frontiers in Immunology

TYPE Original Research
PUBLISHED 06 November 2026
DOI 10.3389/fimmu.2026.1965629

Check for updates

OPEN ACCESS

EDITED BY
Yoonsung Lee,
Kyung Hee University, Republic of Korea

REVIEWED BY
Ren Li,
Shanxi Medical University, China
Weilai Zhou,
Peking University First Hospital, China

*CORRESPONDENCE
Zhanbing Ma
20130108@nxmu.edu.cn
Juan Chen
chenjuan7419@163.com

†These authors have contributed equally to this work

RECEIVED 12 August 2026
REVISED 12 September 2026
ACCEPTED 13 October 2026
PUBLISHED 06 November 2026

CITATION
Hu Z, Yang Y, Yu J, Nian M, Cui J, Ran P, Cao X, Xin H, Wang X, Yang G, Gui Z, Ma Z and Chen J (2026) Integrative multi-omics, machine learning, computational toxicology, and experimental validation reveal PFAS-associated immune–inflammatory signatures in chronic obstructive pulmonary disease. *Front. Immunol.* 17:1965629. doi: 10.3389/fimmu.2026.1965629

COPYRIGHT
© 2026 Hu, Yang, Yu, Nian, Cui, Ran, Cao, Xin, Wang, Yang, Gui, Ma and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Integrative multi-omics, machine learning, computational toxicology, and experimental validation reveal PFAS-associated immune–inflammatory signatures in chronic obstructive pulmonary disease

Zhengyu Hu[1,2†], Yuanlan Yang[3†], Jialin Yu[1], Miaomiao Nian[1,2], Jieda Cui[4,5], Pixin Ran[4,5], Xia Cao[1,2], Hongxia Xin[1,2], Xiuyan Wang[1,2], Guilan Yang[1,2], Zhenzhen Gui[1,2], Zhanbing Ma[6*] and Juan Chen[1,2*]

[1]Ningxia Key Laboratory of Clinical and Pathogenic Microorganisms, Institute of Medical Science, General Hospital of Ningxia Medical University, Yinchuan, China, [2]Department of Respiratory and Critical Care Medicine, General Hospital of Ningxia Medical University, Yinchuan, China, [3]Guangzhou Medical University–Guangzhou Institutes of Biomedicine and Health Joint School of Life Sciences, Guangzhou Medical University, Guangzhou, Guangdong, China, [4]State Key Laboratory of Respiratory Diseases, National Clinical Research Center for Respiratory Diseases, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [5]Guangzhou National Laboratory, Guangzhou, China, [6]Department of Medical Genetics and Cell Biology, College of Basic Medicine, Ningxia Medical University, Yinchuan, China

Background: Per- and polyfluoroalkyl substances (PFAS) are persistent environmental contaminants increasingly implicated in respiratory disorders. However, the systems-level connections among PFAS-associated molecular perturbations, immune dysregulation, and chronic obstructive pulmonary disease (COPD) remain poorly understood. This study aimed to prioritize candidate links between PFAS-associated target information and COPD-related transcriptional and inflammatory signatures.

Methods: PFAS-associated target information was integrated with bulk transcriptomic differential expression and weighted gene co-expression network analysis to identify overlapping candidate genes. Protein–protein interaction analysis and 113 machine-learning model combinations were used for candidate prioritization and cross-cohort classification evaluation. CIBERSORT and xCell provided computational immune-cell signatures, with Benjamini–Hochberg correction applied to group comparisons and gene–signature correlations; single-cell transcriptomics provided cell-type expression context. Candidate expression was examined in cigarette-smoke-extract-treated cells, clinical samples, and public human and murine datasets. Molecular docking and molecular dynamics simulations explored candidate PFAS–protein interactions.

Results: HMOX1, JUNB, CXCL12, and PTGS2 were prioritized as candidate genes overlapping PFAS-associated target information and COPD-related expression patterns. Enrichment and immune-deconvolution analyses suggested associations with inflammatory and immune-related signatures, while single-cell analysis described their distribution across annotated cell types. Exploratory plasma ELISA measurements supported higher levels of the four candidate proteins in participants with COPD, with additional expression support varying

by gene and experimental model. Docking and molecular dynamics generated structural hypotheses for PFOA/PFOS–protein interactions, including HMOX1.
Conclusions: This integrative systems-immunology framework delineates immune–inflammatory signatures at the PFAS–COPD interface and identifies HMOX1, JUNB, CXCL12, and PTGS2 as convergent molecular nodes. The findings do not establish exposure causality but provide a multi-scale basis for future epidemiological and mechanistic studies of how persistent environmental pollutants may contribute to COPD-associated immune and inflammatory alterations.

KEYWORDS
chronic obstructive pulmonary disease, inflammation, multi-omics, per- and polyfluoroalkyl substances, single-cell RNA sequencing, systems immunology

1 Introduction

Per- and polyfluoroalkyl substances (PFAS) are used in diverse applications, such as coatings, waterproof fabrics, food packaging, and fire-fighting foams (1, 2). Their highly stable chemical structures make them resistant to environmental degradation, earning them the nickname "forever chemicals" (3, 4). These compounds can enter the human biological system through industrial emissions, environmental transport, and bioaccumulation, and then persist in the bloodstream for extended durations (5, 6). Epidemiological and toxicological studies have linked PFAS exposure to health outcomes, including liver disease, immunotoxicity, cancer, endocrine disruption, and lung injury, raising global concerns about their potential hazards (7, 8).

Chronic obstructive pulmonary disease (COPD) is a common chronic respiratory disorder characterized by persistent airflow limitation and a progressive decline in lung function, significantly impacting patients' quality of life and survival prognosis (9). According to the Global Burden of Disease Study 2019, approximately 212 million people worldwide had COPD in 2019, and COPD accounted for approximately 3.3 million deaths that year (10). In addition to tobacco smoking (the main risk factor), COPD development is heavily influenced by environmental factors such as biomass smoke, occupational dust, outdoor air pollution, and chemical gases (11, 12). Recent studies have increasingly focused on the impact of environmental pollutants on COPD development, especially emerging contaminants such as PFAS (13).

Current epidemiological evidence suggests that chronic PFAS exposure may be closely linked to the onset and progression of COPD. Data from the U.S. National Health and Nutrition Examination Survey (NHANES) provide exposure-level evidence linking internal serum PFAS concentrations with COPD. In a NHANES 2007–2018 analysis including 4, 541 participants with available PFAS, COPD, and covariate data, serum PFHxS, PFNA, PFOA, and PFOS were analyzed because of their high detection rates. Owing to the skewed distribution of serum PFAS concentrations, natural logarithm-transformed PFAS levels were used as internal exposure indicators. Individuals with COPD showed higher serum levels of PFHxS, PFNA, PFOA, and PFOS than those without COPD. Increased Ln-PFOA and Ln-PFNA were associated with higher COPD risk (14), with some sex-stratified analyses showing stronger associations among males (15). COPD patients present substantially higher internal levels of perfluorohexanesulfonic acid (PFHxS), PFNA, PFOA, and perfluorooctanesulfonic acid (PFOS) than healthy controls do (16). In pediatric and adolescent populations, elevated PFAS exposure has also been associated with impaired pulmonary function. For example, studies in children and adolescents reported associations between serum PFAS concentrations and reductions in lung-function parameters, including FEV_1 and FVC (17, 18). Experimental studies further demonstrated that PFAS induce abnormal DNA methylation and histone modifications in pulmonary epithelial cells, resulting in cell cycle arrest and apoptosis (19, 20). PFOS, PFOA, and related compounds notably impair pulmonary surfactant function, lower alveolar surface tension, and induce inflammatory cytokine release (21). In a mouse study, 14 weeks of PFOA exposure through drinking water altered the expression of inflammation- and immune-related genes in lung tissue (21). Although these findings support the hypothesis that PFAS may promote COPD through multiple pathways—including epigenetic alterations, surfactant dysfunction, and immune-inflammatory remodeling—the specific molecular mechanisms are still not fully understood.

In recent years, machine-learning algorithms have been increasingly applied to high-dimensional biological datasets for disease classification, biomarker discovery, and target prioritization. For example, machine-learning models have been used to classify idiopathic recurrent spontaneous miscarriage using fused FTIR and Raman spectroscopic signatures (22), to construct a random forest diagnostic model for acute myocardial infarction based on ferroptosis-related genes in circulating endothelial cells (23), and to identify shared gene signatures and candidate therapeutic targets in idiopathic pulmonary fibrosis-associated lung cancer using transcriptomic datasets, supervised learning, and molecular modeling (24). These studies support the utility of machine-learning strategies for extracting biologically informative signatures from complex biomedical data.

Building on existing COPD molecular profiling and PFAS toxicogenomic research, this study examined which COPD-associated genes overlap with predicted or database-supported PFOA/PFOS target sets and the cellular contexts in which these candidates are expressed. We integrated differential expression and co-expression analysis with machine-learning prioritization, bulk expression comparisons, inferred immune signatures, single-cell expression patterns within COPD samples, and COPD-related experimental observations. Docking and molecular dynamics provided additional

structural hypotheses. The intended contribution is the COPD-focused prioritization and cross-scale characterization of PFAS-associated candidates, with established gene functions providing biological context for subsequent exposure-based investigation. A detailed diagram of the study design is shown in Figure 1.

2 Materials and methods

2.1 Acquisition of PFAS targets

This study focused on identifying and integrating potential molecular targets for two representative PFAS: PFOA and PFOS. Initially, molecular structures and Simplified Molecular-Input Line-Entry System (SMILES) strings for PFOA and PFOS were retrieved from the PubChem database (25). The SwissTargetPrediction online platform was subsequently utilized to predict putative targets for these compounds, yielding preliminary candidate proteins (26). These predictions were supplemented with chemical–gene information from the Comparative Toxicogenomics Database (CTD) and relevant peer-reviewed literature (27, 28). The retrieved entries were standardized to gene symbols and protein identifiers using UniProt, and duplicates were removed. The resulting set was designated the PFAS-associated candidate-gene set. This designation includes predicted targets and reported chemical–gene associations, which do not necessarily represent direct physical binding or PFAS-dependent regulation in lung tissue.

2.2 Identification of COPD-related gene sets and overlapping candidate genes

The primary COPD lung-tissue transcriptomic dataset GSE151052 was obtained from the Gene Expression Omnibus (GEO). According

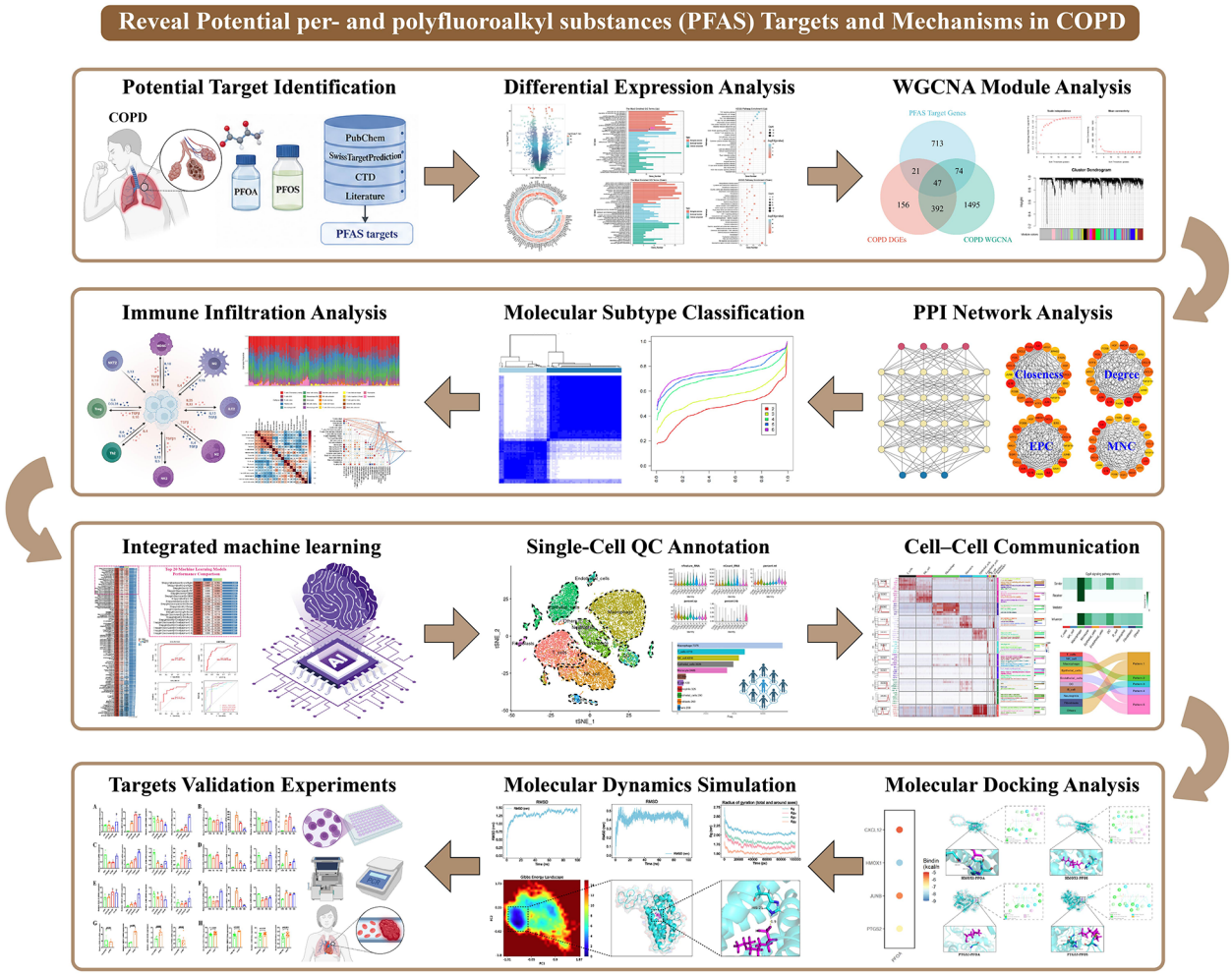

FIGURE 1
Study design overview. The schematic in Figure 1 outlines the multi-omics and computational toxicology approach employed in this study to investigate potential molecular associations between PFAS-related target profiles and COPD-related pathological alterations. The design integrates transcriptomic differential expression analysis, WGCNA, scRNA-seq, feature selection via 113 machine-learning models, immune infiltration profiling, molecular docking, and molecular dynamics simulations. These methodologies are augmented by *in vitro* experiments and the analysis of clinical samples from patients with chronic obstructive pulmonary disease to substantiate the principal findings. The study prioritizes candidate molecular links between PFAS-associated target profiles and COPD-related alterations and explores regional molecular expression patterns within COPD lung tissue. COPD-related expression validation and computational PFAS–protein interaction predictions are presented as distinct forms of evidence. Created in BioRender. Liu, X. (2026) https://BioRender.com/87ttgq4.

to the original GEO submitter record, this dataset was generated from systematic uniform random sampling of lung tissue regions from 10 explanted lungs of patients with very severe COPD, defined as Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage 4 COPD and treated by lung transplantation, and from 5 unused donor control lungs. After quality control in the original study, 117 regional lung tissue samples were retained, including 77 COPD samples and 40 control samples. The analytical observations were regional lung-tissue profiles nested within 15 lungs. In the original workflow, regional profiles were analyzed without explicitly modeling within-donor dependence. Thus, differential-expression statistics and WGCNA module–phenotype associations may be influenced by correlations among regions from the same lung and should be interpreted as exploratory tissue-level findings rather than independent participant-level inference. Differential-expression analysis between COPD and control lung tissue samples in GSE151052 was performed using limma through the tinyarray package, following between-array quantile normalization and probe-to-gene annotation. Gene-wise linear models and empirical Bayes moderation were applied, with differentially expressed genes defined by Benjamini–Hochberg-adjusted P values <0.05 and |log2FC| >0.58. Regional samples were analyzed without accounting for within-donor dependence; therefore, the findings were interpreted as exploratory regional-tissue associations. The data-processing and analysis code is publicly available on GitHub (https://github.com/HTX2023/GSE151052-Microarray-Data-Processing-Code-and-Guide) (29, 30). Functional enrichment analysis, incorporating Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (31), was conducted utilizing ClusterProfiler. Weighted gene coexpression network analysis (WGCNA) was employed to discern COPD-associated modules, with significance determined at $P < 0.05$; these modules were subsequently selected for further investigation (32). A Venn diagram analysis was then performed to intersect the DEG set, the COPD-related module gene set, and the predicted PFAS target set, with the aim of identifying genes common to all three sets. These overlapping genes were considered PFAS-associated candidate targets potentially relevant to COPD, based on their differential expression in COPD lung tissue, membership in COPD-associated co-expression modules, and overlap with predicted PFAS targets.

2.3 Construction of a PFAS-COPD protein interaction network and functional enrichment analysis

Using the overlapping candidate genes, a protein–protein interaction (PPI) network was built with the STRING database (version 11.5) (33), including interactions with a confidence score above 0.4. The network was visualized with Cytoscape (version 3.9.1) (34). Key hub nodes were identified using the cytoHubba plugin by combining multiple topological metrics: degree, maximum neighborhood component (MNC), edge percolated component (EPC), and closeness centrality (35). Proteins that scored highly across several of these algorithms were chosen as hub genes. These hub genes were subjected to GO and KEGG enrichment analyses using clusterProfiler (R package version 4.12.1) to characterize the biological processes and pathways represented in the candidate-gene set (36).

2.4 Prioritization of candidate feature genes using integrated machine learning

To identify COPD-associated feature genes among the PFAS-associated candidates, we developed an integrated machine-learning framework using GSE151052 as the training dataset and GSE76925 and GSE38974 as external validation datasets. Candidate hub-gene expression matrices were extracted from each dataset, and genes with missing values, duplicated symbols, or unavailable expression measurements were excluded. Genes represented in all three datasets were retained to ensure compatible model inputs. The binary outcome was coded as 1 for COPD and 0 for controls. Feature selection, model fitting, internal hyperparameter tuning, and final model selection were conducted using the training dataset; validation-cohort outcomes were reserved for performance evaluation.

To reduce cohort-specific scale differences, the training matrix was standardized globally using z-score transformation, whereas the validation matrices were standardized separately within each cohort. A two-stage machine-learning strategy was then applied. In the first stage, five feature-selection strategies were used to derive candidate feature subsets: LASSO regression, generalized linear model boosting (glmBoost), random forest, bidirectional stepwise logistic regression, and backward stepwise logistic regression. In addition, a "simple" strategy retaining all candidate genes was included as a non-filtered reference strategy. In the second stage, the selected feature subsets were combined with multiple classification algorithms, including elastic net, LASSO, ridge regression, stepwise logistic regression, support vector machine (SVM), linear discriminant analysis (LDA), glmBoost, partial least-squares generalized linear model (plsRglm), random forest, gradient boosting machine (GBM), XGBoost, and Naive Bayes (37, 38). These combinations generated 113 candidate machine-learning models.

For reproducibility, the main hyperparameters were predefined or selected by internal cross-validation. LASSO, ridge regression, and elastic net were fitted using binomial logistic regression with 10-fold cross-validation to select the optimal λ. Stepwise logistic regression was performed using AIC-based selection with $k = 2$ and either bidirectional or backward elimination. Random forest was implemented with 1, 000 trees, nodesize = 1, and mtry set to a maximum of 5 or the number of available variables, whichever was smaller. GBM used Bernoulli loss, interaction depth = 3, shrinkage = 0.001, and cross-validation to select the optimal number of trees. XGBoost used a binary logistic objective, max_depth = 2, eta = 1, nthread = 2, and 5-fold cross-validation to determine the optimal number of boosting rounds. PLS-GLM used logistic PLS regression, with the number of components limited to the smaller of 20 or the number of available features, and 10-fold cross-validation was applied when at least three features were available. SVM was implemented with probability estimation enabled to allow ROC and AUC analysis. Models retaining one or fewer variables were excluded from downstream evaluation.

Each trained model was applied to the training dataset and separately to GSE76925 and GSE38974 to generate continuous prediction scores. Discrimination was assessed using receiver operating characteristic curves and cohort-specific areas under the curve (AUCs). The arithmetic mean of the three cohort-specific AUCs was calculated solely as a descriptive summary for visualization and reporting; it was not used for feature selection, hyperparameter tuning, or final

model selection. Because this average includes training-set performance, it was not interpreted as an external validation estimate. The Stepglm [backward] + plsRglm model retained HMOX1, JUNB, CXCL12, and PTGS2 as the four-gene feature panel. Annotated scripts documenting preprocessing, feature selection, model fitting, prediction, and evaluation are available in the GitHub repository: (https://github.com/HTX2023/Machine-Learning-and-Computational-Toxicology-PFAS-in-COPD).

2.5 Computational inference and statistical analysis of immune-cell signatures

CIBERSORT was applied to the normalized GSE151052 microarray expression matrix using the LM22 reference signature to estimate relative fractions of 22 immune-cell subsets (39). The analysis included 77 COPD and 40 control regional lung-tissue profiles. Quantile normalization was enabled (QN = TRUE), and permutation testing was not performed (perm = 0); therefore, no permutation-derived deconvolution P-value filter was applied.

COPD–control comparisons were performed using two-sided Wilcoxon rank-sum tests, with Benjamini–Hochberg (BH) adjustment across the 22 planned cell-type comparisons. Spearman correlations between HMOX1, JUNB, CXCL12, and PTGS2 expression and estimated immune-cell fractions were calculated separately in COPD samples and in all samples (40). Within each analysis, P values were adjusted jointly across the 88 planned gene–cell combinations. Four zero-variance subsets yielded untestable comparisons, which were recorded as NA; the planned adjustment family sizes of 22 and 88 were retained. For correlations among immune-cell fractions within COPD samples, these four subsets were excluded, and BH adjustment was applied across the 153 unique pairs among the remaining 18 subsets. BH-adjusted $P < 0.05$ was used as the statistical significance threshold for these regional-sample analyses. Complete results are provided in Supplementary Table S18.

As a complementary computational sensitivity analysis, xCell was applied to the same expression matrix using xCellAnalysis with rnaseq = FALSE. Its outputs were interpreted as enrichment scores rather than cell percentages. Group differences were assessed using Wilcoxon rank-sum tests, with BH adjustment across all evaluated output scores. Candidate gene–score correlations were assessed using Spearman tests, with BH adjustment across all evaluated gene–score combinations separately for the pooled and COPD-only analyses. The xCell results are provided in Supplementary Table S17.

Both methods infer immune-related signatures from bulk expression data rather than directly measuring cell abundance. Their reference signatures and output scales differ, and neither analysis explicitly accounted for within-donor dependence among regional samples. The findings were therefore interpreted as exploratory regional-tissue associations.

2.6 Consensus clustering of COPD regional expression profiles

To explore heterogeneity within COPD lung tissue, regional samples were grouped into C1 and C2 using consensus clustering of the candidate-gene expression profiles (41). The unit of clustering was a sampled lung region rather than an individual patient. Candidate-gene expression and ssGSEA-derived immune signatures were compared between the regional clusters (42). The analysis did not establish donor-consistent cluster membership, and C1/C2 were interpreted as regional molecular expression patterns rather than validated patient subtypes or clinical severity categories.

2.7 Single-cell transcriptomic analysis, quality control, batch correction, and cell-type annotation

The COPD-derived subset of the single-cell RNA-sequencing dataset GSE279570 was obtained from GEO, and expression matrices from 11 patients with COPD were imported into Seurat. The analysis was designed to characterize the cell-type distribution and relative transcript abundance of candidate genes within COPD lung tissue. No normal-control single-cell samples were included in the reported analysis, and no COPD-versus-control differential-expression analysis was performed at the single-cell level. For each cell, the number of detected genes, total UMI counts, mitochondrial gene percentage, ribosomal gene percentage, and hemoglobin gene percentage were calculated. Mitochondrial genes were identified using the "^MT-" pattern, ribosomal genes using "^RP[SL]", and hemoglobin genes using "^HB[^(P)]". To remove low-quality cells and cells with abnormal transcript profiles, cells were retained only if they met the following criteria: nFeature_RNA >300 and <3, 000, nCount_RNA <5, 000, mitochondrial gene percentage <20%, and hemoglobin gene percentage <1% (43, 44). After quality-control filtering, 25, 933 cells were retained for subsequent analysis. The upper thresholds for detected genes and UMI counts excluded profiles with unusually high transcript complexity but did not substitute for dedicated doublet detection. No dedicated doublet-calling algorithm or doublet-removal sensitivity analysis was performed. Consequently, the presence and potential influence of residual doublets remain uncertain. The absence of dedicated doublet detection is acknowledged as a methodological limitation in the Discussion. The filtered Seurat object was normalized using NormalizeData, followed by highly variable feature selection using FindVariableFeatures and data scaling using ScaleData. Principal component analysis was performed using variable genes. To reduce sample-level batch effects, Harmony integration was performed using orig.ident as the batch variable (45). The first 20 Harmony dimensions were used for neighbor graph construction, clustering, UMAP visualization, and t-SNE visualization. Clustering was performed using FindClusters at multiple resolutions, including 0.05, 0.1, 0.2, and 0.5, and the final cell-type annotation was determined according to cluster stability, dimensionality-reduction structure, and canonical marker-gene expression. Cell types were manually annotated using established canonical marker genes. Epithelial cells were identified by *EPCAM* and *KRT18*; endothelial cells by *PECAM1* and *VWF*; fibroblasts by *COL1A1* and *COL1A2*; T cells by *CD3D* and *CD3E*; NK cells by *NKG7* and *GNLY*; monocytes by *LYZ* and *CD14*; macrophages by *CD68*, *CD163*, and *MRC1*; B cells by *MS4A1* and *CD79A*; neutrophils by *S100A8*, *S100A9*, and *FCGR3B*; and dendritic cells by *FCER1A* and *CST3*. The expression of PFAS-COPD candidate genes across annotated cell types was

then examined. The COMPARTMENTS database was used to retrieve subcellular-localization annotations for the candidate proteins, providing complementary protein-localization context (46). The CellChat R package was used to analyze single-cell communication and reconstruct ligand–receptor networks and signaling pathways (47).

2.8 Experimental validation, clinical sample recruitment, and sample-size consideration

To examine candidate-gene expression under COPD-related conditions, we performed expression analyses using cell models, clinical samples, and public datasets. A549 cells, 16-HBE cells, and THP-1-derived macrophage-like cells were cultured and exposed to cigarette smoke extract (CSE), followed by RNA extraction and RT-qPCR. Each experimental condition included three independent biological replicates (n = 3), involving separately cultured cells, independent CSE treatment, RNA extraction, and reverse transcription. Each biological replicate was measured in three technical qPCR wells. Technical-replicate values were averaged to obtain one value per biological replicate before statistical analysis. Each plotted data point therefore represents an independent biological replicate.

For clinical RT-qPCR analysis, PBMCs from five patients with COPD and five healthy controls were obtained from the First Affiliated Hospital of Guangzhou Medical University. After thawing, PBMCs were differentiated for seven days with 50 ng/mL human macrophage colony-stimulating factor (M-CSF), followed by RT-qPCR. Each participant constituted one independent biological replicate (n = 5 per group). Each sample was measured in three technical qPCR wells, which were averaged to obtain one participant-level value for analysis.

For plasma protein assessment, fasting plasma samples were collected from 20 patients with stable COPD and 20 healthy volunteers at the General Hospital of Ningxia Medical University. Samples were processed within 2 h of collection and stored at −80 °C until measurement. Commercial human ELISA kits supplied by Jingmei Biotechnology, Nanjing, China, were used to measure HO-1/HMOX1 (JM-04663H1), PTGS2/COX-2 (JM-05273H1), SDF-1/CXCL12 (JM-05313H1), and JUNB (JM-7465H1). The manufacturer's instructions identify human plasma as an applicable sample type for all four kits. Kit information is provided in Supplementary Table S11.

Samples were prepared and diluted according to the corresponding manufacturer's instructions. Standards and prepared samples were added to the assay wells, followed by incubation with the kit-supplied antibody reagents. After washing to remove unbound material, substrate solutions A and B were added, and color development proceeded at 37 °C for 15 min in the dark. The reactions were terminated with stop solution, and absorbance was measured using a microplate reader. Analyte concentrations were determined from the corresponding standard curves, accounting for sample dilution. Each participant constituted one independent biological replicate (n = 20 per group). Each plasma sample was measured in duplicate, with two technical wells per analyte summarized as one participant-level value for statistical analysis.

Clinical characteristics, including sex, age, height, weight, and available pulmonary function status and GOLD classifications, are provided in Supplementary Table S12. Detailed smoking pack-years, comorbidities, and medication histories were incompletely documented in the available clinical records and could not be reliably incorporated into covariate adjustment.

COPD patients were eligible if they were adults with physician-diagnosed stable COPD based on clinical assessment and pulmonary function criteria consistent with GOLD recommendations, including persistent airflow limitation. Stable COPD was defined as the absence of acute exacerbation, acute respiratory infection, or major treatment change within the preceding four weeks. Healthy controls were individuals without COPD or other chronic respiratory diseases. The main exclusion criteria for both groups were acute infection, recent COPD exacerbation, asthma, bronchiectasis, interstitial lung disease, active malignancy, autoimmune disease, severe hepatic or renal dysfunction, recent systemic corticosteroid or immunosuppressive therapy, incomplete clinical information, and poor-quality or hemolyzed blood samples. For public datasets, all samples with clear COPD/control annotations and available expression matrices were included. Samples without definite phenotype annotation, samples failing the original quality-control procedures, unrelated disease groups such as ILD samples in GSE47460, and treatment arms or time points not relevant to the validation comparison in murine datasets were excluded. Dataset-specific source information and inclusion/exclusion rules are summarized in Supplementary Table S15.

Because the main discovery and validation analyses were based on fixed-size public GEO datasets and archived clinical validation samples, prospective sample-size calculation before data acquisition was not applicable. We therefore performed a *post hoc* sample-size and power consideration. For the plasma ELISA validation cohort, 20 COPD patients and 20 controls provided 80% power to detect a large between-group standardized effect size of approximately Cohen's d = 0.91 using a two-sided independent-sample t test with $\alpha = 0.05$ and equal allocation. For the PBMC-derived macrophage RT-qPCR cohort, 5 COPD patients and 5 controls provided 80% power only for very large effects, approximately Cohen's d = 2.02; therefore, this analysis was considered exploratory and supportive. The robustness of the overall findings was primarily supported by concordant evidence from multiple independent public datasets, machine-learning prioritization, immune and single-cell analyses, clinical ELISA validation, and computational toxicology analyses rather than by any single validation cohort alone.

This study followed the Declaration of Helsinki and received approval from the ethics committees of the First Affiliated Hospital of Guangzhou Medical University (Approval No. 2018-53) and the General Hospital of Ningxia Medical University (Approval No. KYLL-2025-1078). All participants provided written informed consent. Experimental data are presented as mean ± standard error of the mean (SEM). Statistical analyses and graph preparation for the experimental validation data were performed using GraphPad Prism 8.0. Two-group comparisons used two-sided unpaired Student's t tests. For the cell-model experiments, CSE-treated cells at each indicated time point were compared with the corresponding untreated controls separately for each gene and cell

model. Data from different genes, cell models, or time points were not pooled as biological replicates. Clinical comparisons were performed separately for each gene or protein using participant-level values. These clinical comparisons were not adjusted for covariates and were interpreted as exploratory associations. Technical qPCR replicates were averaged before analysis. Experimental P values were not adjusted for multiple comparisons, and P <0.05 was interpreted as nominal statistical significance. Statistical procedures for the computational immune-signature analyses, including BH correction, are described separately in Section 2.5.

2.9 Molecular docking and molecular dynamics simulation

Molecular docking and molecular dynamics (MD) simulations were used to explore candidate PFAS-protein interactions and the conformational behavior of docking-derived structures. Predicted three-dimensional models of the candidate proteins were obtained from the AlphaFold Protein Structure Database and prepared for docking using PyMOL (48). The two sentinel PFAS molecules, PFOA and PFOS, were docked individually with AutoDock 4 (GA: population 150, mutation 0.02, crossover 0.8; 10 runs × 2.5 × 10^6 evaluations) and cross-checked with AutoDock Vina (default settings). For each protein–ligand pair, the pose with the lowest predicted docking score was retained, and putative intermolecular contacts were inspected using PyMOL and Discovery Studio (49, 50). Docking scores are reported in kcal mol^{-1} and were interpreted as computational scores rather than experimentally measured binding free energies.

Docking-derived receptor–ligand complexes were preprocessed and segmented using PyMOL prior to MD simulations. Ligand topologies were generated with Sobtop (AMBER parameters); for proteins, we employed amber99sb-ildn in GROMACS 2020.3 (51, 52). After solvation and ion neutralization, the systems were (i) energy-minimized (5,000 steepest-descent step), (ii) heated to 300 K under NVT (100 ps, 2 fs step, positional restraints), and (iii) equilibrated at 1 bar/300 K under NPT (100 ps). One unrestrained 100-ns production trajectory was generated for each complex. Trajectories were characterized using RMSD, RMSF, radius of gyration, and PCA-based conformational free-energy landscapes. Selected structural views and two-dimensional interaction diagrams were visualized using PyMOL and Discovery Studio. These analyses described conformational variation and global compactness within the sampled trajectories. The conformational landscapes were not interpreted as protein–ligand binding free-energy estimates, and contacts illustrated in selected structures were not taken as evidence of their persistence throughout the simulations. The calculations were considered hypothesis-generating and do not independently establish trajectory convergence, persistent ligand binding, or functional modulation of the candidate proteins.

Unless otherwise specified, all online databases and web-based platforms used in this study, including GEO, PubChem, SwissTargetPrediction, CTD, UniProt, STRING, AlphaFold Protein Structure Database, and COMPARTMENTS, were accessed in June 2025.

3 Results

3.1 Overlap between PFAS-associated candidate genes and COPD-related gene sets

Through a literature review and database mining, we assembled a compendium of potential PFAS targets (Supplementary Table S1). Concurrent analysis of publicly accessible COPD transcriptomes identified DEGs between COPD patients and healthy controls in lung tissues (Figures 2A, B; Supplementary Table S2). GO analysis indicated enrichment of leukocyte cell–cell adhesion and T-cell activation among upregulated genes, together with receptor ligand and chemokine activities and the external side of the plasma membrane. Downregulated genes were enriched in amino acid metabolic processes and terms related to carboxylic acid binding and the collagen-containing extracellular matrix. KEGG analysis identified TNF signaling, cytokine–cytokine receptor interaction, and chemokine signaling among upregulated genes, whereas downregulated genes were enriched in metabolic pathways, HIF-1 signaling, complement and coagulation cascades, ECM–receptor interaction, and neuroactive ligand–receptor interaction (Figure 2C; Supplementary Table S2). Furthermore, WGCNA revealed coexpression modules that exhibited significant correlations with the COPD phenotype (Figures 2D, E). Multiple disease-associated modules were detected in the COPD samples, with the blue ($r = 0.86$, $P < 0.05$), turquoise ($r = 0.67$, $P < 0.05$), and green–yellow ($r = 0.53$, $P < 0.05$) modules demonstrating the strongest phenotypic correlations (Figures 2F, G; Supplementary Table S3). Comparison of the PFAS-associated candidate-gene set with the COPD-related gene sets identified 68 genes shared with the DEGs and 121 genes shared with the COPD-associated module genes. The three-way intersection contained 47 genes common to all three sets (Figure 2H). These overlapping genes were retained as candidates for subsequent network analysis and prioritization. The overlap was used as a descriptive selection criterion; no statistical enrichment test was performed for this intersection.

3.2 Nineteen hub genes identified in the PFAS–COPD interaction network

To outline interactions among intersecting genes, a PPI network was constructed using the STRING database. The resulting network exhibited a highly interconnected architecture, suggesting cooperative involvement of these genes in specific biological processes (Figures 3A, B; Supplementary Table S4). By employing four topological algorithms —closeness centrality, degree centrality, EPC, and MNC—we identified 19 hub genes at the network core: *ATF3, FASN, JUN, EGR1, CXCL2, FOSB, CXCL12, ARG1, HMOX1, CXCL1, JUNB, HGF, PTGS2, TNFSF10, IL1B, FOS, IL6, IER3*, and *CXCL10* (Figures 3C–E; Supplementary Table S5). These 19 genes were prioritized on the basis of network topology and are referred to as PPI hub genes in the subsequent analyses. Functional enrichment analysis revealed significant clustering of candidate genes in diverse biological processes and pathways. GO analysis revealed enrichment in cytokine-mediated signaling pathways, responses to bacterial molecular components,

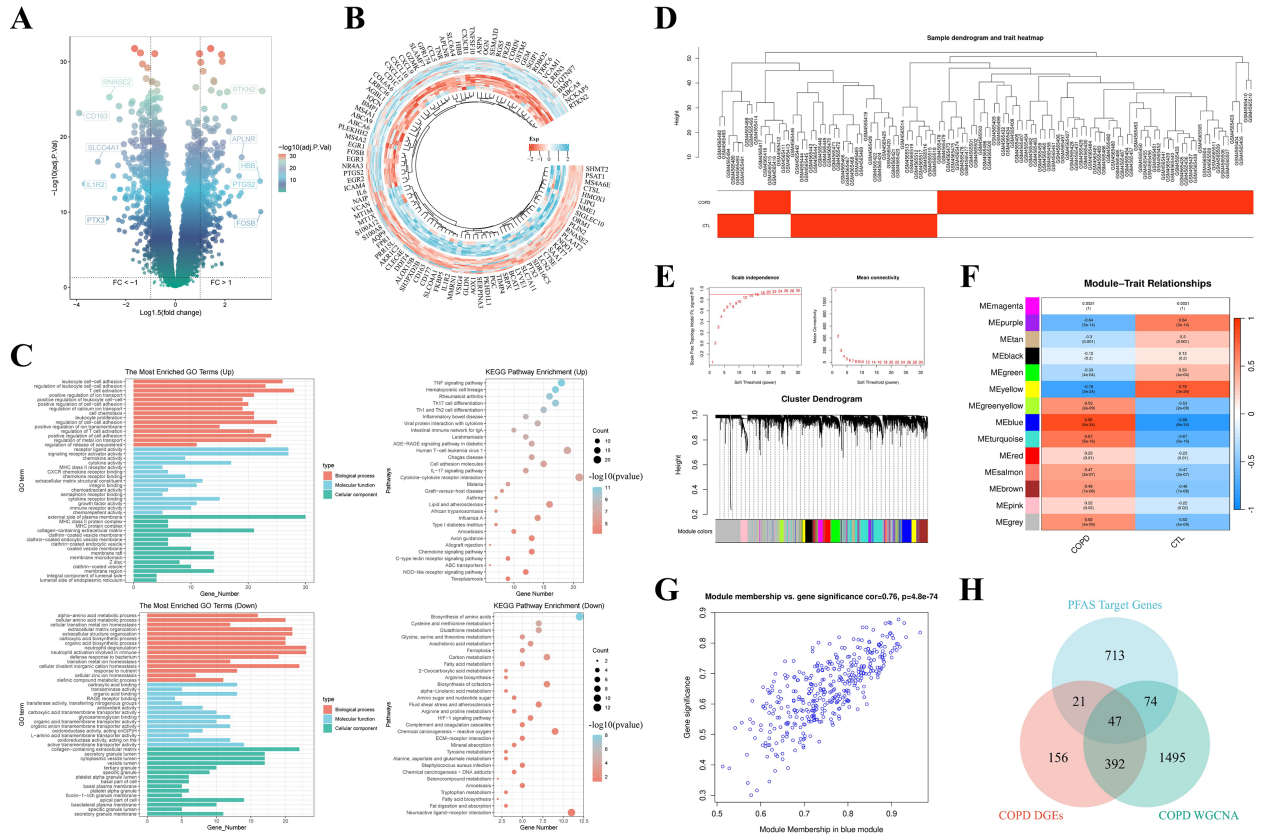

FIGURE 2
PFAS-associated genes and COPD transcriptomic analysis. **(A, B)** Differentially expressed genes (DEGs) between COPD patients and healthy controls were identified through transcriptomic analysis. **(C)** Enrichment analysis of up-regulated and down-regulated DEGs reveals significant clustering in various Gene Ontology (GO) and KEGG pathways. Up-regulated genes were predominantly enriched in chemokine signaling, T-cell activation, and cell adhesion molecule activity, while down-regulated genes were linked to lipid digestion and fatty acid biosynthesis. **(D–G)** Weighted Gene Co-expression Network Analysis (WGCNA) identified co-expression modules showing significant correlations with the COPD phenotype. The blue, turquoise, and green-yellow modules exhibited the strongest phenotypic associations. **(H)** Venn diagram showing the overlap among the PFAS-associated candidate-gene set, COPD DEGs, and COPD-associated module genes. The 47 genes shared by all three sets were retained for subsequent analyses.

and lipopolysaccharide responses (BP); transcription regulator complexes and RNA polymerase II regulatory complexes (CC); and signal receptor activator activity, receptor–ligand activity, and cytokine receptor binding (MF). KEGG pathway analysis revealed significant enrichment of the TNF signaling, IL-17 signaling, cytokine–cytokine receptor interaction, lipid metabolism and atherosclerosis, and rheumatoid arthritis pathways (31) (adjusted $P < 0.05$) (Figures 3F–I; Supplementary Table S6). Overall, these findings suggest that PFAS-associated hub genes are enriched in inflammatory and immune-response pathways, providing a basis for further evaluation of PFAS-COPD hub protein binding affinities and potential molecular associations.

3.3 Integrated machine learning identifies a four-gene candidate signature for COPD classification

A two-stage framework combining feature selection and classification generated 113 candidate machine-learning model combinations using GSE151052 as the training dataset (Figure 4A). The Stepglm [backward] + plsRglm model retained HMOX1, JUNB, CXCL12, and PTGS2 as the four-gene feature panel. ROC curves illustrate the discriminatory performance of the individual candidate genes (Figure 4F). When applied to the two external validation datasets, the trained model achieved AUCs of 0.807 and 0.763. The arithmetic mean AUC across the training dataset and both validation datasets was 0.855, reported solely as a descriptive summary rather than a model-selection criterion or an external validation estimate (Figures 4B–E; Supplementary Table S7).

These findings support the panel's ability to distinguish COPD from control expression profiles in the evaluated datasets. The four genes also overlapped with the PFAS-associated candidate set, supporting their potential relevance to the molecular associations investigated in this study. However, the discovery dataset comprised multiple regional samples from very severe COPD lungs and donor controls. The findings therefore do not establish the panel's utility for early COPD detection, population screening, or prospective risk prediction.

3.4 Candidate-gene expression is associated with inferred immune-cell signatures after multiple-testing correction

CIBERSORT estimated the relative fractions of 22 immune-cell subsets across the 117 regional lung-tissue profiles (Figure 5A).

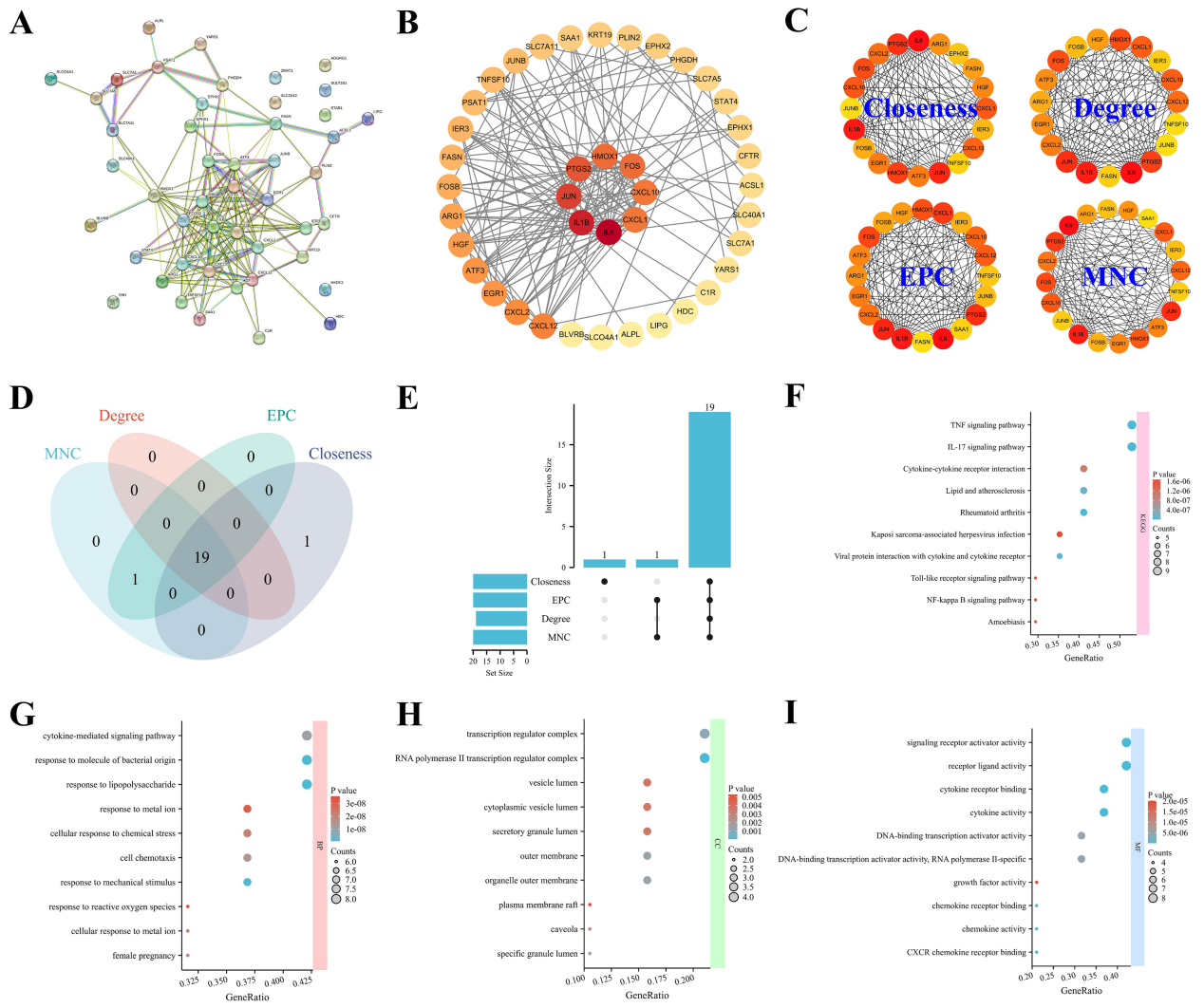

FIGURE 3
PFAS-COPD interaction network and hub gene identification. **(A, B)** Protein-protein interaction (PPI) network constructed using STRING database, revealing an interconnected network of genes related to PFAS exposure and COPD pathogenesis. **(C–E)** Identification of 19 hub genes (e.g., *ATF3, FASN, JUN, EGR1, CXCL2*, etc.) using topological metrics (degree centrality, closeness centrality, EPC, and MNC). These genes were prioritized as network hubs on the basis of their topological scores. **(F–I)** Functional enrichment analysis of the hub genes reveals their involvement in various biological processes (e.g., cytokine-mediated signaling) and pathways (e.g., TNF signaling, IL-17 signaling, lipid metabolism, and atherosclerosis), suggesting their potential relevance to PFAS-associated immune-inflammatory signatures in COPD.

Eleven subsets showed COPD–control differences at BH-adjusted $P < 0.05$. COPD samples had higher median estimated fractions of naive B cells, CD8+ T cells, M1 macrophages, and resting mast cells, and lower median fractions of activated NK cells, monocytes, M0 and M2 macrophages, activated dendritic cells, and neutrophils. Follicular helper T-cell fractions also differed between groups (BH-adjusted $P = 0.0105$), although their median estimated fraction was zero in both groups, indicating a distributional difference rather than a difference in medians (Figure 5B; Supplementary Table S18).

Within the 77 COPD regional samples, 16 gene–immune-cell associations remained significant after BH correction. CXCL12 expression was positively associated with estimated M1 macrophage fractions (Spearman rho = 0.634, BH-adjusted $P = 5.14 \times 10{-}8$) and resting memory CD4+ T-cell fractions, and negatively associated with resting NK cells, monocytes, M0 macrophages, activated dendritic cells, and neutrophils (Figure 5C). HMOX1 expression was positively associated with M0 and M2 macrophage fractions and negatively associated with plasma-cell and resting mast-cell fractions (Figure 5D). JUNB expression was positively associated with naive B-cell and follicular helper T-cell fractions and negatively associated with M2 macrophage fractions (Figure 5E). PTGS2 expression was negatively associated with neutrophil fractions (rho = −0.450, BH-adjusted $P = 4.97 \times 10{-}4$) and resting mast-cell fractions (rho = −0.349, BH-adjusted $P = 0.0110$; Figure 5F). The previously nominally significant HMOX1–M1 macrophage and JUNB–plasma-cell associations did not meet the adjusted significance threshold.

The correlation heatmap summarizes associations among the 18 immune-cell subsets with nonzero variance in COPD samples, while the gene–immune network displays correlation strength and BH-adjusted significance (Figures 5G, H). Complete coefficients and adjusted P values are provided in Supplementary Table S18, and the complementary xCell results are reported in Supplementary Table S17. These findings describe associations between candidate-gene expression and inferred immune signatures; they do not

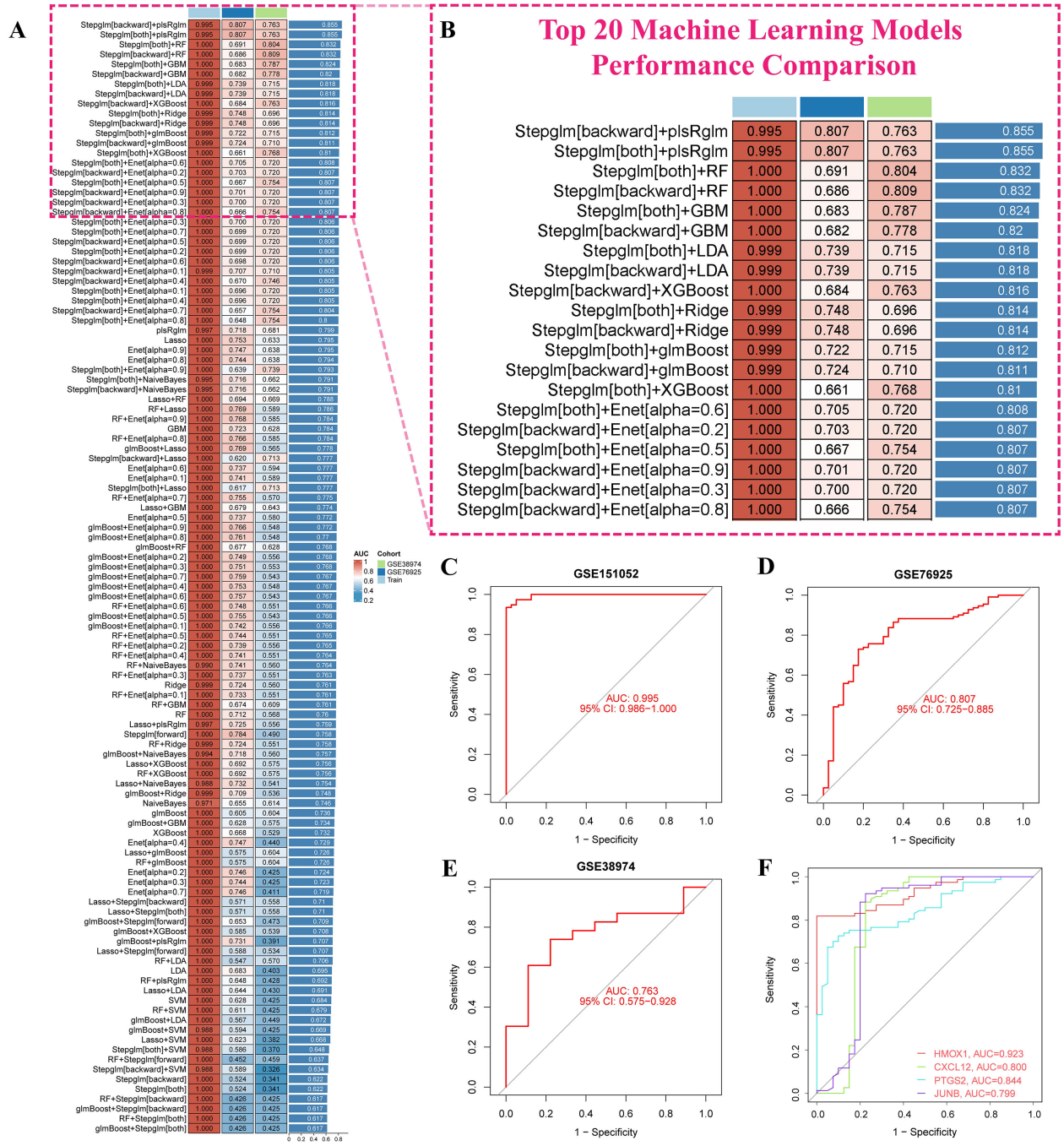

FIGURE 4
Identification and evaluation of a four-gene candidate signature using integrated machine learning. **(A)** Overview of the 113 candidate machine-learning model combinations. **(B–E)** AUC-based performance summaries and ROC curves for model evaluation in the training and external validation datasets. The Stepglm[backward] + plsRglm model achieved AUCs of 0.807 and 0.763 in the two external validation datasets. The mean AUC of 0.855 represents the descriptive arithmetic average across all three datasets and was not used for model selection. **(F)** ROC curves for the individual candidate genes HMOX1, JUNB, CXCL12, and PTGS2.

establish direct cellular interactions, changes in experimentally measured cell abundance, or PFAS-mediated immune regulation.

3.5 Two regional molecular expression patterns identified in COPD lung tissue

Consensus clustering of the candidate-gene expression profiles identified two groups of COPD regional tissue samples, C1 and C2 (Figures 6A–E; Supplementary Table S9). These groups showed different candidate-gene expression patterns. ssGSEA-derived scores for eosinophils, mast cells, natural killer cells, and Th1, Th17, and Th2 cell signatures were higher in C2, whereas macrophage-signature scores were higher in C1 (Figure 6F). These findings describe computationally inferred differences between regional expression patterns. They do not establish mutually exclusive patient subtypes, within-donor consistency, differences in

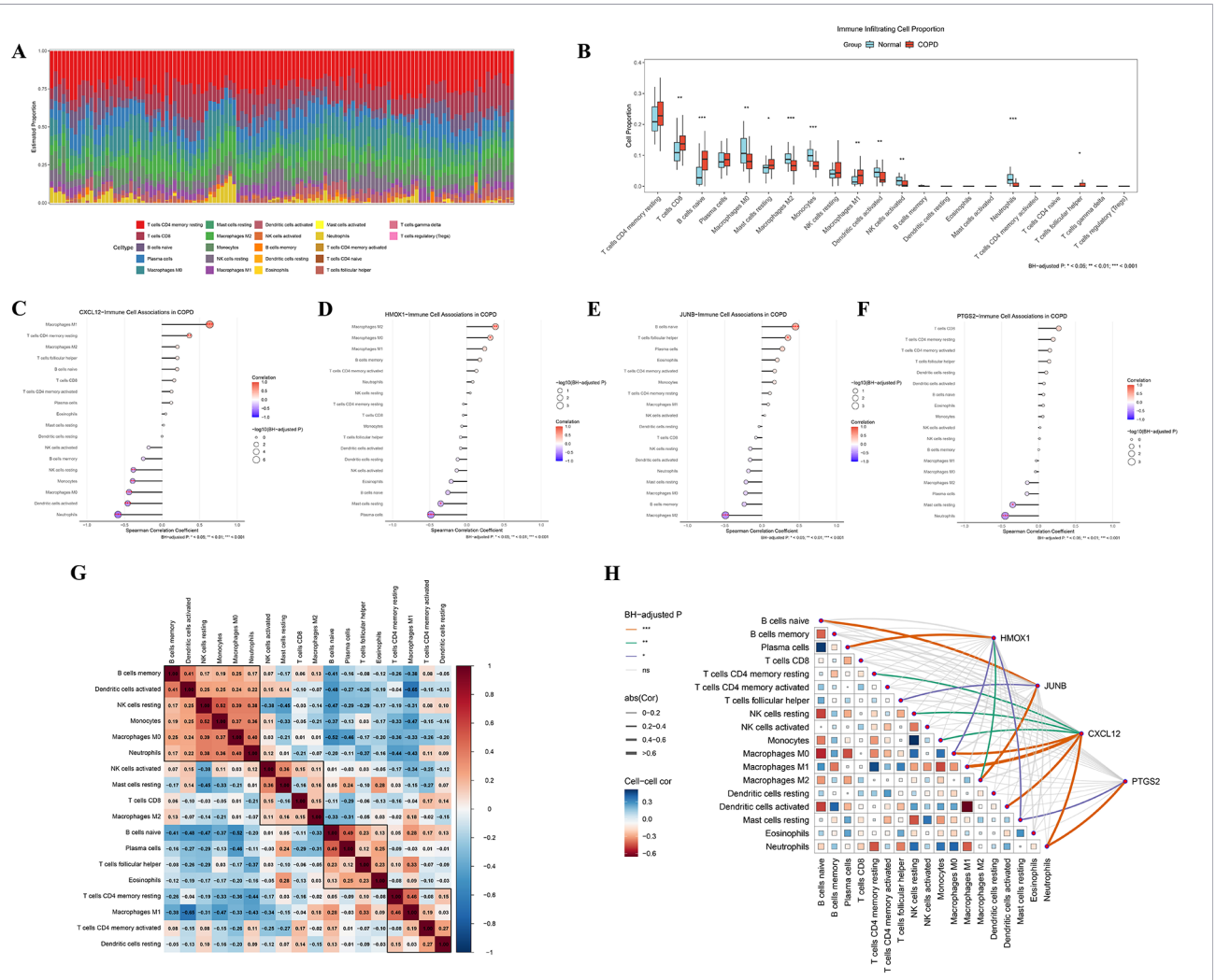

FIGURE 5
CIBERSORT-derived immune-cell signatures and BH-adjusted associations with candidate-gene expression. **(A)** Estimated relative fractions of 22 LM22 immune-cell subsets across 117 regional lung-tissue samples. **(B)** COPD–control comparisons using two-sided Wilcoxon rank-sum tests with BH adjustment across 22 planned comparisons. **(C–F)** Spearman correlations of CXCL12 **(C)**, HMOX1 **(D)**, JUNB **(E)**, and PTGS2 **(F)** expression with estimated immune-cell fractions in the 77 COPD regions. Point position and color represent the correlation coefficient, and point size represents −log10(BH-adjusted P). Adjustment was performed jointly across the 88 planned gene–cell combinations. **(G)** Spearman correlation coefficients among the 18 immune-cell subsets with nonzero variance in COPD samples. **(H)** Gene–immune association network; edge width represents absolute correlation strength, edge color represents BH-adjusted significance, and background tiles display cell–cell correlation coefficients. Where indicated, *BH-adjusted P < 0.05, **<0.01, and ***<0.001; ns, not significant.

measured immune-cell abundance, or clinical disease-severity stratification.

3.6 Cell-type expression patterns and inferred communication in COPD lung tissue

We analyzed 25, 933 cells retained after quality filtering from the COPD-derived subset of GSE279570, representing 11 patients. No normal-control single-cell samples were included. Quality-control metrics are shown in Supplementary Figure S1. Following normalization, Harmony integration, clustering, and marker-based annotation, the dataset contained macrophages, monocytes, epithelial cells, endothelial cells, T cells, NK cells, B cells, neutrophils, dendritic cells, fibroblasts, and smaller annotated populations (Figures 7A–C; Supplementary Table S10). Canonical marker expression supported the broad cell-type annotations (Figure 7E–F). Macrophages constituted the largest proportion of recovered cells, followed by T cells and NK cells (Figure 7D). These proportions describe the analyzed cell collection and should not be interpreted as direct estimates of *in vivo* lung-cell abundance.

Figure 8 combines complementary evidence from three sources. The COPD–control expression comparisons in panels A, E, I, and M were derived from GSE151052 bulk microarray data and showed higher tissue-level expression of the four candidate genes in COPD samples. The single-cell expression maps in panels B, F, J, and N describe expression within COPD tissue only. JUNB expression was detected across macrophages, epithelial cells, neutrophils, and T cells. HMOX1 expression was prominent in macrophages, monocytes, and epithelial cells, whereas PTGS2 expression was most evident in neutrophils, with expression also observed in epithelial cells and macrophages. CXCL12 showed relatively low detected

10.3389/fimmu.2026.1965629

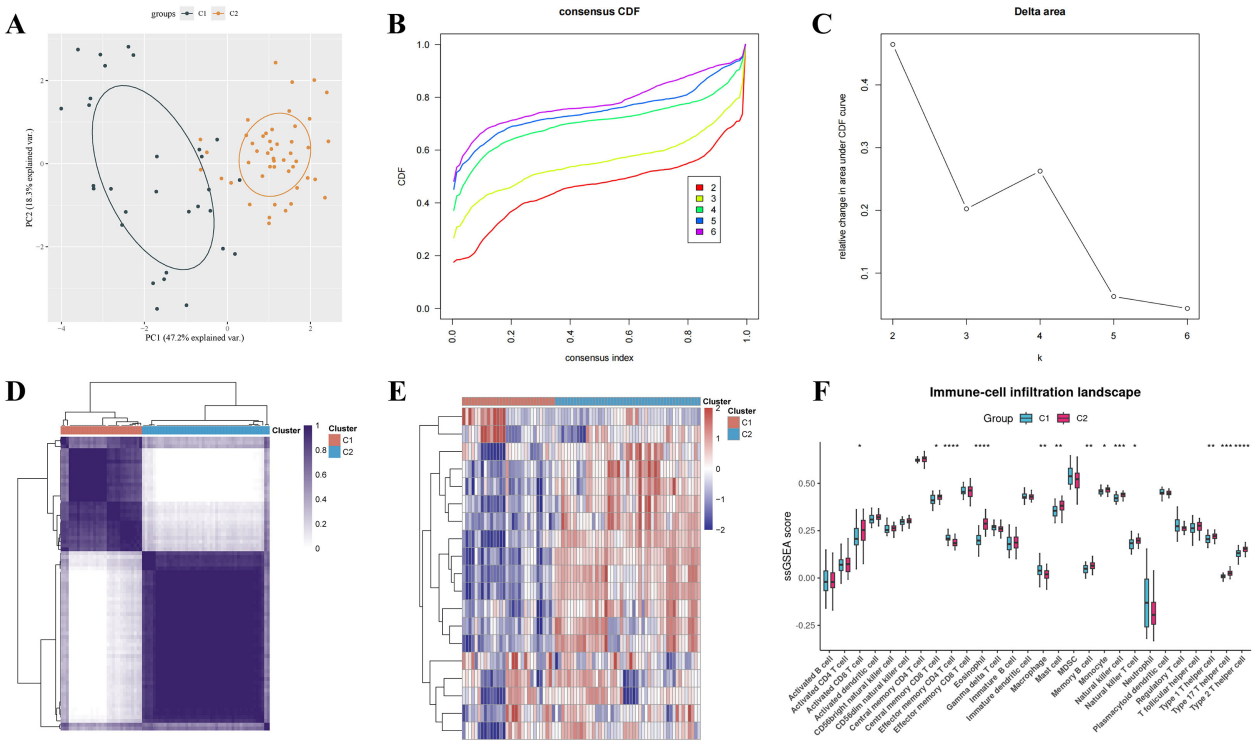

FIGURE 6
Regional molecular expression patterns in COPD lung tissue. **(A–E)** Consensus clustering of regional COPD lung-tissue samples based on candidate-gene expression, identifying patterns C1 and C2. **(F)** Comparison of ssGSEA-derived immune-signature scores between the regional clusters. These scores are computational estimates, and the clusters do not represent validated patient-level subtypes or severity categories. Significance symbols: *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001; ns, not significant.

expression in the analyzed cells. These patterns provide cellular context for the candidate genes but do not demonstrate COPD-associated upregulation within individual cell types.

The remaining Figure 8 panels present subcellular-localization annotations from COMPARTMENTS (Figures 8C, D, G, H, K, L, O, P). These database-derived annotations are distinct from the bulk expression comparisons and single-cell expression maps and do not establish disease-specific protein localization.

Functional enrichment analyses highlighted biological processes associated with the annotated epithelial and macrophage populations (Figure 9). Epithelial-cell-associated terms included alveolar morphogenesis, airway development, secretory-cell differentiation, and lipid-related processes, whereas macrophage-associated terms included ERK1/ERK2 signaling and complement activation. These enrichment patterns describe the analyzed gene sets within COPD samples and do not establish pathway activation relative to normal tissue.

CellChat analysis inferred potential communication between epithelial cells and macrophages through the represented ligand–receptor pairs and signaling pathways (Figures 10A–D). Incoming and outgoing communication patterns included GALECTIN, CypA, UGRP1, and resistin-related signaling (Figures 10E–L). These results provide hypotheses concerning communication among cell populations in COPD tissue. They do not demonstrate increased signaling relative to controls, experimentally confirmed ligand–receptor activity, or a causal contribution to COPD progression.

3.7 Multi-level validation supported COPD-associated dysregulation of target genes

RT-qPCR analyses of the CSE-treated cell models showed higher HMOX1 and PTGS2 mRNA expression than in untreated controls in the indicated comparisons (n = 3 independent biological replicates per condition; nominal P <0.05; Figures 11A–F). In PBMC-derived macrophages, HMOX1 and JUNB expression was higher in the COPD group than in controls (n = 5 participants per group; nominal P <0.05; Figure 11G). Plasma ELISA measurements showed higher ELISA-derived concentrations of HMOX1, JUNB, CXCL12, and PTGS2 in participants with COPD than in controls (n = 20 participants per group; nominal P < 0.05; Figure 11H). These experimental comparisons used P values unadjusted for multiple testing, and the clinical comparisons were additionally unadjusted for covariates. The plasma findings provided supportive, exploratory evidence of COPD-associated differences in circulating immunoreactivity, without establishing the tissue of origin, biological activity, or independence from potential confounding factors. Given the consistently elevated expression of *HMOX1* across cell lines and human samples, public transcriptomic datasets (GSE38974 and GSE47460 for human COPD; GSE125521 and GSE52509 for murine 6-month smoke exposure) were utilized to further validate this finding. The analysis confirmed the significant upregulation of *HMOX1* expression in COPD lung samples and smoke-exposed murine lungs (Figure 11I). Together, these findings

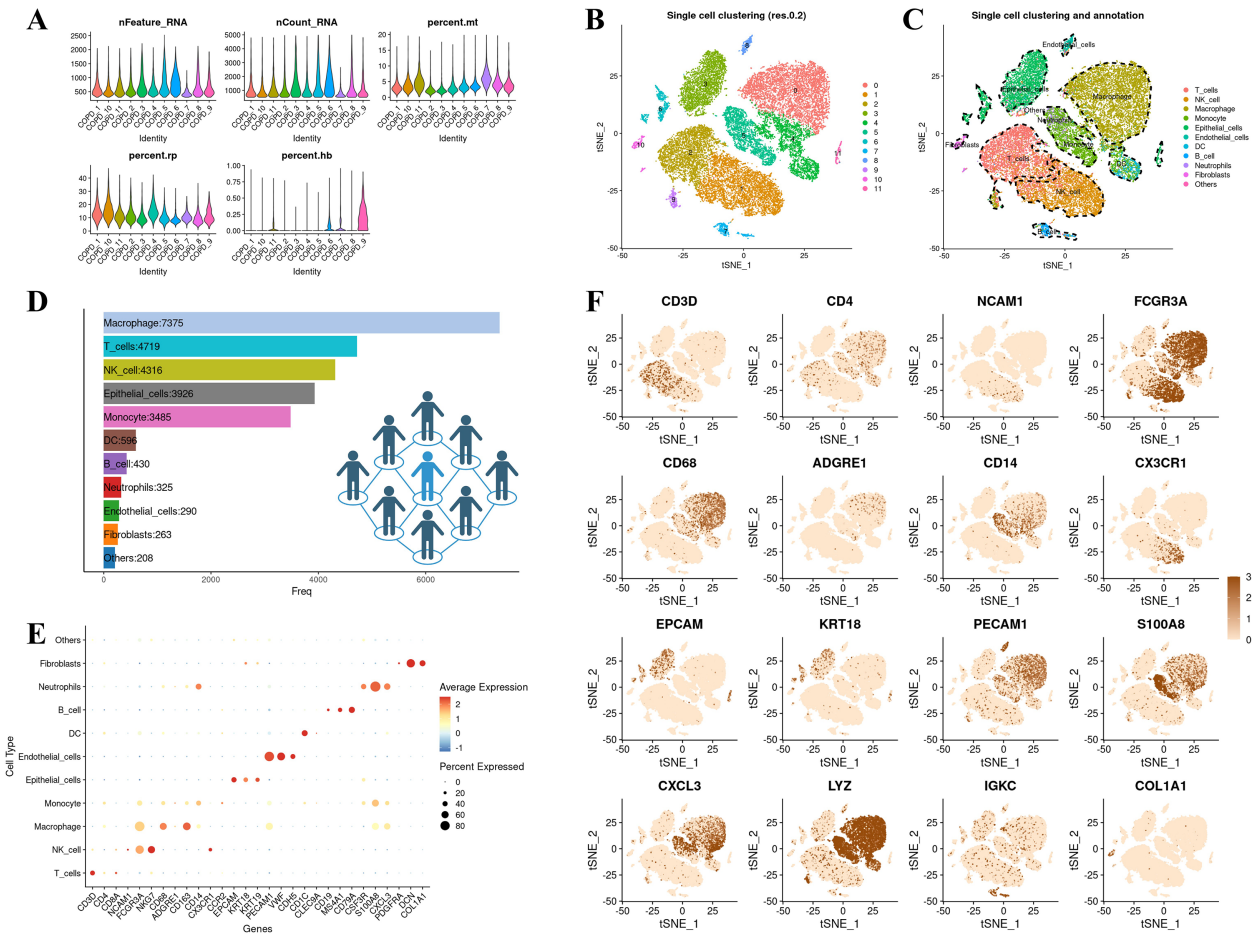

FIGURE 7
Single-cell profiling of the COPD-derived subset of GSE279570, comprising 25,933 cells retained after quality filtering from 11 patients with COPD. **(A)** Violin plots showing the number of detected genes, total RNA counts, and percentages of mitochondrial, ribosomal, and hemoglobin transcripts. **(B, C)** Cell clustering and broad cell-type annotations. **(D)** Numbers of recovered cells in each annotated population. **(E, F)** Canonical marker-gene expression supporting cell-type annotation. No normal-control single-cell samples were included. The displayed cell counts describe the analyzed dataset rather than absolute cell abundance in lung tissue.

support COPD-associated expression alterations in the candidate genes, with HMOX1 receiving additional support across the examined models and datasets. These experiments and public-data comparisons did not test direct regulation by PFOA or PFOS.

3.8 Molecular docking identifies candidate PFAS–protein binding poses

Molecular docking generated candidate binding poses for PFOA and PFOS with the four investigated proteins, with predicted docking scores ranging from −5.1 to −8.4 kcal mol^{-1} (Figure 12A; Supplementary Table S13). HMOX1 yielded the most negative scores among the examined protein–ligand pairs: −8.1 kcal mol^{-1} for PFOA and −8.4 kcal mol^{-1} for PFOS. The corresponding scores for PTGS2, JUNB, and CXCL12 were −6.8, −5.7, and −5.4 kcal mol^{-1} for PFOA and −7.8, −5.1, and −5.5 kcal mol^{-1} for PFOS, respectively. Structural inspection illustrated possible ligand accommodation and putative intermolecular contacts within the predicted HMOX1 binding region (Figure 12B). These docking results supported selecting HMOX1 for exploratory simulations but do not establish experimental binding affinity or persistent complex formation.

3.9 Molecular dynamics simulations characterize conformational changes in predicted PFAS–HMOX1 complexes

HMOX1 was selected for exploratory MD simulations based on its docking scores and complementary COPD-associated expression evidence. One 100-ns production simulation was performed for each docking-derived PFOA–HMOX1 and PFOS–HMOX1 structure (Supplementary Table S14). Potential-energy and equilibration diagnostics are shown in Figures 13A–D, M–P.

The RMSD trajectories showed substantial initial deviation from the reference structures. Following the early increase, the PFOA–HMOX1 RMSD fluctuated at approximately 1.4–1.7 nm and became comparatively more level during the later portion of the trajectory (Figure 13E). The PFOS–HMOX1 RMSD fluctuated at approximately 1.2–1.5 nm after the initial rise, with a residual upward trend (Figure 13Q). These profiles indicate appreciable conformational changes and do not support uniform structural stability throughout the simulations.

The total radius of gyration decreased initially and subsequently fluctuated around approximately 2.0 nm in both systems

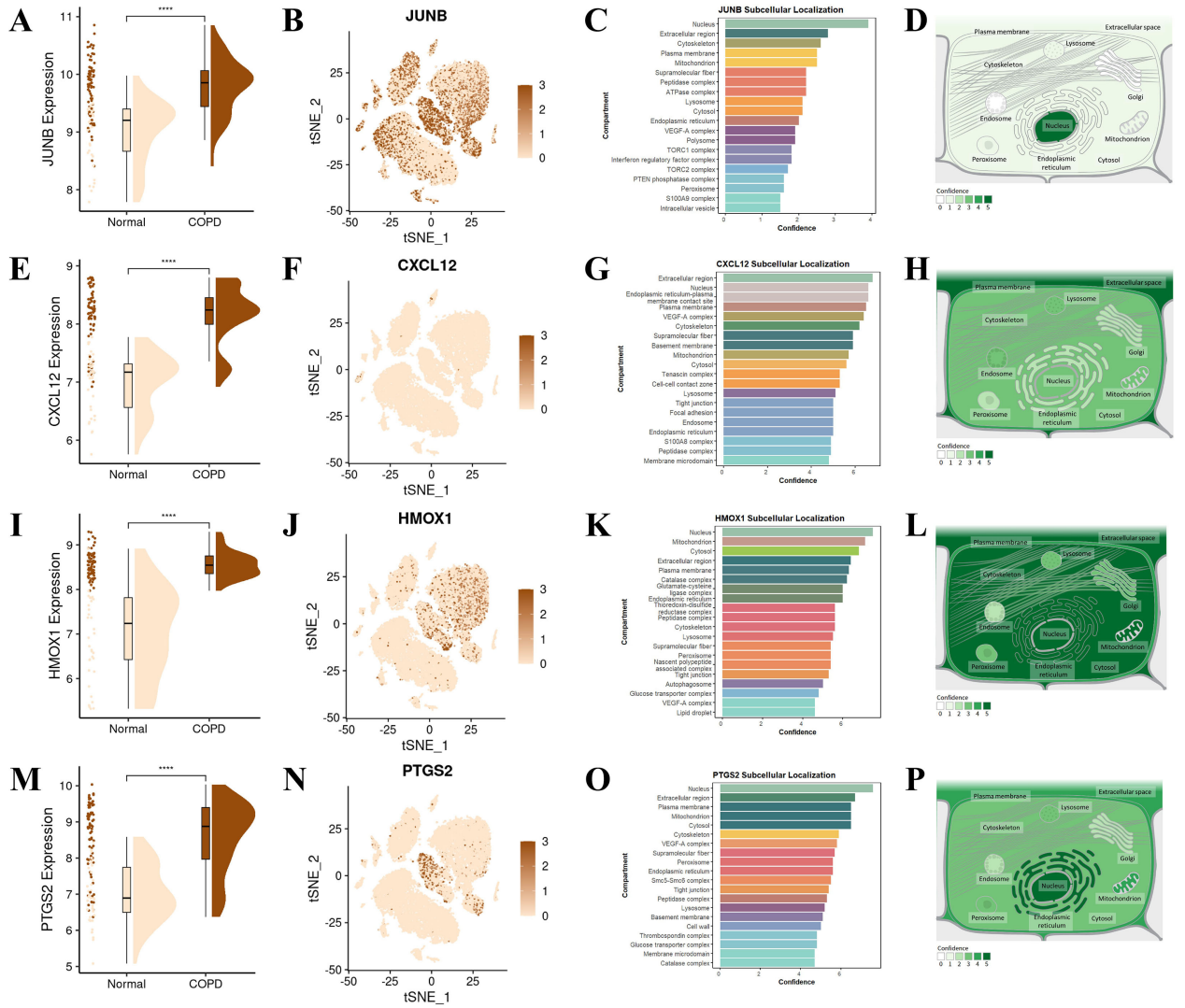

FIGURE 8
Bulk-tissue expression, single-cell expression patterns, and subcellular-localization annotations of candidate genes. **(A, E, I, M)** Normal-versus-COPD expression comparisons derived from the GSE151052 bulk microarray dataset, comprising 40 control and 77 COPD regional lung-tissue samples. These panels display tissue-level expression measurements and are not single-cell comparisons. **(B, F, J, N)** Candidate-gene expression maps in the COPD-derived subset of GSE279570, comprising 25, 933 cells retained after quality filtering from 11 patients. These panels show expression across cells within COPD tissue; no normal-control single-cell group was analyzed. **(C, D, G, H, K, L, O, P)** Subcellular-localization annotations retrieved from COMPARTMENTS. These annotations provide database-derived protein-localization context and do not represent experimental localization measurements in COPD samples. Cellular localization schematics in panels D, H, L, and P are reproduced from COMPARTMENTS (https://compartments.jensenlab.org/; Reference 46) under the Creative Commons Attribution 4.0 International license (CC BY 4.0; https://creativecommons.org/licenses/by/4.0/). ****P < 0.0001 for the bulk-tissue COPD–control expression comparisons.

(Figures 13H, T), consistent with initial compaction followed by relatively maintained global compactness. The coexistence of a comparatively steady Rg and elevated RMSD indicates that limited changes in overall molecular size can accompany substantial deviation from the reference conformation. The RMSF profiles showed heterogeneous atomic fluctuations across the analyzed structures (Figures 13G, S).

The PCA-based conformational landscapes illustrated the distribution of sampled conformations (Figures 13I, U), while selected structural views and interaction diagrams showed possible protein–ligand contacts in individual configurations (Figures 13J–L, V–X). These observations characterize the sampled structures but do not establish trajectory convergence, persistent ligand binding, or binding affinity. Together with the docking results, they provide an exploratory structural context for investigating potential PFAS–HMOX1 interactions.

4 Discussion

As widely used and highly stable organic pollutants, PFAS are referred to as “forever chemicals” because of their persistent presence in the environment and biological systems (53, 54). Their persistent environmental presence and bioaccumulation in biological systems have raised significant concerns (55). Two major NHANES studies linked PFAS exposure to COPD. Wang et al. reported that men with high PFOA levels had a 2.17-fold greater

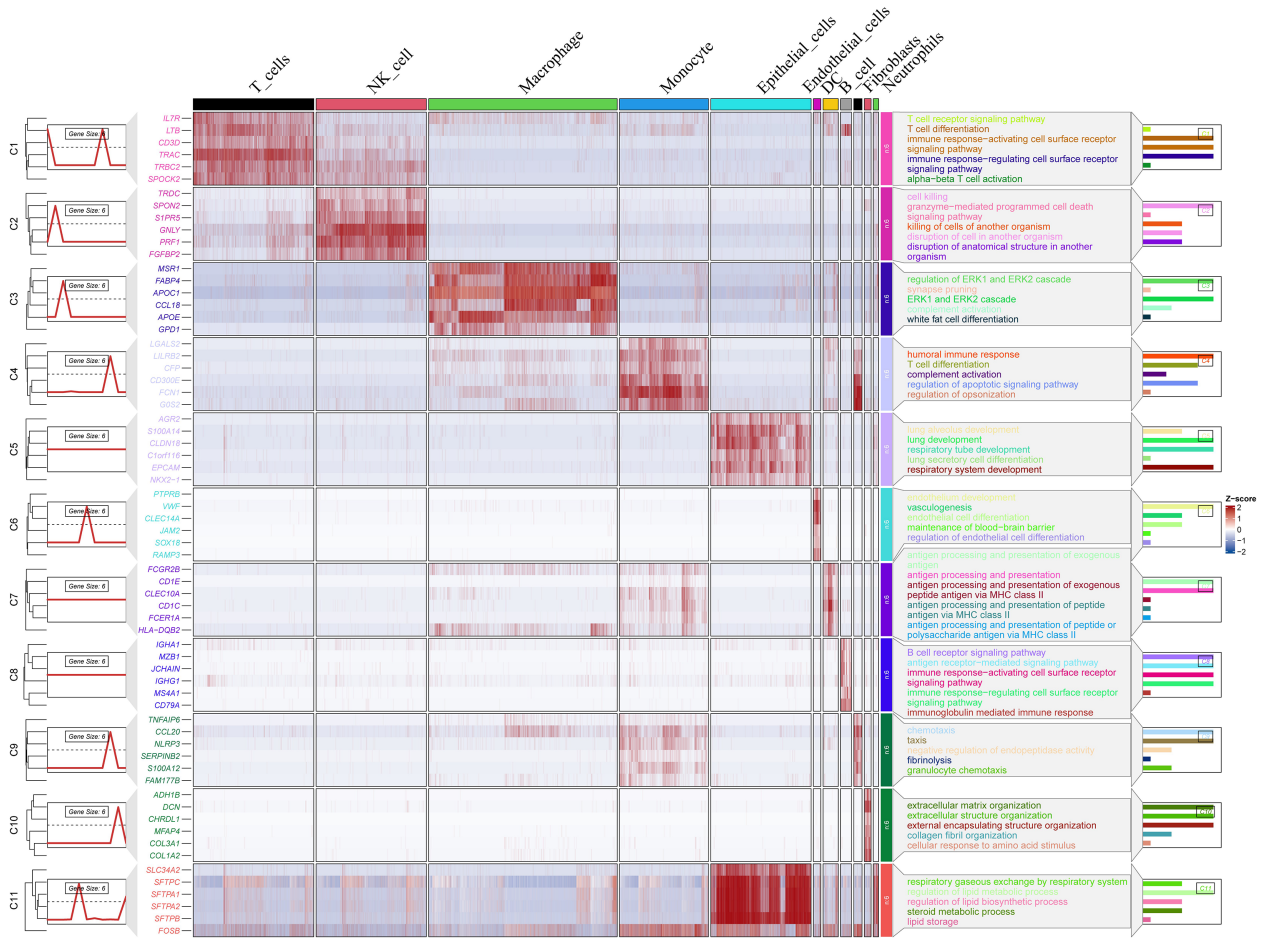

FIGURE 9
Cell-type-associated gene expression and functional enrichment in COPD. The heatmap displays gene-expression patterns across annotated cell populations, with colored annotations indicating gene clusters and cell types. Side panels summarize cluster-specific expression patterns and associated Gene Ontology terms. Epithelial-cell-associated terms include lung development and lipid-related processes, whereas macrophage-associated terms include ERK1/ERK2 signaling and complement activation. These results describe functional associations within the analyzed COPD dataset and do not establish pathway activation relative to normal controls.

COPD risk, with PFNA showing similar results (14). Pan et al. also noted an increased COPD prevalence with increased PFOA and PFOS levels, especially in older adults and former smokers, highlighting PFAS as new environmental risk factors for COPD (15). Furthermore, elevated PFAS burdens have been associated with poorer lung performance in young individuals. Qin et al. reported that asthmatic children with higher serum concentrations exhibit markedly reduced FEV_1, FVC, and related indices (17, 18). Toxicological studies corroborate this risk, demonstrating that PFAS trigger oxidative and immunotoxic stress in the lung epithelium, epigenetically disrupt the cell cycle and apoptotic control, impair surfactant function, and activate the NLRP3 inflammasome with widespread inflammatory gene dysregulation, thereby offering mechanistic support for their emerging link to COPD (19, 21, 56). Dragon et al. reported changes in inflammation- and immune-related gene expression in mouse lungs following 14 weeks of PFOA exposure through drinking water (21). These findings provide pulmonary transcriptomic evidence of an inflammatory response but do not directly establish NLRP3 inflammasome activation in lung tissue. These findings collectively indicate that PFAS can profoundly alter pulmonary biological states, triggering inflammatory cascades and potentially promoting airway hyperreactivity. However, these findings do not by themselves establish a direct causal relationship between PFAS exposure and COPD. Instead, they support the need for systematic analyses of molecular associations linking PFAS-related target profiles with COPD-relevant biological pathways.

By integrating multi-omics analyses and machine learning, we identified four candidate genes linking PFAS-associated target information with COPD-related expression patterns. Their expression showed 16 associations with CIBERSORT-derived immune-cell fractions that met the BH-adjusted $P < 0.05$ threshold within COPD regional samples, supporting their relationship with inferred immune signatures while retaining the limitations of computational inference and repeated regional sampling.

The contribution of these findings should be considered alongside existing COPD and PFAS research. HMOX1, JUNB, CXCL12, and PTGS2 are established participants in COPD-related and/or broader inflammatory and stress-response biology, as discussed below. Previous COPD studies have already combined single-cell profiling, bulk lung expression evidence, and experimental investigations to characterize disease-associated cellular changes (57).

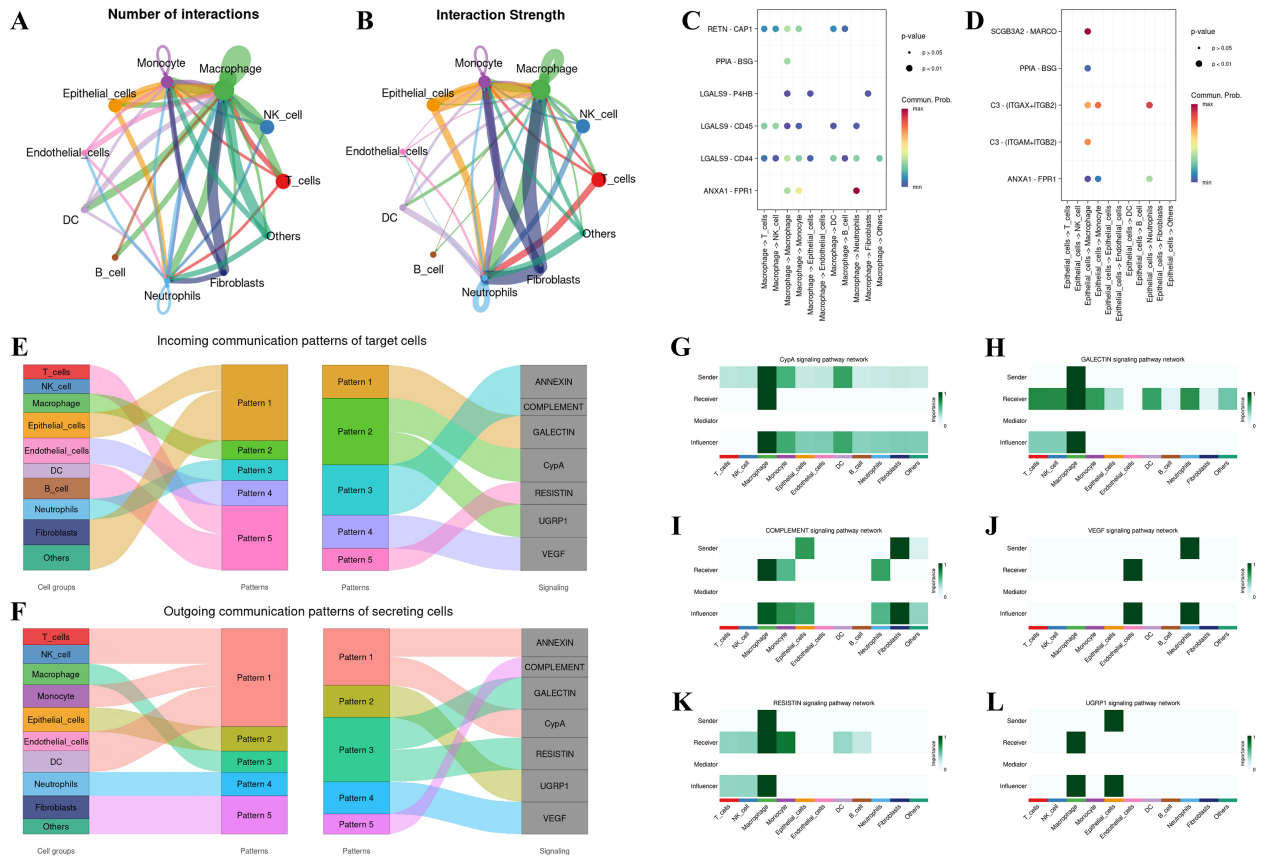

FIGURE 10
CellChat-inferred communication patterns in COPD lung tissue. **(A, B)** Inferred cell–cell communication networks across annotated cell populations, with edges representing the number **(A)** and aggregate strength **(B)** of interactions. **(C, D)** Selected predicted ligand–receptor associations between annotated cell populations, including associations involving LGALS9 and SCGB3A2. **(E, F)** Sankey diagrams depicting inferred incoming **(E)** and outgoing **(F)** communication patterns across annotated cell populations. **(G–L)** Selected signaling pathways represented in the inferred communication networks. All results were computationally inferred from COPD-derived single-cell expression data and do not establish experimentally verified signaling activity, functional effects on barrier or immune processes, or differences relative to normal controls.

PFAS toxicogenomic studies have characterized transcriptional responses under controlled exposure conditions in human hepatocyte models (58), while integrated PFAS-target, machine-learning, and single-cell approaches have also been applied to hepatocellular carcinoma (27). These studies provide precedents for the present analytical framework.

Within this context, our incremental contribution is the use of PFAS-associated target overlap together with COPD differential expression and co-expression modules to prioritize candidates, followed by their characterization across tissue expression, inferred immune signatures, and cellular expression patterns within COPD samples. This connects established disease-associated genes with a defined chemical-associated candidate set and identifies cellular contexts for subsequent investigation. The COPD-related experiments support disease relevance, while docking and molecular dynamics provide structural hypotheses. Thus, the principal advance is contextual candidate prioritization; the findings do not establish PFAS-specific biomarkers, clinical superiority over existing COPD panels, or direct PFAS-mediated mechanisms.

Single-cell RNA sequencing characterized the cellular distribution of the candidate genes within COPD tissue. HMOX1 and JUNB were detected in macrophages and epithelial cells, PTGS2 expression was most evident in neutrophils, and CXCL12 showed relatively low detected expression in the analyzed cells. Bulk-tissue comparisons indicated higher expression of all four candidate genes in COPD, and plasma ELISA measurements provided additional exploratory support. RT-qPCR findings varied by gene and experimental model. Molecular docking identified favorable predicted scores for PFOA/PFOS–HMOX1 poses, while the exploratory MD trajectories showed substantial conformational changes followed by relatively maintained global compactness. These computational observations provide structural hypotheses for further investigation without establishing persistent binding or functional regulation of HMOX1.

Machine-learning analysis prioritized HMOX1, JUNB, CXCL12, and PTGS2 as a four-gene feature panel drawn from the PFAS-associated candidates overlapping COPD-related gene sets. Their potential biological significance and mechanistic roles are herein discussed in light of the extant literature. *HMOX1* encodes heme oxygenase-1, a cytoprotective enzyme integral to the oxidative stress response (59). By degrading heme to yield antioxidant products, heme oxygenase-1 attenuates oxidative injury and confers protection against COPD pathogenesis (60, 61). An elongated $(GT)_n$ repeat within the *HMOX1* promoter diminishes heme oxygenase-1 expression and increases COPD susceptibility (62). Our observations revealed notable upregulation of *HMOX1* in patients with COPD, potentially indicating a compensatory mechanism in response to persistent oxidative stress.

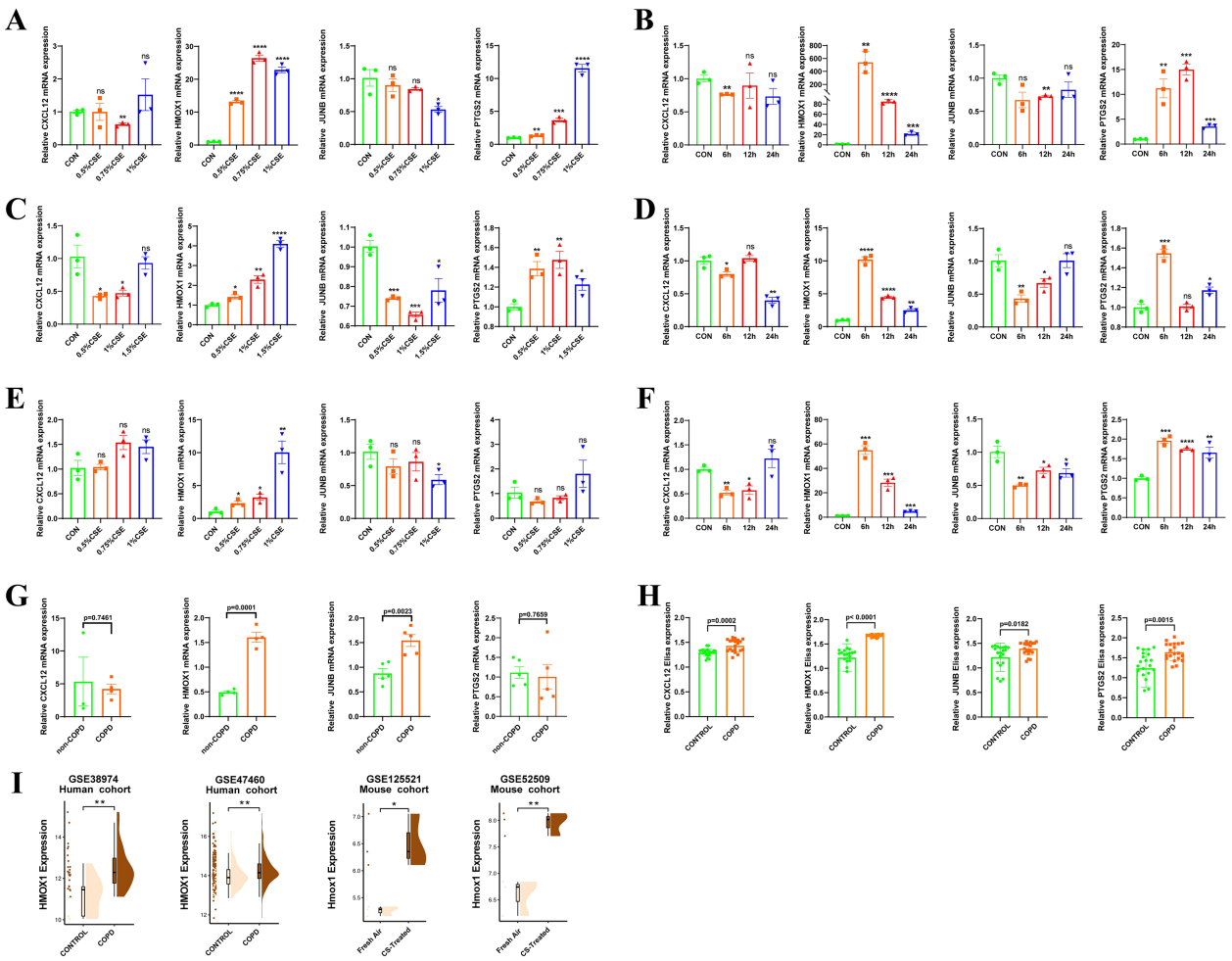

FIGURE 11
Candidate-gene and protein expression in COPD-related experimental models and clinical samples. **(A–F)** RT-qPCR analysis of candidate-gene expression in CSE-treated A549 cells, 16-HBE cells, and THP-1-derived macrophage-like cells and the corresponding untreated controls. Each condition included three independent biological replicates (n = 3), with three technical qPCR wells per biological replicate. Technical-replicate values were averaged before analysis, and each plotted data point represents one biological replicate. **(G)** RT-qPCR analysis of PBMC-derived macrophages from participants with COPD and healthy controls (n = 5 participants per group), with three technical qPCR wells per sample. Each participant represents one biological replicate. **(H)** Plasma ELISA measurements of HMOX1, JUNB, CXCL12, and PTGS2 in participants with COPD and controls (n = 20 participants per group). Each plasma sample was measured in duplicate (two technical wells per analyte), yielding one ELISA-derived concentration value per participant. **(I)** HMOX1 expression in the public human datasets GSE38974 and GSE47460 and murine smoke-exposure datasets GSE125521 and GSE52509. For panels **(A–H)**, data are presented as mean ± SEM. Two-group comparisons were performed using two-sided unpaired Student's t tests, separately for each gene or protein and, where applicable, each cell model and indicated time point. P values were unadjusted, with $P < 0.05$ indicating nominal statistical significance. Significance symbols: *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001; ns, not significant.

The docking results additionally prioritized HMOX1 as a candidate protein for investigating potential PFAS interactions. Direct binding and effects on HMOX1 enzymatic activity or downstream COPD-related processes remain to be established experimentally. These findings suggest a candidate structural interaction that warrants further biochemical validation, but they do not establish that PFAS modulate *HMOX1* enzymatic activity or downstream COPD-related biological processes. *JUNB*, a member of the AP-1 transcription factor family, serves as a critical regulator of stress and inflammatory responses (63). AP-1-driven chronic inflammation and tissue remodeling constitute central features of COPD pathology (64). Studies indicate that AP-1 factors (e.g., c-Jun) are essential for maintaining airway epithelial homeostasis; their deficiency accelerates emphysema development (65). After BH correction, JUNB expression remained positively associated with estimated naive B-cell and follicular helper T-cell fractions and negatively associated with M2 macrophage fractions, whereas its association with plasma-cell fractions did not meet the adjusted significance threshold. These findings place JUNB within COPD-associated immune expression patterns but do not establish that it regulates immune-cell recruitment or behavior. Given that PFAS are established activators of upstream kinases in the AP-1 pathway (66), PFAS-related upstream signaling may be associated with *JUNB*-mediated inflammatory pathways, although direct regulation of *JUNB* by PFAS exposure remains to be experimentally confirmed. *CXCL12*, a critical chemokine signaling molecule via CXCR4, has a context-dependent duality in COPD: physiological levels facilitate tissue repair, whereas excessive expression exacerbates inflammation and remodeling (67, 68). Experimental models have confirmed that dysregulated *CXCL12*/CXCR4 signaling drives emphysematous changes, whereas CXCR4 blockade mitigates lung damage (69, 70).

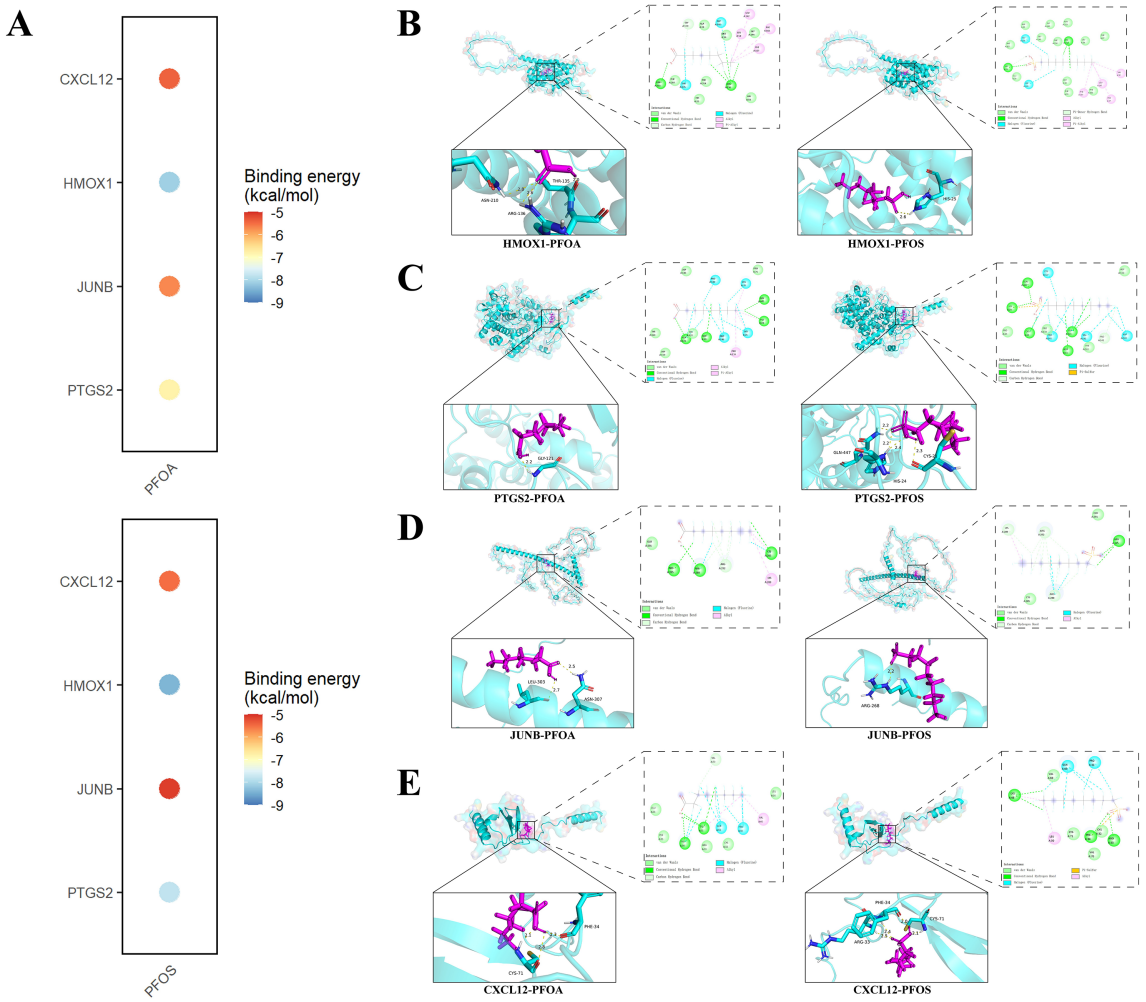

FIGURE 12
Predicted docking poses of PFAS ligands with candidate proteins. **(A)** Predicted docking scores for PFOA and PFOS with HMOX1, JUNB, CXCL12, and PTGS2. Scores are expressed in kcal mol^{-1} and do not represent experimentally measured binding free energies. **(B–E)** Representative docking poses and interaction diagrams for HMOX1 **(B)**, PTGS2 **(C)**, JUNB **(D)**, and CXCL12 **(E)** with PFOA and PFOS, illustrating possible ligand accommodation and putative protein–ligand contacts.

CXCL12 expression showed a positive association with the CIBERSORT-estimated M1 macrophage fraction that remained significant after BH correction (rho = 0.634, BH-adjusted P = 5.14 × 10 −8). This association supports a potential relationship between CXCL12 expression and an inferred macrophage-related signature in COPD tissue, but it does not demonstrate macrophage polarization, cellular recruitment, or direct regulation by PFAS. *PTGS2* (COX-2), the rate-limiting enzyme in prostaglandin synthesis, is markedly induced in COPD lung tissue (71). Elevated levels of its catalytic product PGE_2 promote inflammation and fibroblast senescence, thereby accelerating disease progression (72). Machine-learning analysis designated *PTGS2* as a candidate biomarker, and molecular docking predicted potential PFAS–*PTGS2* binding poses. However, whether PFAS alter *PTGS2* enzymatic activity, prostaglandin production, or inflammatory signaling remains unknown and requires direct biochemical and cellular validation. The plasma ELISA findings require consideration of the predominantly intracellular localization of HMOX1, JUNB, and PTGS2. Although the manufacturer's instructions support the use of the selected assays with human plasma, this analytical application does not establish the biological origin or activity of circulating analytes. Extracellular detection of an intracellular protein is biologically plausible; for example, plasma HO-1 has been measured in experimental and human acute kidney injury (73). Cellular injury or turnover and extracellular-vesicle-associated release may contribute to circulating immunoreactivity, but these possibilities were not investigated here. The plasma findings should therefore be interpreted as supportive observations in clinical samples without inferring pulmonary protein expression or a specific release mechanism.

Several limitations should be considered. Although independent public datasets and local clinical samples supported the COPD relevance of the candidate targets, the findings remain associative and hypothesis-generating. Individual-level PFAS exposure measurements and a prospective exposure-based external cohort were unavailable, and experimental validation examined COPD-related conditions without controlled PFOA or PFOS exposure. Thus, the observed expression changes cannot be attributed directly to PFAS, and docking predictions do not resolve this evidential gap. Larger prospective cohorts with measured PFAS exposure across diverse populations, together with controlled mechanistic studies, are needed to assess the proposed molecular links and their generalizability.

10.3389/fimmu.2026.1965629

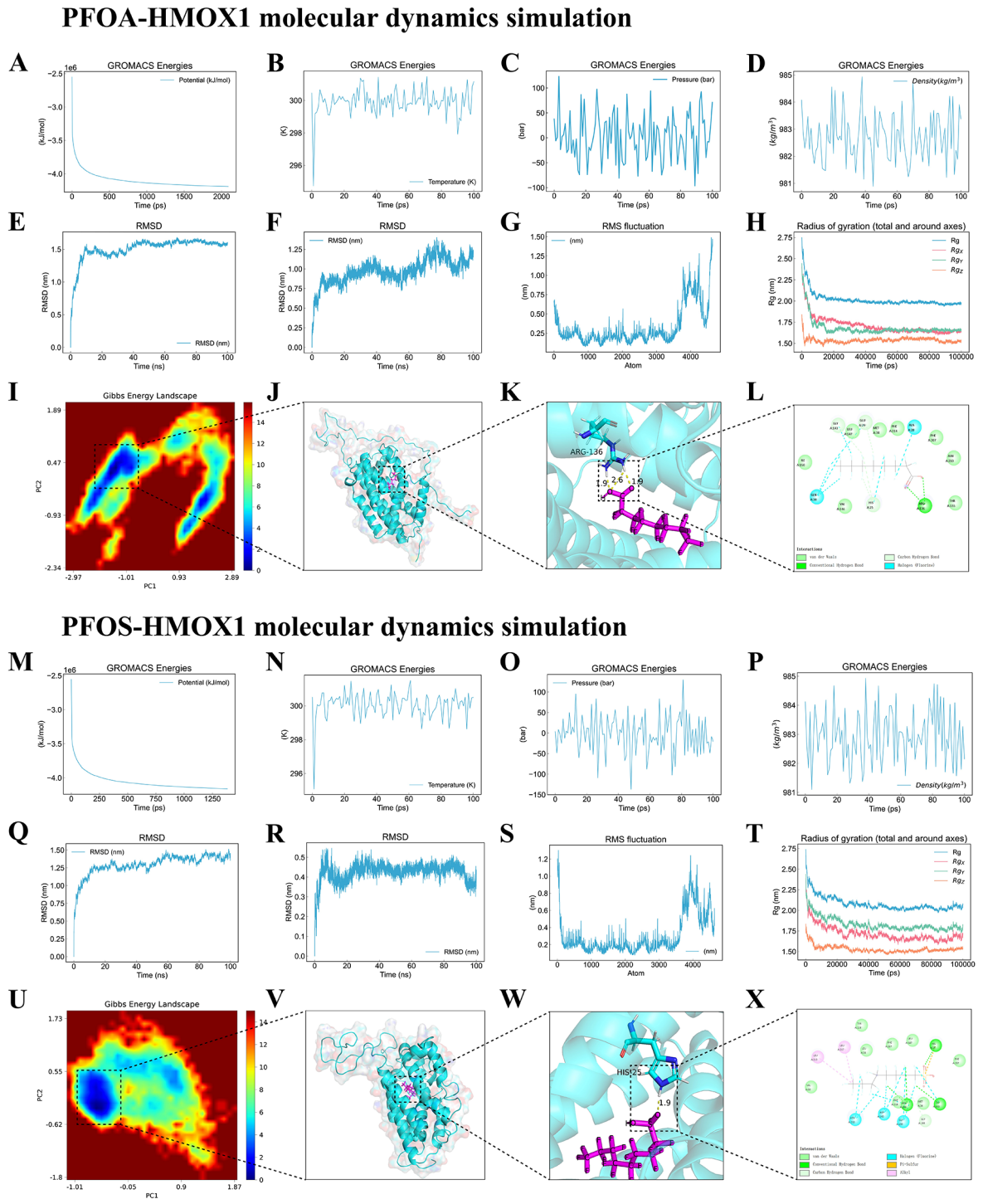

FIGURE 13
Exploratory molecular dynamics analysis of docking-derived PFOA–HMOX1 and PFOS–HMOX1 structures. PFOA–HMOX1 and PFOS–HMOX1 are shown in **(A-L, M-X)**, respectively. One 100-ns production trajectory was generated for each structure. **(A–D, M–P)** Potential-energy and equilibration diagnostics from system preparation. **(E, F, Q, R)** RMSD profiles describing deviations and fluctuations relative to reference structures. **(G, S)** Atomic RMSF profiles. **(H, T)** Radius-of-gyration profiles showing initial compaction followed by comparatively maintained global compactness. **(I, U)** PCA-based conformational free-energy landscapes describing sampled configurations; these are not protein–ligand binding free-energy estimates. **(J–L, V–X)** Selected structural views and two-dimensional interaction diagrams illustrating possible contacts in individual configurations. These exploratory analyses characterize conformational changes within the sampled trajectories and do not establish trajectory convergence or persistent ligand binding.

Second, PFAS-related target prediction has inherent computational limitations. SwissTargetPrediction is a general ligand-based target-prediction platform and has not been specifically validated for PFAS compounds. Although we supplemented these predictions using CTD and literature-derived evidence, the predicted PFAS-related targets should still be interpreted cautiously as candidate computational targets requiring experimental confirmation. The absence of dedicated doublet detection and sensitivity analysis is a

methodological limitation, and the potential influence of residual doublets on cellular interpretation remains uncertain. Future studies should incorporate dedicated doublet identification and removal, together with sensitivity analyses assessing their impact.

Third, the docking and MD analyses were exploratory and have several limitations. Only one 100-ns production trajectory was generated for each complex, limiting assessment of reproducibility and sampling uncertainty. The substantial RMSD changes, despite comparatively steady later Rg values, do not establish preservation of the starting conformation or persistent ligand binding; apparent stabilization of an individual structural metric is insufficient to demonstrate adequate sampling (74). MM/GBSA or MM/PBSA estimates, trajectory-wide contact occupancies, and per-residue energy decomposition were not evaluated. In addition, the ligand parametrization was not benchmarked specifically for PFAS–protein interactions. Research on perfluoroalkanes highlights the importance of nonbonded parameter and partial-charge choices, supporting further evaluation of their applicability to the present systems (75). The causes of the observed RMSD changes were not determined. Limited computational resources precluded additional independent simulations in this revision. Future work should incorporate independent replicates, more extensive sampling, ligand-focused structural and energetic analyses, and assessment of parameter sensitivity and relevant solvent, ionic, and protonation conditions. Experimental binding and functional assays will also be needed to test the proposed interactions. Accordingly, the present calculations support candidate prioritization and structural hypothesis generation rather than confirmed binding or biological regulation.

In addition, the clinical findings remain exploratory given the small cohorts, incomplete covariate information, and limited analytical validation. Although the kits are specified for plasma, study-specific recovery, dilution-linearity, and quantitative hemolysis-interference data were not reported; duplicates and exclusion of hemolyzed samples cannot exclude matrix effects. Baseline characteristics are provided in Supplementary Table S12, but incomplete smoking, comorbidity, and medication records prevented adjustment for these factors. The unadjusted comparisons may therefore reflect residual confounding. Future prospective studies should strengthen assay validation and collect comprehensive clinical data for covariate-adjusted analyses.

Finally, several public datasets comprised fixed archival samples, and GSE151052 included repeated lung regions from only 10 COPD and 5 control lungs. Within-donor dependence was not modeled, and internal machine-learning resampling was not grouped by donor. Unequal regional sampling and donor-specific expression patterns may therefore influence differential expression, co-expression modules, immune associations, and model performance, limiting the strength of the evidence for candidate target prioritization. Future studies should include more independent donors, account for within-donor dependence, and perform model resampling at the donor level to improve the robustness and generalizability of the findings.

5 Conclusion

This study integrates PFAS-associated target information, COPD transcriptomic profiles, machine-learning prioritization, and computational toxicology to identify candidate molecular links between PFAS-associated profiles and COPD-related biological alterations. HMOX1, JUNB, CXCL12, and PTGS2 emerged as candidate genes with COPD-related expression support and associations with inferred immune-inflammatory signatures. Docking and molecular dynamics generated structural hypotheses, particularly for PFOA/PFOS–HMOX1 interactions. The findings do not establish direct PFAS-mediated gene regulation, experimentally confirmed binding, exposure causality, or clinical diagnostic utility. They provide an exploratory basis for future exposure-based and mechanistic studies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the First Affiliated Hospital of Guangzhou Medical University (Approval No. 2018-53) and the Ethics Review Committee of the Medical Research at the General Hospital of Ningxia Medical University (Approval No. KYLL-2025-1078). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

GY: Formal Analysis, Project administration, Writing – review & editing, Software. ZG: Software, Project administration, Conceptualization, Writing – review & editing. XC: Formal Analysis, Writing – review & editing, Project administration, Resources. HX: Writing – review & editing, Formal Analysis, Software, Investigation. PR: Writing – review & editing, Investigation, Methodology, Conceptualization. ZH: Data curation, Formal Analysis, Writing – review & editing, Writing – original draft. JiC: Conceptualization, Writing – review & editing, Supervision, Formal Analysis, Methodology. XW: Resources, Writing – review & editing, Supervision, Funding acquisition. JuC: Formal Analysis, Data curation, Funding acquisition, Writing – review & editing, Conceptualization. YY: Software, Data curation, Writing – original draft, Formal Analysis, Supervision. JY: Formal Analysis, Writing – review & editing, Data curation, Methodology, Project administration. ZM: Project administration, Methodology, Validation, Writing – review & editing, Data curation. MN: Writing – review & editing, Project administration, Methodology, Supervision, Data curation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was funded by Key Research and Development Project of Ningxia (No. 2025BEG02002), and National Science and Technology Major Project of China (No. 2024ZD0528905), and Ningxia Science and Technology Benefits for the People Project (No. 2025CMG03002), and Ningxia Autonomous Region Scientific and Technological Infrastructure Construction Project (No. 2025DPC05027).

Acknowledgments

First and foremost, we sincerely thank the BioRender team for providing high-quality scientific illustration resources and an online graphic design platform, which helped make the illustrations in this article clearer and more vivid. The attribution for Figure 1 is as follows: Created in BioRender. Liu, X. (2026) https://BioRender.com/87ttgq4. Secondly, we extend our heartfelt gratitude to all researchers who uploaded lung tissue transcriptome sequencing data and single-cell sequencing data to the Gene Expression Omnibus database. It is precisely because of their open and shared data that this study was able to conduct in-depth and systematic multi-omics analyses. Additionally, we are especially grateful to Xiyou Cloud Computing (https://www.xiyoucloud.net/) for providing high-performance online servers and technical support. Their outstanding computational capabilities significantly improved the efficiency of this study in large-scale data processing, computational modeling, and molecular dynamics simulations. Once again, we express our sincere appreciation to all institutions, researchers, and technical teams who contributed their hard work and exceptional efforts to this study.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2026.1965629/full#supplementary-material

References

1. Glüge J, Scheringer M, Cousins IT, DeWitt JC, Goldenman G, Herzke D, et al. An overview of the uses of per- and polyfluoroalkyl substances (Pfas). *Environ Sci Process Impacts*. (2020) 22:2345–73. doi: 10.1039/d0em00291g

2. Bayode AA, Emmanuel SS, Akinyemi AO, Ore OT, Akpotu SO, Koko DT, et al. Innovative techniques for combating a common enemy forever chemicals: A comprehensive approach to mitigating per- and polyfluoroalkyl substances (Pfas) contamination. *Environ Res*. (2024) 261:119719. doi: 10.1016/j.envres.2024.119719

3. Skinner JP, Raderstorf A, Rittmann BE, Delgado AG. Biotransforming the "Forever chemicals": Trends and insights from microbiological studies on Pfas. *Environ Sci Technol*. (2025) 59:5417–5430. doi: 10.1021/acs.est.4c04557

4. Miner KR, Clifford H, Taruscio T, Potocki M, Solomon G, Ritari M, et al. Deposition of pfas 'Forever chemicals' on mt. Everest. *Sci Total Environ*. (2021) 759:144421. doi: 10.1016/j.scitotenv.2020.144421

5. Roth K, Imran Z, Liu W, Petriello MC. Diet as an exposure source and mediator of per- and polyfluoroalkyl substance (Pfas) toxicity. *Front Toxicol*. (2020) 2:601149. doi: 10.3389/ftox.2020.601149

6. Gaillard L, Barouki R, Blanc E, Coumoul X, Andréau K. Per- and polyfluoroalkyl substances as persistent pollutants with metabolic and endocrine-disrupting impacts. *Trends Endocrinol Metab*. (2024) 36:249–261. doi: 10.1016/j.tem.2024.07.021

7. Bline AP, DeWitt JC, Kwiatkowski CF, Pelch KE, Reade A, Varshavsky JR. Public health risks of Pfas-related immunotoxicity are real. *Curr Environ Health Rep*. (2024) 11:118–27. doi: 10.1007/s40572-024-00441-y

8. Costello E, Rock S, Stratakis N, Eckel SP, Walker DI, Valvi D, et al. Exposure to per- and polyfluoroalkyl substances and markers of liver injury: A systematic review and meta-analysis. *Environ Health Perspect*. (2022) 130:46001. doi: 10.1289/ehp10092

9. Labaki WW, Rosenberg SR. Chronic obstructive pulmonary disease. *Ann Intern Med*. (2020) 173:Itc17–itc32. doi: 10.7326/aitc202008040

10. GBD 2019 Chronic Respiratory Diseases Collaborators. Global burden of chronic respiratory diseases and risk factors, 1990-2019: An update from the Global Burden of Disease Study 2019. *EClinicalMedicine*. (2023) 59:101936. doi: 10.1016/j.eclinm.2023.101936

11. Yang X, Zhang T, Zhang Y, Chen H, Sang S. Global burden of Copd attributable to ambient Pm2.5 in 204 countries and territories, 1990 to 2019: A systematic analysis for the Global Burden of Disease Study 2019. *Sci Total Environ*. (2021) 796:148819. doi: 10.1016/j.scitotenv.2021.148819

12. Capistrano SJ, van Reyk D, Chen H, Oliver BG. Evidence of biomass smoke exposure as a causative factor for the development of Copd. *Toxics*. (2017) 5:36. doi: 10.3390/toxics5040036

13. Dutta J, Singh S, Greeshma MV, Mahesh PA, Mabalirajan U. Diagnostic challenges and pathogenetic differences in biomass-smoke-induced versus tobacco-smoke-induced Copd: A comparative review. *Diagnostics (Basel)*. (2024) 14:2154. doi: 10.3390/diagnostics14192154

14. Wang Y, Zhang J, Zhang J, Hou M, Kong L, Lin X, et al. Association between per- and polyfluoroalkyl substances exposure and prevalence of chronic obstructive

pulmonary disease: The mediating role of serum albumin. *Sci Total Environ*. (2024) 925:171742. doi: 10.1016/j.scitotenv.2024.171742

15. Pan M, Zou Y, Wei G, Zhang C, Zhang K, Guo H, et al. Moderate-intensity physical activity reduces the role of serum Pfas on Copd: A cross-sectional analysis with Nhanes data. *PloS One*. (2024) 19:e0308148. doi: 10.1371/journal.pone.0308148

16. Aker A, Courtemanche Y, Ayotte P, Robert P, Gaudreau É, Lemire M. Per and poly-fluoroalkyl substances and respiratory health in an Inuit community. *Environ Health*. (2024) 23:83. doi: 10.1186/s12940-024-01126-7

17. Shi S, Ding Y, Wu B, Hu P, Chen M, Dong N, et al. Association of perfluoroalkyl substances with pulmonary function in adolescents (Nhanes 2007-2012). *Environ Sci pollut Res Int*. (2023) 30:53948–61. doi: 10.1007/s11356-023-26119-w

18. Qin XD, Qian ZM, Dharmage SC, Perret J, Geiger SD, Rigdon SE, et al. Association of perfluoroalkyl substances exposure with impaired lung function in children. *Environ Res*. (2017) 155:15–21. doi: 10.1016/j.envres.2017.01.025

19. Jabeen M, Fayyaz M, Irudayaraj J. Epigenetic modifications, and alterations in cell cycle and apoptosis pathway in A549 lung carcinoma cell line upon exposure to perfluoroalkyl substances. *Toxics*. (2020) 8:112. doi: 10.3390/toxics8040112

20. Hailemariam A, Upadhyay S, Srivastava V, Hafiz Z, Zhang L, Tsui WNT, et al. Perfluorooctane sulfonate (Pfos) and related compounds induce nuclear receptor 4a1 (Nr4a1)-dependent carcinogenesis. *Chem Res Toxicol*. (2025) 38:705–16. doi: 10.1021/acs.chemrestox.4c00528

21. Dragon J, Hoaglund M, Badireddy AR, Nielsen G, Schlezinger J, Shukla A. Perfluoroalkyl substances (Pfas) affect inflammation in lung cells and tissues. *Int J Mol Sci*. (2023) 24:8539. doi: 10.3390/ijms24108539

22. Sherpa D, Bhowmick C, Pavan T, Rajwade DA, Halder S, Mitra I, et al. Classification of idiopathic recurrent spontaneous miscarriage using Ftir and Raman spectroscopic fusion technology. *Syst Biol Reprod Med*. (2024) 70:228–39. doi: 10.1080/19396368.2024.2384386

23. Yifan C, Jianfeng S, Jun P. Development and validation of a random forest diagnostic model of acute myocardial infarction based on ferroptosis-related genes in circulating endothelial cells. *Front Cardiovasc Med*. (2021) 8:663509. doi: 10.3389/fcvm.2021.663509

24. Dasgupta S. Identification and molecular modelling of potential drugs targeting the genes involved in the progression of lung cancer in patients with idiopathic pulmonary fibrosis. *Gene Rep*. (2024) 37:102067. doi: 10.1016/j.genrep.2024.102067

25. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. Pubchem 2025 update. *Nucleic Acids Res*. (2025) 53:D1516–d25. doi: 10.1093/nar/gkae1059

26. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. Swisstargetprediction: A web server for target prediction of bioactive small molecules. *Nucleic Acids Res*. (2014) 42:W32–8. doi: 10.1093/nar/gku293

27. Hong Y, Wang D, Liu Z, Chen Y, Wang Y, Li J. Decoding per- and polyfluoroalkyl substances (Pfas) in hepatocellular carcinoma: A multi-omics and computational toxicology approach. *J Transl Med*. (2025) 23:504. doi: 10.1186/s12967-025-06517-z

28. Davis AP, Murphy CG, Saraceni-Richards CA, Rosenstein MC, Wiegers TC, Mattingly CJ. Comparative toxicogenomics database: A knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res*. (2009) 37:D786–92. doi: 10.1093/nar/gkn580

29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for Rna-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007

30. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. (2004) 3:Article3. doi: 10.2202/1544-6115.1027

31. Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. Kegg: Biological systems database as a model of the real world. *Nucleic Acids Res*. (2025) 53:D672–D7. doi: 10.1093/nar/gkae909

32. Langfelder P, Horvath S. Wgcna: An R package for weighted correlation network analysis. *BMC Bioinf*. (2008) 9:559. doi: 10.1186/1471-2105-9-559

33. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. String V11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. (2019) 47:D607–d13. doi: 10.1093/nar/gky1131

34. Simha NA, Patil SM, Chagalamari A, Satish AM, Ramu R. Protocol to identify multiple protein targets and therapeutic compounds using an in silico polypharmacological approach. *STAR Protoc*. (2023) 4:102440. doi: 10.1016/j.xpro.2023.102440

35. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. Cytohubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. (2014) 8 Suppl 4:S11. doi: 10.1186/1752-0509-8-s4-s11

36. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. Clusterprofiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)*. (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

37. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. (2016) 32:2847–9. doi: 10.1093/bioinformatics/btw313

38. Conway JR, Lex A, Gehlenborg N. Upsetr: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*. (2017) 33:2938–40. doi: 10.1093/bioinformatics/btx364

39. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. (2015) 12:453–7. doi: 10.1038/nmeth.3337

40. Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. *Anesth Analg*. (2018) 126:1763–8. doi: 10.1213/ane.0000000000002864

41. Wilkerson MD, Hayes DN. Consensusclusterplus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics*. (2010) 26:1572–3. doi: 10.1093/bioinformatics/btq170

42. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

43. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. (2015) 33:495–502. doi: 10.1038/nbt.3192

44. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: Pre-processing, quality control, normalization and visualization of single-cell Rna-seq data in R. *Bioinformatics*. (2017) 33:1179–86. doi: 10.1093/bioinformatics/btw777

45. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. (2019) 16:1289–96. doi: 10.1038/s41592-019-0619-0

46. Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R, et al. Compartments: Unification and visualization of protein subcellular localization evidence. *Database (Oxford)*. (2014) 2014:bau012. doi: 10.1093/database/bau012

47. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using Cellchat. *Nat Commun*. (2021) 12:1088. doi: 10.1038/s41467-021-21246-9

48. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with Alphafold. *Nature*. (2021) 596:583–9. doi: 10.1038/s41586-021-03819-2

49. Trott O, Olson AJ. Autodock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. (2010) 31:455–61. doi: 10.1002/jcc.21334

50. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. Autodock4 and Autodocktools4: Automated docking with selective receptor flexibility. *J Comput Chem*. (2009) 30:2785–91. doi: 10.1002/jcc.21256

51. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. Gromacs: fast, flexible, and free. *J Comput Chem*. (2005) 26:1701–18. doi: 10.1002/jcc.20291

52. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber Ff99sb protein force field. *Proteins*. (2010) 78:1950–8. doi: 10.1002/prot.22711

53. Wang Z, DeWitt JC, Higgins CP, Cousins IT. A never-ending story of per- and polyfluoroalkyl substances (Pfass)? *Environ Sci Technol*. (2017) 51:2508–18. doi: 10.1021/acs.est.6b04806

54. Evich MG, Davis MJB, McCord JP, Acrey B, Awkerman JA, Knappe DRU, et al. Per- and polyfluoroalkyl substances in the environment. *Science*. (2022) 375:eabg9065. doi: 10.1126/science.abg9065

55. Sunderland EM, Hu XC, Dassuncao C, Tokranov AK, Wagner CC, Allen JG. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (Pfass) and present understanding of health effects. *J Expo Sci Environ Epidemiol*. (2019) 29:131–47. doi: 10.1038/s41370-018-0094-1

56. Sørli JB, Låg M, Ekeren L, Perez-Gil J, Haug LS, Da Silva E, et al. Per- and polyfluoroalkyl substances (Pfass) modify lung surfactant function and pro-inflammatory responses in human bronchial epithelial cells. *Toxicol Vitro*. (2020) 62:104656. doi: 10.1016/j.tiv.2019.104656

57. Sauler M, McDonough JE, Adams TS, Kothapalli N, Barnthaler T, Werder RB, et al. Characterization of the Copd alveolar niche using single-cell Rna sequencing. *Nat Commun*. (2022) 13:494. doi: 10.1038/s41467-022-28062-9

58. Rowan-Carroll A, Reardon A, Leingartner K, Gagné R, Williams A, Meier MJ, et al. High-throughput transcriptomic analysis of human primary hepatocyte spheroids exposed to per- and polyfluoroalkyl substances as a platform for relative potency characterization. *Toxicol Sci*. (2021) 181:199–214. doi: 10.1093/toxsci/kfab039

59. Consoli V, Sorrenti V, Grosso S, Vanella L. Heme oxygenase-1 signaling and redox homeostasis in physiopathological conditions. *Biomolecules*. (2021) 11:589. doi: 10.3390/biom11040589

60. Maestrelli P, El Messlemani AH, De Fina O, Nowicki Y, Saetta M, Mapp C, et al. Increased expression of heme oxygenase (Ho)-1 in alveolar spaces and Ho-2 in alveolar walls of smokers. *Am J Respir Crit Care Med*. (2001) 164:1508–13. doi: 10.1164/ajrccm.164.8.2011083

61. Fredenburgh LE, Perrella MA, Mitsialis SA. The role of heme oxygenase-1 in pulmonary disease. *Am J Respir Cell Mol Biol*. (2007) 36:158–65. doi: 10.1165/rcmb.2006-0331TR

62. Zhou H, Ying X, Liu Y, Ye S, Yan J, Li Y. Genetic polymorphism of heme oxygenase 1 promoter in the occurrence and severity of chronic obstructive pulmonary disease: a meta-analysis. *J Cell Mol Med*. (2017) 21:894–903. doi: 10.1111/jcmm.13028

63. Hsieh T, Sasaki D, Taira N, Chien H, Sarkar S, Seto Y, et al. Junb is critical for survival of T helper cells. *Front Immunol*. (2022) 13:901030. doi: 10.3389/fimmu.2022.901030

64. Yu Y, Yang A, He X, Wu B, Wu Y, Li Y, et al. Soluble epoxide hydrolase deficiency attenuates airway inflammation in Copd via Ire1α/Jnk/Ap-1 signaling pathway. *J Inflammation (Lond)*. (2023) 20:36. doi: 10.1186/s12950-023-00361-y

65. Reddy NM, Vegiraju S, Irving A, Paun BC, Luzina IG, Atamas SP, et al. Targeted deletion of Jun/Ap-1 in alveolar epithelial cells causes progressive emphysema and worsens cigarette smoke-induced lung inflammation. *Am J Pathol*. (2012) 180:562–74. doi: 10.1016/j.ajpath.2011.10.029

66. Ehrlich V, Bil W, Vandebriel R, Granum B, Luijten M, Lindeman B, et al. Consideration of pathways for immunotoxicity of per- and polyfluoroalkyl substances (Pfas). *Environ Health*. (2023) 22:19. doi: 10.1186/s12940-022-00958-5

67. Shen W, Weng Z, Fan M, Wang S, Wang R, Zhang Y, et al. Mechanisms by which the Mbd2/Mir-301a-5p/Cxcl12/Cxcr4 pathway regulates acute exacerbations of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. (2020) 15:2561–72. doi: 10.2147/copd.S261522

68. Dupin I, Henrot P, Maurat E, Abohalaka R, Chaigne S, El Hamrani D, et al. Cxcr4 blockade alleviates pulmonary and cardiac outcomes in early Copd. *Am J Respir Cell Mol Biol*. (2025) 73:530–544. doi: 10.1165/rcmb.2024-0303OC

69. Chen X, Chen L, Chen G, Lv J, Wang J, Yu W, et al. Interleukin-17a promotes airway remodeling in chronic obstructive pulmonary disease by activating C-X-C motif chemokine ligand 12 secreted by lung fibroblasts. *Chronic Obstr Pulm Dis*. (2024) 11:482–95. doi: 10.15326/jcopdf.2024.0495

70. Barwinska D, Oueini H, Poirier C, Albrecht ME, Bogatcheva NV, Justice MJ, et al. Amd3100 ameliorates cigarette smoke-induced emphysema-like manifestations in mice. *Am J Physiol Lung Cell Mol Physiol*. (2018) 315:L382–6. doi: 10.1152/ajplung.00185.2018

71. Roca-Ferrer J, Pujols L, Agusti C, Xaubet A, Mullol J, Gimferrer JM, et al. Cyclooxigenase-2 levels are increased in the lung tissue and bronchial tumors of patients with chronic obstructive pulmonary disease. *Arch Bronconeumol*. (2011) 47:584–9. doi: 10.1016/j.arbres.2011.05.015

72. Dagouassat M, Gagliolo JM, Chrusciel S, Bourin MC, Duprez C, Caramelle P, et al. The cyclooxygenase-2-prostaglandin E2 pathway maintains senescence of chronic obstructive pulmonary disease fibroblasts. *Am J Respir Crit Care Med*. (2013) 187:703–14. doi: 10.1164/rccm.201208-1361OC

73. Zager RA, Johnson AC, Becker K. Plasma and urinary heme oxygenase-1 in Aki. *J Am Soc Nephrol*. (2012) 23:1048–57. doi: 10.1681/asn.2011121147

74. Grossfield A, Patrone PN, Roe DR, Schultz AJ, Siderius DW, Zuckerman DM. Best practices for quantification of uncertainty and sampling quality in molecular simulations [Article V1.0]. *Living J Comput Mol Sci*. (2018) 1:5067. doi: 10.33011/livecoms.1.1.5067

75. Nikitin A, Milchevskiy Y, Lyubartsev A. Amber-Ii: new combining rules and force field for perfluoroalkanes. *J Phys Chem B*. (2015) 119:14563–73. doi: 10.1021/acs.jpcb.5b07233

Glossary

AUC	Area Under the Curve
BP	Biological Process
CC	Cellular Component
COPD	Chronic Obstructive Pulmonary Disease
CTD	Comparative Toxicogenomics Database
DEGs	Differentially Expressed Genes
DNA	Deoxyribonucleic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
EPC	Edge Percolated Component
FEV1	Forced Expiratory Volume in 1 second
FVC	Forced Vital Capacity
GBM	Gradient Boosting Machine
GO	Gene Ontology
HO-1	heme oxygenase-1
Human M-CSF	Human Macrophage Colony-Stimulating Factor
KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
Limma	Linear Models for Microarray Data
MD	Molecular Dynamics
MF	Molecular Function
MNC	Maximum Neighborhood Component
NHANES	National Health and Nutrition Examination Survey
PBMCs	peripheral-blood mononuclear cells
PCA	Principal Component Analysis
PFAS	Per- and Polyfluoroalkyl Substances
PFOA	Perfluorooctanoic Acid
PFHxS	Perfluorohexanesulfonic Acid
PFNA	Perfluorononanoic Acid
PFOS	Perfluorooctanesulfonic Acid
plsRglm	Partial Least-Squares Generalized Linear Model
PPI	Protein–Protein Interaction
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
RF	Random Forest
Rg	Radius of Gyration
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
scRNA-seq	Single-Cell RNA Sequencing
SEM	Standard Error of the Mean
SPSS	Statistical Package for the Social Sciences
ssGSEA	Single-Sample Gene Set Enrichment Analysis
SVM	Support Vector Machine
WGCNA	Weighted Gene Co-Expression Network Analysis
XGBoost	Extreme Gradient Boosting