Time-series expression profile analysis to identify important modules and biomarkers in early phase of acute lung injury based on WGCNA

Jing Li
The First Affiliated Hospital of Zhengzhou University
Yanming Yang
The First Affiliated Hospital of Henan University of Chinese Medicine
Lili Fan
Henan Normal University
Zhengjun Cui (czj0371456@163.com)
The First Affiliated Hospital of Zhengzhou University

Research Article

Keywords: Acute lung injury, WGCNA, Time-series expression profile, circadian rhythms, cell death

Posted Date: February 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1300671/v1

Abstract

Background: The morbidity and mortality associated with ALI continue to be significant. Few medical therapies have demonstrated efficacy in curbing the progression of ALI or improving its outcomes. However, time-series expression data has enhanced our ability to query dynamical processes, and WGCNA and maSigPro have emerged as a promising approach for processing large datasets. Therefore, it is possible for us to explore the molecular mechanism in the progression of ALI.

Methods: Downloaded time series gene expression dataset GSE2565 from the Gene Expression Omnibus (GEO), and normalized the dataset using the “sva” R package. maSigPro were used to screen Differential expressed genes (DEGs) and weighted gene co-expression network analysis (WGCNA) were performed to identify hub modules. Gene Ontology and pathway enrichment analyses of genes identified in the hub module were conducted by Metascape. Integration of module analysis and CytoHubba application for identifying hub genes. Finally, receiver operating characteristic (ROC) curve analysis was to measure the predictive accuracy of the hub genes.

Results: In our study, 3005 DEGs were included in WGCNA and 8 modules were identified. Module-trait analysis presented that red module with the most negative correlation with 8hr mainly involved in the regulation of circadian rhythm; pink module with the most positive correlation with 8hr mainly involved in regulation of cell death. Five hub genes (Bnip3, Cdh11, Fam134b, Sult1a1 and Zbtb16) were identified followed by ROC curve analysis.

Conclusion: Our analysis based on time-series expression data identified significant co-expression modules and pathways correlated with early phase of acute lung injury. The hub genes identified may contribute to provide new insights for the molecular mechanisms in acute lung injury.

Introduction

As the primary target of a diversity of internal and/or external environmental insults including microbe’s infection, autoantibodies, toxic gasses, pollutants, gastric acids and so on, the lung often manifest itself the form of acute lung injury (ALI) after exposing to the above mentioned insults. Acute lung injury can evolve to a more severe condition known as acute respiratory distress syndrome (ARDS), which would result to severe dysregulation of function and lung damage**[1-2].** Despite in-depth investigation in treatment methods, the overall mortality of ALI and ARDS , in the United States, is still a remarkable 38.5 % and 41.1 %, respectively and in Shanghai, the mortality rate of ARDS for patients >15 years is 70%[3]. Hence, therapeutic strategies to inhibit the disease progression or improve prognosis and their underlying molecular mechanisms need to be further studied.

The Progress of ARDS comprised of three phases, which is a continuum rather than a strict chronologic phase including exudative, proliferative and fibrotic. The exudative phase (day 1–7) is considered to play a fundamental role of initiating ALI. In the absence of recovery during the exudative phase, the condition in some patients may evolve to a fibrotic phase which is characterized with fibrosis and other irreversible

pathological changes. So, timely source control should be put at the core in the treatment of ALI. Understanding the pathogenesis of ALI is important for current and future treatments as it is linked to outcomes in ALI[3-5].

Earlier studies have suggested that various biological process and multiple molecular factors involved in the progression of pathogenesis of ALI/ARDS such as inflammatory response, oxidative stress, apoptosis, autophagy, CCN1 and HMGB1 were included[6-12]. Therefore, the pathogenesis of ALI/ARDS was affected by a complex network rather than of a single molecular factor and/or pathway because of the phenotypic complexity of ALI/ARDS engendered by different insults, which indicates that researches into more in-depth and comprehensive understanding of molecular mechanisms underlying ALI/ARDS are in demand[13]. Although several researches have individually focused on the role of the potential biomarkers for the process of ALI and identifying the association of genetic factors with susceptibility to ALI based on Differentially Expressed Genes (DEGs), which revealed that some particular genes have been significantly affected by the progression of acute lung injury such as MMP3 ,Timp1, Ly6i, Cxcl1[14-15]. However, limitation to the context of a specific gene could not fully reveal the mechanism in the relatively whole progression of ALI and some significant molecular regulators or pathways during disease progression might be neglected. In our current study, as for gene expression dataset GSE2565[16], it is archives of change at the early stage(day1-3) in the progression of the acute lung injury over time. To understand the role of gene cluster affected significantly in the process of ALI, the dataset need to be explored, and patterns of change analyzed. Genes in Dataset GSE2565 belong to the big data domain and present time-related pattern, hence traditional methods are not in adequate ability to process them. Weighted gene co-expression network analysis (WGCNA) has emerged as a promising approach for evaluating correlation between gene clusters and traits based on co-expression network. Further analysis can be performed to recognize hub modules and confirm key genes in the module, thus identifying potential candidate biomarkers or worthwhile targets to enhance the ability of rapid intervention and improve clinical outcomes[17-19].

In this study, the DEGs were identified from gene expression dataset GSE2565 by using maSigPro approach[20-21]. The WGCNA application was selected for processing DEGs and hub gene modules associated with the progression of ALI were identified, and pathway analysis on hub modules was also presented. Hub genes were distinguished from the hub modules after CytoHubba analysis based on Cytoscape[22]. To assess the predictive ability of the hub genes, receiver operating characteristic (ROC) curve analysis was preformed.

Materials And Methods

Acquisition of the GEO datasets

Microarray data sets of acute lung injury were screened from database(http://www.ncbi.nlm.nih.gov/geo). GSE2565 Series on the GPL339 platform(Affymetrix Mouse Expression 430A Array) was extracted, which was submitted by Sciuito AM[16]. This acute lung injury

associated with GSE2565 contains 48 phosgene-exposed samples and 56 air-exposed samples. To increase the stability of our analysis, biological duplicates were selected. So, 48 paired biological samples from Male CD-1 mice(at 0.5, 1, 4, 8, 12, 24, 48, and 72 hr post-exposure) were obtained and analyzed by WGCNA (Additional file 1: Table S1).

Microarray data preprocessing

Series matrix files provided by the GEO website were used for further analysis after log2-transformation. Probe sets were annotated with Ensembl gene IDs using the data tables of Annotation File downloaded from the GEO website. The R application, the ComBat function of sva package, was used for time-batch normalization to remove batch effects of the gene expression data from the 48 samples processed on 8 consecutive time points[23-24].

The differentially expressed genes analysis in time-course dataset

MaSigPro software from bioconductor was utilized for identifying differentially expressed genes (DEGs) across different time points in mice ALI[20-21]. MaSigPro is an R package for the analysis of time course microarray experiments and it was used to find genes with significant expression changes with a two step regression approach. The first step, gene selection, performs the least-squares technique to select statistically significant genes between the control group and any other experimental group. The Second step, variable selection, is to identify the conditions for which genes shows statistically significant expression changes. In this work, the first step for gene selection rather than the second step was only performed so that the loss of information was as low as possible. For current work, a four-element regression model was defined and a false discovery rate ($Q < 0.05$) was applied to select significant genes.

Co-expression network construction

In this work, genes selected by masigpro approach was performed to construct a weighted gene co-expression network via the “WGCNA” R package in R Studio (version 3.6.1). An adjacency matrix of genes’similarity by pairwise Pearson correlation analysis was conducted. Using the ‘pickSoftThreshold’ function, a appropriate soft threshold of $\beta = 8$ was obtained for strengthening the matrix to a scale-free co-expression network. The topological overlap matrix (TOM) was transformed from the adjacency matrix, and genes were clustered into different modules which was detected by the dynamic tree cutting algorithm of WGCNA.

Identification of significant hub modules

Module eigengenes (MEs) calculated by principal component analysis(PCA) represented the major component of each gene module. the correlation between MEs and clinical traits was estimated by the module–trait relationship analysis, which allowed identifying modules related to external traits. Gene significance (GS) was defined by the correlation between the gene and the external trait. Module membership (MM) was measured by the correlation between each gene expression and each ME. The

module significance (MS) was calculated as the average absolute gene significance (GS) of the genes within the module, which can be used to identify the significant modules with clinical trait.

Functional Enrichment Analysis of the hub Module

To understand the biological significance of the hub module identified by WGCNA, the genes of hub module were mapped into Metascape (http://metascape.org) for Pathway enrichment analysis and Gene Ontology (GO) enrichment analysis. Terms with a P value <0.01, a minimum overlap of 3, and an minimum enrichment factor >1.5 were chosen to be cutoff value. In this work, the top 20 terms were chosen for visualization when more than 20 terms for GO or pathway annotations were identified[25].

Identification of Hub Genes with Module analysis

The function of “exportNetworkToCytoscape” was used to export a PPI network by edge and node list files generated from hub modules in a format suitable for importing to Cytoscape. Based on the MCC sores, the top 15 highest-scored genes identified by the MCC algorithm of Cytoscape were referred as candidate hub genes. Meanwhile, For WGCNA, Candidate hub genes were also determined by absolute MM $\geq$ 0.8 and absolute GS $\geq$ 0.2. As such, common genes shared in the PPI network and in the WGCNA were considered to be hub genes. Finally , the area under the curve (AUC) of the ROC was calculated via “pROC” package to evaluate the prediction accuracy of the candidate genes. Genes with AUC >0.70 in ROC analysis were referred as the real hub genes [26].

Results

DEGs in ALI

Having been adjusted batch effects, gene expression values of 48 samples from GSE2565 were standardized and the results of boxplot analysis before and after normalization were presented in Figure 1. DEGs were identified by the MaSigPro approach based on the details described in the Methods section. To this end, a total of 3005 DEGs were screened out for further analysis.

Construction of Weighted Gene Co-Expression Network

After evaluation for the microarray quality by sample clustering, no outliers were detected in the clusters and the 48 tissue samples were suitable for constructing a hierarchical clustering tree (dendrogram) (Figure 2). The soft-power threshold β caculated by the function “sft$powerEstimate” was set as 8 to ensure a scale-free network (Figure 3). The result of power estimation was shown in Additional file 2: Table S2. As a result, 3005 genes were clustered in to 9 modules based on the TOM matrix. Among the 9 modules for the ALI data, genes in the grey module were not co-expressed, they were excluded before subsequent analysis[27].

Construction of Module–Trait Relationships and Detection of hub Modules

The modules associated with time point after ALI exposure were confirmed by the analysis of the module-trait relationships (Figure 4). we found that 8hr (R = -0.58, P = 2e−5), 12hr (R = -0.40, P = 0.005), 48hr (R =0.37, P = 0.01) and 72hr (R =0.41, P = 0.004) were all significantly correlated with the red module based on the module-treat relationship analysis (Figure 4) and a correlation in upward trend or downward trend between the MEs and the time point was observed. Especially, the correlation between module and trait exhibited that red module had the most significantly correlation with ALI of 8hr, and then gradually increased from 8hr to 72hr (Figure 4). The correlation of pink module, 8hr (R = 0.51, P = 0.005), 12hr (R = 0.15, P = 0.3), 48hr (R =-0.34, P = 0.02) and 72hr (R =-0.49, P = 4e− 04) , gradually decreased. Further more, both red module and pink module exhibited higher association with 8hr (Figure 5). As a consequence, they were considered to be hub modules linked to the development of ALI for further analysis.

Pathway Enrichment Analysis of Genes in the hub Module

To further evaluate the affected biological functions of the genes of the hub module, GO and KEGG pathway analyses were performed and the results of The GO-BP terms and KEGG pathways of the red module were presented respectively in Figure 6 and Figure 7. Annotation information of gene in hub modules can be seen in Additional file Table S3 and Table S4. The analysis of the top three GO-BP terms showed that genes in the red module were mainly significantly enriched: circadian regulation of gene expression, response to hormone and regulation of cellular response to stress. The terms of KEGG pathway analysis presented in Figures 6b. Circadian rhythm, Propanoate metabolism and PPAR signaling pathway were mainly significantly enriched. For the pink module (Figures 7a and Figures 7b), we identified 17 GO terms and 2 KEGG pathways, which were mainly related with regulation of cytokine production, positive regulation of cell death and glycerophospholipid metabolism.

Module analysis for Screening hub genes

The PPI network of the genes in blue module and pink module was respectively constructed by using the function “exportNetworkToCytoscape” based on WGCNA. The Maximal Clique Centrality (MCC) algorithm of Cytoscape was performed to select candidate hub genes from the PPI network. Based on the MCC sores, the top 15 highest-scored genes were referred as the hub genes. The result of the candidate hub genes detected by MCC from blue and pin module was visualized respectively in Figure 8a and Figure 8b. In addition, using a gene significance (GS) > 0.2 and module membership (MM) > 0.8 in co-expression network, Genes selected from the given modules were referred as hub genes (Table 1 and Table 2). As a consequence, 11 genes shared both in PPI network and co-expression network were chosen as hub genes. Genes shared by both networks are marked in blue color. In addition, to explore the prediction of the candidate hub genes as biomarkers of ALI , ROC curve analysis was performed (Figure 9). Finally, 5 genes with AUC values greater than 0.7 presented a high predictive accuracy for the development of ALI, wich included Bnip3, Cdh11, Fam134b Sult1a1 and Zbtb16. So, these five genes were considered to be the real hub genes .

Discussion

In this work, key modules and genes involved in the development of ALI with different time points were identified by multiple bioinformatics method. Firstly, a total of 3005 significant genes were selected by maSigPro approach. Then, WGCNA was performed to explore the relationship between modules and traits and key modules that were significantly relevant to ALI were confirmed. Finally, hub genes were also confirmed after module analysis. masigpro as a methodology to deal with the analysis of gene expression changes over time manifested itself in a unique statistical advantage[20]. The maSigPro approach in our study was used to select statistically significant genes. Especially, only the first step was performed for the present work when applying the maSigPro procedure in order to minimise the loss of gene information. In addition, WGCNA as an comprehensive method to identify co-expression modules based on the similarity of expression patterns of genes was utilized for identifying key modules and hub genes related to the development of ALI.

Unlike previous works, which had for the most part used a static or a combined static time perspective, this work made use of a dynamic or time-series gene expression data to establish which specific genes can be affected by clinical traits and obtain the number of genes of ALI participation for biological process to the utmost extent. The earlier studies took a static approach in exploring the mechanisms and implications of temporally regulated gene expression of ALI, focusing on either a few gene from a certain time or on gene data set from a specific time frame which was a new data set that integrated data sets at different points in time. The previous works analyzed the genes associated with acute lung injury based on different data sets with different points in time or a new data set that integrated data sets at different points in time. For instance, Chen et al. analyzed ALI related genes by combining dataset GSE2411 containing attribute at 4 hours post-intervention, GSE18341 at 2 hours and GSE17355 at 0,1,4 and 10 days and considered the 3 different datasets with different attribute as a whole. As we known, most biological processes are dynamic and gene expression presents a high degree of temporal and spatial specificity, time-course gene expression data may have the ability to fully explore the patterns of gene expression underlying disease states. It is generally believed that time series gene expression data with the same size, compared to static gene expression data contain more information of gene regulatory network is derived from the perspective of biology. Static gene expression experiment may fail to fully exhibit the temporal trends of gene expression resulting in loss of some co-expressed genes [28]. However, our study identified the key modules and genes that representing the developmental process of ALI based on an 8 time points expression experiment, where we can get the most significantly genes over time.

In current study, by using WGCNA, we confirmed two gene modules that highly correlated to the development of ALI and carried out enrichment analysis for each modules. Since module red and module pink both displayed significant correlation In Module-trait relationships analysis, meanwhile both modules also had greater MS, we treated both modules as hub modules. In our current study, red module and pink module indicated negatively or positively correlation with 8hr, respectively. GO and KEGG enrichment analyses displayed that the genes in the red module had different roles and both were

significantly associated with regulation of circadian rhythm. Previous studies have indicated that cellular and molecular circadian rhythms are elicited by inflammation in the setting of acute lung injury, indicating its important role in the fate of progression related to ALI[29,30]. being related to a variety of biological functions and biological dysfunctions, circadian rhythm related molecules has emerged as a promising target for seeking to rectify biological disorder[31].

For pink module, genes in pink module was mainly related with regulation of cytokine production, positive regulation of cell death based on GO analysis. Regarding to KEGG analysis, genes of pink module revealed enrichment in Glycerophospholipid metabolism and HTLV-I infection. Phosphatidic acid (PA), one of the elements of glycerophospholipid metabolism, is involved in the regulation of mTOR pathway, which affect the process of protein synthesis, autophagy and mitochondrial metabolism[32,33]. Autophagy, a form of programmed cell death, presents dual roles in the pathophysiologic process of ALI for progression/inhibition of the disease according to different cell types, different insults and progression for the disease and so on[34]. As a result, more comprehensive and in-depth evaluation for the role of autophagy in ALI are in demand.

For our current work, we identified 5 susceptibility genes for ALI: Bnip3, Cdh11, Fam134b, Sult1a1 and Zbtb16. Bnip3 have the ability to trigger cell death by the pathway of necrosis, apoptosis and autophagy[35]. Bnip3 exerts a dual effect through autophagy or apoptosis, Homeostasis of mitochondria is achieved to maintain cell Cell survival through autophagy while apoptosis is mediated by the interaction between Bnip3 and mitochondrial fusion protein-optic atrophy1 to induce cell death[36]. Further researches also implied that Bnip3 expression level was significantly increased following acute lung injury [37]. Cdh11 refers to a type II classical cadherin and participates in the regulation of calcium dependent cell-cell adhesion. Cdh11 has been reported to be associated with fibrosis, inflammation, cancer and other pathologic processes[38]. Flow cytometry was used to disclosed that the overall proportion of CDH11+cardiac mesenchymal cells increased from days 3 to days 7 and remained unchanged over time in sham group. Moreover, the significant increase of the expression of Cdh11 is also observed as early as 3 days following injury which in alignment with the time course of inflammation resolution. This observation implicates Cells associated with inflammatory infiltration such as macrophages□neutrophils, and monocytes — may result to Cdh11-mediated fibrotic remodeling in heart[39]. This finding, the expression of Cdh11 presents time dependency, indicates that targeting Cdh11 may provide opportunities for timely intervention to early inflammation and curbing the progression to pulmonary fibrosis.

Since its discovery in mammalian cells, the molecule encoded by Fam134b has been widely recognized as a molecular receptor for endoplasmic reticulum–specific autophagy (ER-phagy). AAA demonstrated that Fam134b activation exerts an important role are in the regulation of ER-phagy. it as a selective autophagy may present a dual function in cell survival. when cells are subjected to diverse insults such as oxidative stress, chemical stimulation and calcium overload, ER-phagy may help ER restore homeostasis or induce cell death according to the timing and magnitude of stimuli[40,41]. The study of Melchiotti R also implicated that Fam134b is a potential molecular target in response to inflammation,

and refers to cytokine secretion [42]. As noted above, being involved in common pathway, It is reasonable to speculate that Fam134b has a strong connection with ALI. However, it remains to be explored whether Fam134b play pivotal role in the progression of ALI.

Sulfotransferase 1a1 (Sult1a1) , known as drug-processing gene, is a phase II metabolic enzyme that exerts extensively functions of metabolism and detoxification of multitudinous drugs and chemicals[43]. Lianxia Guo provided the evidence that the expression and activity of Sult1a1 was manifested in circadian rhythmicity, which was directly regulated by the clock protein Bmal1[44]. Further more, there is a growing body of evidence that drug-processing genes for the most part were expressed in circadian rhythms and rhythmic expressions of which present close association with time dependency of toxicity and tolerance of drug in body[45]. Thereby, elucidating the molecular mechanisms of Sult1a1 in acute lung injury assume considerable significance in working out the best timing for drug administration. Zbtb16 (zinc finger and BTB domain containing 16), also known as the promyelocytic leukemia zinc finger protein (PLZF), was reported to be primarily expressed in the apical membrane of bronchioles and involved in restraining Toll-like receptor-induced inflammation in the hyperglycemic states, therefore, Zbtb16 may serve as moluculer marker in lung tissues[46,47].

The current study has some limitations. The microarray data for analysis were extracted from open access databases with a small sample size, so further studies on time-course are needed to improve the reliability of the result. Although, the mechanism of lung injury induced by various insults in animal models share common biological pathway, the current analysis was limited to the elucidation of animal models exposed to photogas. Thereby, other types of time series studies related to acute lung injury need to be further investigated. To this end, more representative results can be achieved by the analysis for common genes screened out from integrated dataset. In addition, animal model can not fully generate all the pathophysiological feature of ALI in humans, and most of which could only represent partial characteristics of human ALI[48]. Therefore, further researches are required to validate these results.

Conclusion

According to our knowledge, this was the first study that selected the WGCNA method to identify the completely activated gene set in the progression of ALI based on time-series expression data. Circadian rhythm and autophagy were identified as the essential components of pathophysiological mechanisms in the early phase for ALI. Five hub genes, Bnip3, Cdh11, Fam134b, Sult1a1 and Zbtb16, were identified followed by module analysis, PPI network, and ROC analysis, which may contribute to provide new insights for the molecular mechanisms in acute lung injury.

Declarations

Acknowledgments

We acknowledge GEO database for providing their platforms and contributors for uploading their meaningful datasets.

Authors' contributions

LJ and YYM conceived the work. LJ wrote the manuscript. FLL, LJ, and YYM participated in data analysis and interpretation. FLL revised the manuscript and participated in discussion. CZJ supervised the study. All authors approved the submission of the manuscript.

Funding

The present work was supported by grants from the National Natural Science Cooperation Foundation of China (grant no. U1604188)

Availability of data and materials

The dataset analyzed in this study can be derived from public repositories: GSE2565(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2565).

Ethics approval and consent to participate

GEO belongs to public databases. The animals involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. Our study is based on open source data, so there are no ethical issues and other conflicts of interest.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

[1]He YQ, Zhou CC, Yu LY, et al. Natural product derived phytochemicals in managing acute lung injury by multiple mechanisms. Pharmacol Res. 2021;163:105224

[2]Gonçalves-de-Albuquerque CF, Silva AR, Burth P, Castro-Faria MV, Castro-Faria-Neto HC. Acute Respiratory Distress Syndrome: Role of Oleic Acid-Triggered Lung Injury and Inflammation. Mediators Inflamm. 2015;2015:260465. doi:10.1155/2015/260465

[3]He YQ, Zhou CC, Yu LY, et al. Natural product derived phytochemicals in managing acute lung injury by multiple mechanisms. Pharmacol Res. 2021;163:105224. doi:10.1016/j.phrs.2020.105224.

[4]Mokra D, Mikolka P, Kosutova P, Mokry J. Corticosteroids in Acute Lung Injury: The Dilemma Continues. Int J Mol Sci. 2019;20(19):4765. Published 2019 Sep 25. doi:10.3390/ijms20194765

[5] Mowery N T, Terzian W, Nelson A C. Acute Lung Injury[J]. Current Problems in Surgery, 2020, 57(5):100777.

[6]Matthay MA, Zemans RL, Zimmerman GA, et al. Acute respiratory distress syndrome. Nat Rev Dis Primers. 2019;5(1):18. Published 2019 Mar 14. doi:10.1038/s41572-019-0069-0

[7] Zhang H., Wang Z., Liu R., Qian T., Liu J., Wang L., Chu Y. Reactive oxygen species stimulated pulmonary epithelial cells mediate the alveolar recruitment of FasL(+) killer B cells in LPS-induced acute lung injuries. J. Leukoc. Biol. 2018;104(6):1187–1198.3-5; [8]Nadeem A., Al-Harbi N.O., Ahmad S.F., Ibrahim K.E., Siddiqui N., Al-Harbi M.M. Glucose-6-phosphate dehydrogenase inhibition attenuates acute lung injury through reduction in NADPH oxidase-derived reactive oxygen species. Clin. Exp. Immunol. 2018;191(3):279–287.

[9]Fan K., Lin L., Ai Q., Wan J., Dai J., Liu G., Tang L., Yang Y., Ge P., Jiang R., Zhang L. Lipopolysaccharide-induced dephosphorylation of AMPK-Activated protein kinase potentiates inflammatory injury via repression of ULK1-dependent autophagy. Front. Immunol. 2018;9:1464.

[10]Zhang D., Zhou J., Ye L.C., Li J., Wu Z., Li Y., Li C. Autophagy maintains the integrity of endothelial barrier in LPS-induced lung injury. J. Cell. Physiol. 2018;233(1):688–698.

[11]Lee S, Piao C, Kim G, Kim JY, Choi E, Lee M.Production and application of HMGB1 derived recombinant RAGE-antagonist peptide for anti-inflammatory therapy in acute lung injury. Eur J Pharm Sci. 2018 Mar 1; 114():275-284.

[12]Dong N, Ji D, Huang X, Ying Z, Wang X, Chen C. Lipopolysaccharide-induced CCN1 production enhances interleukin-6 secretion in bronchial epithelial cells. Cell Biol Toxicol. 2018 Feb;34(1):39-49

[13] Patel VJ, Biswas Roy S, Mehta HJ, Joo M, Sadikot RT. Alternative and Natural Therapies for Acute Lung Injury and Acute Respiratory Distress Syndrome. Biomed Res Int. 2018;2018:2476824.

[14]Artham S, Verma A, Newsome AS, Somanath PR. Patients with acute respiratory distress syndrome exhibit increased stromelysin1 activity in the blood samples. Cytokine. 2020 Jul;131:155086.

[15]Mao K, Geng W, Liao Y, et al. Identification of robust genetic signatures associated with lipopolysaccharide-induced acute lung injury onset and astaxanthin therapeutic effects by integrative analysis of RNA sequencing data and GEO datasets Aging (Albany NY).2020;12(18):18716-18740. doi:10.18632/aging.104042

[16]Sciuto AM, Phillips CS, Orzolek LD, Hege AI et al. Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation. Chem Res Toxicol 2005 Nov;18(11):1654-60.

[17]Wu Z, Hai E, Di Z, Ma R, Shang F, Wang Y, et al. (2020) Using WGCNA (weighted gene coexpression network analysis) to identify the hub genes of skin hair follicle development in fetus stage of Inner Mongolia cashmere goat. PLoS ONE 15(12): e0243507.

[18]Liang Jia-Wei Fang Zheng-Yu, Yong Huang, et al. Application of Weighted Gene Co-Expression Network Analysis to Explore the Key Genes in Alzheimer's Disease[J]. Journal of Alzheimer's disease 2018; 65(4).

[19]Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis[J]. BMC bioinformatics 2008; 9(1): 559.

[20]Conesa A, Nueda MJ, Ferrer A, Talón M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. Bioinformatics. 2006 May 1;22(9):1096-102.

[21]Comparison of gene expression in liver regeneration and hepatocellular carcinoma formation.Cancer Management and Research 2018:10 5691–5708.

[22]Tong YQ, Song Y, Xia CH and Deng SX (2020) Theoretical and in silico Analyses Reveal MYC as a Dynamic Network Biomarker in Colon and Rectal Cancer.Front. Genet. 11:555540.

[23]Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012).The sva package for removing batch effects and other unwanted variation in high-throughput experiments.Bioinformatics,28(6), 882–883.

[24]Peng XY, Wang Y, Hu H, Zhang XJ, Li Q. Identification of the molecular subgroups in coronary artery disease by gene expression profiles. J Cell Physiol. 2019 Feb 25. doi: 10.1002/jcp.28324.

[25]Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications 10:1523.

[26]Al-Mansour M R, Wu J, Gagnon G, et al. Linear versus volumetric CT analysis in predicting tension-free fascial closure in abdominal wall reconstruction[J]. Hernia, 2021, 25(1):91-98.

[27] Li, J., Zhou, D., Qiu, W. et al. Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. Sci Rep 8, 622 (2018).

[28] Ganesh S K, Joo J, Skelding K, et al. Time course analysis of gene expression identifies multiple genes with differential expression in patients with in-stent restenosis[J]. BMC Medical Genomics,4,1(2011-02-28), 2011, 4(1):1-11.

[29]Haspel J A, Chettima Da S, Shaik R S, et al. Circadian rhythm reprogramming during lung inflammation[J]. Nature Communications, 2014, 5(1):4753.

[30]Dy A, Xf B, Yx C, et al. Rev-erbα can regulate the NF-κB/NALP3 pathway to modulate lipopolysaccharide-induced acute lung injury and inflammation[J]. International Immunopharmacology, 2019, 73:312-320.

[31]Huang S, Jiao X, Lu D, Pei X, Qi D, Li Z. Recent advances in modulators of circadian rhythms: an update and perspective. J Enzyme Inhib Med Chem. 2020;35(1):1267-1286. doi:10.1080/14756366.2020.1772249.

[32]Lutkewitte AJ, Finck BN. Regulation of Signaling and Metabolism by Lipin-mediated Phosphatidic Acid Phosphohydrolase Activity. Biomolecules. 2020;10(10):1386. Published 2020 Sep 29.

[33]Zeng C, Wen B, Hou G, et al. Lipidomics profiling reveals the role of glycerophospholipid metabolism in psoriasis. Gigascience. 2017;6(10):1-11. doi:10.1093/gigascience/gix087.

[34].Wang K, Chen Y, Zhang P, Lin P, Xie N, Wu M. Protective Features of Autophagy in Pulmonary Infection and Inflammatory Diseases. Cells. 2019;8(2):123.

[35] Vishnupriya S, Priya Dharshini LC, Sakthivel KM, Rasmi RR. Autophagy markers as mediators of lung injury-implication for therapeutic intervention. Life Sci. 2020;260:118308.

[36]Liu K.E., Frazier W.A. Phosphorylation of the BNIP3 C-terminus inhibits mitochondrial damage and cell death without blocking autophagy. PLoS One. 2015;10(6).

[37]Wei, Zhang, Jiaqiang, et al. Dexmedetomidine preconditioning protects against lung injury induced by ischemia-reperfusion through inhibition of autophagy[J]. Experimental and Therapeutic Medicine, 2017, 14(2).

[38]Chang SK, et al. Cadherin-11 regulates fibroblast inflammation. Proc Natl Acad Sci USA. 2011;108(20):8402–8407. doi: 10.1073/pnas.1019437108.

[39]Schroer AK, Bersi MR, Clark CR, et al. Cadherin-11 blockade reduces inflammation-driven fibrotic remodeling and improves outcomes after myocardial infarction. JCI Insight. 2019;4(18):e131545. Published 2019 Sep 19.

[40]Jiang X, Wang X, Ding X, et al. FAM134B oligomerization drives endoplasmic reticulum membrane scission for ER-phagy. EMBO J. 2020;39(5):e102608.doi:10.15252/embj.2019102608.

[41]Mo J, Chen J, Zhang B. Critical roles of FAM134B in ER-phagy and diseases. Cell Death Dis. 2020;11(11):983. Published 2020 Nov 16. doi:10.1038/s41419-020-03195-1.

[42]Melchiotti R, Puan KJ, Andiappan AK, et al. Genetic analysis of an allergic rhinitis cohort reveals an intercellular epistasis between FAM134B and CD39. BMC Med Genet. 2014;15:73. Published 2014 Jun 27. doi:10.1186/1471-2350-15-73.

[43]James MO and Ambadapadi S (2013) Interactions of cytosolic sulfotransferases with xenobiotics. Drug Metab Rev 45:401–414.

[44]The Clock Protein Bmal1 Regulates Circadian Expression and Activity of Sulfotransferase 1a1 in Mice Lianxia Guo, Fangjun Yu, Tianpeng Zhang and Baojian WuDrug Metabolism and Disposition October 2018, 46 (10) 1403-1410.

[45] Sukumaran S, Almon RR, DuBois DC, and Jusko WJ (2010) Circadian rhythms in gene expression: relationship to physiology, disease, drug disposition and drug action. Adv Drug Deliv Rev 62:904–917.

[46]Kim JH, Rasaei R, Park S, Kim JY, Na S, Hong SH. Altered Gene Expression Profiles in the Lungs of Streptozotocin-induced Diabetic Mice. Dev Reprod. 2020;24(3):197-205. doi:10.12717/DR.2020.24.3.197.

[47]Sadler AJ, Rossello FJ, Yu L, et al. BTB-ZF transcriptional regulator PLZF modifies chromatin to restrain inflammatory signaling programs. Proc Natl Acad Sci U S A. 2015;112(5):1535-1540.

[48] Mokra D, Mikolka P, Kosutova P, Mokry J. Corticosteroids in Acute Lung Injury: The Dilemma Continues. Int J Mol Sci. 2019;20(19):4765. Published 2019 Sep 25. doi:10.3390/ijms20194765.

Tables

Table 1 Results of GS and MM of hub genes in red module

Hub gene	GS.red	MM.red
Npas2	0.524992056	0.812244171
Bnip3	0.591542468	-0.829989803
Ncbp2	0.466903275	0.843884636
Ints6	0.560423345	-0.833301709
Tef	0.568211177	-0.829765767
Sult1a1	0.53098546	-0.854797912
Pnpla2	0.515505357	-0.855865699
Ucp2	0.538712765	-0.827160254
Gyg	0.54025136	0.913996132
Hp	0.395057092	0.815863357
Map3k6	0.420717642	-0.855985892
Cdh11	0.505636263	0.861182638
H6pd	0.416300045	-0.856911322
Pcbd2	0.59191669	0.812055206
Ckb	0.38803167	0.80548476
Ppl	0.420790992	-0.833715548

Genes marked in bold black are common genes shared in PPI network

Table 2 Results of GS and MM of hub genes in pink module

Hub gene	GS.pink	MM.pink
Bcar3	0.406937702	0.813736076
Zbtb16	0.432140632	0.812118149
Fam134b	0.298895263	0.852403597
Stx3	0.29487702	0.810492385
Mlxip	0.365950339	0.819157543
Klf9	0.576878737	0.929771399
Adrb2	0.362958439	0.868261175
Ctla2a	0.362446535	0.8291722
Map3k6	0.499957606	0.815306015
Chst15	0.296325821	0.837480532
Lpin2	0.505598078	0.876790573

Genes marked in bold black are common genes shared in PPI network

Figures

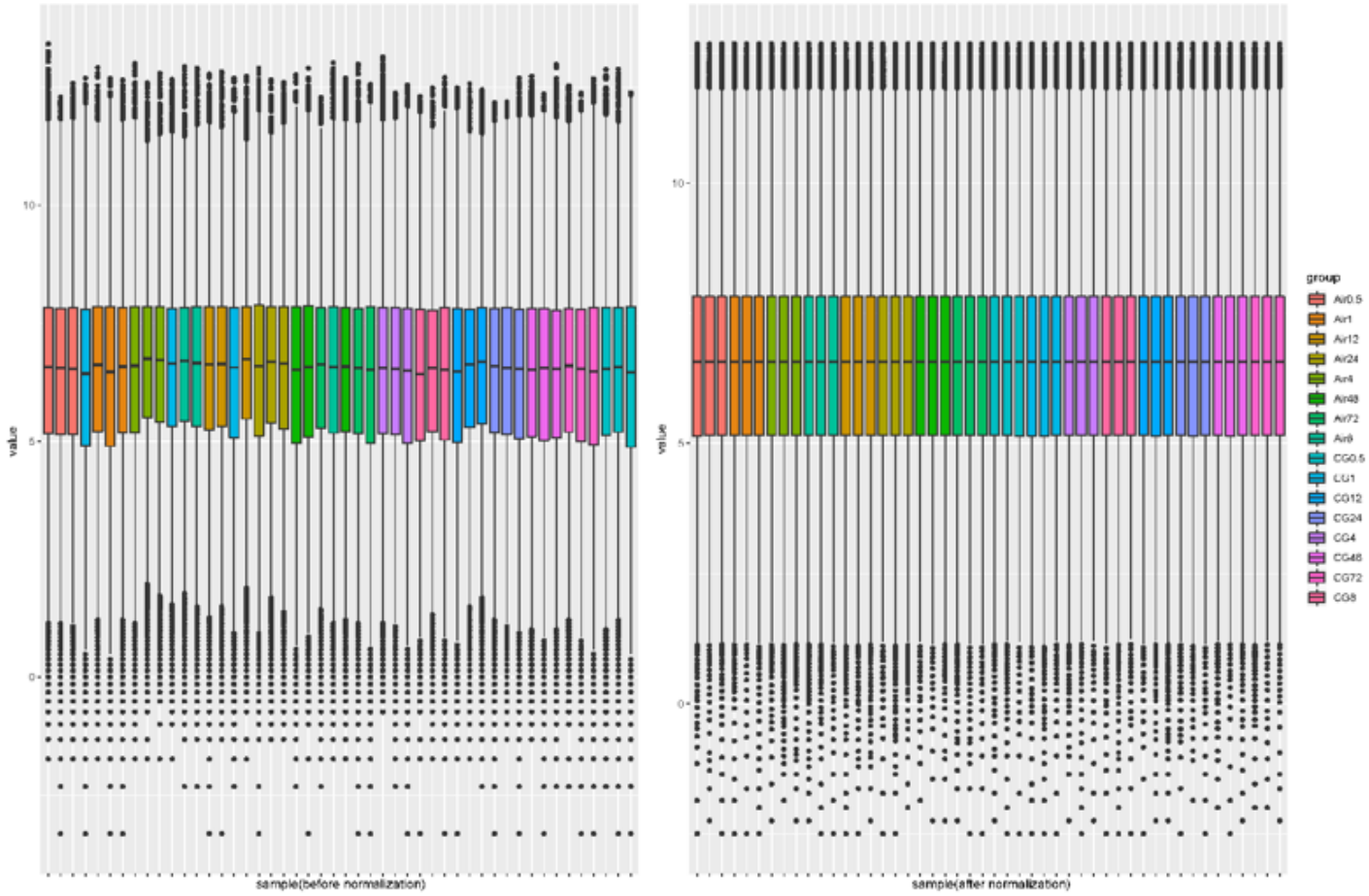

Figure 1

Box plot analysis for unnormalised and normalised data.

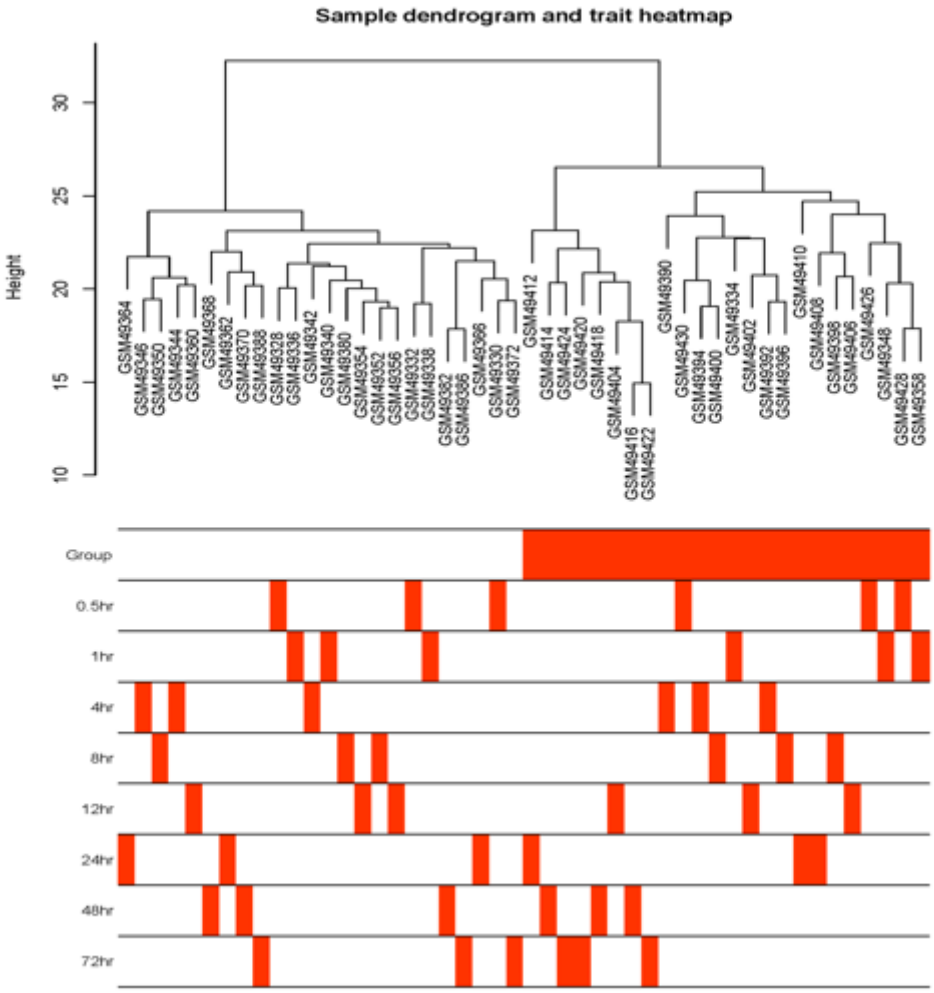

Figure 2

Sample dendrogram and trait heatmap. The clinical traits, the group and the different time points are presented at the bottom.The colors represent the proportion to clinical traits.

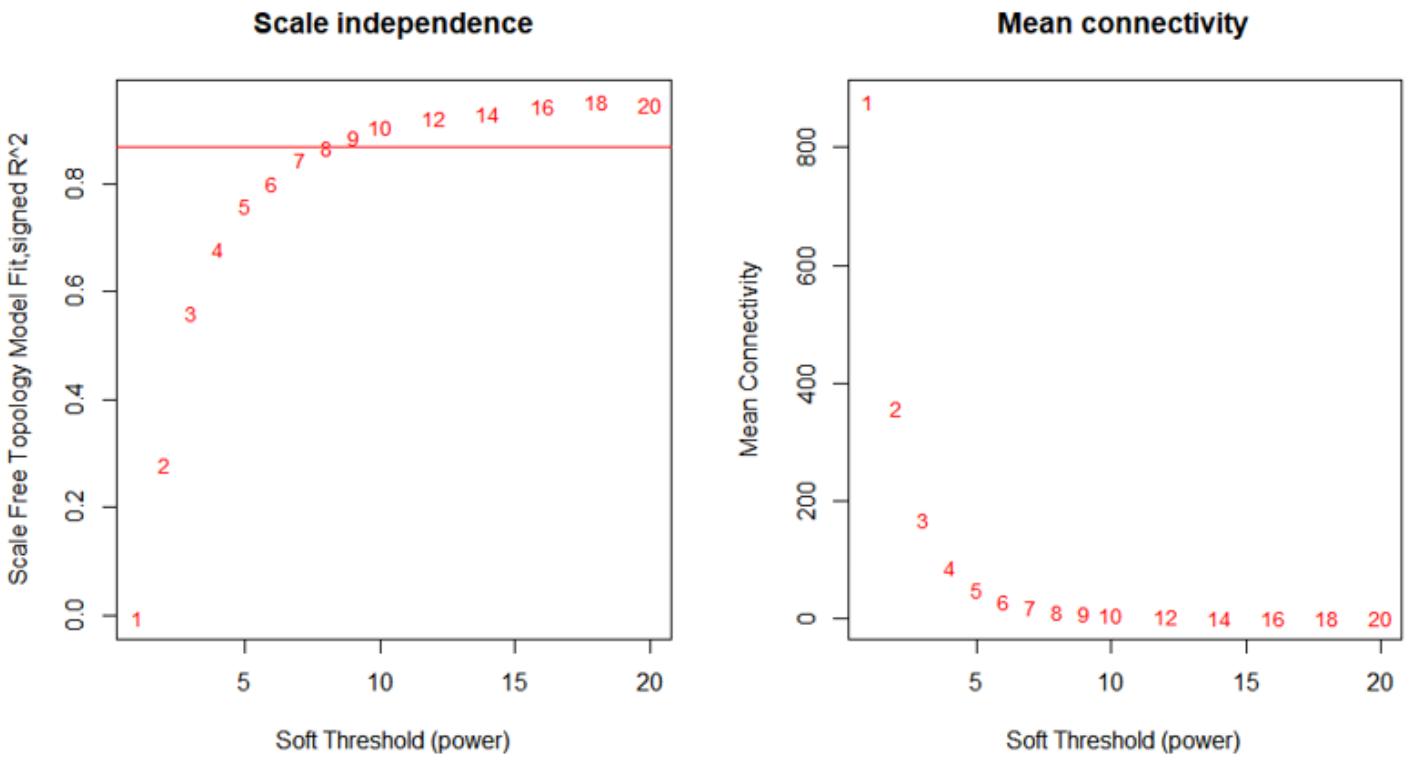

Figure 3

Selection for various soft-thresholding powers (β). Red line represents the correlation coefficient (0.866).

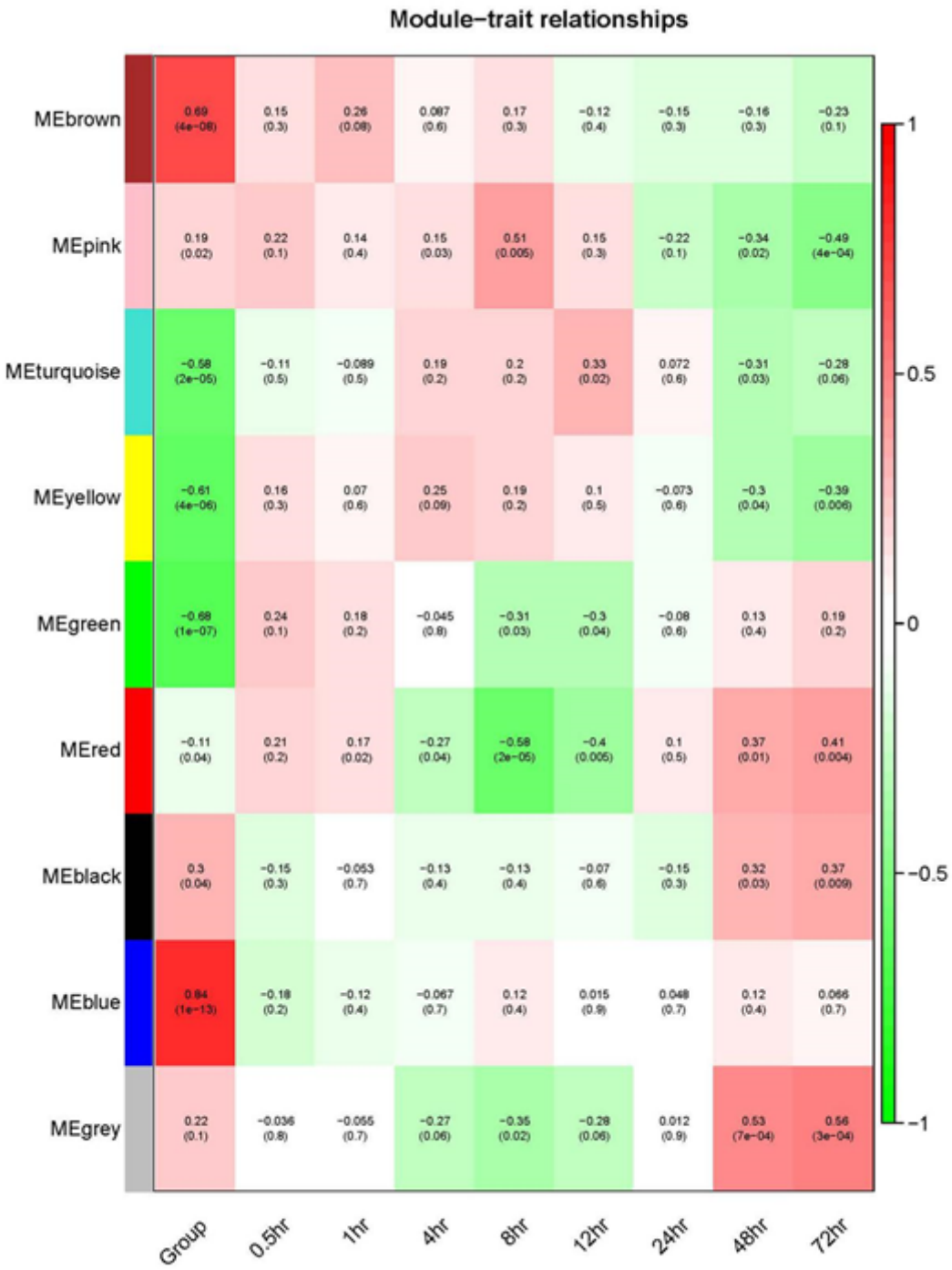

Figure 4

Heatmap of the correlation between module eigengenes and clinical traits. Each row corresponds to a module, and each column corresponds to a trait. Each cell contains the corresponding correlation and P value. The table is color-coded by correlation according to the color legend.

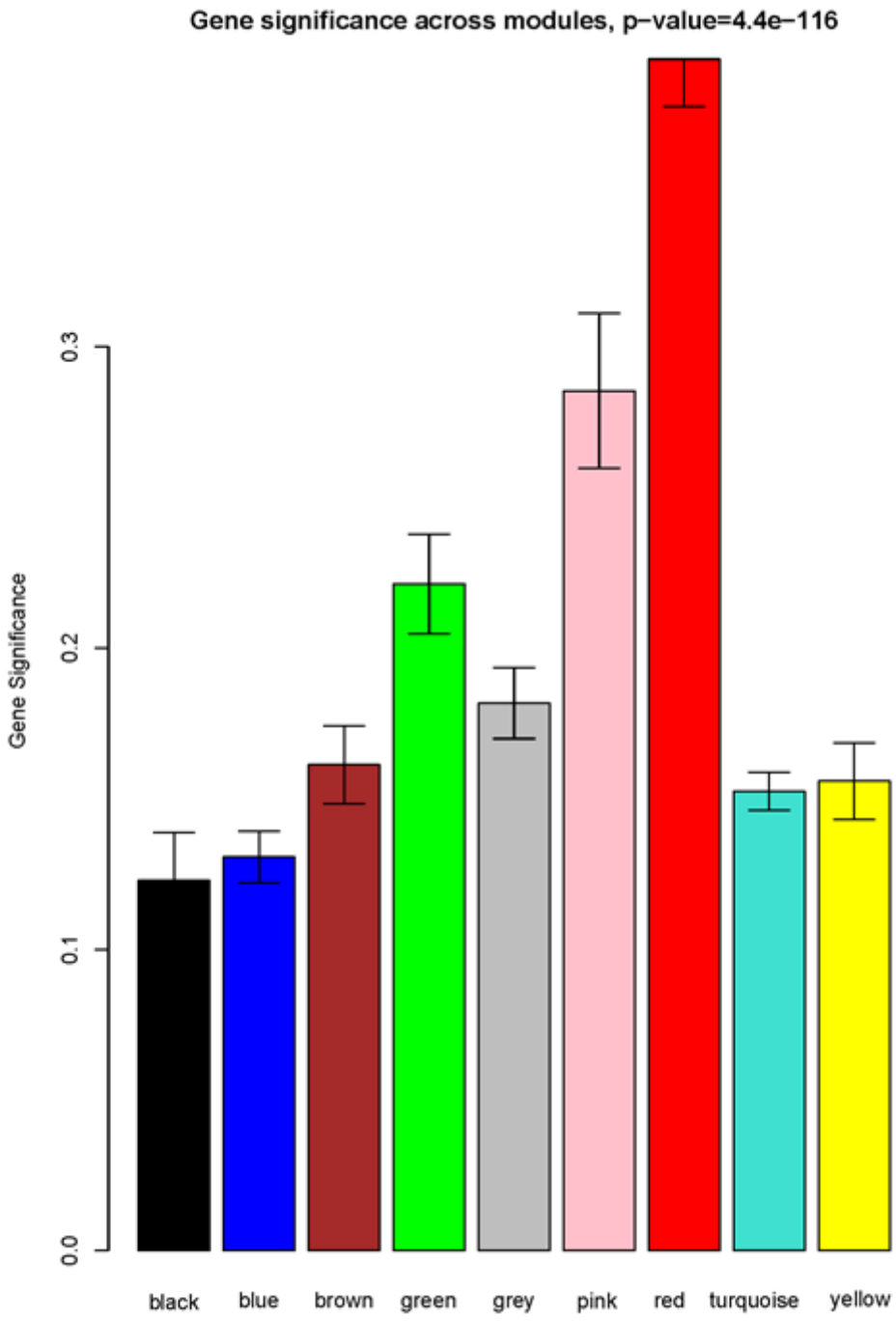

Figure 5

Distribution of average gene significance and errors in the modules associated with the process of ALI

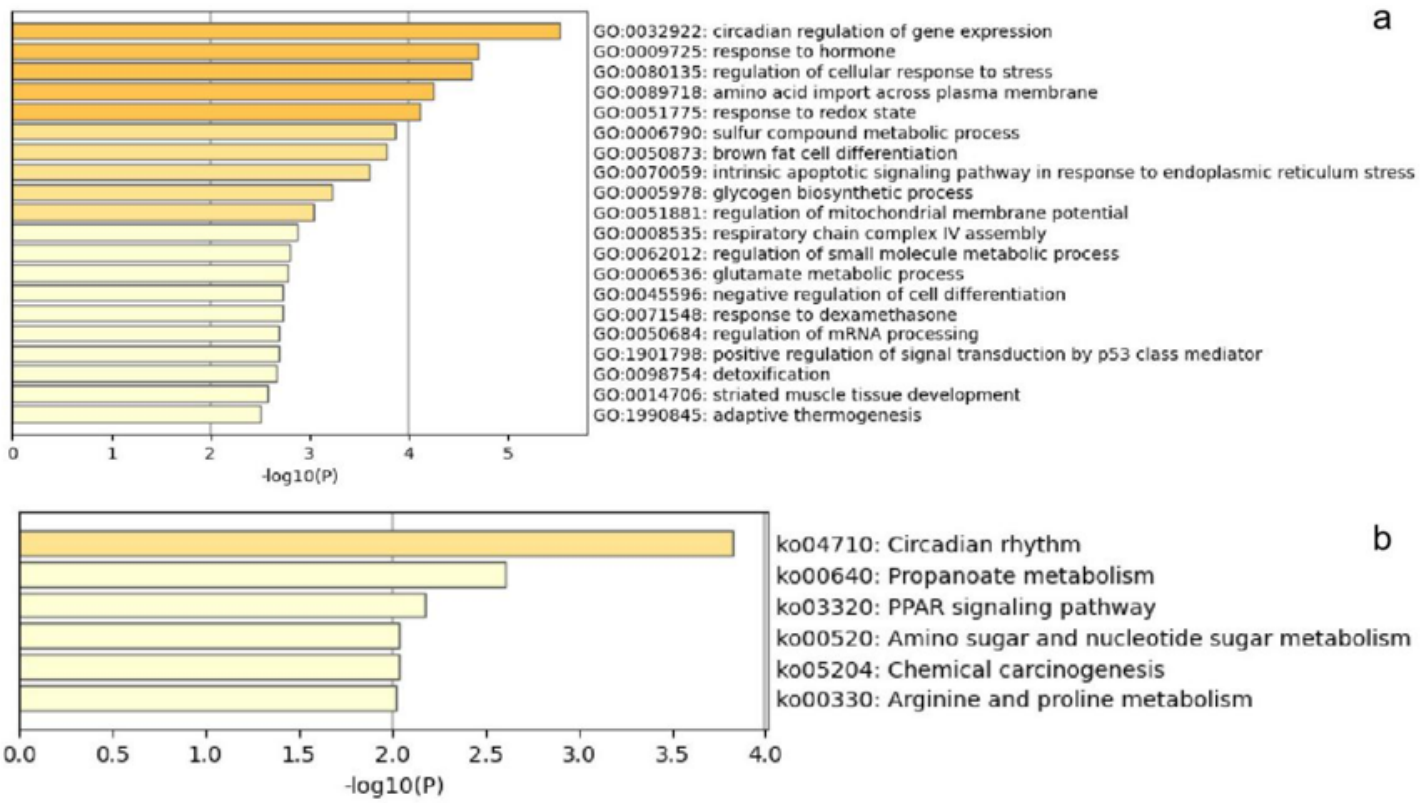

Figure 6

Functional enrichment analysis of the red module genes. (a) Gene Ontology analysis of genes in the red module. (b) KEGG pathway enrichment analysis of genes in the red module.

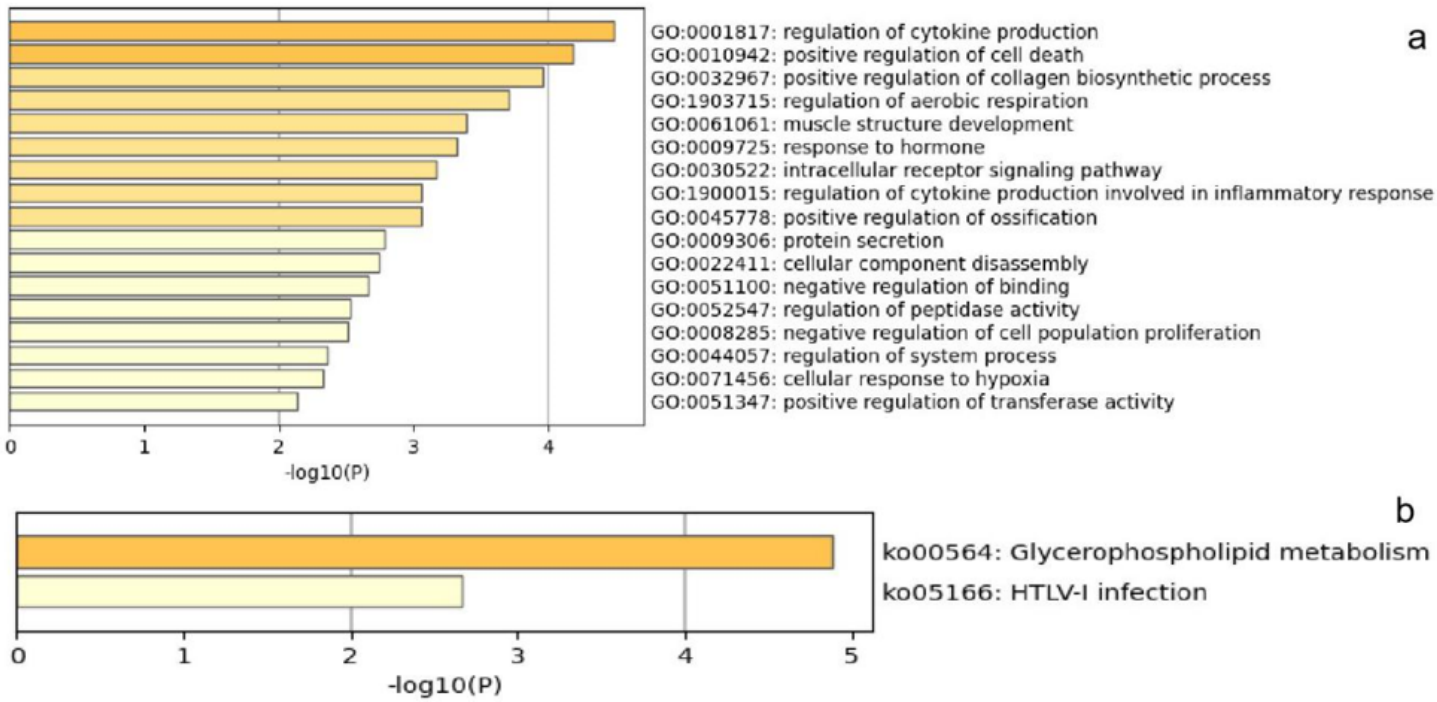

Figure 7

Functional enrichment analysis of the red module genes. (a) Gene Ontology analysis of genes in the red module. (b) KEGG pathway enrichment analysis of genes in the red module.

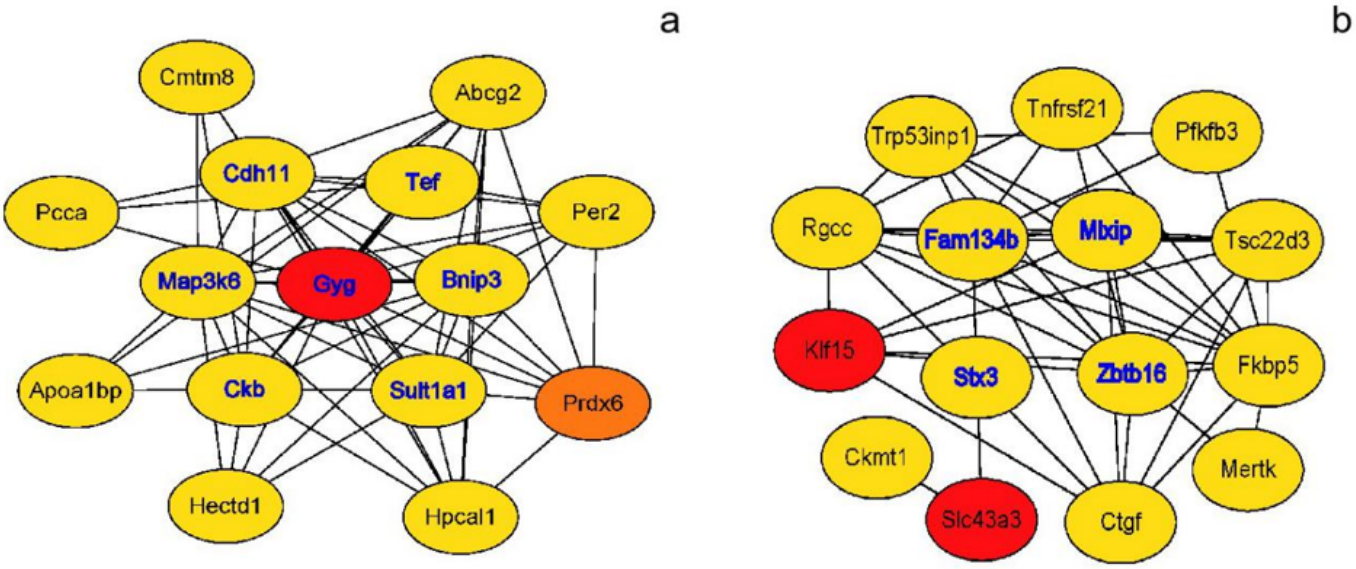

Figure 8

Identification of the hub genes from the PPI network using MCC algorithm. (a) Candidate hub genes in the red module. (b) Candidate hub genes in the pink module. Common hub genes shared in PPI network and co-expression network are marked in blue color.

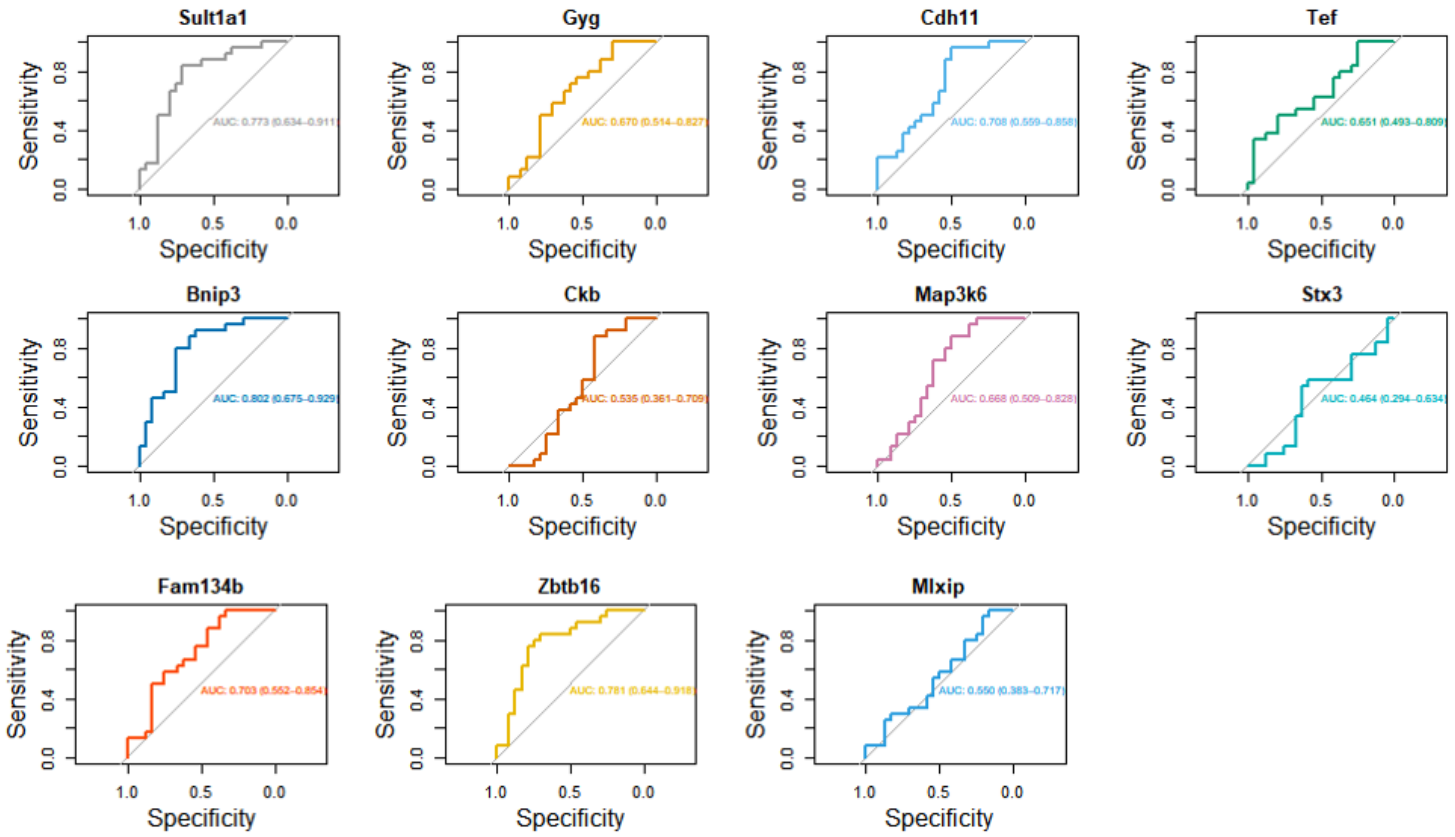

Figure 9

ROC analysis of 7 hub genes in the red module and 4 hub genes in the pink module.

Supplementary Files

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

- TableS1InformationOf48Samples.xlsx
- TableS2TheResultOfPowerEstimation.xlsx
- TableS3Geneanotationinredmodule.xlsx
- TableS4Geneanotationinpinkmodule.xlsx