


Integrating Mendelian Randomization and Machine Learning to Identify Hypoxia-Related Diagnostic Biomarkers and Causal Relationship in COPD

Wenhui Fu, Yangli Liu, Renjie Li, Haiying Jin 

Department of Respiratory Medicine, Jinyun People's Hospital, Lishui, Zhejiang, 321400, People's Republic of China

Correspondence: Haiying Jin, Department of Respiratory Medicine, Jinyun People's Hospital, No. 299 North Ziwei Road, Wuyun Street, Lishui, Zhejiang, 321400, People's Republic of China, Tel +86 13754278979, Email jinhaiying8979@yeah.net

Background: Chronic obstructive pulmonary disease (COPD) involves progressive lung function decline, with hypoxia playing a key pathogenic role. However, systematic investigations focusing on hypoxia-related genes (HRGs) in COPD remain limited.

Methods: We applied machine learning to identify HRG-associated diagnostic biomarkers and evaluated their performance via Receiver Operating Characteristic (ROC) analysis. Mendelian randomization (MR) was conducted to assess causal relationships between candidate genes and COPD. A nomogram model was constructed to evaluate clinical utility, and a ceRNA network was developed using ENCORI database.

Results: Six HRG-based diagnostic biomarkers were identified, including *SLC2A1*, which demonstrated strong diagnostic value (AUC > 0.8). MR analysis revealed a significant causal effect of *SLC2A1* expression on COPD risk (OR = 1.32, 95% CI: 1.02–1.71, P < 0.05). Functional evidence suggests *SLC2A1* promotes hypoxia-induced metabolic reprogramming in airway epithelial cells. The constructed nomogram showed good clinical applicability. ceRNA analysis highlighted *MALAT1*, *NEAT1*, and *XIST* as potential upstream regulators.

Conclusion: Our findings identify *SLC2A1* as a causal and diagnostically relevant gene in COPD, offering novel insight into hypoxia-driven disease mechanisms and supporting future personalized therapeutic strategies.

Keywords: chronic obstructive pulmonary disease, hypoxia-related genes, Mendelian randomization, machine learning

Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory airway disorder and one of the most prevalent respiratory conditions, currently recognized as the third leading cause of death worldwide.^{1,2} The key pathological hallmark of COPD is persistent airflow limitation, which progressively impairs lung function and leads to irreversible airway damage. Common symptoms include chronic cough, sputum production, chest tightness, dyspnea, and respiratory distress.³ As the disease progresses, patients experience a gradual decline in their ability to work and perform daily activities, which significantly reduces their quality of life and imposes a growing economic burden. Consequently, COPD has become a critical global public health concern.⁴ Given these challenges, advancing our understanding of COPD pathogenesis and identifying novel biomarkers are critical for improving therapeutic strategies and enhancing patient outcomes.

Hypoxia-inducible factors (HIFs) play a pivotal role in the pathogenesis of various diseases, including cardiovascular conditions and metabolic disorders.^{5,6} In particular, hypoxic conditions have been implicated in exacerbating the progression of COPD.⁷ In COPD patients, chronic hypoxia not only drives persistent inflammation in the airways, lung parenchyma, and pulmonary vasculature, but also induces a systemic inflammatory response.⁷ This inflammatory state is further amplified by hypoxia-triggered neutrophil elastase, which exacerbates tissue damage. Lodge et al demonstrated that hypoxia intensifies neutrophil-mediated endothelial injury in COPD patients, thereby compounding disease severity.⁸ Additionally, studies have shown that hypoxia-inducible factor (HIF)-1 α is significantly overexpressed

in COPD patients, highlighting its potential as a therapeutic target.⁹ Previous studies have reported significant differential expression of hypoxia-related genes, such as *CXCL9* and *CXCL12*, between patients with COPD-associated pulmonary hypertension (COPD-PH) and healthy controls.¹⁰ Despite these findings, the diagnostic value and causal roles of these genes remain unconfirmed. This gap motivates our integrated approach. While hypoxia has been widely recognized as a key factor in COPD progression, systematic investigations specifically targeting hypoxia-related genes (HRGs) remain scarce. Further research into the interplay between hypoxia and COPD is crucial for enhancing our understanding of the disease's pathogenesis and advancing effective therapeutic strategies.

Bioinformatics technologies enable comprehensive analyses of disease mechanisms by integrating vast amounts of omics data, providing valuable insights into the processes underlying disease development.¹¹ Among these advancements, machine learning, a pivotal branch of artificial intelligence, has demonstrated significant potential when combined with modern precision medicine. This integration facilitating the identification of intricate relationships between genes and diseases, providing critical support for early diagnosis, prognosis evaluation, and personalized treatment strategies.¹¹ Mendelian randomization (MR) is an emerging epidemiological approach that allows precise assessments of potential causal relationships between exposure factors and outcomes.¹² By leveraging genetic variants as instrumental variables, MR minimizes the influence of environmental risk factors, reducing confounding bias and enhancing the reliability of causal inferences.¹² Unlike conventional observational methods, MR addresses confounding and reverse causation, enabling more reliable causal inference. Existing studies are often restricted to conventional differential expression or single-gene analyses, which are insufficient to capture the complex regulatory landscape of HRGs in COPD pathogenesis. Compared with traditional experimental approaches, the integration of high-throughput genomic data and machine learning enables more comprehensive and efficient identification of potential diagnostic genes and evaluation of their clinical relevance, thereby addressing limitations in target discovery and causal inference observed in previous studies.

In this study, we employed bioinformatics techniques and machine learning algorithms to identify diagnostic genes for COPD that are strongly associated with HRGs. Additionally, we combined the risk scores of key feature genes to develop a comprehensive nomogram, which facilitates effective risk stratification for COPD. Importantly, MR analysis was employed to assess the causal relationship between the identified diagnostic genes and COPD. Our findings provide valuable insights into the pathogenesis of COPD and establish a theoretical foundation for the precision medicine in the treatment of COPD patients.

Materials and Methods

Data Acquisition

The microarray data of COPD tissues (GSE19407) were retrieved from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), with 10 normal and 40 COPD samples designated as the training set. The GSE10006 dataset, consisting of 26 normal and 46 case samples, was used as the validation set. A total of 200 HRGs were obtained from the Molecular Signatures Database.¹³

Identification and Preliminary Analysis of Differentially Expressed Genes

The GEO datasets were normalized using the `normalizeBetweenArrays` function from the “limma” package to ensure consistency in expression levels across samples. Differential expression analysis of the GSE19407 dataset was performed using the “limma” package ($|\log \text{fold change (FC)}| > 1$, $\text{adj.p-value} < 0.05$) to identify differentially expressed genes (DEGs, [Supplementary Table S1](#)). Volcano plots and heatmaps were generated for the DEGs, with the heatmap visualizing the top 20 DEGs based on $|\log \text{FC}|$ values. The intersection of DEGs with the HRGs set was used to identify differentially expressed hypoxia-related genes (DEHRGs). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted on the DEHRGs. The protein-protein interaction (PPI) network of the DEHRGs was constructed using STRING (confidence threshold > 0.4), and the network was visualized with Cytoscape v3.10.2.

Screening of Diagnostic Feature Genes for COPD

Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis of the candidate genes was performed using the “glmnet” package, with cross-validation employed to select the optimal penalty parameter for removing highly correlated genes and reducing model complexity. Support Vector Machine-Recursive Feature Elimination (SVM-RFE) analysis was conducted using the “caret” package. Additionally, random forest (RF) analysis was conducted using the “randomForest” package, with the top 10 most important genes were selected based on the MeanDecreaseGini importance ranking. The results derived from the three algorithms are provided in [Supplementary Table S2](#).

Construction and Performance Evaluation of the Diagnostic Model

The “rms” package was used to construct a nomogram and calibration curve for the selected feature genes. To evaluate the diagnostic performance of both the nomogram and the feature genes, the “pROC” package was employed to generate the Receiver Operating Characteristic (ROC) curve. Additionally, the “rmda” package was applied to create the Decision Curve Analysis (DCA) curve for the feature genes and nomogram. The Wilcoxon test was used to analyze the significance of feature gene expression levels.

Immune Landscape Characterization

CIBERSORT analysis was conducted with the “IOBR” package to generate bar plots depicting immune cell abundance in control and COPD samples. The immune infiltration levels for both groups were calculated using the CIBERSORT algorithm, and the results were visualized with box plots. Additionally, the “ggcorrplot” package was employed to construct a correlation heatmap, illustrating the relationships between feature genes and differentially abundant immune cells.

Development of Competing Endogenous RNA (ceRNA) Network

The “multiMiR” package was used to identify microRNAs (miRNAs) interacting with feature genes by screening results from TargetScan and miRDB, and the intersection of predictions from both databases was extracted. Interaction data between Long non-coding RNAs (lncRNAs) and miRNAs were retrieved from ENCORI (<https://rnasyu.com/encori/>), and lncRNAs with a clipExpNum greater than 10 were selected. Visualization of the interactions was performed using the “ggsankey” package.

Analysis of Mendelian Randomization

Expression Quantitative Trait Locus (eQTL) data were retrieved from the MRC IEU OpenGWAS dataset, and outcome data were obtained from the same repository (GWAS ID: bbj-a-103). This Genome-Wide Association Study (GWAS) includes data from 3,315 patients and 201,592 control individuals. The detailed information on the data sources is provided in [Table 1](#).

Mendelian randomization analysis relies on three key assumptions: (1) Instrumental variables (IVs) must be strongly associated with the exposure. (2) IVs must not be correlated with confounders. (3) IVs can influence the outcome solely through the exposure. The criteria for selecting instrumental variables (IVs) are as follows: (1) Single Nucleotide Polymorphisms (SNPs) were selected as IVs. To maximize the number of SNPs available for subsequent analysis,

Table 1 The Detailed Source of the Mendelian Randomization Data

Phenotype	Year	ID	Population	Sample Size	Control	Case
SLC2A1	2018	eqtl-a-ENSG00000117394	European	31,470	/	/
SLC6A6	2018	eqtl-a-ENSG00000131389	European	31,684	/	/
DTNA	2018	eqtl-a-ENSG00000134769	European	31,684	/	/
TMEM45A	2018	eqtl-a-ENSG00000181458	European	31,470	/	/
CDKN1C	2018	eqtl-a-ENSG00000129757	European	14,263	/	/
ADM	2018	eqtl-a-ENSG00000148926	European	31,684	/	/
COPD	2018	bbj-a-103	European	204,907	3,315	201,592

a genome-wide significance threshold of $P < 1 \times 10^{-5}$ was applied; (2) SNPs were clumped to minimize the effects of linkage disequilibrium ($r^2 = 0.001$, region length = 10,000 kb); (3) The F-statistic was employed to assess weak IVs. A higher F-statistic indicates stronger instrument strength, and SNPs with an F-statistic greater than 10 were included to exclude weak instruments; (4) SNPs for both exposure and outcome were harmonized by aligning the allele directions. SNPs with ambiguous or incompatible allele directions were excluded from the analysis ([Supplementary Table S3](#)).

Cochran's Q test was initially applied to evaluate heterogeneity among SNPs, with p-value < 0.05 indicating significant heterogeneity. A random-effects inverse-variance weighted (IVW) model was used to account for variability. Horizontal pleiotropy was assessed using MR-Egger regression, with a statistically significant MR-Egger intercept (p-value < 0.05) suggesting the presence of horizontal pleiotropy. Additionally, MR-PRESSO was employed to identify outliers among the instrumental variables, and the MR analysis was repeated after excluding these outliers to evaluate their influence on causal estimates. To assess the influence of individual SNP on the causal relationship, a leave-one-out sensitivity analysis was conducted, where each SNP was excluded in turn to ensure that the MR results were not disproportionately affected by any single SNP.

Ethics Statement

This study was conducted using only publicly available, de-identified human datasets. According to Article 32 (Items 1 and 2) of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects (effective February 18, 2023, China), research that uses anonymized public databases without involving identifiable personal information or direct human interaction is exempt from additional institutional ethics review. And this study was approved by the Medical Ethics Review Committee of Jinyun People's Hospital (JYLL202504).

Results

Identification and Preliminary Analysis of Candidate Genes for COPD Diagnosis

To identify candidate genes significantly associated with HRGs for COPD diagnosis, and to comprehensively assess the pathogenesis of COPD, differential expression analysis was performed on the GSE19407 dataset. This analysis revealed a total of 3195 DEGs, including 1515 downregulated and 1680 upregulated genes. Volcano plot shown as [Figure 1A](#). The top 20 DEGs, ranked by $|\log_{2}FC|$, are shown in [Figure 1B](#). Subsequently, we intersected the DEGs with the HRGs, resulting in the identification of 45 DEHRGs ([Figure 1C](#)). To explore the interactions among these candidate genes, a PPI analysis was conducted, revealing a highly interconnected network consisting of 32 nodes and 103 edges. Notably, genes such as *SLC2A1*, *CDKN1C*, and *BCL2* exhibited strong interactions with multiple other genes ([Figure 1D](#)). These genes are known to play pivotal roles in hypoxia-related processes, including glucose transport (*SLC2A1*),¹⁴ cell cycle regulation (*CDKN1C*),¹⁵ and apoptosis inhibition (*BCL2*).¹⁶ Additionally, GO enrichment analysis indicated that the DEHRGs are predominantly associated with hypoxia-related biological processes ([Figure 1E](#)). KEGG pathways analysis revealed significant enrichment of DEHRGs in the HIF-1 signaling pathway ([Figure 1F](#)), emphasizing its critical role in mediating hypoxic responses and contributing to COPD pathogenesis.

Identification of Characteristic Genes for COPD Diagnosis

To identify diagnostic feature genes for COPD and eliminate non-essential genes, this study employed three machine learning algorithms: LASSO, SVM-RFE, and RF. LASSO regression was utilized to enhance model accuracy by regularizing the coefficients, thereby mitigating overfitting risk. This approach resulted in the selection of 16 feature genes ([Figure 2A and B](#)). Subsequently, SVM-RFE was applied, leading to the identification of 29 candidate genes ([Figure 2C](#)). RF was employed to identify crucial genes based on the MeanDecreaseGini metric, helping to mitigate overfitting. The top 10 genes were then selected ([Figure 2D and E](#)). Finally, by integrating the genes identified by all three algorithms, we pinpointed six key feature genes—*ADM*, *CDKN1C*, *DTNA*, *SLC2A1*, *SLC6A6*, and *TMEM45A* ([Figure 2F](#))—which may serve as potential biomarkers for COPD.

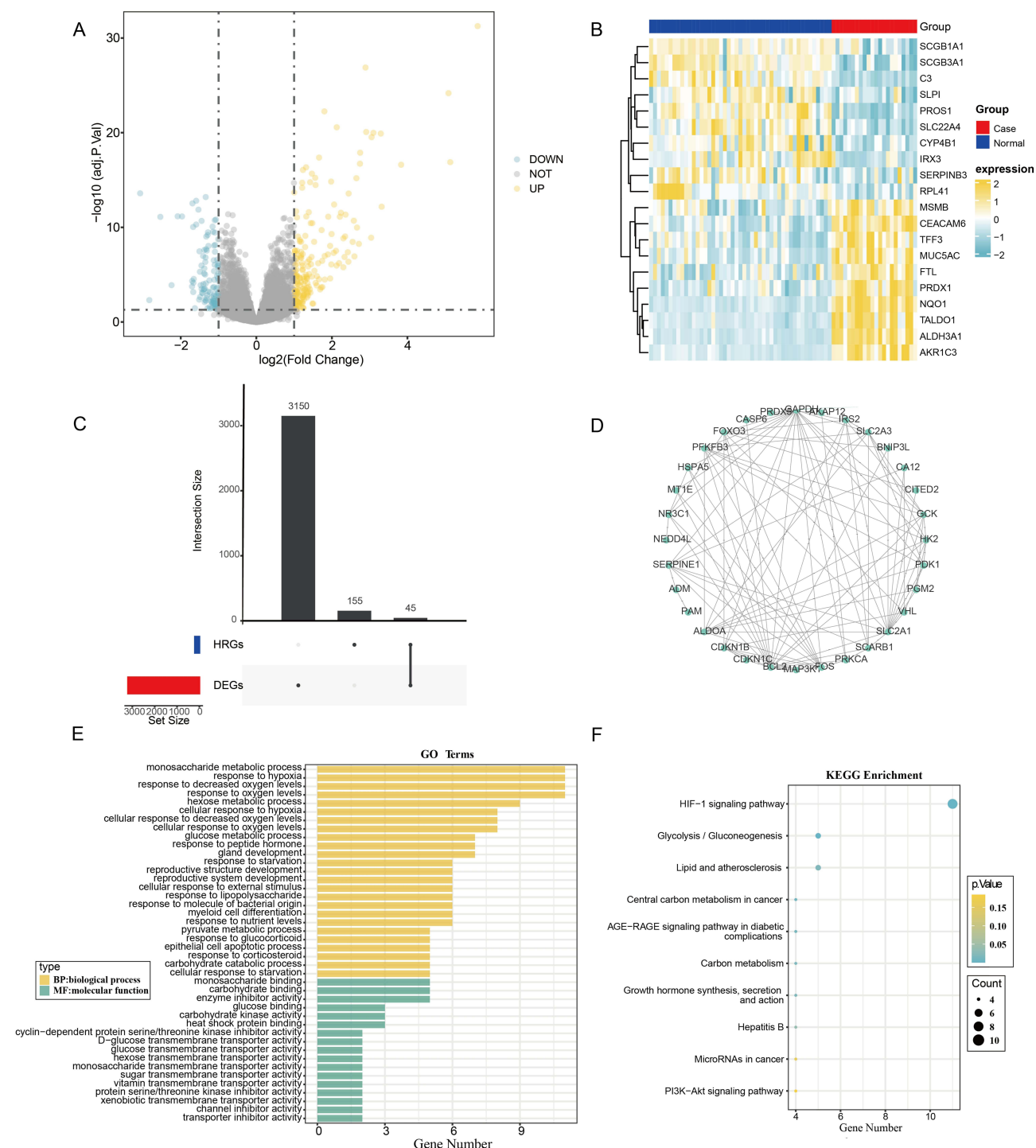


Figure 1 Screening and Preliminary Analysis of Genes Associated with COPD Diagnosis. **(A)** The volcano plot illustrates the differential expression genes. **(B)** The heatmap illustrates the differential expression of the top 20 DEGs, ranked by $\log_2(\text{FC})$. **(C)** The upset plot depicts the intersection between the DEGs and HRGs sets. **(D)** Construction of the PPI network for the DEHRGs. **(E)** GO enrichment analysis of DEHRGs. **(F)** KEGG pathway enrichment analysis for the DEHRGs.

Abbreviations: COPD, chronic obstructive pulmonary disease; DEGs, differentially expressed genes; HRGs, hypoxia-related genes; PPI, protein-protein interaction; DEHRGs, differentially expressed hypoxia-related genes; GO, gene ontology; KEGG, kyoto encyclopedia of genes and genomes.

Diagnostic Performance Evaluation and Expression Pattern Analysis of Feature Genes

To evaluate the diagnostic performance of the identified feature genes for COPD, we conducted ROC curve analysis using both the training and validation datasets. The AUC values of the ROC curves in the training set were all greater than 0.8, indicating that the feature genes exhibit excellent diagnostic performance, as demonstrated by the high

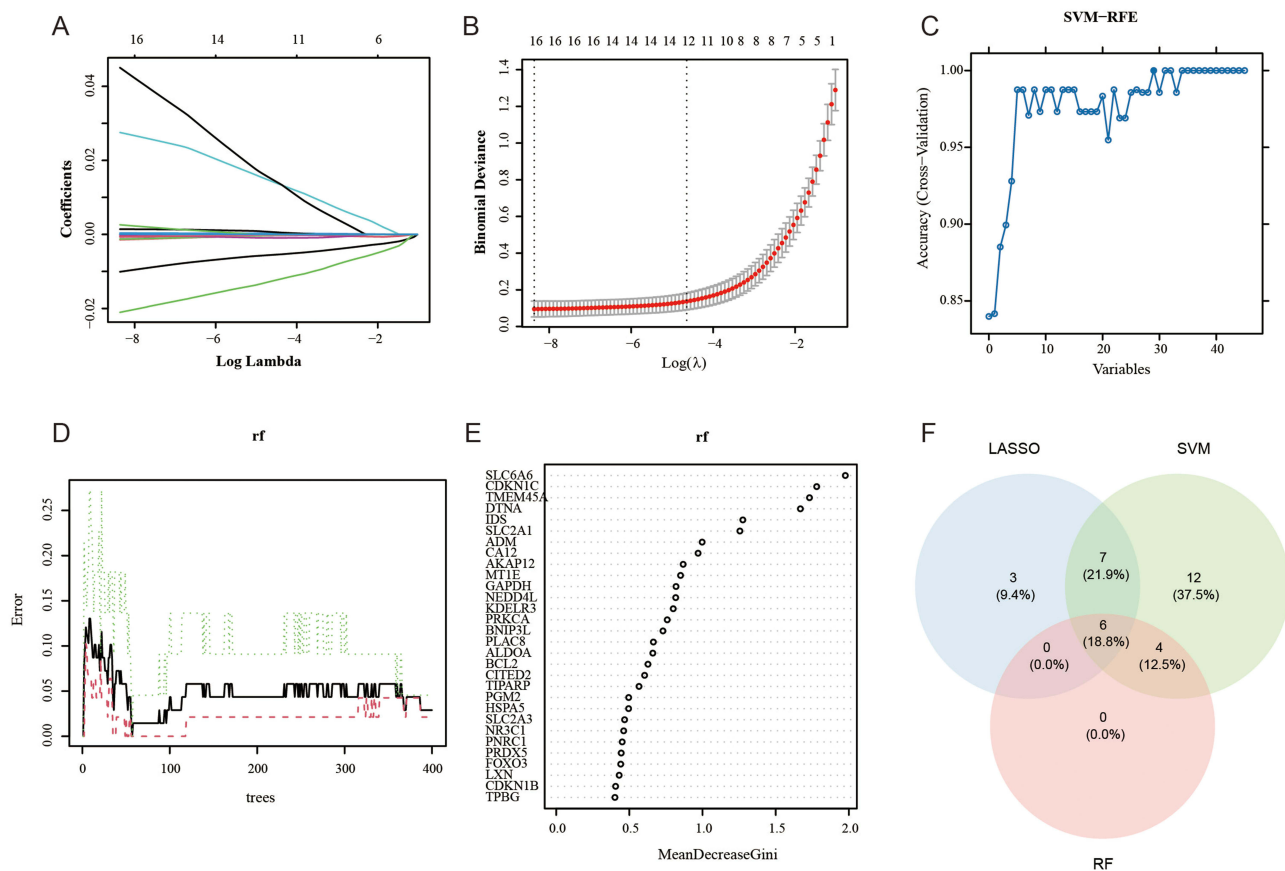


Figure 2 Identification of Potential Biomarkers for COPD Diagnosis. **(A)** LASSO coefficient regression path. **(B)** LASSO cross-validation error plot, where the horizontal axis represents the logarithmic values of the regularization parameter λ , and the vertical axis represents the cross-validation error. A smaller error indicates better predictive performance. **(C)** Relationship between generalization error and the number of features in SVM-RFE. **(D)** Residual distribution plot for RF, with the abscissa representing the number of subtrees and the ordinate representing the error. As the number of subtrees increases, the error gradually decreases. **(E)** Feature importance plot for RF. **(F)** Venn diagram showing the intersection of genes identified by integrating the three algorithms.

Abbreviations: LASSO, least absolute shrinkage and selection operator; RF, random forest; SVM-RFE, support vector machine-recursive feature elimination.

sensitivity and specificity (Figure 3A). Additionally, the AUC values in the validation set GSE10006 further confirmed the robust diagnostic capability of the feature genes, thereby validating the generalizability of the model (Figure 3B). Notably, analysis of the expression patterns of the feature genes revealed that, in the training set, *ADM*, *DTNA*, *SLC2A1*, and *SLC6A6* were significantly overexpressed in COPD tissues, while *CDKN1C* and *TMEM45A* were significantly underexpressed (Figure 3C). These findings were consistent across the validation set, further corroborating the robustness of these feature genes as diagnostic biomarkers (Figure 3D). These results suggest that the six feature genes may serve as potential diagnostic biomarkers for COPD, reflecting the underlying molecular processes of the disease.

Development and Assessment of the COPD Diagnostic Model

To facilitate clinical decision-making and improve risk assessment, we developed a comprehensive nomogram based on diagnostic feature genes. This nomogram enables rapid and accurate prediction of COPD risk by calculating a patient's risk score (Figure 4A). The diagnostic performance of the nomogram was assessed using ROC curves, which yielded an AUC of 1. This exceptional result indicates perfect discrimination capability within the study dataset; however, external validation is required to ensure generalizability and to rule out potential overfitting (Figure 4B). Calibration curve analysis showed excellent concordance between predicted and observed probabilities, confirming the model's reliability for clinical application (Figure 4C). Furthermore, DCA results the highlighted robust diagnostic performance of both the feature genes and the nomogram across a wide range of threshold probabilities, emphasizing their potential value in clinical decision-making (Figure 4D).

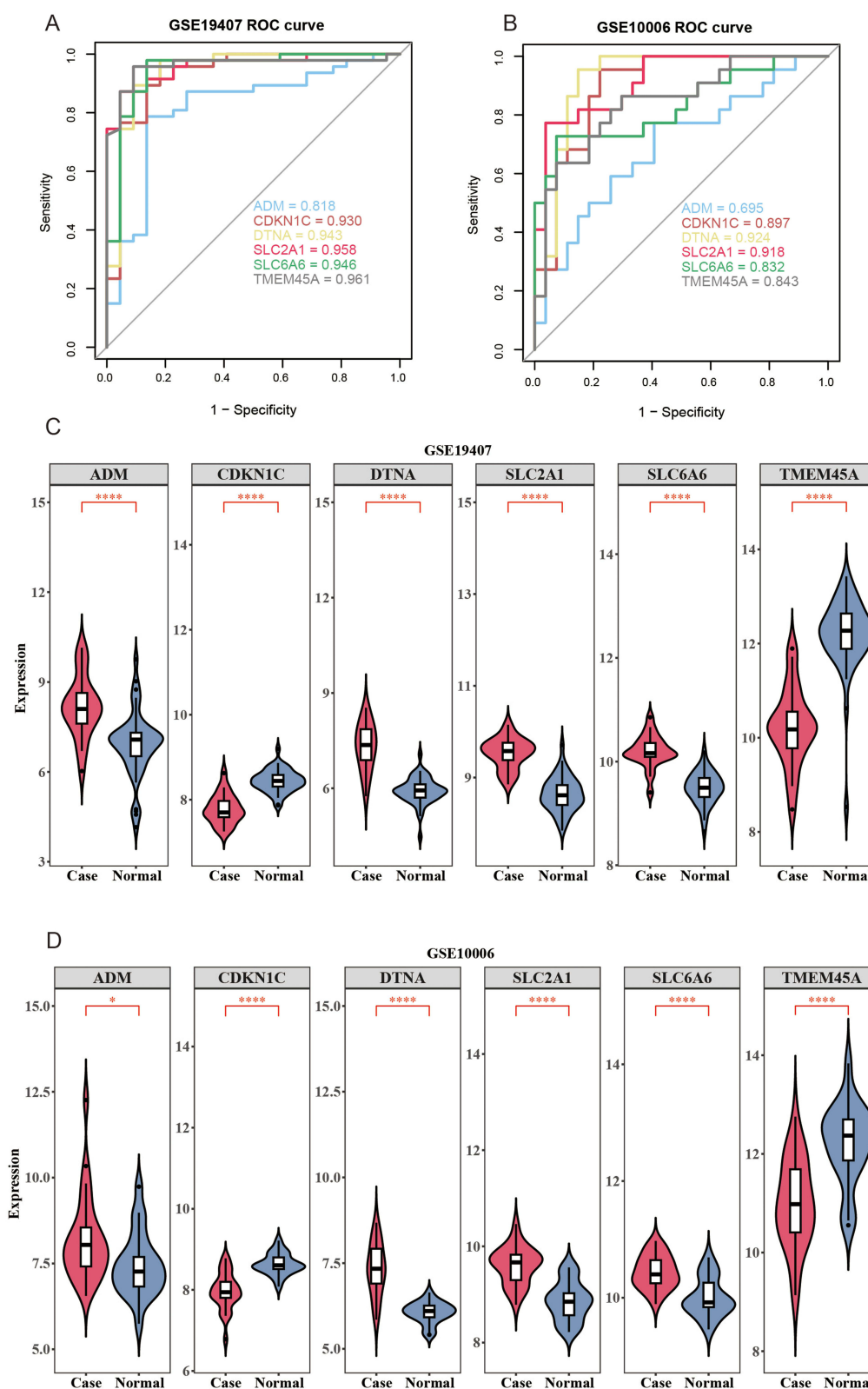


Figure 3 Assessment of Diagnostic Performance and Expression Profiles of Feature Genes. **(A)** ROC curve evaluation of feature genes in the training set GSE19407. **(B)** ROC curve analysis of the diagnostic performance of feature genes in the validation set GSE10006. **(C)** Expression patterns analysis of feature genes in the training set. **(D)** Comparative analysis of feature gene expression levels in the validation set. * $P < 0.05$, and **** $P < 0.0001$.

Abbreviation: ROC: receiver operating characteristic.

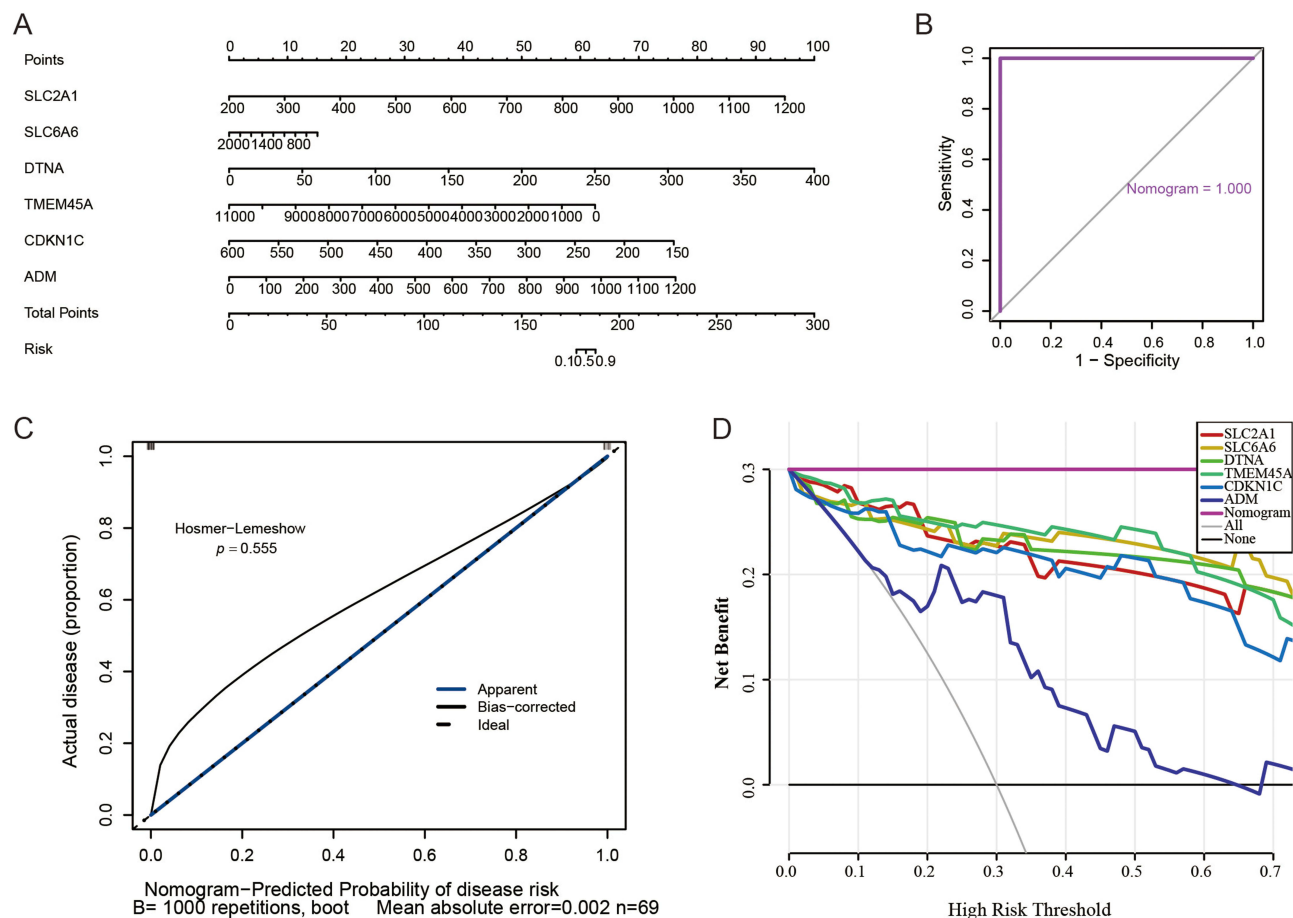


Figure 4 Establishment and Evaluation of the COPD Diagnostic Model. **(A)** Comprehensive risk assessment nomogram. **(B)** The findings of the ROC curve analysis for assessing the diagnostic performance of the nomogram. **(C)** Calibration curve assessment outcomes. **(D)** Visualization of the DCA curve results.

Abbreviations: COPD, Chronic obstructive pulmonary disease; ROC, receiver operating characteristic; DCA, decision curve analysis.

Exploration of the Immune Landscape in COPD

To further investigate the immune landscape in COPD and its underlying biological mechanisms, the CIBERSORT algorithm was employed. This computational tool deconvolutes gene expression profiles to estimate the relative proportions of 22 immune cell types in complex tissues (Figure 5A). Significant differences in immune cell composition were observed between normal and COPD samples, particularly in T cells CD4 memory activated, M0 macrophages, M2 macrophages, and resting dendritic cells. These findings highlight potential roles in the inflammatory and immune dysregulation that are characteristic of COPD (Figure 5B).

Spearman correlation analysis was conducted to investigate associations between feature genes and immune cell infiltration. Significant correlations were defined as $|\text{correlation}| > 0.3$ and $P < 0.05$, a threshold commonly used in similar studies to ensure biological relevance. The results, illustrated in Figure 5C, reveal significant correlations between feature genes and distinct immune cell types, suggesting potential mechanistic pathways that may contribute to COPD pathogenesis.

Establishment of the ceRNA Network for Feature Genes

To elucidate the post-transcriptional regulatory mechanisms of feature genes, we constructed a ceRNA network using data from TargetScan, miRDB, and ENCORI. The ceRNA network comprises 170 validated interactions, including 44 miRNAs and 17 lncRNAs. Notably, key lncRNAs such as *XIST*, *NEAT1*, and *MALAT1* were identified as regulators of multiple miRNAs, highlighting their pivotal roles in the network. This multilayered regulatory structure provides insights into potential mechanisms underlying COPD pathogenesis and positions these lncRNAs and miRNAs as promising targets for therapeutic intervention (Figure 6).

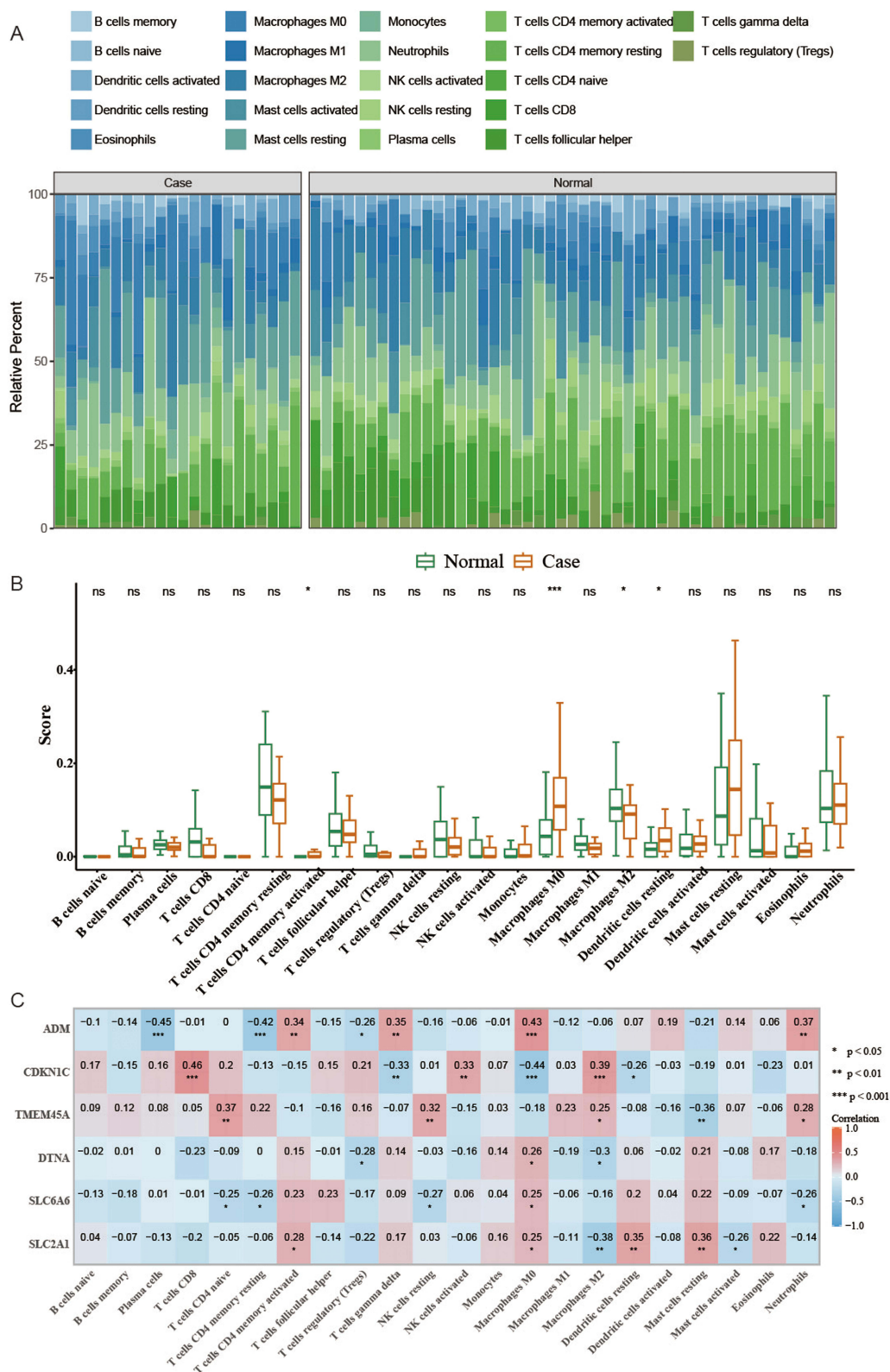


Figure 5 Characterization of the Immune Landscape in COPD. **(A)** Infiltration levels of 22 immune cell types in normal and COPD tissues were analyzed using the CIBERSORT algorithm. **(B)** The box plot presents a comparative analysis of immune cell infiltration levels across different sample groups. **(C)** Spearman correlation analysis was conducted to evaluate the relationship between feature genes and immune cell infiltration. ns indicates no significant difference, * signifies $P < 0.05$, ** denotes $P < 0.01$, and *** means $P < 0.001$.

Abbreviation: COPD, Chronic obstructive pulmonary disease.

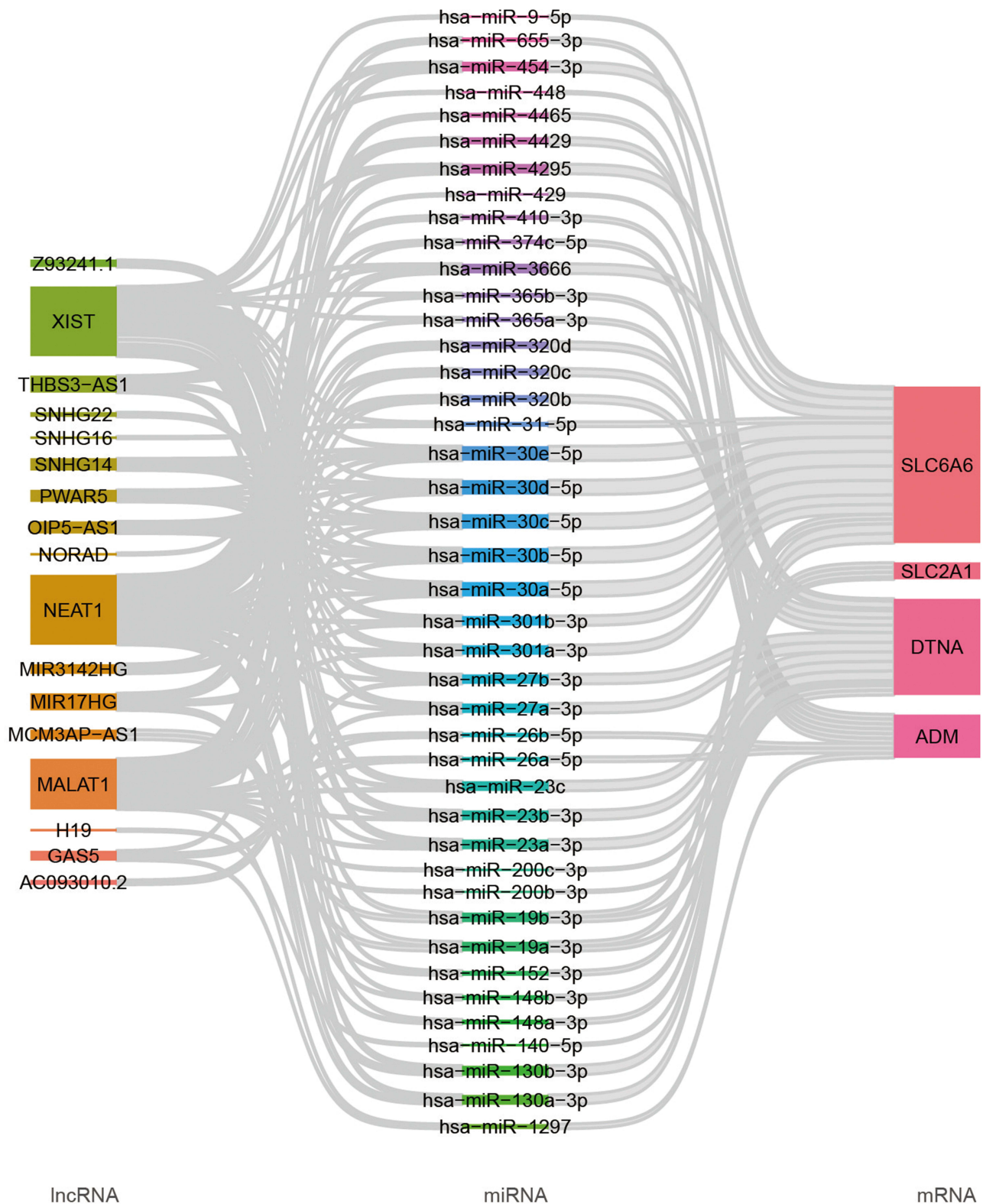


Figure 6 Construction of the ceRNA Regulatory Network for Feature Genes. The ceRNA network for the four feature genes comprises a total of 170 interaction pairs, involving 44 miRNAs and 17 lncRNAs.

Abbreviations: ceRNA, competing endogenous RNA; miRNA, microRNA; lncRNA, long non-coding.

Investigation of the Causal Association Between Feature Genes and COPD

Among the six identified feature genes, only *SLC2A1* demonstrated a statistically significant causal relationship with COPD based on MR analysis (OR = 1.32, 95% CI: 1.02–1.71, $P < 0.05$). The remaining genes did not reach MR significance thresholds, suggesting their roles should be considered hypothesis-generating and warrant further validation (Figure 7A). The MR scatter plot and forest plot further confirmed this causal relationship, demonstrating a positive correlation between *SLC2A1* and COPD (Figure 7B and C). Moreover, the funnel plot (Figure 7D) and leave-one-out sensitivity analysis (Figure 7E) further support the robustness of the findings. Notably, we performed Cochran's Q test to assess the heterogeneity between feature genes and COPD. The results indicated that all p-values exceeded 0.05, except for *TMEM45A*, suggesting no significant heterogeneity. Table 2 shows the results of the heterogeneity analysis. The p-value of Cochran's Q test and the intercept term of MR-Egger were not statistically significant ($P > 0.05$), suggesting no evidence of horizontal pleiotropy.

Discussion

The incidence of COPD has been steadily increasing, driven by factors such as aging, environmental pollution, declining lung function, and genetic predispositions.¹⁷ Hypoxia impairs cellular functions, leading to tissue damage and contributing to disease progression.¹⁸ Hypoxia is strongly associated with the onset of COPD, as demonstrated by significant upregulation of *HIF* expression in COPD patients.¹⁹ To address this, the present study integrates three machine learning techniques with MR analysis to identify potential diagnostic biomarkers of HRGs causally linked to COPD, with a specific focus on *SLC2A1*.

Using LASSO, SVM-RFE, and RF machine learning approaches, six characteristic genes were identified: *ADM*, *CDKN1C*, *DTNA*, *SLC2A1*, *SLC6A6*, and *TMEM45A*. These genes exhibited robust diagnostic performance. Based on these genes, a comprehensive nomogram was developed to facilitate COPD diagnosis. The nomogram converts complex statistical models into intuitive graphical representations, enabling clinicians to make timely and well-informed decisions. Using MR analysis, we provide evidence that *SLC2A1* acts as a causal contributor to COPD pathogenesis rather than being a mere consequence of the disease. These findings indicate the potential of *SLC2A1* as a therapeutic target for COPD. However, further experimental and clinical validation is warranted to confirm its translational applicability.

SLC2A1 encodes the *GLUT1* protein, which is located on the cell membrane and plays a pivotal role in glucose transport.²⁰ Dysregulated expression of *SLC2A1* has been implicated in the proliferation of multiple malignancies, including non-small cell lung cancer,²¹ hepatocellular carcinoma,²² and colorectal cancer.²³ Yao et al demonstrated that inhibition of *SLC2A1/GLUT1* activity by Isoginkgetin induces autophagy in hepatocellular carcinoma cells and suppresses cancer cell proliferation.²² Guan et al found significant upregulation of *SLC2A1* expression in osteoarthritis tissue, where inhibiting its expression through activation of the HIF-1 α pathway promotes apoptosis of osteoarthritis cells.¹⁴ Furthermore, upregulated *SLC2A1* expression has in colorectal cancer suggests its potential as a diagnostic biomarker for this disease.²³ Notably, Berg et al identified aberrant upregulation of *SLC2A1* in COPD,²⁴ which is consistent with our findings. We also observed significant upregulation of *SLC2A1* in COPD tissue. ROC curve analysis further substantiates the potential of *SLC2A1* as a diagnostic biomarker for COPD. These findings provide a theoretical foundation for a more comprehensive understanding of COPD pathogenesis and may facilitate the development of precision treatment strategies.

MR techniques confirmed a causal relationship between *SLC2A1* and COPD, demonstrating a positive correlation. Recent studies have established a causal link between certain gut microbiota and COPD, with gut microbial metabolites influencing *SLC2A1* expression through HIF-1 α regulation.^{14,25} This mechanism implies that gut microbiota regulate *SLC2A1* expression through their metabolites, contributing to COPD pathogenesis. Consequently, *SLC2A1* may serve as a potential therapeutic target for COPD treatment. Additionally, other biomarkers such as *CDKN1C* and *SLC6A6* may also play critical roles in COPD diagnosis and treatment.

CDKN1C encodes the cyclin-dependent kinase inhibitor.²⁶ Previous studies have shown that downregulation of *CDKN1C* expression is significantly associated with poor prognosis in thymic carcinoma.²⁷ *SLC6A6*, a gene involved in taurine transport, has been linked to poor prognosis in cancer due to its dysregulated overexpression.²⁸ Cao et al

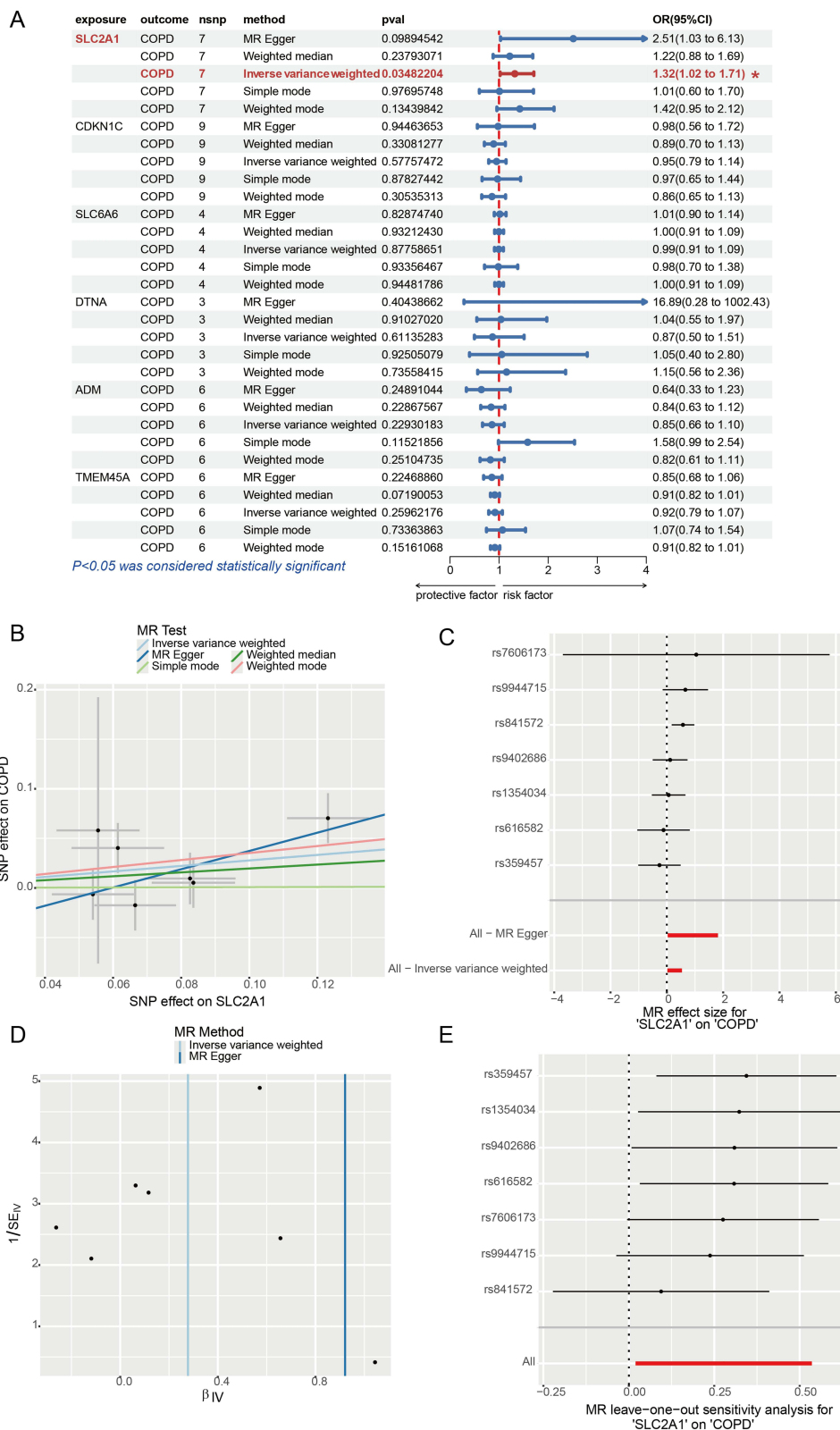


Figure 7 Examination of the Causal Relationship Between HRGs and COPD. **(A)** Forest plot representing MR results between feature genes and COPD. **(B)** The MR scatter plot for *SLC2A1* and COPD. **(C)** Forest plot for *SLC2A1* and COPD. **(D)** Funnel plot for *SLC2A1* and COPD. **(E)** Leave-one-out plot for *SLC2A1* and COPD. * denotes *P* < 0.05. Red-colored text highlights results that are statistically significant (*p* < 0.05).

Abbreviations: HRGs, hypoxia-related genes; COPD, Chronic obstructive pulmonary disease; MR, Mendelian randomization.

Table 2 Sensitivity Analysis of the Causal Association Between Feature Genes and COPD

Outcome	Exposure	Method	Cochran's Q	P. value	Egger Intercept	Standard Error	P. value
COPD	SLC2A1	MR Egger	4.302	0.507	-0.055	0.037	0.201
COPD	SLC2A1	IVW	6.468	0.373			
COPD	CDKN1C	MR Egger	6.369	0.497	-0.004	0.030	0.910
COPD	CDKN1C	IVW	6.383	0.604			
COPD	SLC6A6	MR Egger	0.943	0.624	-0.011	0.019	0.626
COPD	SLC6A6	IVW	1.269	0.737			
COPD	DTNA	MR Egger	0.068	0.794	-0.191	0.133	0.387
COPD	DTNA	IVW	2.136	0.344			
COPD	ADM	MR Egger	3.527	0.474	0.033	0.035	0.395
COPD	ADM	IVW	4.434	0.488			
COPD	TMEM45A	MR Egger	10.122	0.038	0.026	0.029	0.409
COPD	TMEM45A	IVW	12.269	0.031			

reported that overexpression of *SLC6A6* induces T cell exhaustion and dysfunction, thereby facilitating gastric cancer cell invasion.²⁹ In this study, significant downregulation of *CDKN1C* revealed in COPD tissues, while *SLC6A6* exhibiting upexpression. These findings indicate that these characteristic genes are closely associated with COPD progression and may serve as biomarkers for its diagnosis. Previous studies have demonstrated that plasma concentrations of ADM are significantly elevated in COPD patients with hypoxemia,³⁰ with a positive correlation to pulmonary artery pressure,³¹ suggesting its involvement in the pathophysiology of COPD-related pulmonary hypertension. *CDKN1C*, detectable in alternative tissues such as whole blood, serves as an effective biomarker for smoking exposure,³² a key risk factor for COPD, particularly when direct lung tissue access is limited.³³ *TMEM45A*, associated with inflammation in COPD, enriches IL-6-JAK-STAT3, TNF- α , and interferon- γ signaling pathways, which are central to the inflammatory processes in COPD.^{3,34} *SLC6A6* has been implicated in tumor progression by competing with CD8+ T cells for taurine, leading to T cell dysfunction,²⁹ a mechanism relevant to COPD due to immune dysregulation contributing to chronic inflammation and disease exacerbation.³⁵ *DTNA*, involved in synaptic maintenance and blood-brain barrier regulation, modulates TGF β 1 and P53 signaling and may influence HBV-induced hepatocellular carcinoma progression.^{36,37} While its direct link to COPD is unclear, its impact on signaling pathways suggests potential relevance, given the systemic inflammation and multi-organ involvement observed in COPD patients. These genes illustrate the complex interplay between hypoxia, inflammation, immune regulation, and signaling pathways in COPD, highlighting their potential as biomarkers for understanding and managing the disease.

In this study, a ceRNA interaction network was constructed for the characteristic genes to gain further insight into their potential molecular mechanisms. Among these, the lncRNAs *MALAT1*, *NEAT1*, and *XIST* were found to target multiple key miRNAs, highlighting their roles in regulating characteristic genes expression and COPD progression. *MALAT1* was initially identified as overexpressed in non-small cell lung cancer and has since been proposed as a prognostic biomarker.³⁸ Building on this, Sun et al demonstrated that *MALAT1* overexpression impairs lung function in COPD patients by downregulating miR-146a expression.³⁹ Moreover, *MALAT1* exhibits significant clinical value in predicting disease progression in COPD patients.⁴⁰ Specifically, *MALAT1* strongly negatively correlates with its target miRNAs, including miR-125b, miR-146a, and miR-203, and can effectively stratify COPD risk.⁴⁰ Evidence further suggests that *NEAT1* expression is positively associated with increased susceptibility to and severity of COPD.⁴¹ Additionally, upregulation of *XIST* is strongly associated with poor prognosis in COPD patients and has been implicated in the regulation of autoimmunity.⁴²⁻⁴⁴ Previous studies have highlighted the significant regulatory roles of these three lncRNAs in COPD progression. Therefore, targeting lncRNAs and their associated miRNAs could provide a promising therapeutic strategy to modulate characteristic gene expression and improve COPD treatment.

This study identified six potential diagnostic biomarkers for COPD, including *ADM*, *CDKN1C*, *DTNA*, *SLC2A1*, *SLC6A6*, and *TMEM45A*, using three machine learning approaches. Based on these biomarkers, a comprehensive nomogram was developed to improve the precision treatment of COPD patients. MR analysis confirmed a positive

causal relationship between *SLC2A1* and COPD, suggesting that *SLC2A1* may serve as a promising molecular target for COPD treatment. Collectively, these findings provide new insights into controlling COPD progression and advancing personalized treatment strategies.

Several limitations should be acknowledged. First, the study relied entirely on publicly available datasets (GSE19407 and GSE10006), which are relatively small and lack key metadata such as smoking status, BMI, and comorbidities, potentially affecting generalizability. Second, while the nomogram performed well in the training set (AUC = 1.00), this raises concerns about overfitting. Real-world performance should be validated in larger, prospective cohorts. Third, while experimental validation was not conducted in this study, future work will include qPCR and CRISPR-based assays to confirm the biological significance of key genes. Fourth, only *SLC2A1* showed a significant causal association with COPD in the MR analysis; the other genes should be considered hypothesis-generating. Finally, the ceRNA network predictions require further experimental validation. Overall, these findings provide a useful foundation, but further functional and clinical research is needed.

Conclusion

In conclusion, this study demonstrates a positive causal relationship between *SLC2A1* and COPD, indicating that *SLC2A1* could be a promising molecular target for COPD therapy. We further investigated the molecular mechanisms and diagnostic relevance of HRGs, highlighting their involvement in COPD pathogenesis. Given that this study relied exclusively on publicly available datasets and lacked direct clinical or experimental validation, the findings should be interpreted with caution. The absence of key clinical variables may limit the comprehensiveness of the analyses. Moreover, the causal inference for some genes was derived from a limited number of instrumental SNPs, and confidence intervals were relatively wide. Although the association of *SLC2A1* with COPD reached statistical significance, its effect size was modest and may be biologically limited or context-dependent. Nonetheless, we constructed a predictive nomogram to facilitate individualized risk assessment and clinical decision-making in COPD. Collectively, these results enhance our understanding of COPD pathophysiology and lay the groundwork for future personalized management strategies.

Data Sharing Statement

The data and materials in the current study are available from the corresponding author on reasonable request.

Ethics Approval and Consent to Participate

This study was approved by the Medical Ethics Review Committee of Jinyun People's Hospital (JYLL202504).

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors declare that they have no potential conflicts of interest in this work.

References

1. Calverley PMA, Walker PP. Contemporary concise review 2022: chronic obstructive pulmonary disease. *Respirology*. 2023;28(5):428–436. doi:10.1111/resp.14489
2. Ferrera MC, Labaki WW, Han MK. Advances in chronic obstructive pulmonary disease. *Ann Rev Med*. 2021;72:119–134. doi:10.1146/annurev-med-080919-112707

3. Ritchie AI, Wedzicha JA. Definition, causes, pathogenesis, and consequences of chronic obstructive pulmonary disease exacerbations. *Clinics Chest Med.* 2020;41(3):421–438. doi:10.1016/j.ccm.2020.06.007
4. Christenson SA, Smith BM, Bafadhel M, Putcha N. Chronic obstructive pulmonary disease. *Lancet.* 2022;399(10342):2227–2242. doi:10.1016/S0140-6736(22)00470-6
5. Yu B, Wang X, Song Y, et al. The role of hypoxia-inducible factors in cardiovascular diseases. *Pharmacol Ther.* 2022;238:108186. doi:10.1016/j.pharmthera.2022.108186
6. Shobatake R, Ota H, Takahashi N, Ueno S, Sugie K, Takasawa S. The impact of intermittent hypoxia on metabolism and cognition. *Int J Mol Sci.* 2022;23(21):12957. doi:10.3390/ijms232112957
7. Cheng Q, Fan X, Liu Y, et al. miR-455-5p regulates circadian rhythms by accelerating the degradation of Clock mRNA. *IUBMB Life.* 2022;74(3):245–258. doi:10.1002/iub.2587
8. Lodge KM, Vassallo A, Liu B, et al. Hypoxia increases the potential for neutrophil-mediated endothelial damage in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2022;205(8):903–916. doi:10.1164/rccm.202006-2467OC
9. Shukla SD, Walters EH, Simpson JL, et al. Hypoxia-inducible factor and bacterial infections in chronic obstructive pulmonary disease. *Respirology.* 2020;25(1):53–63. doi:10.1111/resp.13722
10. Choudhury P, Dasgupta S, Kar A, et al. Bioinformatics analysis of hypoxia associated genes and inflammatory cytokine profiling in COPD-PH. *Respir Med.* 2024;227:107658. doi: 10.1016/j.rmed.2024.107658
11. MacEachern SJ, Forkert ND. Machine learning for precision medicine. *Genome.* 2021;64(4):416–425. doi:10.1139/gen-2020-0131
12. Smith GD, Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ.* 2005;330(7499):1076–1079. doi:10.1136/bmj.330.7499.1076
13. Zhang JJ, Shao C, Yin YX, et al. Hypoxia-related signature is a prognostic biomarker of pancreatic cancer. *Dis Markers.* 2022;2022:6449997. doi:10.1155/2022/6449997
14. Guan Z, Jin X, Guan Z, Liu S, Tao K, Luo L. The gut microbiota metabolite capsiate regulate SLC2A1 expression by targeting HIF-1 α to inhibit knee osteoarthritis-induced ferroptosis. *Aging Cell.* 2023;22(6):e13807. doi:10.1111/ace1.13807
15. Cinkornpumin JK, Kwon SY, Prandstetter AM, et al. Hypoxia and loss of GCM1 expression prevents differentiation and contact inhibition in human trophoblast stem cells. *bioRxiv.* 2024. doi:10.1101/2024.09.10.612343
16. Liu Y, Zhang H, Liu Y, et al. Hypoxia-induced GPCPD1 depalmitoylation triggers mitophagy via regulating PRKN-mediated ubiquitination of VDAC1. *Autophagy.* 2023;19(9):2443–2463. doi:10.1080/15548627.2023.2182482
17. Singla A, Reuter S, Taube C, Peters M, Peters K. The molecular mechanisms of remodeling in asthma, COPD and IPF with a special emphasis on the complex role of Wnt5A. *Inflamm Res.* 2023;72(3):577–588. doi:10.1007/s00011-023-01692-5
18. Jayaprakash P, Vignali PDA, Delgoffe GM, Curran MA. Hypoxia reduction sensitizes refractory cancers to immunotherapy. *Ann Rev Med.* 2022;73:251–265. doi:10.1146/annurev-med-060619-022830
19. Jiang J, Zheng Z, Chen S, et al. Hypoxia inducible factor (HIF) 3 α prevents COPD by inhibiting alveolar epithelial cell ferroptosis via the HIF-3 α -GPx4 axis. *Theranostics.* 2024;14(14):5512–5527. doi:10.7150/thno.99237
20. Freemerman AJ, Johnson AR, Sacks GN, et al. Metabolic reprogramming of macrophages: glucose transporter 1 (GLUT1)-mediated glucose metabolism drives a proinflammatory phenotype. *J Biol Chem.* 2014;289(11):7884–7896. doi:10.1074/jbc.M113.522037
21. Hao B, Dong H, Xiong R, et al. Identification of SLC2A1 as a predictive biomarker for survival and response to immunotherapy in lung squamous cell carcinoma. *Comput Biol Med.* 2024;171:108183. doi:10.1016/j.compbio.2024.108183
22. Yao J, Tang S, Shi C, et al. Isoginkgetin, a potential CDK6 inhibitor, suppresses SLC2A1/GLUT1 enhancer activity to induce AMPK-ULK1-mediated cytotoxic autophagy in hepatocellular carcinoma. *Autophagy.* 2023;19(4):1221–1238. doi:10.1080/15548627.2022.2119353
23. Liu XS, Yang JW, Zeng J, et al. SLC2A1 is a diagnostic biomarker involved in immune infiltration of colorectal cancer and associated with m6A modification and ceRNA. *Front Cell Develop Biol.* 2022;10:853596. doi:10.3389/fcell.2022.853596
24. Berg T, Myrbäck TH, Olsson M, et al. Gene expression analysis of membrane transporters and drug-metabolizing enzymes in the lung of healthy and COPD subjects. *Pharmacol Res Perspect.* 2014;2(4):e00054. doi:10.1002/prp2.54
25. Wei Y, Lu X, Liu C. Gut microbiota and chronic obstructive pulmonary disease: a Mendelian randomization study. *Front Microbiol.* 2023;14:1196751. doi:10.3389/fmicb.2023.1196751
26. Creff J, Besson A. Functional versatility of the CDK inhibitor p57(Kip2). *Front Cell Develop Biol.* 2020;8:584590. doi:10.3389/fcell.2020.584590
27. Ji H, Tang Z, Jiang K, et al. Investigating potential biomarkers of acute pancreatitis in patients with a BMI>30 using Mendelian randomization and transcriptomic analysis. *Lipids Health Dis.* 2024;23(1):119. doi:10.1186/s12944-024-02102-3
28. Kubo Y, Ishizuka S, Ito T, Yoneyama D, Akanuma SI, Hosoya KI. Involvement of TauT/SLC6A6 in taurine transport at the blood-testis barrier. *Metabolites.* 2022;12(1):66. doi:10.3390/metabo12010066
29. Cao T, Zhang W, Wang Q, et al. Cancer SLC6A6-mediated taurine uptake transactivates immune checkpoint genes and induces exhaustion in CD8 (+) T cells. *Cell.* 2024;187(9):2288–2304.e27. doi:10.1016/j.cell.2024.03.011
30. Kakishita M, Nishikimi T, Okano Y, et al. Increased plasma levels of adrenomedullin in patients with pulmonary hypertension. *Clin Sci.* 1999;96(1):33–39. doi:10.1042/cs0960033
31. Yoshiyoshi M, Kamiya T, Kitamura K, et al. Plasma levels of adrenomedullin in primary and secondary pulmonary hypertension in patients <20 years of age. *Am J Cardiol.* 1997;79(11):1556–1558. doi:10.1016/s0002-9149(97)00195-1
32. Martin F, Talikka M, Hoeng J, Peitsch MC. Identification of gene expression signature for cigarette smoke exposure response--from man to mouse. *Hum Exp Toxicol.* 2015;34(12):1200–1211. doi:10.1177/0960327115600364
33. Chen L, Xiong H, Wen Q, et al. The role of active and passive smoking in chronic obstructive pulmonary disease and systemic inflammation: a 12-year prospective study in China. *J Epidemiol Global Health.* 2024;14(3):1332–1340. doi:10.1007/s44197-024-00290-w
34. Jiang H, Chen H, Wan P, Liang M, Chen N. Upregulation of TMEM45A promoted the progression of clear cell renal cell carcinoma in vitro. *J Inflamm Res.* 2021;14:6421–6430. doi:10.2147/JIR.S341596
35. Ramos Jesus F, Correia Passos F, Miranda Lopes Falcão M, et al. Immunosenescence and inflammation in chronic obstructive pulmonary disease: a systematic review. *J Clin Med.* 2024;13(12):3449. doi:10.3390/jcm13123449
36. Requena T, Cabrera S, Martín-Sierra C, Price SD, Lysakowski A, Lopez-Escamez JA. Identification of two novel mutations in FAM136A and DTNA genes in autosomal-dominant familial Meniere's disease. *Human Mol Genetics.* 2015;24(4):1119–1126. doi:10.1093/hmg/ddu524

37. Hu ZG, Zhang S, Chen YB, et al. DTNA promotes HBV-induced hepatocellular carcinoma progression by activating STAT3 and regulating TGF β 1 and P53 signaling. *Life Sci.* 2020;258:118029. doi:10.1016/j.lfs.2020.118029
38. Willingham AT, Orth AP, Batalov S, et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science.* 2005;309(5740):1570–1573. doi:10.1126/science.1115901
39. Sun L, Xu A, Li M, et al. Effect of methylation status of lncRNA-MALAT1 and microRNA-146a on pulmonary function and expression level of COX2 in patients with chronic obstructive pulmonary disease. *Front Cell Develop Biol.* 2021;9:667624. doi:10.3389/fcell.2021.667624
40. Liu S, Liu M, Dong L. The clinical value of lncRNA MALAT1 and its targets miR-125b, miR-133, miR-146a, and miR-203 for predicting disease progression in chronic obstructive pulmonary disease patients. *J Clin Lab Analysis.* 2020;34(9):e23410. doi:10.1002/jcla.23410
41. Ming X, Duan W, Yi W. Long non-coding RNA NEAT1 predicts elevated chronic obstructive pulmonary disease (COPD) susceptibility and acute exacerbation risk, and correlates with higher disease severity, inflammation, and lower miR-193a in COPD patients. *Int J Clin Exp Pathol.* 2019;12(8):2837–2848.
42. Huang X, Liang J, Li Y, et al. Significance of serum lncRNA XIST in chronic obstructive pulmonary disease and its progression to pulmonary heart disease. *BMC Pulm Med.* 2024;24(1):546. doi:10.1186/s12890-024-03354-6
43. Han H, Hao L. Revealing lncRNA biomarkers related to chronic obstructive pulmonary disease based on bioinformatics. *Int J Chronic Obstr.* 2022;17:2487–2515. doi:10.2147/COPD.S354634
44. Dou DR, Zhao Y, Belk JA, et al. Xist ribonucleoproteins promote female sex-biased autoimmunity. *Cell.* 2024;187(3):733–749.e16. doi:10.1016/j.cell.2023.12.037

International Journal of Chronic Obstructive Pulmonary Disease

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>

Dovepress
Taylor & Francis Group