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

Xiao-Jie An[1], Zi-Feng Wei[2*](Co-first author, Corresponding author), Yi-Ting Huang[2], Jia-Pei Wuzhang[2], Xin-Xin Zhang[3], Hao-Yu Li[2], Run-Ze Liu[2]

[1]*Teaching and Research Office, Liangxiang Hospital of Fangshan District, Beijing, China*

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

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

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

Pheno: +86-17801176584

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

Keywords: Chronic obstructive pulmonary disease, Oral microbiota, East Asian, Mendelian randomization, Multi-omics, Molecular docking

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 48 taxa associated with COPD, primarily from genera such as *Fusobacterium*, *Prevotella*, and *Streptococcus*. Reverse MR detected 79 taxa affected by COPD, mainly involving *Campylobacter*, *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 cells 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: Chronic obstructive pulmonary disease, Oral microbiota, East Asian, Mendelian randomization, Multi-omics, Molecular docking

1. Introduction

COPD ranks as the third leading cause of death worldwide, characterized primarily by airflow limitation and structural lung damage driven by chronic inflammation of the airways and pulmonary parenchyma [1]. The disease is driven by chronic inflammatory responses that cause airway remodeling and alveolar wall destruction, clinically manifesting 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 explain the complex etiology of COPD. Current research has yet to fully elucidate its pathophysiology or yield truly 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]. Notably, 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 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 the oral microbiome has been implicated in COPD severity and progression, the causal relationships 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. Eleven topological algorithms in CytoHubba were then employed to identify key hub genes. The resulting candidates were further refined using 1,000 iterations of Lasso regression combined with 1,000 rounds of 10-fold cross-validation to 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 potential targeted therapeutics and evaluate their binding affinities. 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 overall study workflow is illustrated in Figure 1.

2.2. Data Sources

Summary statistics for oral microbiota were obtained from a GWAS from the CNGD database, which investigated oral microbial traits in an East Asian population using 2,017 dorsum tongue samples and 1,915 saliva samples with high-depth whole-genome sequencing [17]. A comprehensive quality control pipeline was applied, including 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 with minor allele frequency ≥ 0.5%.

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 the limited statistical power of single GWAS, 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 GWAS Catalog database, comprising 7,332 cases and 364,245 controls of East Asian ancestry [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 Supplementary Table 1.

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, 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 estimate causal effects while detecting and adjusting for potential horizontal pleiotropy. The weighted median 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.

All MR analyses were conducted using R (version 4.5.3) with the TwoSampleMR and MR-PRESSO packages.

2.6. Mapping SNPs to genes

We utilized SNPnexus (https://www.snp-nexus.org/v4/), a web-based variant annotation tool supporting both GRCh37/hg19 and GRCh38/hg38 reference assemblies, to map each queried variant to its nearest gene, including overlapping, upstream, and downstream genes [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 COPD-related genes. GO analysis encompassed biological process (BP), cellular component (CC), and molecular function (MF) categories. These analyses were conducted using the Bioinformatics platform (https://www.bioinformatics.com.cn/).

2.8 Construction of PPI network

The PPI network for 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 11 centrality algorithms: Betweenness, Closeness, Clustering Coefficient, Degree, DMNC, EPC, EcCentricity, MCC, MNC, Radiality, and Stress [25]. The top 30 genes from this aggregated ranking were selected. An UpSet plot was applied to pinpoint the genes consistently ranked among the top candidates by all 11 algorithms, thereby confirming their robust centrality.

2.9. DEGs screening

Gene expression data were retrieved from the GSE57148 dataset in the GEO database, comprising 91 controls and 98 COPD patients of East Asian descent [26]. After initial data processing, differential expression analysis was performed using the limma package. Genes were deemed significant if $|\log FC| > 0.25$ and $p < 0.05$, ensuring biological relevance and false positive control [27]. Subsequently, expression heatmaps were generated based on these genes.

Genes identified through CytoHubba screening were then intersected with differentially expressed genes from the bulk transcriptome, and the overlapping candidates were further refined through machine learning.

2.10. Machine learning

To identify key genes associated with COPD driven by oral microbiota, we applied Lasso regression with 1,000 iterations to high-dimensional gene expression data for feature selection, followed by 1,000 rounds of 10-fold cross-validation to ensure model stability. 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 2,000 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" function. The lung tissue module within the scMayoMap package was utilized to annotate subpopulations within each cluster [29].

Additionally, single-cell differential expression analysis was applied to annotated lung tissue cells, enabling a direct comparison of target gene expression between COPD patients and healthy controls.

2.12 Immune infiltration

The CIBERSORT algorithm was utilized to estimate the relative proportions of 22 immune cell types 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 DSigDB (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 (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, 48 oral microbial features demonstrated statistically significant associations with COPD. Of these, 24 taxa (15 from the dorsum tongue and 9 from saliva) exhibited protective effects (OR < 1), while 24 taxa (11 from the dorsum tongue and 13 from saliva) were identified as risk factors (OR > 1). These microorganisms were primarily concentrated within the genera *Fusobacterium*, *Neisseria*, *Pauljensenia*, *Prevotella*, and *Streptococcus* (Figure 2A). In the reverse MR analysis, 79 microbial taxa were identified as having potential reverse causal relationships with COPD. Among these, 37 features (25 from the dorsum tongue and 12 from saliva) showed protective effects (OR < 1), whereas 42 taxa (16 from the dorsum tongue and 26 from saliva) were classified as risk factors (OR > 1). These taxa were predominantly enriched in the genera *Campylobacter_A*, *Rothia*, and *Streptococcus* (Figure 2B). A Sankey diagram illustrating the distribution of microbial taxa at the order, family, and genus levels is presented in Figure 2C.

Furthermore, all five MR methods yielded consistent directional effects and statistical significance. The *F*-statistics for all selected SNPs exceeded 10, ruling out weak instrument bias. The MR-Egger regression intercept did not deviate significantly from zero ($p > 0.05$), and the MR-PRESSO global test detected no outliers, indicating no evidence of horizontal pleiotropy. Cochran's Q test revealed no substantial heterogeneity ($p > 0.05$). The Steiger directionality test ruled out reverse causation (Supplementary Table 2; Supplementary Table 3).

3.2. Genes and functionality

We performed functional annotation on genes mapped by SNPs associated with oral microbiota-related COPD to elucidate their potential biological functions (Supplementary Table 4). KEGG pathway enrichment analysis (Figure 3A) revealed

that the key genes were associated with cell adhesion molecules, bacterial invasion of epithelial cells, Th1 and Th2 cell differentiation, and focal adhesion. For BP (Figure 3B), these genes were primarily involved in cell-cell adhesion via plasma-membrane adhesion molecules and cell junction assembly. CC analysis (Figure 3C) indicated enrichment in ion channel complex, transporter complex, and lamellipodium. MF analysis (Figure 3D) showed enrichment in transmembrane transporter binding, gated channel activity, and cell-cell adhesion mediator activity (Supplementary Table 5).

3.3. Identification of PPI Network and Hub Genes

From the 344 shared driver genes identified through SNP mapping, following the exclusion of disconnected nodes, hub genes were ranked by 11 centrality algorithms implemented in CytoHubba. An UpSet plot subsequently highlighted 10 genes that ranked consistently among the top 30 across all 11 algorithms (Figure 4).

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

In the GSE57148 dataset, comprising COPD patients and healthy controls of East Asian descent, 2,814 DEGs were identified, including 1,326 up-regulated and 1,488 down-regulated genes (Figure 5). Among these, *CDH13* and *MPDZ* were identified as overlapping genes between the CytoHubba hub genes and the bulk RNA-seq DEGs.

3.5. Identification and Validation of Core Genes for COPD

To identify reliable biomarkers for COPD, 1,000 iterations of Lasso regression were applied to the two overlapping candidate genes, consistently selecting *MPDZ* as the most relevant feature. Model performance was then evaluated through 1,000 rounds of 10-fold cross-validation, yielding a mean AUC of 0.762 (95% CI: 0.758-0.766). *MPDZ* was significantly upregulated in COPD patients, indicating its robust diagnostic potential (Figure 5).

3.6. Single-cell analysis results

We analyzed scRNA-seq data from 8 samples, comprising lung cells from three healthy controls and five COPD patients. After stringent quality control, 13 distinct

cell clusters were identified and annotated using the lung tissue module of the scMayoMap package. Major cell types 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 6). Single-cell differential expression analysis revealed that *MPDZ* was significantly upregulated in Ciliated cells from COPD patients compared to healthy individuals (Supplementary Table 6).

3.7. Immuno-infiltration analysis by CIBERSORT

Cell type proportions were estimated using the CIBERSORT algorithm. Compared with normal samples, COPD samples showed significant differences in the proportions of four immune cell types: Eosinophils, Neutrophils, resting NK cells, and T follicular helper cells. Correlation analysis revealed that *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 7).

3.8. Molecular Docking Evaluation

Based on the above results, six active compounds were selected for molecular docking with *MPDZ*: 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. The optimal docking conformations are visualized in Figure 8. All binding energies were < -5 kcal/mol, indicating strong binding affinities.

4. Conclusion

Compared with the extensively studied gut microbiome, research on the oral microbiome in relation to COPD remains relatively limited. Conventional observational studies often suffer from technical and sampling variability, leading to heterogeneous findings that obscure the underlying pathogenesis and potential therapeutic avenues for

COPD. We 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 contributes to COPD pathogenesis and identifying key genes. By integrating bulk and single-cell transcriptomic data with machine learning, we identified potential therapeutic targets for COPD. The robustness of these findings was further validated via drug prediction and molecular docking analyses. Through this multi-omics framework, our research provides novel evidence for the causal role of oral microbiota in COPD and advances the understanding of the 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 Catalog and OpenGWAS project through METAL. The MR results identified significant microbial taxa primarily within the genera *Fusobacterium*, *Neisseria*, *Pauljensenia*, *Prevotella*, and *Streptococcus*, 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*, whether through host-mediated killing or autolysis, 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 components 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]. *Pauljensenia* has been associated with reduced risks of bronchitis and tonsillitis, though the underlying mechanisms remain unclear [41]. However, as a newly established genus, *Pauljensenia* has received limited attention in the context of 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 identified potential causal effects of COPD on 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 relative abundance increases 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 characterize the biological functions of candidate genes, we mapped MR-identified instrumental variables to their genomic loci and performed GO and KEGG enrichment analyses. KEGG analysis identified enrichment in cell adhesion molecules and focal adhesion, both central to airway epithelial-matrix interactions. FAK activation by the COPD-associated glycoprotein YKL-40 drives smooth muscle proliferation and fibrotic remodeling via the MAPK cascade [50, 51]. The enrichment of Th1 and Th2 cell differentiation captures the adaptive immune dysregulation in COPD, where NLRP3-mediated pyroptosis triggers Th1-skewed IFN-γ production, while Th2-derived IL-4 and IL-13 promote mucus hypersecretion and airway hyperresponsiveness

[52, 53]. Bacterial invasion of epithelial cells is directly relevant to the oral microbiota context of this study, given that oral-derived anaerobes reach the lower airways through microaspiration, where their cysteine proteases disrupt the protease-antiprotease balance and sustain chronic inflammation [54, 55]. Refining these findings, BP and CC terms localized these impairments to plasma-membrane adhesion complexes and ion channel machinery. In the COPD airway, cigarette smoke and microbial products compromise adherens and tight junction integrity by epigenetically silencing the *CDH1* enhancer and disrupting zona occludens-1 through aberrant phosphorylation, increasing paracellular permeability and impairing epithelial regeneration [56, 57]. CC enrichment in ion channel complex and transporter complex corresponds to acquired CFTR and BK channel dysfunction, where TGF-β1-mediated transcriptional suppression depletes airway surface liquid and creates a permissive niche for bacterial colonization [58, 59], while enrichment of lamellipodium reflects Rho GTPase-dependent cytoskeletal reorganization that facilitates inflammatory cell infiltration and fibroblast migration during airway remodeling [60]. MF analysis corroborated these observations, with transmembrane transporter binding and gated channel activity reinforcing the ion transport defect, and cell-cell adhesion mediator activity echoing the junctional impairment identified at the BP level. These multi-level enrichments indicate that oral microbiota-associated genetic variants may predispose to COPD through a cascade from epithelial barrier compromise and mucociliary dysfunction to dysregulated immune responses against microbial challenge.

Among the CytoHubba hub genes intersected with DEGs from the East Asian GSE57148 dataset, *CDH13* and *MPDZ* were identified as overlapping candidates, with *MPDZ* consistently selected by Lasso regression across 1,000 iterations. *CDH13* encodes T-cadherin, a GPI-anchored adiponectin receptor whose genetic variants modulate plasma adiponectin levels and COPD susceptibility in Chinese populations [61]. Genome-wide interaction studies have further shown that *CDH13* polymorphisms modify the effect of particulate matter on lung function decline in European and East Asian cohorts, positioning this gene at the intersection of genetic predisposition and

environmental exposure [62, 63]. *MPDZ* is a multi-PDZ domain scaffolding protein that stabilizes tight junction complexes by anchoring claudins and JAM proteins to intracellular signaling partners. Its upregulation in COPD, particularly in ciliated cells as revealed by single-cell analysis, likely reflects a compensatory response to chronic junctional disruption. In lung cancer, *MPDZ* is epigenetically silenced via promoter hypermethylation, and its deficiency promotes metastasis through Hippo-YAP pathway dysregulation [64]. The sustained epithelial injury in advanced COPD may progressively exhaust this compensatory capacity, converging with the *MPDZ* loss observed in malignancy and suggesting a shared vulnerability in tight junction integrity that underlies the epidemiological link between these two conditions.

Drug prediction analysis identified captopril, camptothecin, digitoxigenin, irinotecan, pregnenolone, and trichostatin A as candidate therapeutics targeting *MPDZ*. Captopril, an ACE inhibitor, has demonstrated anti-inflammatory effects in COPD models [65]. Camptothecin analogues exhibit anti-inflammatory properties in pulmonary diseases [66]. Digitoxigenin, through its modulation of ion channels, may alleviate COPD symptoms [67]. Irinotecan, a topoisomerase inhibitor, potentially suppresses fibrotic pathways [68]. Pregnenolone supports steroidogenesis and may counter hormonal imbalances in COPD [69]. Trichostatin A, an HDAC inhibitor, restores diminished HDAC activity in COPD and enhances anti-inflammatory responses [70]. Molecular docking confirmed strong binding affinities for these compounds, suggesting their therapeutic potential.

Nevertheless, several limitations should be acknowledged. The composition of the oral microbiome is influenced by multiple factors, including genetic background, dietary habits, lifestyle, and environmental exposures, all of which could affect the observed associations. Additionally, the MR analysis focused on East Asian populations, and the findings may not be directly generalizable to other ethnic groups. Further studies are warranted to validate and extend these results across diverse populations. The quality of scRNA-seq data is also highly dependent on technical factors such as cell capture efficiency and sequencing depth, which may lead to under-representation

of certain cell types. In the field of drug prediction, although DSigDB provides extensive gene-drug association data, these predictions require experimental validation to confirm efficacy and safety. Furthermore, drug effects observed at the cellular level may not directly translate to in vivo outcomes, and clinical trials are necessary to verify practical applicability.

In summary, this study provides genetic evidence for a causal link between oral microbiota and COPD through bidirectional MR analysis. SNPnexus was utilized to map COPD-associated genes from relevant SNPs, implicating these genes in mediating the influence of oral microbiota on COPD pathogenesis. Single-cell analysis of COPD lung tissues revealed cell-type-specific expression patterns, offering insights into their regulatory roles in COPD. Bulk RNA sequencing identified distinct expression profiles between COPD patients and controls, highlighting candidate biomarkers for disease assessment. Through drug prediction and molecular docking, *MPDZ* was identified as a potential therapeutic target with several predicted candidate drugs. These findings provide genetic evidence and candidate targets to advance personalized therapeutic strategies for COPD.

5. Conclusion

This study presents a comprehensive bioinformatics approach to evaluate the causal relationship between oral microbiota and COPD. Our analysis identified 48 oral microbial taxa with forward causal effects on COPD and 79 taxa demonstrating reverse causal relationships. Through the integration of 11 topological algorithms in CytoHubba, transcriptomic differential expression analysis, machine learning validation, single-cell profiling, immune infiltration assessment, and molecular docking, *MPDZ* was identified as a potential key regulator in COPD-related immune modulation. These findings offer candidate targets and a theoretical basis for future investigations into COPD pathogenesis and therapeutic development.

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. Bahetjan K, Yu-Xia, Lin S, Aili N, Yang H, Du S. Analysis of the bronchoalveolar lavage fluid microbial flora in COPD patients at different lung function during acute exacerbation. *Sci Rep*. 2025;15(1):13179. Published 2025 Apr 16. doi:10.1038/s41598-025-96746-5
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. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. Published 2015 Feb 25. doi:10.1186/s13742-015-0047-8
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. Wei ZF, Wuzhang JP, Huang YT, Li HY, Chen J. Assessing causal relationships between oral microbiota and asthma: evidence from two-sample mendelian randomization analysis. *J Asthma*. Published online March 28, 2026. doi:10.1080/02770903.2026.2647919
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. Higham A, Beech A, Singh D. The relevance of eosinophils in chronic obstructive pulmonary disease: inflammation, microbiome, and clinical outcomes. *J Leukoc Biol*. 2024;116(5):927-946. doi:10.1093/jleuko/qiae153
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. Liu L, Zhong S, Zhou T, et al. Dysregulated ITGA3/FAK/YAP axis mediates impaired alveolar type II epithelial cells function in COPD. *J Adv Res*. Published online November 19, 2025. doi:10.1016/j.jare.2025.11.033

51. Suzuki Y, Saito J, Kikuchi M, et al. Clinical utility of YKL-40 for understanding pathophysiology of obstructive airway disorders. *Allergol Int*. 2026;75(1):52-61. doi:10.1016/j.alit.2025.05.004
52. Chen G, Mu Q, Meng ZJ. Cigarette Smoking Contributes to Th1/Th2 Cell Dysfunction via the Cytokine Milieu in Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis*. 2023;18:2027-2038. Published 2023 Sep 12. doi:10.2147/COPD.S426215
53. Xiong XF, Zhu M, Wu HX, Fan LL, Cheng DY. Immunophenotype in acute exacerbation of chronic obstructive pulmonary disease: a cross-sectional study. *Respir Res*. 2022;23(1):137. Published 2022 May 28. doi:10.1186/s12931-022-02058-x
54. Wang L, Cai Y, Garssen J, Henricks PAJ, Folkerts G, Braber S. The Bidirectional Gut-Lung Axis in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2023;207(9):1145-1160. doi:10.1164/rccm.202206-1066TR
55. 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
56. 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
57. Yeung-Luk BH, Wally A, Swaby C, et al. Epigenetic Reprogramming Drives Epithelial Disruption in Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol*. 2024;70(3):165-177. doi:10.1165/rcmb.2023-0147OC
58. Miravitlles M, Criner GJ, Mall MA, et al. Potential systemic effects of acquired CFTR dysfunction in COPD. *Respir Med*. 2024;221:107499. doi:10.1016/j.rmed.2023.107499
59. Silswal N, Baumlin N, Haworth S, et al. Therapeutic strategies to reverse cigarette smoke-induced ion channel and mucociliary dysfunction in COPD airway

epithelial cells. *Am J Physiol Lung Cell Mol Physiol*. 2025;328(4):L571-L585. doi:10.1152/ajplung.00258.2024

60. Matera DL, Lee AT, Hiraki HL, Baker BM. The Role of Rho GTPases During Fibroblast Spreading, Migration, and Myofibroblast Differentiation in 3D Synthetic Fibrous Matrices. *Cell Mol Bioeng*. 2021;14(5):381-396. Published 2021 Sep 2. doi:10.1007/s12195-021-00698-5
61. Yuan YM, Zhang JL, Xu SC, et al. Genetic variants of CDH13 determine the susceptibility to chronic obstructive pulmonary disease in a Chinese population. *Acta Pharmacol Sin*. 2016;37(3):390-397. doi:10.1038/aps.2015.158
62. Imboden M, Kumar A, Curjuric I, et al. Modification of the association between PM10 and lung function decline by cadherin 13 polymorphisms in the SAPALDIA cohort: a genome-wide interaction analysis. *Environ Health Perspect*. 2015;123(1):72-79. doi:10.1289/ehp.1307398
63. Kim HJ, Min JY, Min KB, et al. CDH13 gene-by-PM10 interaction effect on lung function decline in Korean men. *Chemosphere*. 2017;168:583-589. doi:10.1016/j.chemosphere.2016.11.022
64. 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
65. Mungunsukh O, George J, McCart EA, et al. Captopril reduces lung inflammation and accelerated senescence in response to thoracic radiation in mice. *J Radiat Res*. 2021;62(2):236-248. doi:10.1093/jrr/rraa142
66. Czuwara-Ladykowska J, Makiela B, Smith EA, Trojanowska M, Rudnicka L. The inhibitory effects of camptothecin, a topoisomerase I inhibitor, on collagen synthesis in fibroblasts from patients with systemic sclerosis. *Arthritis Res*. 2001;3(5):311-318. doi:10.1186/ar321
67. Abud EM, Maylor J, Undem C, et al. Digoxin inhibits development of hypoxic pulmonary hypertension in mice. *Proc Natl Acad Sci U S A*. 2012;109(4):1239-1244. doi:10.1073/pnas.1120385109

68. Lapoirie J, Tran L, Piazza L, Contin-Bordes C, Truchetet ME, Bonnet F. Irinotecan and its metabolite SN38 inhibits procollagen I production of dermal fibroblasts from Systemic Sclerosis patients. *Sci Rep*. 2021;11(1):18011. Published 2021 Sep 9. doi:10.1038/s41598-021-97538-3
69. Ambhore NS, Kalidhindi RSR, Sathish V. Sex-Steroid Signaling in Lung Diseases and Inflammation. *Adv Exp Med Biol*. 2021;1303:243-273. doi:10.1007/978-3-030-63046-1_14
70. Zwinderman MRH, de Weerd S, Dekker FJ. Targeting HDAC Complexes in Asthma and COPD. *Epigenomes*. 2019;3(3):19. Published 2019 Sep 7. doi:10.3390/epigenomes3030019

Author contributions

Xiao-Jie An: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review and editing.

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

Hao-Yu Li: Data curation

Run-Ze Liu: Validation

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. Species-level oral microbiota causally associated with COPD identified by bidirectional MR analysis. (A) Dorsal tongue and salivary species with forward causal effects on COPD risk. (B) Dorsal tongue and salivary species with reverse causal effects influenced by COPD. (C) Sankey plot illustrating the taxonomic relationships among causally associated taxa at the order, family, and genus levels.

Figure 3

Figure 3. Functional enrichment analysis of differentially expressed genes associated with COPD. (A) KEGG pathway enrichment analysis. (B) GO biological process (BP) enrichment analysis. (C) GO cellular component (CC) enrichment analysis. (D) GO molecular function (MF) enrichment analysis.

Figure 4

Figure 4. PPI network analysis and hub gene identification. (A) PPI network constructed via Cytoscape with isolated nodes removed. (B) Upset plot showing the intersection of the top 30 genes ranked by 11 CytoHubba algorithms.

Figure 5

Figure 5. Identification and validation of the hub gene *MPDZ*. (A) Clustering heatmap of GSE57148. (B) Volcano plot of GSE57148. (C) Venn diagram showing the intersection of CytoHubba-derived genes, differentially expressed genes from GSE57148, and LASSO-selected genes from 1,000 iterations. (D) ROC curve of *MPDZ* from 1,000 rounds of 10-fold cross-validation. (E) Box plot of *MPDZ* expression between COPD patients and healthy controls.

Figure 6

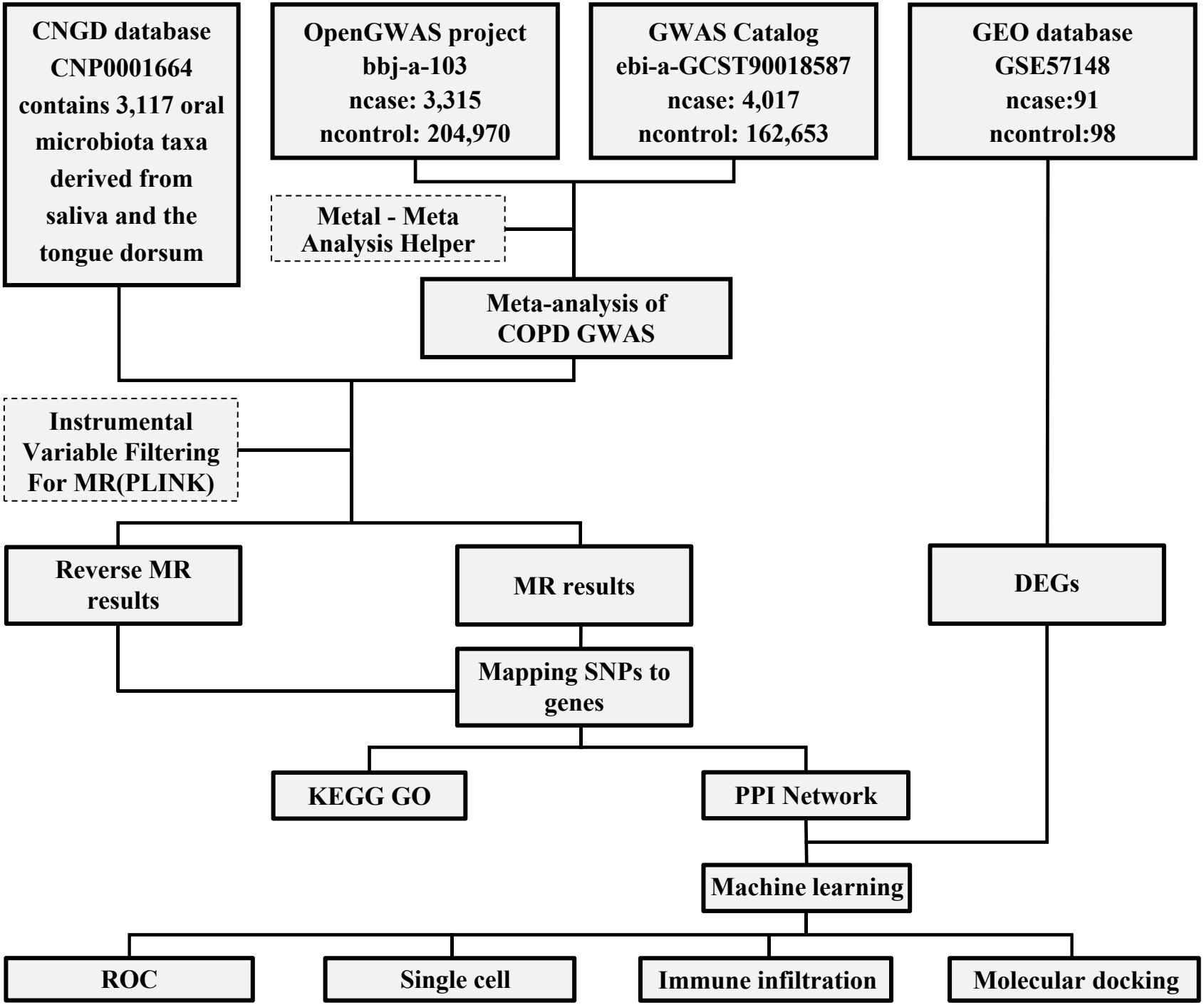
Figure 6. Single-cell RNA sequencing analysis and cell-type-specific localization of MPDZ. (A) Violin plots of quality control metrics including log-transformed feature count, UMI count, and mitochondrial gene percentage. (B) Stacked bar plot showing cell type composition across different samples. (C) t-SNE visualization of 13 cell clusters annotated by scMayoMap. (D) UMAP visualization of 13 cell clusters annotated by scMayoMap. (E) Annotation confidence scores assigned by scMayoMap for lung tissue cell types. (F) Expression distribution of *MPDZ* across cell clusters. (G) Cell-type-specific differential expression landscape, with each dot representing a gene plotted by log2 (FoldChange); *MPDZ* is highlighted in ciliated cells.

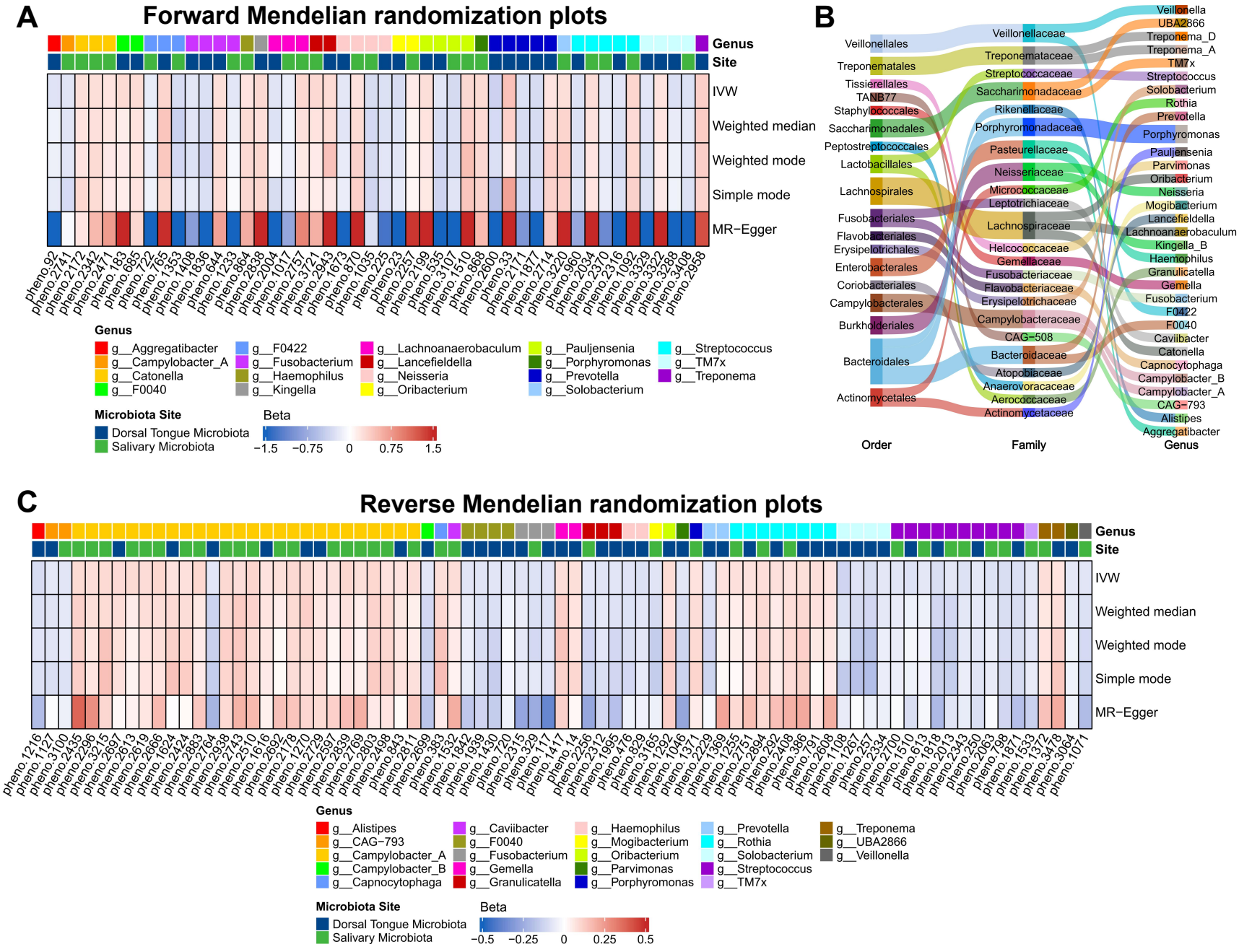
Figure 7

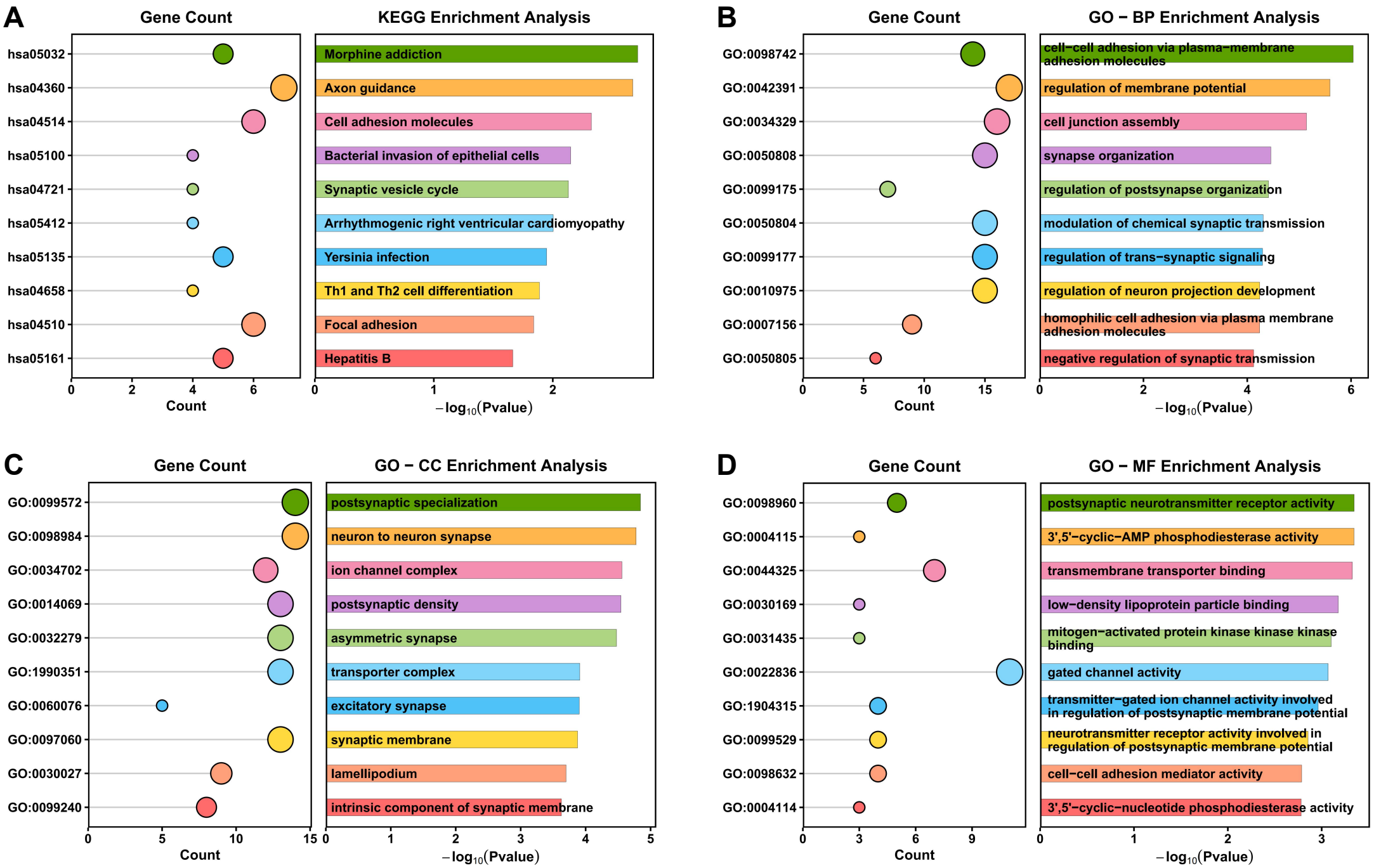
Figure 7. Immune infiltration analysis and correlation with *MPDZ* expression. (A) CIBERSORT-estimated immune cell proportions in healthy controls and COPD patients. (B) Correlation between *MPDZ* expression and 22 immune cell types. (C) Comparison of immune cell proportions between healthy controls and COPD patients. ns, not significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.005$. (D) Chromosomal localization of *MPDZ*.

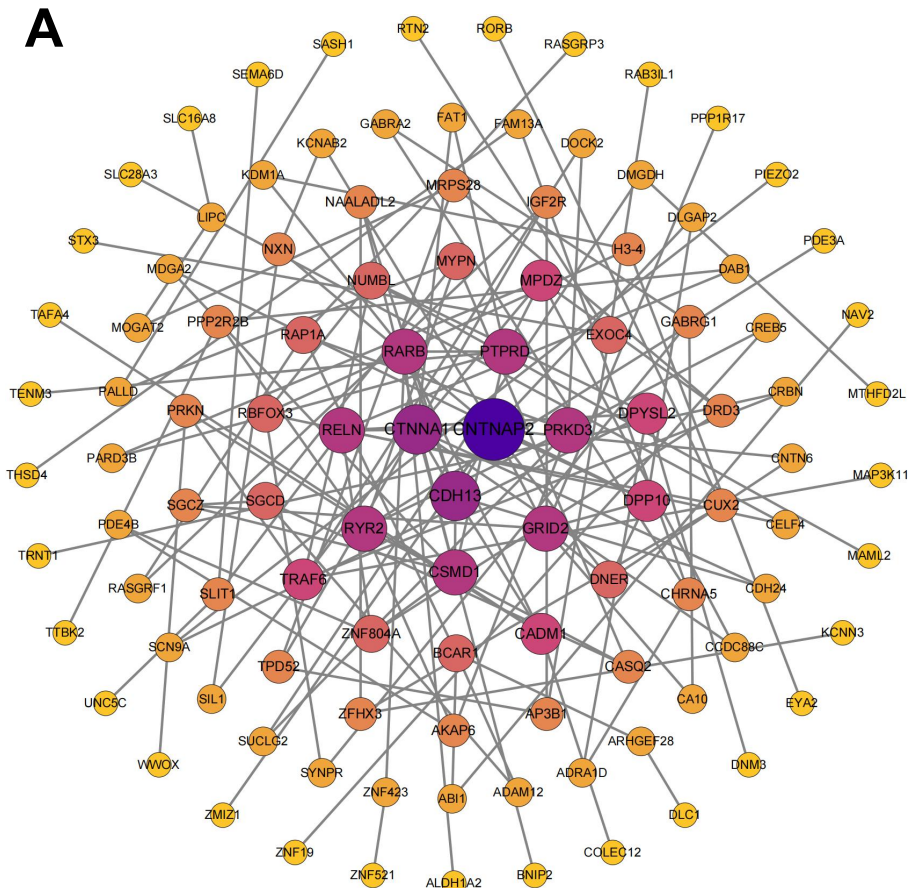
Figure 8

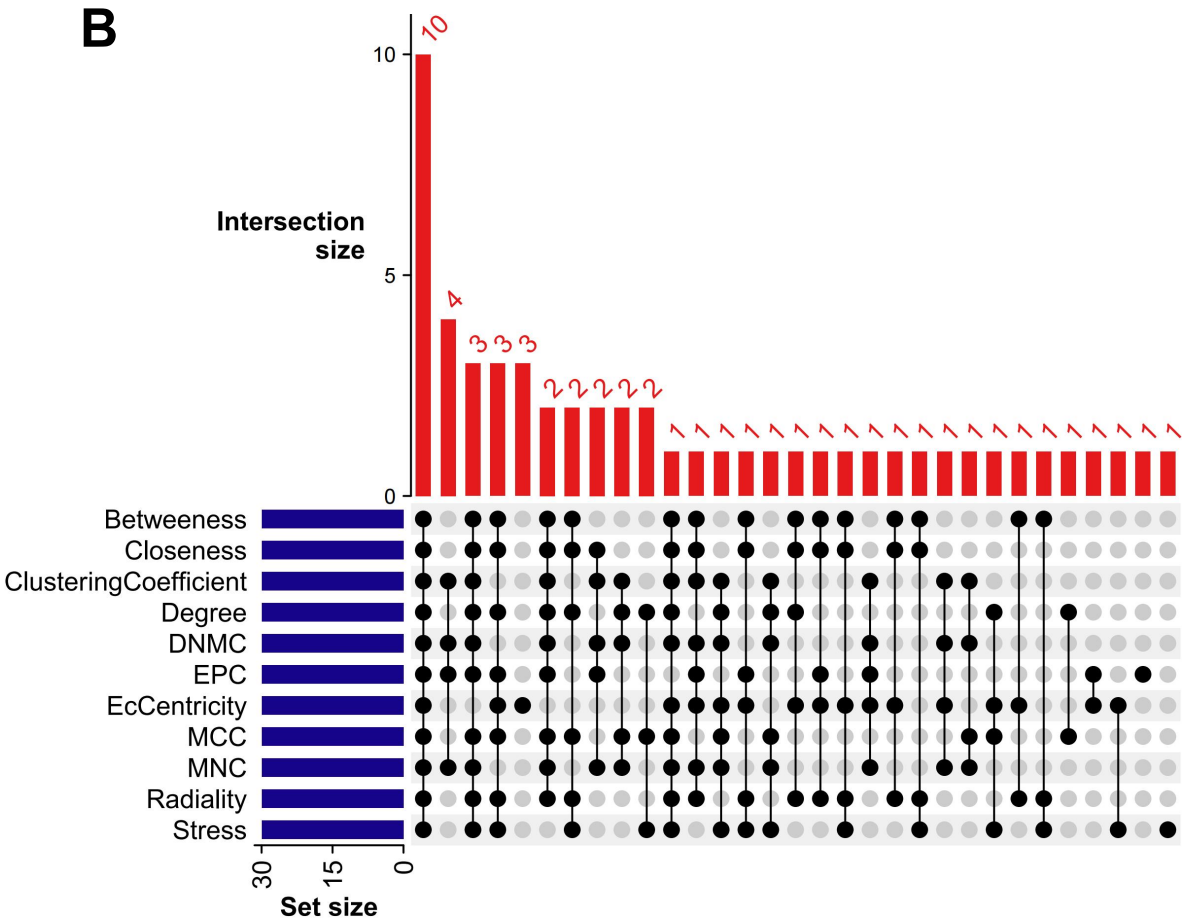
Figure 8. Molecular docking of compounds with *MPDZ*. (A) captopril, (B) Camptothecin, (C) digitoxigenin, (D) coumarin, (E) pregnenolone, and (F) trichostatin A.

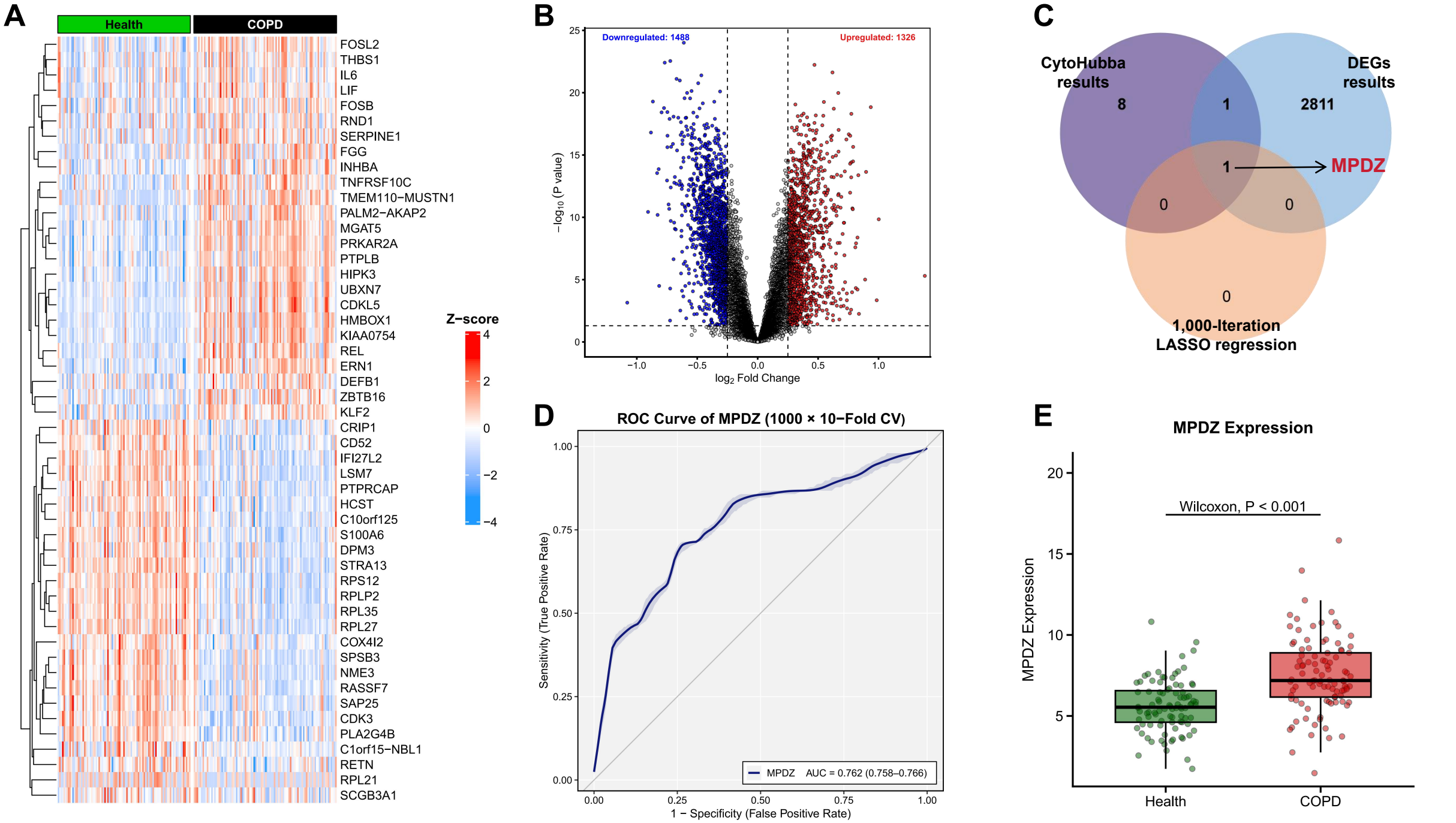

CNGD database
CNP0001664
contains 3,117 oral microbiota taxa derived from saliva and the tongue dorsum
OpenGWAS project
bbj-a-103
ncase: 3,315
ncontrol: 204,970
GWAS Catalog
ebi-a-GCST90018587
ncase: 4,017
ncontrol: 162,653
GEO database
GSE57148
ncase:91
ncontrol:98
Metal - Meta Analysis Helper
Meta-analysis of COPD GWAS
Instrumental Variable Filtering For MR(PLINK)
Reverse MR results
MR results
DEGs
Mapping SNPs to genes
KEGG GO
PPI Network
Machine learning
ROC
Single cell
Immune infiltration
Molecular docking

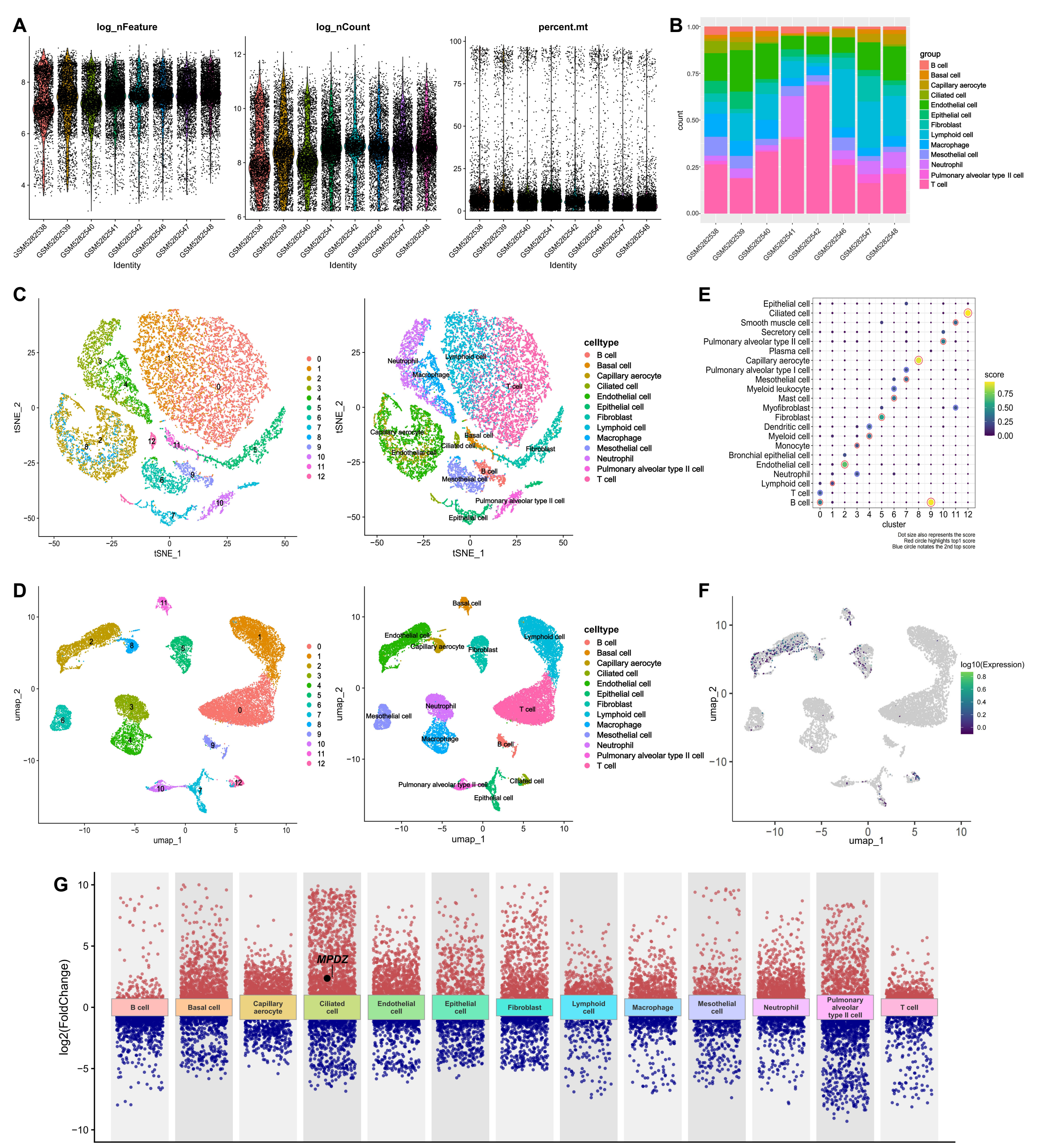

A
Forward Mendelian randomization plots
Genus
Site
IVW
Weighted median
Weighted mode
Simple mode
MR−Egger
Genus
g__Aggregatibacter
g__Campylobacter_A
g__Catonella
g__F0040
g__F0422
g__Fusobacterium
g__Haemophilus
g__Kingella
g__Lachnoanaerobaculum
g__Lancefieldella
g__Neisseria
g__Oribacterium
g__Pauljensenia
g__Porphyromonas
g__Prevotella
g__Solobacterium
g__Streptococcus
g__TM7x
g__Treponema
Microbiota Site
Dorsal Tongue Microbiota
Salivary Microbiota
Beta
−1.5
−0.75
0
0.75
1.5
B
Veillonellales
Treponematales
Tissierellales
TANB77
Staphylococcales
Saccharimonadales
Peptostreptococcales
Lactobacillales
Lachnospirales
Fusobacteriales
Flavobacteriales
Erysipelotrichales
Enterobacterales
Coriobacteriales
Campylobacterales
Burkholderiales
Bacteroidales
Actinomycetales
Veillonellaceae
Treponemataceae
Streptococcaceae
Saccharimonadaceae
Rikenellaceae
Porphyromonadaceae
Pasteurellaceae
Neisseriaceae
Micrococcaceae
Leptotrichiaceae
Lachnospiraceae
Helcococcaceae
Gemellaceae
Fusobacteriaceae
Flavobacteriaceae
Erysipelotrichaceae
Campylobacteraceae
CAG−508
Bacteroidaceae
Atopobiaceae
Anaerovoracaceae
Aerococcaceae
Actinomycetaceae
Veillonella
UBA2866
Treponema_D
Treponema_A
TM7x
Streptococcus
Solobacterium
Rothia
Prevotella
Porphyromonas
Pauljensenia
Parvimonas
Oribacterium
Neisseria
Mogibacterium
Lancefieldella
Lachnoanaerobaculum
Kingella_B
Haemophilus
Granulicatella
Gemella
Fusobacterium
F0422
F0040
Caviibacter
Catonella
Capnocytophaga
Campylobacter_B
Campylobacter_A
CAG−793
Alistipes
Aggregatibacter
Order
Family
Genus
C
Reverse Mendelian randomization plots
Genus
Site
IVW
Weighted median
Weighted mode
Simple mode
MR−Egger
Genus
g__Alistipes
g__CAG−793
g__Campylobacter_A
g__Campylobacter_B
g__Capnocytophaga
g__Caviibacter
g__F0040
g__Fusobacterium
g__Gemella
g__Granulicatella
g__Haemophilus
g__Mogibacterium
g__Oribacterium
g__Parvimonas
g__Porphyromonas
g__Prevotella
g__Rothia
g__Solobacterium
g__Streptococcus
g__TM7x
g__Treponema
g__UBA2866
g__Veillonella
Microbiota Site
Dorsal Tongue Microbiota
Salivary Microbiota
Beta
−0.5
−0.25
0
0.25
0.5

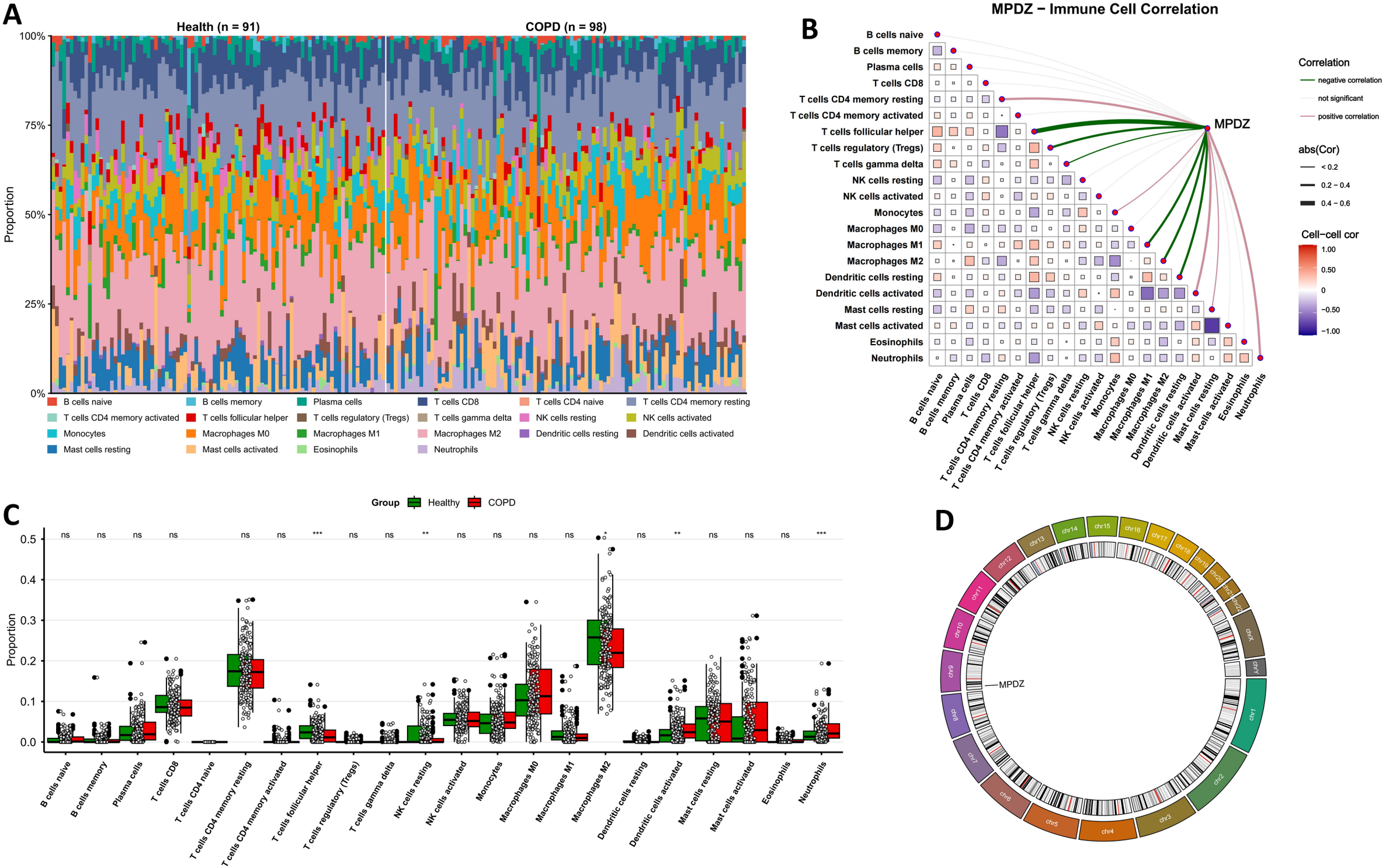

A
Gene Count
KEGG Enrichment Analysis
hsa05032
hsa04360
hsa04514
hsa05100
hsa04721
hsa05412
hsa05135
hsa04658
hsa04510
hsa05161
Morphine addiction
Axon guidance
Cell adhesion molecules
Bacterial invasion of epithelial cells
Synaptic vesicle cycle
Arrhythmogenic right ventricular cardiomyopathy
Yersinia infection
Th1 and Th2 cell differentiation
Focal adhesion
Hepatitis B
0 2 4 6
Count
0 1 2
−log10(Pvalue)
B
Gene Count
GO − BP Enrichment Analysis
GO:0098742
GO:0042391
GO:0034329
GO:0050808
GO:0099175
GO:0050804
GO:0099177
GO:0010975
GO:0007156
GO:0050805
cell−cell adhesion via plasma−membrane adhesion molecules
regulation of membrane potential
cell junction assembly
synapse organization
regulation of postsynapse organization
modulation of chemical synaptic transmission
regulation of trans−synaptic signaling
regulation of neuron projection development
homophilic cell adhesion via plasma membrane adhesion molecules
negative regulation of synaptic transmission
0 5 10 15
Count
0 2 4 6
−log10(Pvalue)
C
Gene Count
GO − CC Enrichment Analysis
GO:0099572
GO:0098984
GO:0034702
GO:0014069
GO:0032279
GO:1990351
GO:0060076
GO:0097060
GO:0030027
GO:0099240
postsynaptic specialization
neuron to neuron synapse
ion channel complex
postsynaptic density
asymmetric synapse
transporter complex
excitatory synapse
synaptic membrane
lamellipodium
intrinsic component of synaptic membrane
0 5 10 15
Count
0 1 2 3 4 5
−log10(Pvalue)
D
Gene Count
GO − MF Enrichment Analysis
GO:0098960
GO:0004115
GO:0044325
GO:0030169
GO:0031435
GO:0022836
GO:1904315
GO:0099529
GO:0098632
GO:0004114
postsynaptic neurotransmitter receptor activity
3',5'−cyclic−AMP phosphodiesterase activity
transmembrane transporter binding
low−density lipoprotein particle binding
mitogen−activated protein kinase kinase kinase binding
gated channel activity
transmitter−gated ion channel activity involved in regulation of postsynaptic membrane potential
neurotransmitter receptor activity involved in regulation of postsynaptic membrane potential
cell−cell adhesion mediator activity
3',5'−cyclic−nucleotide phosphodiesterase activity
0 3 6 9
Count
0 1 2 3
−log10(Pvalue)

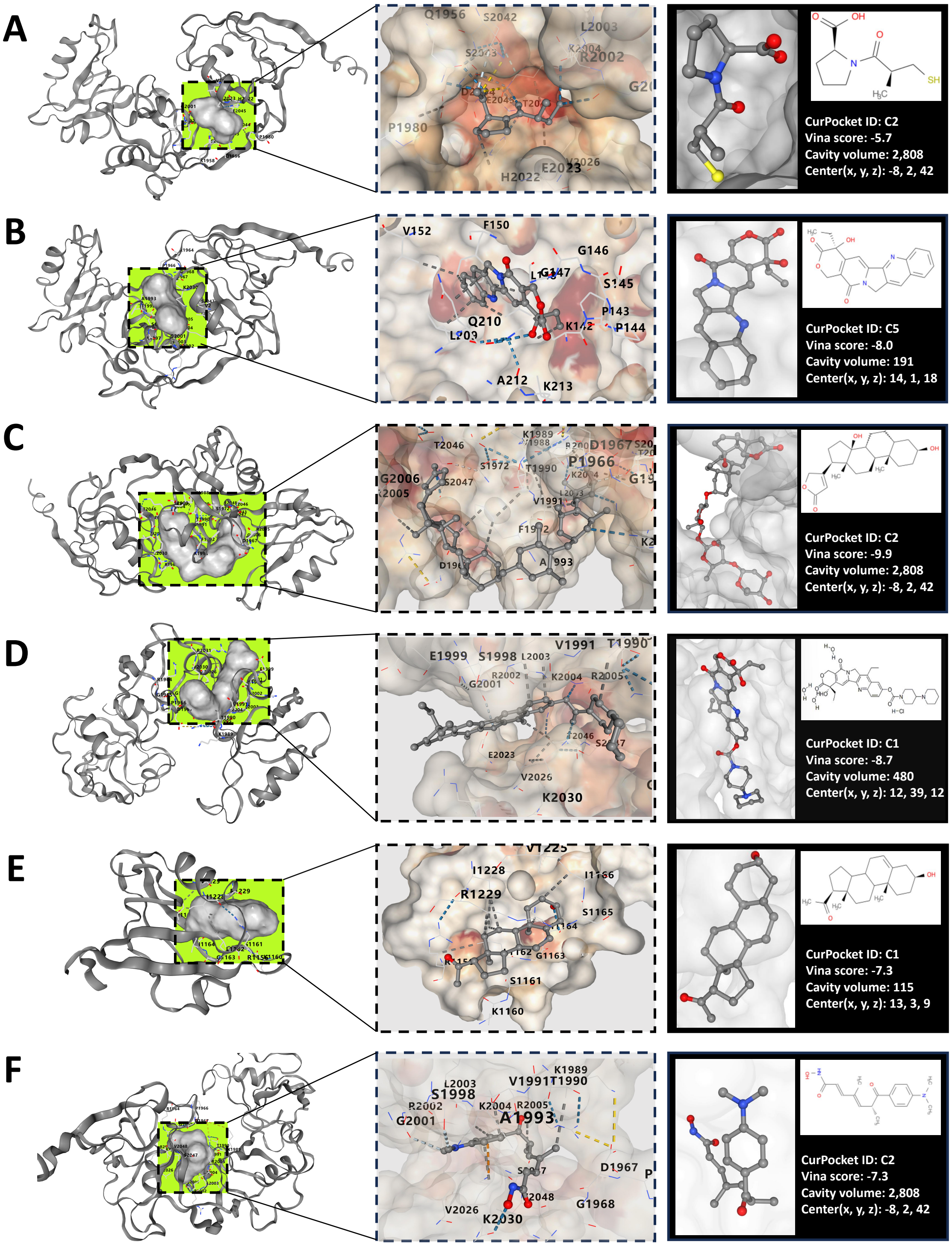

A
SASH1
RTN2
RORB
RASGRP3
SEMA6D
RAB3IL1
SLC16A8
GABRA2
FAT1
FAM13A
PPP1R17
KCNAB2
DOCK2
SLC28A3
KDM1A
MRPS28
DMGDH
PIEZO2
NAALADL2
IGF2R
LIPC
DLGAP2
STX3
NXN
H3-4
PDE3A
MYPN
MDGA2
NUMBL
MPDZ
DAB1
TAFA4
MOGAT2
PPP2R2B
RAP1A
GABRG1
NAV2
EXOC4
CREB5
RARB
PTPRD
TENM3
PALLD
CRBN
MTHFD2L
PRKN
RBFOX3
DPYSL2
ORD3
RELN
CTNNA1
CNTNAP2
PRKD3
THSD4
PARD3B
GNTN6
MAP3K11
CDH13
SGCZ
SGCD
DPP10
CUX2
PDE4B
RYR2
GRID2
CELF4
TRNT1
MAML2
CSMD1
RASGRF1
SLIT1
TRAF6
DNER
CHRNA5
CDH24
TTBK2
ZNF804A
CADM1
KCNN3
SCN9A
BCAR1
CGDC88C
TPD52
CASQ2
UNC5C
SIL1
CA10
EYA2
ZFHX3
AP3B1
AKAP6
SUCLG2
ARHGEF28
WWOX
SYNPR
ADRA1D
DNM3
ZNF423
ABI1
ADAM12
ZMIZ1
DLC1
ZNF19
COLEC12
ZNF521
ALDH1A2
BNIP2


B
Intersection size
10
5
0
10
4
3 3 3
2 2 2 2 2
Betweeness
Closeness
ClusteringCoefficient
Degree
DNMC
EPC
EcCentricity
MCC
MNC
Radiality
Stress
30
15
0
Set size


A
Health
COPD
FOSL2
THBS1
IL6
LIF
FOSB
RND1
SERPINE1
FGG
INHBA
TNFRSF10C
TMEM110−MUSTN1
PALM2−AKAP2
MGAT5
PRKAR2A
PTPLB
HIPK3
UBXN7
CDKL5
HMBOX1
KIAA0754
REL
ERN1
DEFB1
ZBTB16
KLF2
CRIP1
CD52
IFI27L2
LSM7
PTPRCAP
HCST
C10orf125
S100A6
DPM3
STRA13
RPS12
RPLP2
RPL35
RPL27
COX4I2
SPSB3
NME3
RASSF7
SAP25
CDK3
PLA2G4B
C1orf15−NBL1
RETN
RPL21
SCGB3A1
Z−score
4
2
0
−2
−4
B
Downregulated: 1488
Upregulated: 1326
−log10 (P value)
log2 Fold Change
C
CytoHubba results
DEGs results
8
1
2811
1
MPDZ
0
0
0
1,000-Iteration LASSO regression
D
ROC Curve of MPDZ (1000 × 10−Fold CV)
Sensitivity (True Positive Rate)
1 − Specificity (False Positive Rate)
MPDZ AUC = 0.762 (0.758–0.766)
E
MPDZ Expression
Wilcoxon, P < 0.001
MPDZ Expression
Health
COPD


A
log_nFeature
log_nCount
percent.mt
Identity
GSM5282538
GSM5282539
GSM5282540
GSM5282541
GSM5282542
GSM5282546
GSM5282547
GSM5282548
B
count
group
B cell
Basal cell
Capillary aerocyte
Ciliated cell
Endothelial cell
Epithelial cell
Fibroblast
Lymphoid cell
Macrophage
Mesothelial cell
Neutrophil
Pulmonary alveolar type II cell
T cell
C
tSNE_1
tSNE_2
celltype
D
umap_1
umap_2
E
Epithelial cell
Ciliated cell
Smooth muscle cell
Secretory cell
Pulmonary alveolar type II cell
Plasma cell
Capillary aerocyte
Pulmonary alveolar type I cell
Mesothelial cell
Myeloid leukocyte
Mast cell
Myofibroblast
Fibroblast
Dendritic cell
Myeloid cell
Monocyte
Bronchial epithelial cell
Endothelial cell
Neutrophil
Lymphoid cell
T cell
B cell
cluster
score
Dot size also represents the score
Red circle highlights top1 score
Blue circle notates the 2nd top score
F
log10(Expression)
G
log2(FoldChange)
MPDZ
B cell
Basal cell
Capillary aerocyte
Ciliated cell
Endothelial cell
Epithelial cell
Fibroblast
Lymphoid cell
Macrophage
Mesothelial cell
Neutrophil
Pulmonary alveolar type II cell
T cell


A
Health (n = 91)
COPD (n = 98)
Proportion
100%
75%
50%
25%
0%
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
B
MPDZ − Immune Cell Correlation
MPDZ
Correlation
negative correlation
not significant
positive correlation
abs(Cor)
< 0.2
0.2 − 0.4
0.4 − 0.6
Cell−cell cor
1.00
0.50
0
−0.50
−1.00
C
Group
Healthy
COPD
Proportion
ns
*
**
***
D
MPDZ


A
Q1956
S2042
L2003
K2004
R2002
S2043
D2024
E2045
T2026
P1980
E2023
V2026
H2022
CurPocket ID: C2
Vina score: -5.7
Cavity volume: 2,808
Center(x, y, z): -8, 2, 42
B
V152
F150
G146
G147
S145
P143
P144
Q210
K142
L209
A212
K213
CurPocket ID: C5
Vina score: -8.0
Cavity volume: 191
Center(x, y, z): 14, 1, 18
C
T2046
K1989
D1967
P1966
S1972
T1990
G2006
S2047
V1991
F1992
A1993
CurPocket ID: C2
Vina score: -9.9
Cavity volume: 2,808
Center(x, y, z): -8, 2, 42
D
E1999
S1998
L2003
V1991
T1990
R2002
G2001
K2004
R2005
T2046
E2023
V2026
K2030
CurPocket ID: C1
Vina score: -8.7
Cavity volume: 480
Center(x, y, z): 12, 39, 12
E
I1228
R1229
I1166
S1165
G1163
S1161
K1160
CurPocket ID: C1
Vina score: -7.3
Cavity volume: 115
Center(x, y, z): 13, 3, 9
F
K1989
L2003
V1991T1990
S1998
R2002
K2004
R2005
G2001
A1993
D1967
V2026
K2030
G1968
CurPocket ID: C2
Vina score: -7.3
Cavity volume: 2,808
Center(x, y, z): -8, 2, 42