



OPEN

Exploration of the role of oxidative stress-related genes in LPS-induced acute lung injury via bioinformatics and experimental studies

Yuanshui Liu^{1,3,4}✉, Huamei Li^{2,4}✉, Yanhong Ouyang¹, Yan Zhang³✉ & Pinhua Pan³✉

During the progression of acute lung injury (ALI), oxidative stress and inflammatory responses always promote each other. The datasets analyzed in this research were acquired from the Gene Expression Omnibus (GEO) database. The Weighted Gene Co-expression Network Analysis (WGCNA) and limma package were used to obtain the ALI-related genes (ALIRGs) and differentially expressed genes (DEGs), respectively. In total, two biological markers (Gch1 and Tnfaip3) related to oxidative stress were identified by machine learning algorithms, Receiver Operator Characteristic (ROC), and differential expression analyses. The area under the curve (AUC) value of biological markers was greater than 0.9, indicating an excellent power to distinguish between ALI and control groups. Moreover, 15 differential immune cells were selected between the ALI and control samples, and they were correlated to biological markers. The transcription factor (TF)-microRNA (miRNA)-Target network was constructed to explore the potential regulatory mechanisms. Finally, based on the quantitative reverse transcription polymerase chain reaction (qRT-PCR), the expression of Gch1 and Tnfaip3 was significantly higher in ALI lung tissue than in healthy controls. In conclusion, the differences in expression profiles between ALI and normal controls were found, and two biological markers were identified, providing a research basis for further understanding the pathogenesis of ALI.

Abbreviations

ALI	Acute lung injury
OSRGs	Oxidative stress-related genes
ALIRGs	ALI-related genes
DEGs	Differentially expressed genes
DE-ALI-OSRGs	Differentially expressed OSRGs in ALI
PPI	Protein-protein interactions
SVM-RFE	Support vector machine recursive feature elimination
ARDS	Acute respiratory distress syndrome
ROS	Reactive oxygen species
LPS	Lipopolysaccharide
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
GSEA	Gene set enrichment analysis

Acute respiratory distress syndrome (ARDS) is a heterogeneous syndrome of acute lung injury (ALI) in humans with multiple etiologies¹, which can be caused by direct lung injury from pneumonia and aspiration or indirect injury from non-pulmonary sepsis and pancreatitis². ARDS is mainly manifested by direct or indirect alveolar-capillary damage³. Cell death, loss of cell-cell junctions and/or cell-matrix attachment leads to epithelial (direct

¹Department of Emergency Medicine, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou 570311, People's Republic of China. ²Department of Ultrasound, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou 570311, People's Republic of China. ³Department of Respiratory Medicine, Key Cite of National Clinical Research Center for Respiratory Disease, Xiangya Hospital, Central South University, Changsha 410008, People's Republic of China. ⁴These authors contributed equally: Yuanshui Liu and Huamei Li. ✉email: yuanshui@hainmc.edu.cn; huameili@hainmc.edu.cn; zhangy4290@csu.edu.cn; pinhuapan668@csu.edu.cn

ARDS) or endothelial (indirect ARDS) damage and barrier dysfunction⁴, characterized by pulmonary capillary congestion, interstitial edema, alveolar edema, alveolar collapse, alveolar hemorrhage and hyaline membrane formation, resulting in non-cardiogenic pulmonary edema and hypoxic respiratory failure^{5,6}. The mortality rate of ARDS continues to be maintained at a high level of approximately 40% due to the lack of effective drug therapy¹.

Oxidative stress is the result of a disruption in the balance between reactive oxygen species (ROS) and antioxidant defense systems, and the resulting cellular damage is strongly associated with the occurrence and development of many human diseases⁷. Excessive production of ROS causes oxidative damage to molecules and cells under various pathological conditions, including ALI/ARDS, which also upregulates multiple inflammatory cytokines and tends to maintain a vicious malignancy of damage by recruiting more inflammatory cells, which ultimately leading to severe lung tissue damage⁸. Previous findings have shown that destructive oxidants produced by oxidative stress can directly damage tissues during the acute exudative phase of ALI/ARDS⁹. Also through interaction with the inflammatory response, some oxidants (e.g., ROS) act as inflammatory signaling molecules to activate NF- κ B, NLRP3 and other inflammatory pathways, exacerbating ALI/ARDS¹⁰. However, the mechanisms regulating oxidative stress in ALI are not fully elucidated.

Lipopolysaccharide (LPS) is a specific composition of the Gram negative bacterial cell wall, which is one of the main drivers of ALI/ARDS¹¹. It can indirectly damage alveolar epithelial cells by activating the release of inflammatory mediators from alveolar macrophages, neutrophils and other immune cells. Furthermore, it directly acts on TLR4 on alveolar epithelial cells, causing direct oxidative stress and inflammatory response in alveolar epithelial cells, leading to breakage of tight junctions between alveolar epithelial cells, increase in cell gaps and increased permeability¹².

In this study, we performed a series of bioinformatics analyses and basic experiments based on transcriptional profiling data of ALI mice samples and normal control mice lung tissues from the GEO database, aiming to analyze the mechanism of action of oxidative stress-related genes (OSRGs) in LPS-induced lung injury and the relationship with immune cell infiltration. Finally, we obtained the core genes of oxidative stress closely related to LPS-induced ALI, which were used to predict the drugs that may play a therapeutic role and provide a research basis for the study of the pathogenesis and treatment of ALI. The process of data analysis is illustrated in Fig. 1.

Results

Selection of 152 ALI-related genes (ALIRGs) by weighted gene co-expression network analysis (WGCNA)

The samples of the GSE16409, GSE18341 and GSE102016 datasets were discretely distributed before merging, and the sample data (ALI = 21 and control = 14) was uniform after batch processing (Supplementary Fig. 1a,b). To identify the ALIRGs, the WGCNA was performed in the combined dataset. As shown in the Fig. 2a,b, the cluster of samples was performed well with no outlier samples. The soft threshold was equal to 7 when the ordinate R^2 reached the threshold of 0.85 (red line). Simultaneously, the network was closer to a scale-free distribution, and the mean connectivity also close to 0. Thus, optimal soft threshold was selected as 7 (Fig. 2c). In total, 11 candidate modules were selected to obtain key module (Fig. 2d,e). The blue module had the highest correlation with ALI samples, which contained 1642 genes (Fig. 2f). Among 1642 genes, 152 genes were selected as ALIRGs with $|\text{Modulemembership (MM)}| > 0.8$ and $|\text{Genesignificance (GS)}| > 0.2$ (Fig. 2g).

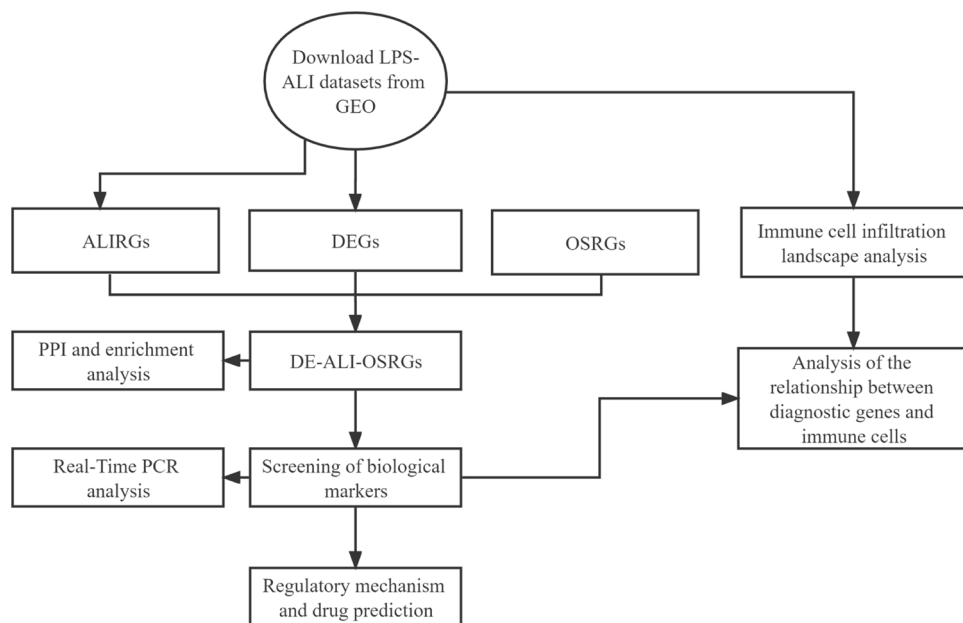


Figure 1. Logical flow of the analysis. ALIRGs ALI-related genes, DEGs differentially expressed genes, OSRGs oxidative stress-related genes, DE-ALI-OSRGs differentially expressed OSRGs in ALI.

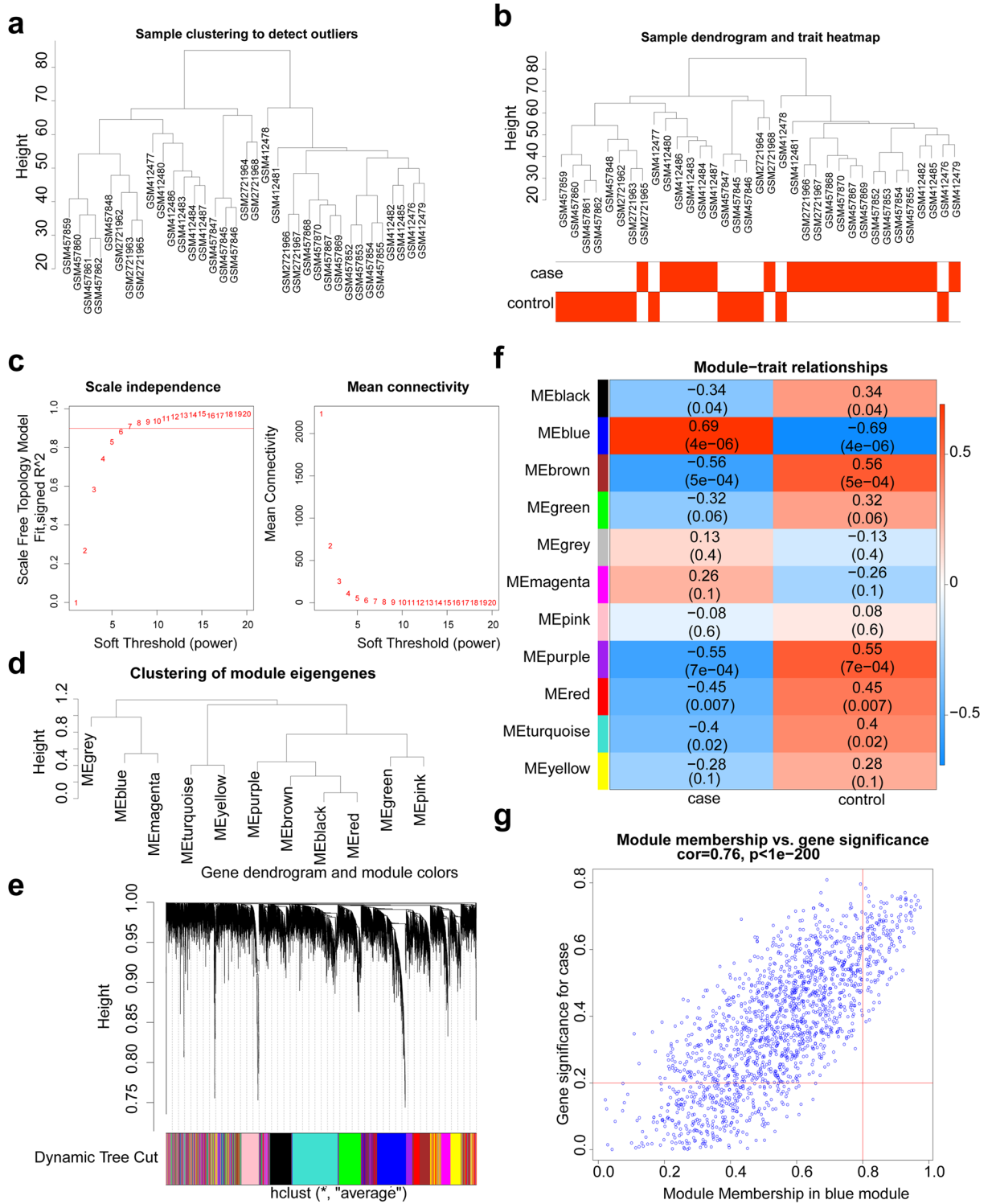


Figure 2. Weighted gene co-expression network analysis (WGCNA). (a) The clustering of samples in the merged dataset to remove outlier. (b) Clustering of merged data samples and phenotype information. (c) The determination of soft threshold. Seven was determined as the optimal soft threshold. (d) The clustering of module eigengenes. (e) Identification of gene co-expression modules. (f) Heatmap of correlation between modules and clinical traits. (g) The module membership (MM) and gene significance (GS scatter) plots of blue module.

Identification of 17 differentially expressed OSRGs in ALI (DE-ALI-OSRGs)

A total of 71 differentially expressed genes (DEGs) were selected with $|\log_2FC| > 1$ and $p < 0.05$ (Fig. 3a). Among them, 69 genes were up-regulated and 2 genes were down-regulated in the ALI samples. Heat map of the expression of up- and down-regulated genes in the ALI and control groups in the combined dataset was plotted (Fig. 3b). Finally, 17 DE-ALI-OSRGs were obtained by overlapping the ALIRGs, DEGs, and OSRGs (Fig. 3c). There were strong correlations among 17 DE-ALI-OSRGs (Fig. 3d). In order to investigate the potential molecular mechanisms of DE-ALI-OSRGs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were implemented^{13,14}. A total of 946 GO terms were enriched with $\text{adj.p} < 0.05$ and $\text{count} > 2$, including 903 BP terms and 43 MF terms. The top10 GO terms were shown in Fig. 3e, such as cellular response to molecule of bacterial origin, cellular response to biotic stimulus and leukocyte migration. Based on the z-score and \log_2FC , these GO terms were significantly enriched by up-regulated, and they were more likely to be increased. Meanwhile, 42 KEGG pathways were enriched with $\text{adj.p} < 0.05$ and $\text{count} > 2$. The top 20 KEGG pathways were shown in the Fig. 3f, such as IL-17 signaling pathway, TNF signaling pathway, and cytokine-cytokine receptor interaction.

Gch1 and Tnfaip3 were screened as biological markers

To screen ALI candidate biological markers, the Least-Absolute Shrinkage and Selection Operator (LASSO) and support vector machine recursive feature elimination (SVM-RFE) models were performed in the combined dataset based on the DE-ALI-OSRGs. Then, 6 candidate biological markers were selected with the $\lambda_{\text{min}} = 0.019$ by LASSO, including Arg2, Ccl4, Gch1, Hpx, Soc3, and Tnfaip3 (Fig. 4a). Moreover, 15 candidate genes were selected by SVM-RFE based on gene importance ranking and error rate (Fig. 4b, Table S1). Finally, the 5 overlapping genes were selected from the results of LASSO and SVM-RFE as candidate biological markers (Fig. 4c), including Arg2, Ccl4, Gch1, Hpx and Tnfaip3.

To further excavate the biological markers, expression analysis was performed and ROC curve was painted in the combined dataset, GSE104214 dataset, and GSE17355 dataset, respectively. Figure 5a–c demonstrated the expression of Gch1 and Tnfaip3 was significantly different between ALI and control samples, and the expression was higher in ALI samples in all three datasets. In addition, the AUC value of 4 genes (Arg2, Ccl4, Gch1 and Tnfaip3) was all greater than 0.7, indicating a decent ability to distinguish between ALI and control samples (Fig. 5d–f). In summary, Gch1 and Tnfaip3 with differential expression and AUC values greater than 0.9 were treated as biological markers for subsequent analysis. To explore the interaction among two biological markers, the protein-protein interaction (PPI) network was constructed. As shown in Fig. 5g, Tnfp2, Gchfr, and Tnfp1 had stronger interaction with biological markers. Additionally, both Gch1 and Tnfaip3 were associated with cellular response to molecule of bacterial origin and response to lipopolysaccharide.

Gene set enrichment analysis (GSEA) enrichment analysis

GSEA enrichment analysis was performed on the four biological markers in the combined dataset using the default background gene set in the org.Mm.eg.db package, and the significance threshold for ssGSEA was $|\text{NES}| > 1$, $p < 0.05$, and $q < 0.2$. The Gch1 was mainly enriched in 1330 GO terms and 115 KEGG pathways (Table S2). As shown in Fig. 6a, Gch1 was associated with toll-like receptor signaling pathway and pathways related to cytokine and immune, such as adaptive immune response, immune effector process, positive/negative regulation of cytokine production and pattern recognition receptor signaling pathway in GO terms. As for KEGG pathways, Gch1 was associated with IL-17 signaling pathway, NOD-like receptor signaling pathway, TNF signaling pathway, and Toll-like receptor signaling pathway (Fig. 6a). Moreover, Tnfaip3 was mainly enriched in 1227 GO terms and 103 KEGG pathways (Table S3). As shown in Fig. 6b, response to virus and bacterium and immune-related pathways (such as adaptive immune response and innate immune response) were enriched in GO terms of Tnfaip3 (Fig. 6b). As for KEGG pathways, Tnfaip3 was associated with IL-17 signaling pathway, NF-kappa B signaling pathway, NOD-like receptor signaling pathway, TNF signaling pathway, and cytokine receptor-related pathways. In conclusion, two biological markers were linked with the occurrence and development of ALI.

Immune cell infiltration landscape analysis

The infiltration levels of 28 immune cells in the ssGSEA algorithm were all higher in ALI samples than in control samples (Fig. 7a). A total of 15 types of immune cells were significantly different between the ALI samples and control samples ($p < 0.05$), including activated CD4 T cell, activated CD8 T cell, activated dendritic cell, effector memory CD8 T cell, gamma delta T cell, macrophage, mast cell, MDSC, natural killer cell, natural killer T cell, neutrophil, regulatory T cell, T follicular helper cell, type 1 T helper cell, and type 17 T helper cell (Fig. 7b). Subsequently, there are correlations among 15 differential immune cells. Among 15 differential immune cells, the mast cell were strongly negatively correlated with natural killer cell, and the regulatory T cell was strongly positively correlated with macrophage (Fig. 7c). The correlations between Gch1 and the immune cells of macrophage and natural killer T cell were greater than 0.8 ($|\text{cor}| > 0.8$); the correlation between Tnfaip3 and the immune cells of natural killer T cell immune cell was greater than 0.8 ($|\text{cor}| > 0.8$) (Fig. 7d).

Regulatory mechanism of biological markers and drug prediction

The TF-miRNA-Target network of two biological markers was obtained using miRNet database (<https://www.mirnet.ca/>) containing 8 TFs, 2 Target genes and 56 miRNAs. The network consisted 66 nodes and 70 relation pairs (Supplementary Fig. 2a). The mmu-let-7c-5p, mmu-mir-124-3p, and mmu-mir-181b-5p were common miRNAs predicted by two biological markers (Supplementary Fig. 2a). Additionally, a Drug-Disease-Target network for the treatment of ALI was constructed, containing 1 disease, 2 target genes, and 3 drugs. The network

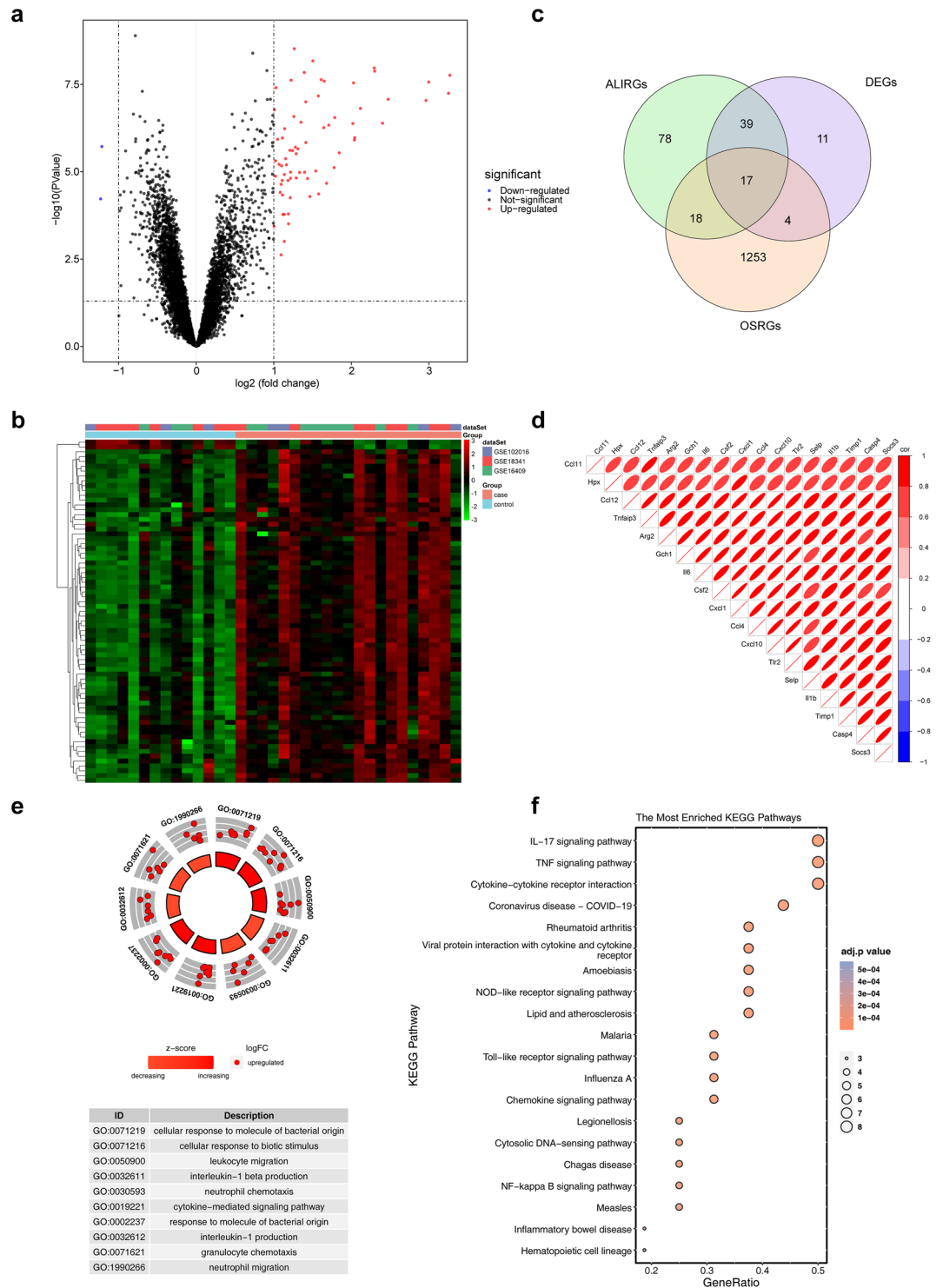


Figure 3. Identification of differentially expressed OSRGs in ALI (DE-ALI-OSRGs) and functional enrichment analysis. **(a)** The volcano map of differentially expressed genes (DEGs) between ALI and control samples. **(b)** The heatmap of DEGs between ALI and control samples. **(c)** Venn diagram for certification of DE-ALI-OSRGs. **(d)** The correlation between DE-ALI-OSRGs. The shadow of the ellipse in each color box represents the size of the correlation between two genes. The greater the correlation, the narrower the ellipse. The smaller the correlation, the rounder the ellipse. Red represents a positive correlation and blue represents a negative correlation. **(e)** The gossip chart of top10 Gene Ontology (GO) terms enriched by DE-ALI-OSRGs. Z-score is an value which give a hint if the biological process (/molecularfunction/cellular components) is more likely to be decreased or increased. LogFC is used to represent the number of up- and down-regulated genes. **(f)** The bubble chart of top20 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched by DE-ALI-OSRGs.

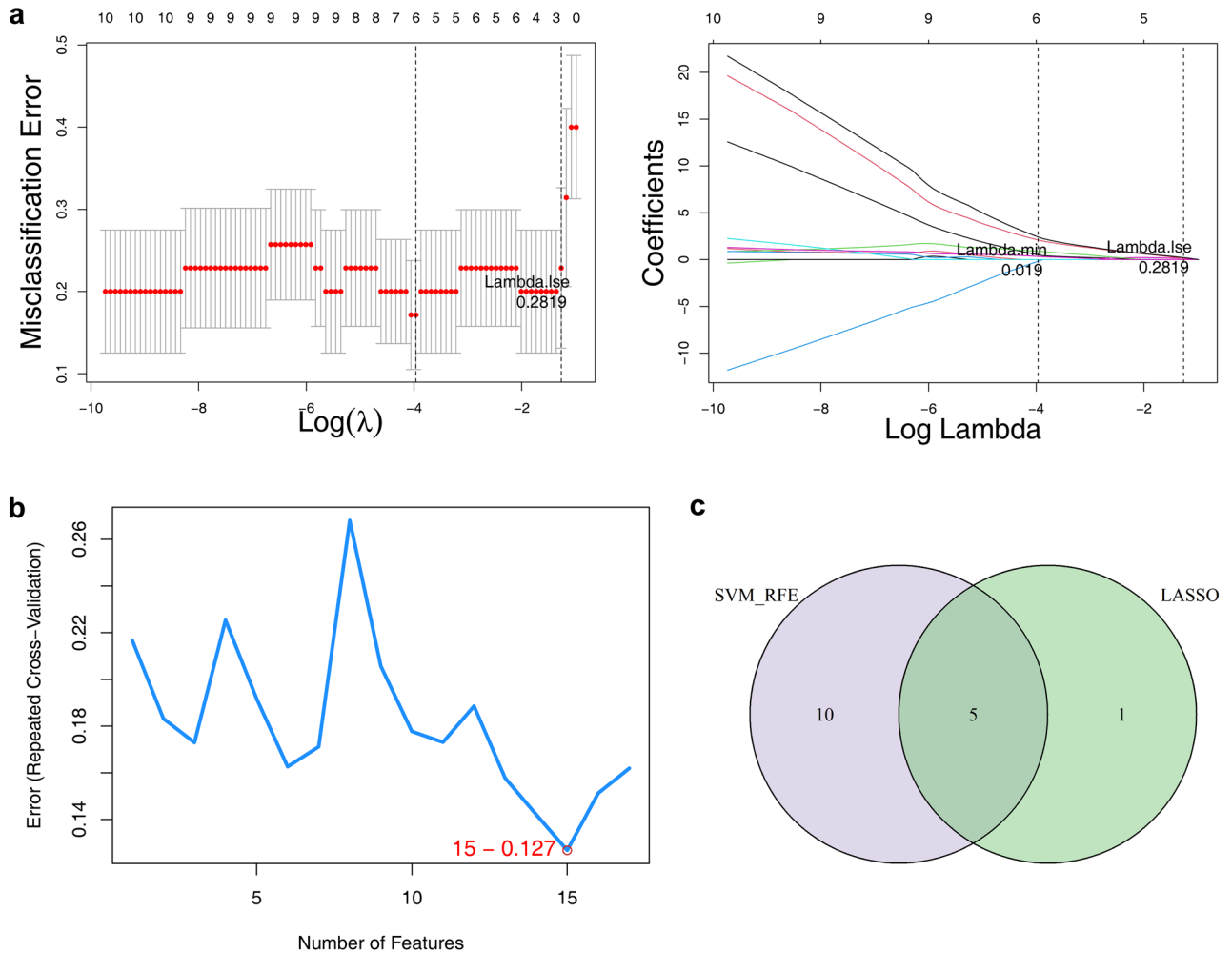


Figure 4. Identification of candidate biological markers. **(a)** Logic factor penalty plot and cross validation error curve of Least-Absolute Shrinkage and Selection Operator (LASSO) model. Each curve represents the trajectory of each independent variable coefficient. **(b)** The accuracy of support vector machine recursive feature elimination (SVM-RFE) model. **(c)** Wayne chart of characteristic genes identified by LASSO and SVM-RFE.

consisted 6 nodes and 5 relationship pairs (Supplementary Fig. 2b). The GUANINE was predicted by Gch1; the USTEKINUMAB and METHOTREXATE were predicted by Tnfaip3 (Supplementary Fig. 2b).

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis

In our mice model, the mRNA levels of the two core genes of oxidative stress, including Gch1 and Tnfaip3, were significantly higher in ALI lung tissue than in healthy controls (Fig. 8).

Discussion

LPS is a specific component of the extracellular membrane of Gram-negative bacteria and is one of the major pathogenic factors causing sepsis and ALI¹⁵. Mice exhibited a systemic inflammatory response 18 h after intra-peritoneal LPS administration¹⁶. Oxidative stress also plays a key role in the development of ALI and ARDS. Intra-tissue homeostasis requires the maintenance of a complex and delicate balance between oxidants and antioxidants. The disruption of this balance could result in the continued generation of ROS and exceeding the capacity of the antioxidant defense system, which would further lead to the damages of DNA, protein and lipid. This situation could contribute apoptosis¹⁷, causing pulmonary edema and excessive inflammatory cell infiltration to lung injury^{10,18}.

In the present study, 17 core genes were obtained from three independent datasets of differential genes from mice containing ALI and normal mice, and functionally they were associated with the roles of cellular responses to molecules of bacterial origin, cellular responses to biological stimuli and leukocyte migration. They were closely involved in signaling pathways such as IL-17 signaling pathway, TNF signaling pathway and cytokine-cytokine receptor interaction, all of which were associated with inflammatory responses and directly involved in the regulation of ALI. The abnormal expression of these genes and the disruption of their regulated signaling pathways may be related to the development of ALI. Subsequently, we further screened two biomarkers, including Gch1 and Tnfaip3, which were associated with cellular responses to bacterial-derived molecules

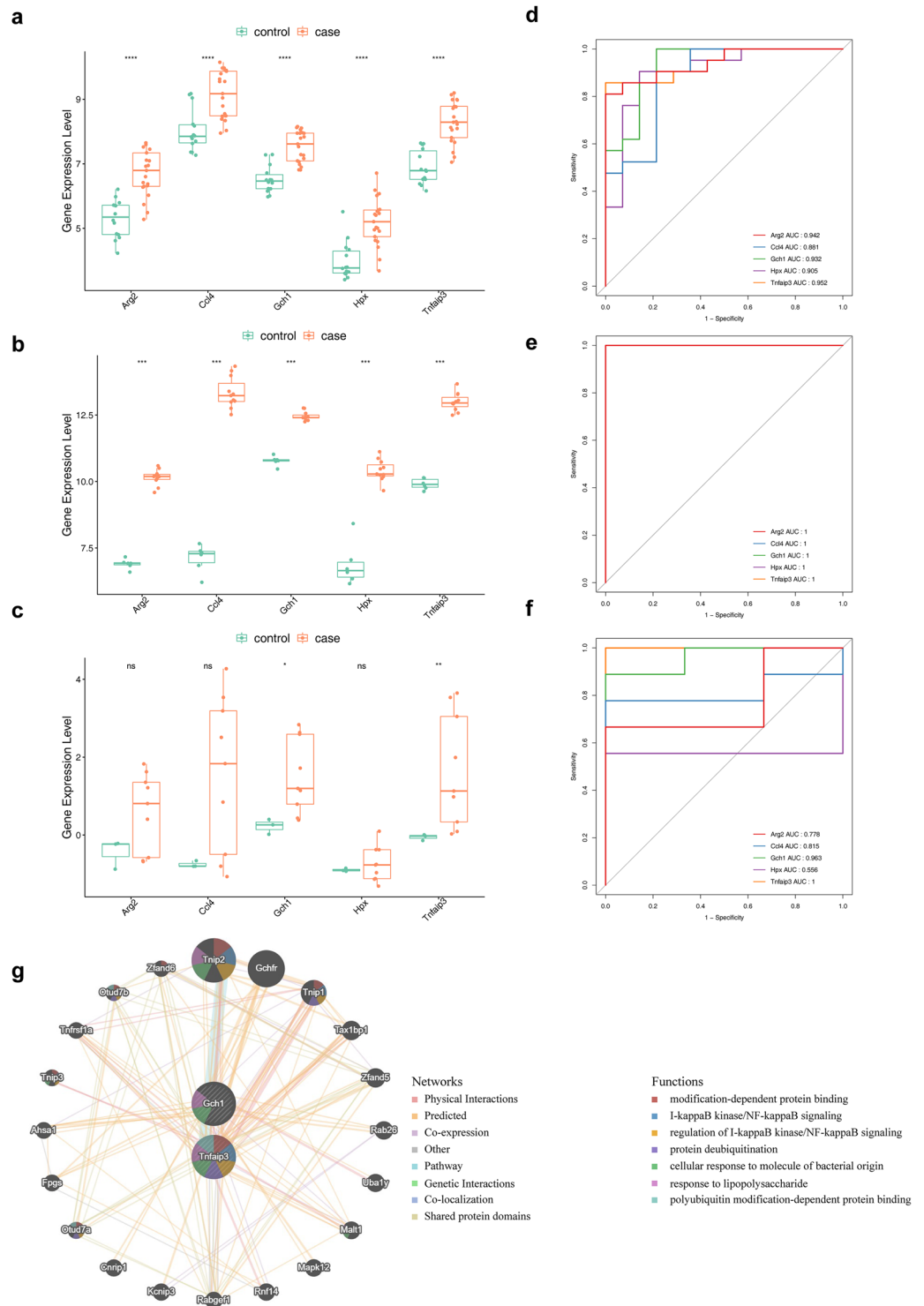


Figure 5. Identification of biological markers. **(a)** The expression levels of key genes in the merged dataset, ****p value < 0.0001. **(b)** The expression levels of key genes in the GSE104214 dataset, ***p value < 0.001. **(c)** The expression levels of key genes in the GSE17355 dataset, ns not significant, *p value < 0.05, **p value < 0.01. **(d)** Receiver Operator Characteristic (ROC) curves of key genes in the merged dataset. AUC area under the curve. **(e)** ROC curves of key genes in the GSE104214 dataset. **(f)** ROC curves of key genes in the GSE17355 dataset. **(g)** The protein–protein interaction (PPI) network of biological markers.

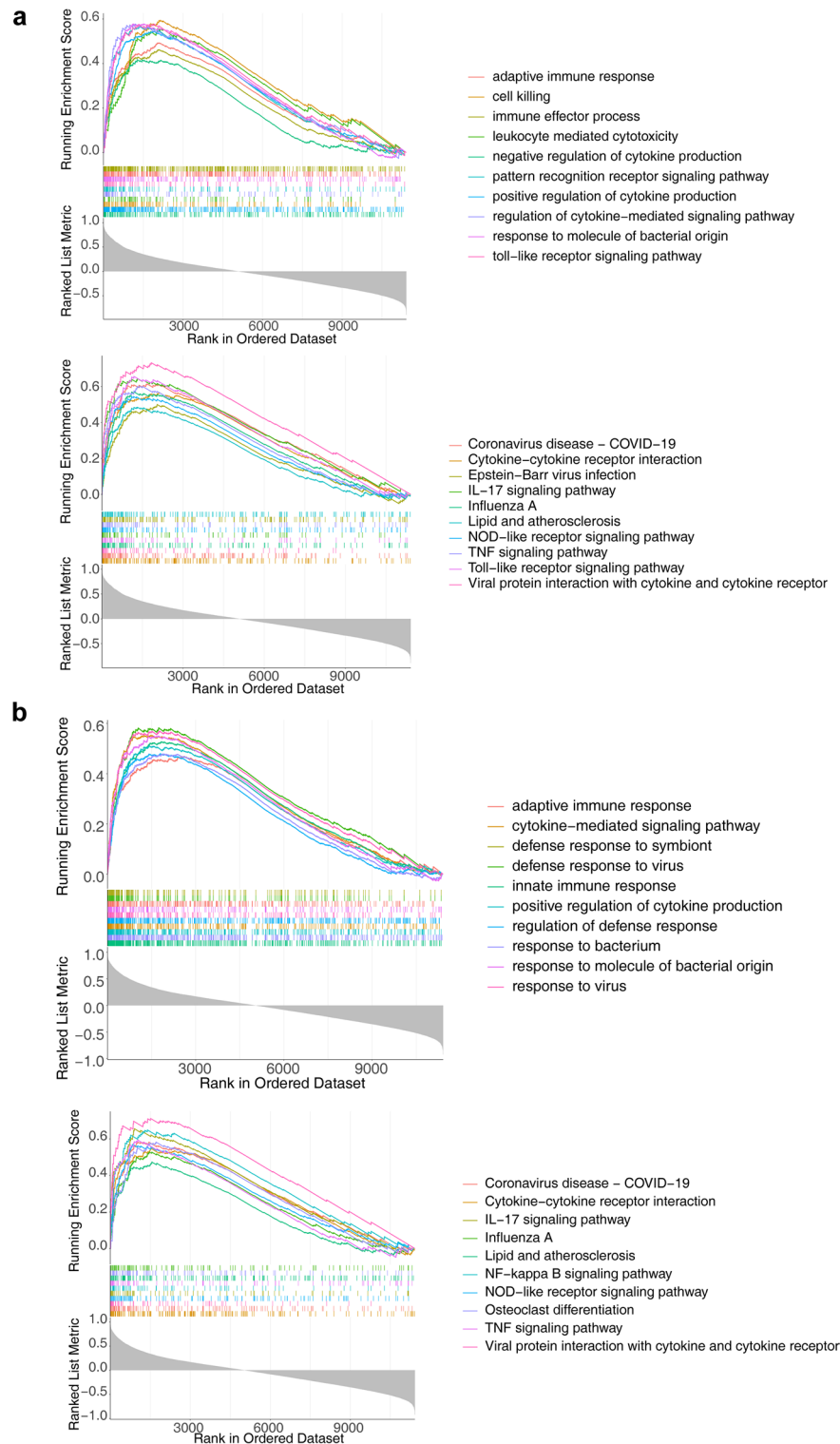


Figure 6. Gene set enrichment analysis (GSEA) of the four biological markers. **(a)** Top10 GO terms and KEGG pathways enriched by Gch1. **(b)** Top10 GO terms and KEGG pathways enriched by Tnfaip3.

and lipopolysaccharides and were probably the hub molecules in the pathogenesis of sepsis-associated ALI. The above results were validated by intraperitoneal injection of LPS to establish a mice ALI model, and the mRNA levels of Gch1 and Tnfaip3 were found to be significantly elevated in the lung tissues of ALI mice.

Gch1 is the rate-limiting enzyme for the production of tetrahydrobiopterin (BH4) in the biosynthetic pathway¹⁹. Higher Gch1 expression contributes to lower levels of oxidative stress²⁰. We found that the levels of Gch1 expression were diametrically opposed in different models of cause-induced lung injury. Gch1 expression

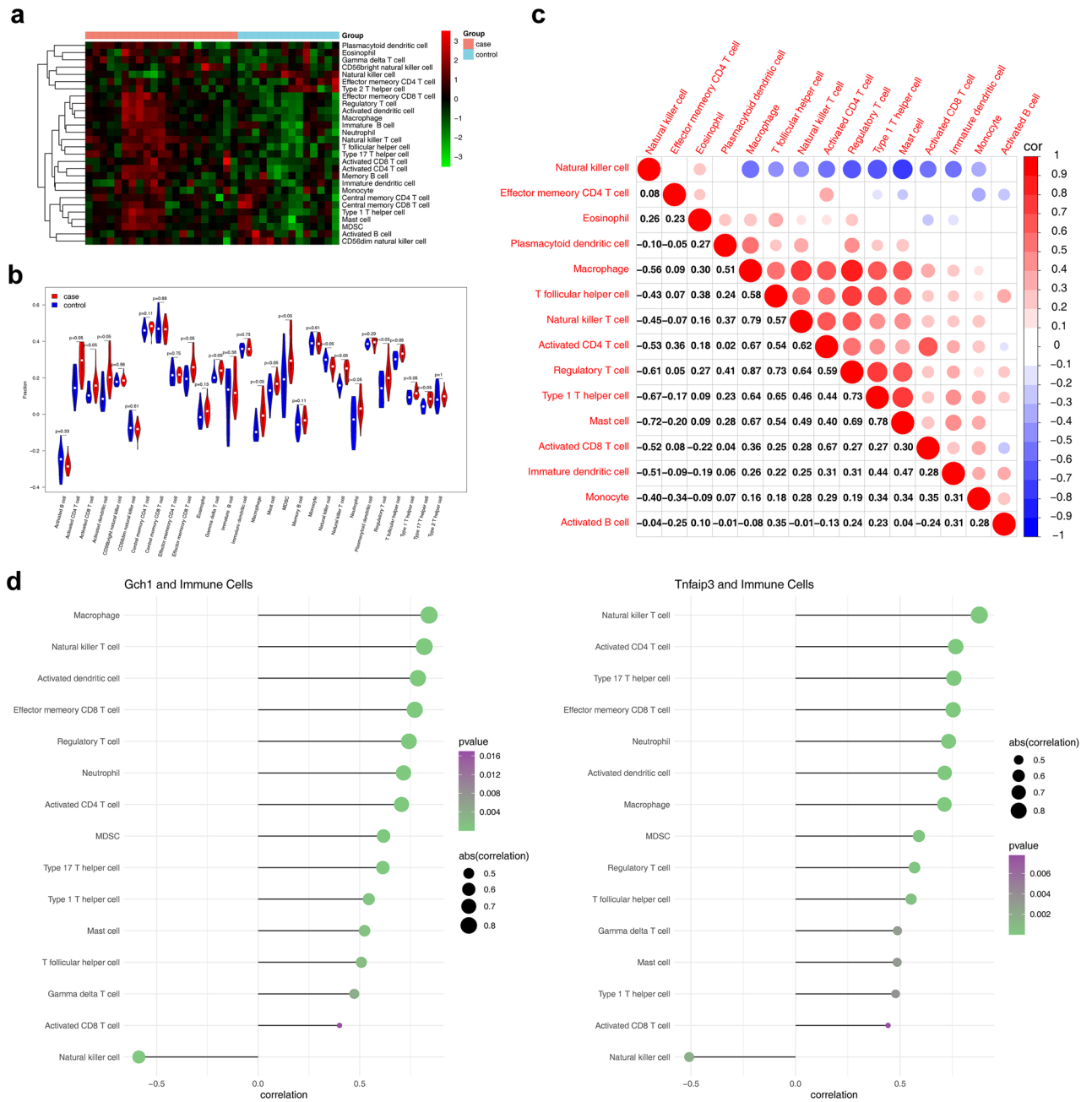


Figure 7. Identification of differential immune cells and analysis of correlation with diagnostic genes. **(a)** The heat map of 28 immune cell infiltration scores. **(b)** The discrepancies of immune cell infiltration between ALI and normal samples. **(c)** The correlation among differential immune cells. Red represents a positive correlation and blue represents a negative correlation. **(d)** The correlation between biological markers and immune cells. The color of the circle represents the direction of the correlation and the size represents the size of the correlation.

was decreased in the lungs of mice with hyperoxia-induced lung injury²¹. However, our results showed that Gch1 levels were elevated in the lungs of mice induced by LPS to mimic lung injury caused by bacterial infection compared to healthy controls. This difference may also be related to the difference in Gch1 expression caused by different stimuli. It has been hypothesized that LPS might better stimulate elevated Gch1 responsiveness in oxidative stress. However, this change could also be related to different time periods in which the material was taken for testing. We speculate that a possibility leading to this phenomenon cannot be excluded, that Gch1 is elevated at the beginning of lung injury and subsequently decreased. Whether increasing Gch1 levels within days of lung injury can alleviate lung injury has not been investigated.

TNFAIP3 is an important inhibitor of the pro-inflammatory factor- κ B pathway and plays a key role in a variety of diseases. Krusche et al.²² performed ex vivo stimulation of Peripheral blood mononuclear cells with LPS and found that activation of the anti-inflammatory process was achieved by increased expression of Tnfaip3.

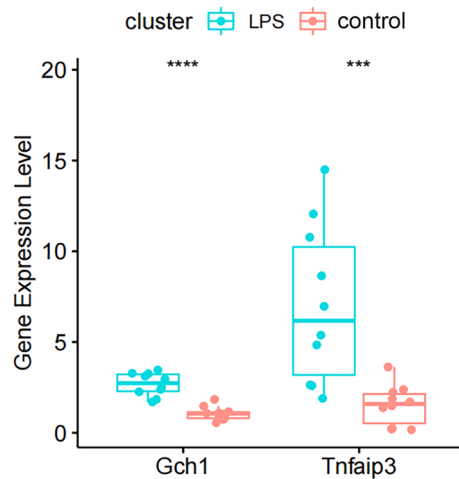


Figure 8. The relative expression of Gch1 and Tnfaip3 in the mice model, *** p value < 0.001, **** p < 0.0001.

After liver injury, TNFAIP3 exerts dual anti-apoptotic, anti-inflammatory and pro-proliferative effects²³. Elevated expression of Tnfaip3 can inhibit the progression of inflammation during an inflammatory attack.

In this study, we analyzed the changes in immune cells in ALI and their interrelationships. Among them, mast cells showed a strong negative correlation with NK cells and Treg showed a strong positive correlation with macrophages. Mast cells are engaged in the promotion of inflammatory responses in LPS-induced lung injury and that inhibition of mast cell activation contributes to the suppression of pro-inflammatory gene expression during LPS-induced ALI²⁴. During LPS-induced ALI, mice NK cells promote chemokine-mediated neutrophil recruitment and promote inflammation, while depletion of NK cells ameliorates this outcome²⁵. After receiving of viral invasion, mast cells secrete cytokines that chemotactic NK cell aggregates and cause them to activate and secrete IFN- γ ²⁶. This view cannot explain the results of the negative correlation between mast cells and NK cells in this study. Whether the activation of NK cells can form negative feedback on mast cells by inhibiting the production level of mast cells, and the reasons for the negative correlation between the NK cells and mast cells needs further research to explore.

This study revealed correlations between each of the two pivotal genes and immune cells. Gch1 and Tnfaip3 showed the strongest positive correlation with macrophages and NK T cells, respectively. Meanwhile, Gch1 and Tnfaip3 were significantly negatively correlated with NK cells. GCH1 induces immunosuppression of TNBC through metabolic reprogramming and IDO1 upregulation²⁷; Xiao²⁸ experimentally demonstrated that GCH1 reduces LPS-induced macrophage polarization and inflammation. Furthermore, it has been shown that selective deletion of TNFAIP3 in mice leads to worsening of systemic inflammation and inflammatory dermatoses under homeostatic conditions²⁹. Thus, we suggest that GCH1 and TNFAIP3 play a negative regulatory role in the body's immune system. In summary, we hypothesized that in the LPS-induced ALI model, the abnormally elevated expression of Gch1 and Tnfaip3 negatively regulated NK cells, which disorganized the body's immune system, and consequently led to the development of the disease.

The medication that acts on Gch1 is probably GUANINE, which has analogs such as VALACYCLOVIR HYDROCHLORIDE. Guanine inhibits the activity of both Gch1 and Gch1 feedback regulatory proteins. However, Gch1 is expected to be increased thus reducing the level of oxidative stress to alleviate the condition in ALI. Therefore, we believe that GUANINE does not alleviate ALI.

Drugs that may act on Tnfaip3 might be USTEKINUMAB and METHOTREXATE (MTX)³⁰. The response of psoriasis patients to USTEKINUMAB was associated with Tnfaip3 gene polymorphism³¹, implying that the pharmacotherapeutic effect exerted by USTEKINUMAB might be depend to some extent on the expression of Tnfaip3 or its protein function. Tnfaip3 acts as a negative regulator of nuclear factor-kB, which regulates the inflammatory response of tumor necrosis factor (TNF) by inhibiting the upstream signaling of kB kinase (IKK)³². In contrast, MTX³⁰ inhibited nuclear factor-kB activation by inhibiting IKK and did not reveal the possibility of altering Tnfaip3 expression or protein function, but may contribute to mitigate the effects of ALI.

This research was the first time to systematically investigate the role of oxidative stress-related genes in ALI and perform immune infiltration-related analysis by bioinformatics technology based on the data in the GEO database. In addition, we also performed preliminary validation of our findings by constructing an animal model. However, there were also shortcomings in this study. First, our analysis was developed based on a limited sample of public databases, and expanding the sample size was an urgent issue. Although we obtained biological markers related to oxidative stress for ALI screening, their roles and mechanisms needed to be further investigated and validated. In addition, we had predicted potential drugs based on biomarkers, their effectiveness needed to be validated in the clinic.

Conclusion

In conclusion, 2 OSRGs (Gch1 and Tnfaip3) with higher expression in ALI samples than in control samples were identified as biological marker for LPS-induced ALI, and they may be involved in the immune-related pathways. In addition, 15 differential immune cells might play an important role in the development and progression of ALI, especially macrophage, natural killer T cell, and natural killer cell. USTEKINUMAB, MTX, and GUANINE may be potential therapeutic agents to alleviate ALI. Thus, we believe that our findings will provide a new theoretical basis for further research on the role of oxidative stress in ALI, and will also provide new targets for the diagnosis and treatment of ALI. We will continue to focus on the role of Gch1 and Tnfaip3 in ALI and further explore their mechanisms of action.

Materials and methods

Data source

The GSE102016, GSE104214, GSE16409, GSE17355, and GSE18341 datasets were acquired from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Among these datasets, the untreated samples from the control group and wild-type samples treated with LPS were selected for subsequent analysis, and the data information was shown in Table S4. Additionally, 1399 OSRGs were downloaded from the GeneCards database (<https://www.genecards.org>) with Relevance score ≥ 7 , and then transformed these human genes to obtain 1292 mice homologous genes. The GSE16409, GSE18341 and GSE102016 datasets were background corrected, quantile normalized and merged as a combined dataset, then batch effects were removed using the SVA package (version 3.42.0)³³.

Screening of ALIRGs

The R package WGCNA (version 1.7-3)³⁴ was used to construct a co-expression network in the combined dataset, and the ALI samples and control samples were used as the trait data of WGCNA to search for ALIRGs. Firstly, the samples were clustered and outlier samples were removed to ensure the accuracy of the further analysis. Then a sample cluster and the heatmap of clinical traits were constructed. The soft threshold of the data determined to ensure that the interaction between genes conformed to the scale-free distribution to the greatest extent. The phylogenetic clustering tree among genes was obtained on the basis of the adjacency relationship and the similarity between genes. The minimum number of genes in each gene module was set to 200 according to the criteria of the hybrid dynamic tree cutting algorithm. Subsequently, modules with the highest disease relevance and key ALIRGs they contained were selected.

Screening of DE-ALI-OSRGs

The limma package (version 3.50.0)³⁵ was used to compare the differences in gene expression between the ALI samples and control samples in the combined dataset. The ggplot2 package (version 3.3.5)³⁶ was used to draw volcano plots to show the DEGs. The VennDiagram (version 1.7.1)³⁷ was used to obtain the DE-ALI-OSRGs, that was the intersection of the ALIRGs, DEGs and OSRGs. In addition, the correlations between intersecting genes were calculated.

Enrichment analysis

The clusterProfiler package (version 4.2.2)³⁸ was used to implement the GO and the Kyoto Encyclopedia of KEGG enrichment of DE-ALI-OSRGs, and the enrichment results were visualized using GPlot (version 1.0.2)³⁹ and enrichplot packages (version 1.10.2)⁴⁰.

Screening of biological markers

The LASSO and SVM-RFE machine learning models were constructed in the combined dataset to screen candidate biological markers in the ALI based on the DE-ALI-OSRGs. The LASSO algorithm was implemented by “glmnet” package⁴¹ (version 4.0-2) with parameters set to `famil = binomial` and `type.measure = class`. The e1071 package⁴² (version 1.7-9) was employed to implement SVM-RFE algorithm. The overlapping genes were selected from the results of the LASSO and SVM-RFE, and these genes were considered as candidate biological markers.

In the combined dataset, GSE104214 dataset and GSE17355 dataset, the `wilcox.test` was performed to verify the differential expression of biological markers between the ALI samples and control samples, and scatter points graphs were visualized to verify the expression levels of the candidate biological markers by the `ggpubr` package (version 0.40)⁴³.

To explore the prediction of candidate biological markers on sample traits and select the biological markers, ROC analysis was performed on the combined dataset, GSE104214 dataset, and GSE17355 datasets for candidate biological markers using the `pROC` package⁴⁴ (version 1.18.0). The genes with differential expression in both datasets and area under the curve (AUC) values greater than 0.9 were treated as biological markers. Finally, the STRING (<https://string-db.org>) website was used to construct a PPI network with the confidence = 0.4 to explore the interaction among biological markers. The PPI network was drawn by Cytoscape (version 3.8.2)⁴⁵ software.

GSEA enrichment analysis

The clusterProfiler (version 4.2.2)³⁸ was applied to perform the GSEA enrichment analysis to find the functions and pathways of the biological markers. GSEA enrichment analysis was performed based on the default background gene set in the `org.Mm.eg.db` package.

Immune cell infiltration landscape analysis

In order to study the difference of immune infiltration between the ALI samples and control samples, ssGSEA was used to estimate the infiltration of immune cell types in the *gsva* package (version 1.42.0)⁴⁶. The *wilcox.test* was used to compare the differences between different immune cells in the combined dataset between the ALI samples and control samples, and the violin plot was drawn using the *vioplot* package (version 0.3.7) to visualize the comparison results. Pearson correlation analysis was performed on the different immune cells using the R language and plotted correlation using the *corrplot* package (version 1.0.12)⁴⁷. In addition, the Pearson correlation coefficient between biological markers and immune cells based on the combined dataset were delved.

Regulatory mechanism of diagnostic genes and drug prediction

The transcriptional regulatory network of TF-miRNA-Target was predicted by biological markers in the miR-Net database (<https://www.mirnet.ca/>). The drug-gene interaction database (DGIdb; <https://dgidb.genome.wustl.edu/>) identified biological markers for therapeutic drugs. The networks mentioned above were drawn by Cytoscape software (version 3.8.2)⁴⁵.

Animal experiments

Mice

The Mice (C57BL/6 mice, male, aged 8–10 weeks) of this study were purchased from the Hunan Shrek Jingda Experimental Animal Co., Ltd. (Changsha, China).

LPS/ALI model

To establish an LPS-induced ALI model, 20 C57BL/6 mice were divided equally into 2 groups, 10 of which were injected intraperitoneally with a single dose of PBS premixed with LPS (10 mg/kg, Sigma Aldrich, Cedex, France) while the other 10 with PBS liquid⁴⁸. LPS-induced ALI mice were molded for 20 h, 1% pentobarbital (50 mg/kg) was injected intraperitoneally for anesthesia, then the eyes were removed and bled, the mice were sacrificed after cervical dislocation, and lung tissue was taken for real-time quantitative PCR analysis. The studies involving animal experiments were reviewed and approved by the Animal Care and Use Committee of Central South University (CSU-2023-0080), all methods were carried out in accordance with relevant guidelines and regulations. All procedures performed in this study involving animal experiments were in accordance with the ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments)⁴⁹. Moreover, the execution of animals was according to the American Veterinary Medical Association (AVMA) Guidelines for the Euthanasia of Animals (2020).

Validation of the biological markers by qRT-PCR

Mice lung tissue were treated with TRIzol reagent (50 mg tissue/1 mL, Invitrogen, Carlsbad, CA) at room temperature for 5 min, then centrifuged at 4000×g for 15 min at 4 °C with chloroform. The upper aqueous phase was aspirated into a new EP tube, treated with isopropyl alcohol for 10 min, and centrifuged at 4000×g for 10 min (both at 4 °C). The precipitate was washed with 75% ethanol, and the RNA was solubilized by adding RNase-free DEPC and stored at –80 °C. After the detection of RNA purity and concentration using the spectrophotometer (Thermo Fisher, Waltham, MA, USA), TransScript® Reverse Transcriptase [M-MLV, RNaseH-] (AT101-02, Transgen, Beijing, China) was used for the reverse transcription. The qPCR reactions were then performed on an ABI 7900HT Fast platform using the TransStart® Green qPCR SuperMix (AQ101-01, TransGen, Beijing, China) according to the manufacturer's instructions⁵⁰. The sequences of the forward and reverse primers were displayed in Table 1 (Tsingke Biotech, Beijing, China). The relative expression of each biological markers was determined by the $2^{-\Delta\Delta C_t}$ method with GAPDH as the internal reference.

Statistical analysis

All analyses were conducted using R language (<https://www.r-project.org/>). The *wilcox.test* was utilized to evaluate the differential expression levels of biological markers between the ALI samples and control samples. All experimental data was expressed as mean ± standard deviation. Comparisons between the two groups were implemented using the Student's t-test. If not specified above, $p < 0.05$ was regarded as statistically significant.

Primer	Sequence
Mice GAPDH forward	5'-ACGGCACAGTCAAGGCAGA-3'
Mice GAPDH reverse	5'-GTGATGGCGTGGACAGTGG-3'
Mice Gch1 forward	5'-GTCCTTGGTCTCAGTAAACTTGCCAGG-3'
Mice Gch1 reverse	5'-GCCAGCCAAGGATAGATGCAG-3'
Mice Tnfaip3 forward	5'-TCAACTGGTGTGCGAGAAGTCC-3'
Mice Tnfaip3 reverse	5'-CAAGTCTGTGTCCTGAACGC-3'

Table 1. The sequences of primers for biological markers.

Ethics statement

The GEO database is a public dataset. The animals involved in the database have been approved by ethics. Users can download the relevant data for free to conduct research and publish related articles. Our research is based on open-source data. The animal experiments in this study followed the ethical requirements of the Animal Care and Use Committee of Central South University (No. 2017sydw00284), the ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments) and the American Veterinary Medical Association (AVMA) Guidelines for the Euthanasia of Animals (2020).

Data availability

The datasets generated and analyzed during the present study are available in the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) database (Accession Number: GSE102016, GSE104214, GSE16409, GSE17355 and GSE18341)/Supplementary Material.

Received: 19 July 2023; Accepted: 5 December 2023

Published online: 09 December 2023

References

- Bellani, G. *et al.* Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA* **315**(8), 788–800. <https://doi.org/10.1001/jama.2016.0291> (2016).
- Sinha, P. & Calfee, C. S. Phenotypes in acute respiratory distress syndrome: Moving towards precision medicine. *Curr. Opin. Crit. Care* **25**(1), 12–20. <https://doi.org/10.1097/mcc.0000000000000571> (2019).
- Hawiger, J., Veach, R. A. & Zienkiewicz, J. New paradigms in sepsis: From prevention to protection of failing microcirculation. *J. Thromb. Haemost.* **13**(10), 1743–1756. <https://doi.org/10.1111/jth.13061> (2015).
- Coletta, C. *et al.* Endothelial dysfunction is a potential contributor to multiple organ failure and mortality in aged mice subjected to septic shock: Preclinical studies in a murine model of cecal ligation and puncture. *Crit. Care* **18**(5), 511. <https://doi.org/10.1186/s13054-014-0511-3> (2014).
- Fan, E., Brodie, D. & Slutsky, A. S. Acute respiratory distress syndrome: Advances in diagnosis and treatment. *JAMA* **319**(7), 698–710. <https://doi.org/10.1001/jama.2017.21907> (2018).
- Englert, J. A., Bobba, C. & Baron, R. M. Integrating molecular pathogenesis and clinical translation in sepsis-induced acute respiratory distress syndrome. *JCI Insight* **4**(2), 4061. <https://doi.org/10.1172/jci.insight.124061> (2019).
- Pizzino, G. *et al.* Oxidative stress: Harms and benefits for human health. *Oxid. Med. Cell Longev.* **2017**, 8416763. <https://doi.org/10.1155/2017/8416763> (2017).
- Kellner, M. *et al.* ROS signaling in the pathogenesis of acute lung injury (ALI) and acute respiratory distress syndrome (ARDS). *Adv. Exp. Med. Biol.* **967**, 105–137. https://doi.org/10.1007/978-3-319-63245-2_8 (2017).
- Huang, X., Xiu, H., Zhang, S. & Zhang, G. The role of macrophages in the pathogenesis of ALI/ARDS. *Mediat. Inflamm.* **2018**, 1264913. <https://doi.org/10.1155/2018/1264913> (2018).
- Yang, H., Lv, H., Li, H., Ci, X. & Peng, L. Oridonin protects LPS-induced acute lung injury by modulating Nrf2-mediated oxidative stress and Nrf2-independent NLRP3 and NF- κ B pathways. *Cell Commun. Signal* **17**(1), 62. <https://doi.org/10.1186/s12964-019-0366-y> (2019).
- Leiva-Juárez, M. M., Kolls, J. K. & Evans, S. E. Lung epithelial cells: Therapeutically inducible effectors of antimicrobial defense. *Mucosal Immunol.* **11**(1), 21–34. <https://doi.org/10.1038/mi.2017.71> (2018).
- Nova, Z., Skovierova, H. & Calkovska, A. Alveolar-capillary membrane-related pulmonary cells as a target in endotoxin-induced acute lung injury. *Int. J. Mol. Sci.* **20**(4), 831. <https://doi.org/10.3390/ijms20040831> (2019).
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070> (2016).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
- Welbourn, C. R. & Young, Y. Endotoxin, septic shock and acute lung injury: Neutrophils, macrophages and inflammatory mediators. *Br. J. Surg.* **79**(10), 998–1003. <https://doi.org/10.1002/bjs.1800791006> (1992).
- Li, W. *et al.* Circulating BMP9 protects the pulmonary endothelium during inflammation-induced lung injury in mice. *Am. J. Respir. Crit. Care Med.* **203**(11), 1419–1430. <https://doi.org/10.1164/rccm.202005-1761OC> (2021).
- Hu, Y. *et al.* Pioglitazone protects compression-mediated apoptosis in nucleus pulposus mesenchymal stem cells by suppressing oxidative stress. *Oxid. Med. Cell Longev.* **2019**, 4764071. <https://doi.org/10.1155/2019/4764071> (2019).
- Stangherlin, A. & Reddy, A. B. Regulation of circadian clocks by redox homeostasis. *J. Biol. Chem.* **288**(37), 26505–26511. <https://doi.org/10.1074/jbc.R113.457564> (2013).
- Thöny, B., Auerbach, G. & Blau, N. Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem. J.* **347**, 1–16 (2000).
- Tran, A. N. *et al.* Reactive species balance via GTP cyclohydrolase I regulates glioblastoma growth and tumor initiating cell maintenance. *Neuro Oncol.* **20**(8), 1055–1067. <https://doi.org/10.1093/neuonc/noy012> (2018).
- Jing, X. *et al.* Caffeine ameliorates hyperoxia-induced lung injury by protecting GCH1 function in neonatal rat pups. *Pediatr. Res.* **82**(3), 483–489. <https://doi.org/10.1038/pr.2017.89> (2017).
- Krusche, J. *et al.* TNF-alpha-induced protein 3 is a key player in childhood asthma development and environment-mediated protection. *J. Allergy Clin. Immunol.* **144**(6), 1684–1696. <https://doi.org/10.1016/j.jaci.2019.07.029> (2019).
- Studer, P. *et al.* Significant lethality following liver resection in A20 heterozygous knockout mice uncovers a key role for A20 in liver regeneration. *Cell Death Differ.* **22**(12), 2068–2077. <https://doi.org/10.1038/cdd.2015.52> (2015).
- Hiroshima, Y. *et al.* S100A8 induces IL-10 and protects against acute lung injury. *J. Immunol.* **192**(6), 2800–2811. <https://doi.org/10.4049/jimmunol.1302556> (2014).
- Hoegl, S. *et al.* NK cells regulate CXCR2+ neutrophil recruitment during acute lung injury. *J. Leukoc. Biol.* **101**(2), 471–480. <https://doi.org/10.1189/jlb.3A0516-227R> (2017).
- Portales-Cervantes, L., Haidl, I. D., Lee, P. W. & Marshall, J. S. Virus-infected human mast cells enhance natural killer cell functions. *J. Innate Immun.* **9**(1), 94–108. <https://doi.org/10.1159/000450576> (2017).
- Wei, J. L. *et al.* GCH1 induces immunosuppression through metabolic reprogramming and IDO1 upregulation in triple-negative breast cancer. *J. Immunother. Cancer* **9**(7), 2383. <https://doi.org/10.1136/jitc-2021-002383> (2021).
- Xiao, Y. *et al.* GCH1 reduces LPS-induced alveolar macrophage polarization and inflammation by inhibition of ferroptosis. *Inflamm. Res.* <https://doi.org/10.1007/s00011-023-01785-1> (2023).
- Devos, M. *et al.* Keratinocyte expression of A20/TNFAIP3 controls skin inflammation associated with atopic dermatitis and psoriasis. *J. Invest. Dermatol.* **139**(1), 135–145. <https://doi.org/10.1016/j.jid.2018.06.191> (2019).

30. Plant, D. *et al.* A genetic marker at the OLIG3/TNFAIP3 locus associates with methotrexate continuation in early inflammatory polyarthritis: Results from the Norfolk Arthritis Register. *Pharmacogenom. J.* **12**(2), 128–133. <https://doi.org/10.1038/tpj.2010.80> (2012).
31. van den Reek, J. *et al.* Polymorphisms in CD84, IL12B and TNFAIP3 are associated with response to biologics in patients with psoriasis. *Br. J. Dermatol.* **176**(5), 1288–1296. <https://doi.org/10.1111/bjd.15005> (2017).
32. Werner, S. L. *et al.* Encoding NF-kappaB temporal control in response to TNF: Distinct roles for the negative regulators Ikap-paBalpha and A20. *Genes Dev.* **22**(15), 2093–2101. <https://doi.org/10.1101/gad.1680708> (2008).
33. Leek, J. T. *et al.* *sva: Surrogate Variable Analysis. R Package Version 3.42.0* (2021).
34. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
35. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47. <https://doi.org/10.1093/nar/gkv007> (2015).
36. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
37. Chen, H. *VennDiagram: Generate High-Resolution Venn and Euler Plots. R Package Version 1.6.20.* <https://CRAN.R-project.org/package=VennDiagram> (2018).
38. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omic* **16**(5), 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
39. Walter, W., Sánchez-Cabo, F. & Ricote, M. GOpplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* **31**(17), 2912–2914. <https://doi.org/10.1093/bioinformatics/btv300> (2015).
40. Zhao, Y., Huang, T. & Huang, P. Integrated analysis of tumor mutation burden and immune infiltrates in hepatocellular carcinoma. *Diagnostics (Basel)* **12**(8), 1918. <https://doi.org/10.3390/diagnostics12081918> (2022).
41. Engebretsen, S. & Bohlin, J. Statistical predictions with glmnet. *Clin. Epigenet.* **11**(1), 123. <https://doi.org/10.1186/s13148-019-0730-1> (2019).
42. Jiang, H. *et al.* A 21-gene support vector machine classifier and a 10-gene risk score system constructed for patients with gastric cancer. *Mol. Med. Rep.* **21**(1), 347–359. <https://doi.org/10.3892/mmr.2019.10841> (2020).
43. Kassambara, A. *ggpubr: ggplot2 Based Publication Ready Plots. R Package Version 0.4.0.* <https://CRAN.R-project.org/package=ggpubr> (2020).
44. Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77. <https://doi.org/10.1186/1471-2105-12-77> (2011).
45. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
46. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7. <https://doi.org/10.1186/1471-2105-14-7> (2013).
47. Wei, T. & Simko, V. *R Package 'corrplot': Visualization of a Correlation Matrix (Version 0.92).* <https://github.com/taiyun/corrplot> (2021).
48. Cui, Y. *et al.* Microglia and macrophage exhibit attenuated inflammatory response and ferroptosis resistance after RSL3 stimulation via increasing Nrf2 expression. *J. Neuroinflamm.* **18**(1), 249. <https://doi.org/10.1186/s12974-021-02231-x> (2021).
49. Percie du Sert, N. *et al.* The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLoS Biol.* **18**(7), e3000410. <https://doi.org/10.1371/journal.pbio.3000410> (2020).
50. Liu, Y., Yang, H., Zhu, F., Ouyang, Y. & Pan, P. Inhibition of STAT3 phosphorylation by colchicine regulates NLRP3 activation to alleviate sepsis-induced acute lung injury. *Inflammopharmacology.* <https://doi.org/10.1007/s10787-023-01199-9> (2023).

Acknowledgements

The authors thank Hang Yang (Xiangya Medical College) and Fei Zhu (Xiangya Medical College) for generously sharing their experience and warm-hearted help.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Y.L., H.L., Y.O., Y.Z. and P.P. The first draft of the manuscript was written by Y.L. and H.L., and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 81770080, 82100037), the Xiangya Hospital-Beida Weiming Clinical Rehabilitation Research Fund (No. XYWM2015I20), Key R&D Program of Hunan Province (No. 2022SK2038), Project Program of central south university graduate education teaching reform (No. 2022JGB025), Research Project on Education and Teaching Reform of Central South University (No. 2021 jy139-2), Peking Union Medical Foundation—Ruiyi Emergency Medical Research Fund (No. R2021004) and Hainan Provincial Medical and Health Research Project (No. 20A200340).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-49165-3>.

Correspondence and requests for materials should be addressed to Y.L., H.L., Y.Z. or P.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023