

Integrating Bioinformatics Analysis with RT-qPCR Experimental Validation to Investigate Immune Cell and Telomere-Related Biomarkers in Chronic Obstructive Pulmonary Disease

Shengwei Wang , Weiwei Tang, Haixia Yang

Department of Emergence Medicine, Nanjing Brain Hospital, Nanjing, Jiangsu, 210029, People's Republic of China

Correspondence: Haixia Yang, Email yanghx8801@163.com

Purpose: Chronic obstructive pulmonary disease (COPD) is one of the most widespread diseases. Previous research has found that immune cells and telomeres may affect COPD's pathogenesis, but their combined mechanism in COPD remains unclear. This study aims to investigate the diagnostic value of telomere-associated genes and immune cells in COPD, as well as their synergistic mechanisms, thereby providing novel insights for the clinical management of COPD.

Patients and Methods: Data comprising 19 COPD cases, 24 control samples, and 2086 telomere-related genes (TRGs) were obtained from public databases. The differentially expressed genes (DEGs) between COPD and control were obtained by differential expression analysis. The key module genes related to different immune cells (DICs) were obtained via weighted gene co-expression network analysis (WGCNA). Subsequently, biomarkers were further identified by intersecting all genes, utilizing machine learning algorithm, and verifying the expression level. Furthermore, the nomogram was constructed, and gene set enrichment analysis (GSEA) of biomarkers was adopted. The transcription factors (TFs), microRNAs (miRNAs) and drugs linked to biomarkers were obtained from the databases. The expression of biomarkers in 10 clinical samples was validated via reverse transcription quantitative polymerase chain reaction (RT-qPCR).

Results: In this study, ALDH2 and HNMT were identified as biomarkers. The nomogram results demonstrated that the model had an outstanding predictive ability for COPD (area under curve (AUC) = 0.88). Besides, ALDH2 and HNMT were enriched in junction, starch, and sucrose metabolism. In addition, a total of 6 TFs such as ELF3, and 2 miRNAs, such as miR-206, were linked to ALDH2 and HNMT, and clozapine was the drug that had been found to be associated with both ALDH2 and HNMT. Finally, the RT-qPCR results were consistent with bioinformatics analysis.

Conclusion: This study identified 2 biomarkers (ALDH2 and HNMT), which might serve as potential targets for COPD. A nomogram model constructed based on biomarkers was employed for the clinical auxiliary diagnosis of COPD. This study provided new scientific evidence for improving the diagnostic process and individualized treatment strategies for COPD.

Keywords: chronic obstructive pulmonary disease, telomeres, immune cells, biomarkers

Introduction

Chronic obstructive pulmonary disease (COPD) is a common chronic airway disease characterized by persistent airflow limitation and corresponding respiratory symptoms, which seriously endangers human health. It is currently the fourth leading cause of death in the world and will become the third leading cause of death by 2030.¹ The 2018 "Chinese Adult Lung Health Study" led by Academician Wang Chen revealed an 8.6% COPD prevalence among Chinese adults aged 20+ and 13.7% in those over 40. With nearly 100 million estimated patients, COPD remains prevalent in China. Multiple factors contribute to COPD development, primarily toxic particle inhalation and smoking. Prolonged exposure to harmful substances causes abnormal inflammation, permanent respiratory damage, and irreversible pathological changes.²

Smoking is the primary risk factor for COPD. Air pollution (such as PM_{2.5}), occupational exposure (dust, chemicals), genetic factors, and indoor biofuel use (low-income countries) also have a significant impact on COPD. Current treatments include inhaled corticosteroids and bronchodilators. Still, these therapies cannot effectively prevent disease progression. Due to the complexity of its pathophysiology, other more practical and specific treatments are needed.³ Given the growing disease burden of COPD and the limitations of existing treatment strategies, in-depth research into the molecular pathological mechanisms of COPD and the development of novel targeted therapies is of great significance for improving patients' quality of life, reducing mortality, and alleviating the global healthcare burden.

Telomere biology has emerged as a significant area of interest among the complex pathophysiological mechanisms underlying COPD. These include DNA sequences at the ends of chromosomes that protect chromosomes from damage and shorten with age.⁴ Studies have shown that telomere maintenance mechanisms and telomerase activity regulation are closely related to the activation and differentiation of T cells and B cells, and telomere length also has a causal effect on the phenotypes of various immune cells (including natural killer cells, T cells, and B cells, etc).⁵ In addition, COPD is associated with a variety of pathological mechanisms, such as telomere loss. Therefore, telomeres may also play an important role in COPD, but the potential mechanism of action of telomere-related genes (TRGs) in COPD is still unclear.

Immune cells, vital for organismal survival, develop from bone marrow to become mature cells like dendritic cells, natural killer cells, lymphocytes, neutrophils, and macrophages. These cells protect the host by eliminating toxic substances and environmental pathogens, reducing infection risk through various immune response mechanisms.⁶ Research shows plasma cells, resting NK cells, activated mast cells, eosinophils, and various T and B cell types are crucial in COPD development. CD8 T cells, which are elevated in COPD lung tissue, primarily secrete IL-4 and IL-5 cytokines associated with lung parenchymal damage. Additionally, NK cells demonstrate cytotoxic effects on lung epithelial cells in COPD patients.⁷ Inhaled toxins trigger oxidative stress and airway inflammation, attracting immune cells through chemokines for defense. However, chronic inflammation disrupts immune regulation, reducing antigen response and impairing immune cell function. This damages respiratory defenses, resulting in recurrent lower respiratory infections and progressive disease deterioration.⁸ Therefore, further exploration of the mechanism of action of immune cells in COPD will help to gain a deeper understanding of the potential pathogenesis of COPD and obtain new potential therapeutic targets.

Although the roles of immune cells and telomeres in the pathogenesis of COPD have been explored separately, the mechanism of their interaction remains unclear. Furthermore, while bioinformatics analysis has been widely applied in disease research, integrated bioinformatics analysis combining telomeres and immune cells is still relatively lacking in the field of COPD. Then, immune cell-related key module genes were obtained through immune infiltration and weighted gene co-expression network analysis (WGCNA). The candidate genes were obtained by taking the intersection of these two genes with the differentially expressed genes in the dataset. Biomarkers were then screened out through machine learning and expression verification methods. Finally, a series of bioinformatics analyses were conducted to identify the potential molecular mechanisms of telomeres and immune cell-related genes as biomarkers of COPD, providing a new reference for the clinical treatment of COPD.

Materials and Methods

Data Collection

COPD's transcriptome data (GSE100153 and GSE42057) were fetched from the gene expression omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). These two datasets have been extensively utilised in previous relevant studies, with their data quality validated by numerous investigations, rendering them highly reliable.^{9,10} The GSE100153 dataset (GPL6884) was used as the training set, including 19 COPD and 24 control blood samples. The GSE42057 dataset (GPL570) was used as the validation set, including 94 COPD and 42 control PBMCs samples. The 2086 TRGs were extracted from the TelNet database (<http://www.cancertelsys.org/telnet/>) ([Supplementary Table 1](#)).

Immune Infiltration Analysis and Weighted Correlation Network Analysis

In this study, the infiltration abundance of the 64 immune and stromal cells¹¹ between all COPD and control samples in GSE100153 was elucidated by the xCell algorithm of the “xCell” package,¹² and the results were displayed in the heat map. Then, DICs were identified by the Wilcoxon test ($P < 0.05$) and the results were displayed utilizing the “ggplot2” package.¹³

Then, the WGCNA analysis was executed utilizing the “WGCNA” package,¹⁴ and the modules related to the traits (DICs) were identified. Firstly, the “GoodSamplesGenes” function was adopted to perform hierarchical clustering analysis on all samples in the GSE100153 dataset to eliminate outlier samples. Then, the R2 was set to 0.85 to obtain soft thresholds (β), and the topology overlap and adjacency matrices were established. Finally, the gene adjacency was calculated to create a hierarchical clustering tree of genes. The modules were obtained by setting the minimum number of genes for each gene module to 30, and setting the mergeCutHeight to 0.50. After that, Pearson correlation analysis was employed to calculate correlations between modules and DICs ($|\text{cor}| > 0.5$ and $P < 0.05$), identifying key modules exhibiting the highest correlations with any DIC.

Identification of Candidate Genes

In all samples of GSE100153, the raw expression profiles were first normalised using the normaliseBetweenArrays() function within the limma package,¹⁵ employing the “quantile” normalisation method. Subsequently, the differentially expressed genes (DEGs) between COPD and control samples (COPD vs control) were obtained utilizing this package ($|\log_2\text{fold change (FC)}| > 0.5$, $P < 0.05$). According to the $\log_2\text{FC}$ value, DEGs were visualized by the volcano plot utilizing the “ggVolcano” package (<https://CRAN.R-project.org/package=ggvolcano>), with the names of the top 10 up- or down-regulated genes labeled. Moreover, a heat map was used to display all DEGs utilizing “ComplexHeatmap” package.¹⁶ Lastly, the candidate genes were extracted by intersecting the DEGs, TRGs and key module genes using the “VennDiagram” package.¹⁷

Enrichment Analysis and Protein-Protein Interaction (PPI) Network

The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) were utilized to analyze the biological functions involved in candidate genes utilizing “clusterProfiler” package¹⁸ ($P < 0.05$) and the top 10 markedly terms of GO and all enriched terms of KEGG were visualized. The protein level interactions of candidate genes were explored using PPI network constructed with the STRING database (<https://string-db.org/>) (confidence ≥ 0.15). The degree scores of these genes were obtained, and the results were displayed utilizing Cytoscape software.¹⁹

Machine Learning and Expression Level Verification

The support vector machine-recursive feature elimination (SVM-RFE) algorithm and expression level of genes were considered for further obtaining biomarkers in this study. Firstly, the “e1071” package²⁰ was employed to perform the SVM-RFE algorithm through 10-fold cross-validation, and the candidate biomarkers were obtained by finding the optimal combination with the lowest error rate in all samples of GSE100153. Furthermore, the expression differences and tendency of the candidate biomarkers were explored by expression level verification in all samples of GSE100153 and GSE42057. The expression disparities between COPD and control samples were evaluated via the Wilcoxon test ($P < 0.05$). Lastly, the genes with notable disparities between COPD and control samples and consistent expression trends in both datasets were considered biomarkers. After that, the correlation between DICs and biomarkers were identified by spearman analysis via “psych” package²¹ ($|\text{cor}| > 0.3$ and $P < 0.05$), and the results were visualized utilizing “ggplot2” package.¹¹

Construction of the Nomogram Model

The nomogram model was employed to explore the diagnostic capability of biomarkers for COPD. In all samples of GSE100153, the nomogram was constructed based on biomarkers via the “rms” package.²² According to the nomogram, biomarkers were scored separately, with each biomarker corresponding to a point, and the points of each biomarker were

added together to get the total points. Furthermore, the accuracy of the nomogram model was validated by constructing a calibration curve via the “rms” package. The predictive ability of the nomogram was assessed by an receiver operating characteristic (ROC) curve (area under curve (AUC) > 0.7) utilizing the “pROC” package²³ and a DCA curve utilizing “rmda”.²⁴

Gene Set Enrichment Analysis (GSEA)

The GSEA explored the biological functions of biomarkers in all samples of GSE100153. Firstly, the MsigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/>) was utilized to get the “c2.cp.kegg.v7.4.symbols.gmt” gene set which was used as a reference gene set. Then, the Spearman correlation between each biomarker and other genes was calculated via the “psych” package.²¹ After that, the genes were sorted with their correlation coefficients in descending order. Lastly, the GSEA was performed by “clusterProfiler” package (NES > 1, P < 0.05, FDR < 0.25), and the top 3–5 pathways were displayed by the P value in ascending order via “enrichplot” package.²⁵

Prediction of Transcription Factors (TFs), microRNAs (miRNAs) and Drugs

In this study, the TFs that regulated biomarkers were predicted utilizing the KnockTF (<http://www.licpathway.net/KnockTF/index.html>) database. In addition, the miRWalk (<http://mirwalk.umm.uni-heidelberg.de>) database and miRcode (<http://www.mircode.org/index.php>) database were utilized to predict the miRNAs, and the miRNAs linked to biomarkers were obtained by overlapping the miRNAs in the two databases. The potential drugs that targeted biomarkers were predicted utilizing the Drug-Gene Interaction database (DGIdb) (<http://dgidb.org/>). These results were visualized via Cytoscape.²⁶

The Assessment of Biomarker Expression

The expression of biomarkers in clinical blood samples was detected using RT-qPCR. This study was conducted in strict accordance with the principles of the Declaration of Helsinki and was approved by the Medical Ethics Committee of Nanjing Brain Hospital (Approval No. 2025-KY078-02). Written informed consent was obtained from all participants prior to their enrollment. Blood samples, including those from 5 patients with COPD and 5 healthy controls, were obtained from Nanjing Chest Hospital. After that, the total RNA of all blood samples was extracted using the TRIzol reagent (Vazyme, China). Then, the RNA concentrations were computered by NanoPhotometer N50, and mRNA was reversely transcribed into cDNA utilizing SureScript-First-strand-cDNA-synthesis-CREB5B test kit (Yesen, Shanghan, China). RT-qPCR reactions were conducted in a 10μL system comprising 3μL cDNA template, 5μL 2× Universal Blue SYBR Green qPCR Master Mix, and 0.2μM forward and reverse primers. The reaction protocol comprised: 1 minute pre-denaturation at 95°C; followed by 40 cycles of 20 seconds at 95°C, 20 seconds at 55°C, and 30 seconds at 72°C. GAPDH served as the internal control gene, with relative expression levels of the target gene calculated using the $2^{-\Delta\Delta C_t}$ method. Detailed information on primers and machine testing conditions was listed in [Supplementary Table 2](#). Student’s t-tests were employed to compare gene expression differences between the COPD group and the control group, with statistical significance set at P < 0.05. All statistical analyses and graphical representations were completed using Graphpad Prism 5.²⁷

Statistical Analysis

Bioinformatics analyses were conducted using the R programming language (v 4.2.2). The specific software packages and their versions employed in the analyses are detailed in [Supplementary Table 3](#). Wilcoxon test or Student’s *t* test was utilized to assess the differences between the two groups. P < 0.05 was statistically significant.

Results

Identification of Key Module Genes

The abundance of 64 types of immune and stromal cells between COPD and control samples was shown in [Figure 1A](#). Wilcoxon test results indicated that there were 13 types of cells that had notable differences between COPD and control

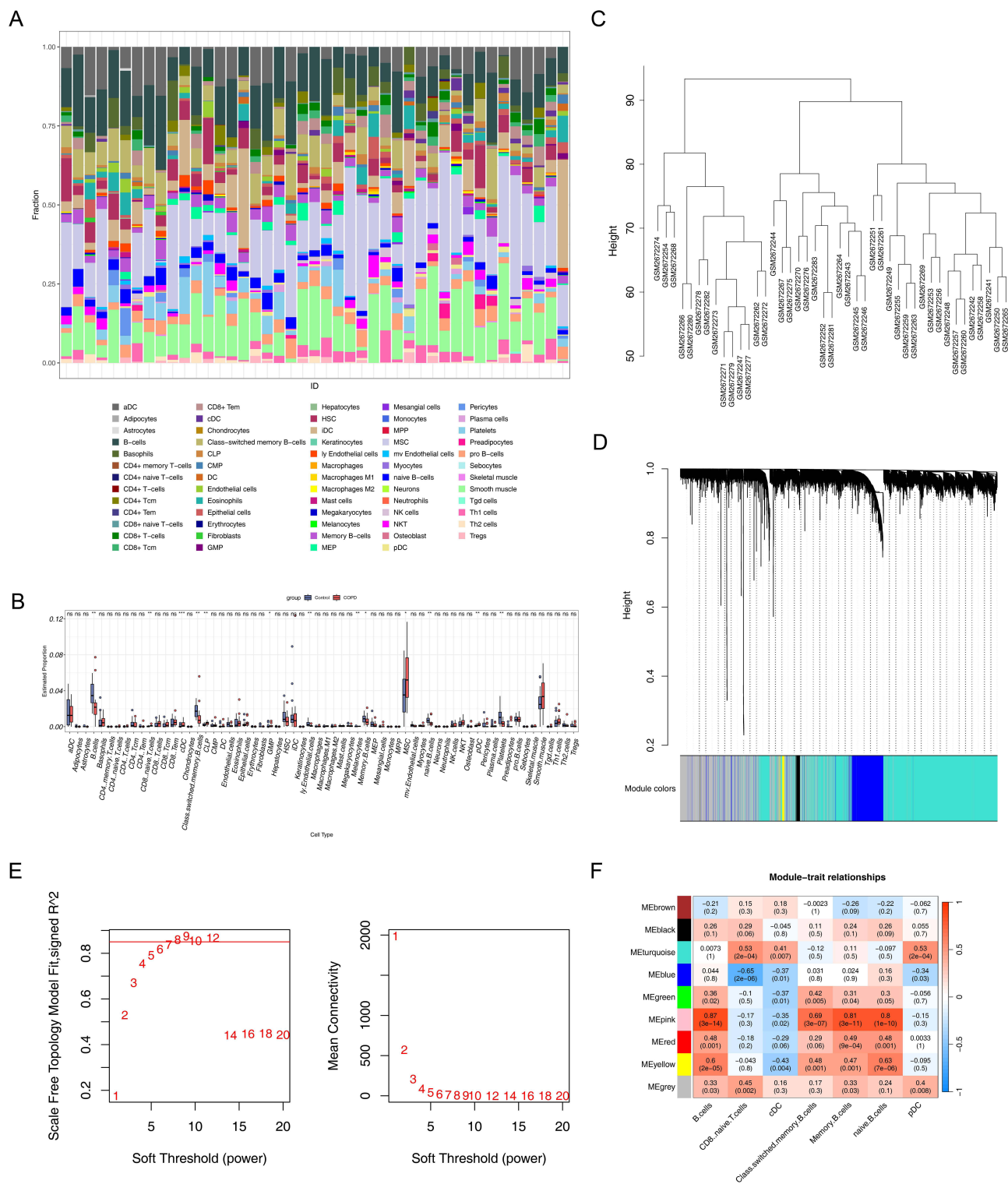


Figure 1 Analysis of immune cell infiltration and gene co-expression network in chronic obstructive pulmonary disease (COPD). **(A)** Relative proportions of 64 cell types in COPD samples and control samples. The horizontal axis represented sample IDs, and the vertical axis represented the proportion of immune cells in each sample. **(B)** Differences in immune cell infiltration between COPD samples and control samples. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; "ns" indicates no statistical significance. **(C)** Sample clustering dendrogram. The horizontal axis represented sample names, and the vertical axis represented the height of the clustering tree, which was used to identify sample outliers and assist in data quality control. **(D)** Gene module clustering dendrogram. The horizontal axis represented the color identifiers of gene modules, and the vertical axis represented the height of the gene clustering tree. **(E)** Soft threshold screening. Red font indicates soft thresholds. In the left graph, the horizontal axis represented soft threshold values (1–20), and the vertical axis represented the scale-free network evaluation coefficient R^2 , with the red cut-off line corresponding to the screening threshold ($R^2 = 0.85$); in the right graph, the horizontal axis represented soft threshold values (1–20), and the vertical axis represented the connectivity of the scale-free network, which was used to determine the optimal parameters for constructing the co-expression network. **(F)** Heatmap of correlations between module genes and immune cells.

samples ($P < 0.05$). Especially, 7 immune cells, including B-cells, naive CD8(+) T-cells, class-switched memory B-cells, conventional dendritic cells (cDCs), memory B-cells, plasmacytoid dendritic cells (pDCs), and naive B-cells, which had notable differences between COPD and control samples (Figure 1B). Then, the WGCNA analysis was performed in all samples. No significant outlier samples were identified, and thus all samples were retained for subsequent network construction (Figure 1C). After that, 9 co-expression modules (excluding the grey module) were obtained when the soft-threshold was 9 (Figure 1D and E). The correlation analysis indicated that the MEturquoise modules ($P < 0.05$, $cor > 0.5$), MEblue modules ($P < 0.05$, $cor < 0.5$), MEpink modules ($P < 0.05$, $cor > 0.5$), and MEyellow modules ($P < 0.05$, $cor > 0.5$) were identified, and 5203 key module genes in the 4 modules were obtained (Figure 1F).

Identification and Exploration of Candidate Genes

In the GSE100153 dataset, a total of 341 DEGs were determined. In detail, 235 genes were up-regulated while the remaining genes showed the opposite expression trend in the COPD group (Figure 2A and B). Finally, 23 candidate genes were obtained (Figure 2C).

In GO analysis, a total of 254 biological processes (BPs), 31 cellular components (CCs), and 40 molecular functions (MFs) were obtained ($P < 0.05$) (Figure 2D, [Supplementary Table 4](#)). Regarding BP, the candidate genes were enriched in terms of positive regulation of the regulated secretory pathway. In CC, candidate genes were significantly enriched in RNA polymerase II transcription regulator complexes and ribosomes. In MF, candidate genes were significantly enriched in heat shock protein binding and profilin binding. In KEGG, candidate genes were significantly enriched in 3 pathways, including cellular senescence, histidine metabolism, and lysine degradation (Figure 2E, [Supplementary Table 5](#)). After removing isolated genes, the PPI results indicated that 23 genes had an interactive relationship. GATA1, CARM1 and KLF1 had the highest degree scores (Figure 2F).

Identification and Exploration of Biomarkers

A total of 4 genes (RPS21, HNMT, ALDH2, and E2F5) were obtained utilizing SVM-RFE (Figure 3A and B). After that, E2F5 was removed because its gene name was missing in the GSE100153 dataset. Then, only ALDH2 and HNMT showed notable differences between COPD and control groups ($P < 0.05$), and the 2 genes showed upward trends in the COPD group of both 2 datasets (Figure 3C and D). Therefore, ALDH2 and HNMT were regarded as the biomarkers. The correlation analysis indicated that ALDH2 ($cor = 0.58$, $P < 0.001$) and HNMT ($cor = 0.52$, $P < 0.001$) were significantly positively correlated with cDCs. The ALDH2 had the most notable negative correlation with naive B-cells ($cor = -0.51$, $P < 0.001$) (Figure 3E, [Supplementary Table 6](#)). These results showed that these 2 DICs might have a notable impact on biomarkers in COPD.

High Accuracy of the Nomogram

In the results of the nomogram model, the higher the total points, the higher the risk of COPD, and the results indicated that ALDH2 exhibited higher diagnostic value for COPD than HNMT (Figure 4A). The predicted probability of calibration curves had a high degree of overlap with the reference line, revealing that the nomogram had good accuracy in predicting COPD ($P = 0.143$) (Figure 4B). The results of the ROC curve indicated that the model had excellent predictive performance for COPD patients ($AUC = 0.88$) (Figure 4C). The net benefit of the model in the DCA curve was all better than that in the None while the decision threshold was 0.2 to 0.8, which indicated that the nomogram could have good practical application value for PD patients (Figure 4D). The results above demonstrated that the model had an outstanding predictive ability for COPD.

Enrichment Pathway of Biomarkers

The GSEA results indicated that ALDH2 was enriched in 16 notable pathways, including starch and steroid hormone biosynthesis and valine, leucine and isoleucine degradation (Figure 5A, [Supplementary Table 7](#)). In addition, HNMT was enriched in 3 notable pathways (Figure 5B, [Supplementary Table 8](#)). Notably, ALDH2 and HNMT were both enriched in the tight junction, as well as starch and sucrose metabolism. The above results indicated that biomarkers most likely affected the development of COPD through these 2 pathways.

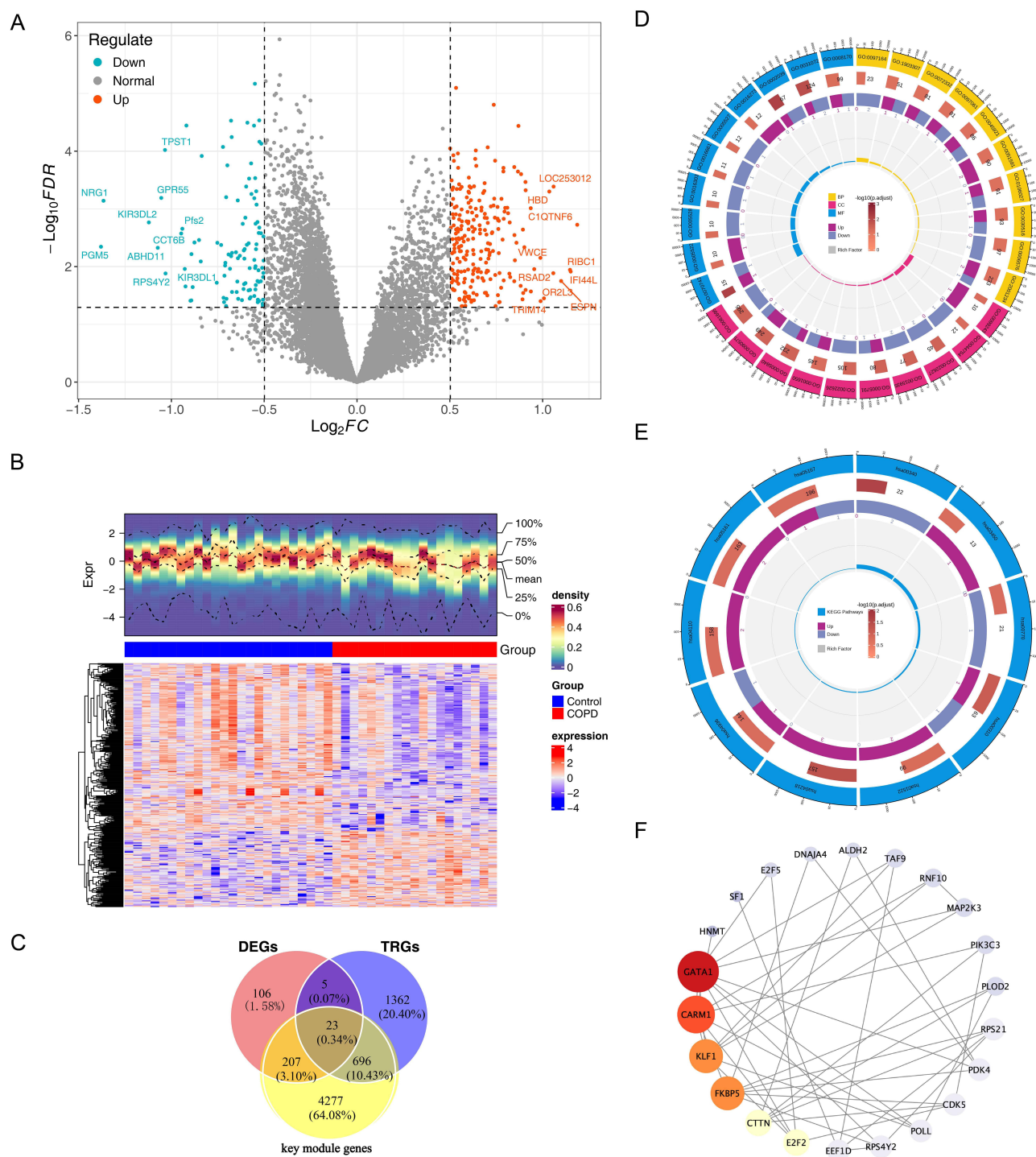


Figure 2 Analysis of Differentially Expressed Genes and Candidate Genes in Chronic Obstructive Pulmonary Disease (COPD). **(A)** Volcano plot of differentially expressed genes (DEGs) in COPD samples and control samples, The horizontal axis represents Log_2FC , and the vertical axis represents $-\text{Log}_{10}(\text{FDR})$. **(B)** Heatmap of differentially expressed genes (DEGs) in COPD samples and control samples, The middle annotation bar indicates red for COPD samples and blue for control samples; the upper density heatmap represents the gene expression density of each sample, with redder colors indicating higher density; the lower heatmap has genes on the vertical axis, with red representing high expression and blue representing low expression. **(C)** Identify candidate genes using Venn diagrams, Red circles represent DEGs, yellow circles represent key module genes from WGCNA, and blue circles represent telomerase-related genes (TRGs). **(D)** GO enrichment analysis of candidate genes. **(E)** KEGG analysis of candidate genes. **(F)** Candidate gene PPI network, Each circle represents a gene, and the lines represent interactions between genes.

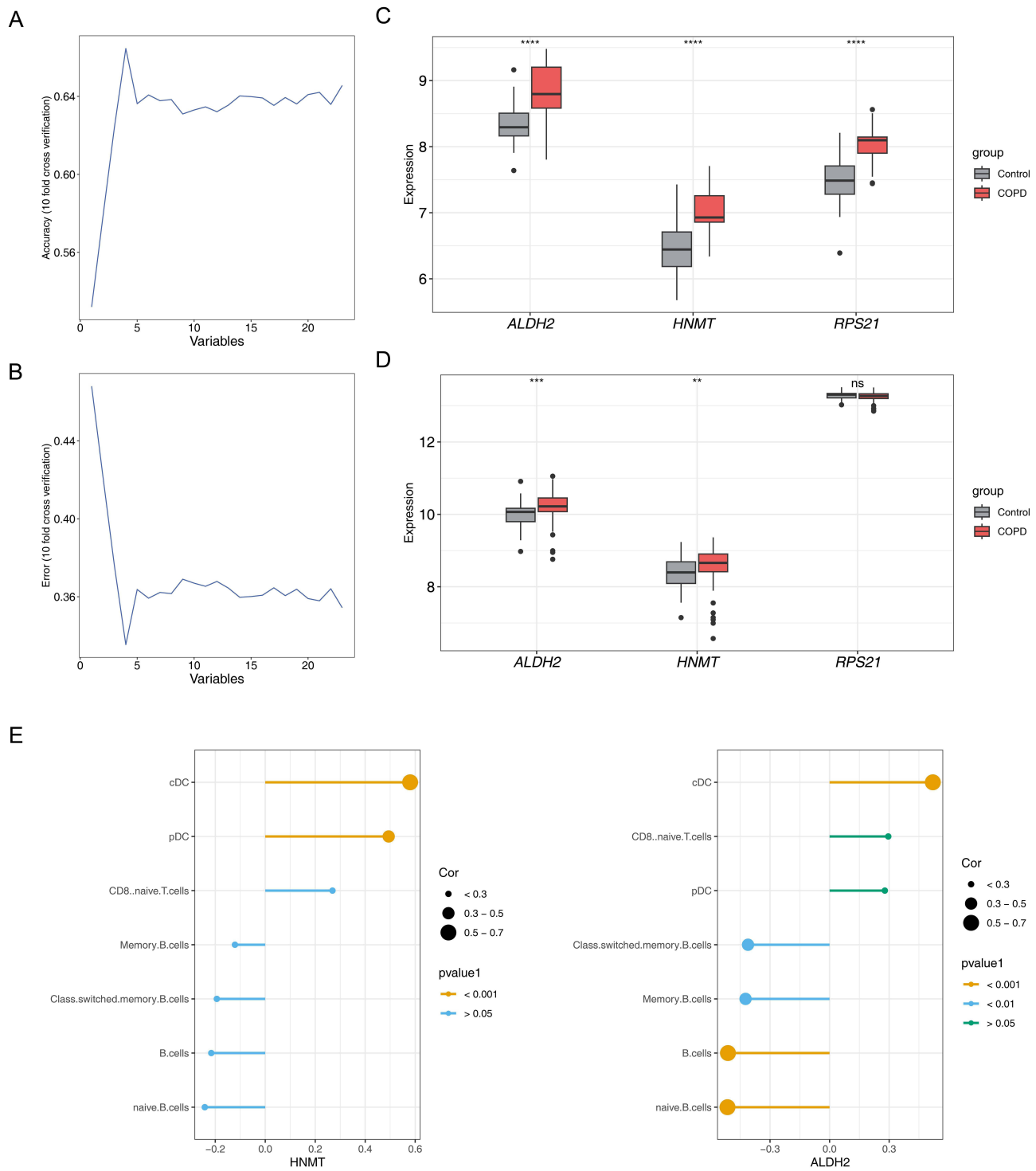
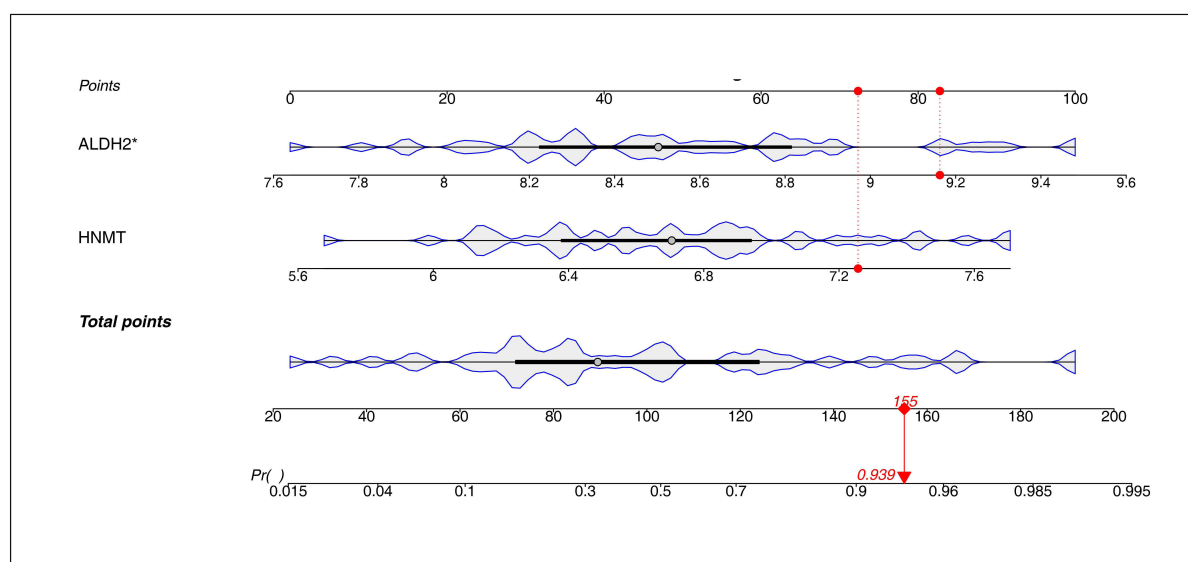
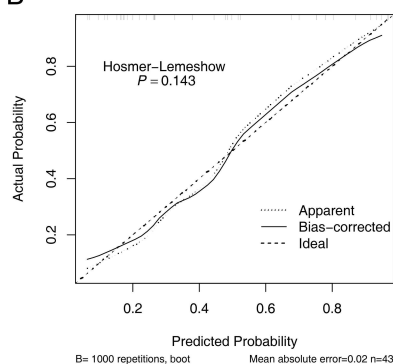


Figure 3 Screening of Candidate Biomarkers and Analysis of Immune Correlation in Chronic Obstructive Pulmonary Disease (COPD). **(A, B)** The SVM-RFE algorithm selects genes with the lowest error and highest accuracy, The horizontal axis represents the number of genes in the model, while the vertical axis shows the accuracy (left graph) and error rate (right graph) of the model. **(C)** Expression of candidate biomarkers in the training set GSE100153, The horizontal axis represents gene names, while the vertical axis shows gene expression values in both patient and control groups. **** $P < 0.0001$. **(D)** Expression of candidate biomarkers in the validation set GSE42057. ** $P < 0.01$, *** $P < 0.001$; “ns” indicates no statistical significance. The horizontal axis represents gene names, while the y-axis shows gene expression values in both patient and control groups. Gray indicates the normal control group, red indicates the COPD patient group. **(E)** Analysis of biomarkers’ correlation with differential immune cells, The left figure shows the correlation between HNMT and seven types of different immune cells, while the right figure displays the correlation between ALDH2 and the same seven types of immune cells. The horizontal axis represents the correlation coefficient, and the vertical axis lists the seven different immune cells.

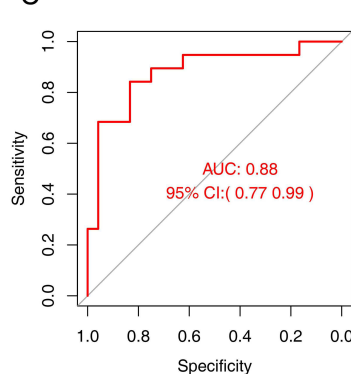
A



B



C



D

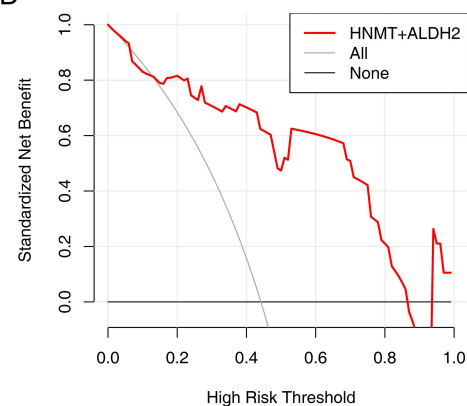
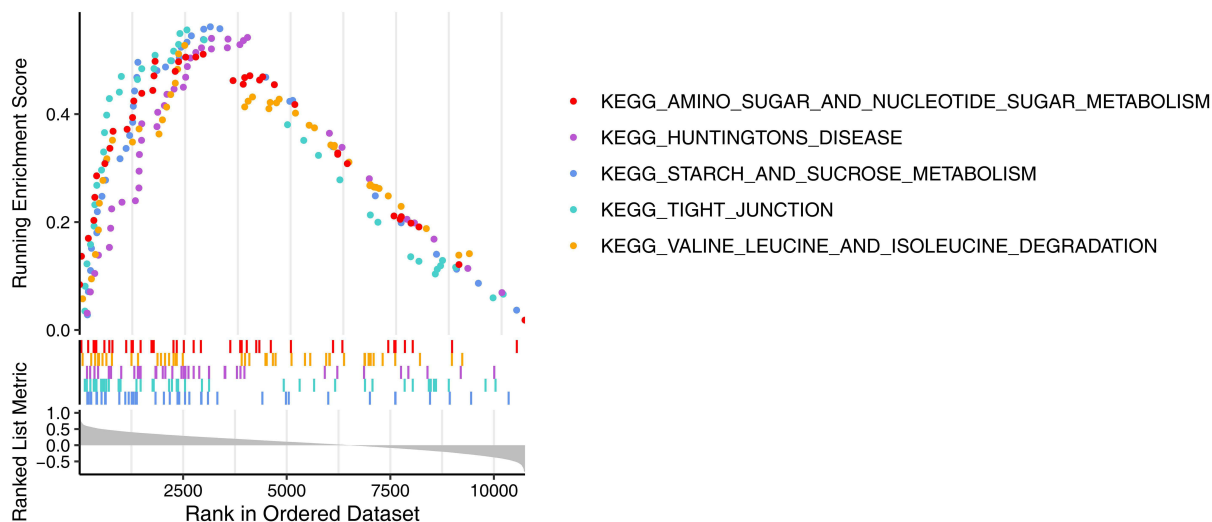


Figure 4 Construction and Efficacy Evaluation of a Prediction Model for Chronic Obstructive Pulmonary Disease (COPD). **(A)** Construction of the nomogram mode. * $P < 0.05$. The first row shows the reference scale for individual scores, the second row displays the expression value of the ALDH2 gene, the third row shows the expression value of the HNMT gene, the fourth row presents the reference scale for total scores, and the fifth row indicates the risk of illness. A red marker provides an example of a specific sample. **(B)** Calibration curve, The horizontal axis represents the predicted probability from the model, while the vertical axis represents the actual probability. In the calibration curve, a straight line passing through the origin with a slope of 1 serves as a reference line indicating that the predicted probabilities from the nomogram perfectly match the actual probabilities. The solid line indicates the corrected values. **(C)** ROC curve, The horizontal axis represents predicted specificity, while the vertical axis represents predicted sensitivity. The area under the red curve is the AUC (Area Under Curve). The AUC value and the 95% confidence interval are labeled in the center of the graph. **(D)** DCA curve, The red line represents the model curve, while the other two lines are models for decision support: one gray line indicates all COPD outcomes, and the other black line indicates that none of the patients have COPD. The horizontal axis is the threshold probability, and the vertical axis is net benefit.

TFs, miRNAs, and Drugs Associated with Biomarkers

A total of 18 TFs such as HIC1, were linked to ALDH2, and 10 TFs, such as KLF5, were linked to HNMT. Notably, a total of 6 TFs, including ELF3, CREB1, TP53, RELB, MYC, and RELA were linked to ALDH2 and HNMT (Figure 6A). A total of 14 miRNAs were found to be linked to HNMT, and a total of 7 miRNAs were found to be associated with ALDH2 (Figure 6B). The miRNAs related to HNMT included miR-507 and miR-4770, while those related to ALDH2 included miR-761 and miR-184. Moreover, miR-206 and miR-490-3p were both linked to ALDH2 and HNMT. These results indicated that these TFs and miRNAs associated with ALDH2 and HNMT might be the key factors affecting COPD. Finally, a total of 21 drugs linked to ALDH2 and 3 drugs linked to HNMT were obtained from the DGIdb database, respectively (Figure 6C). The drugs related to HNMT included dabigatran and diphenhydramine,

A



B

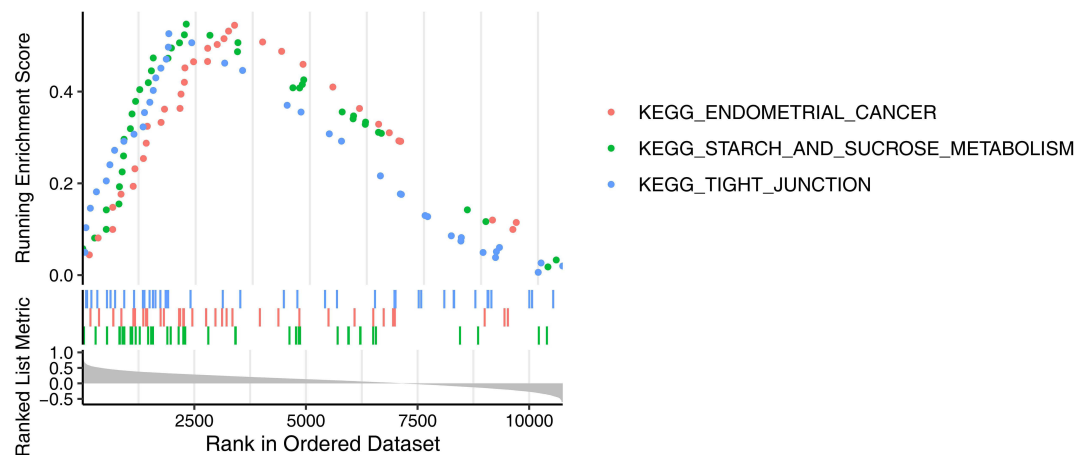


Figure 5 GSEA Enrichment Analysis of COPD Biomarkers ALDH2 and HNMT. **(A)** GSEA enrichment analysis of biomarker ALDH2, The horizontal axis represents reference gene sets sorted by Spearman correlation coefficients from largest to smallest between ALDH2 and all other genes. The vertical axis shows the enrichment score, with higher scores indicating greater enrichment. **(B)** GSEA enrichment analysis of biomarker HNMT, The horizontal axis represents reference gene sets sorted by Spearman correlation coefficients from highest to lowest between HNMT and all other genes. The vertical axis shows the enrichment score, with higher scores indicating greater enrichment.

while those related to ALDH2 included dopamine. Notably, clozapine had been found to be associated with both ALDH2 and HNMT, which indicates that the drug is a potential treatment for COPD.

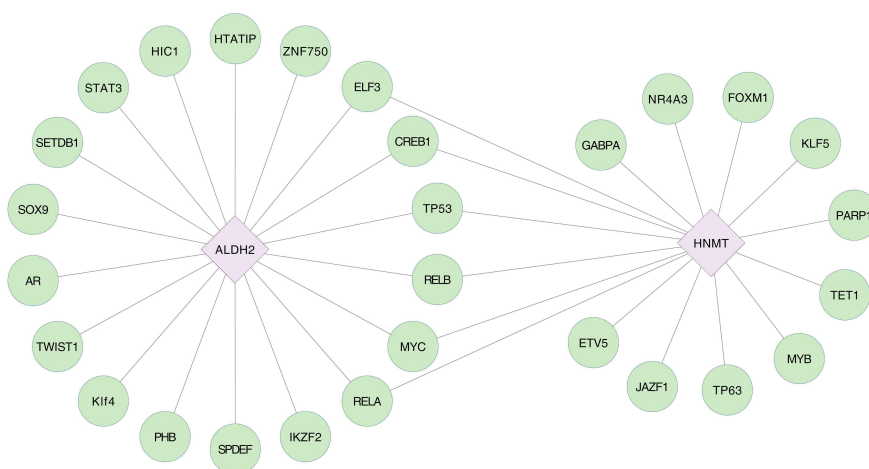
RT-qPCR Experiments of Biomarkers

In RT-qPCR experiments with clinical samples, the expression of these 2 genes showed notable differences between COPD and control samples ($P < 0.05$). Specifically, ALDH2 and HNMT were up-regulated in the COPD samples (Figure 7A and B). This result was consistent with our previous bioinformatics analysis, indicating that our results were reliable.

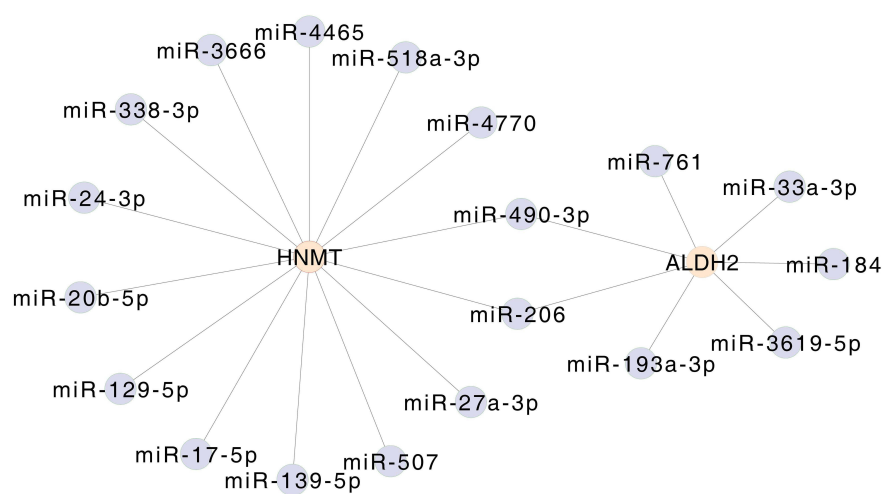
Discussion

COPD seriously endangers human health. By detecting the telomere length in bronchoalveolar lavage fluid, the study found that the proportion of ultra-short telomeres ($<1.5\text{kb}$) in lung tissue was significantly higher in patients with greater smoking and lower lung function (FEV1%). This indicates that smoking-induced telomere shortening is directly involved in the pathological process of COPD.²⁸ Cellular senescence, characterized by telomere dysfunction, drives the

A



B



C

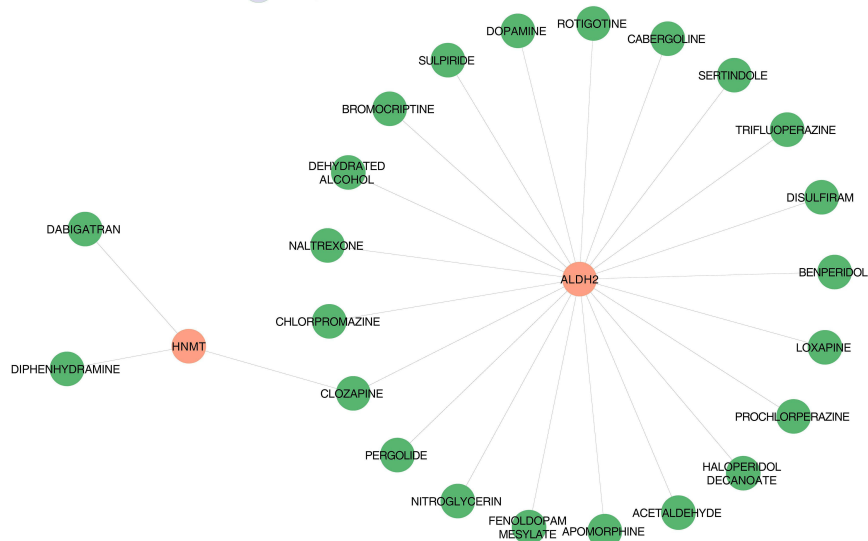


Figure 6 Analysis of Regulatory Networks of COPD Biomarkers and Drug Associations. **(A)** TF-biomarker regulatory network, Green circles represent TFs, purple diamonds represent genes, and lines indicate interactions between TFs and genes. **(B)** miRNA-biomarker regulatory network, Gray circles represent miRNA, Orange circles represent genes, and lines represent interactions between miRNA and genes. **(C)** Drug-biomarker network, Green circles represent drugs, orange circles represent genes, and lines indicate interactions between drugs and genes.

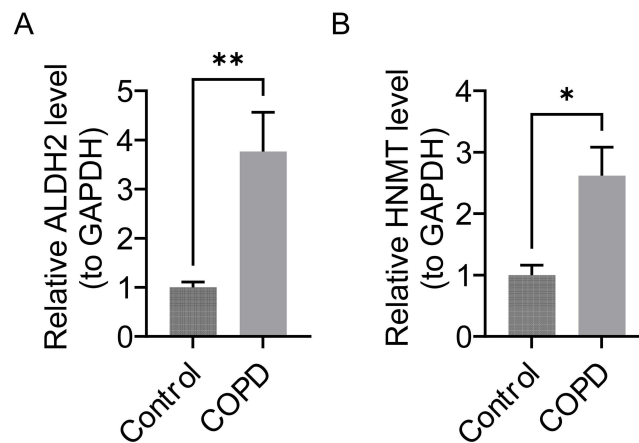


Figure 7 ALDH2 and HNMT gene expression differences between the COPD group and the control group. **(A)** ALDH2 gene expression difference. **P < 0.01. **(B)** HNMT gene expression difference. *P < 0.05.

pathological process of COPD by secreting proinflammatory factors.²⁹ This study integrated immune cell and telomere-related genes using bioinformatics approaches to identify ALDH2 and HNMT as potential biomarkers for COPD. The nomogram constructed based on these findings demonstrated significant predictive ability (AUC = 0.88), providing new insights for the early diagnosis of COPD. Notably, a diagnostic model built by Wang et al³⁰ based on mitochondrial metabolism-related genes also exhibited high discriminative capacity (AUC = 0.8). In future research, systematically integrating the biomarkers identified in this study, such as ALDH2, HNMT, and mitochondrial metabolism-related genes, is expected to enable the development of a more universally applicable and robust combined diagnostic model, thereby further enhancing the ability for early screening and risk stratification of COPD. ALDH2 is a member of the ALDH family. Key alcohol metabolism enzymes, mainly located in mitochondria, play a vital role in detoxifying acetaldehyde and endogenous lipid aldehydes. Approximately 30%-50% of the East Asian population (about 8% of the world's population) carries the ALDH2 rs671 (Glu504Lys) polymorphism. Studies have shown that ALDH2 has a protective effect in various cardiovascular models. Functionally, the activation of ALDH2 leads to improvements in cardiac hemodynamic parameters and myocardial damage. Previous studies have also confirmed that in East Asia, the high frequency of inactive ALDH2 alleles may exacerbate the effects of environmental acetaldehyde exposure on lung function and may exacerbate the impact on COPD. This study found that ALDH2 was significantly upregulated in the COPD group. We speculate that ALDH2 may play a certain role in maintaining lung function by reducing the damage to lung function caused by harmful substances such as acetaldehyde through its active metabolic detoxification effect, which helps to slow the progression of COPD and provides a potential target for the development of new treatments for COPD.³¹

HNMT primarily functions as a cytoplasmic enzyme metabolizing intracellular histamine by facilitating methyl group transfer from S-adenosyl-L-methionine to histamine. When growth factor receptors are activated, HNMT translocates from the cytoplasm to the plasma membrane, interacting with organic cation transporters (histamine transporters). Research indicates HNMT's potential as an auxiliary biomarker for predicting breast cancer patients' responsiveness to anti-HER2 therapy. Additionally, studies have linked HNMT to cardiac toxicity and Parkinson's disease. Although direct evidence connecting HNMT to Chronic Obstructive Pulmonary Disease (COPD) is currently lacking, research has demonstrated that allergic reactions and viral infections in airway epithelium can suppress HNMT and cholinesterase production, thereby intensifying bronchial constriction.³² This suggests that HNMT reduction may disrupt histamine homeostasis, potentially promoting airway inflammation and hyperreactivity, which could contribute to COPD pathophysiology. Future investigations should explore HNMT's role in COPD pathogenesis and evaluate its potential as a therapeutic target.

Tight junctions are the main barrier for the diffusion of epithelial cells through the intercellular space, recruiting various cytoskeleton and signaling molecules on the cytoplasmic surface.³³ Tight junctions comprise transmembrane and

junctional adhesion molecules (JAMs), forming a selective permeability barrier beside the cell. Studies have found that damage to tight junctions is the main cause of epithelial barrier destruction during lung inflammation.³⁴ Cigarette smoke destroys the tight junctions of human airway epithelium, and transcription factors fight against this smoke-induced COPD. Tight junction proteins may protect airway epithelial homeostasis during COPD.³⁴ It has been reported that the level of tight protein CLDN4 in the plasma of patients with stable COPD is significantly lower than that of the control group. In contrast, the plasma CLDN4 level of patients with acute exacerbation of COPD is significantly increased and negatively correlated with lung function indicators.³⁵

In this study, the biomarker ALDH2 was enriched in the tight junction pathway, and it is speculated that ALDH2 may affect the occurrence and development of COPD by participating in tight junction-related physiological processes. Given the importance of tight junctions in maintaining airway epithelial barrier function and the relationship between their damage and COPD, ALDH2 may protect tight junctions and maintain airway epithelial homeostasis, perhaps by regulating tight junction proteins' expression, localization, or function.

Based on combined analyses of regulatory factors and drug prediction results, six TFs, including ELF3, CREB1, TP53, and RELB, were identified through database prediction to be associated with ALDH2 and HNMT. Research has demonstrated that CREB1 regulates GSTCD expression via the rs80245547 and rs72673891 polymorphisms by enhancing DNA binding affinity within enhancer regions, facilitating remote chromatin interactions that upregulate GSTCD transcription. This CREB1-mediated mechanism protects against COPD susceptibility by increasing GSTCD expression, thus protecting role against pulmonary inflammation and oxidative stress responses.³⁶ Studies involving ovalbumin sensitization and airway challenge in ELF3-deficient mice revealed impaired interleukin-6 (IL-6) production by dendritic cells (DCs), resulting in inhibited T helper 17 (Th17) response induction. Simultaneously, these mice exhibited an exacerbated T helper 2 (Th2) response, characterized by increased interleukin-4 (IL-4) production, elevated ovalbumin-specific immunoglobulin E (IgE) and immunoglobulin G1 (IgG1) antibody titers, and heightened Th2 cytokine levels, accompanied by airway inflammation and increased mucus secretion. These findings indicate that ELF3 regulates allergic airway inflammation by modulating DC-driven differentiation of Th1, Th2, and Th17 cells. Given the similar inflammatory mechanisms between COPD and allergic airway inflammation, ELF3 likely participates in COPD inflammatory regulation processes.³⁷ These findings suggest that ALDH2 and HNMT may influence COPD pathogenesis by interacting with these transcription factors. Therefore, the biomarker-related factors may play an essential role in the progression of COPD, indicating that these factors may regulate the expression of biomarkers and thus affect the progression of the disease.

This study screened ALDH2 and HNMT as COPD biomarkers by analyzing COPD transcriptome data in public databases. Nomogram validation confirmed the diagnostic ability of these markers for COPD. Detection methods based on ALDH2 and HNMT are expected to achieve more accurate early diagnosis and timely intervention to delay disease progression. At the same time, differences in its expression levels may be related to treatment response. Through detection, drug efficacy and adverse reactions can be predicted, which helps to formulate targeted treatment plans, improve treatment effects and reduce medication risks. It is worth noting that pulmonary rehabilitation can reduce the resting circumference of the upper chest wall in patients with COPD, improve thoracic mobility,³⁸ and alleviate PaCO₂ levels in patients with hypercapnia³⁹ this benefit also extends to patients with post-COVID-19 syndrome, significantly improving their dyspnea, physical function, quality of life and psychological state,^{40–42} while integrating evidence-based eHealth education. The combination of tool and pulmonary rehabilitation can significantly improve the psychological problems of rehabilitation.⁴³ Therefore, in the future, we can explore the dynamic changes in the expression levels of ALDH2 and HNMT during pulmonary rehabilitation and integrate them with key physiological indicators such as the range of thoracic movement, aiming to build a comprehensive model that can more accurately evaluate and predict the efficacy of rehabilitation. Ultimately, by revealing the intrinsic relationship between biomarkers, clinical outcomes of pulmonary rehabilitation and patients' psychological state, it provides a new perspective and powerful tool for the formation of a personalized and physical and mental COPD management strategy, thereby comprehensively improving patients' diagnosis and treatment benefits and rehabilitation quality. However, this study still has certain limitations. First, the clinical validation cohort is limited in size (only 5 samples per group), which restricts statistical power and limits the generalizability of the research conclusions and their potential for clinical translation. Second, the study heavily relies on

bioinformatics analysis, and the predicted results of key molecular markers (such as transcription factors, miRNAs, and drug targets) still need to be further validated by experimental techniques such as ChIP-seq and CLIP-seq. Additionally, pathway enrichment analysis only reveals statistical associations in gene expression but has not clarified their causal regulatory mechanisms through functional experiments such as gene knockout and overexpression. Finally, the ALDH2 gene rs671 locus has extremely low distribution outside East Asian populations, and its population specificity may limit the cross.

Conclusions

This study screened 23 candidate genes from the COPD transcriptome and preliminarily identified ALDH2 and HNMT as potential biomarkers for COPD. The study showed the two participated in the pathological process of COPD through the tight junction and starch and sucrose metabolism pathways. Six transcription factors (ELF3, CREB1,53, RELB, etc.) and 2 miRNAs (miR-206, miR-490-3p) were found to associated with the dual biomarkers in terms of regulation. Drug prediction suggests that clozapine can target both, but its targeting efficacy and potential side effects in real physiological environments still need to be further verified through subsequent experiments. This discovery provides a new theoretical basis for the participation of telomere and immune cell-related genes in the mechanism of COPD, while the clinical diagnostic value and regulatory mechanism of biomarkers still need to be further verified by large sample clinical.

Abbreviations

COPD, Chronic obstructive pulmonary disease; TRGs, telomere-related genes; DEGs, differentially expressed genes; DICs, different immune cells; WGCNA, weighted gene co-expression network analysis; GSEA, gene set enrichment analysis; TFs, transcription factors; miRNAs, microRNAs; RT-qPCR, reverse transcription quantitative PCR; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; SVM-RFE, support vector machine-recursive feature elimination; DGIdb, Drug-Gene Interaction database; cDCs, conventional dendritic cells; ALDH2, Aldehyde dehydrogenase 2; HNMT, Histamine N-methyltransferase; JAMs, junctional adhesion molecules; IL-6, interleukin-6; DCs, dendritic cells; Th17, T helper 17; Th2, T helper 2; IL-4, interleukin-4; IgE, immunoglobulin E; IgG1, immunoglobulin G1.

Data Sharing Statement

The datasets analyzed in this study are available in the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) and TelNet (<http://www.cancertelsys.org/telnet/>) databases. Datasets used in this study can be obtained from the corresponding author upon request.

Ethics Approval and Informed Consent

This study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the Medical Ethics Committee of Nanjing Brain Hospital (Approval No. 2025-KY078-02, 2025-07-14). For human subjects, written informed consent was obtained after detailed explanation of the study content, risks, and benefits. Clinical blood samples were collected from Nanjing Chest Hospital, comprising 5 patients with chronic obstructive pulmonary disease (COPD) and 5 healthy controls.

Acknowledgments

We would like to express our sincere gratitude to all individuals and organizations who supported and assisted us throughout this research. Special thanks to the following authors: In conclusion, we extend our thanks to everyone who has supported and assisted us along the way. Without your support, this research would not have been possible.

Funding

The authors did not receive support from any organization for the submitted work.

Disclosure

The authors declare that they have no competing interests.

References

- Ren CS, Wang GS, Qian GS. Nosogenesis of chronic obstructive pulmonary disease and the perplexity and hope of treatment. *Chin J Lung Dis.* 2019;12(02):127–141.
- Alfahad AJ, Alzaydi MM, Aldossary AM, et al. Current views in chronic obstructive pulmonary disease pathogenesis and management. *Saudi Pharm J.* 2021;29(12):1361–1373. doi:10.1016/j.jsps.2021.10.008
- Koniali L, Hadjisavvas A, Constantinidou A, et al. Risk factors for breast cancer brain metastases: a systematic review. *Oncotarget.* 2020;11(6):650–669. doi:10.18632/oncotarget.27453
- Zhu Y, Liu X, Ding X, et al. Telomere and its role in the aging pathways: telomere shortening, cell senescence and mitochondria dysfunction. *Biogerontology.* 2019;20(1):1–16. doi:10.1007/s10522-018-9769-1
- Li Y, Lai S, Kan X. Causal relationship between immune cells and telomere length: mendelian randomization analysis. *BMC Immunol.* 2024;25(1):19. doi:10.1186/s12865-024-00610-6
- Como M, Koppala BR, Hasan MN, Han VL, Arora I, Sun D. Cell volume regulation in immune cell function, activation and survival. *Cell Physiol Biochem.* 2021;55(S1):71–88.
- Zhang Y, Xia R, Lv M, et al. Machine-learning algorithm-based prediction of diagnostic gene biomarkers related to immune infiltration in patients with chronic obstructive pulmonary disease. *Front Immunol.* 2022;13:740513. doi:10.3389/fimmu.2022.740513
- Pang X, Liu X. Immune dysregulation in chronic obstructive pulmonary disease. *Immunol Invest.* 2024;53(4):652–694. doi:10.1080/08820139.2024.2334296
- Jing X, Yueqin L. Identification and experimental validation of biomarkers related to MiR-125a-5p in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2025;8(20):581–600. doi:10.2147/COPD.S493749
- Bahr TM, Hughes GJ, Armstrong M, et al. Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol.* 2013;49(2):316–323. doi:10.1165/rcmb.2012-0230OC
- Liu R, Song P, Gu X, et al. Comprehensive landscape of immune infiltration and Aberrant pathway activation in ischemic stroke. *Front Immunol.* 2021;12:766724. doi:10.3389/fimmu.2021.766724
- Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220. doi:10.1186/s13059-017-1349-1
- Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics.* 2022;38(15):3844–3846. doi:10.1093/bioinformatics/btac409
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 2008;9:559. doi:10.1186/1471-2105-9-559
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. doi:10.1186/s13059-014-0550-8
- Gu Z. Complex heatmap visualization. *Imeta.* 2022;1(3):e43. doi:10.1002/imt.2.43
- Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinf.* 2011;12:35. doi:10.1186/1471-2105-12-35
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic.* 2012;16(5):284–287. doi:10.1089/omi.2011.0118
- Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–2504. doi:10.1101/gr.1239303
- Shi H, Yuan X, Liu G, Fan W. Identifying and validating GSTM5 as an immunogenic gene in diabetic foot ulcer using bioinformatics and machine learning. *J Inflamm Res.* 2023;16:6241–6256. doi:10.2147/JIR.S442388
- Robles-Jimenez LE, Aranda-Aguirre E, Castelan-Ortega OA, et al. Worldwide traceability of antibiotic residues from livestock in wastewater and soil: a systematic review. *Animals.* 2021;12(1). doi:10.3390/ani12010060
- Xu J, Yang T, Wu F, Chen T, Wang A, Hou S. A nomogram for predicting prognosis of patients with cervical cerclage. *Heliyon.* 2023;9(11):e21147. doi:10.1016/j.heliyon.2023.e21147
- Wang P, Chen Q, Tang Z, et al. Uncovering ferroptosis in Parkinson's disease via bioinformatics and machine learning, and reversed deducing potential therapeutic natural products. *Front Genet.* 2023;14:1231707. doi:10.3389/fgene.2023.1231707
- Liu C, He Y, Luo J. Application of chest CT imaging feature model in distinguishing squamous cell carcinoma and adenocarcinoma of the lung. *Cancer Manag Res.* 2024;16:547–557. doi:10.2147/CMAR.S462951
- Li H, Gao L, Du J, Ma T, Ye Z, Li Z. Differentially expressed gene profiles and associated ceRNA network in ATG7-Deficient lens epithelial cells under oxidative stress. *Front Genet.* 2022;13:1088943. doi:10.3389/fgene.2022.1088943
- Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. *J Proteome Res.* 2019;18(2):623–632. doi:10.1021/acs.jproteome.8b00702
- Chang J, Wu H, Wu J, et al. Constructing a novel mitochondrial-related gene signature for evaluating the tumor immune microenvironment and predicting survival in stomach adenocarcinoma. *J Transl Med.* 2023;21(1):191. doi:10.1186/s12967-023-04033-6
- Al-Kalisi M, Al-Hajri M, Al-Rai S. Relationship between undernutrition and periodontal diseases among a sample of Yemeni population: a cross-sectional study. *Int J Dent.* 2022;2022:7863531. doi:10.1155/2022/7863531
- Nurwidya F, Damayanti T, Yunus F. The role of innate and adaptive immune cells in the immunopathogenesis of chronic obstructive pulmonary disease. *Tuberc Respir Dis.* 2016;79(1):5–13. doi:10.4046/trd.2016.79.1.5
- Wang W, Lanxiang W, Ouyang C, et al. Identification and validation for biomarkers associated with mitochondrial metabolism in chronic obstructive pulmonary disease. *Front Med Lausanne.* 2025;25(12):1612390. doi:10.3389/fmed.2025.1612390

31. Cheng TC, Hung MC, Wang LH, et al. Histamine N-methyltransferase (HNMT) as a potential auxiliary biomarker for predicting adaptability to anti-HER2 drug treatment in breast cancer patients. *Biomark Res.* 2025;13(1):7. doi:10.1186/s40364-024-00715-5
32. Sasaki H, Sekizawa K. [Clinical strategy of chronic obstructive pulmonary disease in the elderly]. *Nihon Ronen Igakkai Zasshi.* 1993;30(12):999–1004. doi:10.3143/geriatrics.30.999
33. Schulzke JD, Günzel D, John LJ, Fromm M. Perspectives on tight junction research. *Ann N Y Acad Sci.* 2012;1257:1–19. doi:10.1111/j.1749-6632.2012.06485.x
34. Wittekindt OH. Tight junctions in pulmonary epithelia during lung inflammation. *Pflugers Arch.* 2017;469(1):135–147. doi:10.1007/s00424-016-1917-3
35. Park S, Lee PH, Baek AR, et al. Association of the tight junction protein Claudin-4 with lung function and exacerbations in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2021;16:2735–2740. doi:10.2147/COPD.S330674
36. Li JX, Huang XZ, Fu WP, et al. Remote regulation of rs80245547 and rs72673891 mediated by transcription factors C-Jun and CREB1 affect GSTCD expression. *Iscience.* 2023;26(8):107383. doi:10.1016/j.isci.2023.107383
37. Oliver JR, Kushwah R, Wu J, et al. Elf3 plays a role in regulating bronchiolar epithelial repair kinetics following Clara cell-specific injury. *Lab Invest.* 2011;91(10):1514–1529. doi:10.1038/labinvest.2011.100
38. Corbellini C, Rossino E, Massaccesi R, et al. Improvements in perimeter thoracic mobility on patients with COPD after pulmonary rehabilitation: a case series. *Electron J Gen Med.* 2022;19(3):em361. doi:10.29333/ejgm/11671
39. Corbellini C, Tavella S, Gugliotta E, et al. Hypercapnia and functional improvements during pulmonary rehabilitation. *Eur Respir J.* 2021;58(suppl 65):PA1827.
40. Martínez-Pozas O, Corbellini C, Cuenca-Zaldívar JN, et al. Effectiveness of telerehabilitation versus face-to-face pulmonary rehabilitation on physical function and quality of life in people with post COVID-19 condition: a systematic review and network meta-analysis. *Eur J Phys Rehabil Med.* 2024;60(5):868–877. doi:10.23736/S1973-9087.24.08540-X
41. Martínez-Pozas O, Meléndez-Oliva E, Martínez Rolando L, et al. The pulmonary rehabilitation effect on long covid-19 syndrome: A systematic review and meta-analysis. *Physiother Res Int.* 2024;29(2):e2077. doi:10.1002/pri.2077
42. Corbellini C, Villafane J, Gugliotta E, et al. Late breaking abstract-Pulmonary rehabilitation in post-COVID subjects with moderate lung restriction, a case series. *Eur Respir J.* 2021;58(suppl 65):PA2003.
43. Sánchez-Romero EA, García-Barredo-Restegui T, Martínez-Rolando L, et al. Addressing post-COVID-19 musculoskeletal symptoms through pulmonary rehabilitation with an evidence-based eHealth education tool: Preliminary results from a pilot randomized controlled clinical trial. *Medicine.* 2025;104(10):e41583. doi:10.1097/MD.0000000000041583

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