Journal of Inflammation Research

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Identification of Biomarkers for Sepsis-Induced Acute Lung Injury Through Bioinformatics and Machine Learning Approaches, with Experimental Validation

Yannian Luo[1,*], Juan Xu[2,*], Nannan He[1], Wen Cao[1]

[1]Department of Critical Care Medicine, The Second Hospital of Lanzhou University, Lanzhou, Gansu Province, People's Republic of China; [2]Department of Pediatric Neurology, The Second Hospital of Lanzhou University, Lanzhou, Gansu Province, People's Republic of China

*These authors contributed equally to this work

Correspondence: Wen Cao, Email smilecao2009@126.com

Background: Sepsis-induced acute lung injury (ALI) remains a life-threatening condition due to the lack of reliable early diagnostic biomarkers. Machine learning offers powerful tools for analyzing high-dimensional gene expression data and identifying potential biomarkers and therapeutic targets.

Methods: Five datasets (GSE10474, GSE32707, GSE66890, GSE10361, GSE3037) were obtained from the GEO database. After assessment and normalization, GSE10474, GSE32707, and GSE66890 were combined as a training set to identify differentially expressed genes (DEGs). DEGs were intersected with genes from key modules identified by weighted gene co-expression network analysis (WGCNA), yielding 213 overlapping genes. These were analyzed via Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment. Eight machine learning algorithms (RF, SVM, GLM, GBM, KNN, NNET, LASSO, DT) were used to develop diagnostic models, which were validated on GSE10361 and GSE3037. Model performance was evaluated using a nomogram, calibration curves, and decision curve analysis (DCA). Immune and inflammatory states were assessed using the CIBERSORT algorithm. Potential therapeutic compounds were identified through the DSigDB database via the Enrichr platform. Molecular docking and molecular dynamics simulations examined interactions between Resveratrol and selected targets. In vitro experiments validated these findings.

Results: A total of 213 candidate genes were identified by intersecting DEGs with WGCNA-derived MEblue module genes. GO and KEGG analyses indicated associations with immune activation and bacterial infection. Four key genes (DDAH2, PNPLA2, STXBP2, TCN1) were selected using eight machine learning algorithms. The diagnostic model showed good performance via nomogram, calibration curve, and DCA. Molecular docking revealed stable binding of Resveratrol to these genes. In vitro, Resveratrol pretreatment alleviated LPS-induced ALI by modulating the core genes.

Conclusion: The four genes may serve as diagnostic biomarkers for sepsis-ALI. Resveratrol represents a potential therapeutic strategy by targeting these genes.

Keywords: sepsis-induced ALI, machine learning algorithms, WGCNA analysis, CIBERSORT analysis, molecular docking

Introduction

Sepsis is a critical condition characterized by organ dysfunction resulting from an imbalanced immune response to infection, which can eventually progress to multiple organ failure and mortality.[1,2] ALI is a frequent and deadly complication of sepsis and may further develop into acute respiratory distress syndrome (ARDS).[3] A large international study has shown that approximately 75% of ALI cases are sepsis-induced.[4] Reports from the United States indicate that over 210,000 cases of sepsis-induced ALI occur each year. Additionally, patients suffering from sepsis-related ALI tend to experience more severe disease, slower lung recovery, and higher mortality rates compared to those with ALI from other causes.[5] ALI progresses rapidly following the initial injury, and there is no consensus on biomarkers that can directly diagnose ALI or assess lung damage. Therefore, identifying diagnostic biomarkers for ALI is crucial.

Received: 3 June 2025
Accepted: 30 August 2025
Published: 4 October 2025

Machine learning, as a rapidly advancing field, has become indispensable in the analysis of gene expression profiles, which reflect the collective expression patterns of all genes within a cell or tissue under specific conditions.[6] These data are typically high-dimensional, encompassing inherent complexity and potential noise, posing challenges for traditional statistical methods in their interpretation. Machine learning algorithms exhibit unique advantages in handling such high-dimensional, non-linear, and noisy data, making them particularly crucial for identifying key genes associated with diseases. By integrating bioinformatics tools with advanced machine learning techniques, researchers can efficiently identify genes closely linked to disease progression, ultimately providing critical insights for disease prevention, diagnosis, and treatment.[7,8] For instance, Ming et al[9] applied machine learning techniques, including LASSO, to identify critical genes involved in sepsis-associated ARDS and explored their relationship with immune cell infiltration, which has been recognized as a key factor in the pathogenesis of sepsis.[10] Moreover, computational approaches such as CIBERSORT provide powerful tools for analyzing immune cell populations from bulk gene expression data, contributing to a deeper understanding of disease mechanisms.

In this study, three datasets related to sepsis-induced ALI were retrieved from the GEO database. DEGs were determined and overlapped with key gene modules identified via WGCNA. To screen for biomarkers linked to sepsis-related ALI, eight distinct machine learning methods were applied. Immune infiltration characteristics were assessed using the CIBERSORT algorithm, and the selected biomarkers were further validated with an external dataset. Finally, molecular docking was performed to identify potential therapeutic compounds, and validation was carried out in cell models.

Materials and Methods

Data Sources Used for Analysis

We queried the GEO database (https://www.ncbi.nlm.nih.gov/geo/) using relevant keywords including "sepsis-induced acute lung injury", "sepsis", and "human" to identify appropriate transcriptomic datasets. Based on predefined inclusion criteria, five datasets (GSE10474, GSE32707, GSE66890, GSE10361, and GSE3037) were selected for downstream analysis. Among these, GSE10474, GSE32707, and GSE66890 were integrated to construct a diagnostic model for sepsis-induced ALI, while GSE10361 and GSE3037 were used to evaluate the robustness and generalizability of the model. The overall analytical workflow is illustrated in Figure 1.

Identificated DEGs

To minimize batch effects between datasets, gene expression matrices were combined using the "sva" package in R. Principal component analysis (PCA), a commonly used method for feature extraction and reducing dimensionality, was then applied to detect underlying patterns, simplify the data structure, and reveal hidden correlations among variables.[11,12] The expression matrices from GSE10474, GSE32707, and GSE66890 were merged into a unified dataset. Subsequently, The "limma" R package was employed to carry out differential expression analysis. DEGs were identified based on the following thresholds: $|\log2FoldChange| > 0.6$ and a false discovery rate (FDR) < 0.05.

WGCNA Analysis

WGCNA utilizes gene expression data to build a scale-free network, clustering genes with comparable expression trends into separate modules. These modules can then be correlated with specific phenotypic traits or clinical variables, enabling the identification of gene networks potentially linked to biological processes or disease states. This methodology provides a powerful framework for uncovering candidate biomarkers or therapeutic targets. Modules are defined by applying a soft-thresholding power to emphasize strong gene-gene correlations, facilitating the detection of tightly co-expressed gene clusters.

Functional Enrichment Analysis

Functional gene analysis is a cornerstone of biological research, aiming to elucidate the roles of genes and their expression products in cellular processes and disease mechanisms. With continuous advancements in computational tools, increasingly sophisticated methods have emerged to interpret gene function at the systems level. In this research, GO functional annotation and KEGG pathway enrichment were conducted using the R package clusterProfiler, based on the normalized gene expression data matrix.[13,14]

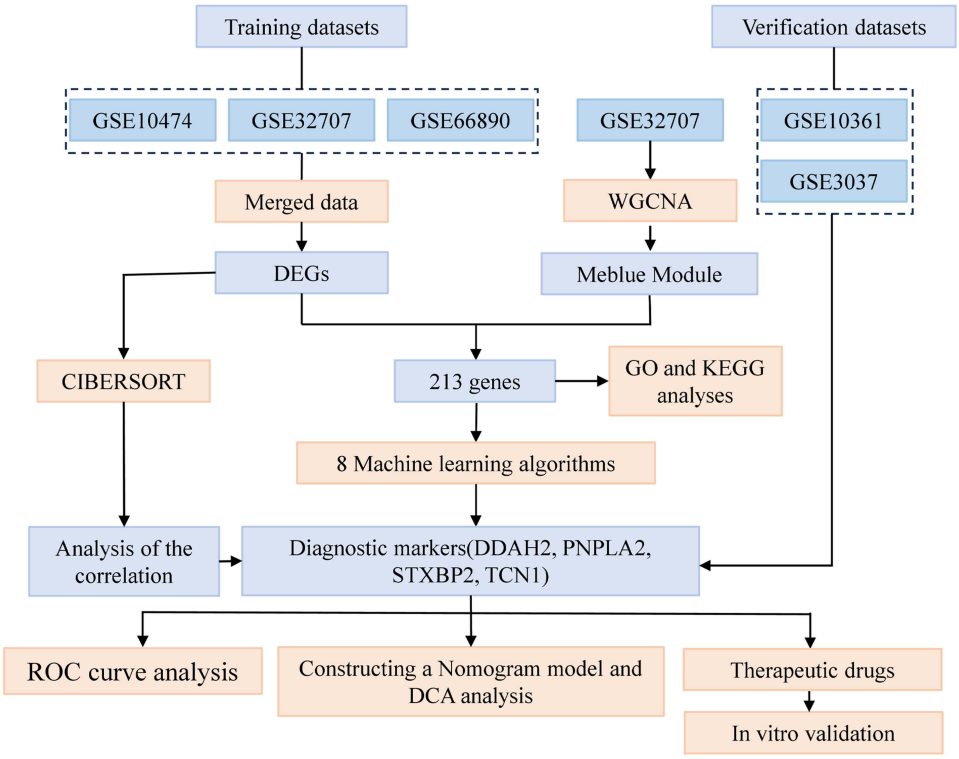

Figure 1 The flow chart portraying the investigation procedure.
Abbreviations: GEO, Gene Expression Omnibus; CIBERSORT, Cell-Type Identification by Estimating Relative Subsets of RNA Transcripts; DEGs, Differentially Expressed Genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ROC, Receiver Operating Characteristic Curve; DCA, Decision Curve Analysis.

Using Machine Learning to Screen Hub Genes

In this study, eight different machine learning models were established. Their performance was assessed by computing the area under the receiver operating characteristic curve (AUC), with the model demonstrating the highest predictive accuracy chosen for subsequent analyses. Feature genes, referred to as hub-DEGs, were defined as those identified by at least three of the machine learning algorithms. The intersection of selected genes across models was visualized using the R package Upset.[15]

Construction of Nomogram and Analysis of ROC Curve

A nomogram, commonly utilized for outcome prediction based on multiple variables, was developed in this study using the rms package in R. The diagnostic utility of the identified hub genes was evaluated through receiver operating characteristic (ROC) curve analysis. To further validate the model's reliability, ROC curves were also plotted using the GSE10361 and GSE3037 datasets as independent validation sets. AUC values exceeding 0.5 were regarded as indicative of acceptable predictive performance.[16] All ROC-related computations were conducted with the pROC package in R.

Analysis of Immune Infiltration

To explore the association between immune cell infiltration and sepsis-induced ALI, the CIBERSORT algorithm[17] was applied to estimate the relative proportions of 22 immune cell subsets. Subsequently, we analyzed the correlations between the expression levels of critical genes and immune cell distribution in both healthy individuals and patients with sepsis-related ALI.

Drugs Screened and Molecular Docking

We performed drug-gene interaction analysis via the DSigDB database on the Enrichr platform to identify candidate drugs targeting key genes (DDAH2, PNPLA2, STXBP2, and TCN1). The structural information of bioactive compounds was sourced from the Traditional Chinese Medicine Active Compound Database. Using ChemBio3D 14.0, the spatial conformations of these molecules were optimized for energy and saved in mol2 format. Subsequently, the compounds

https://doi.org/10.2147/JIR.S539899

were prepared in pdbq format with AutoDockTools 1.5.6.[18] The 3D crystal structures of target proteins were retrieved from UniProt, where water and organic molecules were removed using Notepad2. The protein structures were further processed in AutoDockTools 1.5.6 to add hydrogens, assign charges, and define atom types,[19] then saved in pdbqt format. Molecular docking was conducted with AutoDock Vina, and the docking poses were visualized using PyMOL 2.6.[20]

Cell Culture and Viability Assay

The human bronchial epithelial cell line BEAS-2B (Cat No.: SCSP-5067) was sourced from the Chinese Academy of Sciences. Cells were maintained in high-glucose DMEM supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin, incubated at 37°C with 5% CO_2. BEAS-2B cells were plated in 96-well plates at a density of 5×10^3 cells per well. After adherence, cells were exposed to varying concentrations of lipopolysaccharide (LPS; Sigma, USA) at 0, 2.5, 5, and 10 μg/mL for 24 hours.[21] One hour before LPS exposure, cells were pretreated with 10 μM resveratrol (Selleck, Shanghai, China).[22] Following treatment, 10 μL of Cell Counting Kit-8 (CCK-8; Solarbio, China) reagent was added to each well and incubated for 2 hours at 37°C. Cell viability was assessed by measuring absorbance at 450 nm using a microplate reader.

Quantitative Real-Time PCR (qRT-PCR)

Total RNA was isolated from cells using TRIzol reagent following the manufacturer's protocol. Complementary DNA (cDNA) synthesis was performed by reverse transcription using a commercial kit. The relative expression levels of genes were determined by the $2^{-\Delta\Delta CT}$ method, with GAPDH as the internal control. Primer sequences used were: DDAH2 (forward: 5′-GTCTCCACTGTGCCAGTCTC-3′; reverse: 5′-TCAGGGAGGCATATGGGTGA-3′); PNPLA2 (forward: 5′-CTCCAAGGACGAGCTCATCC-3′; reverse: 5′-GAAGGGGGACACTGTGATGG-3′); STXBP2 (forward: 5′-CACTGGCACAAGAACAAGGC-3′; reverse: 5′-CCAATGAGCACCTCCCACTT-3′); TCN1 (forward: 5′-GGGAACTTCTGAGTGGAGGC-3′; reverse: 5′-CTCCAGCGAACCTCCAAGTT-3′); GAPDH (forward: 5′-GGAGCGAGATCCCTCCAAAAT-3′; reverse: 5′-GGCTGTTGTCATACTTCTCATGG-3′).

Western Blot Assays

Proteins were extracted from cells using RIPA lysis buffer containing 1% protease and phosphatase inhibitor cocktails. The protein concentration of each sample was measured with the BCA Protein Assay Kit following the manufacturer's protocol. To achieve consistent protein loading, samples were mixed with suitable amounts of 4× loading buffer and additional RIPA buffer to equalize protein amounts across all samples. Proteins were separated by SDS-PAGE using gels of 7.5%, 10%, or 12.5% concentration according to their molecular sizes, then transferred onto 0.22 μm PVDF membranes. The membranes were blocked with 5% non-fat milk in TBST at room temperature for 2 hours, followed by overnight incubation at 4 °C with primary antibodies targeting DDAH2 (1:3000, 14966-1-AP, Proteintech), PNPLA2 (1:5000, 55190-1-AP, Proteintech), STXBP2 (1:1000, 15312-1-AP, Proteintech), and TCN1 (1:1000, 16078-1-AP, Proteintech). After washing, membranes were incubated with HRP-linked anti-rabbit secondary antibodies at room temperature for one hour. Protein signals were detected using an enhanced chemiluminescence (ECL) system. Band intensities were quantified using ImageJ software and normalized against β-actin.

Statistical Analysis

Data were analyzed using one-way ANOVA with GraphPad Prism 9.0, considering p-values less than 0.05 as statistically significant.

Results

Screening for DEGs

In this study, the GSE10474, GSE32707, and GSE66890 datasets were selected as the training cohort. Given the differences in sample sources and experimental conditions across these datasets, initial analyses revealed substantial baseline batch effects (Figure 2A), which could introduce systematic bias into downstream analyses. To address this issue, principal component analysis (PCA) was first applied to assess and correct inter-dataset variation. All raw

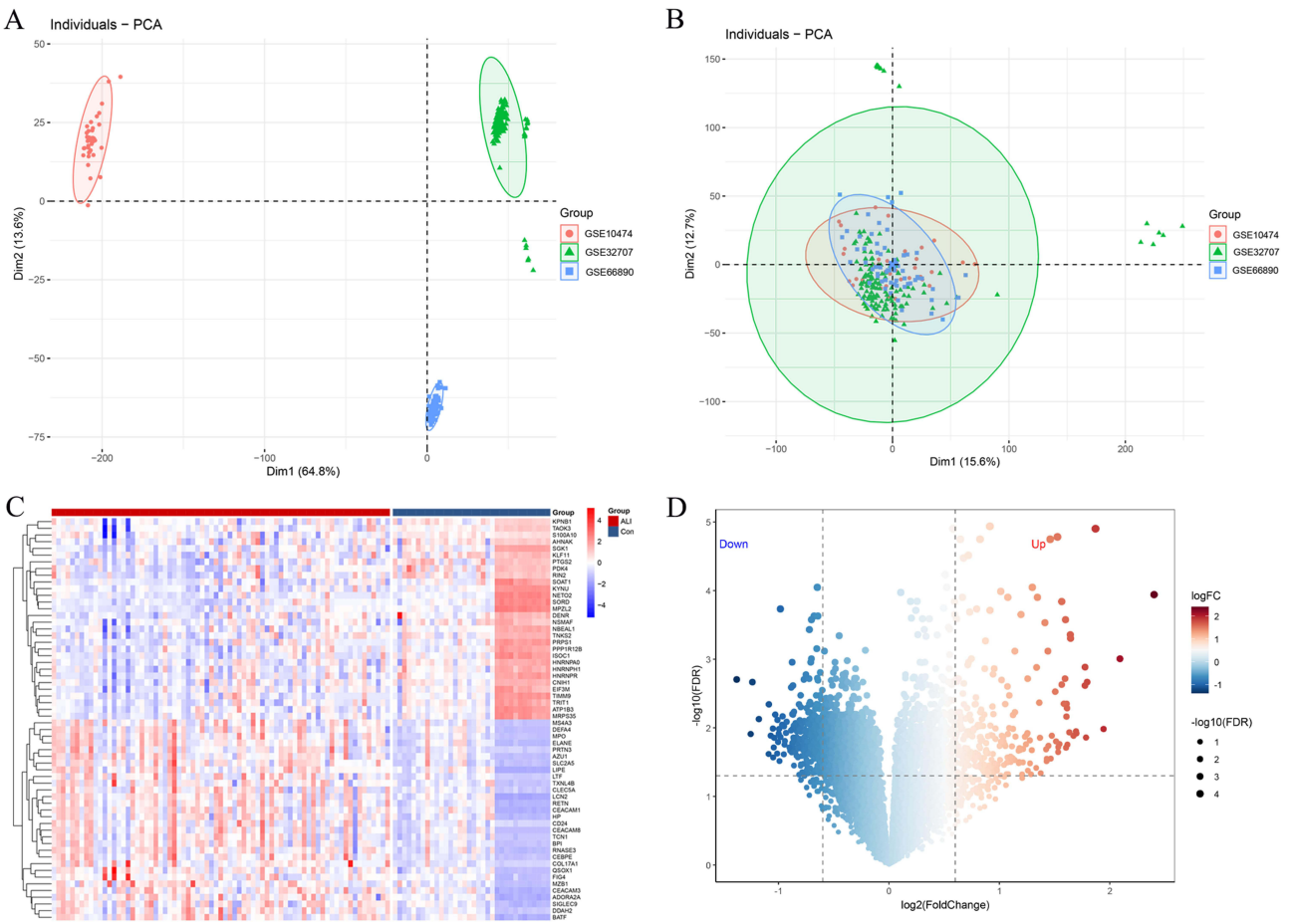

Figure 2 Identification of differentially expressed genes (DEGs). (**A**) Principal component analysis (PCA) of datasets GSE10474, GSE32707, and GSE66890 before batch effect correction. (**B**) PCA of the same datasets after batch effect correction. (**C**) Heatmap illustrating the expression profiles of DEGs. (**D**) Volcano plot depicting the distribution of DEGs.

microarray data were uniformly preprocessed using the Robust Multi-array Average (RMA) normalization method. PCA was repeated after normalization, revealing a significant reduction in batch effects and more coherent clustering of samples, indicating successful correction (Figure 2B). Subsequently, differential expression analysis was performed on the combined and normalized dataset using the *limma* package. Differentially expressed genes were defined based on a false discovery rate (FDR) < 0.05 and $|\log_2$ fold change$| > 0.6$. As a result, a total of 771 DEGs were identified, including 343 upregulated and 428 downregulated genes (Figure 2C and D). These DEGs provided a solid foundation for subsequent feature selection, pathway enrichment analysis, and mechanistic exploration.

WGCNA Analysis

To identify gene modules with extensive co-expression patterns, we applied WGCNA to the GSE32707 dataset. To achieve a scale-free network structure, the soft-thresholding power (β) was selected at the lowest value where the scale-free fit index neared 1, resulting in a final choice of β = 14 (Figure 3A and B). Hierarchical clustering combined with a dynamic tree-cutting algorithm identified four distinct co-expression modules (Figure 3C and D). Among them, the MEblue module demonstrated the strongest correlation with sepsis-induced ALI (R = 0.46, P = 1e−04). A scatter plot of module membership versus gene significance (GS) further underscored the robust association between the MEblue module and clinical phenotype (Figure 3E). Based on this, 1,616 genes from the MEblue module were chosen for further analysis.

https://doi.org/10.2147/JIR.S539899

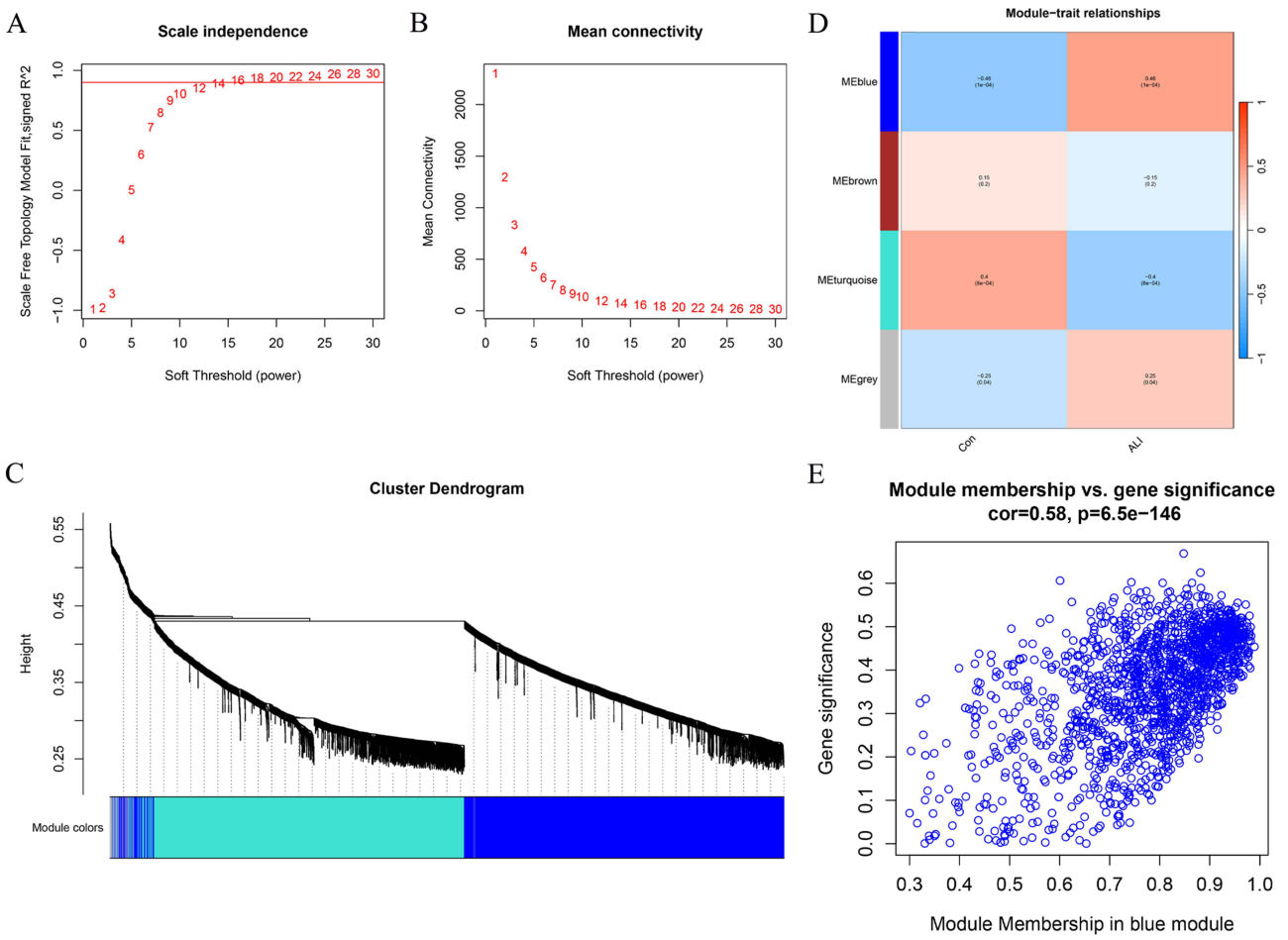

Figure 3 WGCNA analysis. (**A**) Analysis of the scale-free index for various soft-threshold powers (β). (**B**) Analysis of the mean connectivity for various soft-threshold powers. (**C**) WGCNA cluster dendrogram and module assignment. The branches refer to clusters of genes that are highly connected. The colors in the horizontal bar represent the modules. (**D**) The correlation between module eigengenes and sepsis-induced ALI is depicted as a heatmap, with each row showing the correlation and *P*-value. (**E**) The scatter plot of the blue module depicts the link between module membership and gene significance.

Underlying Biological Mechanisms

To further identify key genes implicated in sepsis-induced ALI, we intersected genes from the MEblue module identified by WGCNA with the DEGs, yielding 213 candidate genes for downstream functional annotation (Figure 4A). GO and KEGG enrichment analyses were performed to elucidate the potential molecular mechanisms underlying sepsis-related ALI. GO enrichment revealed significant overrepresentation of immune- and inflammation-related biological processes, including regulation of inflammatory response, leukocyte migration, response to bacterium-derived molecules, phagocytosis, response to lipopolysaccharide, and immune receptor activity (Figure 4B). KEGG analysis further indicated that these candidate genes were enriched in key inflammatory and immune signaling pathways, such as NF-κB signaling, NOD-like receptor signaling, natural killer cell-mediated cytotoxicity, Fc gamma receptor-mediated phagocytosis, and chemokine signaling pathways (Figure 4C). Together, these findings suggest that immune responses and bacteria-associated signaling cascades may play central regulatory roles in the development of sepsis-induced ALI.

Application of Machine Learning for Screening Hub Genes

To comprehensively evaluate the link between DEGs and sepsis-induced ALI pathogenesis, we utilized eight different machine learning methods for feature selection on the 213 DEGs. Predictive performance and error metrics for each algorithm were evaluated using boxplots of absolute residuals and the area under the receiver operating characteristic curve (AUC). Among all models, the GLM exhibited the poorest performance, displaying the largest prediction error and

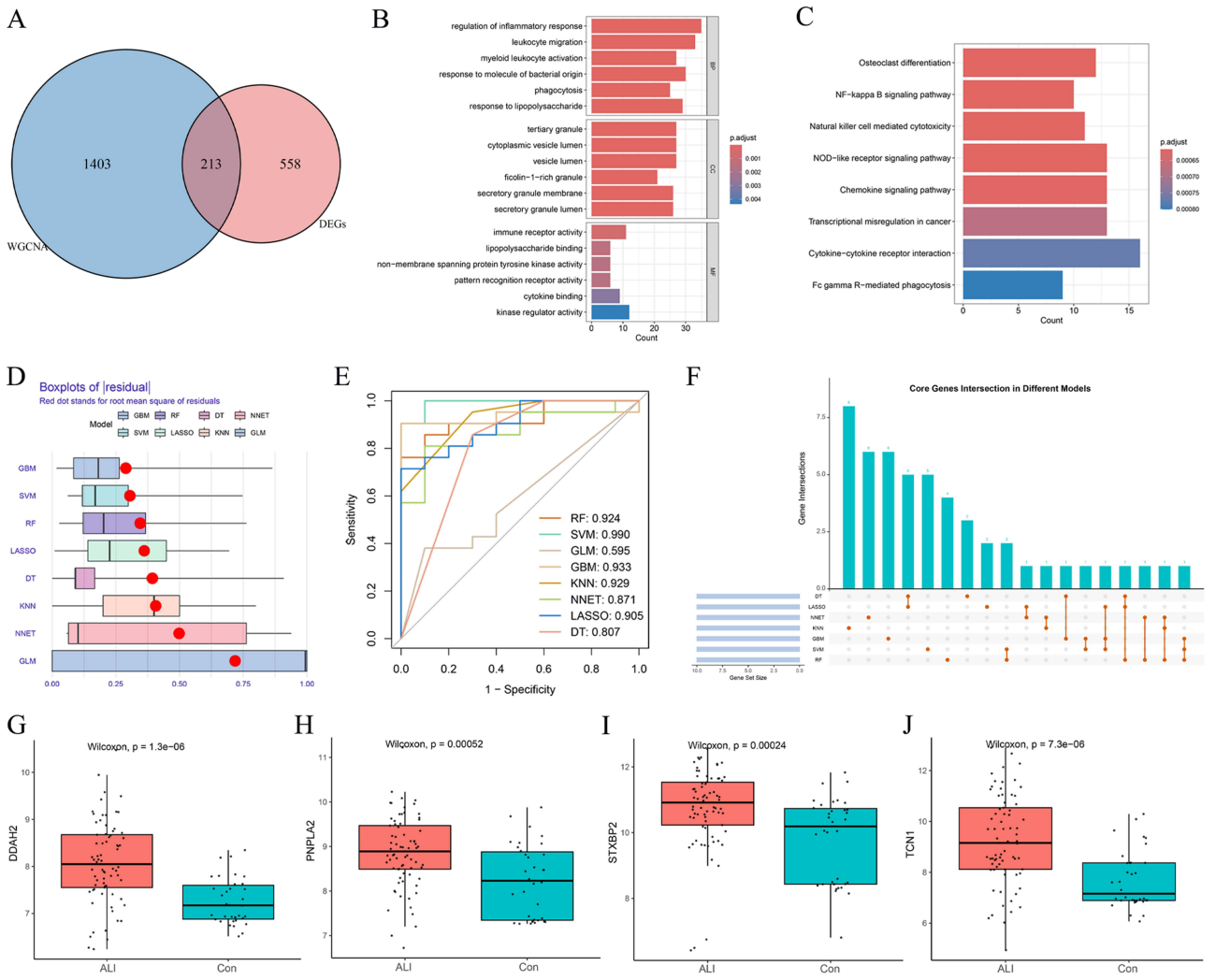

Figure 4 Application of machine learning for screening hub genes. (**A**) Intersection gene between module gene and DEGs in WGCNA analysis. (**B**) GO. (**C**) KEGG. (**D**) Boxplots of absolute residuals for eight machine learning algorithms, reflecting the prediction error of each model. (**E**) ROC curves illustrating the diagnostic performance of each machine learning algorithm. (**F**) Upset plot showing the intersection of genes identified by the top-performing models. Hub-DEGs were defined as genes selected by at least three different algorithms. (**G-J**) Differential expression of the four hub DEGs between sepsis-induced ALI patients and controls in the training cohort.

the lowest AUC value (0.595); therefore, it was excluded from subsequent analyses (Figure 4D–E). After removing GLM, we performed an intersection analysis among the other seven algorithms to determine consensus feature genes, defined as those selected by a minimum of three models. An Upset plot was created to illustrate the shared feature genes identified by each method (Figure 4F). This integrative approach ultimately yielded four hub DEGs strongly associated with sepsis-induced ALI: DDAH2, PNPLA2, STXBP2, and TCN1. Validation within the training cohort demonstrated that all four genes were markedly upregulated in sepsis-induced ALI patient samples relative to controls (Figure 4G–J), indicating these hub genes likely have critical functions in the development and progression of the disease.

Establishment of Diagnostic Prediction Model

To assess the diagnostic value of the key genes in sepsis-induced ALI, a nomogram model was constructed integrating the four hub genes (DDAH2, PNPLA2, TCN1, and STXBP2) (Figure 5A). In this model, each gene is assigned a score proportional to its expression level, and the cumulative score is used to estimate the individual probability of developing sepsis-associated ALI. Higher total scores indicate increased disease risk. Model performance was first assessed using a calibration curve, which demonstrated a strong concordance between predicted probabilities and actual outcomes, indicating excellent calibration and predictive accuracy (Figure 5B). Subsequently, DCA was conducted to evaluate the

https://doi.org/10.2147/JIR.S539899

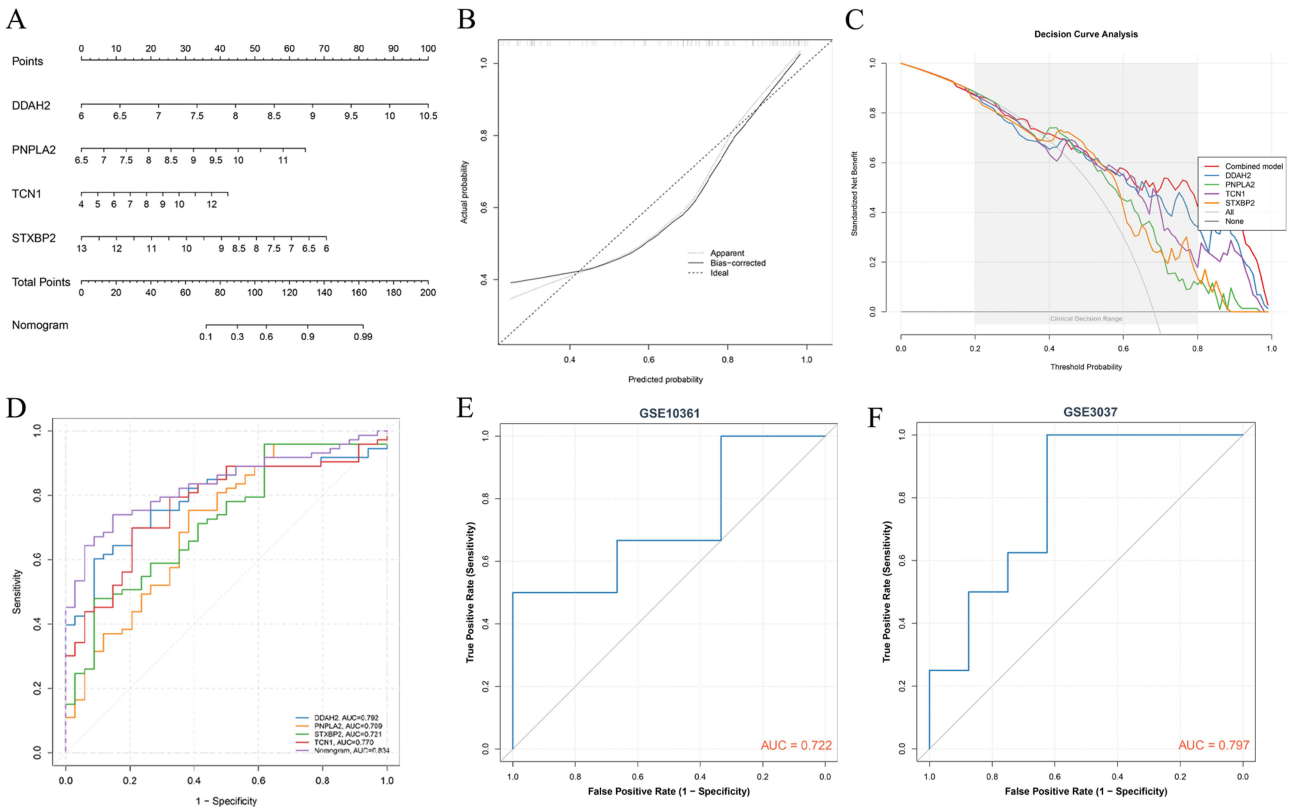

Figure 5 Establishment of diagnostic prediction model. (**A**) Nomogram to evaluate the risk of the occurrence of sepsis-induced ALI. (**B**) Calibration curves of the nomogram prediction. (**C**) DCA curves of the nomogram prediction. (**D**) ROC curves analysis of the feature genes in the training dataset. (**E** and **F**) ROC curves of the feature genes in the validation datasets (GSE10361 or GSE3037).

nomogram's clinical usefulness over various threshold probabilities. The nomogram consistently outperformed the gray line (representing net benefit of treating all or none), and the individual DCA curves for DDAH2, PNPLA2, STXBP2, and TCN1 showed that patients would derive tangible clinical benefit from the nomogram within the high-risk threshold range of 0.4 to 1.0. Notably, the overall net benefit of the nomogram surpassed that of models based on each single gene alone (Figure 5C). ROC curve analysis further supported the discriminative power of the model. The nomogram achieved an AUC of 0.834, while the individual genes also exhibited good diagnostic performance: DDAH2 (AUC = 0.792), PNPLA2 (AUC = 0.709), STXBP2 (AUC = 0.721), and TCN1 (AUC = 0.770) (Figure 5D). To assess the model's generalizability, we validated it in two independent external datasets (GSE10361 and GSE3037), yielding AUCs of 0.722 and 0.797, respectively (Figure 5E and F). These results underscore the robustness and clinical applicability of the diagnostic model for distinguishing sepsis-induced ALI patients from healthy individuals.

Infiltration of Immune Cells Results

CIBERSORT analysis demonstrated notable changes in immune cell infiltration patterns when comparing septic ALI samples to controls. Specifically, septic ALI tissues showed significantly increased proportions of plasma cells, naïve $CD4^+$ T cells, activated memory $CD4^+$ T cells, monocytes, M0 macrophages, resting mast cells, and neutrophils (Figure 6A and B). In contrast, naïve B cells, memory B cells, $CD8^+$ T cells, regulatory T cells (Tregs), resting NK cells, M2 macrophages, and activated dendritic cells were relatively reduced in the septic ALI group. Correlation analysis further revealed significant associations between the expression of DDAH2, PNPLA2, STXBP2, and TCN1 and the infiltration levels of various immune cell subsets (Figure 6C–F). Specifically, DDAH2 expression was positively correlated with M0 macrophages ($r = 0.27$, $P = 0.019$) and showed a borderline positive association with naïve $CD4^+$ T cells ($r = 0.22$, $P = 0.067$), while negatively correlating with $CD8^+$ T cells ($r = -0.30$, $P = 0.010$) and resting NK cells ($r = -0.27$, $P = 0.019$) (Figure 6C). PNPLA2 expression showed a positive correlation with Tregs ($r = 0.47$, $P < 0.001$) and M0 macrophages ($r = 0.35$, $P = 0.002$), while it was inversely

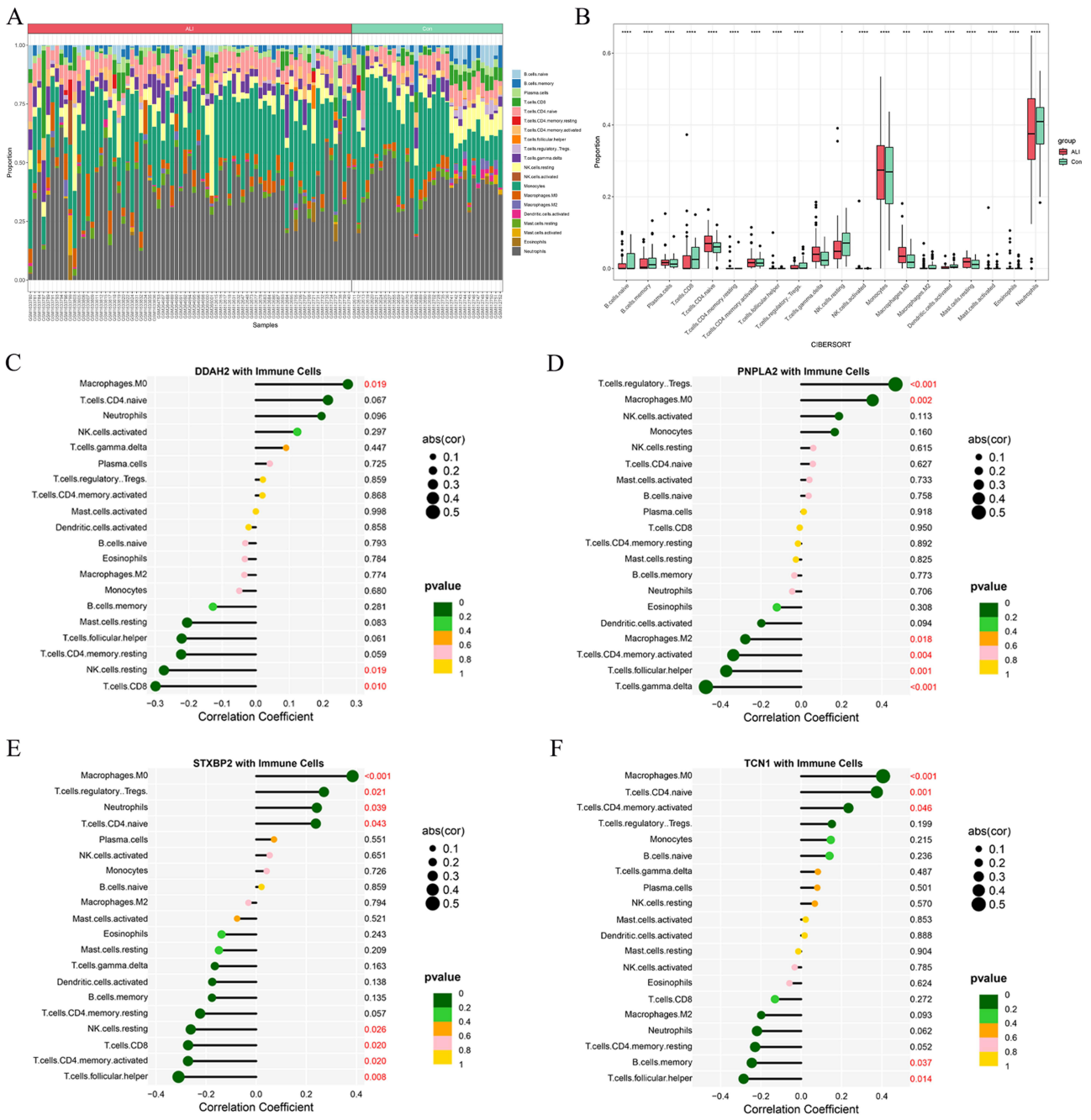

Figure 6 Immune cell infiltration in Normal and sepsis-induced ALI samples. (**A**) The percentage of the 22 immunocytes identified via the CIBERSORT arithmetic. (**B**) The diversities in the architecture of immunocytes between Normal and sepsis-induced ALI samples. (**C-F**) Correlation analysis between feature genes and immune cells. A larger circle shows a stronger correlation. A higher density of green color shows a more robust association.

correlated with γδ T cells (r = −0.47, P < 0.001) and follicular helper T (Tfh) cells (r = −0.37, P = 0.001) (Figure 6D). STXBP2 expression was positively linked to M0 macrophages (r = 0.38, P < 0.001) and Tregs (r = 0.27, P = 0.020), while showing negative correlations with Tfh cells (r = −0.31, P = 0.008) and activated memory $CD4^+$ T cells (r = −0.27, P = 0.020) (Figure 6E). Similarly, TCN1 expression correlated positively with M0 macrophages (r = 0.40, P < 0.001) and naïve $CD4^+$ T cells (r = 0.37, P = 0.001), but negatively with Tfh cells (r = −0.29, P = 0.014) and memory B cells (r = −0.25, P = 0.037) (Figure 6F). Overall, these results indicate that these hub genes may play important roles in septic ALI development by influencing immune cell infiltration and their functional landscape in the lung microenvironment.

https://doi.org/10.2147/JIR.S539899

Screening of Therapeutic Targets

To identify potential therapeutic compounds targeting the hub genes DDAH2, PNPLA2, STXBP2, and TCN1, we performed a drug–gene interaction analysis using the DSigDB database via the Enrichr platform. This analysis identified resveratrol as a promising candidate compound for the treatment of septic ALI. Resveratrol is a naturally occurring polyphenolic compound known for its anti-inflammatory, antioxidant, and anti-apoptotic properties, suggesting its therapeutic potential may be mediated through modulation of the expression or function of these hub genes. To further assess this potential, we conducted molecular docking analyses to evaluate the binding affinities between resveratrol and the four candidate targets. The docking scores were as follows: DDAH2, −8.0 kcal/mol; PNPLA2, −7.6 kcal/mol; STXBP2, −7.0 kcal/mol; and TCN1, −7.7 kcal/mol (Figure 7A–D). Among these, DDAH2 exhibited the strongest binding affinity (Figure 7A). Further examination of the interaction interfaces revealed that resveratrol engages in multiple non-covalent interactions with the target proteins, including conventional hydrogen bonds, carbon–hydrogen bonds, and pi–alkyl interactions (Figure 7A–D).

Molecular Dynamic Simulation Analyses

To assess the structural stability and conformational flexibility of the Resveratrol–DDAH2 complex at the atomic level, a 100-nanosecond all-atom molecular dynamics simulation (MDS) was performed using the GROMACS software suite. Post-simulation, trajectory analyses were conducted with built-in tools to calculate key structural and energetic parameters, including the root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), number of hydrogen bonds, radius of gyration (Rg), and free energy landscape. RMSD analysis revealed minimal global conformational fluctuations throughout the simulation, indicating that the Resveratrol–DDAH2 complex maintained a stable structural conformation over time (Figure 8A). RMSF profiling further demonstrated low flexibility across most amino acid residues, suggesting a high degree of rigidity within the protein backbone (Figure 8B). Hydrogen bond analysis identified persistent interactions between Resveratrol and DDAH2, which likely contribute to the overall stability of the complex (Figure 8C). The Rg, a metric used to assess the compactness of the protein structure during simulation, remained within a relatively narrow range, indicating that the complex retained its spatial integrity without significant expansion or collapse (Figure 8D). Moreover, the free energy landscape illustrated a well-defined and energetically favorable conformational basin, further supporting the stability of the complex (Figure 8E and F). Collectively, these molecular dynamics results suggest that the Resveratrol–DDAH2 complex exhibits robust conformational stability and favorable dynamic behavior over the simulated timescale.

Resveratrol Reversed the Expression of Target Genes to Alleviate LPS-Induced ALI

We next evaluated the protective effects of Resveratrol against LPS-induced injury in BEAS-2B epithelial cells. To establish an in vitro model of ALI, cells were exposed to increasing concentrations of LPS (0, 2.5, 5, and 10 μg/mL) for 24 hours. LPS treatment reduced cell viability in a dose-dependent manner, with a pronounced decline to 49.2% observed at 5 μg/mL (Figure 9A), which was subsequently selected as the optimal concentration for model induction. To assess the therapeutic potential of Resveratrol, cells were pretreated with 10 μM Resveratrol for 1 hour prior to LPS exposure. CCK-8 assays demonstrated that Resveratrol significantly restored cell viability following LPS stimulation (Figure 9B). Morphological examination revealed that LPS exposure induced pronounced cytopathic changes in BEAS-2B cells, including cellular shrinkage and reduced density, whereas Resveratrol pretreatment mitigated these effects and partially preserved cellular integrity (Figure 9C). We further investigated whether Resveratrol modulates the expression of the identified hub genes DDAH2, PNPLA2, STXBP2, and TCN1 under inflammatory stress. Both mRNA and protein levels of these genes were quantified by RT-qPCR and Western blotting, respectively. As expected, LPS stimulation significantly upregulated the expression of all four genes compared to untreated controls. Notably, pretreatment with Resveratrol markedly reversed these LPS-induced transcriptional and translational changes (Figure 9D–L), suggesting that Resveratrol may exert its protective effects, at least in part, through regulation of these targets.

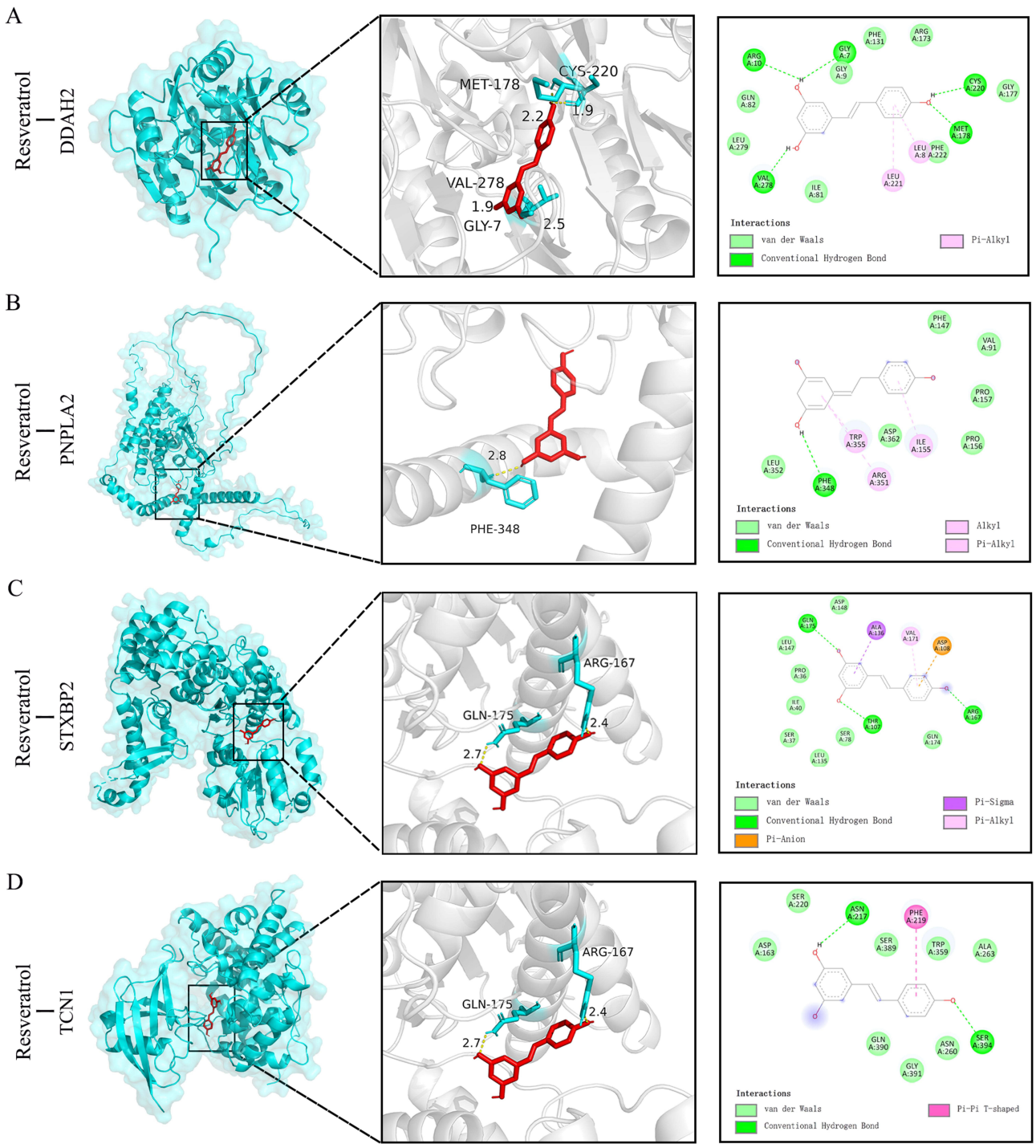

Figure 7 Two-and three-dimensional visualization of the binding interactions between resveratrol and four target proteins. (**A**) DDAH2, (**B**) PNPLA2, (**C**) STXBP2, (**D**) TCN1. Resveratrol is shown in red, and the target proteins are depicted in green. Binding affinities are indicated in kcal/mol.

Discussion

Sepsis is a major cause of mortality in intensive care units, with ALI being one of its most common and fatal complications, often progressing to multiorgan dysfunction. Despite its clinical significance, the pathogenesis of sepsis-induced ALI remains poorly understood, and effective diagnostic or therapeutic strategies are still lacking. Although several biomarkers have been proposed, early diagnosis remains challenging due to the disease's heterogeneity and

https://doi.org/10.2147/JIR.S539899

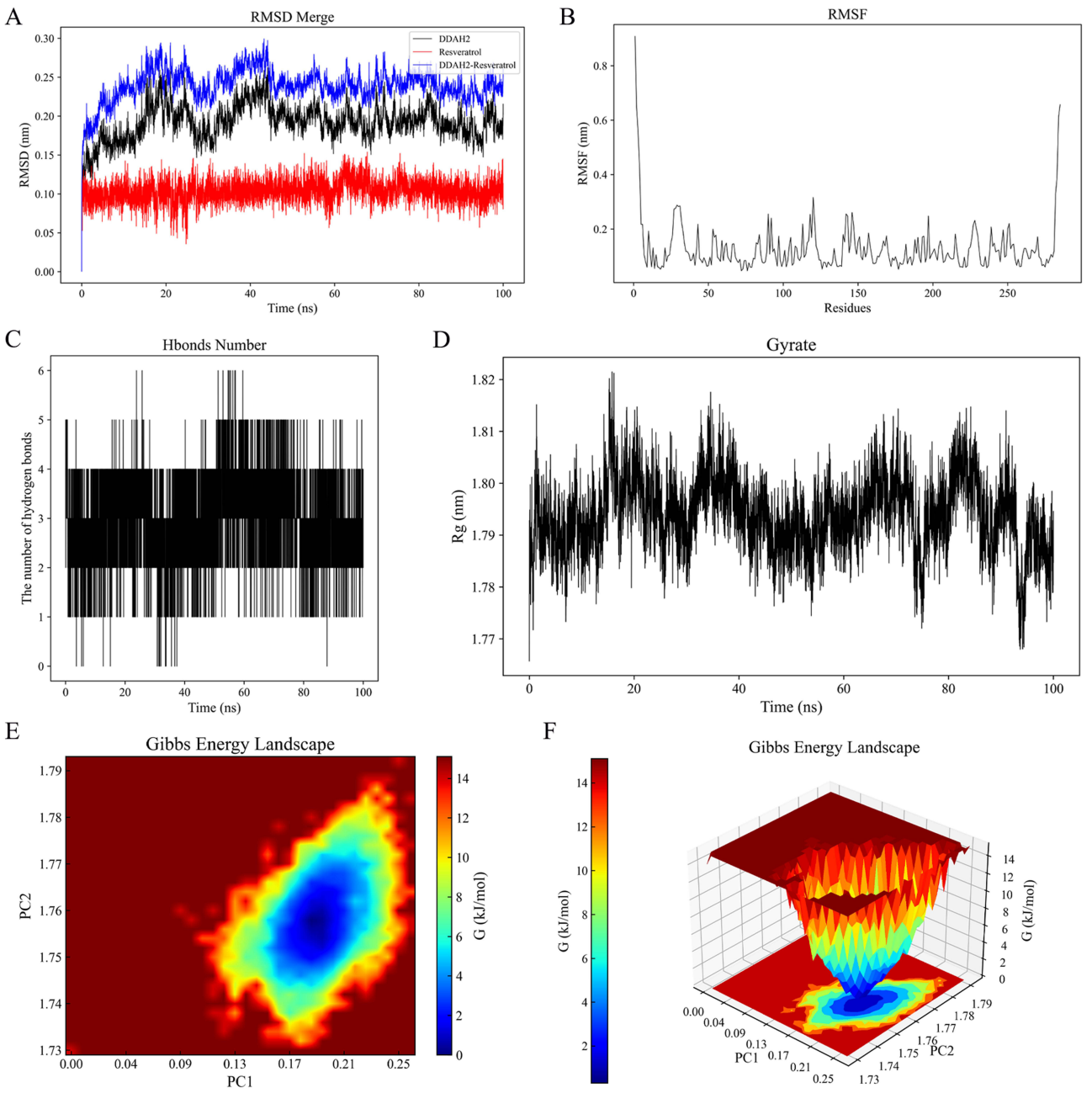

Figure 8 Results of molecular dynamics simulations (MDS) involving Resveratrol and DDAH2. (**A**) The RMSD values for DDAH2-Resveratrol complex. (**B**) Variations in protein flexibility throughout the Resveratrol simulation. (**C**) Dynamics of hydrogen bonding as observed in the molecular dynamics simulations. (**D**) Rg rate curve of the DDAH2-Resveratrol complex. (**E-F**) Two-dimensional and three-dimensional mappings of the free energy landscape.

Abbreviations: RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; ΔMMGBSA, Δ molecular mechanics-generalized-Born surface area; Rg, radius of gyration; PC, principal component.

nonspecific early symptoms.[23,24] Therefore, identifying reliable biomarkers and understanding the underlying immune mechanisms are critical for improving early diagnosis and treatment.

In recent years, the integration of bioinformatics with machine learning approaches has offered novel strategies for the identification of disease-associated biomarkers.[25,26] In this study, we combined these computational techniques to identify potential diagnostic markers for sepsis-induced ALI and systematically characterized the landscape of immune cell infiltration, thereby providing new insights into the immunosuppressive mechanisms underlying this condition. We integrated three publicly available gene expression datasets from the GEO database (GSE10474, GSE32707, and

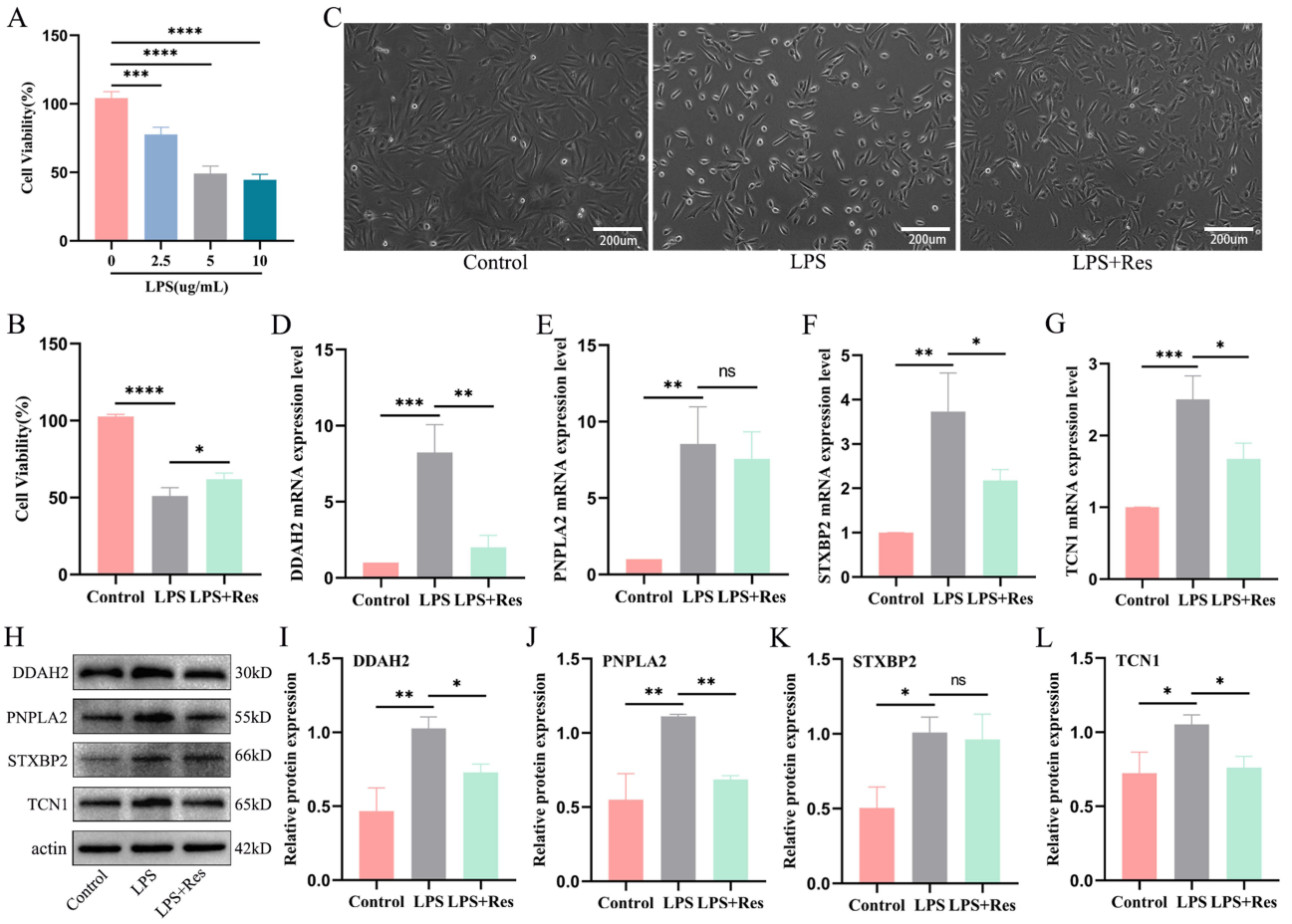

Figure 9 Resveratrol reversed the expression of target genes to alleviate LPS-induced ALI. (**A**) Cell activity of BEAS-2B cell lines treated with LPS for 24 h. (**B**) Cell activity of BEAS-2B cell lines treated with Resveratrol for 1 h. (**C**) The changes in cell morphology after LPS administration with or without pretreatment with Resveratrol at a concentration of 10 μM. (**D-G**) mRNA expression levels of DDAH2, PNPLA2, STXBP2, and TCN1 in BEAS-2B cells. (**H-L**) Expression of DDAH2, PNPLA2, STXBP2, and TCN1 proteins in BEAS-2B cells.

GSE66890) and applied batch effect correction prior to downstream analysis. A total of 771 DEGs were identified between sepsis-induced ALI and control samples. Intersecting DEGs with co-expression modules derived from WGCNA yielded 213 candidate genes. GO enrichment analysis revealed that these genes are predominantly involved in the regulation of inflammatory responses, leukocyte migration, bacterial pattern recognition, phagocytosis, lipopolysaccharide responses, and immune receptor activity. Additionally, KEGG pathway analysis demonstrated significant enrichment in immune-related pathways, including the NF-κB pathway, NOD-like receptor signaling, NK cell-mediated cytotoxicity, Fcγ receptor-mediated phagocytosis, and chemokine signaling.

Subsequently, using eight distinct machine learning algorithms, we further identified four candidate biomarkers closely associated with sepsis-induced ALI: DDAH2, PNPLA2, STXBP2, and TCN1. To our knowledge, this is the first study to experimentally validate the altered expression of these genes in an LPS-induced in vitro model of ALI. The experimental results were consistent with the bioinformatic predictions, and ROC curve analyses further demonstrated their potential diagnostic value. Among these, DDAH2, located in the major histocompatibility complex (MHC) class III region on chromosome 6, is the only isoform of DDAH expressed in immune cells.[27,28] Polymorphisms in the DDAH2 promoter have been linked to sepsis prognosis,[29,30] implicating its role in immune regulation. PNPLA2, also known as adipose triglyceride lipase (ATGL), is essential for triglyceride hydrolysis, yielding free fatty acids (FFAs) and glycerol. Although ATGL has traditionally been implicated in cold-induced thermogenesis through the activation of uncoupling protein 1 (UCP1) in brown adipose tissue, emerging evidence suggests that this function may not fully explain its biological role.[31] For instance, Jha et al[32] reported that adipose-specific ATGL deletion in mice reduces fat browning, lowers circulating FFAs, and

https://doi.org/10.2147/JIR.S539899

mitigates burn-induced hepatomegaly, indicating broader immunometabolic effects. STXBP2 is a core component of the platelet SNARE complex[33] and has been implicated in platelet activation, NETosis, and immunothrombosis during sepsis; however, its role in neutrophils and in the pathogenesis of ALI remains largely unexplored. TCN1 was previously shown to be significantly upregulated in neutrophil degranulation–related gene sets from bronchoalveolar lavage fluid of patients with severe influenza,[34] though its functional role in sepsis-induced ALI has not been fully elucidated.

A key pathological feature of sepsis-induced ALI is the excessive and dysregulated inflammatory response. In addition to the activation of innate immune cells, adaptive immune cells such as B cells and T cells infiltrate the lung in substantial numbers and actively contribute to disease progression. Notably, immune cell infiltration is increasingly recognized as a central driver of ALI pathogenesis in sepsis.[35,36] $CD4^+$ T cells are particularly susceptible to apoptosis during sepsis, leading to immune paralysis and secondary infections.[37,38] A Moreover, intestinal activation of $CD4^+$ T cells can reduce regulatory T cell (Treg) populations, thereby exacerbating systemic immune activation and epithelial barrier disruption.[39] Recurrent episodes of sepsis have been associated with $CD4^+$ T cell exhaustion, which correlates with poor clinical outcomes.[40] Other immune cell subsets may play context-dependent protective or deleterious roles. Thus, a comprehensive understanding of immune cell infiltration dynamics is critical for elucidating sepsis pathology and for identifying potential immunotherapeutic targets. In our study, we found that the expression of DDAH2, PNPLA2, STXBP2, and TCN1 was closely associated with distinct immune cell infiltration patterns, suggesting that these genes may modulate immune responses and influence the progression of ALI. Further preclinical and clinical studies are warranted to explore their mechanistic roles and validate their clinical utility.

To identify potential therapeutic compounds, we queried the DSigDB database and identified resveratrol as a small molecule capable of interacting with the four candidate biomarkers. Molecular docking and molecular dynamics simulations supported strong binding affinity between resveratrol and each target. Resveratrol (trans-3,5,4′-trihydroxystilbene), a naturally occurring polyphenolic compound found in grapes and berries, is well recognized for its potent antioxidant activity.[41–43] In addition to its antioxidative effects, resveratrol has demonstrated anti-inflammatory and anti-proliferative properties in diverse disease models.[44–46] Previous studies have shown that resveratrol alleviates pulmonary edema, airway inflammation, endotoxin-induced acute-phase responses, and joint degeneration. In our in vitro experiments, resveratrol pretreatment significantly improved the viability of LPS-treated BEAS-2B cells. Mechanistically, LPS exposure suppressed the mRNA and protein levels of DDAH2, PNPLA2, STXBP2, and TCN1, whereas resveratrol effectively reversed these alterations. These results suggest a potential cytoprotective role for resveratrol in mitigating LPS-induced lung epithelial injury. Nonetheless, further mechanistic studies are necessary to delineate the specific signaling pathways involved.

Conclusion

In this study, we systematically integrated multiple GEO datasets to identify differentially expressed genes associated with sepsis-induced ALI. By combining WGCNA with eight distinct machine learning algorithms, we identified four core diagnostic genes: DDAH2, PNPLA2, STXBP2, and TCN1. A diagnostic model constructed based on these genes exhibited robust predictive performance and effectively discriminated patients with sepsis-associated ALI from healthy controls. Subsequent drug screening suggested that Resveratrol may exert protective effects by targeting these four hub genes. Both molecular docking and in vitro cellular assays validated its therapeutic potential. Together, our findings provide a novel molecular framework for early diagnosis and targeted intervention in sepsis-induced ALI. Nonetheless, further clinical and mechanistic investigations are essential to confirm and extend these observations.

Data Sharing Statement

The data underlying this study's conclusions can be obtained from the corresponding author upon reasonable request. This aligns with the principles of open science, enabling independent verification and facilitating future research by others in the field.

Ethics Statement

This study used only publicly available, de-identified human gene expression data from the GEO database, with prior ethical approval and informed consent obtained by the original investigators. In accordance with item 1 and item 2 of

Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects (February 18, 2023, China), this research is exempt from Institutional Review Board (IRB) or ethics committee review.

Acknowledgments

The authors extend sincere appreciation to all team members whose commitment and effort were instrumental to the success of this study. We are grateful to the staff and researchers involved in data acquisition, analysis, and interpretation. Special thanks also go to the reviewers for their insightful feedback, which significantly enhanced the quality of the manuscript. We acknowledge the financial support provided by relevant funding agencies.

Author Contributions

YNL was responsible for the execution of the study, conducting experiments, and drafting the manuscript. YNL and JX jointly participated in study design, project supervision, and key experimental work. NNH handled data preprocessing, statistical analysis, and interpretation of results. WC contributed by critically revising the manuscript for important intellectual content and polishing the language to ensure scientific rigor and coherence. All authors were involved in the discussion of the manuscript, reviewed and approved the final version, and agreed to be accountable for all aspects of the work.

Funding

This study was funded by the Natural Science Foundation of Gansu Province Project (25JRRA1026) and the Cuiying Scientific and Technological Innovation Program of Lanzhou University Second Hospital (Project No. CY2023-QN-B10)
.

Disclosure

The authors declare that they have no competing interests.

References

1. Angus DC, Van Der Poll T. Severe sepsis and septic shock. *N Engl J Med*. 2013;369(9):840–851. doi:10.1056/NEJMra1208623
2. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801–810. doi:10.1001/jama.2016.0287
3. Vandini S, Bottau P, Faldella G, et al. Immunological, Viral, Environmental, and Individual Factors Modulating Lung Immune Response to Respiratory Syncytial Virus. *Biomed Res Int*. 2015;2015:875723. doi:10.1155/2015/875723
4. Bellani G, Laffey JG, Pham T, 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
5. Sheu CC, Gong MN, Zhai R, et al. Clinical characteristics and outcomes of sepsis-related vs non-sepsis-related ARDS. *Chest*. 2010;138 (3):559–567. doi:10.1378/chest.09-2933
6. Toh TS, Dondelinger F, Wang D. Looking beyond the hype: applied AI and machine learning in translational medicine. *EBioMedicine*. 2019;47:607–615. doi:10.1016/j.ebiom.2019.08.027
7. Hira ZM, Gillies DF. A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Adv Bioinformatics*. 2015;2015:198363. doi:10.1155/2015/198363
8. Schrider DR, Kern AD. Supervised Machine Learning for Population Genetics: a New Paradigm. *Trends Genet*. 2018;34(4):301–312. doi:10.1016/j.tig.2017.12.005
9. Ming T, Dong M, Song X, et al. Integrated Analysis of Gene Co-Expression Network and Prediction Model Indicates Immune-Related Roles of the Identified Biomarkers in Sepsis and Sepsis-Induced Acute Respiratory Distress Syndrome. *Front Immunol*. 2022;13:897390. doi:10.3389/fimmu.2022.897390
10. Delano MJ, Ward PA. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol Rev*. 2016;274(1):330–353. doi:10.1111/imr.12499
11. Wu MJ, Gao YL, Liu JX, et al. Integrative Hypergraph Regularization Principal Component Analysis for Sample Clustering and Co-Expression Genes Network Analysis on Multi-Omics Data. *IEEE J Biomed Health Inform*. 2020;24(6):1823–1834. doi:10.1109/JBHI.2019.2948456
12. Lenz M, Müller FJ, Zenke M, et al. Principal components analysis and the reported low intrinsic dimensionality of gene expression microarray data. *Sci Rep*. 2016;6(1):25696. doi:10.1038/srep25696
13. A AS, Balhoff J, Carbon S, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1). doi:10.1093/genetics/iyad031.
14. 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
15. Lex A, Gehlenborg N, Strobelt H, et al. UpSet: visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph*. 2014;20(12):1983–1992. doi:10.1109/TVCG.2014.2346248
16. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12(1):77. doi:10.1186/1471-2105-12-77
17. Chen B, m S K, Liu CL, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243–259.

18. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785–2791. doi:10.1002/jcc.21256
19. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–461. doi:10.1002/jcc.21334
20. Eberhardt J, Santos-Martins D, Tillack AF, et al. AutoDock Vina 1.2.0: new Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model*. 2021;61(8):3891–3898. doi:10.1021/acs.jcim.1c00203
21. Mo S, Yi Q, Bei X, et al. INTERFERING HSA_CIRC_0001226 MITIGATES LPS-INDUCED BEAS-2B CELL INJURY BY REGULATING MIR-940/TGFBR2 PATHWAY. *Shock*. 2023;60(4):565–572. doi:10.1097/SHK.0000000000002196
22. Wang S, Gong L, Mo Y, et al. Resveratrol attenuates inflammation and apoptosis through alleviating endoplasmic reticulum stress via Akt/mTOR pathway in fungus-induced allergic airways inflammation. *Int Immunopharmacol*. 2022;103:108489. doi:10.1016/j.intimp.2021.108489
23. Pierrakos C, Velissaris D, Bisdorff M, et al. Biomarkers of sepsis: time for a reappraisal. *Crit Care*. 2020;24(1):287. doi:10.1186/s13054-020-02993-5
24. Barichello T, Generoso J, Singer M, et al. Biomarkers for sepsis: more than just fever and leukocytosis-a narrative review. *Crit Care*. 2022;26(1):14.
25. Meng XW, Cheng ZL, Lu ZY, et al. MX2: identification and systematic mechanistic analysis of a novel immune-related biomarker for systemic lupus erythematosus. *Front Immunol*. 2022;13:978851. doi:10.3389/fimmu.2022.978851
26. Zhang C, Dong N, Xu S, et al. Identification of hub genes and construction of diagnostic nomogram model in schizophrenia. *Front Aging Neurosci*. 2022;14:1032917. doi:10.3389/fnagi.2022.1032917
27. Leiper JM. The DDAH-ADMA-NOS pathway. *Ther Drug Monit*. 2005;27(6):744–746. doi:10.1097/01.ftd.0000179849.42395.11
28. Tran CT, Fox MF, Vallance P, et al. Chromosomal localization, gene structure, and expression pattern of DDAH1: comparison with DDAH2 and implications for evolutionary origins. *Genomics*. 2000;68(1):101–105. doi:10.1006/geno.2000.6262
29. O'dwyer MJ, Dempsey F, Crowley V, et al. Septic shock is correlated with asymmetrical dimethyl arginine levels, which may be influenced by a polymorphism in the dimethylarginine dimethylaminohydrolase II gene: a prospective observational study. *Crit Care*. 2006;10(5):R139. doi:10.1186/cc5053
30. Weiss SL, Yu M, Jennings L, et al. Pilot study of the association of the DDAH2 −449G polymorphism with asymmetric dimethylarginine and hemodynamic shock in pediatric sepsis. *PLoS One*. 2012;7(3):e33355. doi:10.1371/journal.pone.0033355
31. Schreiber R, Diwoky C, Schoiswohl G, et al. Cold-Induced Thermogenesis Depends on ATGL-Mediated Lipolysis in Cardiac Muscle, but Not Brown Adipose Tissue. *Cell Metab*. 2017;26(5):753–763.e7. doi:10.1016/j.cmet.2017.09.004
32. Jha P, Claudel T, Baghdasaryan A, et al. Role of adipose triglyceride lipase (PNPLA2) in protection from hepatic inflammation in mouse models of steatohepatitis and endotoxemia. *Hepatology*. 2014;59(3):858–869. doi:10.1002/hep.26732
33. Yang M, Jiang H, Ding C, et al. STING activation in platelets aggravates septic thrombosis by enhancing platelet activation and granule secretion. *Immunity*. 2023;56(5):1013–1026.e6. doi:10.1016/j.immuni.2023.02.015
34. Dunning J, Blankley S, Hoang LT, et al. Author Correction: progression of whole-blood transcriptional signatures from interferon-induced to neutrophil-associated patterns in severe influenza. *Nat Immunol*. 2019;20(3):373. doi:10.1038/s41590-019-0328-y
35. Duan S, Jiao Y, Wang J, et al. Impaired B-Cell Maturation Contributes to Reduced B Cell Numbers and Poor Prognosis in Sepsis. *Shock*. 2020;54(1):70–77. doi:10.1097/SHK.0000000000001478
36. Lee KA, Gong MN. Pre-B-cell colony-enhancing factor and its clinical correlates with acute lung injury and sepsis. *Chest*. 2011;140(2):382–390. doi:10.1378/chest.10-3100
37. Martin MD, Badovinac VP, Griffith TS. CD4 T Cell Responses and the Sepsis-Induced Immunoparalysis State. *Front Immunol*. 2020;11:1364. doi:10.3389/fimmu.2020.01364
38. Brady J, Horie S, Laffey JG. Role of the adaptive immune response in sepsis. *Intensive Care Med Exp*. 2020;8(Suppl 1):20. doi:10.1186/s40635-020-00309-z
39. Assimakopoulos SF, Eleftheriotis G, Lagadinou M, et al. SARS CoV-2-Induced Viral Sepsis: the Role of Gut Barrier Dysfunction. *Microorganisms*. 2022;10(5):1050. doi:10.3390/microorganisms10051050
40. He W, Xiao K, Xu J, et al. Recurrent Sepsis Exacerbates CD4(+) T Cell Exhaustion and Decreases Antiviral Immune Responses. *Front Immunol*. 2021;12:627435. doi:10.3389/fimmu.2021.627435
41. Pervaiz S, Holme AL. Resveratrol: its biologic targets and functional activity. *Antioxid Redox Signal*. 2009;11(11):2851–2897. doi:10.1089/ars.2008.2412
42. Pervaiz S. Resveratrol: from grapevines to mammalian biology. *FASEB j*. 2003;17(14):1975–1985. doi:10.1096/fj.03-0168rev
43. Baur JA, Sinclair DA. Therapeutic potential of resveratrol: the in vivo evidence. *Nat Rev Drug Discov*. 2006;5(6):493–506. doi:10.1038/nrd2060
44. Berman AY, Motechin RA, Wiesenfeld MY, et al. The therapeutic potential of resveratrol: a review of clinical trials. *NPJ Precis Oncol*. 2017;1:1. doi:10.1038/s41698-017-0038-6
45. Meng T, Xiao D, Muhammed A, et al. Anti-Inflammatory Action and Mechanisms of Resveratrol. *Molecules*. 2021;26(1):229. doi:10.3390/molecules26010229
46. Clément MV, Hirpara JL, Chawdhury SH, et al. Chemopreventive agent resveratrol, a natural product derived from grapes, triggers CD95 signaling-dependent apoptosis in human tumor cells. *Blood*. 1998;92(3):996–1002. doi:10.1182/blood.V92.3.996