Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Molecular Mechanisms of Feizhangning Targeting Programmed Cell Death-Related Genes in COPD Intervention: A Multimodal Mendelian Randomization and Multi-Omics Integration Study

Cheng Luo
Chengdu University of Traditional Chinese Medicine
Xiuzhi Peng
Chengdu University of Traditional Chinese Medicine
Bo Ning
Guangdong Provincial Hospital of Chinese Medicine, the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, University of Chinese Medicine
Jiayu Zhang
Chengdu University of Traditional Chinese Medicine
Ziyi Jiang
Chengdu University of Traditional Chinese Medicine
Junwen Tan
Chengdu University of Traditional Chinese Medicine
Jiajie Li
Yunnan University of Chinese Medicine
Fei Wang
Chengdu University of Traditional Chinese Medicine
Ying Ye
Chengdu University of Traditional Chinese Medicine
Yuanhang Ye
yyh421492693@outlook.com
Chengdu University of Traditional Chinese Medicine
Jia Ke
Hubei Provincial Hospital of Traditional Chinese Medicine
Wanting Qin
Chengdu University of Traditional Chinese Medicine

Research Article

Keywords: Chronic obstructive pulmonary disease, Programmed cell death, Mendelian randomization, Single-cell transcriptome, Feizhangning, Network pharmacology, LGMN, NT5E

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11650241/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Molecular Mechanisms of Feizhangning Targeting Programmed Cell Death-Related Genes in COPD Intervention: A Multimodal Mendelian Randomization and Multi-Omics Integration Study

Cheng Luo[a,b,c,1], Xiuzhi Peng[a,b,1], Bo Ning[d,*], Jiayu Zhang[a,1], Ziyi Jiang[a,b], Junwen Tan[a,b], Jiajie Li[e], Fei Wang[a,b], Ying Ye [a], Yuanhang Ye[a,f,1], Jia Ke[g,h,i,j,*], Wanting Qin[a,b,*]

[a]Chengdu University of Traditional Chinese Medicine, Chengdu, 610075, China;
[b]Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, 610075, China;
[c]China Joint Graduate School of Traditional Chinese Medicine, Suzhou, 215105, China;
[d]Department of Cardiovascular Surgery, Guangdong Provincial Hospital of Chinese Medicine, the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, the Second Clinical College of Guangzhou University of Chinese Medicine, Guangzhou, 510120, China;
[e]Yunnan University of Chinese Medicine, Kunming, 650500, China;
[f]Pengzhou Third People's Hospital, Chengdu, 611931, China;
[g]Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan, 430061, China;
[h]Hubei Key Laboratory of theory and application research of liver and kidney in traditional Chinese medicine, Affiliated Hospital of Hubei University of Chinese Medicine, Wuhan, 430061, China;
[i]Hubei Shizhen Laboratory, Wuhan, 430074, China;
[J]Hubei Province Academy of Traditional Chinese Medicine, Wuhan, 430074, China
[*]**Corresponding author: Yuanhang Ye, Jia Ke, Wanting Qin**
[*]*E-mail addresse*s: Yuanhang Ye (yyh421492693@outlook.com), kejia@hbhtcm.com (Jia Ke), m18974400757@163.com (Wanting Qin)
[1] These authors contributed equally: Cheng Luo (lc1121174213@outlook.com), Xiuzhi Peng (1915211753@qq.com), Bo Ning (ningbo9803@163.com)

Abstract:

Objective: Chronic obstructive pulmonary disease (COPD) is the third leading cause of respiratory disease-related mortality worldwide, characterized by persistent, irreversible airflow limitation, chronic airway inflammation, and pulmonary tissue remodeling. Dysregulation of programmed cell death (PCD) is a significant driving factor mediating sustained lung injury, but the genetic causal evidence linking PCD-related genes to COPD remains insufficient. The multi-target molecular mechanisms of the traditional Chinese medicine compound Feizhangning (FZN) in the intervention of COPD still need to be elucidated. This study aims to identify the causal core genes related to PCD in COPD and to elucidate the mechanism by which FZN improves COPD through the regulation of the PCD pathway.

Methods: Using a multi-omics integration strategy, we employed two-sample Mendelian randomization (MR), summary-data-based Mendelian randomization (SMR) analysis, multi-cohort transcriptomics, WGCNA co-expression analysis, various machine learning algorithms, single-cell transcriptomics, cell communication analysis, scTenifoldKnk/scTenifoldNet's LGMN virtual bidirectional perturbation, network pharmacology,

molecular docking, and molecular dynamics simulations, alongside validation using an in vivo model of COPD in rats.

Results: A total of 439 PCD-related genes with robust genetic causal associations with COPD were screened. Through multi-dimensional intersection analysis, 16 core candidate genes were identified, and further screening identified 5 optimal diagnostic markers: NT5E, SPP1, BLK, PLAT, and LGMN. Single-cell analysis confirmed that COPD lung cell subpopulations underwent significant remodeling, with five core genes exhibiting cell type-specific expression. Further cellular communication analysis indicated that macrophages, epithelial cells, and fibroblasts serve as communication hubs within the pulmonary microenvironment. Virtual perturbation experiments demonstrated that LGMN played a bidirectional regulatory role: LGMN deficiency activated the NF-κB/TNF/IL-17 pro-inflammatory pathway, while LGMN overexpression activated pathways related to pulmonary lipid homeostasis and barrier repair. FZN contained 27 active components that could stably bind to five target proteins, with Tanshinone IIA and Dihydrotanshinone exhibiting the strongest binding affinity to NT5E. Animal experiments showed that FZN could dose-dependently reduce the levels of inflammatory factors IL-1β, IL-6, and TNF-α in rat serum and bronchoalveolar lavage fluid (BALF), and reverse the abnormal expression of the five core genes at both mRNA and protein levels.

Conclusion: This study establishes a multi-omics integration research system, confirming that FZN inhibits excessive PCD, alleviates chronic inflammatory responses, and restores alveolar structural homeostasis by targeting core PCD genes, thereby improving COPD and providing a theoretical basis for the development of COPD diagnostic markers and the translational application of FZN.

Keywords: Chronic obstructive pulmonary disease; Programmed cell death; Mendelian randomization; Single-cell transcriptome; Feizhangning; Network pharmacology; LGMN; NT5E

Introduction:

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory lung disease characterized by persistent and irreversible airflow limitation. It has become the third leading cause of death globally, with over 400 million affected people, posing a severe public health challenge[1]. Its pathological changes involve chronic airway inflammation, airway remodeling, destruction of alveolar structures, and the formation of emphysema, with clinical manifestations primarily including chronic cough, sputum production, and dyspnea after exertion[2]. According to the Global Burden of Disease Study 2021 (GBD 2021), there are approximately 213 million COPD patients worldwide, with the highest prevalence among men. The number of deaths has increased from 2.5 million in 1990 to 3.72 million in 2021, and the disability-adjusted life years (DALY) rate has increased from 926.1 per 100,000 in 2019 to 940.7 per 100,000 in 2021[3]. The onset of COPD is associated with factors such as age, smoking, air pollution, body mass index, and long-term exposure to dust, although the specific mechanisms remain not fully elucidated[4].Currently, clinical treatment primarily relies on symptomatic relief with bronchodilators, inhaled corticosteroids, and oxygen therapy, which can only alleviate symptoms without preventing or reversing structural damage to lung tissue. Therefore, exploring the core pathways and genetic targets driving chronic lung injury is of significant clinical value for the development of new therapeutic strategies.

Programmed Cell Death (PCD) is a form of cell death that is strictly regulated by genes, covering various subtypes such as apoptosis, pyroptosis, ferroptosis, necroptosis, autophagy, and

cuproptosis. It serves as the biological foundation for maintaining normal lung development, epithelial homeostasis, and immune balance[5-7]. Abnormal activation of PCD can lead to persistent damage and death of airway epithelial cells, alveolar epithelial cells, vascular endothelial cells, and immune cells, triggering a persistent inflammatory cascade which accelerates the progression of COPD[8, 9]. Basic research has confirmed that long-term exposure to cigarette smoke (CS) can activate PCD pathways such as apoptosis, autophagy, and ferroptosis, thereby impairing the regenerative capacity of type II alveolar epithelial cells and worsening lung tissue damage[10, 11]. COPD is not mediated by a single type of cell death, but rather is a complex phenotype shaped by multiple factors. Under different stimuli, death programs initiated by different cell types have varying impacts on inflammation and the immune system, and may even antagonize each other[12].

The traditional Chinese medicine (TCM) compound, characterized by multi-component, multi-pathway, and multi-target synergistic regulation, can intervene in the pathological progression of COPD by regulating PCD pathways[13]. Feizhangning (FZN) is a clinical experience compound developed by Professor Wang Fei from the Affiliated Hospital of Chengdu University of Traditional Chinese Medicine for the treatment of COPD. It consists of 16 traditional Chinese medicinal herbs, including Huangqi *(Astragalus),* Dangshen *(Codonopsis),* Sangbaipi *(Mori Cortex),* Huangqin *(Scutellaria),* Danshen *(Salvia),* and so on. The combination of various medicinal herbs takes into account both the symptoms and the root causes, alleviating airway obstruction while also regulating the body's internal environment to treat the underlying issues. Modern pharmacological research indicates that the active components of herbs such as Huangqi, Sangbaipi, Huangqin, and Danshen can improve chronic inflammation, oxidative stress, and tissue remodeling in COPD by regulating PCD[14-17]. However, the specific molecular targets of FZN in regulating PCD and the causal evidence at the genetic level remain unclear and require further elucidation from a systems biology perspective.

To elucidate the genetic causal roles and functional mechanisms of PCD-related genes regulated by FZN in the pathogenesis and progression of COPD, this study constructs a multimodal integrative analysis system that integrates human whole-genome genetic data, multi-cohort lung tissue transcriptomic data, and single-cell transcriptomic data related to COPD. Through the two-sample Mendelian randomization (MR) combined with summary-data-based Mendelian randomization (SMR) analysis, this study screened for PCD-related genes that had a genetic causal association with COPD. And by integrating Weighted Gene Co-expression Network Analysis (WGCNA) co-expression analysis, machine learning, and Shapley Additive Explanations (SHAP) interpretability analysis, we analyzed and identified core genes associated with COPD and developed a risk prediction model. Utilizing single-cell transcriptomics and cell communication, this study analyzed the characteristics of core gene cell-specific expression and interactions within the pulmonary microenvironment. Additionally, it elucidates the regulatory functions through LGMN virtual bidirectional perturbation experiments. Furthermore, network pharmacology, molecular docking, and molecular dynamics simulations were used to analyze the targeted binding characteristics of the active components of FZN. Finally, in vivo experiments in COPD rat models were conducted to validate the mechanisms. This study deepens the understanding of the molecular etiology of COPD from the perspective of PCD, uncovers key molecular targets for intervention, and elucidates the multi-target action mechanisms of FZN in the treatment of COPD, providing theoretical and experimental support for early screening and

drug translation in COPD.

1 Materials and Methods

1.1 Data Sources

1.1.1 PCD-Related Gene Set Data

We integrated items of PCD-related genes from recently published studies, resulting in a total of 3079 unique PCD-related genes after merging and deduplication[18-22].

1.1.2 COPD Transcriptome Data

We retrieved lung tissue COPD transcriptome cohorts from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), and selected a total of five independent datasets: GSE13932, GSE21359, GSE162635, GSE212331, and GSE324136. The datasets GSE13932 (29 cases of COPD / 69 controls) and GSE21359 (23 cases of COPD / 12 controls) were integrated to form the training set, and the remaining three cohorts (a total of 632 cases of COPD / 92 controls) served as an external independent validation set.

1.1.3 Mendelian Randomization Data for COPD

The summary data for two-sample MR on cis-expression quantitative trait loci (cis-eQTL) was obtained from the eQTLGen Consortium. The data for cis-protein quantitative trait loci (cis-pQTL) was integrated from publicly available datasets of the Finnish Biobank and the UK Biobank Pharma Proteomics Project (UKB-PPP). The SMR used blood cis-eQTL from eQTLGen as the exposure data source, and the genome-wide association study (GWAS) summary data for COPD phenotypes was sourced from the Medical Research Council Integrative Epidemiology Unit (MRC IEU) OpenGWAS database (ebi-a-GCST90018807), which included 13,530 COPD cases and 454,945 population controls.

1.2 Genetic Instrumental Variable Selection Criteria

Candidate genetic instrumental variables (IVs) were derived from the cis-eQTL/cis-pQTL loci of the PCD gene. The MR analysis utilized the cis-eQTL predicted gene expression levels and cis-pQTL predicted protein abundance as exposure phenotypes and selected single nucleotide polymorphism (SNP) loci that are significantly associated with the aforementioned quantitative molecular traits as genetic IVs. The cis-eQTL data were first screened for significant loci using a false discovery rate (FDR) of < 0.05, and then retained SNPs with genome-wide significance at $P < 5\times10^{-8}$. For the cis-pQTL data, the genome-wide threshold of $P < 5\times10^{-8}$ was applied directly for screening. Based on the reference data from the third phase of the 1,000 Genomes Project (v5) for the European population, we performed linkage disequilibrium (LD) clustering and filtering using a 1,000 kb LD window and the correlation coefficient $r^2 < 0.001$ to obtain the final IVs.

1.3 Two-Sample MR Analysis

Using the TwoSampleMR method, we analyzed the potential causal relationship between the genetic expression characteristics of the PCD gene as the exposure and COPD as the outcome. The core effect model employed the inverse-variance weighted (IVW) method, supplemented by MR-Egger regression, the weighted median method, and the weighted mode method to validate the robustness of the results. And the causal effect was quantified using odds ratios (OR) and 95%

confidence intervals (CI). The criteria for evaluation were that a potential causal relationship between the gene and COPD existed only when the IVW model yielded $P < 0.05$ and the effect directions of all auxiliary models were consistent. Sensitivity analyses included Cochran's Q test to assess heterogeneity among SNPs ($P > 0.05$ indicated no heterogeneity), MR-Egger intercept to test horizontal pleiotropy ($P > 0.05$ indicated no pleiotropy), and leave-one-out analysis to validate the driving effect of individual SNPs on the results.

1.4 SMR Analysis

The analysis was conducted using the SMR v1.3.1 software, which integrated GWAS and eQTL summary statistics to differentiate the potential causal effects and non-specific genetic associations of gene expression on COPD, thereby to further eliminate pleiotropy bias. Throughout the analysis, the default parameters of the SMR software were consistently applied to ensure the uniformity of the analytical process.

1.5 Screening of Active Ingredients and Targets of FZN

The chemical components and action targets of Dangshen, Huangqi, Beiwuweizi (*Schisandra chinensis*), Shudihuang (*Rehmannia glutinosa*), Sangbaipi, Ziwan (*Aster tataricus*), Chuanxiong (*Ligusticum chuanxiong*), Chishao (*Paeonia lactiflora*), Fangfeng (*Saposhnikovia divaricata*), Huangqin, Danshen, Fuling (*Poria cocos),* and Baizhu (*Atractylodes macrocephala*) were retrieved through the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP, https://next.tcmsp-e.com/). Major active components were screened based on the criteria of oral bioavailability (OB) ≥ 30% and drug likeness (DL) ≥ 0.18. Shanzha (*Crataegus pinnatifida*), Dilong (*Pheretima*), and Buguzhi (*Psoralea corylifolia*) did not get relevant data from the TCMSP database. Therefore, we conducted the search using the Bioinformatics Analysis Tool for Molecular mechANism of Traditional Chinese Medicine (BATMAN-TCM, https://bionet.ncpsb.org.cn/batman-tcm/#/home) database. The parameters were set with a score cutoff of 0.84, and multiple testing corrections were performed using the Benjamini-Hochberg method, with a corrected $P < 0.05$ as the screening criterion. Subsequently, the evaluation was conducted using the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) and the SwissADME platform (https://www.swissadme.ch/), with gastrointestinal absorption scores classified as 'high' and the DL which passed at least two 'Yes' responses as the criteria for screening active ingredients. All the screened active components of traditional Chinese medicine were imported into the Swiss Target Prediction platform (https://www.swisstargetprediction.ch/) to predict their action targets. Finally, the potential target set of FZN was obtained by correcting and excluding non-human genes using the UniProt database (https://www.uniprot.org/).

1.6 Differentially Expressed Genes Analysis

Differentially expressed genes (DEGs) between the COPD group and the healthy control group were analyzed using the limma and dplyr packages. The GSE101929 and GSE244647 transcriptome datasets were integrated, and batch effects were corrected using the ComBat algorithm from the sva package. And the effectiveness of the correction was validated through principal component analysis (PCA) and boxplots of sample expression. After merging duplicate probes using the avereps function, a linear model without intercept was constructed, and the eBayes empirical Bayes method was employed to shrink gene variance and enhance the statistical

stability of low-expressed genes. Then multiple testing correction was performed using the Benjamini-Hochberg method, with the screening criteria as: $|\log_2 FC| > 0.585$ and corrected $P < 0.05$.

1.7 Weighted Gene Co-expression Network Analysis

Weighted Gene Co-expression Network Analysis (WGCNA) was employed to construct a co-expression network and to identify gene modules which were highly associated with the COPD phenotype. Using the batch-corrected standardized expression matrix as input, genes with a standard deviation greater than 0.5 were retained after quality control, while genes with expression loss and abnormal variations, and outlier samples were removed. The optimal soft threshold β was selected using a scale-free network fitting index $R^2 \geqslant 0.8$. We calculated the Pearson correlation coefficients among genes and constructed a topological overlap matrix. And then gene modules were delineated through average linkage hierarchical clustering. We calculated the correlation coefficients of each module with the COPD phenotypes and selected the modules with $P < 0.05$ which had the highest correlation strength as key modules for the disease.

1.8 Functional Enrichment Analysis

Using the clusterProfiler package, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis on the feature gene set. Gene symbols were batch converted to ENTREZ IDs to ensure complete annotation, with the species set as human. The GO analysis encompasses three aspects: biological process (BP), cellular component (CC), and molecular function (MF).

1.9 Multi-dimensional Intersection Analysis

By intersecting the causal genes of PCD identified through MR and SMR, the differentially expressed genes from the transcriptome, the pivotal hub genes from the WGCNA key modules, and the drug target set of FZN, we obtained a key candidate gene set which possesses genetic causal effects, disease abnormal expression characteristics, a central regulatory position in the network, and can be targeted by FZN.

1.10 Machine Learning Model Construction and Causal Feature Gene Screening

This study integrated 14 machine learning algorithms, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Generalized Boosted Regression Models (GBM), Least Absolute Shrinkage and Selection Operator (Lasso), Generalized Linear Model Boosting (GlmBoost), Bidirectional Stepwise Generalized Linear Regression (Stepglm [both]), Backward Stepwise Generalized Linear Regression (Stepglm [backward]), Naive Bayes, Linear Discriminant Analysis (LDA), Forward Stepwise Generalized Linear Model (Stepglm [forward]), Partial Least Squares Regression Generalized Linear Model (PLS-RGLM), Ridge Regression, Elastic Net (Enet), and Support Vector Machine (SVM). Based on the caret package, we standardized the training process using the training sets (GSE13932 + GSE21359) for modeling, and validated the model with three independent external cohorts. The input gene expression profiles were subjected to Z-score normalization, with a random seed set to 123 to ensure reproducibility of the results. The analysis of machine learning consisted of two steps: feature initial screening and model construction. Inefficient models with fewer than five features are eliminated to reduce the risk of

overfitting. Subsequently, we calculated the area under the curve (AUC) of the receiver operating characteristic (ROC) for each model in both the training set and multiple validation cohorts. The models were ranked based on the average AUC across the multiple cohorts, allowing us to select the optimal ensemble model and extract the core feature genes.

1.11 SHAP Model Interpretability Analysis

The SHAP algorithm was employed to analyze the internal decision logic of the optimal model and quantify the marginal contribution of individual gene expression to the risk prediction of COPD. Positive SHAP values indicated that high gene expression increased the risk of the disease, while negative values suggested a protective effect. Based on the kernelshap tool, we calculated the SHAP values for all samples and features in bulk. The average absolute SHAP values were used to measure the importance of features, and the top five ranked genes were selected as core predictive biomarkers.

1.12 Single-Gene Diagnostic Performance Evaluation and Nomogram Model Construction with Calibration, Validation, and Decision Curve Analysis

Using the pROC package, we calculated the ROC curves and AUC values for the top five core genes distinguishing COPD patients from healthy controls. The curves were integrated and plotted to assess the independent diagnostic value of each individual gene. A binary logistic regression nomogram was constructed based on the top five core genes. The visualization was achieved using the rms package, which converted gene expression levels into risk scores and realised the rapid quantification of disease probability for individual samples. The model was calibrated using 1,000 bootstrap resamplings, and a calibration curve was plotted to evaluate predictive accuracy. Subsequently, decision curve analysis (DCA) was conducted using the rmda package to assess the clinical net benefit of the model.

1.13 Single-cell Transcriptomic Cell Subpopulation and Gene Expression Analysis

Using the GSE279570 single-cell transcriptome dataset of lung tissue, which included 11 COPD patients and 7 healthy controls (Table 1), quality control and downstream analysis were performed using the Seurat package. The quality control criteria included: the number of detected genes per cell $\geq$ 100, and the exclusion of low-quality cells where the proportion of mitochondrial-encoded genes exceeded 10%. Qualified cells were normalized for expression levels using the LogNormalize method (scaling factor 10,000), and 1,500 highly variable genes were selected for subsequent analysis through variance stabilizing transformation (VST). After dimensionality reduction using PCA and determining the effective dimensions with the JackStraw permutation test, the Harmony algorithm was employed to correct for batch effects among samples. Subsequently, cell clustering was performed at a resolution of 0.6. The FindAllMarkers function was used to identify cell cluster-specific marker genes, with the filtering criteria set as: min.pct = 0.25, $|\log_2 FC| > 1$, and adjusted $P < 0.05$. And the top 10 upregulated genes for each cell cluster were extracted to generate a feature heatmap. We integrated multiple human single-cell reference databases, including the Human Primary Cell Atlas, Blueprint, ENCODE, and so on, to perform automatic annotation of cell clusters using the SingleR unsupervised algorithm, and to standardize the naming of subpopulations such as CD4+ T cells, CD8+ T cells, macrophages, and B cells. The t-distributed stochastic neighbor embedding (t-SNE) algorithm was employed to

visualize cell clustering and compare the distribution differences of cell subpopulations between the COPD group and the control group. Differential expression analysis of the same cell subpopulations between groups was conducted using the scRNAtoolVis toolkit, with a minimum of 5 cells per group to ensure statistical reliability, applying the same screening criteria mentioned above. Subsequently, a bar chart depicting the proportion of cell subpopulations was generated to quantify the differences in cellular components between the two groups.

Table 1. Detailed information of single-cell RNA-seq dataset for COPD

Dataset ID	Sample ID	Sequencing type	Tissue source	Platform	Group
GSE279570	GSM8576142	scRNA-seq	Lung tissue	10X Genomics	healthy control
	GSM8576143				healthy control
	GSM8576144				healthy control
	GSM8576145				healthy control
	GSM8576146				healthy control
	GSM8576147				healthy control
	GSM8576148				healthy control
	GSM8576149				COPD
	GSM8576150				COPD
	GSM8576151				COPD
	GSM8576152				COPD
	GSM8576153				COPD
	GSM8576154				COPD
	GSM8576155				COPD
	GSM8576156				COPD
	GSM8576157				COPD
	GSM8576158				COPD
	GSM8576159				COPD

1.14 Analysis of Intercellular Communication Interactions

Based on single-cell data from the COPD group, the CellChat algorithm was employed to construct a cell-to-cell communication network. We integrated the standardized expression matrix and cell subpopulation annotation and applied a human-specific ligand-receptor interaction database, covering various pathways, such as secreted signaling, membrane surface receptors, and extracellular matrix interactions. According to standard procedures, we identified the specific ligand and receptor molecules for each cell subtype. After correcting for background expression interference, we quantitatively estimated the ligand-receptor binding probabilities and signaling interaction strengths among different cell subtypes. To ensure reliability, we filtered out

communication results involving subpopulations with fewer than 10 cells and summarized the overall communication activity of each signaling pathway to dissect the intercellular regulatory patterns in the COPD pathological microenvironment.

1.15 Analysis of Network Perturbation Induced by Virtual Gene Knockout

Focusing on the core candidate gene LGMN, we performed a single-cell level virtual gene knockout analysis using the scTenifoldKnk algorithm based on single-cell transcriptome data from the COPD group. An analysis matrix was constructed using 5,000 highly variable genes selected from the original count matrix combined with VST, and quality control parameters were set as follows: mitochondrial gene proportion < 0.1 and detected genes per cell ≥ 500. Ten subnetworks were established, each randomly sampling 500 cells, to reconstruct a high-precision gene co-expression regulatory network through tensor decomposition. By setting all downstream regulatory connection weights of LGMN to zero, a knockout network was constructed and projected together with the wild-type network into the same low-dimensional space. The spatial shift *Z*-score and adjusted *P* value were then calculated for each gene. Genes with an adjusted $P < 0.05$ were identified as significantly differentially regulated genes and further subjected to GO functional and KEGG pathway enrichment analyses.

1.16 Analysis of Regulatory Network Perturbation Induced by Virtual Gene Overexpression

The scTenifoldNet algorithm was employed to simulate the transcriptional regulatory effects of LGMN overexpression, complementing the validation provided by virtual knockout experiments, thereby revealing the functional roles of target genes. Only single-cell samples from the COPD group were included, and the original count matrix was extracted from the standardized Seurat object. An analysis dataset was constructed based on 5,000 highly variable genes selected via VST in conjunction with target genes. Ten iterations of subnetwork construction were set, with each subnetwork randomly sampling 500 cells while retaining three core principal components and enabling multi-core parallel computation for accelerated processing. Through tensor decomposition, both the wild-type and overexpression perturbation networks were reconstructed, quantifying the transcriptional shift degree and *Z*-score for each gene. Significantly differentially regulated genes were selected with an adjusted $P < 0.05$, followed by GO functional annotation and KEGG pathway enrichment analysis.

1.17 Molecular Docking Validation

The amino acid sequence of the target protein was obtained from the UniProt database (https://www.uniprot.org/). The three-dimensional structure of the target protein was preferentially obtained from the PDB database (https://www.rcsb.org/) by downloading the experimentally resolved crystal structures. In the absence of a suitable structure, the high-accuracy prediction model from the AlphaFold database (https://www.alphafold.ebi.ac.uk/) was used. The core compound SDF format file was acquired from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) and converted into three-dimensional conformation using Chem3D 22.0, followed by energy minimization using the MM2 force field. Semi-flexible molecular docking was performed using AutoDock Vina 1.5.7, with the docking box covering all potential active sites of the protein to evaluate binding affinity. The docking results were visualized using PyMOL 3.1 to analyze the best binding conformation, hydrogen bonding

interactions, and characteristics of the binding pocket.

1.18 Molecular Dynamics Simulation

The simulation was performed in a WSL2 environment running Ubuntu 24.04 using GROMACS 2026.2, with computations accelerated by the NVIDIA CUDA architecture. Ligands were preprocessed with Open Babel, and the topology file was generated using ACPYPE with the General Amber Force Field (GAFF). The protein was parameterized using the AMBER99SB-ILDN force field, and the solvent system employed the TIP3P water model, with Na^+/Cl^- ions added to neutralize the net charge. After energy minimization using the steepest descent method, we performed 100 ps of isothermal-isochoric (NVT) and 100 ps of isothermal-isobaric (NPT) equilibration simulations, followed by a 100 ns unconstrained production simulation. After periodic boundary corrections of the trajectory, key parameters such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), the number of hydrogen bonds between the protein and ligand, solvent-accessible surface area (SASA), and free energy landscape (FEL) were calculated to assess the binding stability and dynamic conformational characteristics of the compound.

1.19 Construction and Grouping of the COPD Rat Model

Male SPF-grade SD rats, weighing 180-220 g, were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd. (Animal Production License No.: SYXK (Chuan) 2024-0029). The experimental protocol was approved by the Experimental Animal Welfare and Ethics Committee of the Sichuan Province Experimental Animal Association (Ethics Approval No.): SCAF20260618001), and all procedures were performed in strict accordance with the Guide for the Care and Use of Laboratory Animals. The rats were housed in the SPF-grade Animal Experiment Center of Chengdu University of Traditional Chinese Medicine, under controlled environmental conditions: the temperature of 22±2 °C, the relative humidity of 50%±10%, with a 12-hour light/dark cycle and the free access to food and water. After 7 days of acclimatization, 36 rats were randomly divided into 6 groups (6 rats per group): Control group, Model group (COPD, CS+LPS), Dexamethasone group (DEX, DEX+CS+LPS), Low-dose FZN group (L-FZN, 12.68 g • kg-1+CS+LPS), Medium-dose FZN group (M-FZN, 25.35 g • kg-1+CS+LPS), and High-dose FZN group (H-FZN, 50.70 g • kg-1+CS+LPS).

The total daily clinical dose of FZN for adults was 195 g (for a standard adult weighing 70 kg), which corresponded to a dose of 2.786 $g \cdot kg^{-1}$ based on body weight. Using the conversion factor of 6.3 for surface area between humans and rats, the clinical equivalent medium dose for rats was calculated to be 17.55 $g \cdot kg^{-1}$. The low dose was set at half of the medium dose (8.78 $g \cdot kg^{-1}$), while the high dose was double the medium dose (35.10 $g \cdot kg^{-1}$). The administration volume for each group was standardized to 10 $mL \cdot kg^{-1}$, with a daily gavage. The control group and the model group received an equal volume of physiological saline. The DEX group was administered dexamethasone phosphate injection via intraperitoneal injection daily at a dose of 1.0 $mg \cdot kg^{-1} \cdot d^{-1}$.

A rat model of COPD was established by exposing the animals to intratracheal instillation of lipopolysaccharide (LPS) using a CS exposure method. On days 1, 14, 28, 42, 56, and 71 of the experiment, the rats were anesthetized with isoflurane and administered 200 μL of a 1 mg/mL LPS solution intratracheally, while the control group received an equivalent volume of physiological

saline. Except for the days of LPS instillation, the model group and each treatment group of rats were exposed to a C−100 automatic cigarette smoke generation device for 40 minutes daily, with 6 cigarettes each time, continuing until the end of the experiment (a total of 65 days). Starting from day 72, treatment interventions were initiated and continued for 28 days.

1.20 Sample Collection and Preparation

Before the experiment concluded, rats were fasted for 24 hours, and sampling was synchronized 24 hours after the last administration. Following deep anesthesia with 5% isoflurane, painless euthanasia was performed using a gradual CO_2 inhalation method, in accordance with the American Veterinary Medical Association (AVMA) guidelines for the euthanasia of animals. For serum preparation, blood was collected from the abdominal aorta. After standing at 4 °C for 30 minutes, the samples were centrifuged at 1500 rpm for 10 min at 4 °C to separate the upper serum, which was then aliquoted and stored at −80 °C until analysis. For the preparation of bronchoalveolar lavage fluid (BALF), the right main bronchus was ligated, and the left lung was intubated. The left lung was lavaged three times with 0.5 mL of pre-cooled sterile phosphate-buffered saline (PBS) each time, with a recovery rate of ≥80%. The collected BALF was centrifuged at 1500 rpm for 10 minutes at 4 °C, and the supernatant was aliquoted and stored at −80 °C. For lung tissue preparation, the right middle lobe was collected and rinsed with pre-cooled sterile PBS. One portion was placed in RNAlater and stored at −80 °C for RNA extraction, while the remaining portion was snap-frozen in liquid nitrogen for protein extraction.

1.21 Detection of Inflammatory Factor Levels by ELISA

The concentrations of IL-1β, IL-6, and TNF-α in serum and BALF were measured using the enzyme-linked immunosorbent assay (ELISA) method. All experimental procedures were conducted strictly according to the manufacturer's instructions. The absorbance at 450 nm was measured using a Synergy H1 microplate reader (BioTek), and the concentrations of cytokines in each sample were calculated based on the standard curve.

1.22 Real-time quantitative polymerase chain reaction (RT-qPCR)

Total RNA was extracted from rat lung tissue using TRIzol reagent, and the purity and concentration of RNA were assessed using a NanoDrop 2000 spectrophotometer. cDNA was synthesized via reverse transcription using the ReverTra Ace qPCR RT Kit. Amplification was performed on the QuantStudio 5 PCR system using SYBR Green qPCR Master Mix. GAPDH was used as the internal reference gene, and all primers were custom synthesized by Wuhan Saiweier Biotechnology Co., Ltd., with the primer sequences listed in Table 2. The amplification program included an initial denaturation at 95 °C for 30 seconds, followed by 40 cycles of denaturation at 95 °C for 5 seconds and annealing/extension at 60 °C for 30 seconds. Melting curve analysis was performed immediately after amplification to verify product specificity. The relative mRNA expression levels of each gene were calculated using the $2^{-\Delta\Delta Ct}$ method. All biological samples were analyzed in triplicate.

Table 2 Primer Sequences

Gene	Upstream primer (5'→3')	Downstream primer (5'→3')
NT5E	GATGGTGAAGGTGGTGCTAA	AGTAGCCACGATGTTGGTGA
SPP1	CCAAGTAAGTCCAACGAAGC	GGTGATGTCCTCGTCTGTCA

BLK	GCAGAGCAAGATCCAAGAGC	GCCGATGATGGTGAAGATGT
PLAT	ACCGCAAATGGAACAGAAAC	GCAGGGTCATCCATAGAGGA
LGMN	TGGCTGACTGGATTGGAATG	GCTTGATGGTGCTGATGTGA
GAPDH	AGGTCGGTGTGAACGGATTTG	TGTAGACCATGTAGTTGAGGTCA

1.23 Western blot

A suitable amount of frozen lung tissues was weighed, minced with scissors, and lysed in an appropriate volume of radioimmunoprecipitation assay (RIPA) buffer for 30 minutes, followed by centrifugation at 12,000 rpm for 15 minutes at 4 °C to collect the supernatant. The total protein concentration was determined using a BCA protein assay kit. Equal amounts of protein samples were mixed with the sample loading buffer and denatured in a water bath at 100 °C for 10 minutes. After separation by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), the proteins were transferred to a polyvinylidene fluoride membrane (PVDF) at a constant current of 300 mA for 90 minutes. The membranes were blocked with 5% non-fat dry milk at room temperature for 2 hours and then incubated with diluted specific primary antibodies against NT5E (1:2,000), SPP1 (1:2,000), BLK (1:2,000), PLAT (1:2,000), LGMN (1:2,000), and GAPDH (1:5,000) overnight at 4 °C with gentle shaking. After washing three times with TBST, the membranes were incubated with the corresponding secondary antibodies at room temperature for 2 hours. Protein bands were visualized using an ECL chemiluminescence kit, and images were captured with the ChemiScope 6,000 imaging system. The grayscale values of the target bands were analyzed using ImageJ software and normalized to GAPDH as an internal control.

1.24 Statistical Analysis

All statistical analyses and plotting were performed using R version 4.5.3 and GraphPad Prism version 10. Quantitative data were presented as mean ± standard deviation (SD). For comparisons between two groups, normally distributed data were analyzed using independent samples t-test. For non-normally distributed data, the Mann-Whitney U test (Wilcoxon rank-sum test) was employed. For comparisons among multiple groups, normally distributed data with equal variances were analyzed using one-way ANOVA, followed by Tukey's HSD method for multiple comparisons. For non-normally distributed data, the Kruskal-Wallis H test was applied, followed by the Wilcoxon rank-sum test with FDR correction for multiple comparisons. Correlation analyses were conducted using Spearman's rank correlation. All tests were two-sided, and $P < 0.05$ was considered statistically significant.

2 Results

2.1 MR Identifies Core PCD-Related Genes Genetically Associated with COPD

Using the cis-eQTL dataset provided by the eQTLGen Consortium, we performed a two-sample MR analysis and preliminarily identified 308 PCD-related genes with potential genetic associations with COPD ($P < 0.05$). To further validate these genetic associations at the protein expression level, this study performed stratified two-sample MR analyses based on the cis-pQTL datasets from the Finnish biobank and the UKB-PPP database, identifying 13 and 65 PCD genes with significant genetic associations with the onset of COPD, respectively ($P < 0.05$). The genetic causal association between the MR-selected PCD genes and COPD was confirmed to be robust against heterogeneity and pleiotropy, as evidenced by Cochran's Q test for heterogeneity,

the MR-Egger intercept test for horizontal pleiotropy, leave-one-out analysis, and funnel plot diagnostics, indicating that the overall conclusions were robust.

2.2 SMR Identifies Core PCD-Related Genes Genetically Associated with COPD

After data preprocessing and quality control, a total of 15,631 SNPs were included for subsequent analysis. Following the annotation of gene IDs for the SNPs, $P_{SMR} < 0.05$ and $P_{HEIDI} > 0.05$ were used as significance screening criteria, and the results were intersected with PCD-related genes. A total of 152 PCD-related genes genetically associated with COPD were identified. To visually present the distribution characteristics of the SMR analysis results across the entire genome, a Manhattan plot was constructed based on chromosome position and the significance levels of the SMR analysis (Figure 1A). Additionally, a circular plot was employed to display the positioning of significant genes within the genome and their significance levels from a global perspective (Figure 1B).

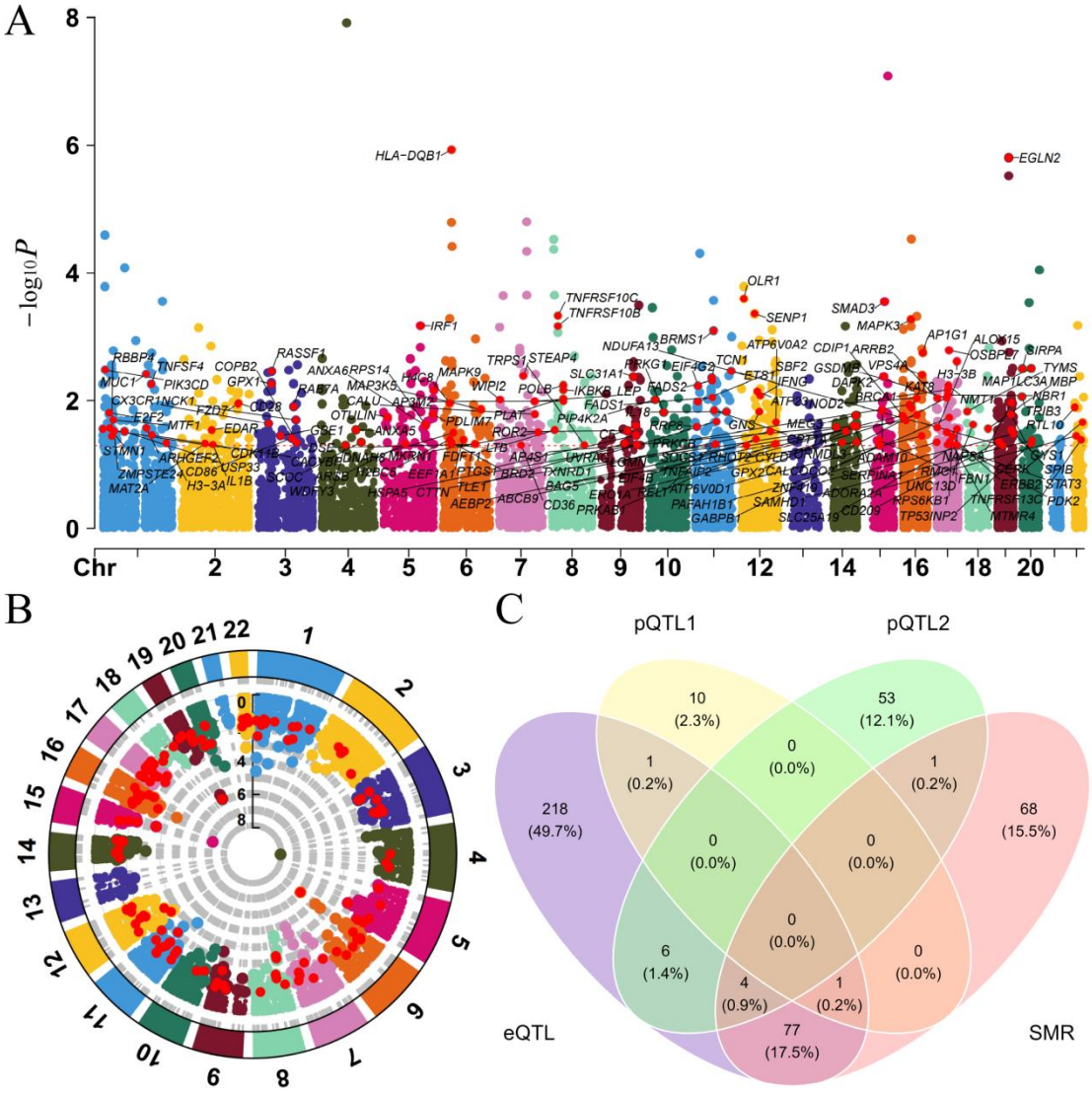

Figure 1 Results of SMR Analysis.

A: Manhattan plot of 152 PCD genes genetically associated with COPD. B: Circular plot of the 152 genes. C: Venn diagram of the results from four MR analyses, where pQTL1 represents the analysis based on cis-pQTL data from the Finnish biobank, and pQTL2 represents the analysis

based on cis-pQTL data from the UKB-PPP.

2.3 Integration of Multi-Omics MR Analysis Results

To integrate genetic causal evidence from different molecular levels, this study combined and deduplicated PCD genes selected from eQTL analysis, the Finnish population biobank pQTL analysis, UKB-PPP pQTL analysis, and SMR analysis. Ultimately, a total of 439 PCD-related genes with a genetic causal association with COPD were summarized (Figure 1C).

2.4 Identification of Active Ingredients in FZN and Prediction of Their Potential Targets

Based on the TCMSP 5.0 database, the active components of 16 traditional Chinese medicinal herbs in the FZN formula were screened, resulting in a total of 371 potential active components. The number of components contained in each herb is as follows: Dangshen has 17 components, Huangqi has 18, Beiwuweizi has 9, Shudihuang has 2, Sangbaipi has 25, Ziwan has 13, Chuanxiong has 6, Dilong has 23, Chishao has 14, Fangfeng has 20, Huangqin has 33, Buguzhi has 62, Danshen has 59, Fuling has 6, Baizhu has 4, and Shanzha has 60, with a total of 31 overlapping components. After conducting target retrieval and standardization of the names for the aforementioned components, a total of 1230 relevant target genes were obtained after merging and removing duplicates. On this basis, a "Traditional Chinese Medicine - Active Component - Target" network diagram was constructed, which contains 1563 nodes and 31366 edges (Figure 2).

Based on network topology analysis, the active components in various traditional Chinese medicines were ranked according to their degree centrality. For each traditional Chinese medicine, the top two key components were selected (if fewer than two existed, all were selected). Additionally, non-specific components that were widely distributed in plants, such as quercetin, kaempferol, luteolin, sitosterol, isorhamnetin, stigmasterol, and β-sitosterol, were excluded. The key components identified in the final selection are presented in Table 3: Bavacoumestan B and Corylidin from Buguzhi, Suchilactone and 2-(3,4-Dichlorophenyl) Succinic Acid from Shanzha, L-Tyrosine and Uridine from Dilong, Wogonin and Baicalein from Huangqin, Wogonin and Phellopterin from Fangfeng, 7-O-methylisomucronulatol and Formononetin from Huangqi, Tanshinone IIA and Dihydrotanshinlactone from Danshen, Baicalein and (+)-catechin from Chishao. Hederagenin and (2R)-2-[(3S,5R,10S,13R,14R,16R,17R)-3,16-dihydroxy-4,4,10,13,14-pentamethyl-2,3,5,6,12,15,16,17-octahydro-1H-cyclopenta[a]phenanthren-17-yl]-6-methylhept-5-enoic acid from Fuling, 3β-Hydroxymethyllenetanshiquinone and Frutinone A from Dangshen, Galangin and 5,7-dihydroxy-2-(3-hydroxy-4-methoxyphenyl)chroman-4-one from Ziwan, Iristectorigenin and Sanggenone F from Sangbaipi, Arnebin 7 and Gomisin R from Beiwuweizi, 3β-acetoxyatractylone and 8β-ethoxy atractylenolide III from Baizhu (Atractylodes macrocephala), and Myricanone and Perlolyrine from Chuanxiong, Shudihuang contained only sitosterol and stigmasterol, both of which were non-specific components and were therefore not included in this screening.

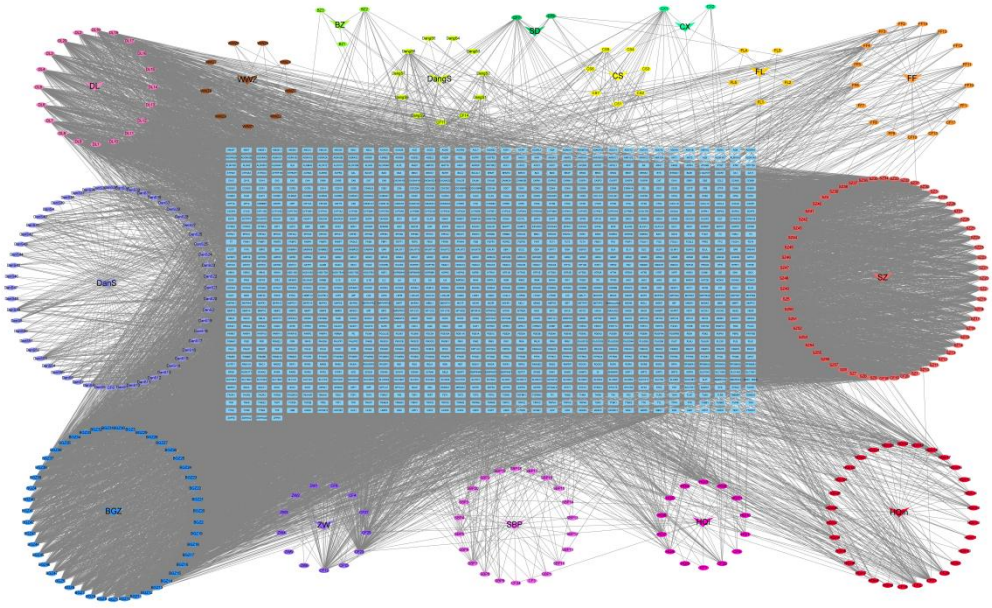

Figure 2 “TCM - Active Component - Target” network

Table 3 Summary of Active Ingredients in FZN

PubChem CID	English Name of Active Ingredient	Herb	Abbreviation	Degree
132350840	Suchilactone	Shanzha	SZ41	103
5321820	Bavacoumestan B	Buguzhi	BGZ8	101
444899	Arachidonic acid	Buguzhi	BGZ11	101
612736	2-(3,4-Dichlorophenyl)Succinic Acid	Shanzha	SZ11	103
6057	L-Tyrosine	Dilong	DL9	99
6029	Uridine	Dilong	DL17	99
5281703	Wogonin	Huangqin、Fangfeng	CF21	60
15689652	7-O-methylisomucronulatol	Huangqi	HQi3	42
164676	Tanshinone IIA	Danshen	DanS53	41
5281605	Baicalein	Chishao、Huangqin	CF2	37
5280378	Formononetin	Huangqi	HQi7	35
177827442	Dihydrotanshinone	Danshen	DanS	30

			30	
98608	Phellopterin	Fangfeng	FF4	26
161748	Myricanone	Chuanxiong	CX1	24
5488781	Iristectorigenin (9CI)	Sangbaipi	SBP2	20
73299	Hederagenin	Fuling	CF22	19
5318290	3-beta-Hydroxymethyllenetanshiquinone	Danshen、Dangshen	CF14	17
5281616	Galangin	Ziwan	ZW1	17
163082914	3β-acetoxyatractylone	Baizhu	BZ2	16
42608038	Sanggenone F	Sangbaipi	SBP15	16
98914	Arnebin 7	Wuweizi	WWZ3	15
441965	Frutinone A	Dangshen	DangS1	14
9064	(+)-catechin	Chishao	CS2	11
676152	5,7-dihydroxy-2-(3-hydroxy-4-methoxyphenyl)chroman-4-one	Ziwan	ZW2	9
11495015	Gomisin R	Wuweizi	WWZ8	8
160179	Perlolyrine	Chuanxiong、Dangshen	CF11	6
14448075	8β-ethoxy atractylenolide III	Baizhu	BZ3	5
10743008	(2R)-2-[(3S,5R,10S,13R,14R,16R,17R)-3,16-dihydroxy-4,4,10,13,14-pentamethyl-2,3,5,6,12,15,16,17-octahydro-1H-cyclopenta[a]phenanthren-17-yl]-6-methylhept-5-enoic acid	Fuling	FL1	3

2.5 DEG Screening and Analysis

This study integrated the GEO transcriptome datasets GSE148004 and GSE162635 for analysis, retaining a total of 233 samples after batch correction. Before batch correction, there was a significant shift in the median levels of gene expression between these two datasets. As a result, the PCA clustering results showed that samples clustered strictly according to the source of the sequencing datasets, highlighting the pronounced interference effects of batch bias. After completing batch correction, the distribution of the expression matrix became more uniform, with

samples from different datasets sufficiently mixed in the PCA plot. Sample clustering was no longer dominated by sequencing batches, confirming the elimination of batch effects across datasets (Figure 3A-B). Based on the corrected normalized expression profiles, inter-group differential expression analysis identified a total of 499 DEGs. A heatmap was generated for the top 20 representative differentially expressed genes (Figure 3C), and a comprehensive volcano plot illustrated the distribution pattern of all genes (Figure 3D).

2.6 WGCNA

Based on the batch-corrected joint transcriptome expression matrix, WGCNA was conducted, and module clustering was completed according to gene co-expression similarity. Ultimately, six independent gene modules were obtained: MEbrown, MEturquoise, MEblue, MEgreen, MEblack, and MEyellow (Figure 3E). The MEblack module exhibited the strongest positive correlation with the COPD phenotype (r=0.53, P=3× 10^{-18}), while the MEturquoise module showed the strongest negative correlation with COPD (r=-0.19, P=0.004). A total of 6,668 candidate genes from the two core modules were extracted after removing duplicate gene entries for subsequent research.

2.7 Multi-dimensional Gene Set Intersection Screening Analysis

We integrated all genes from the DEG and WGCNA core modules while removing duplicate genes, resulting in 6,746 candidate genes related to COPD transcriptomics. An intersection analysis was conducted between this transcriptomic gene set and the potential active targets of the FZN compound, as well as the genetic causal genes related to PCD identified through multi-omics MR screening. A total of 16 genes were found to be both differentially expressed and causally associated with COPD (Figure 3F).

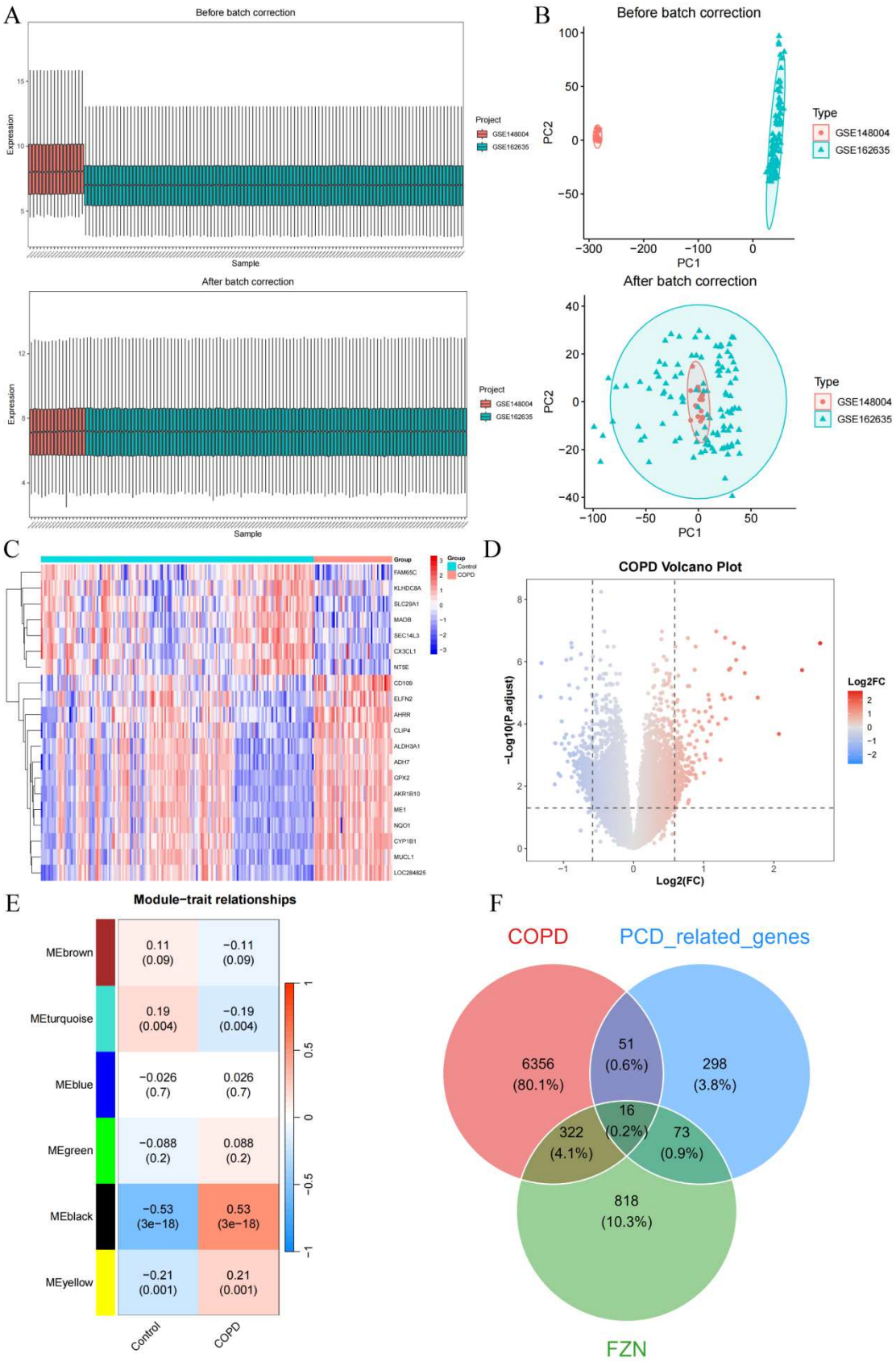

Figure 3. Integration of COPD transcriptome data and differential expression analysis. A: Box plot of gene expression distribution before and after batch correction. B: PCA scatter plot of samples before and after batch correction. C: Hierarchical clustering heatmap of the top 20

genes ranked by change magnitude among DEGs. D: Volcano plot of differential gene expression for all genes. E: Heatmap of the correlation between WGCNA modules and COPD phenotypes. F: Venn diagram of the intersection of multidimensional candidate genes.

2.8 Functional Enrichment Analysis

The BP enrichment terms focused on functional units related to steroid metabolic processes, amyloid-β responses, neuronal apoptosis regulation, negative regulation of multicellular growth, and amino acid-stimulated cellular responses. This suggested that the core genes were primarily involved in regulating programmed cell death, maintaining lipid metabolic homeostasis, and responding to external damage stimuli. The CC enrichment results pointed to membrane structures and organelle microdomains such as the outer side of the plasma membrane, lysosomes, endoplasmic reticulum lumen, lysosomal lumen, and apical regions of cells, indicating that the gene functions were mainly localized to the cell membrane and lysosome-related degradation systems. The MF enrichment terms predominantly involved platelet-derived growth factor binding, endopeptidase activity, serine-type endopeptidase activity, and integrin binding, reflecting that the core genes generally exhibited protease catalytic and cytokine receptor binding activities. The KEGG pathway enrichment analysis identified pathways related to prostate cancer, human cytomegalovirus infection, lipid metabolism, atherosclerosis, melanoma, glioma, rheumatoid arthritis, the PI3K-Akt signaling pathway, the Apelin signaling pathway, hepatitis C, and hepatitis B. These pathways could be categorized into three main groups: lipid metabolism, atherosclerosis, PI3K-Akt, and Apelin signaling pathways, which dominated lipid metabolism, cell survival, and the regulation of inflammatory homeostasis; rheumatoid arthritis and virus infection-related pathways indicated chronic persistent inflammation activation; and various tumor-related pathways reflected the dysregulation of cell proliferation and apoptosis (Figure 4A-B). The aforementioned enrichment characteristics suggested that the potential mechanism of FZN intervention in COPD was closely related to the PCD pathway, lipid metabolism, inflammatory signaling, and protease regulation. The compound's active components might target 16 core genes, synchronously regulating multiple biological pathways such as apoptosis, lysosomal degradation, lipid homeostasis, and chronic inflammation, thereby intervening in the process of lung tissue damage in COPD from multiple perspectives.

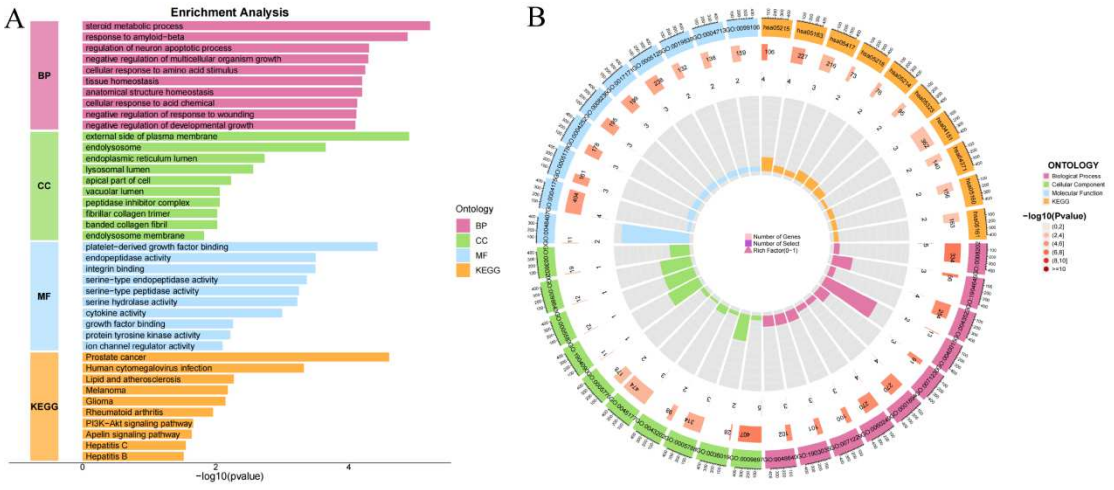

Figure 4. Functional enrichment analysis of core intersecting genes.

A: The three major classifications of GO and the KEGG pathway enrichment bar chart. B:

Multi-dimensional enrichment circular visualization chart.

2.9 Construction and Performance Evaluation of Diagnostic Models Based on Machine Learning

Based on the intersection of multi-omics data, 16 core genes related to COPD were identified. This study constructed 87 machine learning diagnostic models encompassing 14 different algorithms. The models were fitted using the internal training cohorts GSE13932 and GSE21359, and the generalization performance was validated using three independent external datasets: GSE162635, GSE212331, and GSE324136. The cross-cohort AUC values of each model were visualized in a heatmap (Figure 5A), where the AUC value of the GBM model reached 0.754, demonstrating the best discrimination performance among all models, and it could serve as a potential molecular diagnostic model for COPD.

2.10 SHAP-Based Interpretability Analysis of the Optimal Diagnostic Model

The global feature importance bar chart was sorted in descending order based on the average absolute SHAP values of each gene (Figure 5B). NT5E, SPP1, BLK, PLAT, and LGMN were the top five core predictive markers contributing to this study, with corresponding average SHAP weights of 0.125, 0.087, 0.076, 0.065, and 0.043, respectively. These genes had a significantly greater driving effect on the model's discrimination results for COPD compared to other targets, making them key molecular features for distinguishing between diseased and healthy samples. The SHAP swarm plot (Figure 5C) indicated that when the expression levels of SPP1, PLAT, and LGMN increase, the SHAP values shifted positively, significantly enhancing the model's probability of classifying the sample as COPD. Conversely, high expressions of NT5E and BLK corresponded to negative SHAP contributions, which could reduce the risk of COPD prediction. The single-sample SHAP waterfall plot and force plot further illustrated the predictive decomposition logic of a typical individual (Figure 5D-E). The baseline predicted value E[f(x)] for this sample was 0.213. The features NT5E, BLK, PLAT, and SPP1 sequentially contributed negative predictions, collectively lowering the overall predicted score. Meanwhile, the remaining secondary features collaboratively provided positive SHAP gains, ultimately reducing the comprehensive predicted score f(x) to -0.0277. The results of the SHAP visualization indicated that NT5E, SPP1, BLK, PLAT, and LGMN were the core genes driving the model's risk assessment for COPD. This validated the molecular value of these core genes in predicting the onset of COPD from the perspective of model interpretability, while also providing key candidate targets for the FZN compound's targeted intervention in COPD.

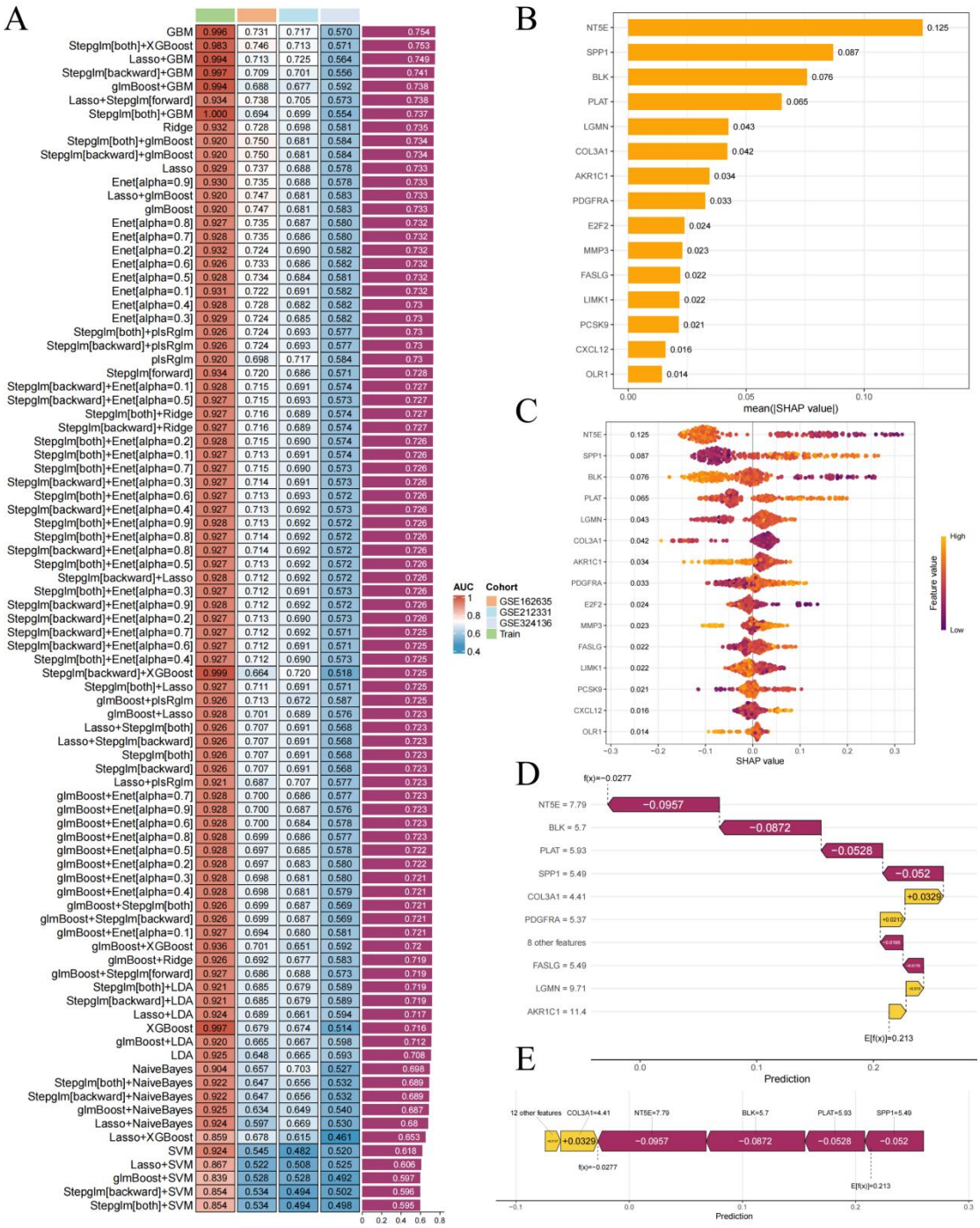

Figure 5 Comprehensive Visualization of the Performance and SHAP Interpretability of the Optimal COPD Prediction Model

A: Heatmap of AUC performance for the multi-algorithm fusion COPD diagnostic model. B: Bar chart of global feature importance. C: SHAP swarm summary plot. D: SHAP waterfall plot. E: SHAP force plot.

2.11 ROC Validation of Core Genes, Nomogram Construction, and Evaluation of Clinical Decision-Making Performance

The results showed (Figure 6A) that the areas under the ROC curve for NT5E, SPP1, BLK, PLAT, and LGMN were 0.803, 0.804, 0.710, 0.781, and 0.606, respectively. Among these, SPP1

and NT5E demonstrated the optimal efficacy in distinguishing between COPD patients and healthy samples. All genes could serve as potential independent molecular diagnostic markers for COPD. To achieve a visual and quantitative assessment of individual COPD risk, a multi-gene combined diagnostic nomogram was constructed by integrating the aforementioned five core genes (Figure 6B). The nomogram assigned a specific score scale for each gene, allowing the corresponding individual score to be read based on the gene expression levels. By summing all scores, a total score is obtained, which could be directly converted into the quantitative risk probability of the subject developing COPD based on the total score scale. A calibration curve was employed to verify the predictive accuracy of the nomogram using 1,000 bootstrap resampling (Figure 6C), where the apparent fit curve and bias-corrected curve closely aligned with the ideal reference diagonal line, indicating minimal deviation between the model's predicted probability of incidence and the actual incidence probability of the samples, demonstrating excellent prediction stability and accuracy. DCA (Figure 6D) was used to quantify the clinical value of the model: within a wide range of threshold probability intervals, the net benefit of the five-gene combined model consistently exceeded that of the two baseline reference strategies, 'population-wide intervention' and 'no population intervention'. Furthermore, under various cost-benefit ratios, the model maintained a positive net benefit, demonstrating that this polygenic diagnostic model possessed considerable clinical translational value and could provide an objective quantitative basis for early screening of COPD and stratification of high-risk populations.

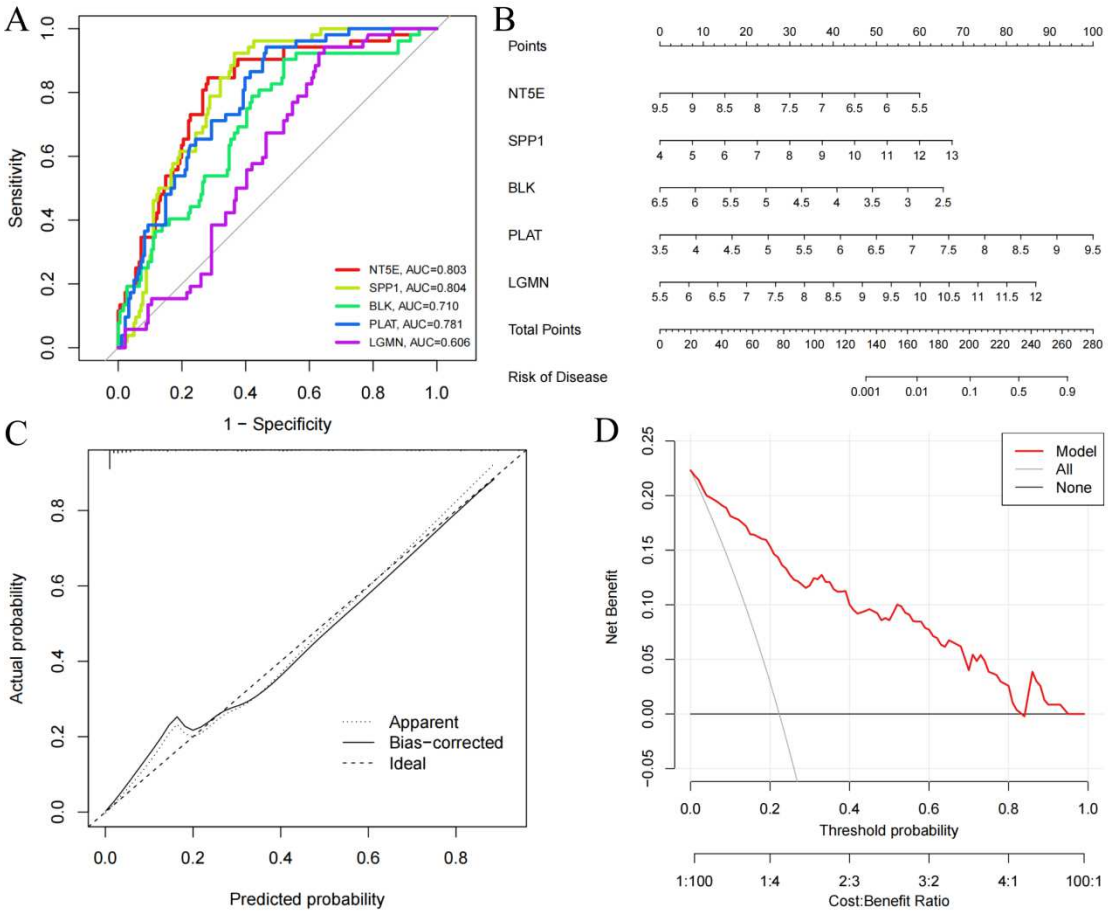

Figure 6 Evaluation of Core Gene Diagnostic Efficacy and Clinical Prediction Models.

A: ROC curves for individual genes NT5E, SPP1, BLK, PLAT, and LGMN. B: Nomogram for the risk diagnosis of COPD based on the combination of five genes. C: Calibration curve of the

nomogram model. D: DCA.

2.12 Analysis of the Single-Cell Transcriptomic Atlas in COPD and the Cell-Specific Expression of Core Genes

The t-SNE analysis of the single-cell transcriptome in COPD identified a total of 17 cell subpopulations with significant transcriptomic characteristics (Figure 7A). Based on this, classic cell marker genes were utilized for annotation, ultimately determining ten main types of lung tissue cells: epithelial cells, CD4+ T cells, CD8+ T cells, T cells, NK cells, B cells, monocytes, macrophages, endothelial cells and fibroblasts (Figure 7B). A comparison of the t-SNE distributions between the control group and the COPD group revealed that, compared to the control group, the relative proportions of CD4+ T cells, CD8+ T cells, epithelial cells, NK cells and T cells were significantly reduced in the COPD group, while the relative proportions of macrophages, monocytes, B cells, endothelial cells and fibroblasts were increased. The bubble plot and violin plot further illustrated the expression characteristics of five core genes across various cell types (Figure 7C–H). Among them, LGMN was widely expressed in epithelial cells, macrophages, monocytes and fibroblasts; BLK was primarily expressed in B cells; PLAT exhibited high expression in endothelial cells and fibroblasts; the expression of SPP1 was concentrated in macrophages; and NT5E was mainly expressed in endothelial cells, fibroblasts, epithelial cells and B cells.

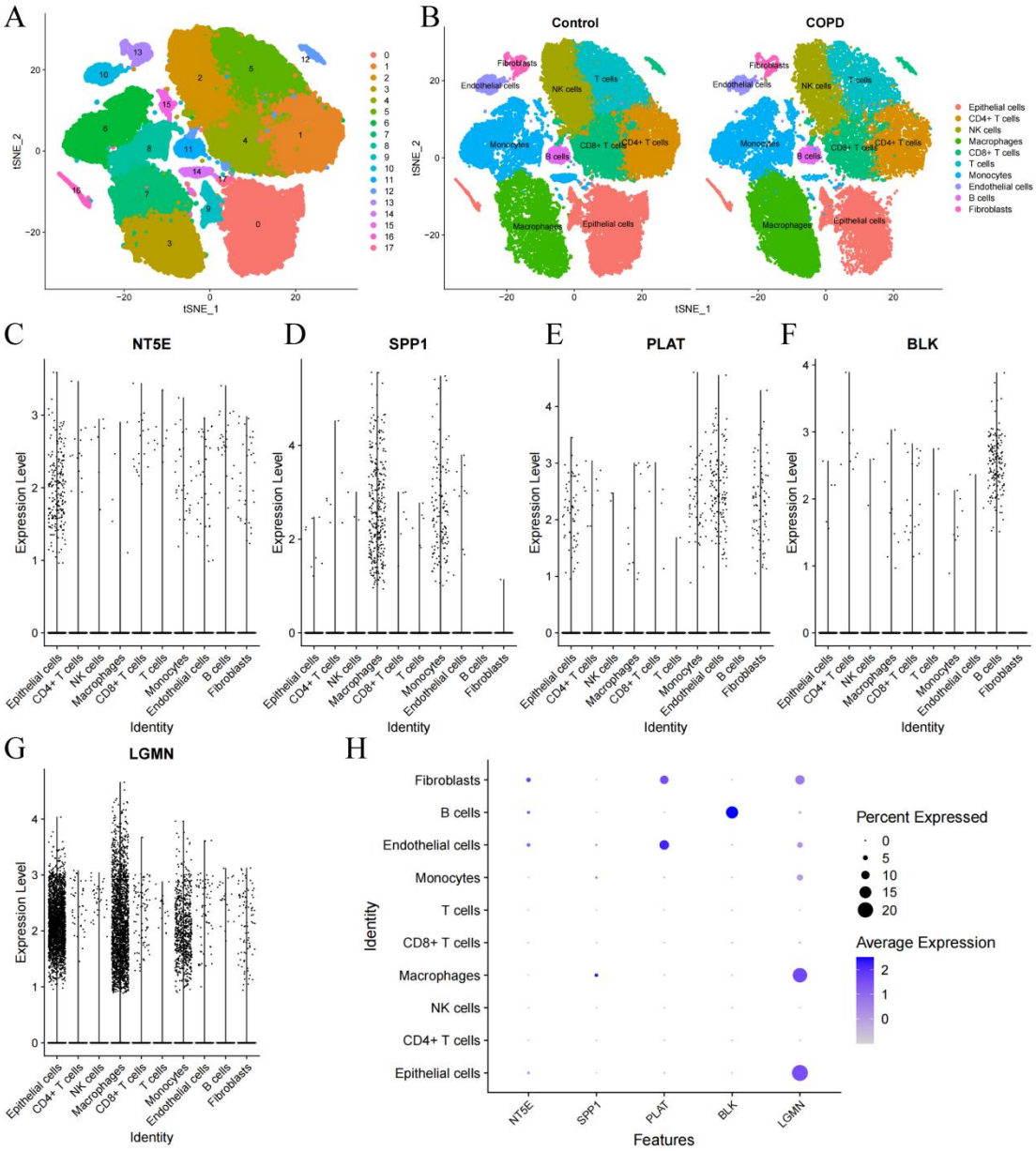

Figure 7: Single-cell atlas of lung tissue in COPD and cell-type-specific expression of core genes. A: t-SNE clustering diagram of single cells from healthy control and COPD groups. B: t-SNE diagram of unsupervised clustering of single cells. C–G: Violin plots showing the expression of five core genes across different cell types. H: Bubble plot depicting the expression of five core genes.

2.13 Analysis of the Cellular Communication Interaction Network in the COPD Lung Microenvironment

The results of the signal transmission types of cell interaction signals are shown in Figure 8A. The proportion of secretory signals accounted for 61.8%, making it the predominant form of cell communication; interactions involving extracellular matrix receptors accounted for 21.7%, while direct cell contact signals represented the lowest proportion at only 16.5%. The global cell interaction network diagram intuitively displayed all ligand-receptor communication connections between 10 types of cells (Figures 8B-C). It could be observed that macrophages, fibroblasts, and

endothelial cells possessed the richest intercellular signal interactions, serving as communication hub cells in the COPD microenvironment. Among them, there was a significant amount of bidirectional signal transmission between macrophages and various immune cells, suggesting that during the airway remodeling process, macrophages regulated the inflammatory microenvironment through extensive signaling. Further, single-source cellular communication networks were constructed for each cell type. The results indicated (Figure 8D–M) that epithelial cells, as the core of the airway barrier, could release a large amount of signals that acted on macrophages, endothelial cells, and various types of T lymphocytes, serving as upstream signal sources for the initiation of airway inflammation. Macrophages possessed the strongest paracrine signaling output capability, widely regulating the other nine cell types and mediating the amplification of persistent chronic inflammation. Fibroblasts formed a strong bidirectional communication loop with macrophages and epithelial cells, synchronously receiving inflammatory signals and secreting matrix-related ligands, collaboratively driving airway fibrotic remodeling. CD4+ T cells, CD8+ T cells, NK cells, B cells, monocytes, and endothelial cells exhibited specific directional signal transmission, forming a complex cross-regulatory network among immune cells, which collectively maintained the imbalanced inflammatory microenvironment.

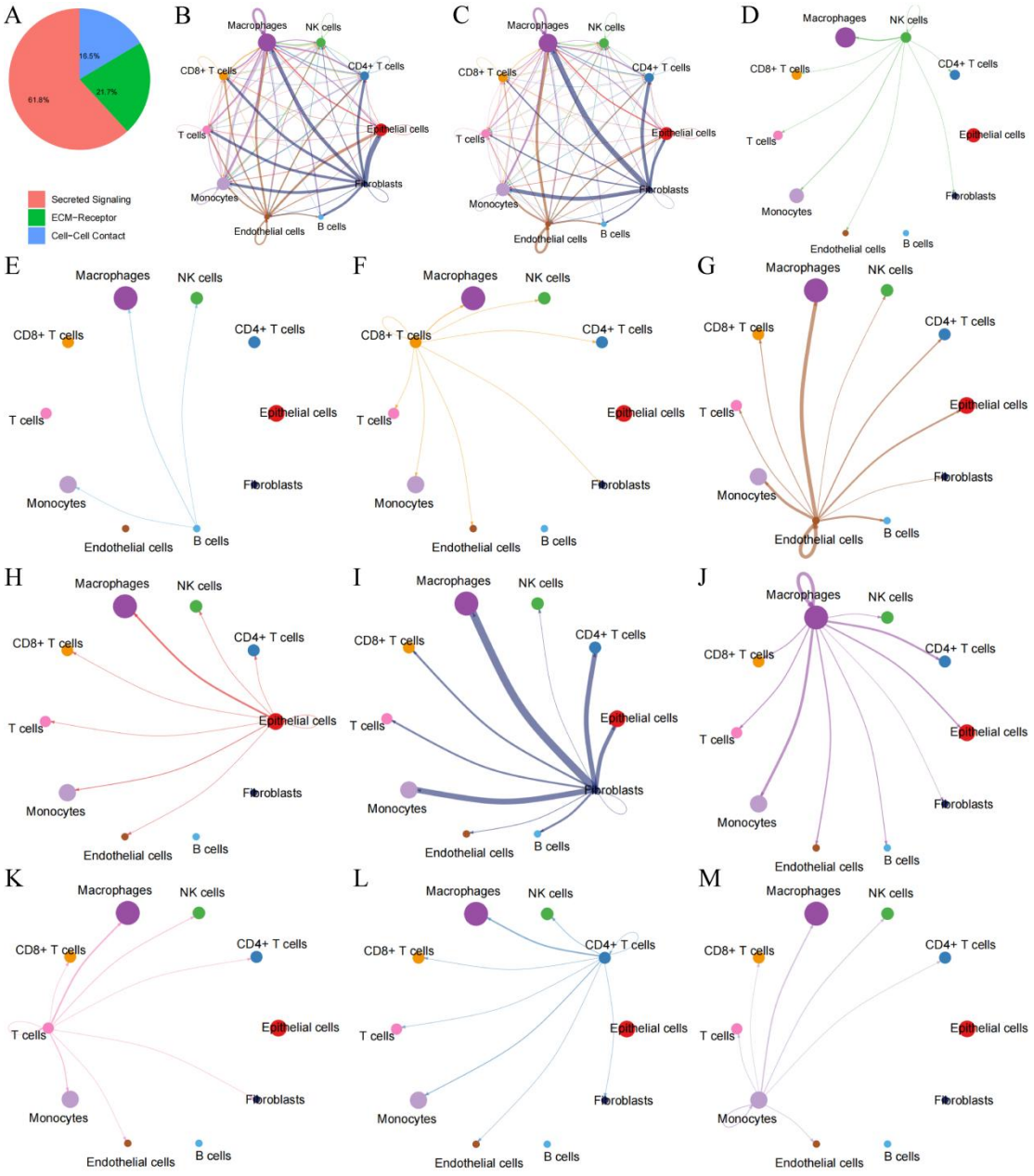

Figure 8 Analysis of the Cell Communication Network in COPD Lung Tissue.
A: Classification diagram of cell communication ligands and receptors. B: Interaction quantity diagram of the global interaction network among all cell subpopulations. C: Interaction intensity diagram of the global interaction network among all cell subpopulations. D-M: Directed communication network diagrams for individual cell types, corresponding to NK cells, B cells, CD8+ T cells, endothelial cells, epithelial cells, fibroblasts, macrophages, T cells, CD4+ T cells, and monocytes.

2.14 DGE and Functional Enrichment Analysis Following LGMN Gene Knockout

To investigate the core regulatory function of LGMN in the pathological microenvironment of COPD, this study performed a virtual knockout of LGMN based on single-cell transcriptomic data and systematically assessed its impact on genome-wide transcriptional expression changes and functional enrichment. Differential expression analysis revealed that only 1.4% of all genes

exhibited significant expression changes following LGMN knockout, while 98.6% of genes showed no notable transcriptional fluctuations (Figure 9A). Among these, genes such as *NFKBIA*, *IER3*, *IL1B*, *NAMPT*, and *FCN1* demonstrated the most pronounced changes post-LGMN knockout (Figure 9B-C). The GO enrichment bar chart indicated that BP was primarily enriched in inflammation-related entries such as lipopolysaccharide response, regulation of inflammatory response, and cellular response to molecules of bacterial origin. KEGG pathway analysis shows significant enrichment in classical pro-inflammatory signaling axes, including the NF-κB signaling pathway, TNF pathway, IL-17 pathway, and NOD-like receptor pathway, suggesting that the absence of LGMN may broadly amplify the body's inflammatory cascade (Figure 9D).

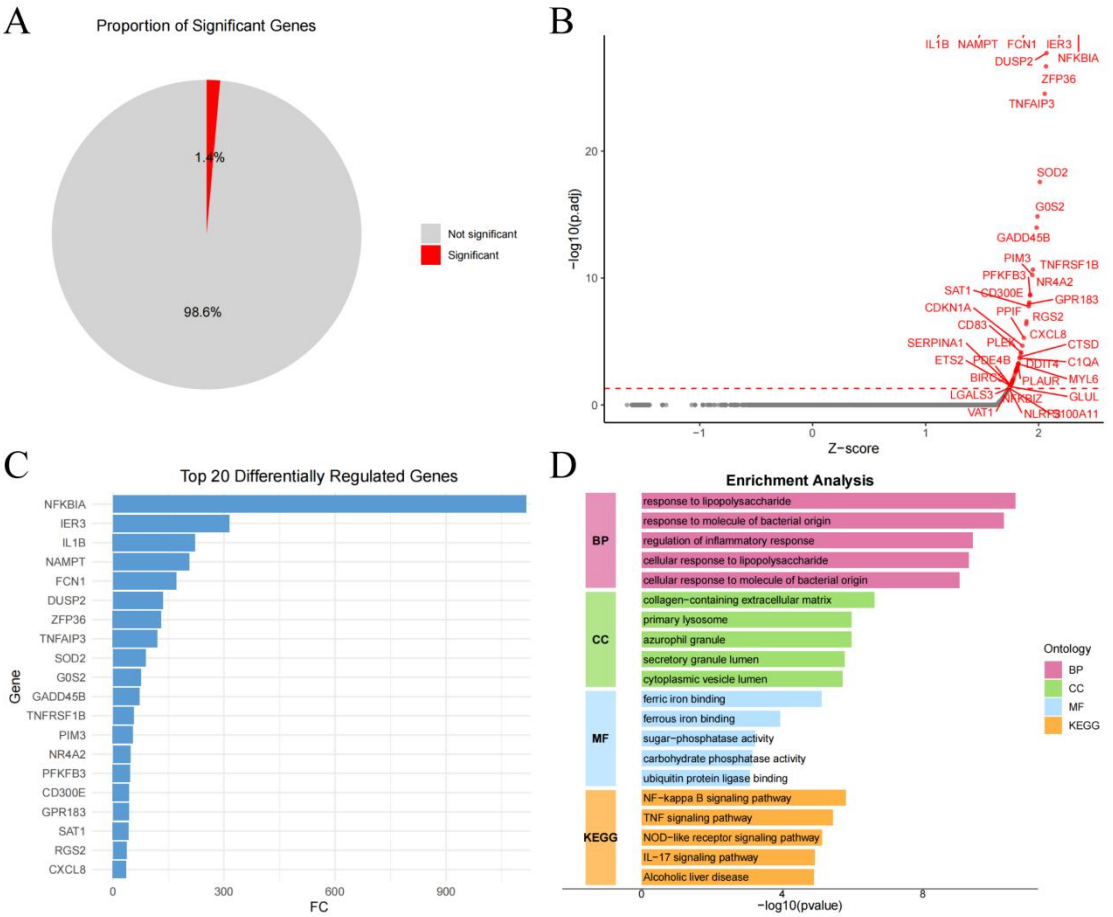

Figure 9 Visualization of Differential Gene Distribution and GO/KEGG Functional Enrichment after LGMN Virtual Knockout.

A: Pie chart of the significance distribution of all genes. B: Volcano plot of differential genes. C: Bar chart of the top 20 differentially expressed genes after LGMN knockout. D: Bar chart of GO/KEGG enrichment.

2.15 Analysis of Transcriptomic Changes and Functional Enrichment Following Virtual Overexpression of the LGMN Gene

To comprehensively elucidate the bidirectional regulatory role of LGMN in the lung microenvironment, we further conducted a single-cell transcriptomic simulation analysis of LGMN virtual overexpression, systematically analyzing the genome-wide transcriptomic changes and functional enrichment under high expression levels of LGMN. The pie chart depicting the

significant distribution of genes shows that 0.2% of the expressed transcripts exhibited significant expression differences following LGMN overexpression, while 99.8% of gene transcription levels remained stable (Figure 10A). Among them, genes such as *SFTPC*, *SFTPA1, SFTPA2*, *SFTPB*, *NAPSA*, and *SFTPD* exhibited the most pronounced changes after LGMN overexpression (Figures 10B-C). The bar chart of GO enrichment indicated that the BP terms were primarily enriched in pathways related to surfactant homeostasis, respiratory gas exchange, and phospholipid metabolism regulation, which were crucial for the physiological maintenance of alveoli. The KEGG pathways significantly enriched phagocytosis, ABC transporter proteins, and ether lipid metabolism, confirming that LGMN overexpression primarily regulated alveolar epithelial lipid homeostasis and the maintenance of airway barrier processes (Figure 10D).

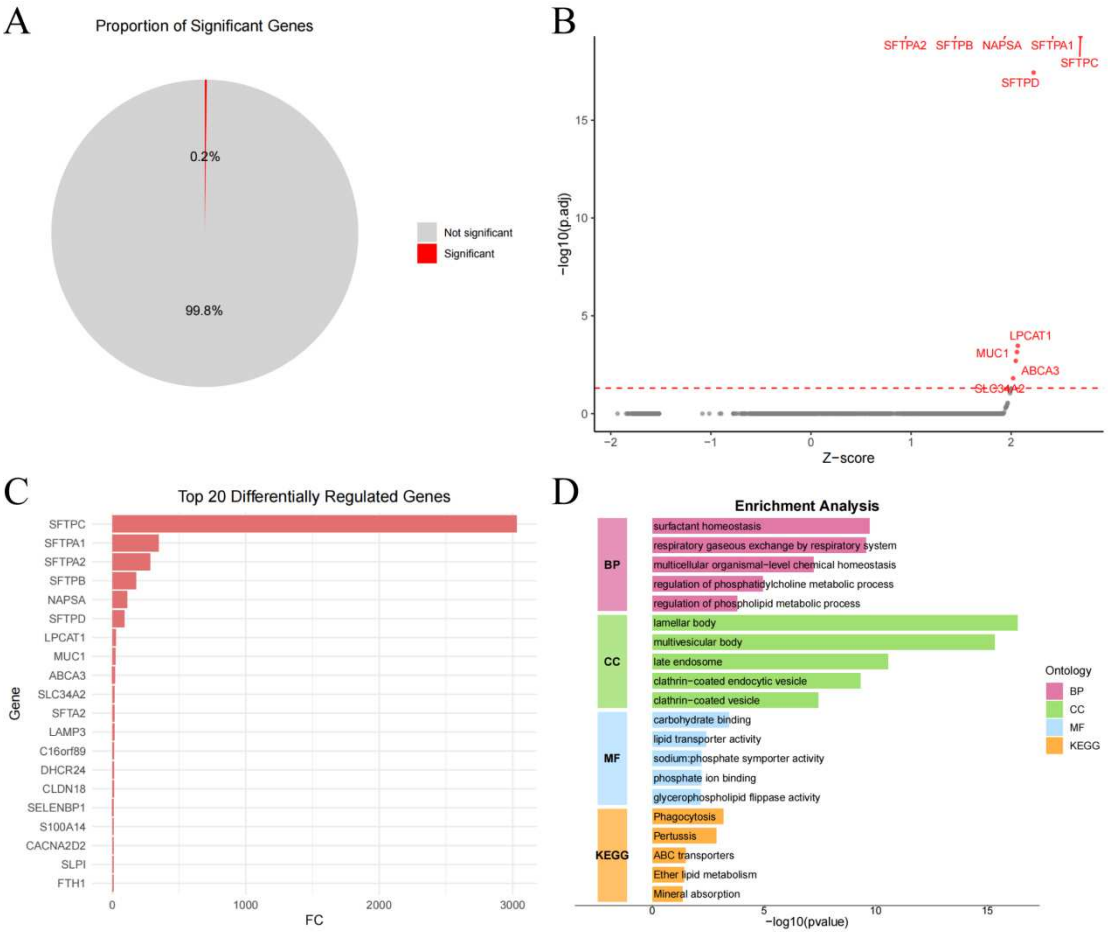

Figure 10 Visualization of Differential Gene Distribution and GO/KEGG Functional Enrichment after LGMN Overexpression.

A: Pie chart of gene significance proportion. B: Volcano plot of differential genes. C: Bar chart of the top 20 differentially expressed genes after LGMN overexpression. D: Bar chart of GO/KEGG enrichment for differential genes.

2.16 Validation of Active Ingredients and Core Target Protein Molecular Docking

The two-dimensional structures of 27 candidate active ingredients were obtained from the PubChem database and imported into Chem3D 22.0.0 software. Geometric optimization and energy minimization were performed using the MM2 molecular mechanics force field to eliminate high-energy unstable conformations. The three-dimensional structures of the core target proteins

were sourced from the PDB database: NT5E (PDB ID: 4H1S), PLAT (PDB ID: 1A5H), and LGMN (PDB ID: 4AW9); the monomer models for SPP1 (AF-Q9BX95-F1-model_v6) and BLK (AF-P51451-F1-model_v6) were predicted using AlphaFold2.

Binding affinity was evaluated based on the ligand–receptor binding free energy. Binding energies less than −4.25 kcal/mol indicated stable binding; energies below −5.0 kcal/mol suggested good efficacy; and values below −7.0 kcal/mol indicated the presence of very strong interactions. The docking results show (Figure 11A-F) that the binding energies of 27 candidate active components with 5 core target proteins are all less than −4.25 kcal/mol, suggesting that stable complexes can be formed. Among them, Dihydrotanshinlactone and Tanshinone IIA exhibit binding energies as low as −10.6 kcal/mol with NT5E, indicating very strong interactions with the NT5E pocket (Figure 11D-E). This suggests that the active components derived from Danshen are key pharmacological substances for targeting NT5E and intervening in COPD.

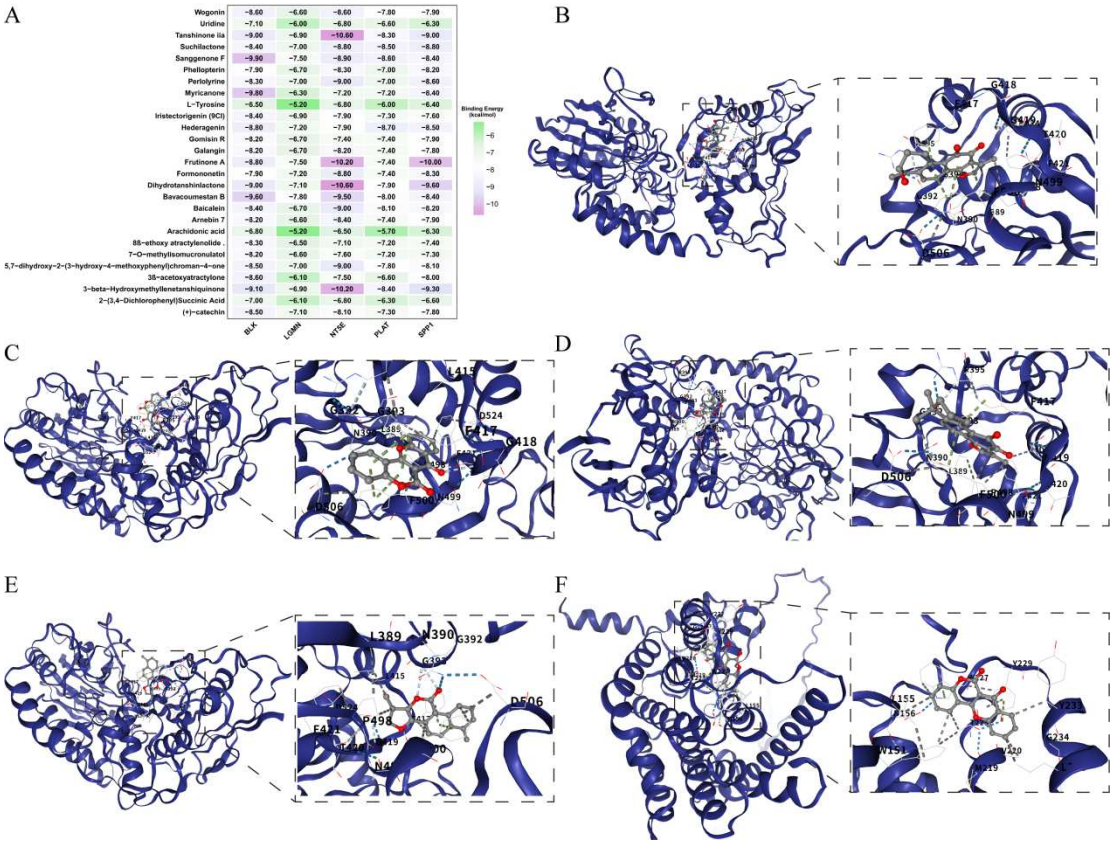

Figure 11 Molecular docking results.

A: Heatmap of the docking of the candidate active ingredient FZN with five core target proteins. B: Docking mode of NT5T with 3-beta-Hydroxymethyllenetanshiquinone. C: Docking mode of NT5T with Frutinone A. D: Docking mode of NT5E with Tanshinone IIA. E: Docking mode of NT5E with Dihydrotanshinlone. F: Docking mode of SPP1 with Frutinone A.

2.17 Molecular Dynamics Simulation

Through a 100 ns molecular dynamics simulation, the dynamic binding characteristics of the NT5E-Tanshinone IIA compound were analyzed (Figure 12A–G). The RMSD curve indicates that neither the free NT5E protein nor the ligand-protein complex exhibited significant mutation fluctuations. After a brief relaxation period, the system quickly reached equilibrium, suggesting

that the binding of Tanshinone IIA does not induce drastic conformational changes in the protein backbone (Figure 12A). The radius of gyration (Rg) results showed that after binding with the ligand, both the total Rg and the Rg along each axis exhibited minor fluctuations without sustained increases or decreases, indicating that the protein did not undergo unfolding and the overall tertiary structure remained compact (Figure 12B). RMSF analysis suggested that the flexibility of the residues in the binding pocket region was relatively low. After Tanshinone IIA embedded into the cavity, it could anchor the local conformation and reduce the fluctuations of the residues, ensuring the structural stability of the compound (Figure 12C).

The SASA maintained a slight oscillation between 215 – 235 nm^2 after a rapid adjustment in the early simulation phase, further confirming that the protein conformation tended to stabilize (Figure 12D). The number of hydrogen bonds fluctuated between 400-460, with the average number of hydrogen bonds in the mid to late stages being higher than in the early stages, and there was no phenomenon of complete hydrogen bond disappearance, indicating that the two could continuously maintain their binding through hydrogen bonds (Figure 12E). The two-dimensional and three-dimensional FEL free energy surfaces exhibited distinct low-energy conformational basins, with most conformations concentrated in the region of minimal energy, suggesting that this compound tended to form a dominant stable conformation (Figure 12F-G).

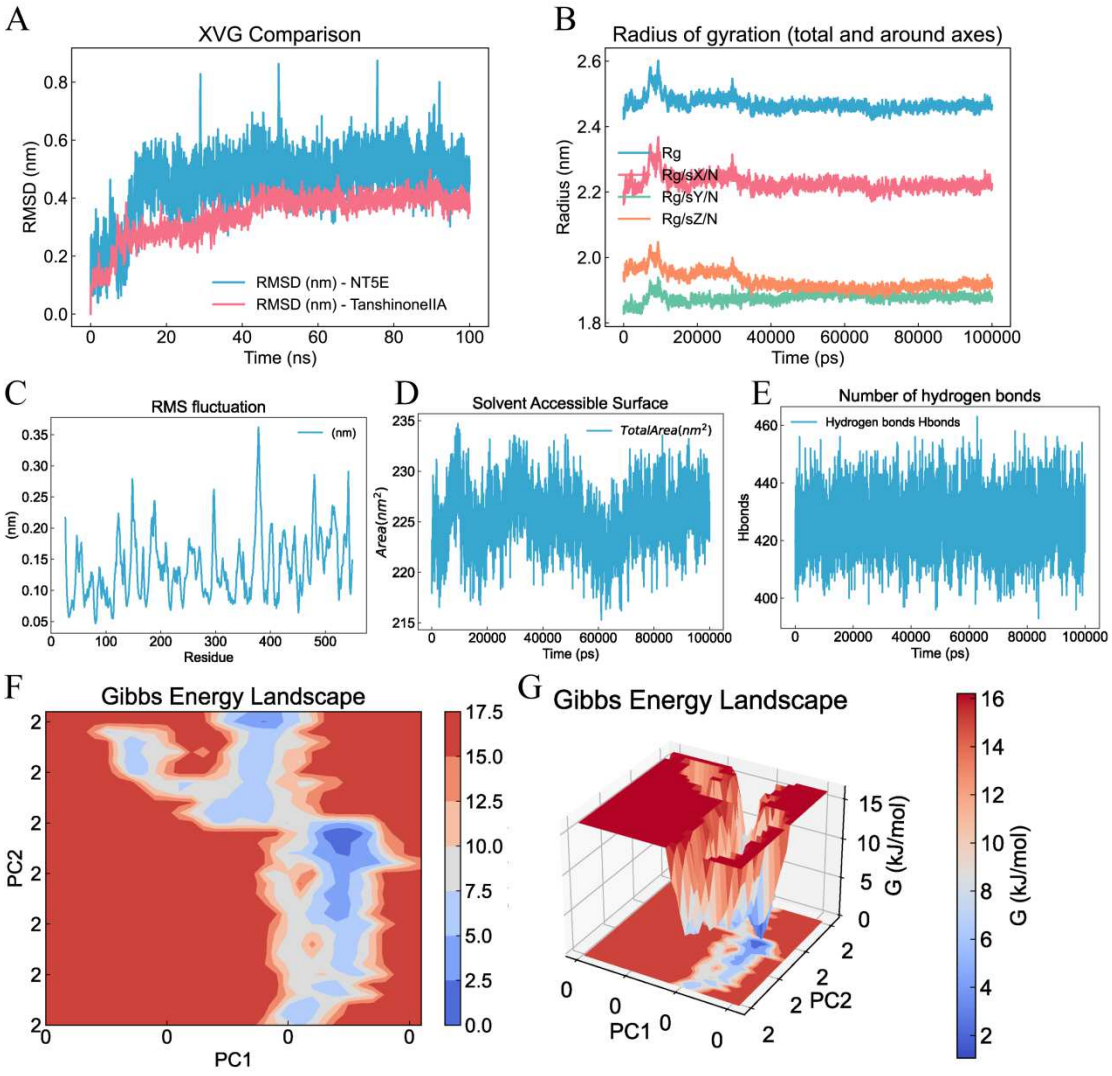

Figure 12 Molecular dynamics simulation analysis of the binding stability of Tanshinone IIA and NT5E protein.

A: Time-series change curve of RMSD. B: Dynamic evolution curve of gyration radius Rg. C: RMSF curve of root mean square fluctuation of residue levels. D: SASA time-series change curve. E: Distribution of the number of intermolecular hydrogen bonds as a function of simulation duration. F: Two-dimensional Gibbs FEL. G: Map of the 3D Gibbs free energy surface.

2.18 Effects of FZN on Inflammatory Factors in Serum and BALF of COPD Rats

To evaluate the anti-inflammatory effects of FZN in a COPD rat model, we measured the levels of IL-1β, IL-6, and TNF-α in serum and BALF. The results showed (Figure 13A-F) that, compared to the control group, the levels of IL-1β, IL-6, and TNF-α were significantly increased in the model group ($P < 0.0001$). In contrast, the levels of IL-1β, IL-6, and TNF-α in each treatment group were significantly reduced compared to the model group ($P < 0.0001$), with H-FZN demonstrating the most pronounced anti-inflammatory effect. This indicates that FZN improves the inflammatory response in COPD rats in a dose-dependent manner.

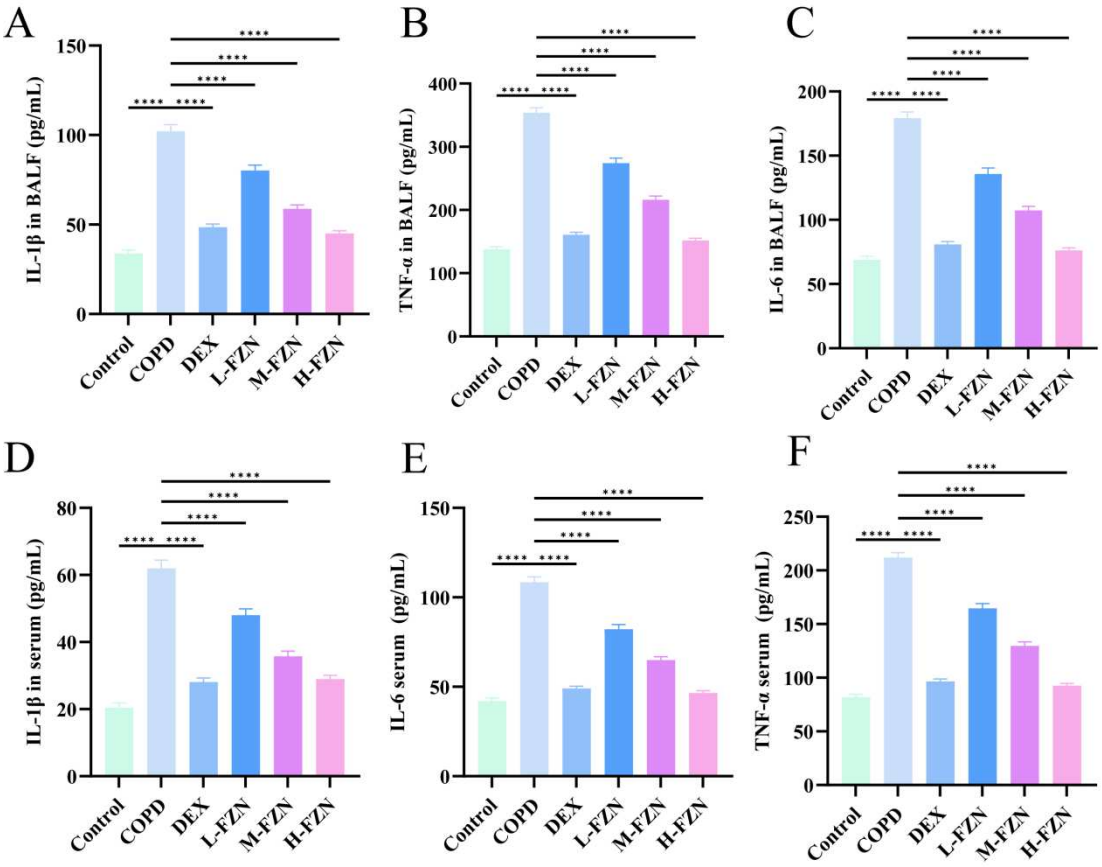

Figure 13 Effects of FZN on inflammatory cytokines in serum and BALF of COPD rats.

A-C: Levels of IL-1β, TNF-α, and IL-6 in the BALF of rats. D-F: Levels of IL-1β, TNF-α, and IL-6 in the serum of rats. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

2.19 Effects of FZN on the mRNA Expression of NE5T, SPP1, BLK, PLAT, and LGMN in COPD Rats

To validate the candidate genes identified through multi-omics screening and to explore the

molecular mechanisms of FZN action, we examined the mRNA expression of five core PCD-related genes (NT5E, SPP1, BLK, PLAT, and LGMN) in the lung tissues of rats. The results showed (Figure 14A-E) that, compared to the control group, the mRNA expression level of NT5E was significantly downregulated in the COPD model group ($P < 0.0001$), while the mRNA expression levels of SPP1, BLK, PLAT, and LGMN were significantly increased ($P < 0.0001$). After intervention with different doses of FZN, these trends in gene expression were significantly reversed, with the H-FZN showing the best effect. This indicates that FZN can dose-dependently regulate the abnormal expression of key genes in the lung tissues of COPD.

2.20 Effects of FZN on the Protein Expression of NT5E, SPP1, BLK, PLAT, and LGMN in COPD Rats

To further validate the changes in the expression of candidate genes at the protein level, this study employed Western blot analysis to detect the protein expression levels of NT5E, SPP1, BLK, PLAT, and LGMN in rat lung tissues (Figure 14F). The results indicated (Figures 14G-K) that, compared to the control group, the COPD model group exhibited a significant decrease in NT5E protein expression ($P < 0.0001$), while the protein expressions of SPP1, BLK, PLAT, and LGMN were significantly increased ($P < 0.0001$). Following intervention with different doses of FZN, the abnormal expression trends of the aforementioned genes were reversed ($P < 0.05$), with the regulatory effect of H-FZN being the most pronounced. This suggests that FZN can dose-dependently correct the abnormal expression of NT5E, SPP1, BLK, PLAT, and LGMN in the lung tissues of COPD rats at the protein level.

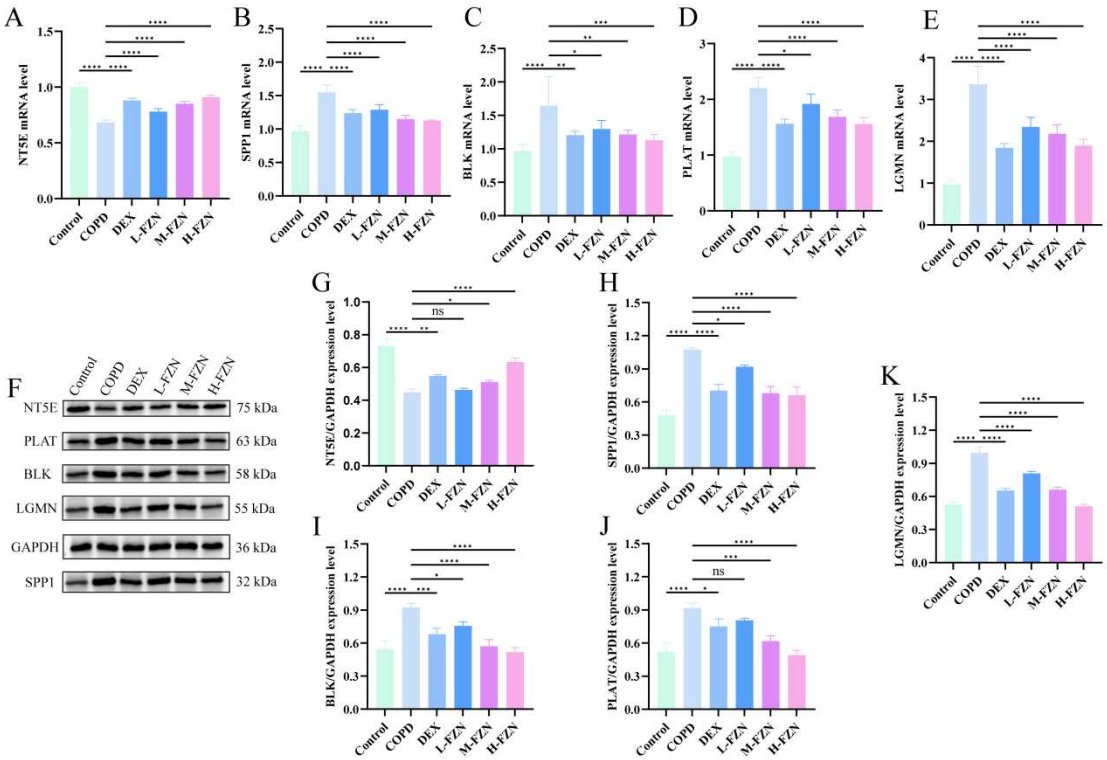

Figure 14 Effects of FZN on inflammatory cytokines in serum and BALF of COPD rats. A-E: mRNA expression levels of NT5E, SPP1, BLK, PLAT, and LGMN. F: Protein bands of NT5E, SPP1, BLK, PLAT, and LGMN. G-K: Quantitative analysis of the relative expression levels of NT5E, SPP1, BLK, PLAT, and LGMN proteins, with GAPDH as the internal reference

protein. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

3 Discussion

COPD, as the third leading cause of death globally among chronic respiratory diseases, is characterized by irreversible and persistent airflow limitation, chronic airway inflammation, and pulmonary tissue remodeling. Current clinical interventions can only alleviate symptoms and cannot halt the progressive damage to lung structure. Therefore, there is an urgent need to explore novel molecular targets that possess genetic causal evidence, clinical diagnostic value, and drug targetability to develop new strategies for multi-target natural compound interventions. PCD encompasses various regulatory modes of cell death, including apoptosis, pyroptosis, and ferroptosis, which are core biological processes maintaining the homeostasis of pulmonary epithelium and immune cells. The disruption of PCD pathways mediates persistent cellular damage and amplifies inflammation, which are central molecular mechanisms driving the progression of COPD. This study integrated two-sample MR, SMR multifactor analysis, multi-cohort transcriptomics, single-cell transcriptomics, network pharmacology, molecular docking, and molecular dynamics simulations, combined with in vivo animal validation experiments. It systematically elucidated the regulatory mechanisms of PCD-related genes in COPD from multiple dimensions, including genetic causal inference, transcriptomic feature analysis, single-cell microenvironment interactions, and targeted validation of traditional Chinese medicine compound intervention. This research also analyzed the multi-target molecular mechanisms of the clinical empirical prescription FZN in intervening COPD, providing a comprehensive theoretical foundation and candidate molecular targets for precise targeted prevention and treatment of COPD.

This study provided population genetic evidence for the causal relationship between PCD pathway dysregulation and susceptibility to COPD through MR analysis of blood cis-eQTL and cis-pQTL data. A robust genetic association was found between 439 PCD-related genes and COPD, supporting the view that abnormal cell death signaling was not merely a downstream consequence of chronic inflammation, but might serve as a genetically driven initiating or sustaining factor in the pathogenesis of COPD. By intersecting this genetic evidence with the action targets of FZN and genes identified through differential analysis of the COPD transcriptome, 16 key candidate genes were obtained. Further screening through various machine learning algorithms and SHAP value analysis identified five genes (NT5E, SPP1, BLK, PLAT, and LGMN) with the highest discrimination efficacy for COPD. Within this multimodal framework, these five core genes demonstrated potential mechanistic significance related to the pathogenesis of COPD, warranting further investigation.

NT5E (CD73) is a key glycoprotein on the cell membrane surface and serves as the rate-limiting enzyme for the generation of extracellular adenosine[23]. In the classical purinergic signaling axis, CD39 hydrolyzes extracellular pro-inflammatory ATP/ADP to AMP, and NT5E further catalyzes the conversion of AMP to adenosine. Adenosine regulates multiple downstream signaling pathways, including cAMP-PKA, MAPK, PI3K-Akt, and HIF-1α, by binding to G protein-coupled receptors[24, 25]. The adenosine signaling mediated by NT5E intersects with PCD regulation at multiple levels. Adenosine elevates intracellular cAMP through A2A receptors, activating the PKA pathway, which leads to the phosphorylation and inactivation of Bad, thereby upregulating anti-apoptotic proteins such as Bcl-2 to protect alveolar epithelial and endothelial

cells from apoptotic damage[26]. Meanwhile, A2B adenosine receptor signaling can also regulate the mTOR pathway to participate in the homeostasis of autophagy[27]. Studies have confirmed that CD73 can inhibit GSDMD-mediated pyroptosis via the A2B/NF-κB pathway, alleviating central inflammatory damage[28]. Clinical cohort studies have found that NT5E expression in lung tissues of non-obstructive smokers is significantly upregulated, whereas this compensatory upregulation effect disappears in COPD patients, suggesting that NT5E is a key protective molecule that blocks the transition of smoking damage to COPD. The imbalance of NT5E expression is involved in airway inflammation and epithelial barrier disruption in COPD[29].This study's single-cell results indicate that NT5E was predominantly expressed in endothelial cells, fibroblasts, epithelial cells, and B cells. MR analysis suggested that NT5E had a protective genetic effect against COPD. Animal experiments confirmed that NT5E mRNA and protein levels were significantly downregulated in the lung tissues of COPD model rats, while FZN intervention could significantly restore NT5E expression. Combined with molecular docking results, Tanshinone II A and Dihydrotanshinlactone exhibited a strong binding affinity to NT5E. Molecular dynamics simulations further confirmed that the conformation of the Tanshinone II A-NT5E compound remains stable over time, suggesting that the active components derived from Danshen in the FZN compound might target NT5E to restore adenosine signaling homeostasis and regulate the PCD pathway, thereby exerting a protective effect on the lungs.

SPP1 (osteopontin) is a secreted phosphoglycoprotein primarily secreted by alveolar macrophages, airway epithelial cells, and fibroblasts, participating in multiple biological processes such as immune regulation, PCD, and lung tissue remodeling[30-32]. SPP1 exhibits a bidirectional regulatory effect on PCD: on one hand, SPP1 can regulate the autophagy process through the PI3K/Akt-mTOR pathway[33]; on the other hand, SPP1 binds to integrin αvβ3, activating the PI3K/Akt and NF-κB pathways, downregulating Bcl-2, upregulating Bax, and thereby inducing apoptosis[34, 35]. Additionally, SPP1 is involved in the regulation of ferroptosis in macrophages, where the ubiquitination of SPP1 mediated by SMURF2 can inhibit the Nrf2-GPX4 pathway, promoting lipid peroxidation and the occurrence of ferroptosis[36, 37]. Clinical studies have confirmed that the expression of SPP1 is significantly elevated in the distal airways and sputum of COPD patients, and the expression level is positively correlated with disease severity[38]. Our single-cell analysis revealed that SPP1 was highly specifically expressed in macrophages. Cell communication analysis indicated that SPP1 acted as a key secretory ligand participating in the cellular interactions within the COPD lung microenvironment. In vivo experiments also confirmed that the expression of SPP1 mRNA and protein in the lung tissue of COPD model rats was significantly upregulated, and FZN intervention could reverse its abnormally high expression. Collectively, the evidence suggested that SPP1-mediated macrophage inflammation and PCD dysregulation were important aspects of lung injury in COPD.

BLK is a non-receptor tyrosine kinase belonging to the Src family, primarily enriched in B lymphocytes. It plays a crucial role in B cell development, differentiation, and the regulation of adaptive immune homeostasis, while also participating in the regulation of oxidative stress-mediated PCD[39]. Knockdown of BLK significantly reduces the expression of antioxidant and anti-apoptotic molecules such as HO-1, SOD, and Bcl-2, exacerbating oxidative stress-induced apoptosis[40]. Additionally, BLK can regulate the expression of LC3B-II, negatively modulating autophagic activity[39]. Clinical transcriptomic studies have found that BLK was significantly upregulated in the lung tissues of lung cancer patients with smoking-related

emphysema, accompanied by excessive activation of the IL-17 pathway, suggesting that BLK was involved in smoking-induced pulmonary inflammatory damage. Our single-cell data indicated that BLK was highly expressed specifically in B cells. Animal experiments confirmed that BLK expression was significantly elevated in the lung tissues of COPD models, and FZN intervention could downregulate BLK expression levels, indicating that BLK mediated the excessive activation of B lymphocytes and participated in the adaptive immune disorder and PCD imbalance in COPD.

The PLAT-encoded tissue-type plasminogen activator (tPA) is primarily synthesized by endothelial cells, and its core function is to regulate the fibrinolytic system and vascular homeostasis[41]. It also participates in the regulation of epithelial cell and immune cell PCD, serving as a key molecule linking pulmonary microvascular injury and airway inflammation[42, 43]. The absence of PLAT reduces the activity of the Erk1/2 signaling pathway, increases levels of reactive oxygen species (ROS) and lipid peroxidation, and enhances cellular sensitivity to apoptosis and ferroptosis stimuli[44]. In non-small cell lung cancer, the knockout of PLAT exacerbates mitochondrial damage and activates the caspase cascade, inducing apoptosis[45]. Epidemiological studies suggest that PLAT is a critical immune gene linking smoking-related osteoporosis and COPD[46]. The single-cell results indicated that PLAT was primarily expressed in endothelial cells and fibroblasts. Animal experiments confirmed that PLAT expression was significantly upregulated in the lung tissue of COPD model rats, and FZN intervention could reverse this abnormal change, indicating that PLAT-mediated microvascular endothelial injury was involved in the pathological process of COPD.

LGMN is an aspartic acid-specific cysteine endopeptidase that is predominantly expressed in epithelial cells, macrophages, monocytes, and fibroblasts. It plays a role in extracellular matrix degradation, regulation of inflammatory signaling, and remodeling of the immune microenvironment[47-49]. LGMN exhibits a bidirectional regulatory characteristic towards PCD: it can inhibit mechanical stress-induced ferroptosis through the integrin-Hippo-YAP1 pathway[50]. However, it can also promote the accumulation of ROS through the HSP90/GPX4 pathway, exacerbating ferroptosis while mediating endoplasmic reticulum calcium release that triggers apoptosis[51]. Clinical studies have found that LGMN+ macrophages are significantly associated with mortality in patients with respiratory failure[52]. In the process of diesel particle exposure-induced pulmonary inflammation, LGMN is involved in the release of pro-inflammatory cytokines[53]. This study confirmed through single-cell virtual knockout and overexpression bidirectional perturbation experiments that LGMN knockout significantly activated pro-inflammatory signaling pathways such as NF-κB, TNF, and IL-17, amplifying inflammatory cascade damage. Whereas LGMN overexpression enriched protective pathways such as alveolar surfactant synthesis, lipid homeostasis, and phagocytosis, suggesting a dual-edged sword effect of LGMN in COPD. The single-cell communication analysis revealed that macrophages with high expression of LGMN served as communication hubs in the COPD pulmonary microenvironment, mediating extensive inflammatory signaling. In vivo animal experiments confirmed a significant increase in LGMN mRNA and protein levels in the lung tissues of COPD rats, and FZN intervention could downregulate the abnormal expression of LGMN. In total, these results suggested that LGMN-mediated macrophage hyperactivation and PCD dysregulation were important driving factors in COPD-related lung injury.

In summary, the five core genes selected in this study (NT5E, SPP1, BLK, PLAT, and LGMN) are associated with various forms of PCD, including apoptosis, autophagy, and ferroptosis. In the

COPD microenvironment, the imbalance in the expression of these genes disrupts the PCD homeostasis of lung epithelial cells, endothelial cells, macrophages, and lymphocytes, leading to persistent airway inflammation, destruction of alveolar structures, microvascular damage, and airway remodeling. This provides a comprehensive explanation from the perspective of PCD regarding the molecular genetic mechanisms driving the pathogenesis of COPD due to the imbalance of PCD pathways. Furthermore, it offers solid theoretical support for the targeted regulation of these core genes by the FZN compound, aiming to restore cellular survival homeostasis and intervene in the progression of COPD.

The network pharmacology analysis system elucidated the material basis of the pharmacological effects of FZN, with core active components including Tanshinone II A, Dihydrotanshinone, Baicalin, and Psoralen B. Molecular docking indicated that these active ingredients could form stable bindings with the five core genes, among which Dihydrotanshinone and Tanshinone II A from Danshen exhibit the strongest binding affinity with NT5E. The study found that Tanshinone II A can reduce the levels of TNF-α, IL-6, IL-1β, and MDA in the lung tissues of COPD rats, increase SOD activity, and downregulate the mRNA and protein expression of PINK1 and Parkin, thereby alleviating lung injury in COPD. The underlying mechanism may be related to the reduction of inflammatory factor expression, inhibition of oxidative stress, and regulation of the mitochondrial autophagy PINK1/Parkin pathway[17]. In vivo animal experiments demonstrated that FZN could dose-dependently reduce the levels of inflammatory factors IL-1β, IL-6, and TNF-α in the BALF and serum of COPD rats. Simultaneously, it reversed the abnormal expression of NT5E, SPP1, BLK, PLAT, and LGMN mRNA and proteins, validating the predicted mechanisms of action from bioinformatics analysis and fully reflecting the characteristics of traditional Chinese medicine compounds with multi-component, multi-target, and multi-pathway holistic regulation.

This study still has certain limitations. First, the MR analysis exposure data is derived from peripheral blood eQTL/pQTL, and the lung tissue-specific genetic regulatory effects still require further validation using lung tissue eQTL datasets. Second, the virtual gene knockout/overexpression in single cells is based on computational simulations, and in vitro cell and in vivo animal gene intervention experiments are needed to corroborate the biological function of LGMN. Third, molecular docking has only completed static binding affinity assessments, and subsequent studies could complement this with molecular dynamics simulations to analyze the dynamic stability characteristics of the compounds. Fourth, the conclusions of this study are all based on bioinformatics analysis from public databases, and in vivo and in vitro pharmacodynamic experiments of the FZN compound and its core active components need to be conducted to verify the true regulatory effects of the compound on the PCD pathway and LGMN targets. Subsequent research will establish COPD cell and animal models, overexpress/silence LGMN to conduct functional experiments, and administer Tanshinone II A and FZN compound interventions to validate the regulatory mechanisms predicted by this study at the molecular, cellular, and tissue levels, providing experimental evidence for the clinical translation of FZN and the development of new targeted drugs for COPD.

4 Conclusion

This study integrates dual-sample MR-SMR genetic causal analysis, multi-cohort transcriptomics, WGCNA co-expression analysis, machine learning, single-cell transcriptomics,

cell communication, single-cell bidirectional virtual perturbation, network pharmacology, molecular docking-molecular dynamics simulations, and in vivo validation experiments in COPD rats. We identifies 439 PCD-related genes with robust genetic causal associations with COPD and further focuses on 16 core genes that exhibit genetic causality, disease differential expression, and are targetable by FZN. Among these, we identify five optimal COPD discriminative biomarkers: NT5E, SPP1, BLK, PLAT, and LGMN, and construct a five-gene combined nomogram to quantitatively assess individual COPD disease risk. Single-cell analysis indicates that there is a remodeling of cellular subpopulations in COPD lung tissues, with five core genes exhibiting cell-specific expression. Macrophages, epithelial cells, and fibroblasts act as communication hubs mediating the crosstalk of inflammation and airway remodeling signals. Furthermore, it is confirmed that LGMN has a bidirectional regulatory effect; the absence of LGMN activates the pro-inflammatory pathways of NF-κB, TNF, and IL-17, exacerbating lung damage, while high expression of LGMN can activate pathways related to alveolar lipid homeostasis and barrier repair. Network pharmacology and simulation analyses demonstrate that all 27 active components of FZN can stably bind to the aforementioned five target proteins, with Tanshinone IIA and Dihydrotanshinone showing the strongest binding affinity to NT5E. In vivo animal experiments validate that FZN can dose-dependently reduce the levels of inflammatory factors and reverse the abnormal mRNA and protein expression of core genes. In summary, this study establishes an integrated research system of "genetic causal screening - multi-omics validation - single-cell microenvironment analysis - traditional Chinese medicine compound targeted simulation - in vivo animal validation," clarifying that LGMN is a key regulatory target in COPD. It elucidates the multi-target mechanism by which FZN inhibits excessive PCD, alleviates chronic airway inflammation, and repairs alveolar structural damage by targeting core PCD genes, providing a theoretical basis and candidate pharmacological substances for the development of novel diagnostic biomarkers and targeted interventions with natural TCM compounds in COPD.

Competing interests： The authors declare no competing interests.

Clinical trial number： Not applicable.

Data availability statements： Data will be made available on request.

Ethics approval and consent to participate： The animal experiments in this study have been approved by the Laboratory Animal Welfare and Ethics Work Committee of Sichuan Laboratory Animal Society (SCAF20260618001).

Acknowledgments： We acknowledge support from Joint Innovation Fund of Health Commission of Chengdu and Chengdu University of Traditional Chinese Medicine, the Natural Science Foundation of Hubei Province, Major Science and Technology Special Projects in Hubei Province, the Chief Scientist Research Project of Hubei Shizhen Laboratory, Doctoral Program of China Association for Science and Technology, The National Natural Science Foundation of China, Sichuan Provincial Natural Science Foundation Project, and China Postdoctoral Science Foundation funded project.

Funding Statement： This research was supported by the National Natural Science Foundation of China [Grant No. 82505509], the China Postdoctoral Science Foundation [Grant No. 2024MD753905], the Doctoral Student Program of the Young S&T Talents Cultivation Project, the Natural Science Foundation of Hubei Province [Grant No. 2023AFD173], the Major Science and Technology Special Project of Hubei Province [Grant No. 2023BCA005], the Chief Scientist Research Project of Hubei Shizhen Laboratory [Grant No. HSL2024SX0002], the Natural Science Foundation of Sichuan Province [Grant No. 2025ZNSFSC1853], the Joint Innovation Fund of Health Commission of Chengdu Municipality and Chengdu University of Traditional Chinese Medicine [Grant No. WXLH202403266], and the "Xinglin Scholar" Scientific Research Promotion Program of Chengdu University of Traditional Chinese Medicine [Grant No. BSZ202413].

Author contributions：Cheng Luo: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Xiuzhi Peng:** Data curation, Methodology, Resources, Writing – original draft. **Bo Ning:** Methodology, Software, Visualization, Writing – original draft. **Jiayu Zhang:** Data curation,Writing – original draft. **Ziyi Jiang:** Data curation,Writing – original draft. **Junwen Tan:** Data curation,Writing – original draft. **Jiajie Li:** Data curation,Writing – original draft. **Fei Wang:** Data curation, Writing – original draft. **Ye Ying:** Data curation, Writing – original draft. **Yuanhang Ye:** Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Jia Ke:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing. **Wanting Qin:** Funding acquisition, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.

References

[1] COLLABORATORS G B D C O D.Global burden of 292 causes of death in 204 countries and territories and 660 subnational locations, 1990-2023: a systematic analysis for the Global Burden of Disease Study 2023[J].Lancet，2025，406（10513）：1811–1872.

[2] PAN M，ZHOU X.Airway remodeling in chronic obstructive pulmonary disease: characteristics and opportunities[J].Front Med (Lausanne)，2025，12: 1556868.

[3] WANG Z，LIN J，LIANG L，et al. Global, regional, and national burden of chronic obstructive pulmonary disease and its attributable risk factors from 1990 to 2021: an analysis for the Global Burden of Disease Study 2021 [J]. Respir Res. 2025: 2.

[4] FAN Y，CHEN Q，SUN H, et al. Economic evaluations of screening and case-finding for Chronic Obstructive Pulmonary Disease (COPD): a systematic review [J]. NPJ Prim Care Respir Med. 2026: 7.

[5] LUO C, WU X, ZHANG S, et al. Cuproptosis: a novel therapeutic mechanism in lung cancer [J]. Cancer Cell Int. 2025: 231.

[6] WU M, WANG W, CHEN Z, et al.Attenuated FTO induces necroptosis of alveolar epithelium via the m(6)A/CYP1B1/ROS/MLKL axis to promote the aggravation of pulmonary emphysema[J].Redox Biol，2026，95104237.

[7] HUANG A, LIU S, LAN J.Convergent death programs in chronic obstructive pulmonary disease: how pyroptotic and ferroptotic crosstalk reshapes therapeutic paradigms[J].Front Immunol，2026，171827181.

[8] ZHUANG Y, YAO L, QI W.Molecular Mechanisms of Programmed Cell Death Pathways in Chronic Obstructive Pulmonary Disease[J].J Cell Physiol，2026，241（5）：e70177.

[9] XUE J, WANG C, FAN H.The role of programmed cell death in chronic obstructive pulmonary disease: from pathogenesis to treatment[J].Front Immunol，2026，171715665.

[10] XIA H, WU Y, ZHAO J, et al. N6-Methyladenosine-modified circSAV1 triggers ferroptosis in COPD through recruiting YTHDF1 to facilitate the translation of IREB2 [J]. Cell Death Differ. 2023: 1293–1304.

[11] ZHAO J, XIA H, WU Y, et al. CircRNA_0026344 via miR-21 is involved in cigarette smoke-induced autophagy and apoptosis of alveolar epithelial cells in emphysema [J]. Cell Biol Toxicol. 2023: 929–944.

[12] SAAOUD F, XU K, LU Y, et al.Chronic obstructive pulmonary disease reprograms the lung into an immune organ through trained immunity, cell death networks, and immune checkpoint dysregulation[J].Front Med (Lausanne)，2026，131721780.

[13] Peng X, Li YH, Liang L, et al. Molecular mechanisms of programmed cell death in chronic obstructive pulmonary disease and research progress on traditional Chinese medicine intervention[J]. Chinese Journal of Experimental Traditional Medical Formulae, 2026, 32(3): 304-313. (in Chinese)

[14] Gong L, Liu T, Li RT, et al. Research progress on the mechanism of Astragalus and its compound prescriptions against chronic obstructive pulmonary disease[J]. Modern Chinese Medicine, 2026, 28(6): 1279-1288. (in Chinese)

[15] Peng LP, Zhang A, Tan W, et al. Effect of Astragalus–Morus alba containing serum on inflammation in human bronchial epithelial cells[J]. Traditional Chinese Medicine Journal, 2024, 23(9): 53-57. (in Chinese)

[16] Dou WJ, Guo SJ, Sun ZT. Research progress on the pharmacological mechanism of active components of Scutellaria baicalensis in the prevention and treatment of chronic obstructive pulmonary disease[J]. Global Traditional Chinese Medicine, 2026, 19(3): 653-658. (in Chinese)

[17] Ye XX, Yang C, Zhang YL, et al. Mechanism of tanshinone IIA in improving acute lung injury in rats with chronic obstructive pulmonary disease by regulating the PINK1/Parkin pathway[J]. Zhejiang Journal of Integrated Traditional Chinese and Western Medicine, 2025, 35(6): 519-524. (in Chinese)

[18] YE L, LONG C, XU B, et al. Multi-omics identification of a novel signature for serous ovarian carcinoma in the context of 3P medicine and based on twelve programmed cell death patterns: a multi-cohort machine learning study [J]. Mol Med. 2025: 5.

[19] LEI K, ZHAO Y, LI S, et al.Integrative spatial and single-cell transcriptomics elucidate programmed cell death-driven tumor microenvironment dynamics in hepatocellular carcinoma[J].Front Immunol，2025，16: 1589563.

[20] LI M M, YANG Y X, HUANG Y L, et al.Programmed cell death signatures-driven microglial transformation in Alzheimer's disease: single-cell transcriptomics and functional validation[J].Front Immunol，2025，161610717.

[21] LIU F, LUO J, YU P, et al. Uncovering the Targets of Pueraria Associated with Programmed Cell Death and the Construction of a Diagnostic Model in Septic Cardiomyopathy [J]. Biomedicines. 2026.

[22] LIU J, HU Z, XU Z, et al. Integrated Bulk and Single-Cell Transcriptomic Analysis Followed by Clinical Validation Reveal Programmed Cell Death-Related Shared Molecular Signatures in OA and MDD [J]. Int J Mol Sci. 2026.

[23] MYINT Y Y K, LIU A, GAO J, et al. hsa-let-7c-3p Restrains the Proliferation-to-Osteogenesis Transition of Human Periodontal Ligament Mesenchymal Stem/Stromal Cells via NT5E [J]. Stem Cell Rev Rep. 2026: 2863–2878.

[24] SATYAM S M, EL-TANANI M, ALGHOUL W I, et al. Adenosine Signaling as a Central Integrative Network in Cellular Stress Responses and a Therapeutically Actionable Target in Human Disease [J]. Biomolecules. 2026.

[25] ANTONIOLI L, BLANDIZZI C, PACHER P, et al. Immunity, inflammation and cancer: a leading role for adenosine [J]. Nat Rev Cancer. 2013: 842–857.

[26] PLIYEV B K, DIMITRIEVA T V, SAVCHENKO V G.Diadenosine diphosphate (Ap(2)A) delays neutrophil apoptosis via the adenosine A2A receptor and cAMP/PKA pathway[J].Biochem Cell Biol，2014，92（5）：420–424.

[27] GUO X, OKUKA M, SHORT B, et al.Spatially restricted ecto-5'-nucleotidase expression promotes the growth of uterine leiomyomas by modulating Akt activity[J].FASEB J，2024，38（19）：e70084.

[28] ZHUANG H, LEI W, WU Q, et al.Overexpressed CD73 attenuates GSDMD-mediated astrocyte pyroptosis induced by cerebral ischemia-reperfusion injury through the A2B/NF-kappaB pathway[J].Exp Neurol，2025，386115152.

[29] CARETA O, CUEVAS E, MUNOZ-ESQUERRE M, et al. Imbalance in the Expression of Genes Associated with Purinergic Signalling in the Lung and Systemic Arteries of COPD Patients [J]. Sci Rep. 2019: 2796.

[30] CHEN X, LI T, WANG J, et al.SPP1 promotes fatty acid synthesis in macrophages to drive silicosis[J].Cell Signal，2026，146112693.

[31] LIU H，COARFA C，CHARANIA A N，et al. Secreted Phosphoprotein 1 in Lung Diseases [J]. Metabolites. 2025.

[32] FU H, LIU X, SHI L, et al. Regulatory roles of Osteopontin in lung epithelial inflammation and epithelial-telocyte interaction [J]. Clin Transl Med. 2023: e1381.

[33] LIN H, CHENG S, YANG S, et al. Isoforskolin modulates AQP4-SPP1-PIK3C3 related pathway for chronic obstructive pulmonary disease via cAMP signaling [J]. Chin Med. 2023: 128.

[34] WANG D, JIANG K, FENG H, et al. Lapatinib Promotes Apoptosis in Hepatoma Cells by Regulating SPP1 Expression [J]. Anticancer Agents Med Chem. 2026.

[35] WANG W, DONG Y, YI H, et al. Single-Nucleus RNA Sequencing Reveals SPP1(+) Macrophages Induce Cardiomyocyte Apoptosis to Promote Atrial Fibrillation Susceptibility [J]. J Cardiovasc Dev Dis. 2026.

[36] CAO J, YANG X, GUO J, et al.Based on WGCNA and machine learning studies, SMURF2 drives NSCLC malignant transformation, ferroptosis, and macrophage polarization by ubiquitinating SPP1[J].Gene，2026，994150093.

[37] LIU P, WU D, DUAN J, et al.NRF2 regulates the sensitivity of human NSCLC cells to cystine deprivation-induced ferroptosis via FOCAD-FAK signaling pathway[J].Redox Biol，2020，37101702.

[38] PAPAPORFYRIOU A, LOUKIDES S, KOSTIKAS K, et al.Increased levels of osteopontin in sputum supernatant in patients with COPD[J].Chest，2014，146（4）：951–958.

[39] ZHOU R S, ZHAO J Z, GUO L M, et al. The novel antitumor compound clinopodiside A induces cytotoxicity via autophagy mediated by the signaling of BLK and RasGRP2 in T24 bladder cancer cells [J]. Front Pharmacol. 2022: 982860.

[40] ZHANG F, WANG T, WEI L, et al. B-Lymphoid Tyrosine Kinase Crosslinks Redox and Apoptosis Signaling Networks to Promote the Survival of Transplanted Bone Marrow Mesenchymal Stem Cells [J]. Research (Wash D C). 2025: 0660.

[41] LI W, HE J, LI J, et al.High-fat diets influence breast cancer progression by modulating CAF-immune cell interactions through PLAT signaling[J].J Pathol，2025，267（3）：249–260.

[42] YANG R, ZHENG T, XIANG H, et al.Lung single-cell RNA profiling reveals response of pulmonary capillary to sepsis-induced acute lung injury[J].Front Immunol，2024，15: 1308915.

[43] VELIKANOVA E A, SINITSKY M Y, SINITSKAYA CAPITAL A C, et al.Evaluation of the Feasibility of Endothelial Colony-Forming Cells to Develop Tissue-Engineered Vascular Grafts Based on the Gene Expression Profile Analysis[J].Sovrem Tekhnologii Med，2022，14（3）：15–19.

[44] HE X, ZHANG H, WANG Y, et al. Plat safeguards maternally aged oocytes against programmed cell death through activating the Erk1/2 pathway [J]. Sci China Life Sci. 2026: 506–522.

[45] YAN M, WANG W, ZHOU J, et al. Knockdown of PLAT enhances the anticancer effect of gefitinib in non-small cell lung cancer [J]. J Thorac Dis. 2020: 712–723.

[46] WANG H, LI S, CHEN B, et al. Exploring the shared gene signatures of smoking-related osteoporosis and chronic obstructive pulmonary disease using machine learning algorithms [J]. Front Mol Biosci. 2023: 1204031.

[47] XIA Y, ZHOU F, MA B, et al.LGMN promotes crosstalk between macrophages and fibroblasts in pulmonary fibrosis: a potential therapeutic target[J].Front Immunol，2026, 17: 1789907.

[48] LI Z, ZHANG Y, YANG K, et al. Monocyte-Derived LGMN(+) Macrophages Divert Lung Injury Outcomes toward Fibrosis through Matrix Remodeling [J]. Research (Wash D C). 2026: 1341.

[49] ZHANG W, CHEN D, LI J, et al.LGMN(+) macrophage promotes the formation of a tumor-supportive microenvironment in gastric cancer[J].Front Immunol，2026，17: 1810794.

[50] WANG P, XIE Z, DENG L, et al. Macrophage-derived legumain ameliorates excessive mechanical stress-induced ferroptosis of nucleus pulposus cells and intervertebral disc degeneration via integrin alphavbeta3-Hippo signaling [J]. Cell Mol Biol Lett. 2026.

[51] LEE J K, RIABOWOL K, LEE K Y. Lgmn targets two distinct GPCRs, PAR2 and micro-OR1, and induces cell death in acute lymphoblastic leukemia through an intracellular Ca(2)(+) imbalance triggered by ER Ca(2)(+) release [J]. Cell Death Discov. 2026.

[52] MORRELL E D, HOLTON S E, LAWRANCE M, et al. The transcriptional and phenotypic characteristics that define alveolar macrophage subsets in acute hypoxemic respiratory failure [J]. Nat Commun. 2023: 7443.

[53] MCCARRICK S, MALMBORG V, GREN L, et al. Pulmonary exposure to renewable diesel exhaust particles alters protein expression and toxicity profiles in bronchoalveolar lavage fluid and plasma of mice [J]. Arch Toxicol. 2025: 797–814.

Supplementary Files

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

- TheuncroppedGelsandBlotsimage.docx