The Association Between Oral Microbiota and Chronic Obstructive Pulmonary Disease: An Integrated Study of Genetic Causal Inference and Bioinformatics Analysis

Zi-Feng Wei[1*](First author, Corresponding author), Yi-Ting Huang[1], Jia-Pei Wuzhang[1], Xin-xin Zhang[2]

[1]*Department of Clinical Medicine, Capital Medical University, Beijing, China*

[2]*Department of Yanjing, Capital Medical University, Beijing, China*

*Corresponding author: Zi-Feng Wei, Department of Clinical Medicine, Capital Medical University, Beijing 100069, China.

Pheno: +86-178-01176584

Email: w2004zfw2004zf@mail.ccmu.edu.cn; w2004zf@qq.com

Keywords: Oral microbiota, Chronic Obstructive Pulmonary Disease, Mendelian randomization, Transcriptomics, Single-cell RNA-seq, Network pharmacology

The Association Between Oral Microbiota and Chronic Obstructive Pulmonary Disease: An Integrated Study of Genetic Causal Inference and Bioinformatics Analysis

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is the third leading cause of global mortality. Emerging evidence suggests the oral microbiome may contribute to COPD progression, though causal relationships remain elusive.

Methods: Using bidirectional Mendelian randomization (MR) on East Asian genome-wide association study (GWAS) summary data, we assessed causal links between oral microbial taxa and COPD risk. Subsequently, hub genes in COPD bulk RNA sequencing data were identified by integrating the Protein-Protein Interaction (PPI) network with machine learning, followed by target validation using single-cell RNA sequencing, immune infiltration analysis, and molecular docking.

Results: Forward MR identified 43 taxa associated with COPD, primarily from genera such as *Fusobacterium*, *Prevotella*, and *Streptococcus*. Reverse MR detected 73 taxa affected by COPD, mainly involving *Campylobacter_A*, *Rothia*, and *Streptococcus*. Through the PPI network, machine learning screening, and multi-omics analysis validation, *MPDZ* emerged as a key hub gene, upregulated in Ciliated and Endothelial cells during early stages of COPD and linked to immune dysregulation. Molecular docking revealed six candidate drugs with strong binding affinity to *MPDZ.*

Conclusion: Our study provides insights for the development of personalized treatment strategies for COPD and offers preliminary candidate targets and drugs for future drug development.

Keywords: Oral microbiota, Chronic Obstructive Pulmonary Disease, Mendelian randomization, Transcriptomics, Single-cell RNA-seq, Molecular docking

1. Introduction

COPD ranks as the third leading cause of death worldwide, characterized primarily by airflow limitation and structural damage to the lungs resulting from chronic inflammation in the airways and pulmonary parenchyma [1]. The disease manifests with chronic inflammatory responses, leading to airway remodeling and destruction of alveolar walls, which clinically present as chronic cough, expectoration, and progressive dyspnea [2]. Recent global epidemiological data indicate that the overall burden of COPD has been increasing since 1990, with rising prevalence and mortality rates observed annually [3, 4]. Environmental exposures, genetic susceptibility, pulmonary inflammation, and oxidative stress are well-recognized mechanisms contributing to COPD pathogenesis [5]. However, these pathways only partially elucidate the complex etiology of COPD. Current research has yet to provide a comprehensive understanding of its pathophysiology or fully effective interventions. Notably, the oral microbiota has emerged as a potential influencing factor and is increasingly becoming a focus of novel investigation in COPD research.

The human oral ecosystem contains over 700 species of microorganisms, including bacteria, viruses, and fungi, whose composition is closely linked to distinct ecological niches such as teeth, saliva, and the dorsum of the tongue [6]. Through complex interspecies interactions such as symbiosis, competition, and antagonism, these microorganisms collectively maintain oral microecological homeostasis with the host. Furthermore, the microbiota plays a significant role in modulating inflammatory signaling pathways, epigenetic modifications, immune responses, and the production of microbial metabolites, influencing the onset and progression of host diseases [7, 8]. Noteworthily, the oral cavity and respiratory tract share a continuous anatomical pathway, providing a route for the translocation of oral microorganisms to the lungs. Oral dysbiosis or the presence of oral diseases may promote or exacerbate respiratory conditions [9, 10]. Recent observational studies have identified characteristic alterations in oral microbial diversity and core genus abundance among COPD

patients across different age groups [11-13]. During acute exacerbations, reduced microbial diversity and increased microbial load are frequently observed in the lungs of affected individuals [14]. Although associations between the oral microbiome and COPD severity and progression have been implicated, the causal relationship and underlying mechanisms require further validation.

Observational studies are often susceptible to confounding variables and reverse causality. In contrast, MR employs single-nucleotide polymorphisms (SNPs) as instrumental variables (IVs) to estimate the exposure-outcome relationship [15]. Since SNPs are randomly allocated during meiosis, with allele transmission being fixed and independent of disease status, MR analysis effectively mitigates the influence of confounding factors and reverse causation. Furthermore, bulk RNA sequencing techniques can be utilized to comprehensively elucidate the potential mechanisms of the oral microbiome in COPD pathogenesis, thus providing further validation for MR findings [16].

The current study performed a bidirectional MR analysis using summary statistics from two GWAS in East Asian populations, systematically evaluating the causal direction between oral microbial traits and COPD. We mapped SNPs associated with relevant microbial taxa to corresponding genes and constructed a PPI network. We also employed multiple topological algorithms to identify key hub genes. Using a combination of Boruta feature selection and artificial neural networks, we applied machine learning to construct a high-performance predictive model and pinpoint core functional genes.

To further validate the biological mechanisms of these genes, we integrated single-cell RNA sequencing data and performed immune infiltration analysis. Finally, molecular docking was employed to screen and evaluate potential targeted therapeutics along with their binding properties. Through this multi-omics integrative approach, our study aims to elucidate the causal role and molecular basis of the oral microbiota in COPD pathogenesis, offering novel insights and potential intervention targets for disease prevention and treatment.

2. Materials and methods

2.1. Study Design

This study employed a bidirectional two-sample MR approach to investigate potential causal relationships between oral microbiota and COPD. All genetic association summary statistics were derived from GWAS conducted in East Asian populations. The MR design relies on three core assumptions: 1. The genetic variants used as instrumental variables must be strongly associated with the specific oral microbial taxa; 2. The selected SNPs should not be associated with any known or unknown confounders influencing the exposure–outcome relationship; 3. The genetic instruments influence COPD risk exclusively through their effect on oral microbiota, rather than via alternative biological pathways [15].

All analyses were performed using publicly available summary-level data. As no individual-level data were collected and no new experiments were conducted, additional ethical approval was not required. The study research process is illustrated in Figure_1.

2.2. Data Sources

Summary statistics for oral microbiota were obtained from a GWAS published by the CNGD database, which investigated oral microbial traits in individuals of East Asian ancestry. This GWAS represents the first large-scale effort of its kind in an East Asian population, incorporating 2,017 dorsum tongue samples and 1,915 saliva samples subjected to high-depth whole-genome sequencing [17]. To ensure data reliability, we applied a comprehensive quality control pipeline for the selection of samples. Inclusion criteria included a minimum 98% variant call rate, >20× mean sequencing depth, principal component analysis confirming no population stratification, and identity-by-descent-based exclusion of related samples. The final dataset comprised approximately 10 million genetic variants, including both common

and low-frequency variants with minor allele frequency⩾0.5%, which were retained for subsequent analysis.

Due to variations in key demographic characteristics (e.g., age, sex, living environment) across individual GWAS, which are known to affect genotype-phenotype associations, and single GWASs are underpowered, we performed a meta-analysis to combine GWAS summary statistics from two independent East Asian studies. This approach enhances statistical power and reduces the likelihood of false-positive findings [18]. We incorporated COPD GWAS summary data from the OpenGWAS project and the EBI database, comprising 7,332 cases and 364,245 controls of East Asian [19, 20]. A fixed-effects meta-analysis was conducted using the METAL software [21]. A summary of the GWAS included in this MR study is presented in Table S1.

2.3. Selection of Instrumental Variables

In the MR analysis, we utilized the clumping procedure in PLINK to identify an adequate number of IVs while minimizing linkage disequilibrium (LD) among associated genetic variants. This approach mitigates multicollinearity arising from LD and reduces potential bias caused by weak instruments [22]. SNPs significantly associated with each oral microbial taxon were selected as IVs. Given the limited number of IVs identified under a strict genome-wide significance threshold ($p < 5\times10^{-8}$), a more inclusive criterion ($p < 1\times10^{-5}$, $r^2 = 0.001$, $kb = 10,000$) was applied to enhance statistical power and improve the robustness of the results. To minimize potential bias, SNPs with a minor allele frequency (MAF) below 0.01 in the original GWAS were excluded. Specific SNPs, including palindromic SNPs and those with an *F*-statistic below 10, were excluded to avoid weak instrumental variables and reduce bias, where the F-statistic is calculated using the formula: $F = (beta/se)^2$. Finally, Steiger filtering was conducted to retain SNPs where the exposure's R-squared was greater than that of the outcome, ensuring the instrumental variables did not exhibit reverse causality.

2.4. Bidirectional MR Analysis

The study employed five complementary statistical methods to examine bidirectional causal relationships between the oral microbiome and COPD: inverse variance weighted (IVW), MR-Egger, weighted median (WM), weighted mode, and simple mode. In analyses with multiple instrumental variables, the IVW method served as the primary analytical approach. The IVW method combines Wald ratio estimates of individual SNPs using meta-analysis techniques to derive an overall causal estimate of the effect of oral microbiota on COPD [23]. The MR-Egger method was applied to detect and adjust for potential horizontal pleiotropy, in addition to estimating causal effects. The WM approach provides consistent causal estimates even when up to 50% of the instrumental variables are invalid. A causal association was considered statistically significant only when the direction of effect estimates from the other four MR methods aligned consistently with that of the IVW method.

2.5. Sensitivity analysis

We conducted sensitivity analyses to validate the accuracy and robustness of the findings. Pleiotropy was assessed using the MR-PRESSO global test and the MR-Egger intercept test, both of which yielded *p*-values greater than 0.05, indicating no evidence of pleiotropy. Heterogeneity was evaluated with Cochran's Q test, which also produced a *p*-value greater than 0.05, suggesting the absence of heterogeneity. Additionally, a leave-one-out sensitivity analysis was performed to determine whether the exclusion of any SNP would alter the inference of the causal relationship.

All MR analyses were conducted using R (version 4.5.1) with the TwoSampleMR and MR-PRESSO packages.

2.6. Mapping SNPs to genes

We utilized SNPnexus to map each queried variant to its nearest gene, including overlapping and upstream/downstream genes. This approach enabled precise gene

annotations, revealing the potential biological significance of variants across different genomic contexts.

SNPnexus (https://www.snp-nexus.org/v4/) is a robust, web-based tool for variant annotation, providing comprehensive and efficient functional analysis of SNPs. By leveraging extensive genomic databases and predictive models, SNPnexus offers high-precision functional annotation, integrating data from resources such as RefSeq and Ensembl [24].

2.7. Functional enrichment analysis of key genes

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to explore the biological functions of diabetic nephropathy (DN)-related genes. GO analysis covers three functional categories: biological process (BP), cellular component (CC), and molecular function (MF), while KEGG provides a publicly available resource on genomes, pathways, diseases, and drugs. These analyses were conducted using the Bioinformatics (https://www.bioinformatics.com.cn/) website.

2.8 Construction of PPI network

The PPI network for a set of 346 potential shared driver genes was generated using the STRING database (https://string-db.org/) with a medium confidence threshold (0.400). Disconnected nodes were excluded from the network, which was subsequently constructed and visualized in Cytoscape. Hub genes within this network were identified and ranked using the CytoHubba plugin by integrating five distinct centrality algorithms: BottleNeck, Betweenness, Degree, EPC, and MCC [25]. The top 30 genes from this aggregated ranking were selected. A Venn diagram was applied to pinpoint the genes consistently ranked among the top candidates by all five algorithms, thereby confirming their robust centrality.

2.9. DEGs screening

Gene expression data were retrieved from the GSE57148 dataset in the GEO database [26], comprising 91 controls and 98 COPD patients of East Asian descent. After initial data processing, differential expression analysis was performed using the 'limma' package. Genes were deemed significant if |logFC| > 0.25 and adjusted *p*-value < 0.05, ensuring biological relevance and false positive control [27]. Subsequently, expression heatmaps and principal component analysis (PCA) plots were generated based on these genes.

Genes identified through CytoHubba screening were intersected with those derived from differential expression analysis of the transcriptome to determine core genes. Visualization was performed using a Venn diagram.

2.10. Machine learning

To identify key genes associated with COPD driven by oral microbiota, machine learning approaches were applied to high-dimensional gene expression data. In this study, the Boruta algorithm was utilized, which operates by comparing the importance of original features with that of randomly generated shadow features and iteratively eliminating irrelevant variables to ensure statistical reliability. The selected genes were modelled using an artificial neural network (ANN) to capture complex relationships. Predictive performance was assessed through receiver operating characteristic (ROC) curve analysis, with the area under the curve (AUC) quantifying model robustness.

2.11. Single-cell analysis

Lung tissue samples from five East Asian COPD patients and three non-smokers were obtained from the GEO database under accession number GSE173896 [28]. The Seurat package initially filtered scRNA-seq data to obtain higher-quality cells. After quality control and data filtering, PCA was performed on the top 2000 highly variable genes. Utilizing UMAP, we achieved unsupervised clustering and unbiased visualization of cell subpopulations. Differential gene expression between clusters was assessed using the "FindAllMarkers" package. The lung tissue module within the

"scMayoMap" package was utilized to annotate subpopulations within each cluster [29].

Additionally, single-cell DEG was applied to annotated lung tissue cells, enabling a direct comparison of target gene expression between COPD patients and healthy controls. Monocle3 was utilized to reconstruct cellular differentiation trajectories. The expression patterns of candidate genes were projected onto the inferred pseudo-temporal trajectory to visualize their dynamic changes during cellular differentiation.

2.12 Immune infiltration

The CIBERSORT tool, employing a gene expression-based deconvolution algorithm, was utilized to assess the relative proportions of twenty-two infiltrating lymphocyte subtypes in each sample group. Pearson correlation analysis was conducted to examine the associations between key genes and immune cell subtypes.

2.13. Drug-gene interactions and protein-protein docking analysis

Drug-gene interactions were explored using DisgDB (https://dsigdb.tanlab.org/DSigDBv1.0/). Drug structures were obtained from the PubChem database (https://pubchem.ncbi.nlm.nih.gov), while protein structures of the candidate genes were retrieved from the Protein Data Bank (PDB, http://www.rcsb.org/). CB-Dock2 (https://cadd.labshare.cn/cb-dock2/index.php) was used for protein-ligand blind docking, and the cavity size and Vina score were evaluated.

3. Results

3.1. Causal impact of the oral microbiome on COPD development

Figure_2 presents the results of the correlation analysis between genetically predicted oral microbiota and COPD risk. Among 3,117 bacterial taxa examined in the forward MR analysis, 43 oral microbial features demonstrated statistically

significant associations with COPD. Of these, 22 taxa (15 from the dorsum tongue and 7 from saliva) exhibited protective effects (OR < 1, Figure_2A, Table S2). While 21 taxa (10 from the dorsum tongue and 11 from saliva) were identified as risk factors (OR > 1, Figure_2B, Table S3). These microorganisms were primarily concentrated within the genus *Streptococcus*, *Prevotella*, *Pauljensenia*, and *Fusobacterium* (Figure_2C). In the reverse MR analysis, 73 microbial taxa were identified as having potential reverse causal relationships with COPD. Among these, 36 features (23 from the dorsum tongue and 13 from saliva) showed protective effects (OR < 1; Figure_2D, Table S4), whereas37 taxa (15 from the dorsum tongue and 22 from saliva) were classified as risk factors (OR > 1; Figure_2E, Table S5). These taxa were predominantly enriched in the genus *Streptococcus*, *Rothia*, and *Campylobacter_A* (Figure_2F).

Notably, we employed five distinct MR analytical approaches, all yielding consistent directional effects and statistical significance. This uniformity underscores the robustness of the causal link between the oral microbiome and COPD. The *F*-statistics for all selected SNPs surpassed the conventional threshold of 10, thereby ruling out weak instrument bias (Table S6-S9). Pleiotropy assessments showed that the MR-Egger regression intercept did not deviate significantly from zero ($p > 0.05$), indicating no evidence of horizontal pleiotropy across the examined taxa. Furthermore, the MR-PRESSO global test detected no outliers, reinforcing that the findings were free from pleiotropic influences. Cochran's Q test revealed no substantial heterogeneity ($p > 0.05$), aligning with the instrumental variable homogeneity assumption. In additional sensitivity checks, the leave-one-out method, in which SNPs were sequentially excluded, produced no major shifts in the causal estimates, suggesting that the results were not driven by any individual SNP (Table S10-S13). Finally, the Steiger directionality test dismissed the likelihood of reverse causation, bolstering confidence in the established causal pathway. Forest, funnel, scatter, and leave-one-out plots for forward MR and reverse MR are shown in Supplementary Figures S14 and S15.

3.2. Genes and functionality

We performed functional annotation on genes mapped by SNPs associated with oral microbiota-related COPD acquisition to elucidate their potential biological functions. As shown in Figure_3A, regarding biological processes, these genes were primarily involved in ameboidal-type cell migration, cell–cell junction assembly, and immune response–regulating signaling pathway, among others. For cellular components, the genes were associated with caveola, catenin complex, cell–substrate junction, and lamellipodium (Figure_3B). Figure_3C illustrates the molecular functions, indicating enrichment in activities such as 3′,5′-cyclic-AMP phosphodiesterase activity, 3′,5′-cyclic-nucleotide phosphodiesterase activity, and low-density lipoprotein particle binding (Table S16). Furthermore, as shown in Figure_3D, pathway enrichment analysis revealed that the key genes were associated with bacterial invasion of epithelial cells, leukocyte transendothelial migration, and morphine addiction, among other pathways (Table S17).

3.3. Identification of PPI Network and Hub Genes

To investigate PPIs among genes, a PPI network was constructed using key module genes identified through SNP mapping, utilizing the STRING database. The resulting network was visualized with Cytoscape software. Hub genes were subsequently identified using the CytoHubba plugin, with the top 30 genes ranked based on multiple topological algorithms, including Bottleneck, Betweenness, Degree, Edge Percolated Component (EPC), and Maximal Clique Centrality (MCC). Through integrated bioinformatics analysis, 22 common hub genes were identified: *AKAP6*, *BCAR1*, *CADM1*, *CDH13*, *CNTNAP2*, *CSMD1*, *CTNNA1*, *DPYSL2*, *DRD3*, *DSCAM*, *ERBB4*, *FBN1*, *H3-4*, *HDAC4*, *MPDZ*, *PRKD3*, *PTPRD*, *RAP1A*, *RARB*, *RBFOX1*, *RBFOX3*, and *RYR2* (Figure_4A-F, Table S18).

3.4. Validation of Differentially Expressed Hub Genes in a GSE Dataset

In the GSE57148 dataset, comprising samples from COPD patients and healthy controls of East Asian descent, a total of 3,644 DEGs were identified, including 1,867

up-regulated and 1,777 down-regulated genes. Five genes (CDH13, FBN1, MPDZ, CADM1, and CTNNA1) overlapping between the DEGs and those derived from the PPI network were identified and visualized in a Venn diagram (Figure_4H-K).

3.5. Identification and Validation of Core Genes for COPD

To identify reliable and critical biomarkers for COPD, the Boruta algorithm was employed, resulting in the selection of three candidate genes: *FBN1*, *MPDZ*, and *CTNNA1*(Figure_5A, Table S19). Based on their respective gene weights, an ANN diagnostic model was developed (Figure_5B). The performance of the model was assessed through receiver operating characteristic (ROC) curve analysis. Typically, an area under the curve (AUC) value ranging from 0.5 to 0.7 indicates low predictive accuracy, while an AUC value between 0.7 and 0.9 suggests moderate predictive accuracy. In the training cohort, the AUC values for *FBN1*, *MPDZ*, and *CTNNA1* were 0.688, 0.764, and 0.620, respectively (Figure_5C and D). These findings suggest that MPDZ demonstrates diagnostic and therapeutic potential for COPD.

3.6. Single-cell analysis results

We analyzed scRNA-seq data from a total of 8 samples, comprising human lung cells from three healthy controls and five patients with COPD. Following preprocessing with stringent quality control criteria, we visualized the high-dimensional scRNA-seq data using the UMAP technique based on the top 13 principal components (Figure_5E, Table S20). We classified the cells into 12 distinct subclusters and annotated them into recognized cell types using the "scMayoMap" package. Major cell types identified included B cells, Basal cells, Capillary aerocytes, Ciliated cells, Endothelial cells, Epithelial cells, Fibroblasts, Lymphoid cells, macrophages, Mesothelial cells, Neutrophils, Pulmonary alveolar type II cells, and T cells (Figure_5F-I).

Furthermore, through single-cell DEG, we observed that *MPDZ* expression was upregulated in Ciliated cells and Endothelial cells from COPD patients compared to healthy individuals (Figure_5J-L, Table S21). Monocle 3 analysis reflects the changes

in *MPDZ* expression along the pseudotime in individual cells (Figure_5M and N, Table S22). In the early stages of COPD, the expression levels of the *MPDZ* gene increase in Ciliated cells and Epithelial cells, and decrease as the disease progresses.

3.7. Immuno-infiltration analysis by CIBERSORT

The cell type proportions in each sample were calculated by the CIBERSORT algorithm. In COPD samples and normal samples, there were differences in the proportions of four types of immune cells, including Eosinophils, Neutrophils, resting NK cells, and T follicular helper cells. We calculated the correlation between the 22 types of infiltrating immune cells. Furthermore, correlation analysis revealed that the key gene *MPDZ* was significantly positively correlated with activated Dendritic cells, resting Mast cells, Monocytes, Neutrophils, and resting memory CD4 T cells, but negatively correlated with memory B cells, resting Dendritic cells, regulatory T cells (Tregs), follicular helper T cells, M1 macrophages, and M2 macrophages in COPD (Figure_6A-E).

3.8. Molecular Docking Evaluation

Based on the above results, six active compounds (Captopril, -5.7 kcal/mol; Camptothecin, -8.0 kcal/mol; Digitoxin, -9.9 kcal/mol; Irinotecan, -8.7 kcal/mol; Pregnenolone, -7.3 kcal/mol; and Trichostatin A, -7.3 kcal/mol) and MPDZ were selected for molecular docking. The optimal docking conformation between the active compounds and the target proteins is visualized in Figure_7A-E. The molecular docking results demonstrated that the binding energies of the effective compounds to the target proteins were consistently less than or equal to -5 kcal/mol, indicating strong binding interactions.

4. Conclusion

Compared with the extensively studied gut microbiome, research on the oral microbiome in relation to COPD remains relatively limited. Observational studies and

conventional microbiome sequencing are often constrained by technical and sampling variability, which can lead to heterogeneous findings, thereby limiting insights into the complex pathogenesis of COPD and the exploration of potential therapeutic strategies. This study employed bidirectional MR to investigate the causal relationship between the oral microbiome and COPD, with a focus on elucidating the biological mechanisms through which oral microbiota contribute to COPD pathogenesis and identifying key genes. Through in-depth analyses of transcriptomic, single-cell, and immune infiltration data, potential therapeutic targets for COPD were identified. The robustness of these findings was further validated via drug prediction and molecular docking analyses. By integrating multi-omics approaches, this research provides a detailed inference regarding the causal role of oral microbiota in COPD and refines the understanding of its underlying pathogenic mechanisms.

This study performed a bidirectional MR analysis using oral microbiome data from the CNGD database and a meta-analyzed COPD GWAS dataset that integrated East Asian population statistics from both the GWAS Catalogue and OpenGWAS project through METAL. The MR results identified significant microbial taxa primarily within the genera *Streptococcus*, *Prevotella*, *Fusobacterium*, and *Pauljensenia*, with notable heterogeneity observed at the species level. The pathogenesis of COPD has been closely associated with dysregulation of pro-inflammatory cytokines. *Streptococcus* polysaccharide capsule reduces the effectiveness of host defense mechanisms, which normally cause its clearance [30]. However, the death of pneumococci, either by the host or the autolysis of itself, not only results in bacterial clearance but also in the release of inflammatory agents such as the pore-forming toxin, pneumolysin, and bacterial cell wall particles that intensify local inflammation [31, 32]. While substantial evidence has established *Streptococcus* as a risk factor for COPD, emerging studies indicate that specific streptococcal species may exert protective effects. For instance, pulmonary *Streptococcus sanguinis* has been shown to suppress NF-κB pathway activation and inhibit IL-8 production triggered by lipopolysaccharide or H_2O_2, attenuating inflammation and oxidative stress in COPD [33]. Similarly, oral

Streptococcus oxalis activates innate immunity and reduces influenza-induced pulmonary inflammation [34]. The heterogeneous associations observed for *Streptococcus* in our study align with this evolving understanding, reinforcing the notion that microbial functions are often species-specific. In contrast, *Prevotella* is generally less frequently implicated as a risk factor for respiratory diseases, with only limited strains demonstrating opportunistic pathogenic potential [35]. Accumulating evidence reveals that *Prevotella* abundance was inversely related to COPD severity in terms of symptoms and positively related to lung function and exercise capacity [36]. Mechanistically, *Prevotella* contributes to the process of airway epithelial gene expression and suppresses inflammatory signaling pathways. Furthermore, within the respiratory microbial ecosystem, *Prevotella* exhibits a low pro-inflammatory immunophenotype. In healthy individuals, it collaborates with other commensal bacteria to induce immune tolerance, facilitating microbial colonization and maintaining mucosal homeostasis [37]. Regarding *Fusobacterium*, existing research presents seemingly contradictory findings. On one hand, *Fusobacterium. nucleatum* (*F. n*) has been identified as a pathogen that induces the production of IL-6 and IL-8, exacerbating respiratory diseases, especially COPD [38, 39]. On the other hand, *F. n* infection has been shown to alleviate aortic atherosclerosis symptoms in animal models [40]. There are indications that *Pauljensenia* may be associated with reduced risks of bronchitis and tonsillitis, though the underlying mechanisms remain unclear [41]. However, as a newly established taxonomic genus, current research has not yet focused on the correlation between *Pauljensenia* and COPD, and its mechanistic role in COPD pathogenesis requires further validation. At the species level, the relationships between most bacterial taxa and disease states remain insufficiently investigated. The complexity of oral microbial communities and substantial inter-individual variability present significant challenges for research. Therefore, further studies are essential to elucidate the role of oral bacteria in respiratory diseases such as COPD, which will facilitate the development of potential therapeutic strategies.

Furthermore, the reverse MR analysis established correlations between COPD and specific oral microbial taxa. Our findings identified significant associations primarily involving *Rothia*, *Campylobacter_A*, and *Streptococcus*. Specifically, COPD appears to promote the proliferation of *Campylobacter_A* and *Rothia*, while simultaneously suppressing *Streptococcus* abundance. These results align with existing research evidence. Multiple cohort studies have demonstrated a significant increase in the relative abundance of *Campylobacter* among COPD patients, with its levels positively correlating with metabolic characteristics of eosinophilic COPD [42, 43]. Time-series analyses further confirm that fluctuations in *Campylobacter* abundance closely parallel dynamic changes in inflammatory markers [44]. In COPD patients with HIV co-infection, *Campylobacter* abundance similarly shows marked elevation, suggesting that compromised immune function and inflammation collectively promote the ecological advantage of this bacterium [45]. *Rothia*, as a common oral-derived bacterium, has been observed to be significantly enriched in lung samples from COPD patients and associated with disease stability or inflammatory indicators [45]. Specifically, its growth rate increases by 2.38-fold in COPD patients compared to healthy controls and shows positive correlations with cytokines such as IL-4 and IL-5, indicating that the inflammatory microenvironment in COPD drives ecological niche expansion of *Rothia* and its potential beneficial roles [46, 47]. Multiple studies have reported reduced abundance of *Streptococcus* in both oral and respiratory tracts of COPD patients, potentially attributable to enhanced inflammation, increased antimicrobial peptide production, or therapeutic interventions. This suppression appears more pronounced during stable disease phases and in high-risk patients [48, 49].

To investigate the biological functions of these genes in COPD, we performed GO and KEGG enrichment analyses. These genes were primarily enriched in amoeboid-like cell migration, cell-cell junction assembly, and immune response regulatory signaling pathways in biological processes. Amoeboid cell migration plays a critical role in immune cell motility and tissue remodeling. Its dysregulation in COPD leads to excessive protease release, exacerbating tissue damage, persistent inflammation, and

airway obstruction. Previous studies have demonstrated that enhanced amoeboid migration of neutrophils and macrophages within lung tissue positively correlates with COPD severity [50]. Cell-cell junction assembly is crucial for maintaining epithelial barrier integrity. Impairment of this process in COPD leads to increased barrier permeability and heightened susceptibility to infection. Cigarette smoke has been shown to disrupt junctional complexes, promoting airway epithelial-mesenchymal transition and accelerating fibrotic progression [51]. Immune response regulatory signaling pathways play a central role in COPD pathogenesis, involving cytokine network-driven chronic inflammation [52]. At the cellular component level, the relevant genes were enriched in Caveolin, junctional protein complexes, and cell-matrix adhesions. Caveolin participates in signal transduction and mechanical protection, with the functional impairment in COPD being associated with oxidative stress and endothelial injury. Reduced expression of caveolin-1 has been correlated with disease severity [53]. Junctional protein complexes maintain intercellular adhesion, and their disruption promotes airway remodeling, accompanied by β-catenin signaling alterations that induce fibrotic responses [54]. Cell-matrix adhesions regulate interactions between cells and the extracellular matrix. Dysregulation of these adhesions in COPD enhances fibroblast activity and promotes collagen deposition [55]. Molecular function analysis revealed significant enrichment in terms such as 3′,5′-cyclic AMP (cAMP) phosphodiesterase activity and low-density lipoprotein (LDL) particle binding. cAMP phosphodiesterase modulates inflammatory responses by degrading cAMP, and its inhibitors have been shown to reduce acute exacerbations in COPD [56]. LDL binding capacity influences lipid metabolism, and oxidized LDL can promote macrophage activation in COPD [57]. KEGG pathway analysis further revealed that these genes are primarily involved in pathways such as bacterial invasion of epithelial cells, leukocyte transendothelial migration, and endocytosis. Bacterial invasion of epithelial cells in COPD contributes to pathogen persistence and exacerbates inflammation [58]. Impaired endocytic clearance of apoptotic cells in COPD leads to persistent inflammatory responses [59]. We identified pathways related

to purine metabolism, mTOR signaling, and various nucleotide sugar biosynthesis. Aberrant purine metabolism in COPD leads to elevated uric acid levels and exacerbates oxidative stress [60]. Hyperactivation of the mTOR signaling pathway may impair cellular autophagy and repair mechanisms [61]. Nucleotide sugar biosynthesis influences glycosylation modifications, thereby regulating mucus secretion and barrier function in COPD [62]. The pathways involving parathyroid hormone (PTH) synthesis, secretion, and action may participate in calcium homeostasis, with elevated PTH levels being associated with COPD severity [63]. The enrichment of the morphine addiction pathway may reflect the interplay between opioid signaling and pain-inflammation regulation, though its specific mechanistic role in COPD requires further investigation [64]. In summary, these enriched pathways reveal complex interactions among inflammatory responses, tissue remodeling, and metabolic regulation during COPD pathogenesis, providing a theoretical foundation for the development of targeted therapeutic strategies.

In this study, machine learning approaches identified *FBN1*, *MPDZ*, and *CTNNA1* as core genes, with *MPDZ* demonstrating diagnostic potential. *FBN1* encodes fibrillin-1, a key component in maintaining extracellular matrix integrity. In COPD, dysregulation of fibrillin-1 exacerbates emphysematous pathology by promoting elastin degradation [65]. *CTNNA1* contributes to epithelial barrier integrity through enhanced cell adhesion, and its reduced expression in COPD is associated with accelerated airway remodeling [66]. *MPDZ*, a regulator of tight junction integrity, was found to be upregulated in ciliated and endothelial cells in COPD, suggesting its role in barrier dysfunction. Single-cell analysis revealed elevated *MPDZ* expression in these cell types. Pseudotime trajectory analysis further demonstrated that *MPDZ* expression is highest during early COPD stages and gradually declines as the disease progresses, aligning with the characteristic epithelial injury observed in COPD. Consistent with the pseudotime findings in advanced COPD, *MPDZ* expression is also significantly reduced in lung cancer tissues, where its decreased expression is associated with poorer patient prognosis. Furthermore, *MPDZ* deficiency has been shown to promote tumor

metastasis and is associated with reduced overall survival in genetically modified mice [67].

Drug prediction analysis identified captopril, camptothecin, digoxin, irinotecan, pregnenolone, and trichostatin A as candidate therapeutics targeting *MPDZ*. Captopril, an ACE inhibitor, has demonstrated anti-inflammatory effects in COPD models [68]. Camptothecin analogues exhibit anti-inflammatory properties in pulmonary diseases [69]. Digoxin, through its modulation of ion channels, may alleviate COPD symptoms [70]. Irinotecan, a topoisomerase inhibitor, potentially suppresses fibrotic pathways [71]. Pregnenolone supports steroidogenesis and may counter hormonal imbalances in COPD [72]. Trichostatin A, an HDAC inhibitor, restores diminished HDAC activity in COPD and enhances anti-inflammatory responses [73]. Molecular docking confirmed strong binding affinities for these compounds, suggesting their therapeutic potential.

Nevertheless, several limitations of this study should be acknowledged. Firstly, the composition of the oral microbiome is multifactorially influenced, including genetic background, lifestyle choices, dietary habits, and environmental exposures, all of which could potentially affect associations. Secondly, the MR analysis focused on East Asian populations, implying that the findings may not be directly generalizable to other ethnic groups, such as those of European ancestry. Additional studies are warranted to validate our results and extend these investigations to diverse populations. Thirdly, the quality and resolution of scRNA-seq data are highly dependent on sampling and technical factors, such as cell capture efficiency and sequencing depth. This dependency may lead to under-representation of certain cell types or subpopulations. Finally, in the field of drug prediction, although the DSigDB provides extensive gene-drug association information, these predictions require further experimental validation to confirm their efficacy and safety. The accuracy of predictive models is also constrained by the coverage of existing datasets and the assumptions inherent to the algorithms employed. Furthermore, drug effects observed at the cellular level may not directly translate to in vivo outcomes. Therefore, further in vivo studies and clinical trials are necessary to verify their practical applicability.

In summary, employing Mendelian randomization, this study pioneers the validation of a novel causal link between oral microbiota and COPD. Furthermore, we utilized SNPnexus to identify COPD-associated genes from relevant SNPs, implicating their role in mediating the influence of oral microbiota on COPD pathogenesis. We also performed single-cell analysis of COPD lung tissues to delineate gene expression across different cell types, revealing cell-type-specific patterns and providing insights into their functional and regulatory roles in COPD development. Using bulk RNA sequencing data, we compared distinct expression profiles in COPD patients versus controls, informing both pathological changes and candidate biomarkers for prognosis and severity assessment. Finally, through drug prediction and molecular docking, we explored several genes as potential therapeutic targets and predicted corresponding candidate drugs. These findings provide genetic evidence to advance personalized COPD intervention and identify promising targets for therapeutic development.

5. Conclusion

The study presents a comprehensive bioinformatics approach to evaluate the relationship between oral microbiota and COPD. Our analysis revealed significant causal associations, with 43 oral microbial taxa identified as having forward causal effects on COPD pathogenesis, and 73 taxa demonstrating reverse causal relationships with the disease. Through the integration of five topological algorithms in CytoHubba, transcriptomic differential expression analysis, single-cell profiling, immune infiltration assessment, and molecular docking, we identified *MPDZ* as a potential key regulator in COPD-related immune modulation. We anticipate that this research will provide precise directions and a solid theoretical foundation for subsequent investigations into the molecular mechanisms of COPD, while also offering clues for discovering novel diagnostic biomarkers and therapeutic strategies.

References

1. Safiri S, Carson-Chahhoud K, Noori M, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019. *BMJ*. 2022;378:e069679. Published 2022 Jul 27. doi:10.1136/bmj-2021-069679
2. 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
3. Wang Z, Lin J, Liang L, et al. Global, regional, and national burden of chronic obstructive pulmonary disease and its attributable risk factors from 1990 to 2021: an analysis for the Global Burden of Disease Study 2021. *Respir Res*. 2025;26(1):2. Published 2025 Jan 2. doi:10.1186/s12931-024-03051-2
4. Adeloye D, Song P, Zhu Y, et al. Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. Lancet Respir Med. 2022;10(5):447-458. doi:10.1016/S2213-2600(21)00511-7
5. Yang IA, Jenkins CR, Salvi SS. Chronic obstructive pulmonary disease in never-smokers: risk factors, pathogenesis, and implications for prevention and treatment. *Lancet Respir Med*. 2022;10(5):497-511. doi:10.1016/S2213-2600(21)00506-3
6. Baker JL, Mark Welch JL, Kauffman KM, McLean JS, He X. The oral microbiome: diversity, biogeography and human health. *Nat Rev Microbiol*. 2024;22(2):89-104. doi:10.1038/s41579-023-00963-6
7. Liu X, Shi F, Zeng J, et al. Oral microbiota and respiratory diseases: advances and perspectives. *Clin Microbiol Rev*. Published online April 2, 2025. doi:10.1128/cmr.00150-24
8. Kaan AMM, Kahharova D, Zaura E. Acquisition and establishment of the oral microbiota. *Periodontol 2000*. 2021;86(1):123-141. doi:10.1111/prd.12366

9. Lin Z, Jiang Y, Liu H, et al. Airway microbiota and immunity associated with chronic obstructive pulmonary disease severity. *J Transl Med.* 2025:23(1):962. Published 2025 Aug 26. doi:10.1186/s12967-025-06986-2
10. Yan Z, Chen B, Yang Y, et al. Multi-omics analyses of airway host-microbe interactions in chronic obstructive pulmonary disease identify potential therapeutic interventions. *Nat Microbiol*. 2022;7(9):1361-1375. doi:10.1038/s41564-022-01196-8
11. Bouquet J, Tabor DE, Silver JS, et al. Microbial burden and viral exacerbations in a longitudinal multicenter COPD cohort. *Respir Res*. 2020;21(1):77. Published 2020 Mar 30. doi:10.1186/s12931-020-01340-0
12. Taherkhani H, KavianFar A, Aminnezhad S, et al. Deciphering the impact of microbial interactions on COPD exacerbation: An in-depth analysis of the lung microbiome. *Heliyon*. 2024;10(4):e24775. Published 2024 Feb 7. doi:10.1016/j.heliyon.2024.e24775
13. Gao J, Yang Y, Xiang X, et al. Human genetic associations of the airway microbiome in chronic obstructive pulmonary disease. *Respir Res*. 2024;25(1):165. Published 2024 Apr 16. doi:10.1186/s12931-024-02805-2
14. Burcham ZM, Garneau NL, Comstock SS, et al. Patterns of Oral Microbiota Diversity in Adults and Children: A Crowdsourced Population Study. *Sci Rep*. 2020;10(1):2133. Published 2020 Feb 7. doi:10.1038/s41598-020-59016-0
15. de Leeuw C, Savage J, Bucur IG, Heskes T, Posthuma D. Understanding the assumptions underlying Mendelian randomization. *Eur J Hum Genet*. 2022;30(6):653-660. doi:10.1038/s41431-022-01038-5
16. Yazar S, Alquicira-Hernandez J, Wing K, et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science*. 2022;376(6589):eabf3041. doi:10.1126/science.abf3041
17. Liu X, Tong X, Zhu J, et al. Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. *Cell Discov*. 2021;7(1):117. Published 2021 Dec 7. doi:10.1038/s41421-021-00356-0

18. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet*. 2013;14(6):379-389. doi:10.1038/nrg3472
19. Ishigaki K, Akiyama M, Kanai M, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*. 2020;52(7):669-679. doi:10.1038/s41588-020-0640-3
20. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415-1424. doi:10.1038/s41588-021-00931-x
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
23. Su Y, Xu S, Ma Y, et al. A Modified Debiased Inverse-Variance Weighted Estimator in Two-Sample Summary-Data Mendelian Randomization. *Stat Med*. 2024;43(29):5484-5496. doi:10.1002/sim.10245
24. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res*. 2020;48(W1):W185-W192. doi:10.1093/nar/gkaa420
25. Ma H, He Z, Chen J, Zhang X, Song P. Identifying of biomarkers associated with gastric cancer based on 11 topological analysis methods of CytoHubba. *Sci Rep*. 2021;11(1):1331. Published 2021 Jan 14. doi:10.1038/s41598-020-79235-9
26. Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM. Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. *Int J Genomics*. 2015;2015:206937. doi:10.1155/2015/206937
27. Shen B, Yi X, Sun Y, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell*. 2020;182(1):59-72.e15. doi:10.1016/j.cell.2020.05.032

28. Watanabe N, Fujita Y, Nakayama J, et al. Anomalous Epithelial Variations and Ectopic Inflammatory Response in Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol*. 2022;67(6):708-719. doi:10.1165/rcmb.2021-0555OC
29. Yang L, Ng YE, Sun H, et al. Single-cell Mayo Map (scMayoMap): an easy-to-use tool for cell type annotation in single-cell RNA-sequencing data analysis. *BMC Biol*. 2023;21(1):223. Published 2023 Oct 20. doi:10.1186/s12915-023-01728-6
30. Leung JM, Tiew PY, Mac Aogáin M, et al. The role of acute and chronic respiratory colonization and infections in the pathogenesis of COPD. *Respirology*. 2017;22(4):634-650. doi:10.1111/resp.13032
31. Jahn K, Kohler TP, Swiatek LS, Wiebe S, Hammerschmidt S. Platelets, Bacterial Adhesins and the Pneumococcus. *Cells*. 2022;11(7):1121. Published 2022 Mar 25. doi:10.3390/cells11071121
32. Kruckow KL, Zhao K, Bowdish DME, Orihuela CJ. Acute organ injury and long-term sequelae of severe pneumococcal infections. *Pneumonia (Nathan)*. 2023;15(1):5. Published 2023 Mar 5. doi:10.1186/s41479-023-00110-y
33. Goeteyn E, Grassi L, Van den Bossche S, et al. Commensal bacteria of the lung microbiota synergistically inhibit inflammation in a three-dimensional epithelial cell model. *Front Immunol*. 2023;14:1176044. Published 2023 Apr 21. doi:10.3389/fimmu.2023.1176044
34. Zou X, Cao H, Hong L, et al. Enrichment of Streptococcus oralis in respiratory microbiome enhance innate immunity and protects against influenza infection. *Signal Transduct Target Ther*. 2025;10(1):272. Published 2025 Aug 27. doi:10.1038/s41392-025-02365-x
35. Russo C, Colaianni V, Ielo G, Valle MS, Spicuzza L, Malaguarnera L. Impact of Lung Microbiota on COPD. *Biomedicines*. 2022;10(6):1337. Published 2022 Jun 6. doi:10.3390/biomedicines10061337
36. Larsen JM, Musavian HS, Butt TM, Ingvorsen C, Thysen AH, Brix S. Chronic obstructive pulmonary disease and asthma-associated Proteobacteria, but not commensal Prevotella spp., promote Toll-like receptor 2-independent lung

inflammation and pathology. *Immunology*. 2015;144(2):333-342. doi:10.1111/imm.12376

37. Ramsheh MY, Haldar K, Esteve-Codina A, et al. Lung microbiome composition and bronchial epithelial gene expression in patients with COPD versus healthy individuals: a bacterial 16S rRNA gene sequencing and host transcriptomic analysis. *Lancet Microbe*. 2021;2(7):e300-e310. doi:10.1016/S2666-5247(21)00035-5
38. Brennan CA, Garrett WS. Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. *Nat Rev Microbiol*. 2019;17(3):156-166. doi:10.1038/s41579-018-0129-6
39. Li Q, Wang H, Tan L, et al. Oral Pathogen *Fusobacterium nucleatum* Coaggregates With *Pseudomonas aeruginosa* to Modulate the Inflammatory Cytotoxicity of Pulmonary Epithelial Cells. *Front Cell Infect Microbiol*. 2021;11:643913. Published 2021 Mar 19. doi:10.3389/fcimb.2021.643913
40. Velsko IM, Chukkapalli SS, Rivera-Kweh MF, et al. Fusobacterium nucleatum Alters Atherosclerosis Risk Factors and Enhances Inflammatory Markers with an Atheroprotective Immune Response in ApoE(null) Mice. *PLoS One*. 2015;10(6):e0129795. Published 2015 Jun 16. doi:10.1371/journal.pone.0129795
41. He J, Mao N, Lyu W, et al. Association between oral microbiome and five types of respiratory infections: a two-sample Mendelian randomization study in east Asian population. *Front Microbiol*. 2024;15:1392473. Published 2024 Apr 10. doi:10.3389/fmicb.2024.1392473
42. Wang Z, Locantore N, Haldar K, et al. Inflammatory Endotype-associated Airway Microbiome in Chronic Obstructive Pulmonary Disease Clinical Stability and Exacerbations: A Multicohort Longitudinal Analysis. *Am J Respir Crit Care Med*. 2021;203(12):1488-1502. doi:10.1164/rccm.202009-3448OC
43. Tao M, Lu G, Zhang S, Li P. Root causes of long-term complications of gestational diabetes mellitus: Metabolic disturbances of the host and gut microbiota. *Clin Chim Acta*. 2023;548:117490. doi:10.1016/j.cca.2023.117490

44. Xiao W, Chen YL, Du LY, et al. Bacterial interactome disturbance in chronic obstructive pulmonary disease: clinical stability and exacerbations. *Respir Res*. 2024;25(1):173. Published 2024 Apr 20. doi:10.1186/s12931-024-02802-5
45. Kayongo A, Bartolomaeus TUP, Birkner T, et al. Sputum Microbiome and Chronic Obstructive Pulmonary Disease in a Rural Ugandan Cohort of Well-Controlled HIV Infection. *Microbiol Spectr*. Published online February 15, 2023. doi:10.1128/spectrum.02139-21
46. Bowerman KL, Rehman SF, Vaughan A, et al. Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nat Commun*. 2020;11(1):5886. Published 2020 Nov 18. doi:10.1038/s41467-020-19701-0
47. Xue Q, Xie Y, He Y, et al. Lung microbiome and cytokine profiles in different disease states of COPD: a cohort study. *Sci Rep*. 2023;13(1):5715. Published 2023 Apr 7. doi:10.1038/s41598-023-32901-0
48. Yang CY, Li SW, Chin CY, et al. Association of exacerbation phenotype with the sputum microbiome in chronic obstructive pulmonary disease patients during the clinically stable state. *J Transl Med*. 2021;19(1):121. Published 2021 Mar 23. doi:10.1186/s12967-021-02788-4
49. Kou Z, Liu K, Qiao Z, et al. The alterations of oral, airway and intestine microbiota in chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Front Immunol*. 2024;15:1407439. Published 2024 May 8. doi:10.3389/fimmu.2024.1407439
50. Imamoto T, Kawasaki T, Sato H, et al. Different Transcriptome Features of Peripheral Blood Mononuclear Cells in Non-Emphysematous Chronic Obstructive Pulmonary Disease. *Int J Mol Sci*. 2023;25(1):66. Published 2023 Dec 20. doi:10.3390/ijms25010066
51. Huang ZQ, Liu J, Sun LY, et al. Updated epithelial barrier dysfunction in chronic rhinosinusitis: Targeting pathophysiology and treatment response of tight junctions. *Allergy*. 2024;79(5):1146-1165. doi:10.1111/all.16064

52. Qi Y, Yan Y, Tang D, et al. Inflammatory and Immune Mechanisms in COPD: Current Status and Therapeutic Prospects. *J Inflamm Res*. 2024;17:6603-6618. Published 2024 Sep 20. doi:10.2147/JIR.S478568
53. Zhang T, Shang F, Ma Y, Xu Y, Sun W, Song H. Caveolin-1 Promotes the Imbalance of Th17/Treg in Chronic Obstructive Pulmonary Disease by Regulating Hsp70 Expression. *Int J Chron Obstruct Pulmon Dis*. 2023;18:565-574. Published 2023 Apr 12. doi:10.2147/COPD.S398780
54. Shi J, Li F, Luo M, Wei J, Liu X. Distinct Roles of Wnt/*β*-Catenin Signaling in the Pathogenesis of Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis. *Mediators Inflamm*. 2017;2017:3520581. doi:10.1155/2017/3520581
55. Wang T, Dong Y, Fang L, Zhou H. Patterns and Underlying Mechanisms of Airway Epithelial Cell Death in COPD. *COPD*. 2025;22(1):2542153. doi:10.1080/15412555.2025.2542153
56. Nourian YH, Salimian J, Ahmadi A, et al. cAMP-PDE signaling in COPD: Review of cellular, molecular and clinical features. *Biochem Biophys Rep*. 2023;34:101438. Published 2023 Feb 16. doi:10.1016/j.bbrep.2023.101438
57. Raby KL, Michaeloudes C, Tonkin J, Chung KF, Bhavsar PK. Mechanisms of airway epithelial injury and abnormal repair in asthma and COPD. *Front Immunol*. 2023;14:1201658. Published 2023 Jul 13. doi:10.3389/fimmu.2023.1201658
58. Love ME, Proud D. Respiratory Viral and Bacterial Exacerbations of COPD-The Role of the Airway Epithelium. *Cells*. 2022;11(9):1416. Published 2022 Apr 22. doi:10.3390/cells11091416
59. Marzec JM, Nadadur SS. Inflammation resolution in environmental pulmonary health and morbidity. *Toxicol Appl Pharmacol*. 2022;449:116070. doi:10.1016/j.taap.2022.116070
60. Esther CR Jr, Lazaar AL, Bordonali E, Qaqish B, Boucher RC. Elevated airway purines in COPD. *Chest*. 2011;140(4):954-960. doi:10.1378/chest.10-2471

61. Goyal A, Chopra V, Garg K, Sharma S. Mechanisms coupling the mTOR pathway to chronic obstructive pulmonary disease (COPD) pathogenesis. *Cytokine Growth Factor Rev*. 2025;82:55-69. doi:10.1016/j.cytogfr.2024.12.005
62. Xie X, Kong S, Cao W. Targeting protein glycosylation to regulate inflammation in the respiratory tract: novel diagnostic and therapeutic candidates for chronic respiratory diseases. *Front Immunol*. 2023;14:1168023. Published 2023 May 15. doi:10.3389/fimmu.2023.1168023
63. Amado CA, Muñoz P, García-Unzueta M, et al. High parathyroid hormone predicts exacerbations in COPD patients with hypovitaminosis D. *Respir Med*. 2021;182:106416. doi:10.1016/j.rmed.2021.106416
64. Fu Y, Chapman EJ, Boland AC, Bennett MI. Evidence-based management approaches for patients with severe chronic obstructive pulmonary disease (COPD): A practice review. *Palliat Med*. 2022;36(5):770-782. doi:10.1177/02692163221079697
65. Tun MH, Borg B, Godfrey M, Hadley-Miller N, Chan ED. Respiratory manifestations of Marfan syndrome: a narrative review. *J Thorac Dis*. 2021;13(10):6012-6025. doi:10.21037/jtd-21-1064
66. Werder RB, Liu T, Abo KM, et al. CRISPR interference interrogation of COPD GWAS genes reveals the functional significance of desmoplakin in iPSC-derived alveolar epithelial cells. *Sci Adv*. 2022;8(28):eabo6566. doi:10.1126/sciadv.abo6566
67. Liu W, Huang Y, Wang D, et al. MPDZ as a novel epigenetic silenced tumor suppressor inhibits growth and progression of lung cancer through the Hippo-YAP pathway. *Oncogene*. 2021;40(26):4468-4485. doi:10.1038/s41388-021-01857-8
68. Hosseini FS, Malektojari A, Ghazizadeh S, et al. The efficacy and safety of Ivermectin in patients with mild and moderate COVID-19: A structured summary of a study protocol for a randomized controlled trial. *Trials*. 2021;22(1):4. Published 2021 Jan 4. doi:10.1186/s13063-020-04988-7

69. Shang XF, Morris-Natschke SL, Yang GZ, et al. Biologically active quinoline and quinazoline alkaloids part II. *Med Res Rev*. 2018;38(5):1614-1660. doi:10.1002/med.21492
70. Zhang K, Zhou F, Zhu C, et al. Role of Digoxin in Preventing Cigarette Smoke-Induced COPD via HIF-1α Inhibition in a Mouse Model. *Int J Chron Obstruct Pulmon Dis*. 2025;20:1665-1678. Published 2025 May 23. doi:10.2147/COPD.S493856
71. Yamaguchi H, Soda H, Kitazaki T, et al. Serum progastrin-releasing peptide levels followed by whole-body positron emission tomography detects early recurrence of small-cell lung cancer. *Respirology*. 2007;12(1):137-139. doi:10.1111/j.1440-1843.2006.00966.x
72. Sathish V, Martin YN, Prakash YS. Sex steroid signaling: implications for lung diseases. *Pharmacol Ther*. 2015;150:94-108. doi:10.1016/j.pharmthera.2015.01.007
73. Ding J, Li F, Cong Y, et al. Trichostatin A inhibits skeletal muscle atrophy induced by cigarette smoke exposure in mice. *Life Sci*. 2019;235:116800. doi:10.1016/j.lfs.2019.116800

Author contributions

Zi-feng Wei: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review and editing. Yi-ting Huang: Data curation, Methodology, Supervision, and Visualization. Jia-pei Wuzhang: Writing - original draft, Writing - review and editing. Xin-xin Zhang: Data curation

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The GWAS summary data for COPD analyzed in the current study are sourced from bbj-a-103 and GCST90018587 (OpenGWAS project: https://opengwas.io/datasets/; GWAS Catalog: https://www.ebi.ac.uk/gwas/). The GWAS summary data for salivary and dorsal tongue microbiota are sourced from CNP0001664 (CNGB database: https://ftp2.cngb.org/). The RNA bulk and single-cell data for COPD are sourced from GSE57148 and GSE173896 (GEO database: https://www.ncbi.nlm.nih.gov/geo/).

Declarations

Competing interests:

The authors declare no competing interests.

Ethics statement:

The present study does not require ethical approval.

Figure_1

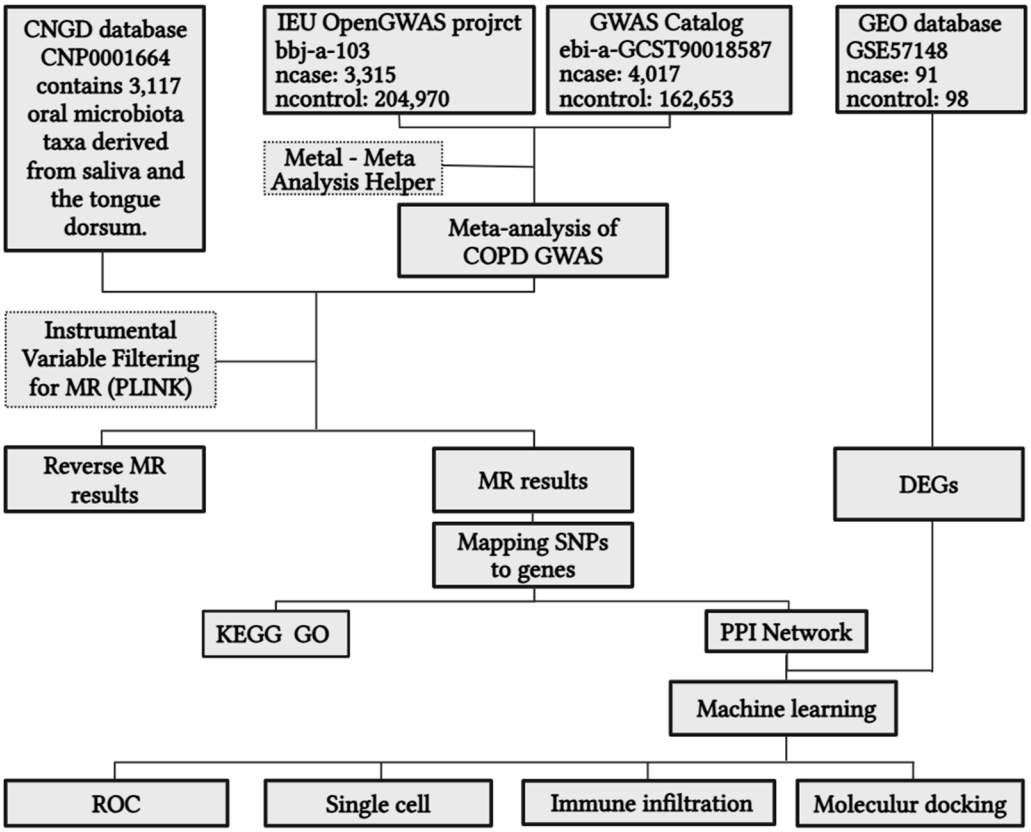

Figure_1 The flowchart of the study.

Figure_2

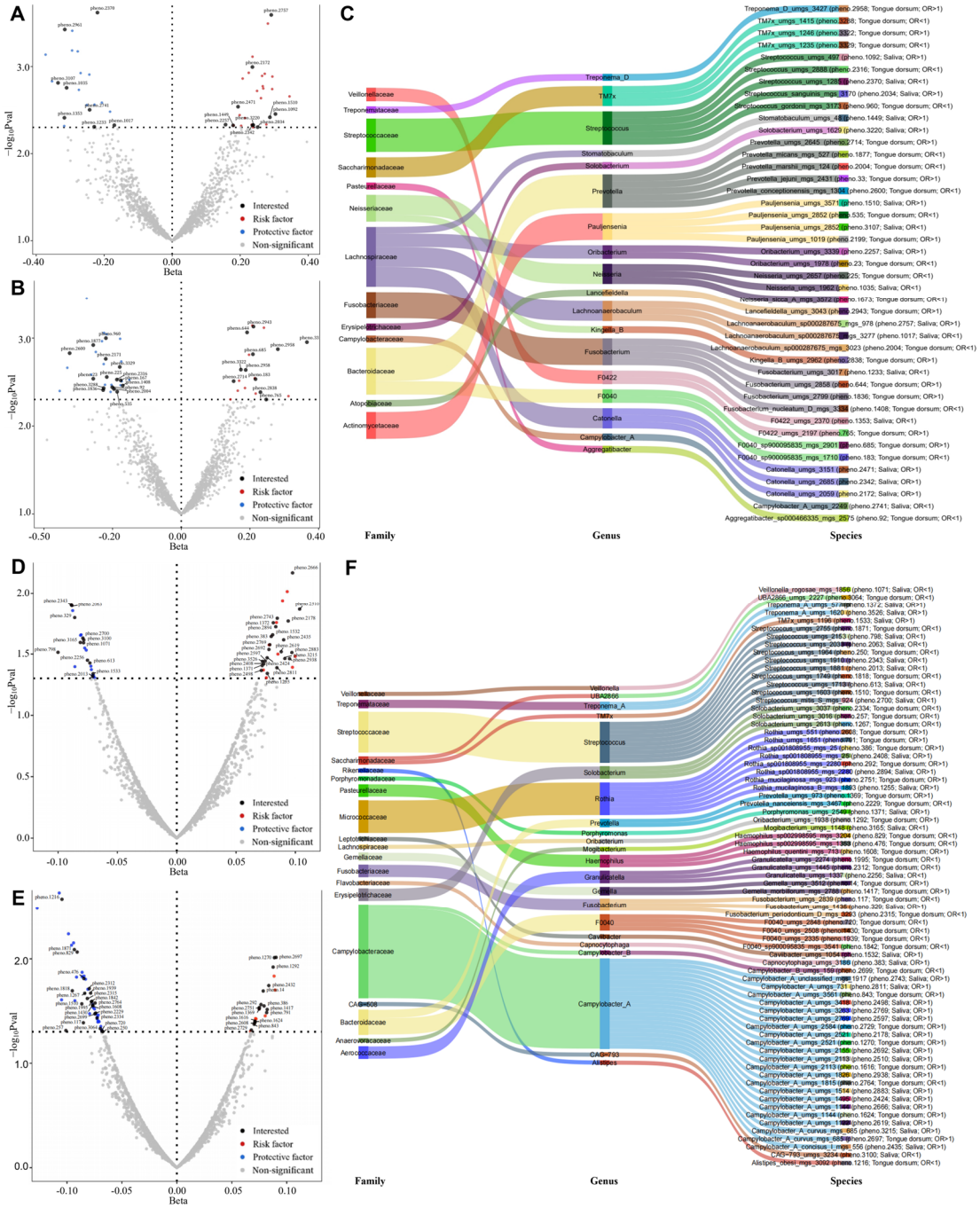

Figure_2 Volcano plots illustrating the MR results for associations between salivary microbiota (A), dorsal tongue microbiota (B), and COPD (gray: Non-significant; black: Interested; blue: Protective factor; red: Risk factor). (C) Taxonomic summary of salivary and dorsal tongue microbiota (family, genus, species levels) from MR analyses. Volcano plots illustrating the reverse MR results for associations between salivary microbiota (D), dorsal tongue microbiota (E), and COPD. (F) Taxonomic

summary of salivary and dorsal tongue microbiota (family, genus, species levels) from reverse MR analyses.

Figure_3

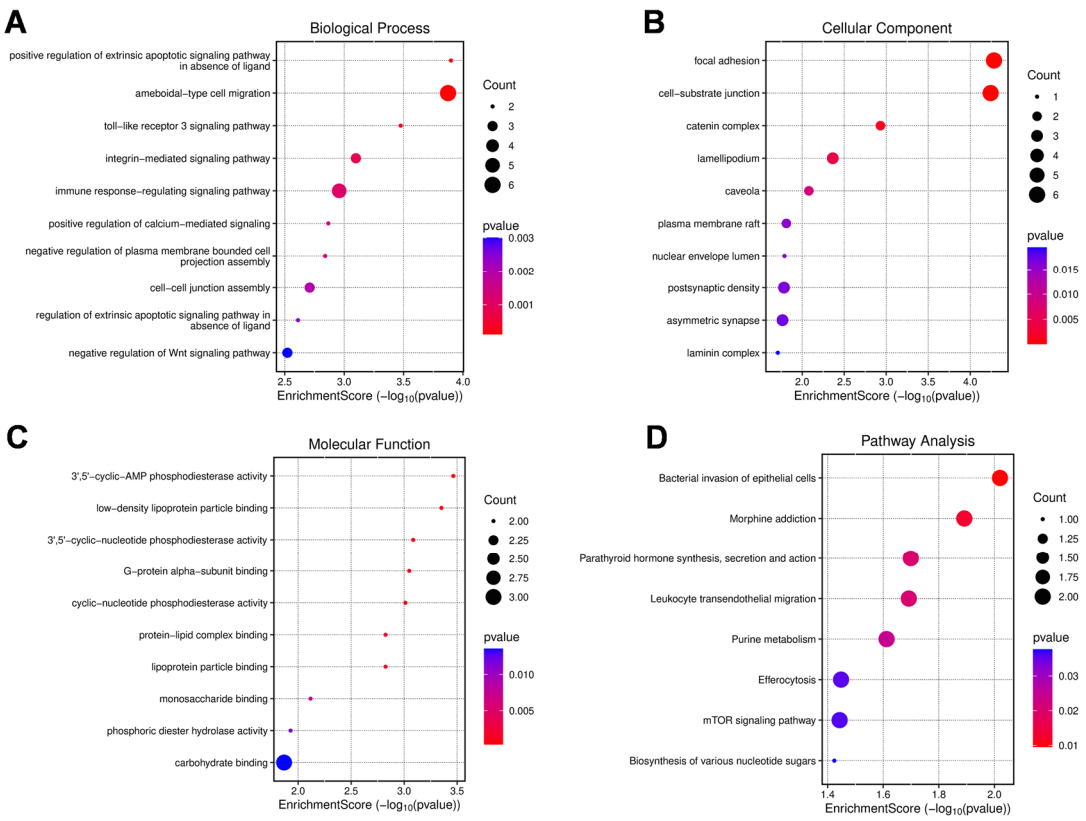

Figure_3 GO and KEGG enrichment analysis of differentially expressed genes. Bubble plots where size denotes gene count and color indicates -log10 (adjusted P value), with darker hues for greater significance. (A) GO biological processes. (B) GO cellular components. (C) GO molecular functions. (D) KEGG pathways.

Figure_4

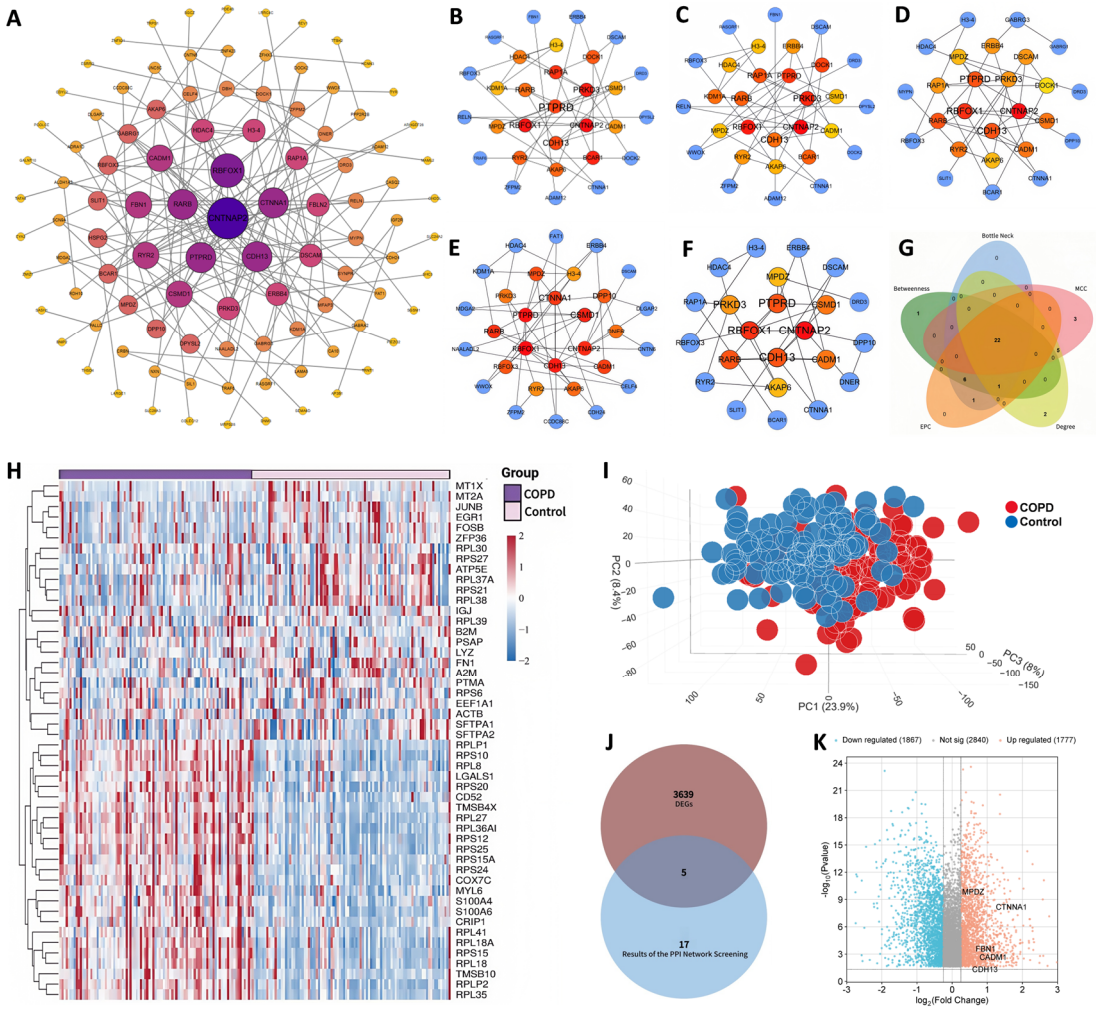

Figure_4 (A) Cytoscape visualization of interconnected protein nodes, excluding isolates. Top 30 hub genes from CytoHubba, ranked by (B) Betweenness, (C) Bottleneck, (D) EPC, (E) Degree, and (F) MCC scores. (G) Venn diagram of overlaps across the five algorithms. (H) Heatmap of DEGs, clustered by COPD versus control expression. (I) PCA plot separating COPD and control samples along PC1, PC2, and PC3. (J) Venn diagram of PPI screening results, with five genes at key intersections. (K) Volcano plot of DEGs, showing log2FC versus -log10 (Pvalue), with upregulated, downregulated, and non-significant genes.

Figure_5

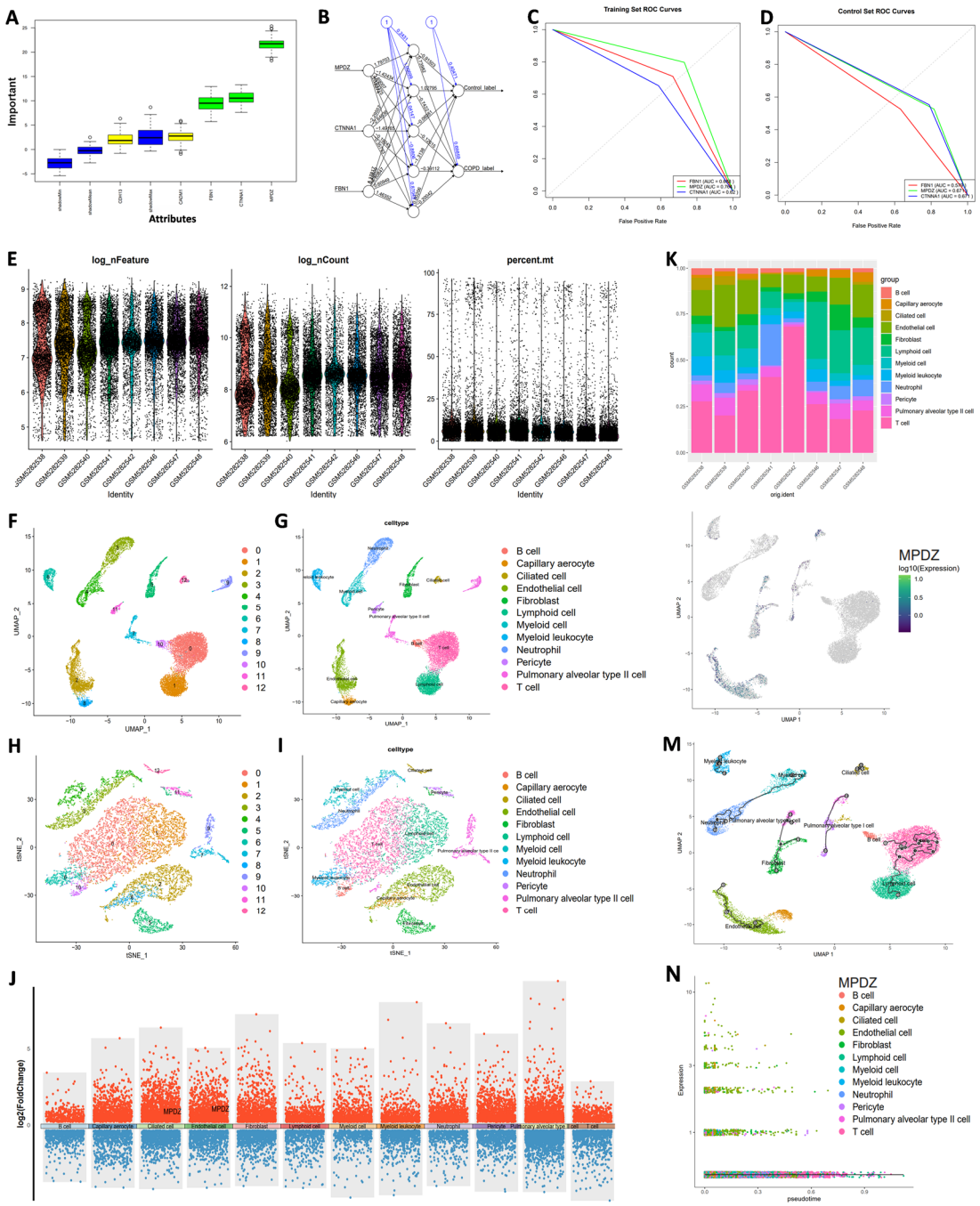

Figure_5 (A) Boruta feature importance bar plot with cross-validation error bars; green: confirmed, yellow: tentative, blue: rejected. (B) The Artificial Neural Networks (ANN) model of the key genes. (C) Training set ROC curves, with area under the curve (AUC) values labeled for each gene. (D) Control set ROC curves. (E) Sample quality control violin plots: log_nFeature, log_nCount, percent.mt. (F) UMAP plot displaying 13 distinct clusters from single-cell RNA sequencing. each identified by a

unique color. (G) UMAP annotated with cell types using scMayoMap. (H) t-SNE colored and clustered as in F. (I) t-SNE plot annotated with cell types using scMayoMap, with colors matching G. (J) Dot plot of single-cell differential expression across cell types in patients, highlighting *MPDZ* distribution. (K) Stacked bar chart representing the cell type composition across different patient samples. (L) *MPDZ* Expression in Specific Cell Types. (M) UMAP pseudotime trajectory annotated by cell types. (N) *MPDZ* expression along pseudotime by cell type.

Figure_6

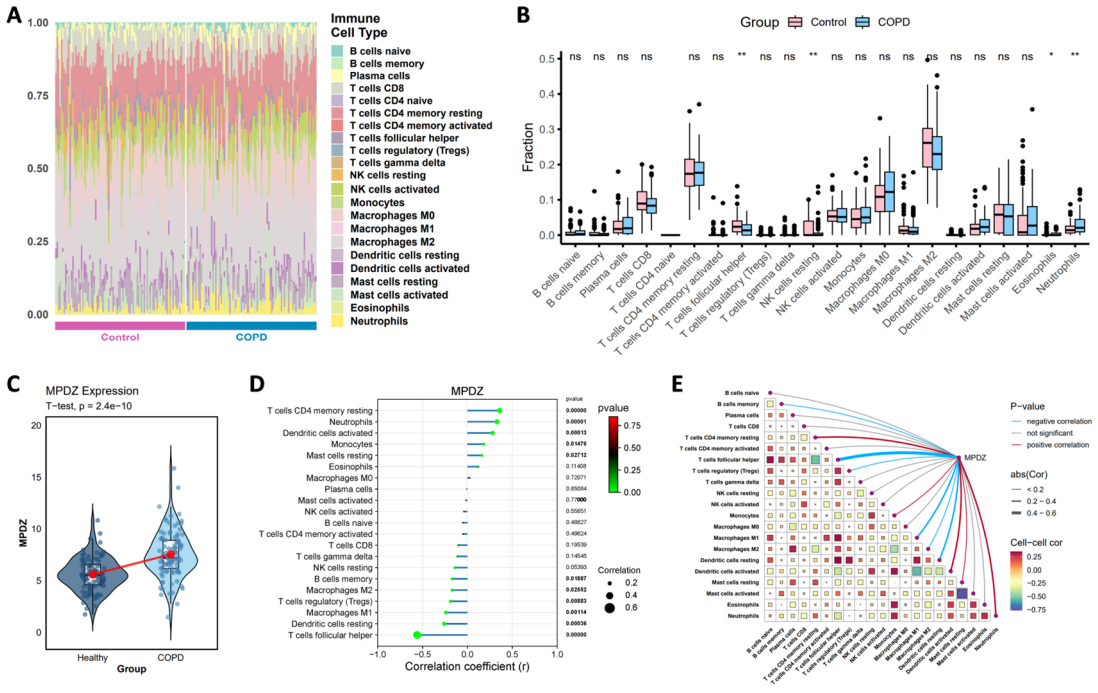

Figure_6 (A) Stacked bar plots illustrating the proportional distribution of immune cell types across individual samples in control and COPD cohorts. (B) Bar plots showing mean fractions of specific immune cell subsets in control versus COPD groups, with individual data points overlaid; statistical significance denoted by ns (not significant), * ($P < 0.05$), and ** ($P < 0.01$). (C) Violin plot of *MPDZ* gene expression levels in healthy and COPD samples. (D) Lollipop plot depicting Pearson correlation coefficients (r) between MPDZ expression and immune cell fractions, ordered by strength. (E) Radial correlation plot visualizing associations between MPDZ and immune cell types, with line thickness and color indicating correlation magnitude (red for positive, blue for negative) and direction; p-value significance shown by color gradient ($P < 0.05$ in shades, non-significant in gray).

Figure_7

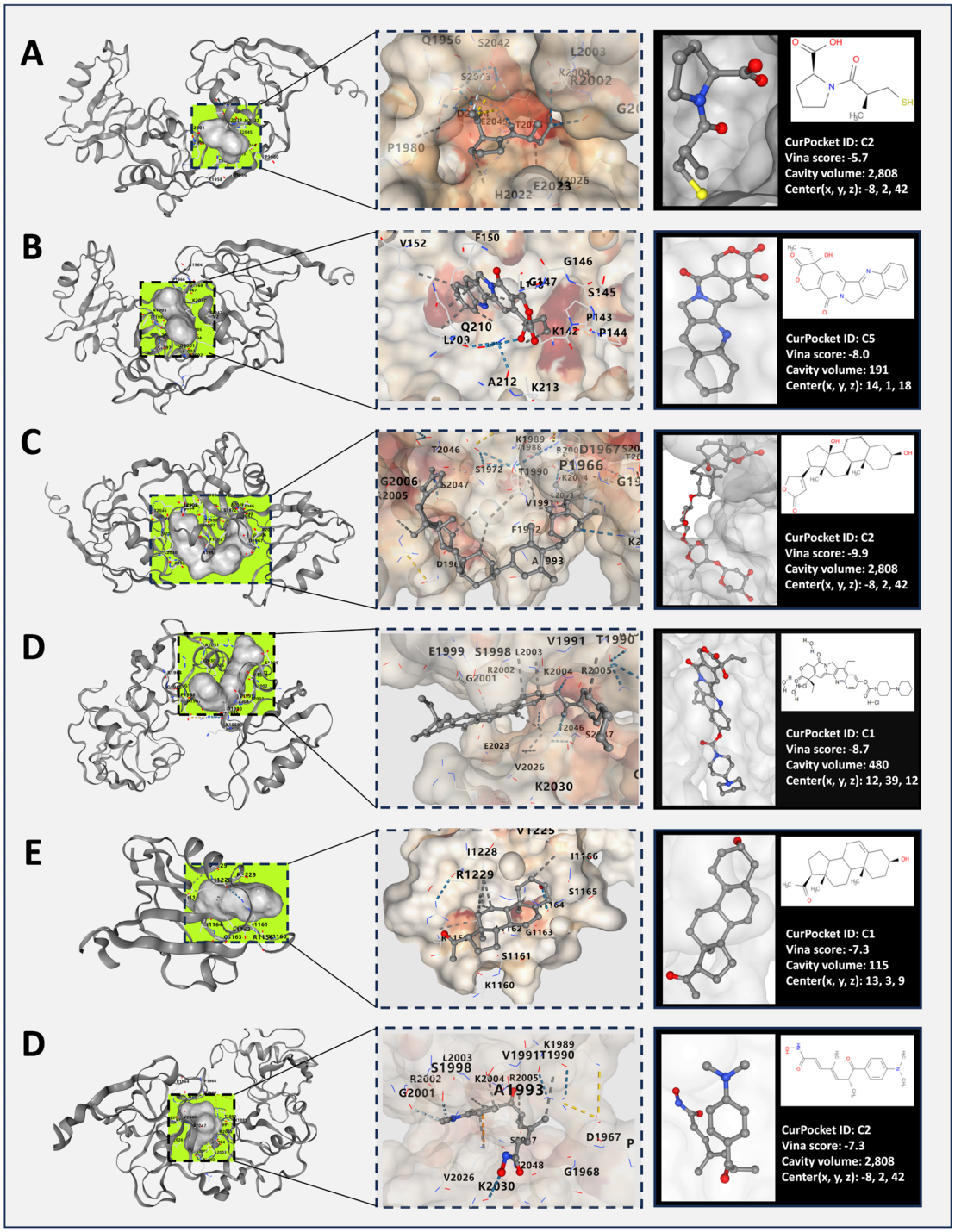

Figure_7 Molecular docking of compounds with *MPDZ*. (A) Captopril, (B) Camptothecin, (C) Digitoxin, (D) Coumarin, (E) Pregnenolone, and (F) Trichostatin A.