



OPEN ACCESS

EDITED BY

Shuming Pan,
Shanghai Jiao Tong University, China

REVIEWED BY

Yonghao Xu,
First Affiliated Hospital of Guangzhou
Medical University, China
Xiao-min Lu,
Jiangsu Provincial Hospital of Traditional
Chinese Medicine, China
Yeming Wang,
China-Japan Friendship Hospital, China

*CORRESPONDENCE

Fengmei Guo
✉ fmguo2003@163.com

RECEIVED 21 April 2023

ACCEPTED 09 October 2023

PUBLISHED 23 October 2023

CITATION

Xia F, Chen H, Liu Y, Huang L, Meng S,
Xu J, Xie J, Wang G and Guo F (2023)
Development of genomic phenotype and
immunophenotype of acute respiratory
distress syndrome using autophagy and
metabolism-related genes.
Front. Immunol. 14:1209959.
doi: 10.3389/fimmu.2023.1209959

COPYRIGHT

© 2023 Xia, Chen, Liu, Huang, Meng, Xu, Xie,
Wang and Guo. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Development of genomic phenotype and immunophenotype of acute respiratory distress syndrome using autophagy and metabolism-related genes

Feiping Xia¹, Hui Chen^{1,2}, Yigao Liu¹, Lili Huang¹,
Shanshan Meng¹, Jingyuan Xu¹, Jianfeng Xie¹,
Guozheng Wang³ and Fengmei Guo^{1*}

¹Jiangsu Provincial Key Laboratory of Critical Care Medicine, Department of Critical Care Medicine, Zhongda Hospital, School of Medicine, Southeast University, Nanjing, China, ²Department of Critical Care Medicine, The First Affiliated Hospital of Soochow University, Soochow University, Suzhou, China, ³Department of Clinical Infection Microbiology and Immunology, University of Liverpool, Liverpool, United Kingdom

Background: Distinguishing ARDS phenotypes is of great importance for its precise treatment. In the study, we attempted to ascertain its phenotypes based on metabolic and autophagy-related genes and infiltrated immune cells.

Methods: Transcription datasets of ARDS patients were obtained from Gene expression omnibus (GEO), autophagy and metabolic-related genes were from the Human Autophagy Database and the GeneCards Database, respectively. Autophagy and metabolism-related differentially expressed genes (AMRDEGs) were further identified by machine learning and processed for constructing the nomogram and the risk prediction model. Functional enrichment analyses of differentially expressed genes were performed between high- and low-risk groups. According to the protein-protein interaction network, these hub genes closely linked to increased risk of ARDS were identified with CytoHubba. ssGSEA and CIBERSORT was applied to analyze the infiltration pattern of immune cells in ARDS. Afterwards, immunologically characterized and molecular phenotypes were constructed according to infiltrated immune cells and hub genes.

Results: A total of 26 AMRDEGs were obtained, and CTSB and EEF2 were identified as crucial AMRDEGs. The predictive capability of the risk score, calculated based on the expression levels of CTSB and EEF2, was robust for ARDS in both the discovery cohort (AUC = 1) and the validation cohort (AUC = 0.826). The mean risk score was determined to be 2.231332, and based on this score, patients were classified into high-risk and low-risk groups. 371 differential genes in high- and low-risk groups were analyzed. ITGAM, TYROBP, ITGB2, SPI1, PLEK, FGR, MPO, S100A12, HCK, and MYC were identified as hub genes. A total of 12 infiltrated immune cells were differentially expressed and have correlations with hub genes. According to hub genes and implanted immune cells, ARDS patients were divided into two different molecular phenotypes (Group 1: n = 38;

Group 2: $n = 19$) and two immune phenotypes (Cluster1: $n = 22$; Cluster2: $n = 35$), respectively.

Conclusion: This study picked up hub genes of ARDS related to autophagy and metabolism and clustered ARDS patients into different molecular phenotypes and immunophenotypes, providing insights into the precision medicine of treating patients with ARDS.

KEYWORDS

autophagy, metabolism, genomic phenotype, immunophenotype, ARDS

Introduction

Acute respiratory distress syndrome (ARDS) is a profound pulmonary inflammatory reaction resulting from multiple potential causes associated with extremely high morbidity and mortality (1–3). ARDS is a global health problem, accounting for 10.4% of the admissions to ICU, and its mortality rate was about 30% to 60% (1). There is no specific treatment proved to be effective in decreasing the mortality or morbidity of ARDS (4) because ARDS is a heterogeneous syndrome rather than an explicit pathological disease.

Heterogeneity in ARDS is increasingly recognized as a principal barrier to developing efficacious targeted therapies. Phenotyping schemas are considered effective approaches in identifying a homogenous population within heterogeneous patients based on a multitude of data. In recent years they have been used in patients with ARDS to thoroughly understand the patients' conditions and identify potential treatments (5). Further, a recent study has demonstrated that the ability to detect heterogeneity in ARDS diminished when phenotyping schema relied on only clinical data (6). Nevertheless, plenty of excellent types of research (7–10) have leveraged multivariate clinical data to phenotype patients of ARDS. On the other hand, triggering of ARDS depends on immune cells, such as neutrophils, monocytes, and macrophages, which propagate uncontrolled inflammation and tissue injury by secreting cytokines (11). Subsequently, adaptive immune responses are involved in the immunopathology of ARDS (12). Hence, recognizing different biological phenotypes (molecular phenotype and immunophenotype) has important implications for the precision medicine of patients with ARDS.

Autophagy is an evolutionarily conserved, lysosomal degradation pathway that is associated with cellular recycling, homeostasis, and elimination of intracellular pathogens (13). Autophagy controls inflammation through regulatory interactions with innate and adaptive immune signaling pathways, eliminating endogenous inflammasome agonists and immune mediators (14, 15). Recently, researchers have indicated that autophagy is essential in regulating the outcome of ARDS (16, 17). The fundamental essence of autophagy lies in its capacity to adapt to metabolic demands (18). Meanwhile, the metabolism data generated from

clinical and experimental studies (19, 20) demonstrate a disturbance in energy and oxidative stress metabolism, which is consistent with the pathology of ARDS. The pathological alterations observed in ARDS may arise from a combination of genetic predisposition and immune response, leading to subsequent modifications in the downstream metabolites implicated in the development of ARDS. Therefore, host-derived metabolites are essential to connect to the pathogenesis of ARDS. Above all, we speculate that the phenotype classification of ARDS based on the characteristics of autophagy and metabolism is of great significance for its precision treatment.

In this study, we aimed to untangle the heterogeneity of ARDS based on autophagy and metabolism by using molecular and immune data in an algorithm of computation to determine homogenous phenotypes, with the aim to provide evidence in support of the concept of precision medicine of treating ARDS.

Materials and methods

Downloading and processing of data

Data containing three transcription profiles (GSE89953 (21), GSE32707 (22), and GSE76293 (23)) of acute respiratory distress syndrome (ARDS) and their corresponding clinical data were downloaded from the NCBI GEO database (24) (<https://www.ncbi.nlm.nih.gov/geo/>) from version 2.62.2 of 'GEOquery' package (25) in R software. The GSE89953 dataset, GSE32707 dataset, and GSE76293 dataset were based on GPL6883, GPL10558, and GPL570 platforms, respectively. The GSE89953 dataset comprised 26 acute respiratory distress syndrome (ARDS) patients. A total of 144 patients were included in dataset GSE32707, including 31 ARDS patients, 34 healthy controls, and 79 other patients, in which ARDS patients and normal patients were selected for this analysis. A total of 12 ARDS patients and 12 healthy controls of the dataset GSE76293 were included for analysis in the present study. The datasets GSE89953 and GSE32707 were set as the discovery cohort and GSE76293 as the validation cohort (Table 1). The datasets GSE89953 and GSE32707 were subjected to batch effect correction using the 'sva' (26) package in R. The Combined Datasets encompassed a total of 57 ARDS patients and 34 healthy controls.

TABLE 1 Acute respiratory distress syndrome data set information list.

	GSE89953	GSE32707	GSE76293
Platform	GPL6883	GPL10558	GPL570
Species	Homo sapiens	Homo sapiens	Homo sapiens
Tissue	alveolar macrophage	whole blood	Neutrophils
Samples in ARDS group	26	31	12
Samples in Normal group	0	34	12
Gender			
Male	15		
Female	11		
Age (years)			
Mean	41.8		
Median	46.5		
Reference	PMID: 28708019	PMID: 22461369	PMID:27064380

GEO, Gene Expression Omnibus; ARDS, Acute Respiratory Distress Syndrome.

HADb (Human Autophagy Database, <http://www.autophagy.lu/index.html>) is a public human autophagy-specific database. Autophagy-related genes (ARGs) were downloaded by HADb, and 222 ARGs were obtained (Supplementary Table S1). Metabolic-Related Genes (MRGs) were downloaded from the GeneCards (<https://www.genecards.org/>) database (27), with “metabolic” as the search keyword, after retaining MRGs with “Protein Coding” and correlation scores greater than 1, 7036 MRGs were obtained (Supplementary Table S2). A total of 159 autophagy and metabolism-related genes (AMRGs) were obtained after merging and deduplication (Supplementary Table S3).

Identification of autophagy and metabolism-related differentially expressed genes

According to the grouping of the integrated GEO dataset (Combined Datasets), the patients were divided into the ARDS group and the normal group. The differential analysis of genes in the ARDS and normal groups was performed using the ‘limma’ package in R (28). Set $|\logFC| > 0.5$ and $\text{adj. } p < 0.01$ as threshold for differential genes. Genes with $\logFC > 0.5$ and $\text{adj. } p < 0.01$ are up-regulated differentially expressed genes, and $\logFC < -0.5$ and $\text{adj. } p < 0.01$ are down-regulated differentially expressed genes. In order to obtain the AMRDEGs associated with ARDS, all $|\logFC| > 0.5$ and $\text{adj. } p < 0.01$ differentially expressed genes (DEGs) in Combined Datasets and AMRGs take the intersection. The results of the differential analyses were displayed as a volcano map by the ‘ggplot2’ package, a heat map by the ‘pheatmap’ package, and a chromosome map by the ‘RCircos’ package (29) in R.

In order to explore the correlation between AMRDEGs, the Spearman algorithm was used. The TOP 3 positively and negatively correlated AMRDEGs are displayed by correlation scatter plots with the ‘ggplot2’ package in R. The ‘PROC’ package in R was used to

plot the receiver operating characteristic (ROC) curve (30) of AMDEGs in the Combined Datasets to distinguish whether the patient is ARDS.

Screening crucial AMRDEGs

To further screen the crucial AMRDEGs and candidate signatures, three machine learning [least absolute shrinkage and selection operator (LASSO) (31) logistic regression, support vector machine (SVM) (32), and random forest (RF) (33) algorithms were adopted. Based on AMRDEGs in the Combined Datasets, the LASSO logistic regression model with parameters $\text{seed} = 3$, family = “binomial” was run by the ‘glmnet’ package (34) in R and run 1000 cycles to prevent overfitting combined. SVM, a supervised machine learning technique, was executed with parameters $\text{seed} = 3$, method = “svmLinear” by the ‘caret’ package. AMRDEGs were screened with parameters $\text{seed} = 3$, $\text{ntree} = 150$ of a random forest model by ‘randomForest’ package. Finally, the genes at the intersection of those screened by LASSO, SVM, and random forest were used to diagnose ARDS.

Constructing the nomogram and the risk prediction model

In order to analyze the diagnostic efficacy of crucial AMRDEGs for ARDS, we performed logistic regression analysis on crucial AMRDEGs in the Combined Datasets and constructed a logistic regression model. Based on the logistic regression analysis results, the ‘rms’ package was used to construct a nomogram (35). The calibration curve was used to assess the nomogram’s performance, and decision curve analysis (DCA) (36) was performed to evaluate the accuracy and resolution of the logistic regression model by the ‘rmda’ package.

To evaluate the predictive power of crucial AMRDEGs for ARDS, LASSO regression analysis was used to establish the prediction model. The Combined Datasets were used as the discovery cohort and the GSE76293 as the validation cohort. Risk scores were calculated from the expression of crucial AMRDEGs and LASSO regression coefficients.

$$\text{riskScore} = \sum_i \text{Coefficient}(\text{gene}_i) * \text{mRNA Expression}(\text{gene}_i)$$

According to the calculated mean risk score, patients were divided into high-risk and low-risk groups. The ROC curve was drawn to analyze the accuracy of the risk score in predicting whether the patient was ARDS by the 'pROC' package. The 'ggplot2' package was used to draw a risk scatter diagram to show the distribution of high and low-risk groups and the occurrence of ARDS. The expression of autophagy and metabolic phenotype-related crucial AMRDEGs in high and low-risk groups was shown in a hot map by the 'pheatmap' package. Taking the expression profile data of GSE76293 as the verification set, the risk scores were calculated by the above formulas. Grouping was based on the mean risk score, then the ROC curve, scatter plot and heat map were drawn.

Screening differentially expressed genes and functional enrichment analysis

To identify genes associated with autophagy and metabolic phenotypes, according to the grouping of the high- and low-risk groups in the Combined Datasets, the 'limma' package of R was used to perform differential analysis. Set $\text{adjPvalue} < 0.01$ and $|\log\text{FC}| > 1$ as the threshold to screen the genes, among which the genes with $\log\text{FC} > 1$ and $\text{adjPvalue} < 0.01$ are the differentially expressed genes (DEGs) with up-regulated expression, and the genes with $\log\text{FC} < -1$ and $\text{adjPvalue} < 0.01$ are the genes with down-regulated expression. The results of the different analyses are displayed by drawing volcano maps and heat maps with the 'ggplot2' and the 'pheatmap' package.

Gene Ontology (GO) analysis (37) is a standard method for large-scale functional enrichment analysis, including biological process (BP), molecular function (MF), and cellular component (CC). The Kyoto Encyclopedia of Genes and Genomes (KEGG) (38) is a widely used database that stores information about genomes, biological pathways, diseases, and drugs. The 'clusterProfiler' package (39) was used to perform GO annotation analysis and KEGG pathway enrichment analysis on the differentially expressed genes of the high- and low-risk groups in the Combined Datasets. The false discovery rate (FDR) < 0.05 was considered statistically significant, and the screening standard was $\text{adjPvalue} < 0.05$ and $q \text{ value} < 0.05$. The P value correction method is the Benjamini-Hochberg (BH).

In order to study the differences in biological processes between high and low-risk groups, we downloaded the "h.all.v7.5.1.symbols" gene set from the MSigDB database (40) and performed Gene set enrichment analysis (GSEA) (41). Adjusted P values less than 0.05 were considered statistically significant.

To further investigate the differences in biological processes between the high- and low-risk groups, the "h.all.v7.5.1.symbols" Gene Set was downloaded from the MSigDB database (40) to perform Gene Set Variation Analysis (GSVA) on the Combined Datasets (42). Adjusted P values of less than 0.05 were considered statistically significant.

Identification of hub genes and construction of the related regulatory network

The protein interaction network is significant for understanding the functional connections between proteins (43). The STRING database (43) (<https://string-db.org/>) searches for proteins and predicts protein interactions. The database contains 67.59 million proteins and 20052.39 million protein-protein interactions. We used the STRING database to construct a protein-protein interaction (PPI) network for the differentially expressed genes of the high-risk group vs. the low-risk group in the Combined Datasets, and the coefficient was set to 0.4. The PPI results were exported from the STRING database and visualized by Cytoscape (44), and the hub genes in the PPI network were analyzed by CytoHubba (45).

miRTarBase (https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2022/php/index.php) database (46) is a collection of microRNA-target interactions (MTI) database supported by experimental evidence. To study the regulatory relationship between hub genes and miRNAs, information about miRNA-mRNA interactions was downloaded from the miRTarBase database. Based on the hub genes obtained from the PPI analysis, the miRTarBase database was used to predict possible regulated miRNAs and construct a miRNA-mRNA regulatory network. Cytoscape software was used to visualize the miRNA-mRNA regulatory network.

Transcription factor (TF) controls gene expression by interacting with target genes at the post-transcriptional stage. TRRUST (Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining) (<https://www.grnpedia.org/trrust/>) is a human-annotated transcriptional regulation network database (47). Transcription factors that bind to hub genes were searched in the TRRUST database. The hub-TF network was subsequently visualized by Cytoscape.

Immune infiltration analysis and construction of immunologically characterized phenotypes

Immune-related genes were downloaded from literature PMID:28052254 (48). The gene set contains 782 genes and 28 cell types. The degree of infiltration of immune cells was analyzed using the single-sample GSEA (ssGSEA) algorithm on the Combined Datasets by the 'GSVA' package (42). CIBERSORT (49) deconvolutes the transcriptome expression matrix based on the

principle of linear support vector regression to estimate the composition and abundance of immune cells in mixed cells. The Combined Datasets were computed to derive the immune cell infiltration matrix by CIBERSORT. Boxplots were drawn to show the differences in the infiltration levels of 22 immune cells between the ARDS group and the Normal group in the Combined Datasets by the 'ggplot' package. For the immune cells with significant differences ($P < 0.05$), the lollipop plot was used to show the correlation between immune cells and hub genes.

Consensus Clustering (50) could divide patients into several phenotypes according to different omics data sets to discover new disease phenotypes or conduct a comparative analysis. Cluster analysis was used by the 'ConsensusClusterPlus' package (51). According to the results obtained by ssGSEA, ARDS patients were divided into different groups according to the infiltration level of immune cells to construct immunologically characterized phenotypes. In this process, clusters were set between 2 and 8, and 80% of the total samples were extracted with 1000 repetitions, with clusterAlg = "km" and distance = "euclidean".

Construction of molecular phenotypes and correlation analysis with immune infiltrating cells

Molecular phenotypes of ARDS were identified using consensus clustering analysis based on ten hub genes using the 'ConsensusClusterPlus' package in R. According to the analysis results of ssGSEA. The 'ggplot' package was used to analyze the difference in the degree of immune cell infiltration among different

phenotypes and the correlation between hub genes and immune cells in different molecular phenotypes.

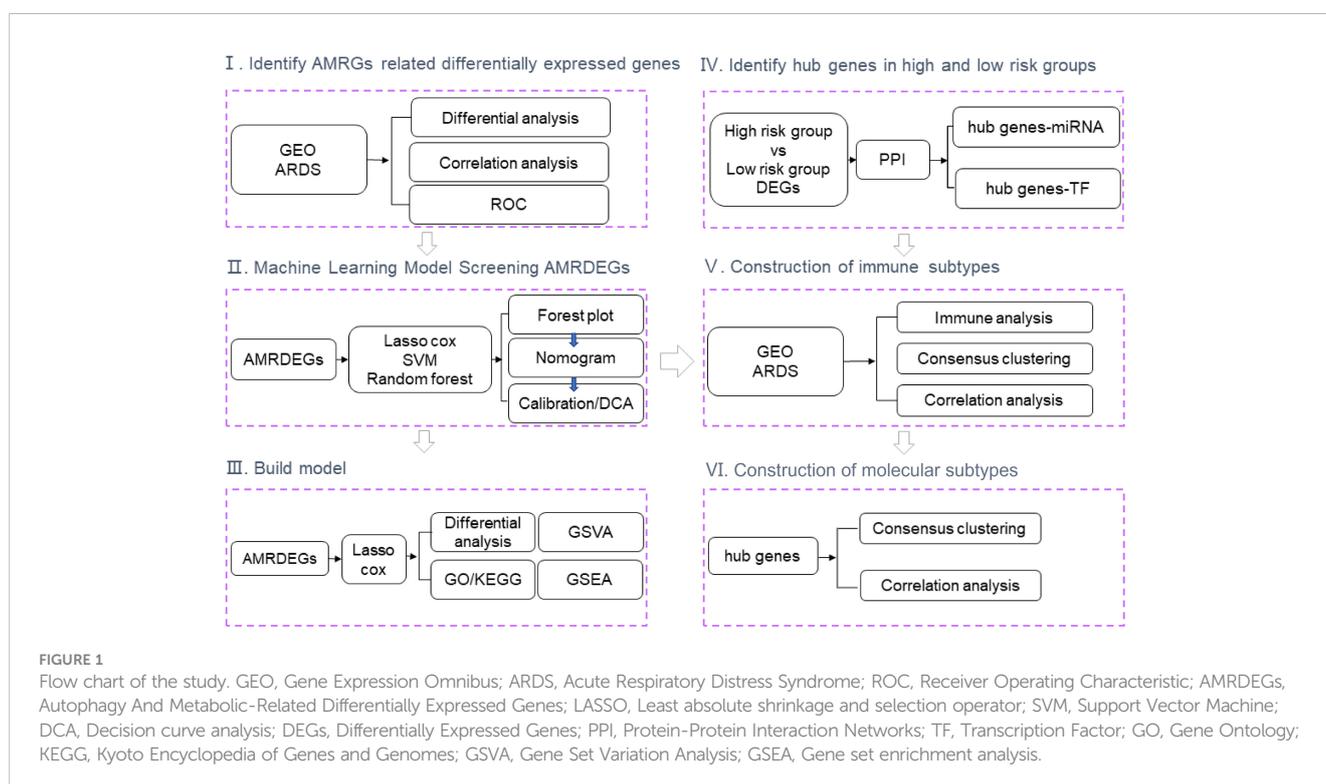
Statistical analysis

All data processing and analysis involved in this study were performed using R software (Version 4.1.1, <https://www.r-project.org/>). For comparisons of continuous variables between the two groups, the statistical significance of normally distributed variables was estimated utilizing an independent Student's t-test, and differences between non-normally distributed variables were analyzed through the Mann-Whitney U test (i.e., the Wilcoxon rank-sum test). LASSO analysis was based on the glmnet package of R (34). All statistical P values are two-sided, and a P value of less than 0.05 was considered statistically significant.

Results

Identification of AMRDEGs between ARDS and control

A flow chart of the study is shown in Figure 1. GSE89953 and GSE32707, two sets of ARDS datasets, were combined to obtain the integrated GEO dataset after removing the batch effect. The batch effect of samples within the integrated GEO dataset was effectively alleviated following the batch removal procedure (Figures 2A, C). The efficacy of batch effect removal was further confirmed through Principal Component Analysis (PCA), which demonstrated the



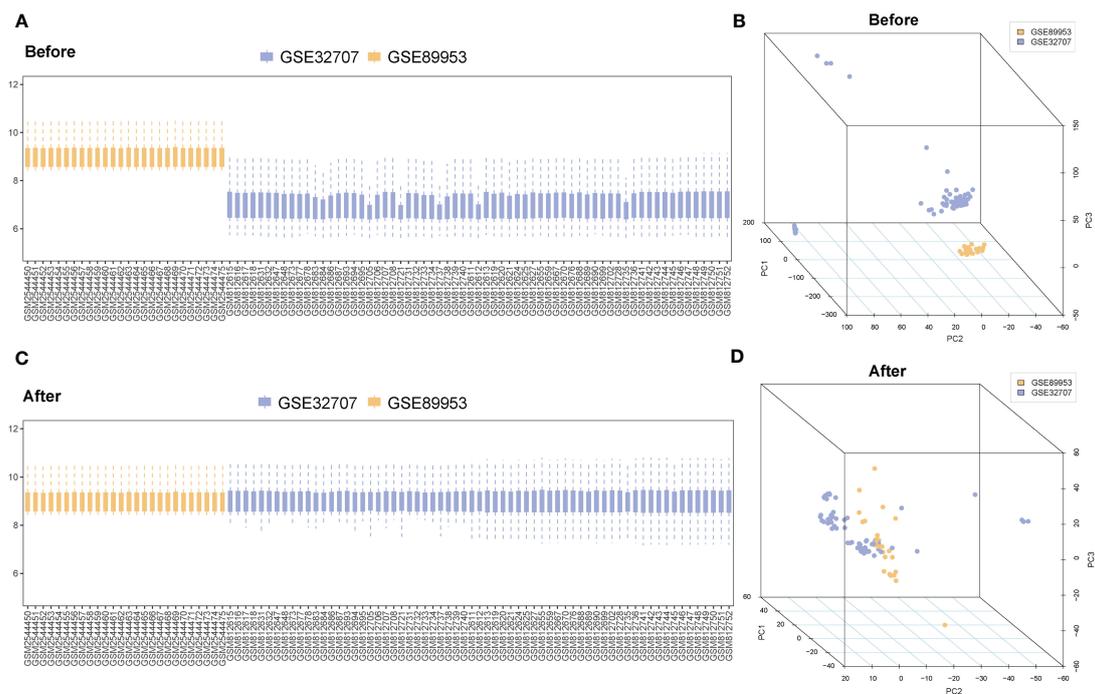


FIGURE 2

Merging and correction of databases. (A) The boxplot of the ARDS datasets before removing the batch effect. (B) Principal component analysis of the ARDS datasets before removing the batch effect. (C) The boxplot of the ARDS datasets after removing the batch effect. (D) Principal component analysis of the ARDS datasets after removing the batch effect. The blue represents the sample in the GSE32707; the orange represents the sample in the GSE89953. ARDS, Acute Respiratory Distress Syndrome.

substantial elimination of batch effects among samples from diverse sources within the integrated GEO dataset (Figures 2B, D).

26 AMRDEGs were obtained by the intersection of AMRGs and differentially expressed genes (Figures 3A–C). The location of AMRDEGs on human chromosomes was shown on the chromosomal map (Figure 3D). All AMRDEGs were significantly different between the two groups ($P < 0.001$) (Figure 3E). Further analysis for AMRDEGs in ARDS showed multiple correlations between the genes (Figure 4A). The degree of correlation was indicated by the r value. The synergistic effect was observed as the strongest between HSP90AB1 and HSPA8 ($R = 0.901$, $P < 2.2 \times 10^{-16}$, Figure 4B), followed by STK11 and PINK1 ($R = 0.897$, $P < 2.2 \times 10^{-16}$, Figure 4C), SIRT1 and MYC ($R = 0.854$, $P < 2.2 \times 10^{-16}$, Figure 4D). In contrast, the competitive effect was found as the strongest between CTSB and BIRC5 ($R = -0.802$, $P < 2.2 \times 10^{-16}$, Figure 4E), followed by CFLAR and BIRC5 ($R = -0.786$, $P < 2.2 \times 10^{-16}$, Figure 4F), STK11 and BIRC5 ($R = -0.762$, $P < 2.2 \times 10^{-16}$, Figure 4G). The ROC curve analysis revealed that 17 of the 26 AMRDEGs (Figures 4H–N) had high predictive accuracy, with a ROC curve above 0.9.

Selection of crucial AMRDEGs via LASSO, SVM, and RF

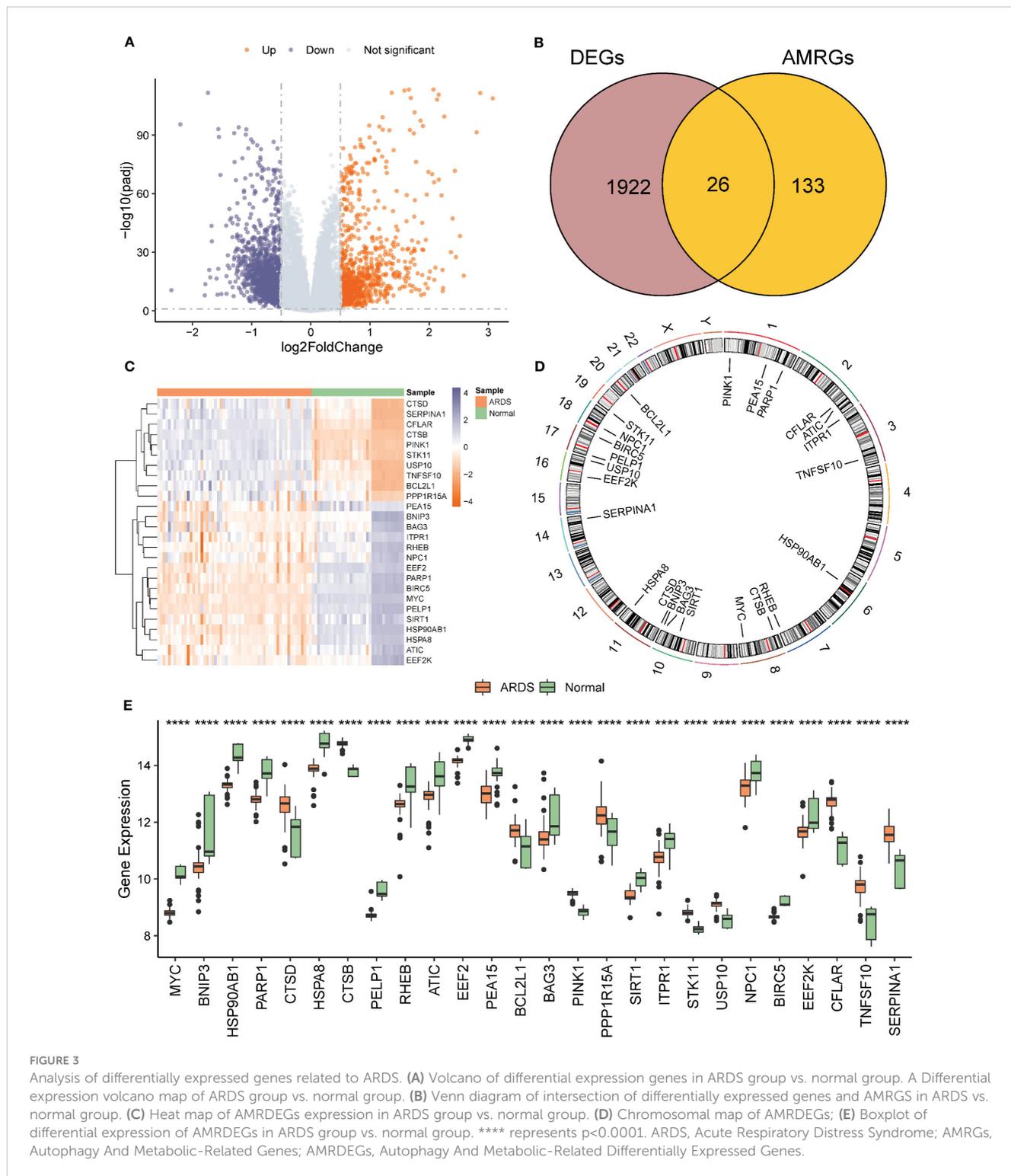
To further optimize the screening of the crucial AMRDEGs, we applied three machine learning algorithms (LASSO regression, SVM-RFE, and RF). A total of 4 genes were screened by LASSO regression (Figures 5A, B). The accuracy of the SVM method reached its peak

when the number of variables was 7 (Figure 5C). RF was used to select the top 10 genes ranked by variable importance (Figures 5D, E). By overlapping the three algorithms, two diagnostic signatures (CTSB and *EEF2*) were identified as crucial AMRDEGs (Figure 5F).

Diagnostic efficacy and predictive power of crucial AMRDEGs in predicting ARDS

Expressions of CTSB and *EEF2* were similar in the Combined Databases (Figure 6A). In addition, a predictive tool for ARDS development, a nomogram, was constructed by including these two signature genes related to autophagy and metabolism. In the nomogram, the value of each significant variable is associated with a score point, and the scores of all characteristic variables are summed to obtain the total score, which represents the risk of ARDS onset (Figure 6B). Calibration curves confirmed the accuracy of the nomogram in the diagnosis of ARDS (Figure 6C). DCA revealed that the application of the nomogram brought benefits to patients with ARDS (Figure 6D). Taken together, it can be concluded that the signature genes related to autophagy and metabolism show better diagnostic efficacy in predicting ARDS progression.

In order to evaluate the predictive power of critical AMRDEGs in predicting ARDS, we integrated the expression of two diagnostic genes to construct a risk-scoring model. The analysis revealed that the risk score exhibited robust predictive capability for ARDS in both the discovery cohort ($AUC = 1$) (Figure 7A) and the validation cohort ($AUC = 0.826$) (Figure 7B). Then, the mean risk score



(2.231332) was used to divide the patients in the discovery cohort into high- and low-risk groups. It was found that patients in the high-risk group were more inclined to be ARDS, CTSB was highly expressed in the high-risk group, and EEF2 was highly expressed in the low-risk group (Figure 7C). Finally, the patients in the validation cohort were divided into high- and low-risk groups using the mean value of the risk score (-25.53986), and the results showed the same finding (Figure 7D).

Identification of differentially expressed genes of high and low-risk groups

The patients were divided into high- and low-risk groups to analyze the influence of the risk model on the occurrence and development of ARDS. A total of 371 differentially expressed genes were obtained, of which 214 genes were significantly up-regulated, and 154 genes were significantly down-regulated (Figures 7E, F).

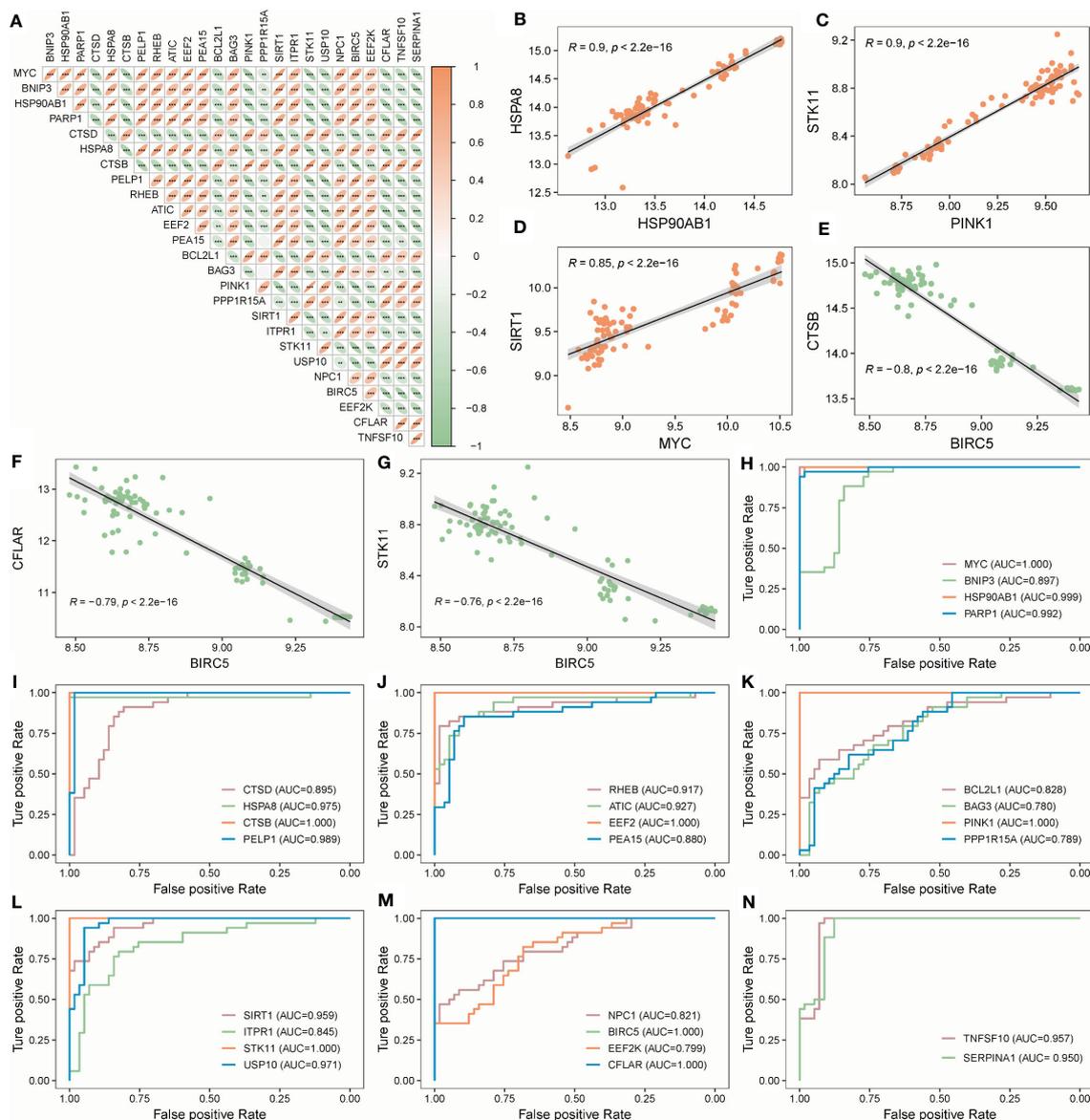


FIGURE 4 Correlation and ROC analysis of AMRDEGs in ARDS. **(A)** Correlation heatmap of 26 AMRDEGs in the Combined Datasets. **(B–D)** Scatter plot of correlation between TOP3 positively correlated genes. **(E–G)** Scatter plot of correlation between TOP3 negatively correlated genes. **(H–N)** ROC curves of 26 AMRDEGs in the Combined Datasets. ** represents $p < 0.01$; *** represents $p < 0.001$. AMRDEGs, Autophagy and Metabolic-Related Differentially Expressed Genes; ROC, Receiver Operating Characteristic.

Based on differentially expressed genes, GO, and KEGG analyses were performed. The results showed that the differentially expressed genes were mainly correlated with neutrophil activation (BP), secretory granule lumen (MF), peptide transmembrane transporter activity (CC), and NF-kappa signal pathway (KEGG) (Figures 8A–D; Supplementary Tables S4–5).

The gene expression in high- and low-risk patients were further processed for functional enrichment with GSEA and GSVA. GSEA showed that pathways such as allograft rejection (Figure 8E) and inflammatory response (Figure 8F) were significantly enriched in high-risk patients, and E2F targets (Figure 8G) were significantly enriched in low-risk patients (Supplementary Table S6). GSVA

showed that the low-risk group was significantly enriched in E2F targets and MYC targets V2, while the high-risk group was significantly enriched in complement, IL6 JAK STAT3 Signaling and allograft rejection (Figures 8H, I).

Selecting hub genes and construction of the related regulatory network

PPI of 371 differential genes in high- and low-risk groups was analyzed using the STRING database and visualized as a network with the Cytoscape (Figure 9A). With CytoHubba, ten candidate

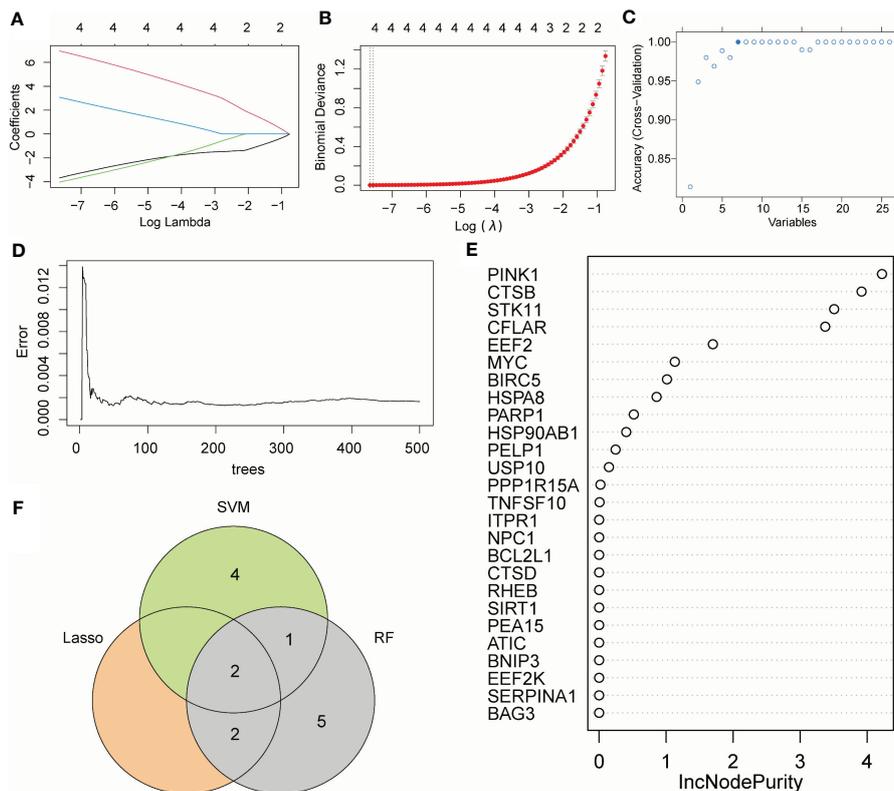


FIGURE 5 Machine learning algorithm for crucial AMRDEGs. **(A)** LASSO plot showed the variations in the size of coefficients for parameters shrank as the value of the k penalty increased. **(B)** Penalty plot of the LASSO model with error bars denoting standard errors. **(C)** Variable screening plot of SVM model. **(D)** The error rate confidence intervals for the random forest model. **(E)** The importance of genes in random forest model. **(F)** The interaction of the LASSO, SVM, and RF algorithms. AMRDEGs, Autophagy And Metabolic-Related Differentially Expressed Genes; LASSO, Least absolute shrinkage and selection operator; SVM, Support Vector Machine; RF, random forest; IncNodePurity, Increase in Node Purity.

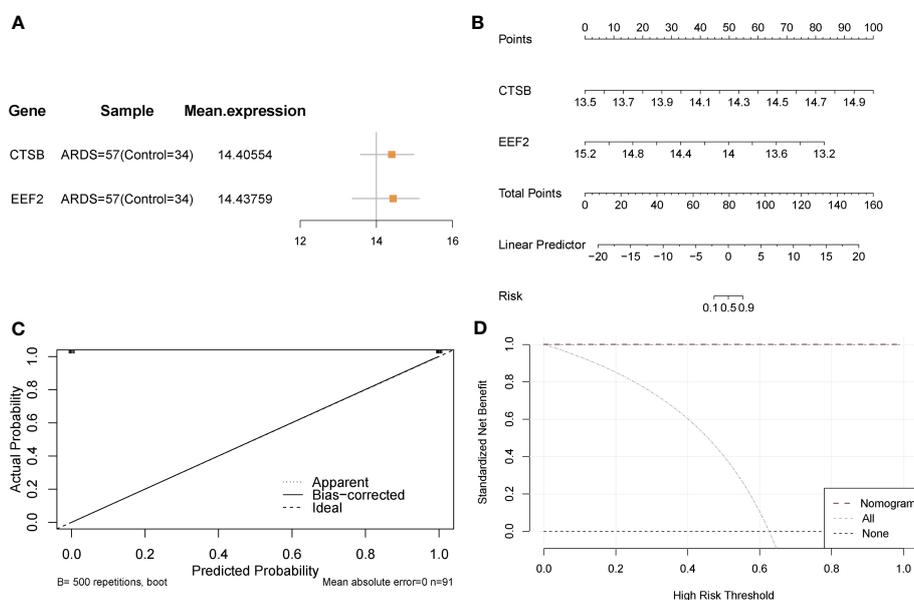


FIGURE 6 Validation of the diagnostic efficacy of crucial AMRDEGs. **(A)** forest plot of crucial AMRDEGs. **(B)** Nomogram showing the predicted risk for ARDS based on crucial AMRDEGs. **(C)** Calibration curve showing predicted performance of the nomogram. **(D)** DCA showing the clinical benefits of the nomogram. DCA, Decision Curve Analysis.

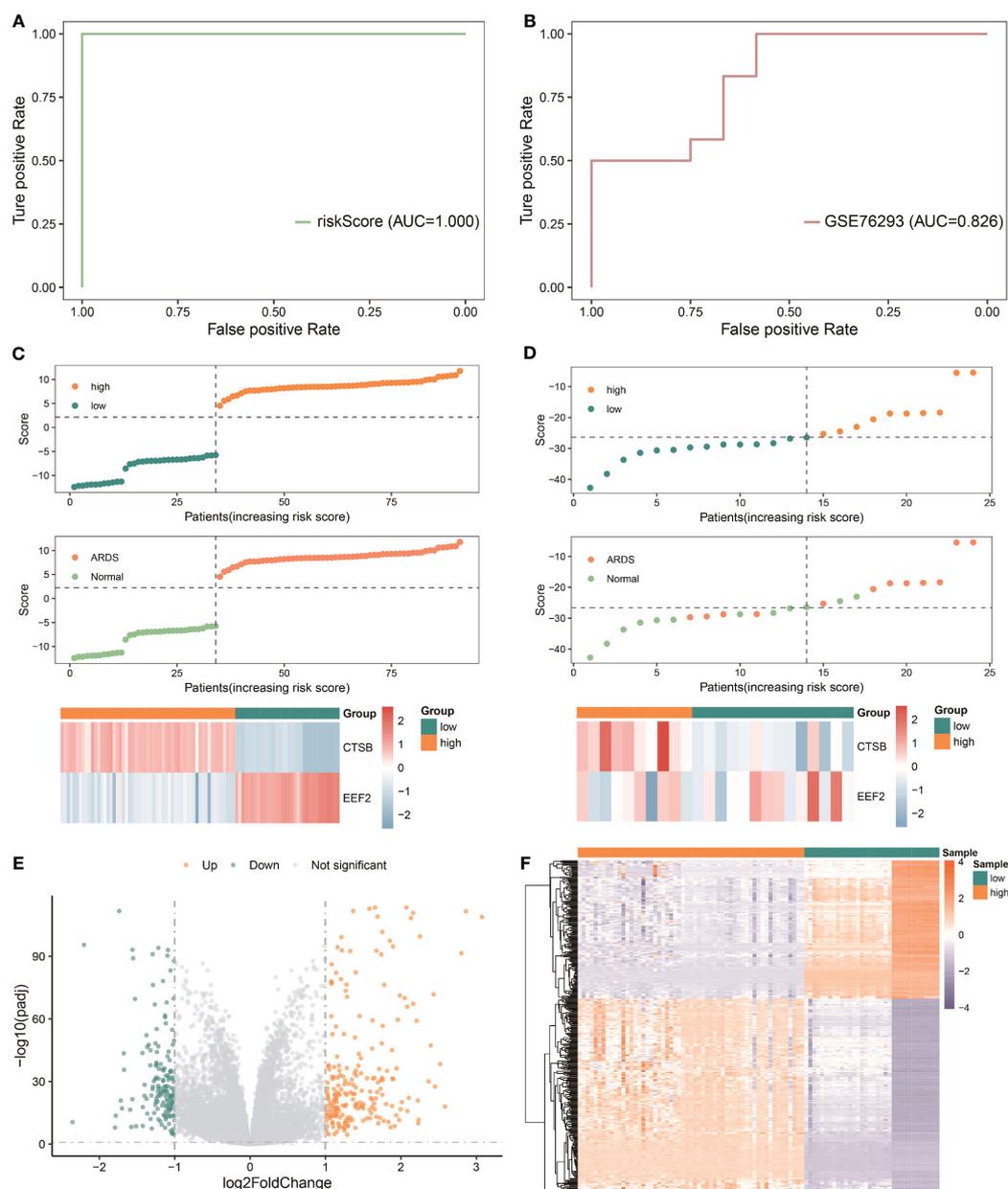


FIGURE 7

Predictive power of the risk model based on critical AMRDEGs and screening differential expression genes in the high-risk group of ARDS. **(A)** ROC curve of the risk score in the Combined Datasets. **(B)** ROC curve of the risk score in the validation set. **(C)** Distribution of risk scores, distribution of ARDS patients, and heat map of crucial AMRDEG expression in the integrated GEO dataset. **(D)** Distribution of risk scores, distribution of ARDS patients, and heat map of crucial AMRDEG expression in the validation set. **(E)** Volcano showed expression of differential expression genes between the high-risk and low-risk groups. **(F)** Heat map of differential expression genes expression in the high-risk and low-risk groups. ROC, Receiver Operating Characteristic; ARDS, Acute Respiratory Distress Syndrome.

hub genes were identified from the PPI network, including ITGAM, TYROBP, ITGB2, SPI1, PLEK, FGR, MPO, S100A12, HCK, and MYC (Figure 9B). Friends analysis found that the HCK gene played an essential role in the hub genes (Figure 9C). Further analysis showed that MYC, SPI1, and PLEK genes played significant regulatory roles in the miRNA-mRNA regulatory network (Figure 9D). For the TF-mRNA regulatory network, MYC, SPI1, and ITGB2 play essential roles in the regulation (Figure 9E).

Analysis of hub genes and infiltrated immune cells

For ARDS patients in Combined Datasets, levels of infiltration of immune cells were shown in Supplementary Table S7. Moreover, 12 of 22 infiltrated immune cells were differentially expressed between the ARDS group and the normal group in the Combined Datasets (Supplementary Table S8; Figure 10A). Most of the

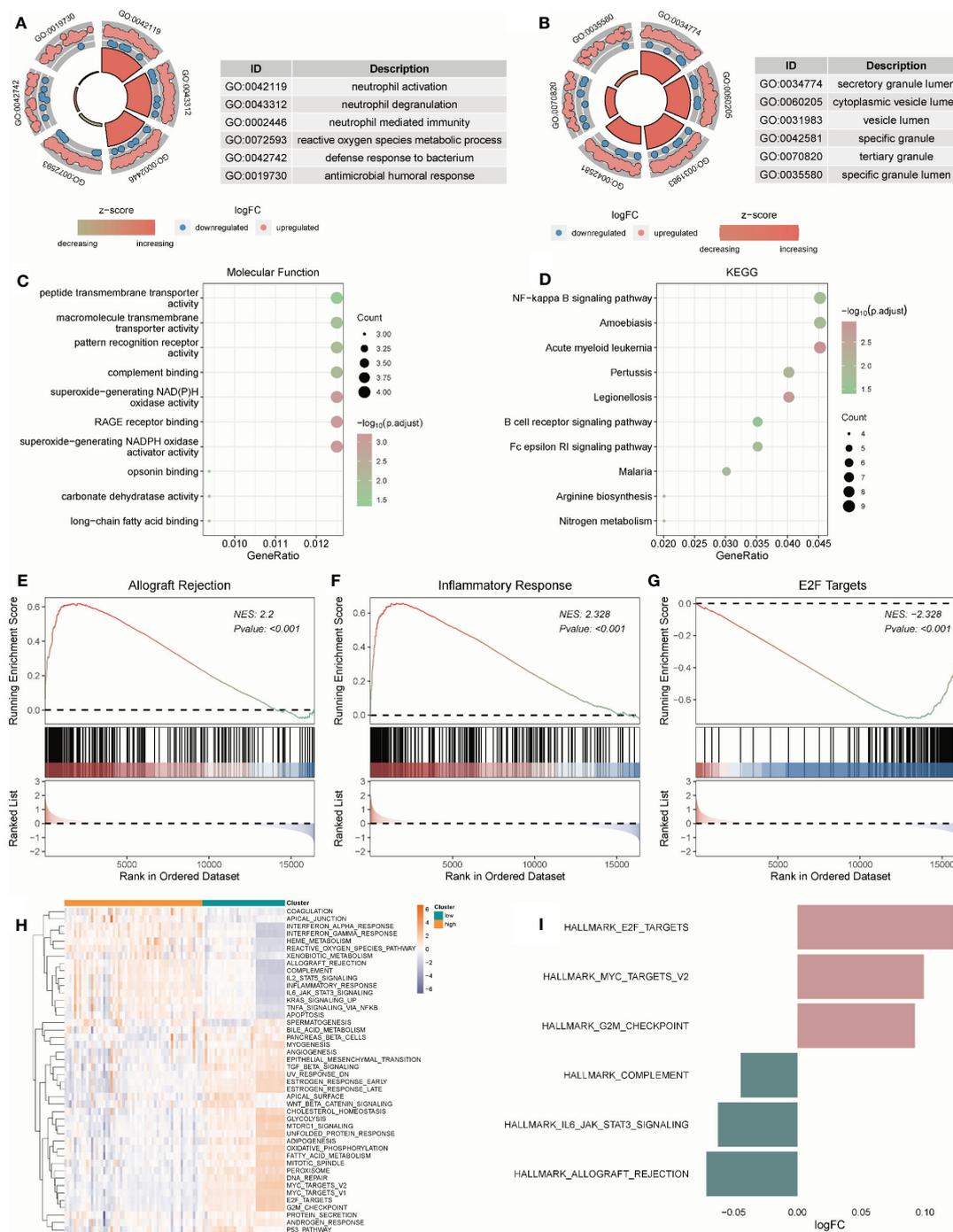


FIGURE 8

Functional enrichment analysis of genes in high and low-risk groups. (A–C) The functional enrichment in BP, CC, and MF analysis, respectively. (D) The KEGG analysis of differential expression genes. (H, I) GSVA-based analysis of biological function enrichment of hallmark gene set by heat map (H) and bar plot (I). (E–G): Allograft Rejection (E), Inflammatory Response (F), and E2F Targets (G) show three pathways related to high and low-risk groups. The abscissa is the gene, sorted according to the logFC value in the differential gene list, logFC greater than 0 indicates up-regulated genes are marked in red, logFC less than 0 indicates down-regulated genes are marked in blue, the upper ordinate is the enrichment score, and the lower ordinate is the logFC value. BP, Biological Process; CC, Cellular Component; MF, Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

different immune cells were negatively correlated, among which the highest positive correlation was between T cells CD4 naive and T cells CD8 ($r = 0.518$), and the highest negative correlation was observed between dendritic cells activated and monocytes ($r = -0.639$) (Figure 10B). These correlations with different immune cells and hub genes are shown in Figures 10C–N.

Construction of immunologically characterized phenotypes

According to the analysis results of ssGSEA, ARDS patients in the Combined Datasets were divided into two different immune phenotypes (Cluster1: $n = 22$; Cluster2: $n = 35$) by unsupervised

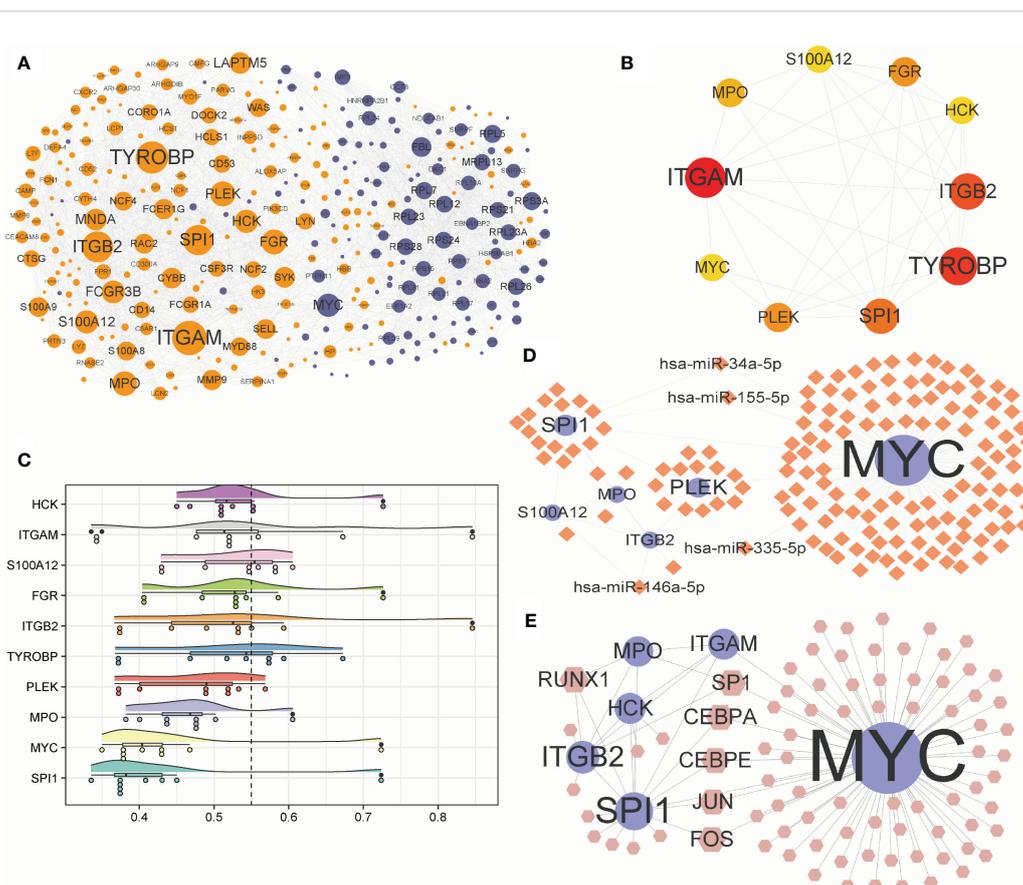


FIGURE 9

Construction of PPI network and interconnected regulatory network. (A) protein-protein interaction network constructed by differential expression genes. (B) Network of the top ten hub genes. (C) Cloud and rain map of hub gene importance. (D, E) miRNA-mRNA (D) and TF-mRNA (E) regulatory network constructed by hub genes. PPI, protein-protein interaction; TF, Transcription Factor.

consensus clustering (Figures 11A, B; Supplementary Table S9). Next, we compared the differences in hub genes (Figure 11C) and immune cells (Figure 11D) in immune signature phenotypes, and analysis correlations of differential immune cells (Figure 11E). The results showed that there were six hub genes with significant expression differences (Figure 11C) and 20 kinds of immune cells with significant infiltration differences (Figure 11D) in immune signature phenotypes (Figure 11C). The heat map of correlations between differential immune cells shows that there is a positive correlation between immune cells, among which Plasmacytoid dendritic cell and Monocyte have the highest positive correlation ($r=0.780$, $P=8.900e-13$). Differential hub genes and immune cells have correlations (Figures 11F-K).

Construction of molecular phenotypes and correlation analysis with immune infiltrating cells

According to the ten hub genes, ARDS patients in the Combined Datasets were divided into two different molecular

phenotypes (Group 1: $n = 38$; Group 2: $n = 19$) by unsupervised consensus clustering when the number of clusters is 2 (Figures 12A-C; Supplementary Table S10). The result of the principal component analysis shows that the two groups of ARDS patients could be well distinguished (Figure 12D).

Most of the hub genes were highly expressed in Group 1 and under-expressed in Group 2 (Figure 13A). There were significant differences in most hub genes between the two subgroups, and the expression value in Group 1 was significantly higher than in Group 2 (Figure 13B). In both the Combined Datasets (Figure 13C) and GSE76293 datasets (Figure 13D), the expression values of most genes in the ARDS group were significantly higher than those in the normal group. Further analysis revealed that more than 70% of immune cells significantly differed between the two groups (Figure 13E). Both in Group1 (Figure 13F) and Group2 (Figure 13G), most of the differential immune cells were positively correlated.

Most hub genes positively correlated with immune cells in groups 1 (Figure 14A) and 2 (Figure 14B). The top ten correlations of genes and immune cells in Group 1 and Group 2 were shown in Figures 14C-L and Figures 14M-V, respectively.

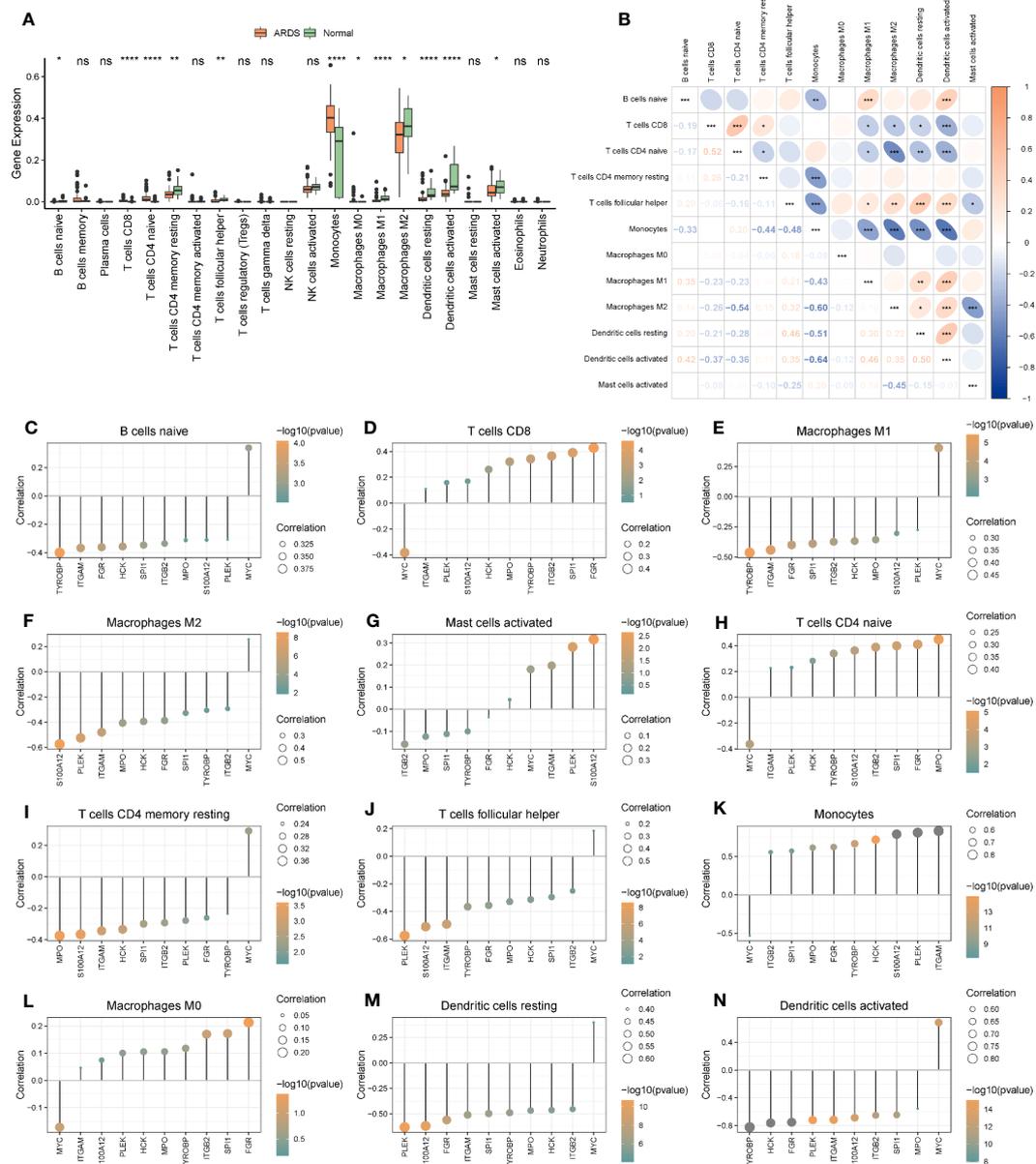


FIGURE 10 The immune cell infiltration and association with hub genes. (A) The immune cell infiltration in the ARDS and the Normal group. (B) Correlations of differently infiltrated Immune cells. (C–N) The association between hub genes and differently infiltrated Immune cells. * represents $p < 0.05$; ** represents $p < 0.01$; *** represents $p < 0.001$; **** represents $p < 0.0001$. ARDS, Acute Respiratory Distress Syndrome; ns, no significance.

Discussion

The discrimination of ARDS phenotypes holds clinical significance in the context of devising tailored therapeutic interventions. However, the identification of phenotypes is currently mainly based on clinical indicators. In this study, we delved into the critical insights gained from analyzing autophagy and metabolism-related genes to identify and differentiate phenotypes of ARDS. We stratified patients into two groups, high-risk and low-risk for ARDS development, which represents a crucial step toward better patient management. Through in-depth bioinformatics analysis, we identified a set of ten candidate hub

genes, namely ITGAM, TYROBP, ITGB2, SPI1, PLEK, FGR, MPO, S100A12, HCK, and MYC, with significantly altered expression. Additionally, we revealed two distinct phenotypes within the ARDS patient population, enhancing our understanding of the intrinsic heterogeneity of this syndrome.

In this study, we employed bioinformatics analysis to identify a panel of 10 hub genes, assessed the infiltration of immune cells using ssGSEA and CIBERSORT algorithms, and performed consensus clustering analysis to classify ARDS subtypes based on the identified hub genes and infiltrating immune cells. The crucial AMRDEGs associated with ARDS were identified by LASSO, SVM, and RF analysis, including CTSB and EEF 2. The application of

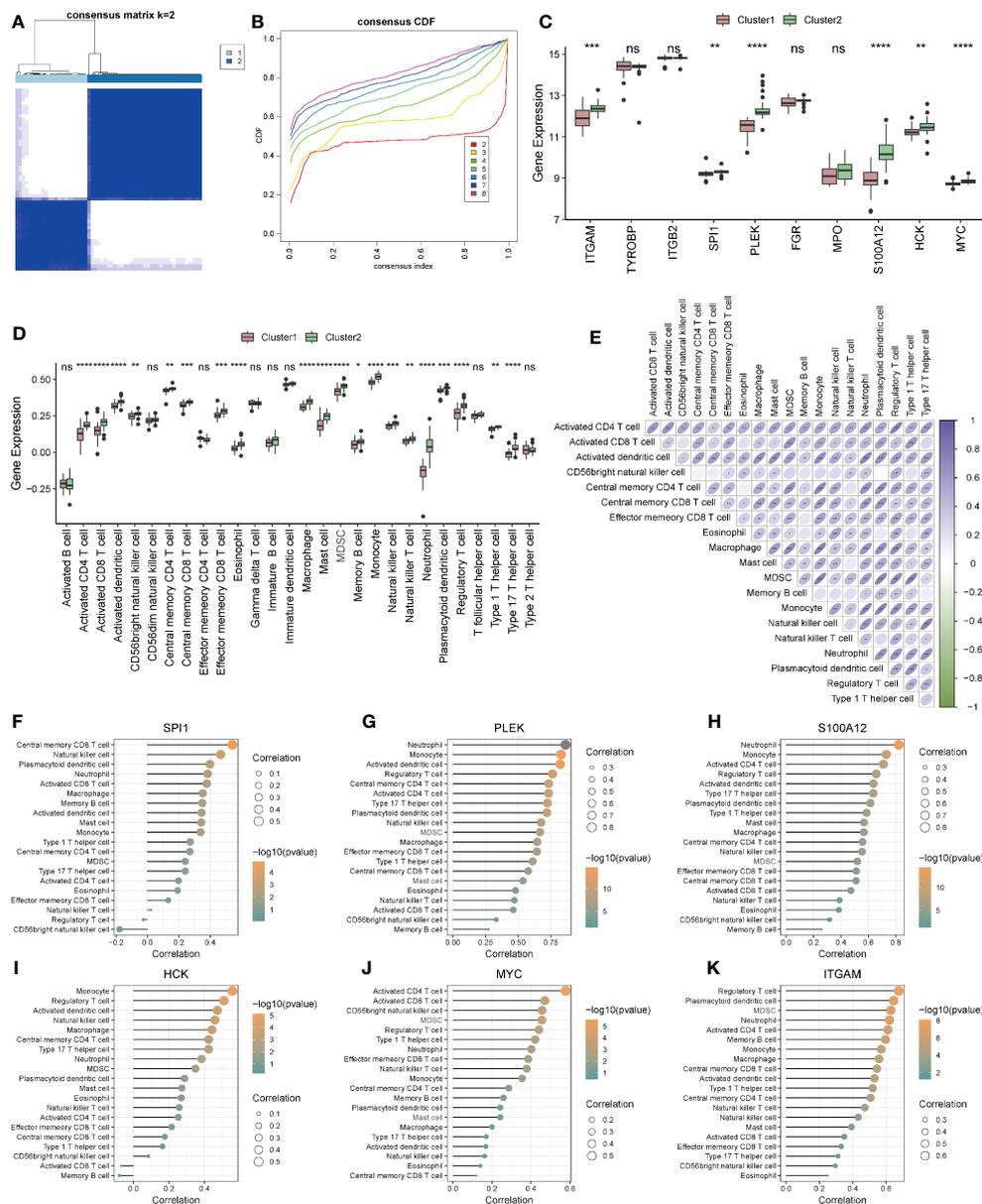
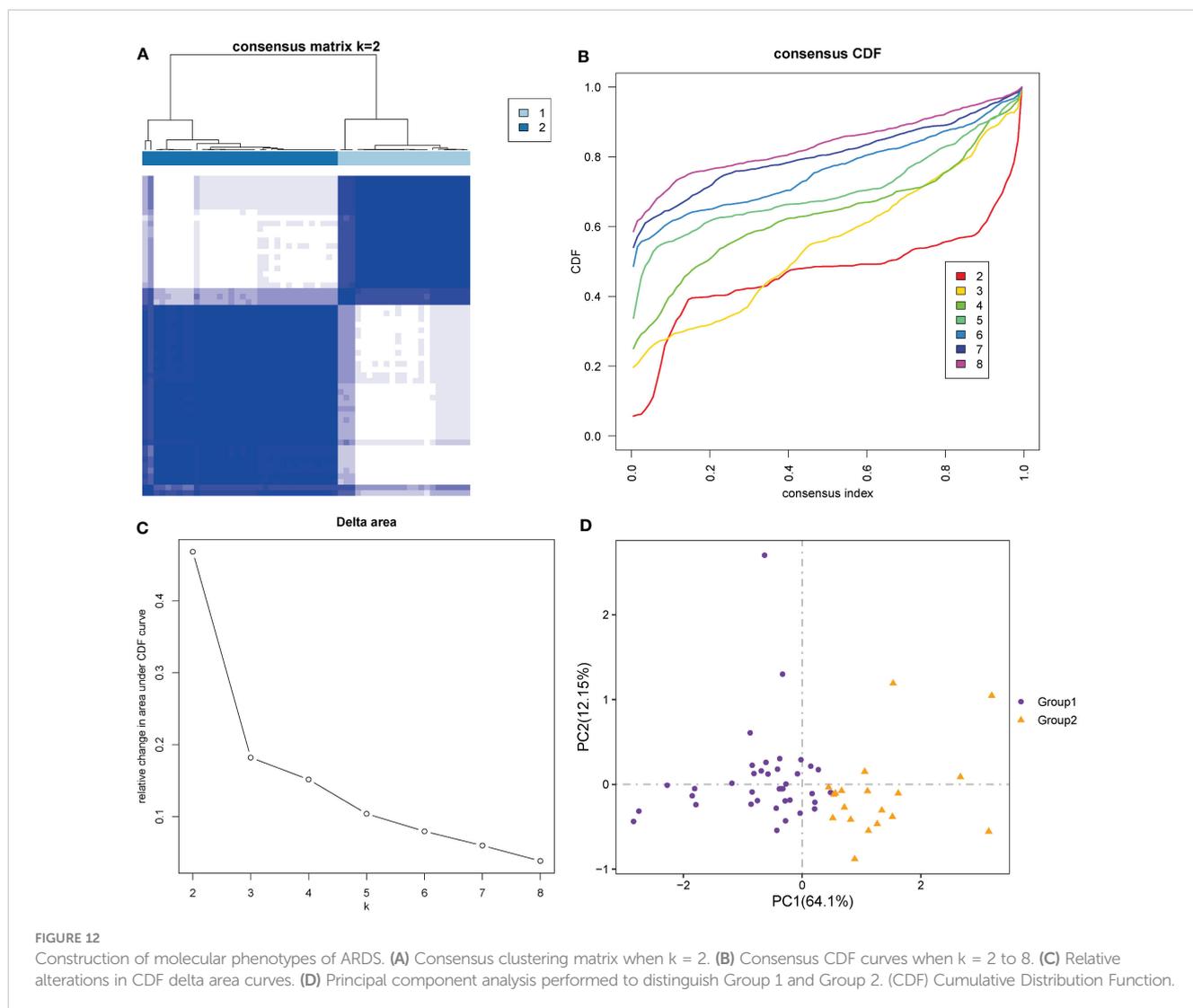


FIGURE 11 Construction and correlation analysis of immune signature phenotypes. (A) Consensus clustering matrix when $k = 2$. (B) Consensus CDF curves when $k = 2$ to 8. (C) Boxplot of hub genes in different immune phenotypes. (D) Boxplot of immune cells in different immune phenotypes. (E) Correlations of differential immune cells. (F–K): Correlation between differential hub genes and differential immune cells. * represents $p < 0.05$; ** represents $p < 0.01$; *** represents $p < 0.001$; **** represents $p < 0.0001$. CDF, Cumulative Distribution Function; ns, no significance.

three machine learning methods further ensured the reliability of screened genes. We then analyzed the diagnostic efficacy of crucial AMEDEGs for ARDS by constructing a nomogram. After that, a risk prediction model was established to evaluate the predictive power of ARDS according to the risk score calculated by the crucial AMERGs and was further verified in the validation cohort. Subsequently, we screened DEGs between the high- and the low-group and explored the function enrichment analysis based on GO, KEGG, GSEA, and GSVA. The hub genes associated with ARDS were selected by PPI analysis and explored the correlation with infiltrated immune cells. Finally, consensus clustering was applied to construct genomic and immune phenotypes, and the correlations

between hub genes and immune cells were further analyzed in different phenotypes. Our findings underscore the pivotal involvement of autophagy, metabolism, and immunity in the underlying pathophysiological mechanisms of ARDS.

As a result of our study, we selected ten autophagy and metabolism-related hub genes. A total of 6 genes, ITGAM, SPI1, PLEK, S100A12, HCK, and MYC, were expressed differently between different phenotypes. ITGAM has been identified to encode integrin αM (CD11b). It's a surface marker of monocytes and mediates various cell functions, including chemotaxis, adhesion, and transendothelial migration (52, 53). CD11b expressions were found enhanced in ARDS alveolar polymorphonuclear neutrophils (23), which is consistent with



our study, that ITGAM was highly expressed in ARDS patients in both the Combined Datasets and the validation set. SPI1 encoding transcription factor PU.1 regulates the innate immune function of alveolar macrophages (54) and functions to initiate inflammatory cascade through activating alveolar macrophages (55, 56). The transcribed protein of PLEK has been identified in platelets and is involved in platelet biology (57, 58). There is currently a lack of data on the relationship between SPI1 and PLEK with ARDS. In our study SPI1 was highly expressed in ARDS patients, while PLEK was not consistently expressed in the two datasets. S100A12, known as a RAGE ligand, elicits a proinflammatory response in leukocytes and endothelial cells. High concentrations have been found in lung tissue and bronchoalveolar lavage fluid in acute lung injury (59), consistent with our study. HCK is a member of the Src family expressed in myelomonocytic cell lineages, ultimately affecting cellular proliferation, differentiation, and migration (60), and has played an essential role in ARDS (61–63). The MYC-encoded protein forms a heterodimer with the transcription factor and has different expressions in ARDS patients and controls (64). Our study showed that the expression of HCK and MYC was controversial in different datasets but was consistent in

phenotype. These differences may be related to the heterogeneity of ARDS, which further illustrates the necessity of phenotypic classification of ARDS patients.

Differentially expressed genes in high- and low-risk groups have been identified. The high-risk group was significantly enriched in regulating immune inflammation, including neutrophil activation, inflammatory response, and complement. The related pathways in the high-risk group mainly focused on IL6-JAK-STAT3 signaling, NF-kappa signal pathway, and E2F targets were increased considerably in low-risk patients. The result revealed that immune inflammation may play a crucial role in the development of ARDS in high- and low-risk groups. This result is consistent with previous phenotypic classifications of ARDS (7, 9, 65). However, the previous phenotype classification based on inflammatory factors was limited to clinical markers, so it is necessary to classify phenotypes based on molecular mechanisms to achieve precise treatment of ARDS.

We subsequently evaluated the roles of hub genes in immune cell infiltration and constructed the immunophenotypes. After we had selected the ten autophagy and metabolism-related hub genes, it was found that they could be used to cluster patients with ARDS.

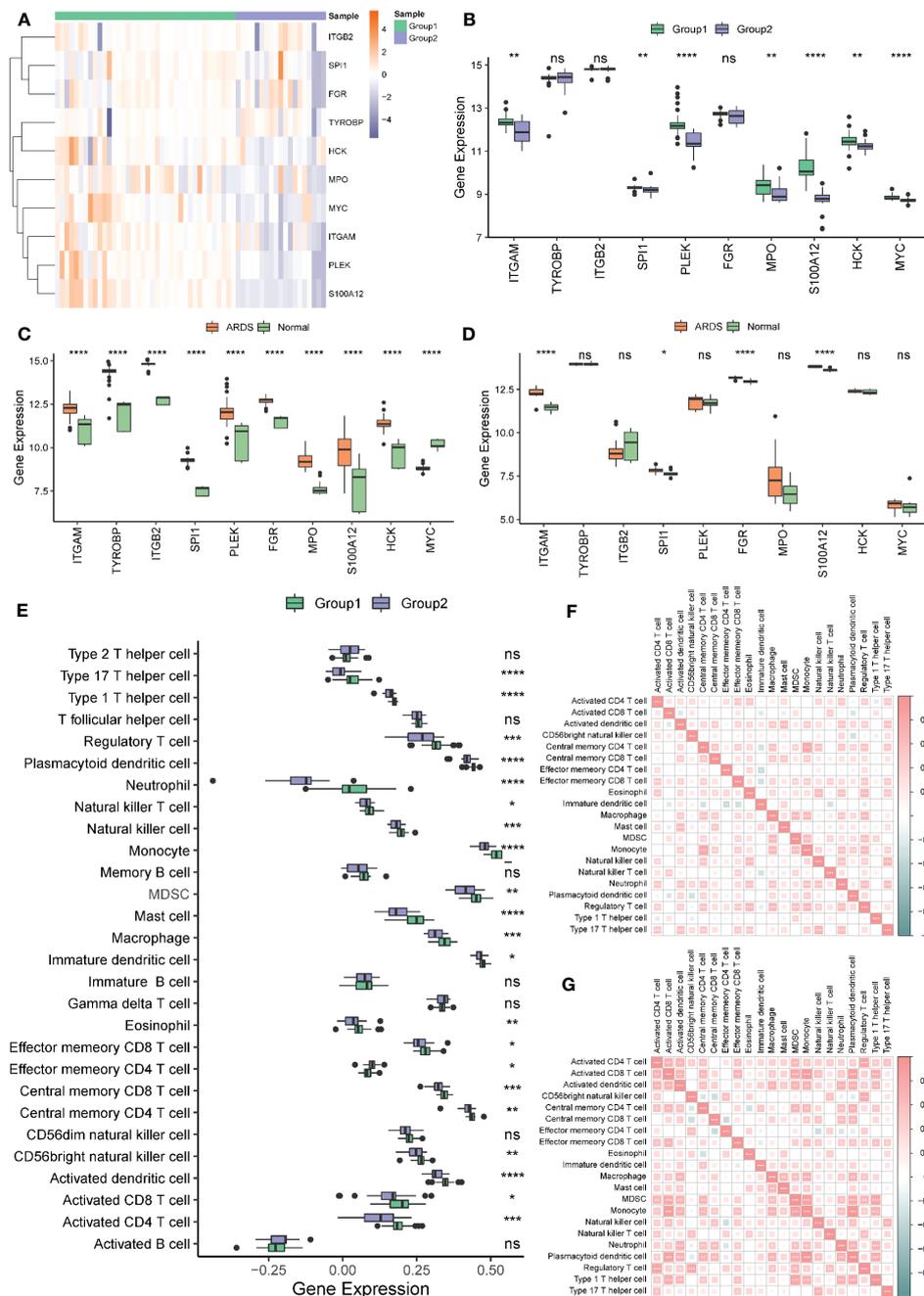


FIGURE 13

Hub genes and immune cell infiltration of molecular signature phenotypes. (A) Heatmap of hub genes expression in Group 1 and Group 2. (B) Boxplot of hub genes in Group 1 and Group 2. (C, D) Boxplot of hub genes between ARDS and Normal group in the Combined (C) and the GSE76293 dataset (D). (E) Immune cell infiltration between Group 1 and Group 2. (F, G) The correlation of the immune cells in Group 1 (F) and Group 2 (G). * represents $p < 0.05$; ** represents $p < 0.01$; *** represents $p < 0.001$; **** represents $p < 0.0001$. ARDS, Acute Respiratory Distress Syndrome; ns, no significance.

Our research found that most of the hub genes and immune cells were highly expressed in cluster 2 and group 1. It means that cluster 2 and group 1 are mainly immune activations. It is consistent with the current classification of ARDS on the high and low-inflammation types (8, 9). Further research is needed to integrate the molecular, immune, and clinical features of ARDS to classify subtypes and guide the precise treatment of ARDS. Patients in the different clusters and groups had differently infiltrated immune

cells, suggesting that autophagy and metabolism in patients with ARDS might regulate immune status, which might lead to a different prognosis.

Our study has limitations. First, all analyses were based on data obtained from public databases. GEO dataset is restricted in terms of species representation, sequencing platforms, molecular types, sample grouping, and sample quality. However, the availability of suitable datasets for our analysis was limited to the dataset used in

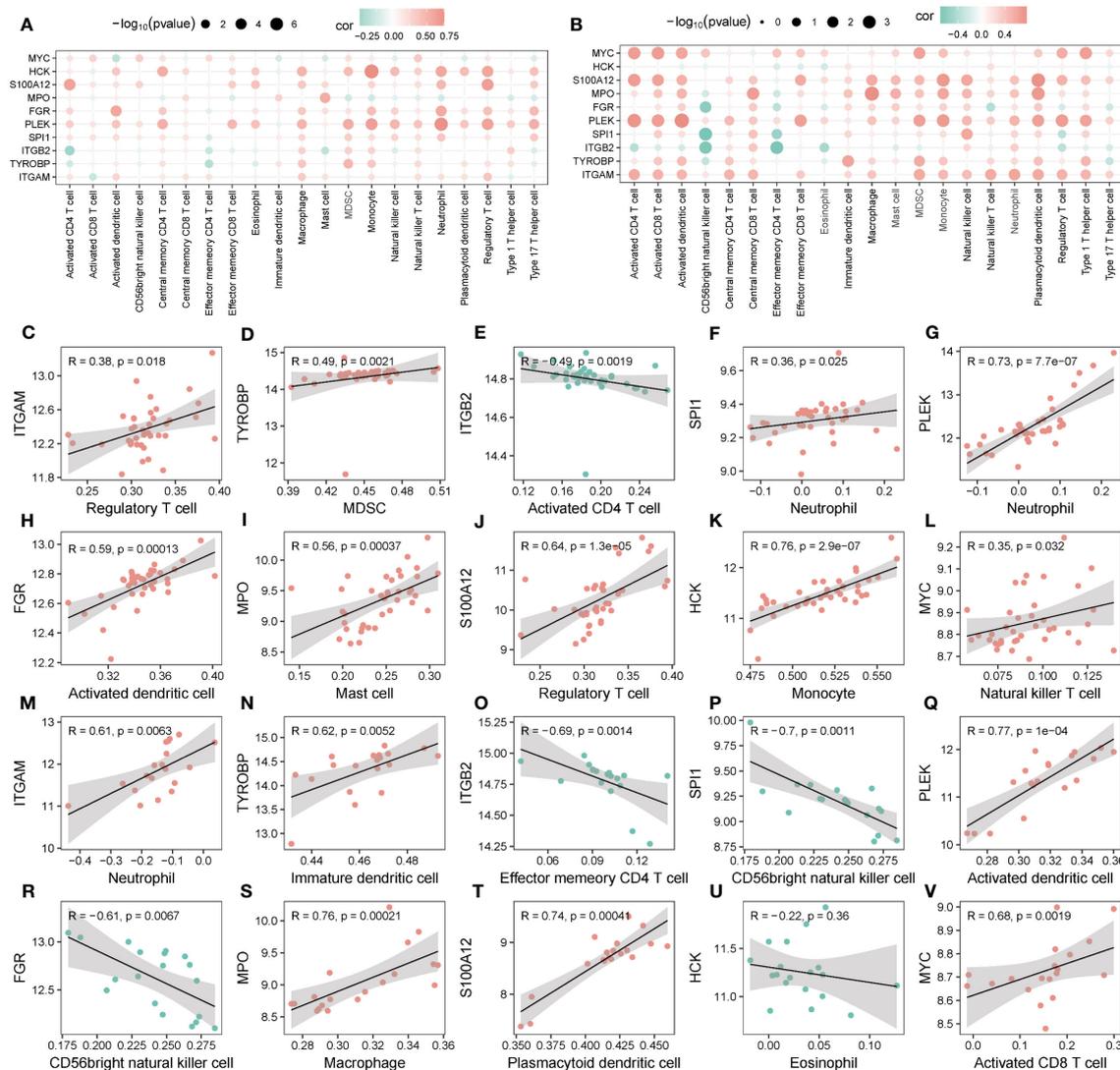


FIGURE 14 Correlation of hub genes and infiltrating immune cells. (A, B) Correlation between hub genes and immune cells in Group 1 (A) and Group 2 (B). (C-V) Scatter plot of top 10 correlations between hub genes and immune cells in Group1 (C-L) in Group2 (M-V).

this study. Although we have set the validated cohort, further validation would be best done in prospective studies. We have also emphasized our commitment to accumulating samples from our institution for sequencing, aiming to supplement the shortcomings of this study in our subsequent research. Additionally, we acknowledge the limitations of different sample types can lead to variations in gene expression profiles and potentially introduce noise if the biological relevance of sample types is not appropriately considered. Secondly, the results do not provide a comprehensive view of the role of autophagy in ARDS because the RNA sequence findings do not distinguish between phosphorylation and dephosphorylation of autophagy-related proteins. Proteomics analysis is required before translational application. Thirdly, it is imperative to conduct experimental investigations to delve deeper into the molecular mechanisms involving the interplay between autophagy, metabolism-related genes, and immune cells. Additionally, thorough experimental

validation is warranted to ascertain the prognostic implications for patients exhibiting distinct ARDS phenotypes.

Our study identified ten candidate hub genes and classified ARDS patients into two distinct phenotypes. These findings hold direct clinical significance, paving the way for more precise diagnosis and personalized treatment strategies. Moreover, our research establishes a foundation for future experimental and clinical investigations, providing valuable directions for further exploration of ARDS and its therapeutic approaches. In future studies, the expression of the hub genes in ARDS patients can be explored in relation to clinical parameters such as disease severity and treatment response. Furthermore, we can delve deeper into the investigation of immune cells associated with the hub genes to better understand their roles in the pathogenesis of ARDS. Specifically, we can explore the distribution and functions of these immune cells in different phenotypes of ARDS. This may aid physicians in better guiding the treatment of ARDS patients.

Conclusion

In conclusion, the present study picked up ten hub genes of ARDS related to autophagy and metabolism, namely ITGAM, TYROBP, ITGB2, SPI1, PLEK, FGR, MPO, S100A12, HCK, and MYC, which could cluster ARDS patients into different molecular phenotypes. Besides, we also explored the infiltrated immune cells of ARDS and clustered ARDS patients into different immunophenotypes. We further analyzed their correlations with hub genes, providing a new perspective on the role of immunity, autophagy, and metabolism in ARDS.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

FX and FG contributed to the conception of the protocol. FX contributed to data acquisition. FX wrote the first version of the

References

- Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *Jama* (2016) 315(8):788–800. doi: 10.1001/jama.2016.0291
- Neto AS, Barbas CSV, Simonis FD, Artigas-Raventós A, Canet J, Determann RM, et al. Epidemiological characteristics, practice of ventilation, and clinical outcome in patients at risk of acute respiratory distress syndrome in intensive care units from 16 countries (PRoVENT): an international, multicentre, prospective study. *Lancet Respir Med* (2016) 4(11):882–93. doi: 10.1016/S2213-2600(16)30305-8
- Huang X, Zhang R, Fan G, Wu D, Lu H, Wang D, et al. Incidence and outcomes of acute respiratory distress syndrome in intensive care units of mainland China: a multicentre prospective longitudinal study. *Crit Care (London England)* (2020) 24(1):515. doi: 10.1186/s13054-020-03112-0
- Prescott HC, Calfee CS, Thompson BT, Angus DC, Liu VX. Toward smarter lumping and smarter splitting: rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. *Am J Respir Crit Care Med* (2016) 194(2):147–55. doi: 10.1164/rccm.201512-2544CP
- Sinha P, Meyer NJ, Calfee CS. Biological phenotyping in sepsis and acute respiratory distress syndrome. *Annu Rev Med* (2023) 74:457–71. doi: 10.1146/annurev-med-043021-014005
- Sinha P, Spicer A, Delucchi KL, McAuley DF, Calfee CS, Churpek MM. Comparison of machine learning clustering algorithms for detecting heterogeneity of

manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by grants from the National Natural Science Foundation of China (grant numbers: 81871602 and 82102300), and a Special fund project for health science and technology development of Nanjing Municipal Health Commission (2022ZXNS07).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1209959/full#supplementary-material>

treatment effect in acute respiratory distress syndrome: A secondary analysis of three randomised controlled trials. *EBioMedicine* (2021) 74:103697. doi: 10.1016/j.ebiom.2021.103697

7. Sinha P, Delucchi KL, Chen Y, Zhuo H, Abbott J, Wang C, et al. Latent class analysis-derived subphenotypes are generalisable to observational cohorts of acute respiratory distress syndrome: a prospective study. *Thorax* (2022) 77(1):13–21. doi: 10.1136/thoraxjnl-2021-217158

8. Maddali MV, Churpek M, Pham T, Rezoagli E, Zhuo H, Zhao W, et al. Validation and utility of ARDS subphenotypes identified by machine-learning models using clinical data: an observational, multicohort, retrospective analysis. *Lancet Respir Med* (2022) 10(4):367–77. doi: 10.1016/S2213-2600(21)00461-6

9. Dahmer MK, Yang G, Zhang M, Quasney MW, Sapru A, Weeks HM, et al. Identification of phenotypes in paediatric patients with acute respiratory distress syndrome: a latent class analysis. *Lancet Respir Med* (2022) 10(3):289–97. doi: 10.1016/S2213-2600(21)00382-9

10. Chen H, Yu Q, Xie J, Liu S, Pan C, Liu L, et al. Longitudinal phenotypes in patients with acute respiratory distress syndrome: a multi-database study. *Crit Care (London England)* (2022) 26(1):340. doi: 10.1186/s13054-022-04211-w

11. Xian H, Liu Y, Rundberg Nilsson A, Gatchalian R, Crother TR, Tourtellotte WG, et al. Metformin inhibition of mitochondrial ATP and DNA synthesis abrogates NLRP3 inflammasome activation and pulmonary inflammation. *Immunity* (2021) 54(7):1463–1477.e1411. doi: 10.1016/j.immuni.2021.05.004

12. Hue S, Beldi-Ferchiou A, Bendib I, Surenaud M, Fourati S, Frapard T, et al. Uncontrolled innate and impaired adaptive immune responses in patients with COVID-19 acute respiratory distress syndrome. *Am J Respir Crit Care Med* (2020) 202(11):1509–19. doi: 10.1164/rccm.202005-1885OC
13. Mizushima N, Komatsu M. Autophagy: renovation of cells and tissues. *Cell* (2011) 147(4):728–41. doi: 10.1016/j.cell.2011.10.026
14. Deretic V, Saitoh T, Akira S. Autophagy in infection, inflammation and immunity. *Nat Rev Immunol* (2013) 13(10):722–37. doi: 10.1038/nri3532
15. Levine B, Mizushima N, Virgin HW. Autophagy in immunity and inflammation. *Nature* (2011) 469(7330):323–35. doi: 10.1038/nature09782
16. Zhao H, Chen H, Xiaoyin M, Yang G, Hu Y, Xie K, et al. Autophagy activation improves lung injury and inflammation in sepsis. *Inflammation* (2019) 42(2):426–39. doi: 10.1007/s10753-018-00952-5
17. Chichger H, Rounds S, Harrington EO. Endosomes and autophagy: regulators of pulmonary endothelial cell homeostasis in health and disease. *Antioxid Redox Signaling* (2019) 31(13):994–1008. doi: 10.1089/ars.2019.7817
18. Mizushima N, Levine B. Autophagy in human diseases. *New Engl J Med* (2020) 383(16):1564–76. doi: 10.1056/NEJMra2022774
19. Bos LD, Weda H, Wang Y, Knobel HH, Nijsen TM, Vink TJ, et al. Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome. *Eur Respir J* (2014) 44(1):188–97. doi: 10.1183/09031936.00005614
20. Stringer KA, McKay RT, Karnovsky A, Quémerais B, Lacy P. Metabolomics and its application to acute lung diseases. *Front Immunol* (2016) 7:44. doi: 10.3389/fimmu.2016.00044
21. Morrell ED, Radella F2nd, Manicone AM, Mikacenic C, Stapleton RD, Gharib SA, et al. : peripheral and alveolar cell transcriptional programs are distinct in acute respiratory distress syndrome. *Am J Respir Crit Care Med* (2018) 197(4):528–32. doi: 10.1164/rccm.201703-0614LE
22. Dolinay T, Kim YS, Howrylak J, Hunninghake GM, An CH, Fredenburgh L, et al. Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am J Respir Crit Care Med* (2012) 185(11):1225–34. doi: 10.1164/rccm.201201-0003OC
23. Juss JK, House D, Amour A, Begg M, Herre J, Storisteanu DM, et al. Acute respiratory distress syndrome neutrophils have a distinct phenotype and are resistant to phosphoinositide 3-Kinase inhibition. *Am J Respir Crit Care Med* (2016) 194(8):961–73. doi: 10.1164/rccm.201509-1818OC
24. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* (2013) 41(Database issue):D991–995. doi: 10.1093/nar/gks1193
25. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioConductor. *Bioinf (Oxford England)* (2007) 23(14):1846–7. doi: 10.1093/bioinformatics/btm254
26. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinf (Oxford England)* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034
27. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. *Database J Biol Database Curation* (2010) 2010:baq020. doi: 10.1093/database/baq020
28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
29. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinf* (2013) 14:244. doi: 10.1186/1471-2105-14-244
30. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol* (2004) 5(1):11–8. doi: 10.3348/kjr.2004.5.1.11
31. Cai W, van der Laan M. Nonparametric bootstrap inference for the targeted highly adaptive least absolute shrinkage and selection operator (LASSO) estimator. *Int J Biostatistics* (2020) 16(2):1–36. doi: 10.1515/ijb-2017-0070
32. Nedaie A, Najafi AA. Support vector machine with Dirichlet feature mapping. *Neural Networks Off J Int Neural Netw Soc* (2018) 98:87–101. doi: 10.1016/j.neunet.2017.11.006
33. Rigatti SJ. Random forest. *J Insurance Med (New York NY)* (2017) 47(1):31–9. doi: 10.17849/inm-47-01-31-39.1
34. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenet* (2019) 11(1):123. doi: 10.1186/s13148-019-0730-1
35. Park SY. Nomogram: An analogue tool to deliver digital knowledge. *J Thorac Cardiovasc Surg* (2018) 155(4):1793. doi: 10.1016/j.jtcvs.2017.12.107
36. Van Calster B, Wynants L, Verbeek JEM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: A guide for investigators. *Eur Urol* (2018) 74(6):796–804. doi: 10.1016/j.eururo.2018.08.038
37. Blake JA, Christie KR, Dolan ME, Drabkin HJ, Hill DP, Ni L, et al. Gene ontology consortium: going forward. *Nucleic Acids Res* (2015) 43(Database issue):D1049–1056. doi: 10.1093/nar/gku1179
38. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27
39. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge Mass)* (2021) 2(3):100141. doi: 10.1016/j.xinn.2021.100141
40. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 1(6):417–25. doi: 10.1016/j.cels.2015.12.004
41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci United States America* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102
42. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
43. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* (2019) 47(D1):D607–d613. doi: 10.1093/nar/gky1131
44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303
45. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* (2014) 8 Suppl 4(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11
46. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res* (2020) 48(D1):D148–d154. doi: 10.1093/nar/gkz896
47. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* (2018) 46(D1):D380–d386. doi: 10.1093/nar/gkx1013
48. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
49. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol (Clifton NJ)* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12
50. Brière G, Darbo É, Thébault P, Uricaru R. Consensus clustering applied to multi-omics disease subtyping. *BMC Bioinf* (2021) 22(1):361. doi: 10.1186/s12859-021-04279-1
51. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinf (Oxford England)* (2010) 26(12):1572–3. doi: 10.1093/bioinformatics/btq170
52. Zhou M, Wang X, Shi Y, Ding Y, Li X, Xie T, et al. Deficiency of ITGAM attenuates experimental abdominal aortic aneurysm in mice. *J Am Heart Assoc* (2021) 10(7):e019900. doi: 10.1161/JAHA.120.019900
53. Solovjov DA, Pluskota E, Plow EF. Distinct roles for the alpha and beta subunits in the functions of integrin alphaMbeta2. *J Biol Chem* (2005) 280(2):1336–45. doi: 10.1074/jbc.M406968200
54. Staitieh BS, Fan X, Neveu W, Guidot DM. Nrf2 regulates PU.1 expression and activity in the alveolar macrophage. *Am J Physiol Lung Cell Mol Physiol* (2015) 308(10):L1086–1093. doi: 10.1152/ajplung.00355.2014
55. Berclaz PY, Carey B, Fillipi MD, Wernke-Dollries K, Geraci N, Cush S, et al. GM-CSF regulates a PU.1-dependent transcriptional program determining the pulmonary response to LPS. *Am J Respir Cell Mol Biol* (2007) 36(1):114–21. doi: 10.1165/rccm.2006-0174OC
56. Qian F, Deng J, Lee YG, Zhu J, Karpurapu M, Chung S, et al. The transcription factor PU.1 promotes alternative macrophage polarization and asthmatic airway inflammation. *J Mol Cell Biol* (2015) 7(6):557–67. doi: 10.1093/jmcb/mjv042
57. Coppinger JA, Cagney G, Toomey S, Kislinger T, Belton O, McRedmond JP, et al. Characterization of the proteins released from activated platelets leads to localization of novel platelet proteins in human atherosclerotic lesions. *Blood* (2004) 103(6):2096–104. doi: 10.1182/blood-2003-08-2804
58. Fröbel J, Cadeddu RP, Hartwig S, Bruns I, Wilk CM, Kündgen A, et al. Platelet proteome analysis reveals integrin-dependent aggregation defects in patients with myelodysplastic syndromes. *Mol Cell Proteomics MCP* (2013) 12(5):1272–80. doi: 10.1074/mcp.M112.023168
59. Wittkowski H, Sturrock A, van Zoelen MA, Viemann D, van der Poll T, Hoidal JR, et al. Neutrophil-derived S100A12 in acute lung injury and respiratory distress syndrome. *Crit Care Med* (2007) 35(5):1369–75. doi: 10.1097/01.CCM.0000262386.32287.29
60. Ernst M, Inglese M, Scholz GM, Harder KW, Clay FJ, Bozinovski S, et al. Constitutive activation of the SRC family kinase Hck results in spontaneous pulmonary inflammation and an enhanced innate immune response. *J Exp Med* (2002) 196(5):589–604. doi: 10.1084/jem.20020873
61. Khadarog RG, He R, Parodo J, Powers KA, Marshall JC, Kapus A, et al. The role of the Src family of tyrosine kinases after oxidant-induced lung injury in vivo. *Surgery* (2004) 136(2):483–8. doi: 10.1016/j.surg.2004.05.029

62. Luo CY, Li Y, Li X, Liang X, Wang Q, Ma YH, et al. Alleviation of lipopolysaccharide-induced acute respiratory distress syndrome in rats by Yiqi Huayu Jiedu Decoction: A tandem mass tag-based proteomics study. *Front Pharmacol* (2020) 11:1215. doi: 10.3389/fphar.2020.01215
63. Kao TI, Chen PJ, Wang YH, Tseng HH, Chang SH, Wu TS, et al. Bletinin ameliorates neutrophilic inflammation and lung injury by inhibiting Src family kinase phosphorylation and activity. *Br J Pharmacol* (2021) 178(20):4069–84. doi: 10.1111/bph.15597
64. Sun M, Yang Q, Hu C, Zhang H, Xing L. Identification and validation of autophagy-related genes in sepsis-induced acute respiratory distress syndrome and immune infiltration. *J Inflammation Res* (2022) 15:2199–212. doi: 10.2147/JIR.S355225
65. Sinha P, Delucchi KL, Thompson BT, McAuley DF, Matthay MA, Calfee CS. Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med* (2018) 44(11):1859–69. doi: 10.1007/s00134-018-5378-3