactions, and successfully obtained six hub genes and the enriched pathways of ARDS at network level. The DEN offered an accurate manner to predict that the six hub genes and the enriched pathways might be potential biomarkers of early detection and therapy for ARDS.

RECOMMENDATIONS

This study provide a system-wide understanding for ARDS and predicted several potential biomarkers for early diagnosis and treatment of ARDS. However, there were some weak points in this study. Further experiments needed to be carried out to verify our results as soon as possible.

ACKNOWLEDGEMENTS

This research received no specific grants from any funding agency in public, commercial, or not-for-profit sectors.

REFERENCES

Agostinelli C, Rizvi H, Paterson J et al. 2014. Intracellular TCR-signaling pathway: Novel markers for lymphoma diagnosis and potential therapeutic targets. *Am J Surg Pathol,* 38(10): 1349-1359.

Capelozzi VL, Allen TC, Beasley MB et al. 2017. Molecular and immune biomarkers in acute respiratory distress syndrome: A perspective from members of the pulmonary pathology society. *Archives of Pathology and Laboratory Medicine,* 141(12): 1719-1727.

Chatr-Aryamontri AB, Breitkreutz J, Oughtred R et al. 2015. The BioGRID interaction database: 2015 update. *Nucleic Acids Res,* 43(Database Issue): D470-D478.

Datta SG, Satten A, Benos DJ et al. 2004. An empirical bayes adjustment to increase the sensitivity of detecting differentially expressed genes in microarray experiments. *Bioinformatics,* 20(2): 235-242.

Earl J, Rico D, Carrillo-de-Santa-Pau E et al. 2015. The UBC-40 urothelial bladder cancer cell line index: A genomic resource for functional studies. *BMC Genomics,* 16(1): 403.

Fitzgerald M, Millar J, Blackwood B et al. 2014. Extracorporeal carbon dioxide removal for patients with acute respiratory failure secondary to the acute respiratory distress syndrome: A systematic review. *Crit Care,* 18(3): 222.

Guerin C, Reignier J, Richard JC et al. 2013. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med,* 368(23): 2159-2168.

Hoeboer SH, Oudemans-van Straaten HM, Groeneveld AB 2015. Albumin rather than C-reactive protein may be valuable in predicting and monitoring the severity and course of acute respiratory distress syndrome in critically ill patients with or at risk for the syndrome after new onset fever. *BMC Pulm Med,* 15: 22.

Huang DW, Sherman BT, Lempicki RA 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols,* 4(1): 44-57.

Kanehisa M, Goto S 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research,* 28(1): 27-30.

Koschützki D, Schreiber F 2008. Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regulation and Systems Biology,* 2: 193.

Ma L, Robinson LN, Towle HC 2006. ChREBP• Mlx is the principal mediator of glucose-induced gene expression in the liver. *Journal of Biological Chemistry,* 281(39): 28721-28730.

Papazian LJ, Forel M, Gacouin A et al. 2010. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med,* 363(12): 1107-1116.

Pepper SD, Saunders EK, Edwards LE et al. 2007. The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics,* 8(1): 273.

Reiner A, Yekutieli D, Benjamini Y 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics,* 19(3): 368-375.

Rifai N, Ridker PM 2001. Proposed cardiovascular risk assessment algorithm using high-sensitivity C-reactive protein and lipid screening. *Clinical Chemistry,* 47(1): 28-30.

Ryu KY, Maehr R, Gilchrist CA et al. 2007. The mouse polyubiquitin gene UbC is essential for fetal liver development, cell-cycle progression and stress tolerance. *EMBO J,* 26(11): 2693-2706.

Scardoni G, Laudanna C 2012. *Centralities Based Analysis of Complex Networks*. Crotia: INTECH Open Access Publisher.

Sinnar SA, Small CL, Evanoff RM et al. 2011. Altered testicular gene expression patterns in mice lacking the polyubiquitin gene Ubb. *Molecular Reproduction and Development,* 78(6): 415-425.

Smyth GK 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol,* 3(1): 3.

Sud S, Friedrich JO, Taccone P et al. 2010. Prone ventilation reduces mortality in patients with acute respiratory failure and severe hypoxemia: Systematic review and meta-analysis. *Intensive Care Med,* 36(4): 585-599.

Sun SY, Liu ZP, Zeng T et al. 2013. Spatio-temporal analysis of type 2 diabetes mellitus based on differential expression networks. *Sci Rep,* 3: 2268.

Vahl J, Drees CC, Heger K et al. 2014. Continuous T cell receptor signals maintain a functional regulatory T cell pool. *Immunity,* 41(5): 722-736.

Vassiliou AG, Maniatis NA, Orfanos SE et al. 2013. Induced expression and functional effects of aquaporin-1 in human leukocytes in sepsis. *Crit Care,* 17(5): R199.

Villar J, Sulemanji D, Kacmarek RM 2014. The acute respiratory distress syndrome: incidence and mortality, has it changed? *Curr Opin Crit Care,* 20(1): 3-9.

Yadav H, Thompson BT, Gajic O 2017. Fifty years of research in ARDS: Is acute respiratory distress syndrome a preventable disease? *Am J Respir Crit Care Med,* 195(6): 725-736.

Ye H, Zhang J, Wang J et al. 2015. CD4 T-cell transcriptome analysis reveals aberrant regulation of STAT3 and Wnt signaling pathways in rheumatoid arthritis: Evidence from a case-control study. *Arthritis Res Ther,* 17(1): 76.

Ye J, Shi H, Shen Y et al. 2015. PP6 controls T cell development and homeostasis by negatively regulating distal TCR signaling. *J Immunol,* 194(4): 1654-1664.

Zhao J, Yang TH, Huang Y et al. 2011. Ranking candidate disease genes from gene expression and protein interaction: A Katz-centrality based approach. *PloS One,* 6(9): e24306.

Paper received for publication on March 2018
Paper accepted for publication on May 2018