scientific reports

OPEN

Multi-modal transcriptomic analysis reveals metabolic dysregulation and immune responses in chronic obstructive pulmonary disease

Xiufang Luo[1,9], Wei Zeng[2,9], Jingyi Tang[3,9], Wang Liu[4,9], Jinyan Yang[5], Haiqing Chen[3], Lai Jiang[3], Xuancheng Zhou[3], Jinbang Huang[3], Shengke Zhang[3], Linjuan Du[6], Xiang Shen[7✉], Hao Chi[3✉] & Huachuan Wang[8✉]

Chronic obstructive pulmonary disease (COPD), a progressive inflammatory condition of the airways, emerges from the complex interplay between genetic predisposition and environmental factors. Notably, its incidence is on the rise, particularly among the elderly demographic. Current research increasingly highlights cellular senescence as a key driver in chronic lung pathologies. Despite this, the detailed mechanisms linking COPD with senescent genomic alterations remain elusive. To address this gap, there is a pressing need for comprehensive bioinformatics methodologies that can elucidate the molecular intricacies of this link. This approach is crucial for advancing our understanding of COPD and its association with cellular aging processes. Utilizing a spectrum of advanced bioinformatics techniques, this research delved into the potential mechanisms linking COPD with aging-related genes, identifying four key genes (EP300, MTOR, NFE2L1, TXN) through machine learning and weighted gene co-expression network analysis (WGCNA) analyses. Subsequently, a precise diagnostic model leveraging an artificial neural network was developed. The study further employed single-cell analysis and molecular docking to investigate senescence-related cell types in COPD tissues, particularly focusing on the interactions between COPD and NFE2L1, thereby enhancing the understanding of COPD's molecular underpinnings. Leveraging artificial neural networks, we developed a robust classification model centered on four genes—EP300, MTOR, NFE2L1, TXN— exhibiting significant predictive capability for COPD and offering novel avenues for its early diagnosis. Furthermore, employing various single-cell analysis techniques, the study intricately unraveled the characteristics of senescence-related cell types in COPD tissues, enriching our understanding of the disease's cellular landscape. This research anticipates offering novel biomarkers and therapeutic targets for early COPD intervention, potentially alleviating the disease's impact on individuals and healthcare systems, and contributing to a reduction in global COPD-related mortality. These findings carry significant clinical and public health ramifications, bolstering the foundation for future research and clinical strategies in managing and understanding COPD.

[1]Geriatric Department, Dazhou Central Hospital, Dazhou 635000, China. [2]Oncology Department, Second People's Hospital of Yaan City, Yaan 625000, China. [3]Department of Clinical Medicine, Clinical Medical College, Southwest Medical University, Luzhou 646000, China. [4]Department of General Surgery, Cheng Fei Hospital, Chengdu 610000, China. [5]School of Stomatology, Southwest Medical University, Luzhou 646000, China. [6]Oncology Department, Dazhou Central Hospital, Dazhou 635000, China. [7]Department of Respiratory and Critical Care Medicine, The Affiliated Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China. [8]Department of Thoracic Surgery, Dazhou Central Hospital, Dazhou 635000, China. [9]These authors contributed equally: Xiufang Luo, Wei Zeng, Jingyi Tang and Wang Liu. ✉email: shenxiangdoctor@163.com; Chihao7511@163.com; 420743070@qq.com

COPD is an increasingly prevalent chronic inflammatory lung disorder, driven by a multifaceted interplay of genetic predispositions and environmental factors. Alarmingly, its global burden is escalating, with projections placing it as the third leading cause of mortality by 2030[1,2]. This trend signifies a substantial impact on individual health and healthcare infrastructures, as evidenced by over three million annual fatalities attributed to COPD[3]. Despite advancements in pharmaceutical technologies enhancing the efficacy of current COPD interventions—namely bronchodilators, anti-inflammatory agents, and antioxidants—in managing symptoms and curtailing acute exacerbations, these therapies have yet to significantly alter the disease's progression or reduce its mortality rate[3,4].

While prior research has predominantly concentrated on linking COPD with cigarette smoke exposure, air pollution, and genetic determinants, the role of cellular senescence in COPD pathogenesis is gaining increasing attention, offering novel perspectives[5,6]. An expanding corpus of evidence underscores the critical association of senescence with the onset and progression of chronic lung conditions. Lung senescence, characterized by heightened vulnerability to both acute and chronic pulmonary diseases, impaired respiratory function, and structural alterations in lung architecture, is becoming increasingly recognized[7,8]. Cellular senescence, in particular, has emerged as a pivotal factor in chronic lung diseases, notably in COPD and idiopathic pulmonary fibrosis[9–11]. The incidence of COPD escalates with age, supporting the hypothesis of cellular senescence as a key contributor to COPD (5.1% in patients aged 35–54 years, 13.3% in patients aged 55–64 years, and 21.7% in patients aged 65 years or older)[12]. Investigations have demonstrated compromised lung function and heightened senescence in individuals exposed to various concentrations of cigarette smoke (CS), encompassing both smokers and non-smokers with intact lung function. This is evident through the expression of senescence markers such as p21 and p16[13–16]. Nevertheless, the intricate interplay between COPD and senescence mechanisms remains to be fully elucidated[17]. Consequently, probing the potential links between senescence-related genes and COPD progression is anticipated to unveil new, disease-specific biomarkers and therapeutic targets for early diagnosis and intervention.

NFE2L1, or nuclear factor E2-related factor 1, emerges as a pivotal regulator in mitigating inflammation and oxidative stress-induced cellular senescence, as evidenced by recent studies[18–20]. Concurrently, the multifaceted bioactive compound resveratrol, a natural polyphenolic constituent, is acclaimed for its comprehensive biological functions encompassing antioxidative, anticancer, anti-obesity, anti-aging, anti-inflammatory, and immunomodulatory properties[21,22]. These attributes suggest a potential regulatory influence on the cellular aging process. Nonetheless, the specific interplay between NFE2L1 and resveratrol in the etiology of chronic obstructive pulmonary disease (COPD) remains an underexplored domain. Addressing this gap, the current study adopts molecular docking methodologies to explore the intricate molecular mechanisms underpinning the role of NFE2L1 and resveratrol in COPD pathogenesis.

This research integrates a robust transcriptome analysis framework, synthesizing the synergistic methodologies of bulk RNA-Seq analysis and single-cell RNA sequencing[23,24]. Bulk RNA-Seq, recognized for its efficiency and classical approach[25], facilitates the quantification of average gene expression by sequencing RNA from entire cell populations, encompassing diverse subpopulations[26]. This technique excels in capturing data on cell populations, indispensable for examining comprehensive biological processes, discerning disease impacts at a global level, and elucidating pivotal regulatory networks[27]. In contrast, single-cell RNA sequencing, a cutting-edge biosequencing technology, surpasses traditional bulk RNA-Seq in detecting subtle intercellular variations[28]. It is instrumental in delineating and reconstructing developmental pathways within tissues and examining intricate cell–cell interactions, thereby illuminating the heterogeneity inherent in cellular assemblies[29,30]. By amalgamating these methodologies, the study achieves a nuanced dissection of cellular populations within tissues and among individual cells, thereby unveiling a detailed transcriptomic landscape. This approach offers an invaluable tool for a holistic understanding of the dynamic nuances of gene expression processes.

With the continuous advancement of bioinformatics, an increasing number of studies are utilizing machine learning and other methods to filter prognostic biomarkers and construct disease prognosis models[31,32]. In our endeavor to elucidate the nexus between COPD and senescence-related genes, this study employed sophisticated bioinformatics approaches, integrating multiple high-throughput genomic data analysis algorithms. This included the utilization of three distinct machine learning algorithms alongside WGCNA, culminating in the identification of four pivotal senescence-related genes—EP300, MTOR, NFE2L1, TXN—implicated in the potential mechanisms underlying COPD. Additionally, we developed a diagnostic model for COPD, leveraging the construction of artificial neural networks and nomograms. Our research further delved into the intricacies of COPD at the cellular level, utilizing single-cell datasets to identify senescence-related cell types within COPD tissues. Comprehensive enrichment analyses, pseudotime analyses, and an exploration of intercellular communication within the COPD milieu were conducted, offering a nuanced understanding of the interplay between COPD and senescence-related genes.

Materials and methods
Acquisition and processing of gene expression data
In this study, bulk RNA-sequencing data from datasets GSE57148 and GSE54837, sourced from the Gene Expression Omnibus (GEO) database, were combined and carefully annotated. Both datasets were used as training sets to enable robust comparative analysis. Identical probes were merged by calculating mean values, and expression profiles were standardized to Transcripts Per Million (TPM). Gene expression datasets were harmonized and corrected for batch effects using the "limma" and "sva" software packages. Additionally, a list of senescent genes was curated from Genecards. Single-cell RNA sequencing (scRNA-seq) data from COPD patients were also obtained from GEO (accession number GSE173896) to enhance our genomic analysis.

Screening of differential genes and enrichment analysis of functional pathways

In our rigorous examination of differential gene expression within the designated training set, the "Limma" package in R was adeptly deployed to identify differentially expressed genes (DEGs) showcasing significant variances in expression ($|logFC| > 0.25$, $FDR < 0.05$)[33]. The visualization of these DEGs was elegantly accomplished using the "pheatmap" and "ggplot2" packages[34], which facilitated the creation of insightful heatmaps and detailed volcano plots. Further, to determine the enrichment of these DEGs across various functional pathways, the "clusterProfiler" package was employed for a comprehensive gene set enrichment analysis (GSEA)[35,36]. The resultant data was meticulously analyzed to select the top five functional pathways exhibiting the highest and lowest levels of enrichment. This strategic selection and subsequent visualization provided a thorough and nuanced depiction of both the most and least enriched functional pathways, offering a deeper understanding of the functional implications of the identified DEGs.

Construction of gene co-expression network

In our quest to pinpoint co-expressed genes related with COPD, a comprehensive WGCNA was executed on the merged dataset. This methodology, pivotal in gene expression studies, was facilitated by the adept application of the R package "WGCNA"[37], enabling the intricate construction and modularization of gene co-expression networks. The initial phase of the analysis entailed sample clustering to identify and subsequently eliminate potential outliers, a process dictated by a pre-defined threshold and focused on removing samples with pronounced anomalies. Following this, an advanced automated network construction approach was adopted to establish the co-expression network. The network's complexity was unraveled through hierarchical clustering integrated with dynamic tree cutting, which efficiently identified distinct modules within the network. The relevance of these modules to COPD was meticulously determined by analyzing module membership (MM) and gene significance (GS). This analytical process culminated in the selection of the most pertinent modules in relation to the patient cohort, thereby isolating a targeted set of disease-related co-expressed genes for subsequent in-depth exploration.

Machine learning screening of diagnostic biomarkers

In this study, genes related with the senescence process were first delineated using Venn diagrams to establish a preliminary set. The refinement of this set to identify diagnostic biomarkers for COPD was achieved through the application of three sophisticated machine learning algorithms: SVM-RFE, Random Forest (RF), and LASSO logistic regression. The SVM-RFE algorithm, facilitated by R packages "caret" and "e1071"[37,38], was instrumental in methodically training and evaluating various subsets of features, with the ultimate goal of isolating the subset exhibiting the lowest cross-validation error rate. This subset was then earmarked as indicative of potential diagnostic biomarkers, with the minimal gene count identified by SVM-RFE representing the core of these biomarkers. Concurrently, the RF algorithm, utilizing the "randomForest" R package[39], played a crucial role in ascertaining the relative importance of candidate genes, with genes scoring a relative importance above 1.25 being categorized as diagnostic biomarkers. Complementing these methods, LASSO logistic regression, executed via the "glmnet" R package[40], incorporated a tenfold cross-validation to meticulously optimize the penalty parameter λ. This optimization was critical for the precise identification of diagnostic biomarkers. The culmination of this integrative approach was the identification of a consensus set of four disease-related feature genes, representing the intersection of the genes pinpointed by these three distinct, yet complementary, machine learning methodologies.

Construction and validation of diagnostic models

In our comprehensive analytical framework, the construction of a sophisticated nomogram model was achieved by harnessing the capabilities of the R packages "neuralnet" and "dplyr"[41]. The "neuralnet" package was instrumental in developing a neural network model, encompassing the intricacies of input, hidden, and output layers. To evaluate the performance of this model, receiver operating characteristic (ROC) curves were generated using the "PROC" package, enabling the precise calculation of the area under the curve (AUC). This assessment was not confined to the training set; the same methodology was rigorously applied to the gene expression matrix of the validation set, culminating in a robust computation of the AUC for this dataset as well, thus ensuring the model's generalizability. Additionally, leveraging the "rms" package, a nomogram was meticulously constructed to prognosticate the outcomes for COPD patients. This process involved extracting the expression levels of each characterized gene to calculate individualized scores for each patient, which were then aggregated to determine a total score. This cumulative scoring system facilitated the prediction of disease risk. The predictive accuracy of the model was thoroughly evaluated using calibration curves, and its clinical applicability was further elucidated through decision curve analysis (DCA), providing a holistic view of the model's efficacy in a clinical setting.

GSEA functional enrichment analysis

The cornerstone of our genomic investigation was the acquisition of essential data from the gene set enrichment analysis (GSEA) website (https://www.gsea-msigdb.org/gsea/index.jsp). In the quest to deeply understand and contrast the expression profiles related to various biological functions in our study cohorts, which included both disease and control groups, we utilized the powerful capabilities of R software packages "GSVA" and "GSEA-Base"[42]. This strategic approach allowed us to conduct a nuanced and detailed analysis, enabling us to distinguish and compare the differential expression profiles across these cohorts, thereby providing valuable insights into the distinct genomic landscapes characterizing the disease and control groups within the context of our investigation.

Immune infiltration analysis

CIBERSORT, an algorithm renowned for its efficacy in estimating the composition of immune cells based on gene expression matrices, was employed in our study to quantify the presence of 22 unique immune cell infiltrates within cohorts comprising both COPD patients and healthy individuals. This analytical approach enabled the precise calculation of immune cell proportions, effectively portraying the varied immune cell landscape across different sample subtypes. The resultant data were visually articulated through the use of bar charts, offering an intuitive representation of the immune cell composition. Furthermore, to explore the intricate interrelationships among the identified immune cell types, the "corrplot" R package was utilized for a detailed correlation analysis[43]. This was graphically depicted in the form of comprehensive heatmaps, providing an in-depth view of the immune cell interactions. Box plots were then employed to highlight and compare the expression levels of these 22 immune cell types across various subtypes, delivering a clearer understanding of their differential expression patterns. The study was rounded off with a heatmap visualization, adeptly illustrating the correlations between specific genes and immune cell types, thus offering a multifaceted view of their interplay within the context of COPD.

Consensus aggregation of diseases and GSVA analysis

In our study, COPD demonstrated the potential for division into distinct subtypes via co-clustering analysis. This process commenced with the amalgamation of expression data from both training and validation sets of COPD patients, setting the stage for a comprehensive analysis. Characteristic gene expressions extracted from all disease samples were instrumental in constructing an initial expression matrix, specifically tailored for co-clustering. Employing the "ConsensusClusterPlus" R package[44], we adeptly classified subtypes within the COPD patient cohort. The determination of the optimal number of clusters was a critical step, achieved by meticulously examining relative changes in the consistency matrix plot, the consistency cumulative distribution function (CDF) plot, and the trajectory plot. Principal component analysis (PCA), facilitated through the use of the "ggplot2" R package, was then conducted to ascertain the robustness of our clustering outcomes. PCA plots provided a visual representation of the distinct separation among various subtypes, evident on the scatter plot. Our analysis further extended to an exploration of expression patterns across different subtypes, employing heatmaps and box plots to focus on the characterization of each gene. Subsequently, gene set variation analysis (GSVA) was performed on the diverse COPD subtypes. Utilizing the molecular signatures database (MSigDB) as our reference set, and feature genes identified in prior steps as the sample gene set, GSVA scores were calculated using the "GSVA" R package. These scores, indicative of the absolute enrichment of gene sets, furnished valuable insights into the differential enrichment of functional pathways across the COPD subtypes, thereby enriching our understanding of the disease's complex molecular framework.

Single-cell data annotation

In our detailed investigation of COPD, the single-cell dataset GSE173896 was meticulously obtained using the Read10X function, with subsequent data conversion into Seurat objects in the R environment accomplished via the "Seurat" package. Notably, two samples, GSM5282538 and GSM7882055, were excluded due to their marked heterogeneity compared to other sample data. The integration of the remaining 10 tumor samples was proficiently executed using the merge function. Quality control was a paramount step, assessed by evaluating the proportion of mitochondrial ribosomes and erythrocytes within the cells. Criteria for inclusion entailed cells with gene counts ranging from 800 to 8000, while cells exceeding a ribosome proportion of 15% (pctMT = 15) were omitted to ensure the exclusion of low-quality cells from our dataset. The consolidated dataset then underwent normalization through the NormalizeData function. In the ensuing analysis of single cells based on RNA sequencing (scRNA-seq) data, Principal Component Analysis (PCA) was employed for dimensionality reduction and clustering. The initial phase involved leveraging the FindVariableFeatures function to pinpoint the top 2000 highly variable genes. The dimensionality reduction step incorporated assigning principal component scores from 1 to 10. A UMAP figure was generated for visualizing the emergent unsupervised cell clusters. Cell type identification for each cluster was achieved using marker genes sourced from previous studies, ensuring accuracy in cell type labeling. Additionally, heatmaps were utilized to exhibit and label the differentially expressed genes specific to each cell cluster, effectively showcasing the characteristic genes of each cluster.

Senescence gene score

In our advanced single-cell transcriptomic analysis, we employed five widely recognized algorithms to evaluate gene sets: AddModuleScore, ssGSEA, AUCell, UCell, and SingScore, each offering unique analytical perspectives. The "AddModuleScore" algorithm, integrated within the "SingleR" R package, is utilized for gene set scoring. This method involves initially calculating the mean expression of all genes within a gene set. Subsequently, the expression matrix is segmented based on this average value, and a set of control genes are randomly selected from each segment to serve as comparative background values. The "ssGSEA" approach, a single-sample gene set enrichment analysis method, focuses on assessing the enrichment level of a gene set within individual samples or cells. It operates on a rank-based ordering of gene expression within a sample, calculating an enrichment score for each gene set. "AUCell" evaluates whether a specified gene set is predominantly enriched among the top 5% of expressed genes in a single sample, based on gene expression rankings. The distribution of AUC scores across cells aids in examining relative expression features. Notably, due to its ranking-based scoring mechanism, AUCell is independent of gene expression units and normalization methods. "UCell" functions as an unsupervised cell type identification method, categorizing individual cell types. Its signature scoring is based on the Mann–Whitney U statistic, offering robustness against dataset size and heterogeneity, and requiring less computational resources compared to other methods. Lastly, "SingScore" assesses cellular states, quantifying the activity level of specific

functions or biological processes in a single sample or cell. It leverages the gene set of a gene expression profile, calculating cell state scores by considering gene weights and orientations. These tools collectively play a pivotal role in facilitating in-depth analysis in the realms of single-cell transcriptomics and gene set enrichment.

Differential gene enrichment analysis

The ClusterGVis R package, a notable contribution from Dr. Jun Zhang of China Pharmaceutical University, stands as an advanced tool in the realm of single-cell data processing and visualization. As part of a series of visualization toolkits, ClusterGVis is uniquely equipped to simultaneously render a combination of clustering, grouping expression trend line graphs, and functional annotation figures. This package is particularly adept at delineating differential genes into distinct clusters via heatmaps, providing a clear visualization of how the expression of each cluster evolves over time. In our study, ClusterGVis was used to identify differential genes in different cell populations for visualisation. The results of the study are elegantly presented through heatmaps, providing an intuitive visual representation. Furthermore, the genes identified were functionally annotated using either gene ontology (GO) or Kyoto encyclopedia of genes and genomes (KEGG) pathways, adding a layer of functional insight to the genomic data. This integrative approach facilitated a comprehensive understanding of the dynamic gene expression patterns and their functional implications within the context of our investigation.

Pseudotime analysis

In our exploration of the role of senescent genes within cellular pseudotemporal trajectories, a rigorous analysis was conducted using the monocle2 package in R software. This entailed the creation of mycds objects, with the initial step being the identification of highly variable genes in the dataset. Criteria were set for this purpose, including a mean expression threshold of ≥ 0.1 and a discrete experience of $\geq 1 *$ discrete fit. Following this, the dimensionality of the mycds object was reduced utilizing the "DDRTree" method, specifying a maxComponent of 2. The subsequent phase involved the application of the "orderCells" function, which adeptly ordered the reduced dimensionality results. To visualize the dynamic expression of senescent genes in the pseudotemporal trajectories of various cell types in COPD, trajectory figures were generated. Further, the slingshot package played a crucial role in inferring cell lineages and pseudotimes. This package utilizes clustering-based minimum spanning trees (MSTs) to discern lineage structures and applies synchronized master curves to fit branching curves to these identified lineages, thereby providing an intricate understanding of the cellular dynamics and gene expression patterns in the context of senescence within COPD.

Cellular communication analysis

To delve into the complexities of intercellular communication and the heterogeneity among individual cell clusters and those identified via Non-negative Matrix Factorization (NMF), we conducted an advanced analysis using the "CellChat" R software package. This in-depth analysis was bolstered by leveraging the CellChatDB.human database (www.cellphonedb.org), a comprehensive repository encompassing a wealth of receptors, ligands, and their interactions, as compiled from a vast array of previously published research data. This rich database enabled us to meticulously assess the intricate exchanges of key signaling molecules between different cell clusters, thereby allowing the identification of specific ligand-receptor pairs that underpin these cellular interactions. For a vivid and informative visualization of the differences in immune communication networks, we employed a variety of visualization techniques available in CellChat, including netVisual heatmaps, netVisual bubble plots, and circle diagrams. These tools provided us with a clear, detailed view of the communication patterns and pathways within and across cell clusters, thereby enriching our understanding of the cellular interplay in the studied context.

Statistical analysis

R software (version 4.3.0) along with its related packages served as the foundation for our statistical evaluations. To rigorously assess the disparities and correlations among various groups of variables, we employed two key statistical tests: the t-test and the Wilcoxon rank-sum test. These methods were pivotal in discerning the statistical nuances within our data. We established the threshold for statistical significance at $P < 0.05$, ensuring that our findings adhered to a rigorous standard of scientific accuracy and relevance. This approach allowed for a robust and reliable interpretation of the data, underpinning the validity of our conclusions within the context of our research objectives.

Results

Analysis of variance and GSEA analysis

Figure 1 in our study delineates the methodological flow chart. We utilized the GSE57148 dataset sourced from the gene expression omnibus (GEO) repository as our training cohort, comprising bulk RNA-sequencing data from 98 individuals diagnosed with chronic obstructive pulmonary disease (COPD) alongside 91 healthy controls[45]. The inclusion criteria were a postbronchodilator FEV1/FVC ratio (ratio of forced expiratory volume in the first second to forced vital capacity) of less than 0.7 for the COPD group and normal spirometry for the control group in accordance with American Thoracic Society/European Respiratory Society criteria[46]. Furthermore, to ensure the replicability and precision of our findings, we validated our results using the GSE54837 dataset, encompassing blood samples from 226 COPD patients across various disease stages[47]. Initially, standardization procedures were applied to the GSE57148 training cohort to guarantee uniformity in both mean and quartile range distributions (Supplementary Fig. 1A, B). We analyzed the difference between the COPD and control groups and obtained 2754 genes with significant differences ($|logFC| > 0.25$ and FDR < 0.05). The volcano plot shows the approximate profile of the differential genes between the two groups, and we labeled the genes with $|logFC| > 0.5$ and FDR < 0.05, of which 10 genes were downregulated and 6 genes were upregulated (Fig. 2A).
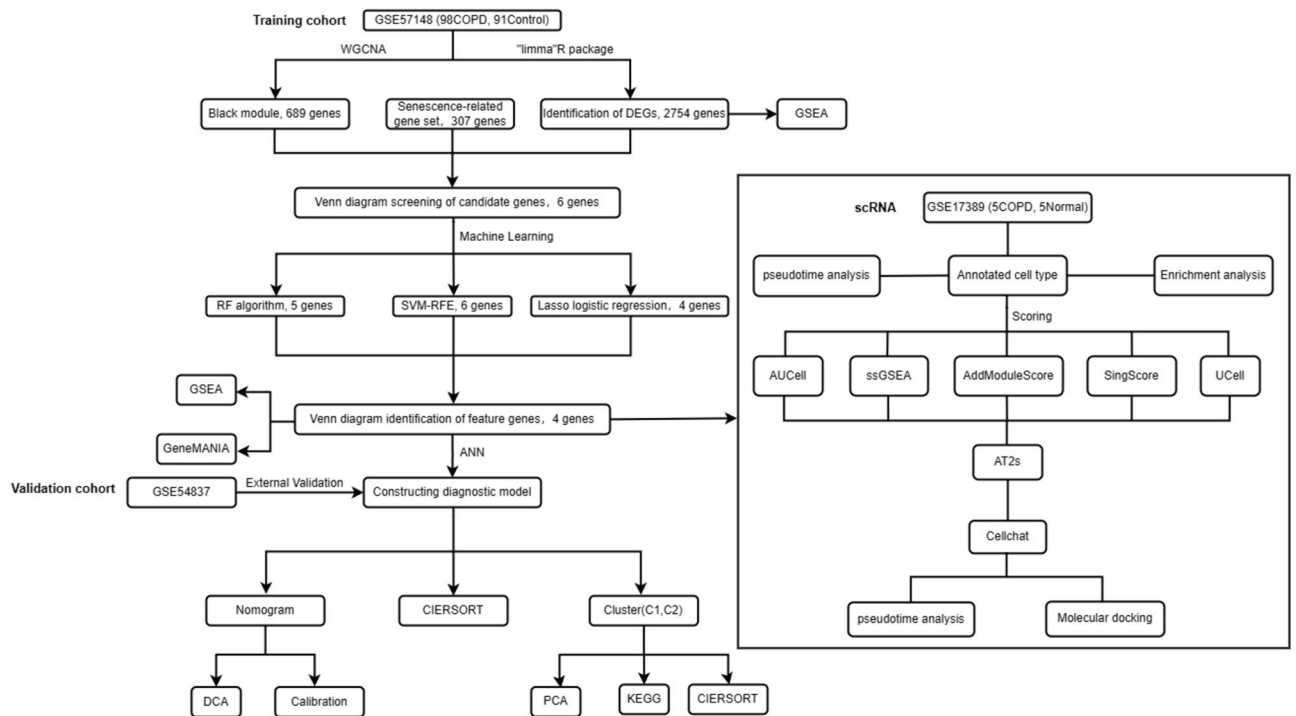
Fig. 1. Flow chart of the design of this study.

In addition to this, the heatmap more specifically demonstrated the variability of gene expression between the COPD and control groups (Fig. 2B). Further, in order to explore the related biological processes and molecular mechanisms, we analyzed the enrichment levels of all 2754 genes with significant differential expression in functional pathways using GSEA. This analysis highlighted the top five most significantly enriched pathways that were both up-regulated and down-regulated (Fig. 2C,D). Intriguingly, DEGs were notably concentrated in the IL-17 and TNF signaling pathways, both of which are frequently implicated as critical regulators in chronic inflammation-related lung diseases, such as COPD and pulmonary fibrosis[44,48]. Additionally, pathways including Chemical Carcinogenesis—DNA Adducts, Drug Metabolism—Other Enzymes, Glutathione Metabolism, Oxidative Phosphorylation, and Ribosome were found to be down-regulated. This pattern suggests that the observed disparities in gene expression between normal and COPD groups may significantly impact COPD development. The modulation of gene expression in these specific pathways emerges as a pivotal factor in the progression of COPD, offering crucial insights for a more profound understanding of the disease's mechanism and informing the development of future therapeutic strategies.

WGCNA to construct gene co-expression networks

In order to pinpoint differentially expressed genes (DEGs) associated with aging, and to identify candidate biomarkers for COPD etiology, we strategically employed WGCNA to delineate the intricate regulatory network and pivotal biological pathways governing COPD pathogenesis. Maintaining an optimal average connectivity of the network, indicated by a scale-free topological fit index $R^2 > 0.9$, we established a soft threshold β of 10 (Supplementary Fig. 2A). To cluster genes with analogous expression patterns, a height of 0.25 was set for dendrogram cutting (Supplementary Fig. 2B). The resulting gene clustering tree elucidated both original and merged modules, leading to the identification of 19 modules with strong correlations for further analysis (Supplementary Fig. 2C). The network heatmap furnished a comprehensive view of the correlations among genes within each module (Supplementary Fig. 2D). Particularly notable in our comparison between the Control and COPD groups was the distinct gene expression observed in two modules, MEsalmon and MEblack (Fig. 3A). A module correlation analysis, represented in a scatter plot, revealed a significant positive correlation between the MEblack module and COPD (cor = 0.67, p = 5.9e–91) (Supplementary Fig. 2E). Based on these findings, we have further narrowed down the candidate biomarker genes to the MEblack module, which we will study in a comprehensive and targeted manner to explore their roles in terms of hypothesized molecular pathways related to COPD pathophysiology.

Machine learning screening of senescence-related diagnostic biomarkers

Delving into the nexus between COPD and senescence genes, a rigorous intersection analysis was conducted. This involved integrating disease-related co-expression module genes, identified via WGCNA, with differentially expressed genes and a defined set of senescence genes, resulting in the extraction of six candidate genes (Fig. 4A). These candidates underwent a further filtering process utilizing three advanced machine learning algorithms to pinpoint reliable biomarkers with potential diagnostic value for COPD. The support vector machine recursive
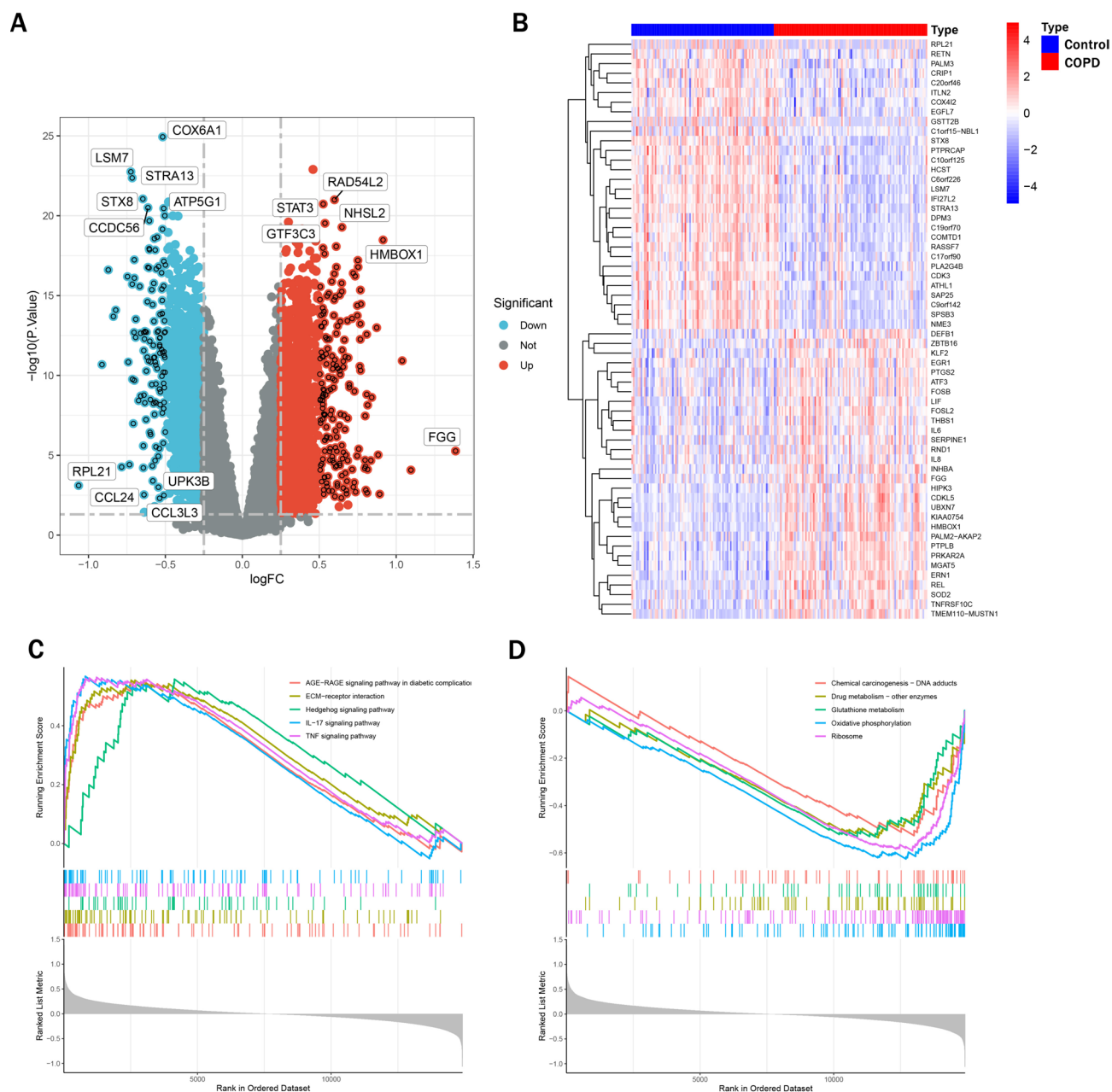
Fig. 2. This study involved the identification and GSEA enrichment analysis of differential genes between Control and COPD. In (**A**), a Volcano plot illustrates differential gene expression, with up-regulation indicated by |log2FC| > 0, p < 0.05 and down-regulation by |log2FC| < 0, p < 0.05. Genes were selected based on |log2FC| > 1, p < 0.05 criteria, with blue denoting down-regulation and red signifying up-regulation. (**B**) Features a heatmap contrasting Control (blue module) with COPD (red module), where red and blue colors correspond to up- and down-regulated expressions, respectively. Finally, (**C,D**) present GSEA enrichment analyses, highlighting the top five and bottom five enriched pathways.

feature elimination (SVM-RFE) algorithm effectively identified the top six feature genes, achieving the highest accuracy (0.773) and the lowest error rate (0.227) (Fig. 4B,C). LASSO regression analysis, focusing on statistically significant univariate variables, culled this list down to four genes (Fig. 4D). Additionally, the Random Forest (RF) algorithm was employed for ranking the relative importance of these candidate genes based on error rates and random forest numbers, identifying five genes with a relative importance exceeding 12.5 (Fig. 4E,F). A Venn diagram, known as a Wayne diagram, was then used to intersect the candidate genes identified by all three machine learning algorithms, ultimately converging on four key genes as biomarkers for the diagnosis of COPD: EP300, MTOR, NFE2L1, and TXN (Fig. 4G). This multi-layered analytical approach provided a refined set of biomarkers, offering significant promise for enhancing diagnostic accuracy in COPD.
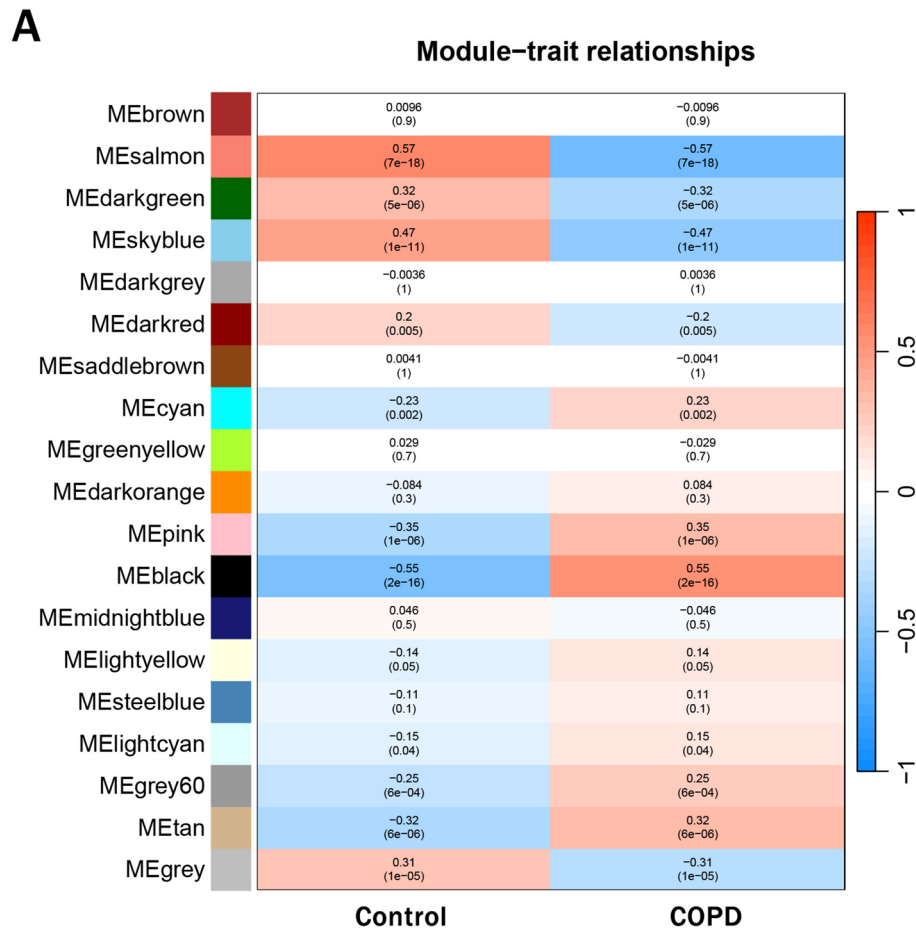
A

Fig. 3. Construction of the WGCNA co-expression network. (**A**) Heatmap of correlation between different modules and associated clinical manifestations.

Construction and validation of diagnostic models

In our investigation, an artificial neural network (ANN) model was constructed using four feature genes as input layers and categorizing samples into Control and COPD groups as output layers. The efficacy of this ANN model was established through a training and validation set test, demonstrating its significant capability to differentiate between Control and COPD groups (Fig. 5A). The model's predictive power was further underscored by the area under the curve (AUC) of the receiver operating characteristic (ROC) curve in our training set, indicating a substantial role in predicting COPD (Fig. 5B). To validate the model's applicability, we employed the identified feature genes and their corresponding gene scores in an independent validation dataset, achieving an AUC value of 0.864. This result affirmed the model's robustness in predicting COPD (Fig. 5C). Additionally, a Nomogram was innovatively developed to enhance the clinical utility of COPD risk assessment (Fig. 5D). The calibration curve of the Nomogram exhibited minimal deviation between the model's predictions and actual COPD risk levels (Fig. 5E). Notably, the decision curve analysis (DCA) results demonstrated the superior diagnostic utility of our model over the baseline "ALL" curves, particularly within the 0–1 threshold range, thereby significantly enhancing clinical diagnostic accuracy (Fig. 5F). This comprehensive approach ensured that our ANN model not only distinguished between Control and COPD groups in the training set effectively but also showcased impressive performance in the validation dataset. The integration of the Nomogram and DCA further highlighted the model's potential clinical application in the diagnosis of COPD.

Correlation analysis of feature genes

Utilizing single-sample gene set enrichment analysis, our analysis illuminated the top five up- and down-regulated enrichment pathways related with each feature gene, thereby unraveling their roles in the pathogenesis of COPD and their expression patterns in diverse biological processes (Fig. 6A–D). Notably, genes EP300, MTOR, and NFE2L1 exhibited significant overexpression in the ECM-receptor interaction pathway, whereas TXN was markedly underexpressed. In stark contrast, the oxidative phosphorylation pathway presented an inverse expression pattern. A further intriguing discovery was the strong positive correlation among the expressions of EP300, MTOR, and NFE2L1, all of which exhibited negative correlations with TXN (Fig. 6F). To probe the potential synergistic genetic effects of these genes, we employed GeneMANIA, a network-based tool, revealing their multifaceted interactions (physical, genetic, etc.). Crucially, EP300 and MTOR were implicated in cellular responses
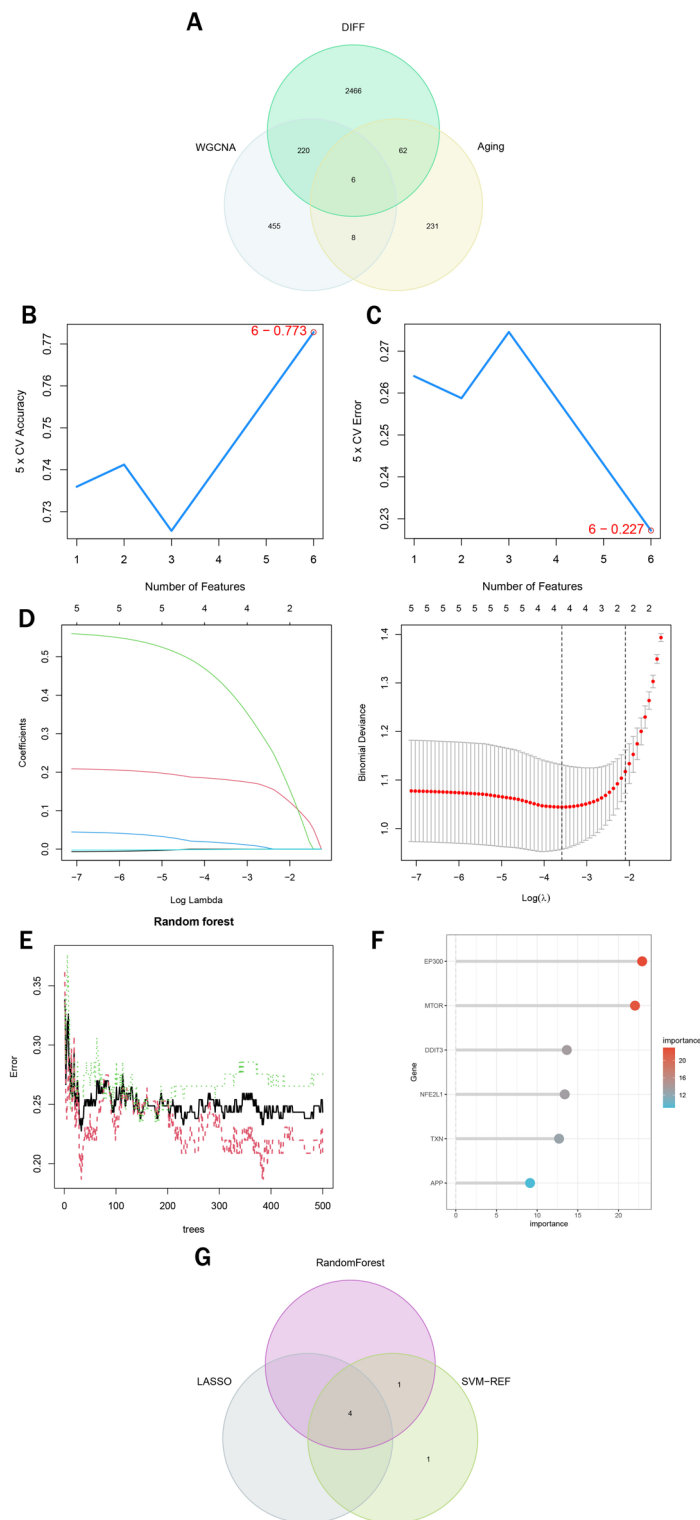
Fig. 4. This study utilized machine learning to identify feature genes associated with COPD. (**A**) Presents Venn diagrams comparing differential genes between Control and COPD, genes from the WGCNA module most correlated with COPD, and senescence-related genes. (**B**,**C**) Describe the use of the support vector machine recursive feature elimination (SVM-RFE) algorithm to select biometric feature genes, identifying the optimal number based on the lowest accuracy and error rate. (**D**) Details the application of the LASSO algorithm for feature gene selection. (**E**) Examines the relationship between the number of decision trees and the error rate in the Random Forest algorithm. (**F**) Highlights the top six genes identified by Random Forest in terms of gene importance. Finally, (**G**) showcases a Venn diagram summarizing feature genes identified by all three machine learning algorithms as key markers for COPD.
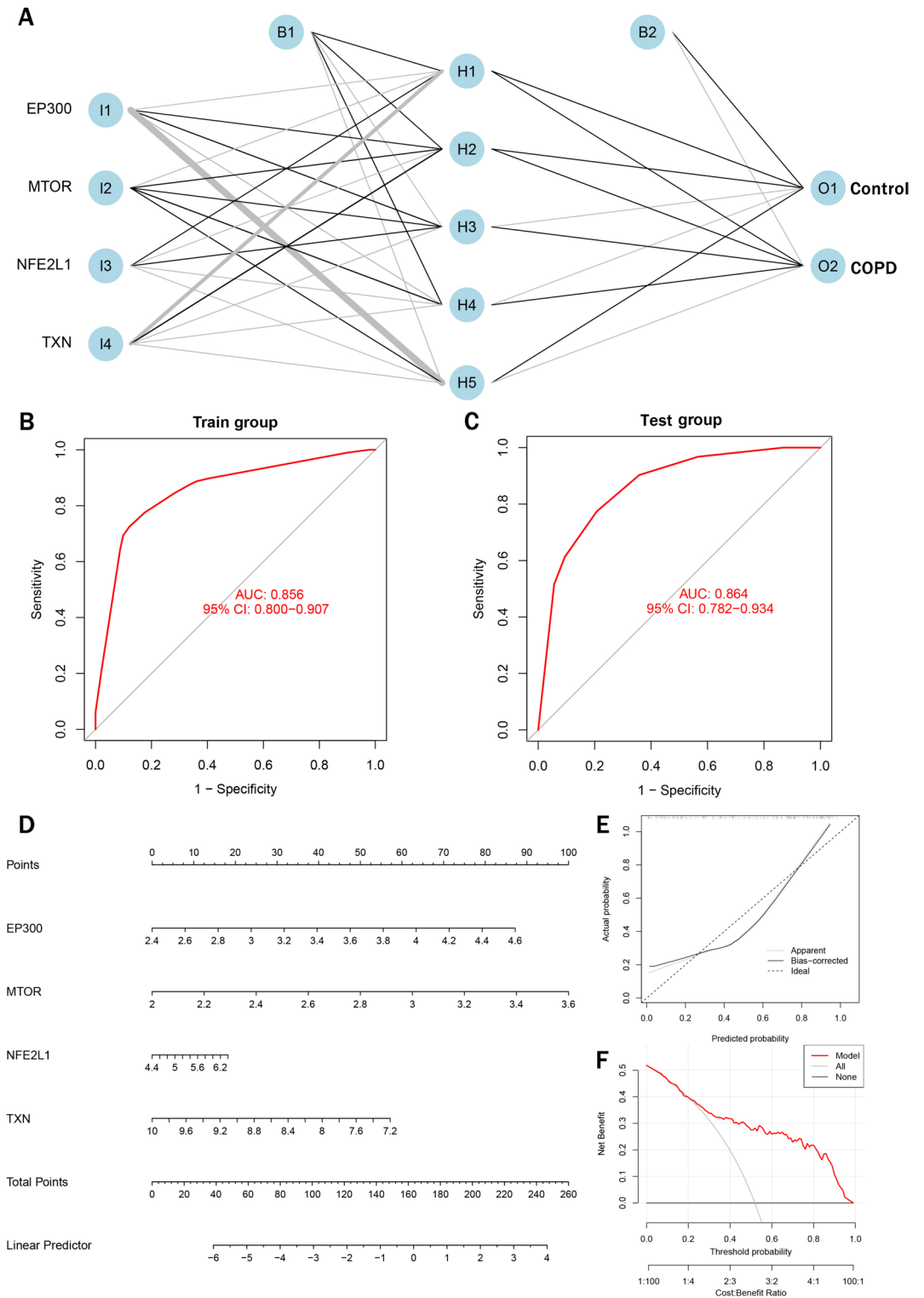
Fig. 5. Construction and validation of ANN models and nomogram models. (**A**) Details the ANN model, structured with input, hidden, and output layers. The model's predictive performance was evaluated using a training set receiver operating characteristic (ROC) curve (**B**), yielding an area under the curve (AUC) of 0.856 (95% CI 0.800–0.907), and a validation set ROC curve (**C**) with an AUC of 0.864 (95% CI 0.782–0.934), assessing model stability and generalizability. (**D**) Describes the constructed nomogram, which integrates feature genes with an assigned scale, summing element scores to predict disease risk. The nomogram's predictive accuracy was assessed through a calibration curve (**E**), where the concurrence of the solid and dashed lines, and proximity to the diagonal, indicated predictive reliability. Finally, (**F**) employs decision curve analysis (DCA) to compare the clinical utility of the nomogram against other diagnostic measures, evaluating benefits across a threshold range of 0 to 1, with higher AUC values denoting greater clinical relevance.
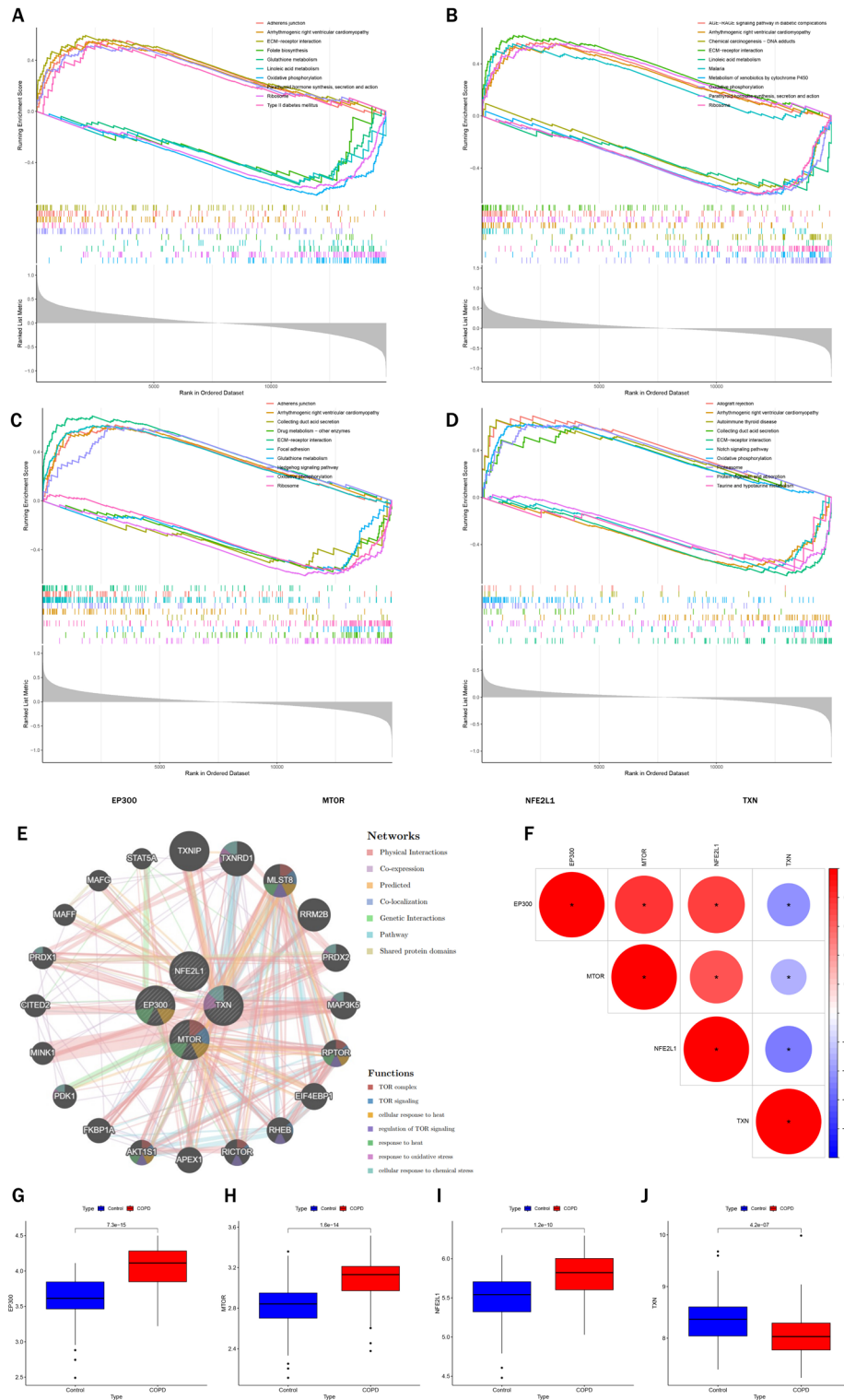
Fig. 6. Feature gene correlation analysis. (**A–D**) detail the single-gene gene set enrichment analysis (GSEA) conducted for EP300, MTOR, NFE2L1, and TXN, respectively, with a focus on the top five pathways of highest and lowest enrichment significance. (**E**) Presents a GeneMANIA network diagram for these genes, featuring 20 closely related genes in the outer circle. The interconnecting lines represent correlations, while varied colored modules within the genes signify different enriched functional pathways. (**F**) Illustrates a heatmap of correlations among the four genes, where red and blue denote positive and negative correlations, respectively, and circle size and color intensity reflect the strength of the correlation. (**G–J**) depict box plots showing the differential expression of each gene in Control and COPD. The range of the box plots indicates gene expression variability, with the black line representing the mean value. Statistical significance is denoted by p-values between the box plots, with $P < 0.05$ indicating significance. Asterisks represent varying levels of statistical significance: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

to thermal stimuli, while TXN played a role in cellular oxidative stress and chemical stress regulation. A notable interaction through MAP3K5, also known as apoptosis signal-regulated kinase 1, was observed among NFE2L1, EP300, MTOR, and TXN (Fig. 6E). MAP3K5, integral in oxidative stress (OS) regulation, cell proliferation, differentiation, and death, and immune responses[49], is inhibited by TXN, a target of ROS[50]. COPD's hallmark features include heightened oxidative stress, mitochondrial dysfunction, and inhibited oxidative phosphorylation[51], with ECM-receptor interactions and MAPK signaling pathways being crucial in its development[52]. Our findings corroborated these insights, revealing upregulation of EP300, MTOR, and NFE2L1 in COPD, contrasted with TXN's downregulation (Fig. 6G–J). This suggests that these genes may influence COPD progression via oxidative stress regulation. Additionally, the upregulation of EP300, MTOR, and NFE2L1 might disrupt lung function by altering ECM generation and degradation, while TXN's downregulation indirectly mirrors oxidative phosphorylation inhibition. Collectively, these insights shed light on how the regulation of senescent genes and their interrelationships contribute to COPD's evolution, laying the groundwork for further elucidation of the disease mechanism.

Association between immune infiltration and feature genes in COPD

CIBERSORT is based on the principle of linear support vector regression for deconvolution, and CIBERSORT outperformed other methods with respect to noise, unknown mixture content, and closely related cell types[45,53]. Therefore, we chose the LM22 signature matrix as the reference file for the CIBERSORT algorithm to analyze the immune infiltration of the gene expression matrix of the training set and speculate on its possible cell type composition, which in turn provides insight into the complexity of immune cell infiltration in COPD. Employing the CIBERSORT algorithm, our analysis delved into the intricacies of immune cell infiltration in COPD, assessing the proportions of 22 immune cell types between Control and COPD groups (Fig. 7A). A box-and-line plot provided a more vivid comparison of immune cell infiltration differences between these groups (Fig. 7C). A noteworthy observation was the predominance of macrophages in both Control and COPD groups. Specifically,
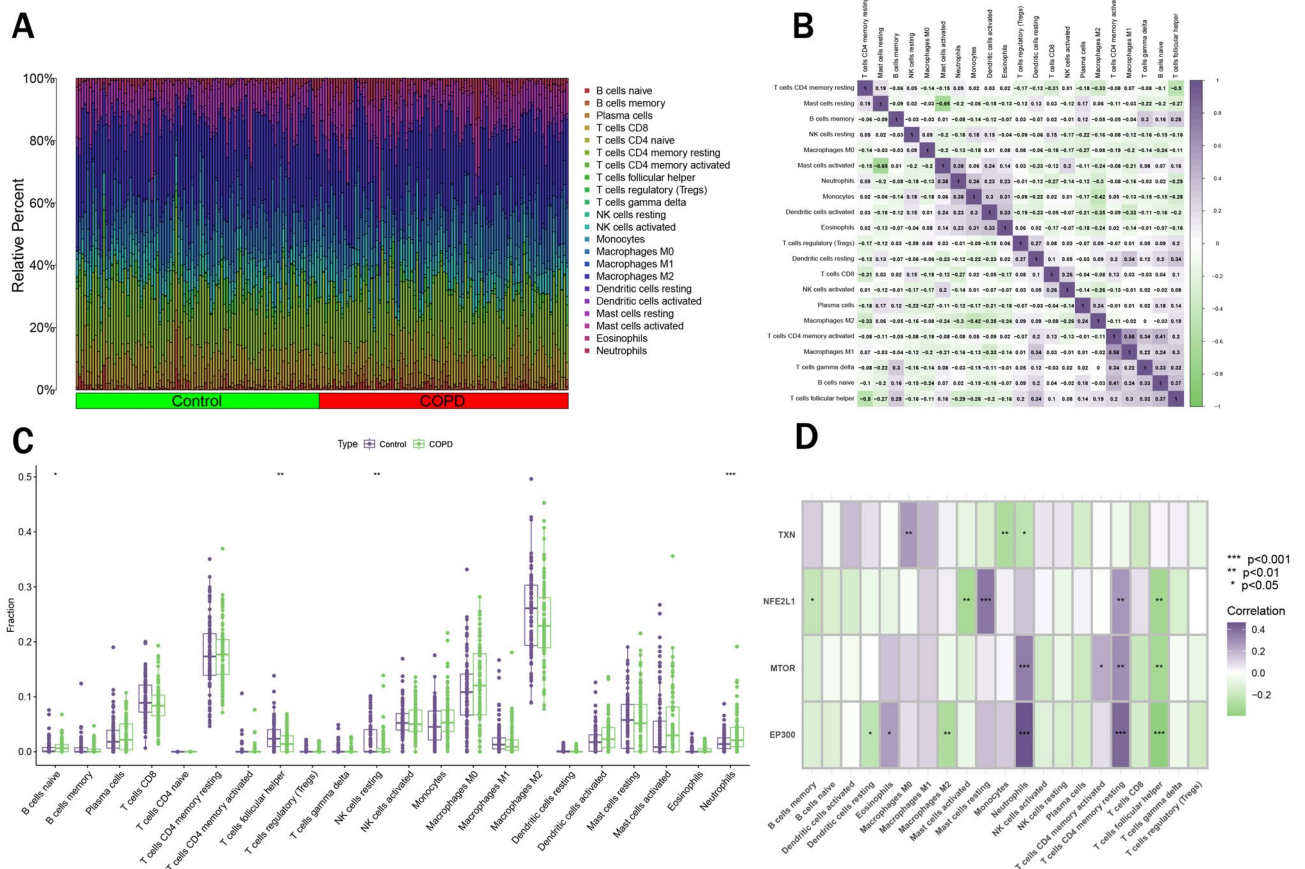

Fig. 7. The analysis focused on immune cell infiltration differences between control and COPD subjects. (**A**) Features a stacked histogram depicting the proportions of 22 immune cell types infiltrating in each sample for both groups. (**B**) Displays a heatmap representing the correlation between these 22 immune cells, with purple indicating positive correlation and green signaling negative correlation; the values in the heatmap correspond to the magnitude of these correlations. (**C**) Introduces a box line plot to illustrate the variation in infiltration levels of the 22 immune cells between the Control and COPD groups. (**D**) Presents another heatmap, this time delineating the correlations between the four feature genes (EP300, MTOR, NFE2L1, TXN) and the 22 types of immune cells, providing a nuanced understanding of gene-immune interactions in COPD. Statistical significance in these correlations is indicated by asterisks: *P < 0.05, **P < 0.01, ***P < 0.001.

Neutrophils (p < 0.001) and B cells native (p < 0.05) showed a significantly higher infiltration in the COPD group. Conversely, T cells follicular helper and NK cells resting were found to be relatively down-regulated compared to the Control group. Our correlation analysis among immune cells revealed a pronounced negative correlation between Mast cells activated and Mast cells resting (r = − 0.65), and a positive correlation between T cells CD4 memory activated and Macrophages M1 (Fig. 7B). Further, to probe the interaction between feature genes and immune cell infiltration, we presented their correlations using a heatmap. Intriguingly, EP300, MTOR, and NFE2L1 were positively correlated with Neutrophils and T cells CD4 memory resting, whereas TXN displayed a negative correlation. Notably, there was a significant negative correlation between EP300, MTOR, NFE2L1 and T cells follicular helper (Fig. 7D). Previous studies have indicated that Treg cells, memory T cells, and T follicular helper cells are differently involved in metabolism, favoring fatty acid oxidation through AMPK while inhibiting MTOR[54]. Furthermore, MTOR is known to regulate Neutrophil chemotaxis and the formation of neutrophil extracellular traps (NETs) by inducing HIF-1α protein expression[55,56]. These insights collectively suggest a profound association between MTOR and its related genes with immune cell infiltration in COPD, offering a new perspective on the immune landscape of this disease.

Consensus clustering of feature genes
COPD samples from the training and validation sets were amalgamated to acquire the necessary expression data for subtyping. Utilizing the expression profiles of the four identified feature genes, COPD was subclassified based on consensus clustering. The optimal subtype number, determined as k = 2 from the consensus matrix plot and the cumulative distribution function (CDF) curves, led to the categorization of the samples into two distinct subtypes, C1 and C2 (Fig. 8A–C). Principal Component Analysis (PCA) validated this classification, distinctly separating the two subtypes on the PCA plot (Fig. 8D). Subsequent Gene Set Variation Analysis (GSVA) of subtypes C1 and C2 revealed that the KEGG_TIGHT_JUNCTION pathway played a pivotal role in differentiating the two subtypes, highlighting its potential significance in the development of COPD (Fig. 8E). A notable observation was the differential expression of the four feature genes between the subtypes: down-regulation in C1 and up-regulation in C2 (Fig. 8F,G). This pattern of gene expression thus provided a basis for classifying COPD into two subtypes, further underscoring the impact of the four feature genes on the progression of COPD.

Consensus clustering of HBV-LC by feature genes
Single-cell sequencing data for COPD were sourced from GSE173896, encompassing lung tissue samples from five pathologically confirmed COPD patients and five normal samples, including smokers and non-smokers. For the initial processing and analysis of this single-cell data, we employed the widely recognized Seurat software package. The data were imported and Seurat objects created for preprocessing, which included filtering cells with gene expression outside the 200 to 6000 range and those with a mitochondrial gene ratio exceeding 20%, aiming to exclude potentially abnormal or low-quality cells from further analysis. Figure 9A and B illustrate the sequencing data features before and after this quality control (QC) process. Following data normalization, dimensionality reduction cluster analysis was performed, identifying highly variable genes and normalizing the expression matrix. Principal Component Analysis (PCA) was utilized to reveal the distribution characteristics of single cells in both disease and normal groups, as depicted in Fig. 9C (left panel). The Elbow Plot (Fig. 9C, right panel) aided in determining the optimal number of dimensions, specifically the first ten principal components, for more detailed cluster analysis. Employing t-distributed stochastic neighbor embedding (tSNE), cells from diseased and normal groups were clustered into 24 distinct clusters and mapped in a uniform coordinate space (Fig. 9D). Cell type identification marker genes were sourced from the comprehensive CellMarker database (http://xteam.xbio.top/CellMarker/index.jsp). Upon completing precise cell type annotation using these markers, the distribution patterns of each cell type in normal and COPD patient samples were visualized via tSNE plots (Fig. 9E). To further investigate the interplay between the inflammatory microenvironment and cellular senescence in COPD, we conducted a gene set scoring analysis on the annotated data. This process incorporated five gene set expression scoring tools—AUCell, UCell, ssgsea, singscore, and addModuleScore—to enhance the robustness of our findings. Each cell type was scored based on a defined set of senescence-related genes, with the results from these five methods being standardized and normalized to compile a composite score. Bubble plots (Fig. 9F) effectively display these scores, with violin plots (Fig. 9G) presenting the combined scores of different cell types in normal versus COPD samples. Notably, T cells, NK cells, AT2 alveolar epithelial cells, mast cells, B cells, and ciliated cells exhibited significant score differences between COPD and normal groups. The augmented senescence-related gene expression in AT2 cells within the COPD group, in particular, suggested their critical role in disease pathology. This comprehensive analysis enabled the clustering of single-cell data, accurate cell type annotation, and identification of cell subpopulations with elevated expression of senescence-related genes in COPD patients, thereby providing vital insights for further exploration of COPD's molecular mechanisms and potential therapeutic targets.

Single-cell RNA sequencing reveals cellular heterogeneity and identification of senescence-related cell types in lung tissues of COPD patients
To more clearly explore the expression and function of aging-related genes in the lung tissues of COPD patients, we have incorporated single-cell transcriptomic data to achieve higher resolution analysis. Single-cell sequencing data for COPD were sourced from GSE173896, encompassing lung tissue samples from five pathologically confirmed COPD patients and five normal samples, including smokers and non-smokers. For the initial processing and analysis of this single-cell data, we employed the widely recognized Seurat software package. The data were imported and Seurat objects created for preprocessing, which included filtering cells with gene expression outside the 200 to 6000 range and those with a mitochondrial gene ratio exceeding 20%, aiming to exclude potentially
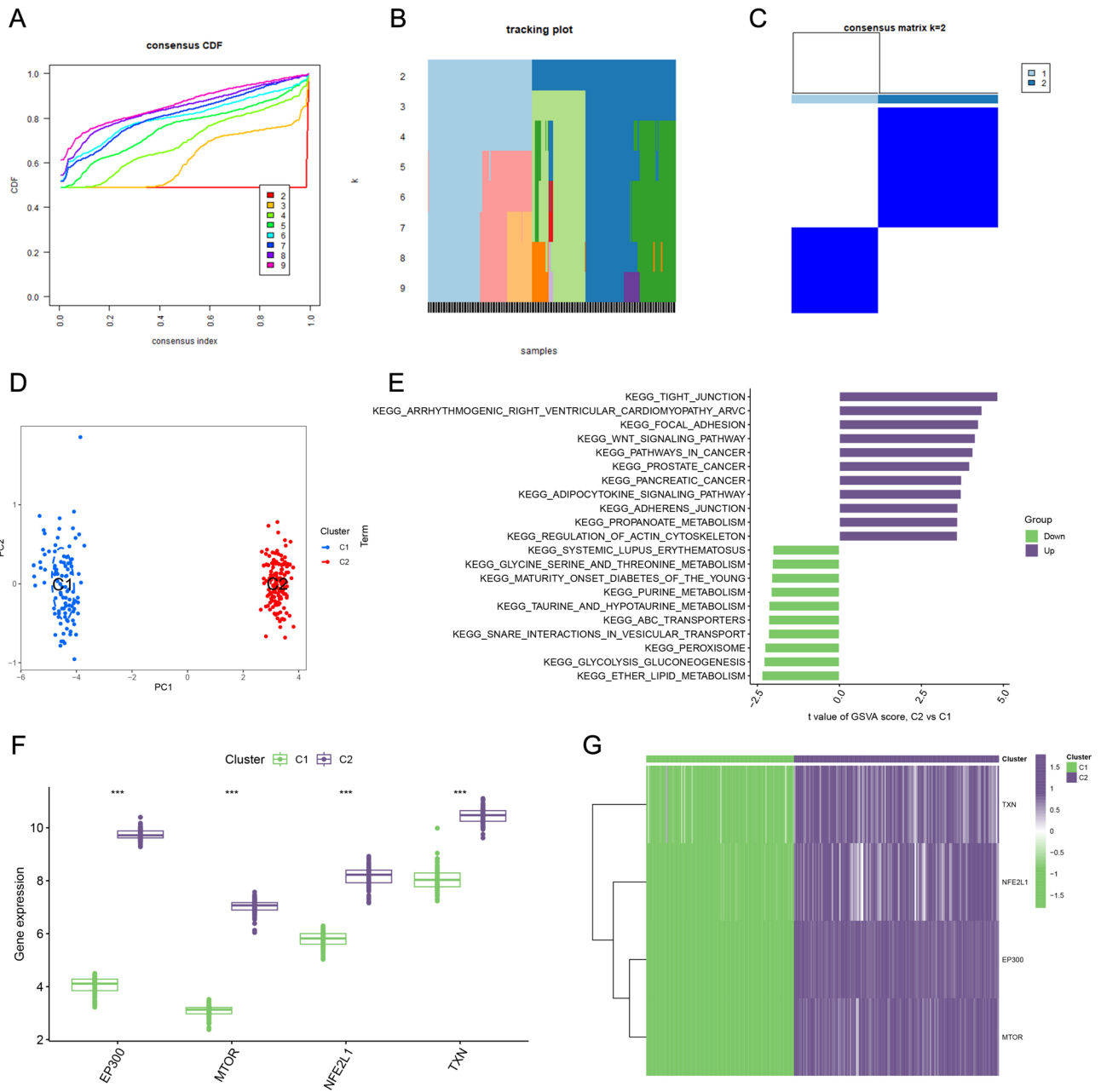
Fig. 8. Consensus clustering of COPD samples. (**A**) presents a cumulative distribution function (CDF) figure for consensus clustering, highlighting the relative change in consensus indices from k = 2 to k = 9, with the curve exhibiting the most stable change indicating the optimal number of clusters. (**B**) Visualizes the clustering trajectory from k = 2 to k = 9. (**C**) Features a consensus matrix, where distinct blank areas between blue modules signify a successful analysis. (**D**) Illustrates the principal component analysis (PCA) of COPD samples, with scatter plots delineating feature genes differentiating COPD into two subtypes, C1 and C2. (**E**) Details a KEGG analysis comparing these subtypes, where purple and green represent up- and down-regulated functional pathways, respectively. (**F**) Shows a box plot contrasting the expression of feature genes between C1 and C2. Finally, (**G**) displays a heatmap correlating C1 and C2 with feature genes, using purple to indicate up-regulation and green for down-regulation, with *P < 0.05, **P < 0.01, ***P < 0.001 denoting statistical significance levels.

abnormal or low-quality cells from further analysis. Figures 9A and B illustrate the sequencing data features before and after this quality control (QC) process. Following data normalization, dimensionality reduction cluster analysis was performed, identifying highly variable genes and normalizing the expression matrix. Principal component analysis (PCA) was utilized to reveal the distribution characteristics of single cells in both disease and normal groups, as depicted in Fig. 9C (left panel). The elbow plot (Fig. 9C, right panel) aided in determining the optimal number of dimensions, specifically the first ten principal components, for more detailed cluster analysis. Employing t-distributed stochastic neighbor embedding (tSNE), cells from diseased and normal groups
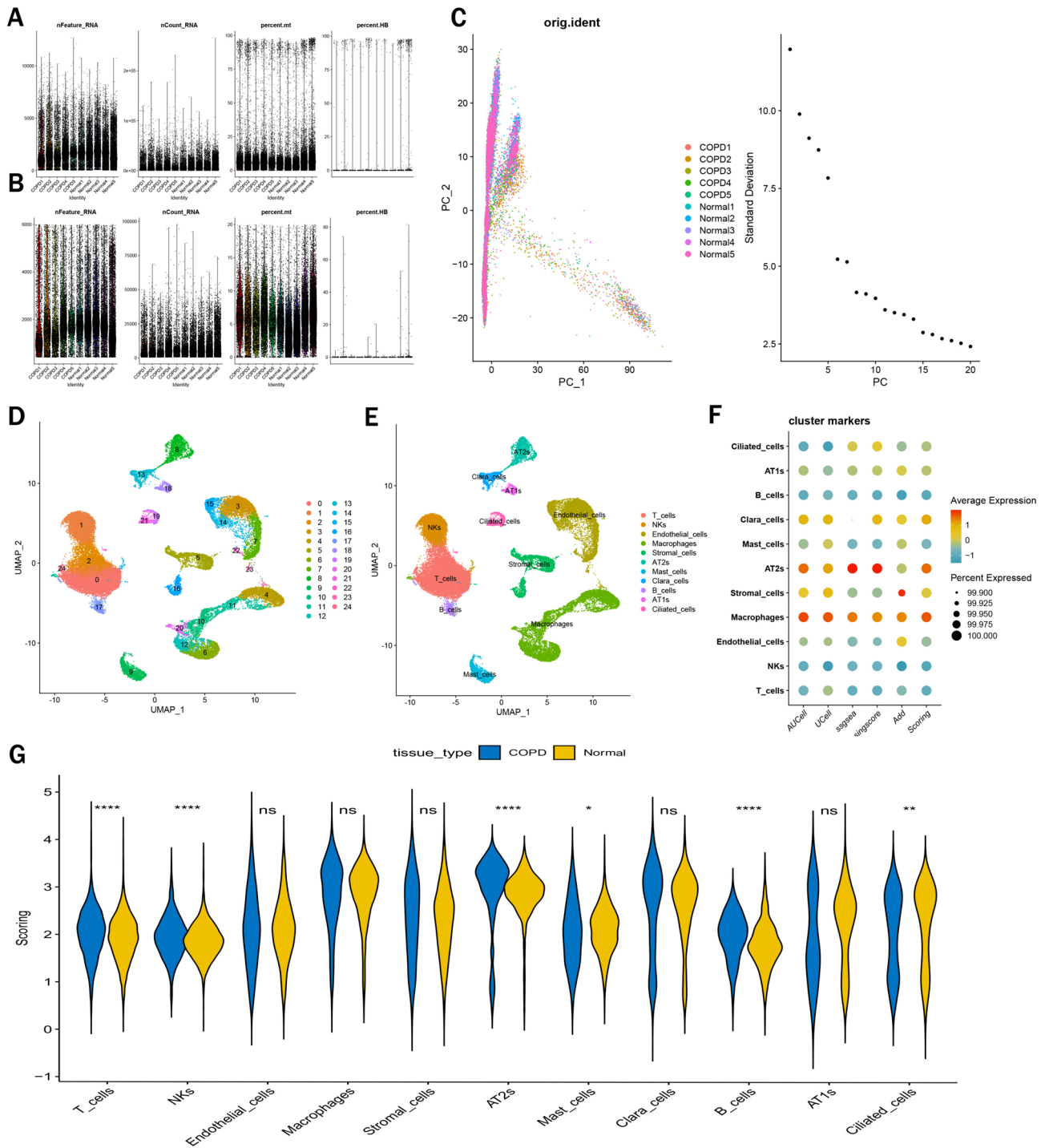
Fig. 9. Single-cell data processing and scoring of senescence-related gene sets. (**A**) and (**B**) depict the single-cell data before and after quality control, respectively, with the latter stage involving the removal of genes exhibiting excessive variation and mitochondrial genes. (**C**) Illustrates the gene screening within the single-cell expression matrix, incorporating linear downscaling through PCA clustering, alongside a distribution chart of the top 20 principal components (PCs). (**D**) Showcases the UMAP dimensionality reduction clustering applied to the quality-controlled single-cell data, segregating it into 25 distinct cell clusters. (**E**) Further refines this UMAP clustering, post-cellular annotation, classifying different cell clusters into specific cell types. (**F**) Details the scoring of these cell types based on senescence-related gene sets, employing five distinct scoring methodologies. Finally, (**E**) (which seems to be a typographical error and should be referred to as (**G**) presents differential violin plots, contrasting the expression of senescence-related genes between the COPD and control groups.

were clustered into 24 distinct clusters and mapped in a uniform coordinate space (Fig. 9D). Cell type identification marker genes were sourced from the comprehensive CellMarker database (http://xteam.xbio.top/CellMarker/index.jsp). Upon completing precise cell type annotation using these markers, the distribution patterns of each cell type in normal and COPD patient samples were visualized via tSNE plots (Fig. 9E). To understand the expression of aging-related genes across various cell types within the pulmonary inflammatory microenvironment of COPD patients, we performed gene set scoring on single-cell data. This process incorporated five gene set expression scoring tools—AUCell, UCell, ssgsea, singscore, and addModuleScore—to enhance the robustness of our findings. Each cell type was scored based on a defined set of senescence-related genes, with the results from these five methods being standardized and normalized to compile a composite score. Bubble plots (Fig. 9F) effectively display these scores, with violin plots (Fig. 9G) presenting the combined scores of different cell types in normal versus COPD samples. Notably, T cells, NK cells, AT2 alveolar epithelial cells, mast cells, B cells, and ciliated cells exhibited significant score differences between COPD and normal groups. The augmented senescence-related gene expression in AT2 cells within the COPD group, in particular, suggested their critical role in disease pathology. This comprehensive analysis enabled the clustering of single-cell data, accurate cell type annotation, and identification of cell subpopulations with elevated expression of senescence-related genes in COPD patients, thereby providing vital insights for further exploration of COPD's molecular mechanisms and potential therapeutic targets.

Enrichment analysis and pseudotime analysis based on single-cell data

Numerous cell types exist within the lung tissues of COPD patients, each performing distinct functions. In an in-depth examination of single-cell sequencing data from COPD patients, we constructed bar graphs to depict the distribution of various cell types across different samples (Fig. 10A). For the eleven identified cell types, average gene expression levels within each cell type were calculated, leading to the identification of genes differentially expressed among these cell types. Gene ontology (GO) enrichment analysis was then performed on these differentially expressed genes, revealing the biological processes and molecular functions enriched in different cell types (Fig. 10B). To elucidate cellular differentiation and evolutionary trends at the single-cell level, pseudotime analysis and cell trajectory analysis were conducted. Employing the "monocle" R package, we mapped the pseudotime distribution of cells and their developmental trajectories, with Fig. 10C depicting pseudo-temporal ordering of cells and Fig. 10D illustrating various cell types alongside developmental trajectories and dendrograms. Additionally, the "Slingshot" R package was utilized to construct single-cell differentiation lineages and perform pseudo-temporal inference. This involved constructing a minimum spanning tree on cell clustering to ascertain the global genealogical structures of cells, and plotting a developmental line graph based on tSNE plots to represent these structures (Fig. 10E). Smooth lineage structures were then created by fitting master curves (Fig. 10F), highlighting cell populations involved in different lineages (Fig. 10G). These comprehensive analyses not only underscored the variances in cell type distribution within lung tissues of COPD patients but also offered a novel perspective on the mechanisms of cell differentiation and evolution in this context.

CellChat deciphered cellular communication in COPD microenvironment

To elucidate the roles of various cell types in the microenvironment of COPD, the CellChat software package was employed for predicting and analyzing intercellular communication networks. Extracted annotated single-cell data from Seurat objects were imported into CellChat, utilizing its human ligand-receptor database. CellChat models the probability of intercellular communication by integrating gene expression data with established knowledge of interactions among signaling ligands, receptors, and cofactors. This simulation approach leverages a variety of action modes, enabling the deduction of the number and strength of interactions between different cell types based on a priori information (Fig. 11A and B). Hierarchical diagrams were crafted to display the relative positions and connectivity of diverse cells in the signaling network (Fig. 11C). Scatter plots were used to depict the signal sending and receiving strengths of each cell type (Fig. 11D), while chordal plots illustrated patterns of AT2s receiving signals from other cell types (Fig. 11E) and Stromal cells transmitting signals (Fig. 11F). These plots underscored the crucial roles of these cell types in intercellular communication. Particular focus was placed on the FN1 ligand receptor family due to its key role in cellular communication. We isolated the FN1 family of ligand-receptor-mediated signaling interactions from the cellular communication network and highlighted the relative contributions of each ligand-receptor pair within this family via bar graphs, identifying the FN1-CD44 pair as significantly impactful (Fig. 11G). String and circle plots were then utilized to represent FN1-CD44 ligand-receptor pair-mediated intercellular communication (Fig. 11H and I), with line thickness indicative of interaction strength. These analyses revealed the significance of communication between Stromal cells and other cell types in FN1-CD44 ligand-receptor pair-mediated interactions, offering novel perspectives on cellular communication within the COPD microenvironment.

Further exploration of senescence-related diagnostic genes

To ascertain the expression patterns of four senescence-related diagnostic genes (EP300, MTOR, NFE2L1, TXN) in single-cell sequencing data, a heatmap on a tSNE plot was generated (Fig. 12A), enabling visualization of gene expression at the individual cell level. Utilizing the "monocle" R package, differentiation trajectories of cell development were constructed, showcasing the dynamic expression of these genes in a chronological sequence (Fig. 12B). This analysis was augmented with pseudo-temporal analysis, mapping a gene expression heatmap to chronicle the expression patterns of these genes over time (Fig. 12C). It was observed that TXN genes manifested higher expression levels in the middle stage of cell development, whereas NFE2L1, MTOR, and EP300 were significantly more expressed in the early stages. Figure 12D precisely delineates the expression of these genes throughout the cell development trajectory, and Fig. 12E displays their expression trends across different
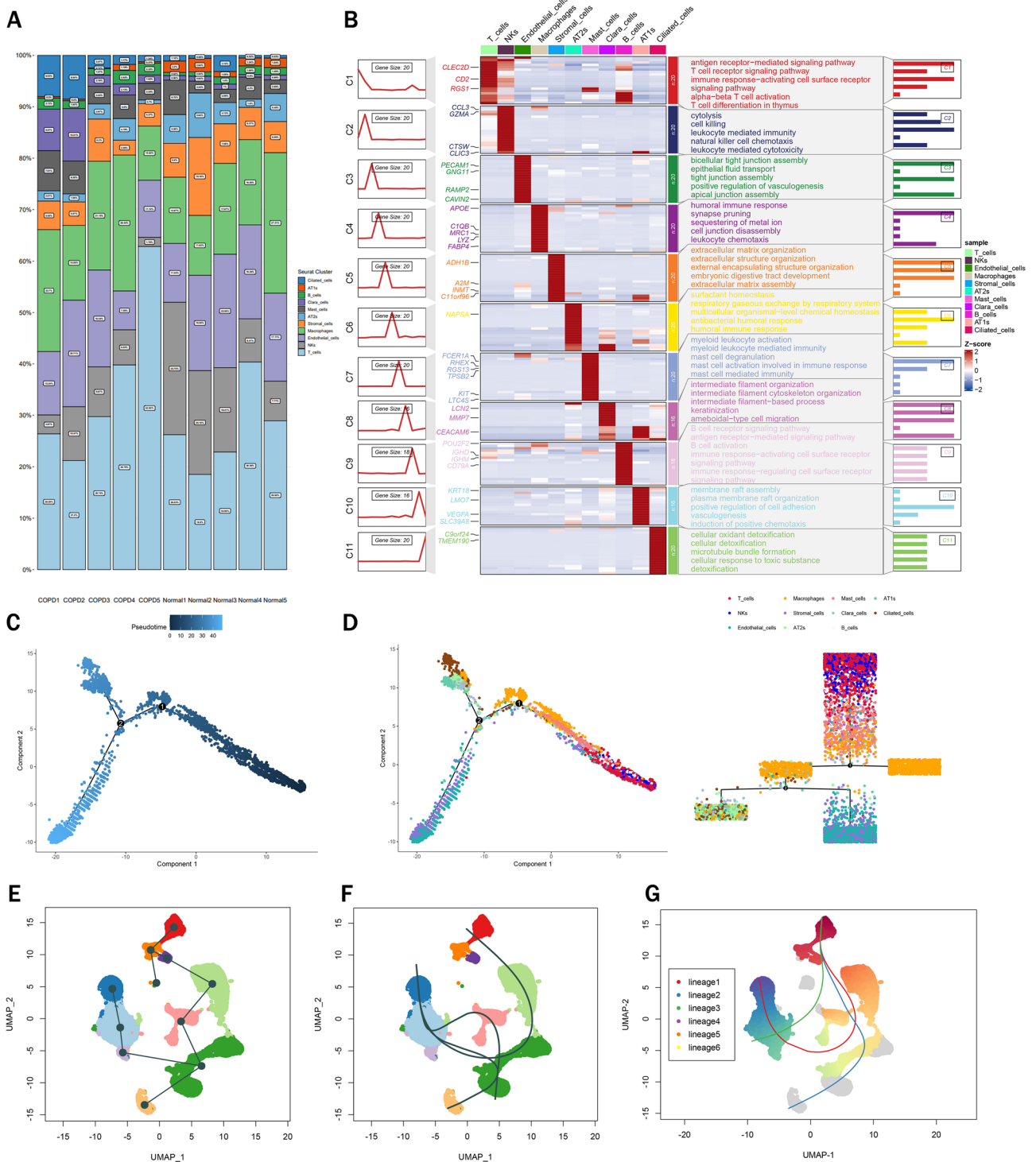
Fig. 10. Enrichment and proposed time-series analysis of single-cell data. (**A**) Presents a histogram depicting cell type proportions across each single-cell sample. (**B**) Features a heatmap of gene ontology (GO) enrichment analysis for various cell types. (**C**) Illustrates the proposed time-series trend for all cells, indicating a temporal progression from black to blue. (**D**) Outlines the proposed time-series analysis for different cell types, aimed at tracking their temporal changes. (**E**) and (**F**) Describe the implementation of monocle2 for a proposed time-series line graph and plot, respectively, of the single-cell data. Lastly, (**G**) showcases a figure representing the proposed time-series analysis, with a focus on the progression of type II alveolar epithelial cells as the initial point of study.
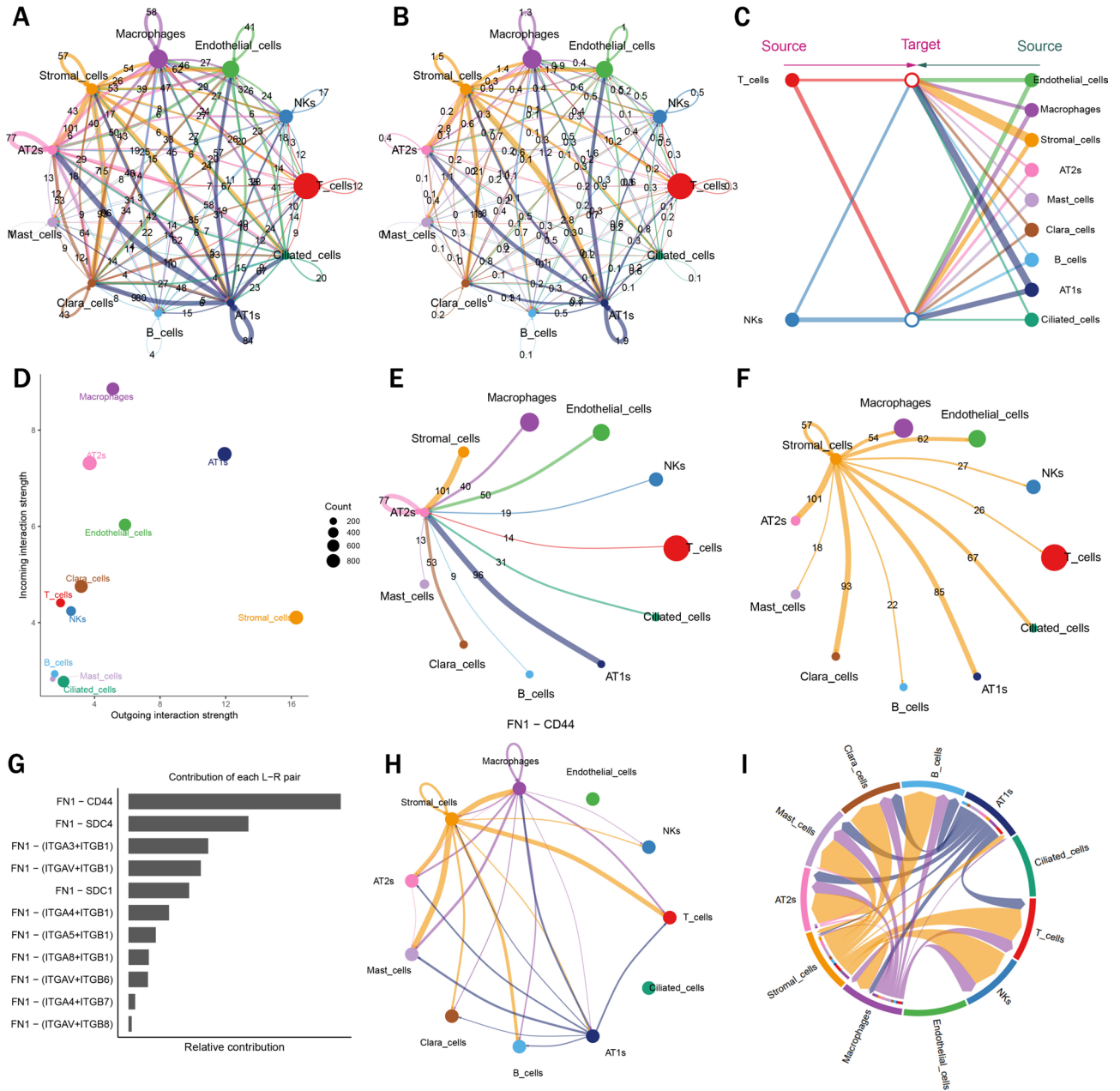
Fig. 11. COPD intercellular communication analysis. (**A**) Details the number of cell–cell interactions, with line thickness and accompanying numbers indicating the quantity of these interactions. (**B**) Focuses on the strength of these interactions, again represented by line thickness and numerical values. (**C**) Depicts a hierarchical figure illustrating the relationships among eleven cell types in the communication network. (**D**) Provides a dot plot of prominent ligand-receptor pairings categorized by their respective families. (**E**) and (**F**) Describe the output signals originating from type II alveolar epithelial cells and stromal cells, respectively, to other cell types. (**G**) Presents a histogram quantifying the contribution of each ligand-receptor pair. (**H**) and (**I**) Both highlight the specific intercellular communication involving the ligand-receptor pair FN1-CD44 across eleven cell types, with (**H**) showing a demonstration and (**I**) presenting chord plots. This comprehensive analysis offers a multi-faceted view of the intricate communication networks in COPD at the cellular level.

developmental stages. In a subsequent molecular docking analysis of these four diagnostic genes, a significant interaction was noted between NFE2L1 and its regulator resveratrol, meriting an in-depth visualization (Fig. 12F). Three-dimensional structural analysis revealed intricate interactions between the resveratrol molecule and the NFE2L1 protein binding pocket. Resveratrol's structure proved highly compatible with NFE2L1's active site, with its hydrophobic portion engaging with the protein's hydrophobic residues and the hydroxyl group forming hydrogen bonds with polar residues. This allows for a stable ligand–protein binding state. A zoomed-in view further elucidated the precise localization of resveratrol at the NFE2L1 binding site, detailing the interactions between key residues and suggesting a potential regulatory mechanism. These molecular docking studies not only affirmed NFE2L1's potential as a therapeutic target for senescence but also provided molecular-level
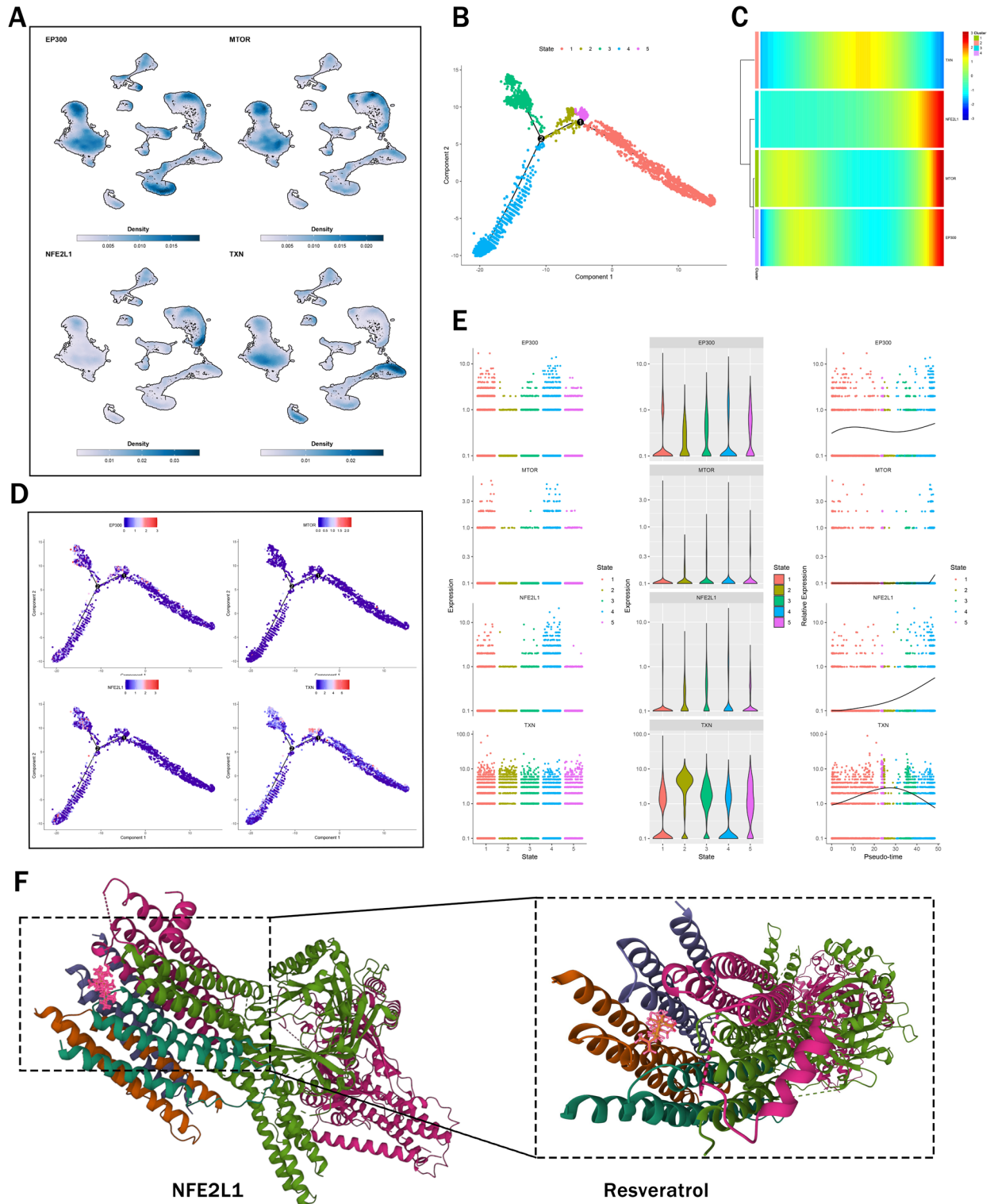
Fig. 12. Analysis and molecular docking of COPD diagnosis-related feature genes. (**A**) Presents the expression visualization of four COPD diagnosis-related feature genes within single-cell data. (**B**) (initially labeled as (**D**)) Proposes a time-series analysis of cell clusters exhibiting significant differences. (**C**) Depicts a time-series heatmap for these feature genes, illustrating temporal progression from blue to red. (**D**) Delves into a time-series analysis of these feature genes in notably differentiated cell clusters. (**E**) Extends this analysis to the expression patterns of these genes in corresponding clusters where they are highly expressed. Finally, (**F**) explores the molecular docking interactions between NFE2L1 and resveratrol, providing insights into the potential molecular mechanisms underpinning their relationship in the context of COPD. This comprehensive approach combines expression analysis with molecular docking to enhance the understanding of gene roles in COPD diagnosis.

insights into resveratrol's possible role in senescence regulation. This finding holds significant implications for drug design and development related to senescence and may offer crucial information for novel therapeutic strategies targeting the senescence process.

Discussion

COPD stands as a complex chronic inflammatory condition, intricately influenced by a myriad of factors, including genetic predispositions, which are pivotal in its pathogenesis. A growing body of evidence underscores a significant link between senescence genes and COPD[56,57]. Research indicates that with senescence, cells increasingly succumb to damage and senescence, leading to alterations in lung structure and function, heightened sensitivity to inflammation, and oxidative stress, all of which cumulatively contribute to COPD's development[58]. Notably, certain senescence genes, such as SIRT1, have been identified to exert a regulatory influence on cellular senescence and antioxidant defenses, with their variants potentially elevating the risk of COPD[59,60]. In this study, a comprehensive exploration of gene expression variances, gene co-expression networks, machine learning-based screening of diagnostic genes, cellular heterogeneity, and intercellular communication networks in lung tissues from COPD patients was undertaken. This was achieved through the synergistic application of bioinformatics and single-cell RNA sequencing technologies. The aim of these multifaceted analyses is to delve into the nexus between senescent genes and COPD, aspiring to unveil novel insights and methodologies that could revolutionize early diagnosis, pathogenesis elucidation, and the crafting of individualized treatment strategies for COPD.

In the study on COPD, employing a comprehensive approach, we elucidate the intricate mechanisms of immune regulation, focusing particularly on senescence-associated genes. The identification of EP300, MTOR, NFE2L1, and TXN through machine learning and weighted gene co-expression network analysis (WGCNA) underscores their pivotal roles in modulating immune responses within the COPD milieu. EP300, a transcriptional co-activator, orchestrates gene expression profiles that influence immune cell functionality and inflammatory pathways, thereby sustaining chronic inflammation[61], and its involvement in inflammatory and oxidative stress responses in COPD has been suggested[62]. MTOR, integral to cellular growth and metabolism, modulates immune cell activation and function; its dysregulation in COPD likely exacerbates immune cell hyperactivity, contributing to persistent inflammation and tissue damage[63,64]. The NFE2L1 gene encodes proteins crucial in the anti-oxidative stress pathway and has been linked with various cancers and diabetes mellitus[65–68], involved in cellular stress responses and antioxidative defense, plays a critical role in maintaining immune homeostasis by mitigating oxidative stress; its impairment may amplify oxidative damage and disrupt immune equilibrium. The TXN gene, coding for proteins with antioxidant and anti-inflammatory properties, is instrumental in regulating inflammation and apoptosis, thus holding promise for COPD prevention and treatment[69]. Single-cell analyses delineate the heterogeneity of immune cell populations in COPD, revealing that senescent macrophages exhibit a pro-inflammatory phenotype, perpetuating chronic inflammation. Concurrently, neutrophils exhibit heightened protease activity, inflicting direct tissue damage, while senescent T cells display an imbalance between effector and regulatory functions, further disrupting immune regulation[70]. Molecular docking studies highlight interactions between NFE2L1 and other proteins, shedding light on their roles in stress response and immune modulation. This comprehensive analysis not only enhances our understanding of the molecular underpinnings of COPD but also identifies potential biomarkers and therapeutic targets for early intervention. These findings are poised to inform future research directions and clinical strategies, ultimately aiming to alleviate the burden of COPD and improve patient outcomes.

The regulation of senescence-related genes within the IL-17 and TNF signaling pathways in COPD likely mirrors complex pathophysiological processes. In the IL-17 pathway, altered senescence gene expression intensify the inflammatory response, as overactivation of IL-17, an inflammatory mediator, is a noted phenomenon in COPD patients[71]. This regulation lead to heightened IL-17 signaling, exacerbating immune cell infiltration and inflammatory mediator release, thereby hastening lung tissue damage[72,73]. Similarly, aberrations in the TNF signaling pathway, involving senescence genes, could augment inflammatory processes. Elevated levels of TNF-α, a crucial inflammatory mediator, are well-documented in COPD, and its over-activation, influenced by senescence gene regulation, propagate the inflammatory response, impairing lung tissue repair and regeneration[73,74]. In the chemical carcinogenesis—DNA adduct pathway, senescence gene down-regulation disrupt the balance between DNA repair and apoptosis, elevating susceptibility to carcinogen-induced DNA damage, leading to cell mutation and cancer[75]. Furthermore, in the Drug Metabolism—Other Enzymes pathway, diminished expression of senescent genes weaken drug metabolism, possibly reducing therapeutic drug responsiveness in COPD patients[76]. The down-regulation of oxidative phosphorylation, glutathione metabolism, and ribosomal pathways cause anomalies in energy metabolism and protein synthesis, impacting normal cell function and regeneration[77]. Consequently, the atypical regulation of senescence-related genes in these pathways exacerbate lung inflammation, cellular damage, and environmental exposure impacts in COPD patients, influencing disease development and progression.

The immune microenvironment in COPD patients is characterized by complex, dynamic changes, including chronic inflammation, aberrant immune cell activation, oxidative stress, and fluctuations in cytokines and chemokines. Notable observations include the hyperactivation of T-cell naïve and neutrophils, coupled with the significant down-regulation of T-cell follicular helper cells and M2 macrophages in immune infiltrates. This pattern underscores a multifaceted dysregulation within the immune system in COPD development. The heightened activation of T-cell naïve indicate an allergic response to pathogenic elements in the airways, instigating a substantial migration of immune cells to lung tissues and perpetuating chronic inflammation and airway mucosal damage[78]. Meanwhile, the active role of neutrophils, potentially linked to chronic infections and inflammatory responses, exacerbated airway obstruction and mucus accumulation[79,80]. Conversely, the decline in T-cell follicular helper cells and M2 macrophages presents challenges in immune regulation and lung tissue repair. T-cell follicular helper cells, typically involved in antibody response regulation and inflammation resolution,

left inflammation unchecked when down-regulated, resulting in exacerbated inflammatory responses[81]. The reduction in M2 macrophages, generally related with anti-inflammatory actions and tissue repair, impaired anti-inflammatory mechanisms and repair processes in the lungs, leading to uncontrolled chronic inflammation and accelerated lung tissue damage[82]. This imbalanced immune microenvironment thus emerges as a pivotal factor in COPD's development and progression.

In the classification of COPD patients into two distinct molecular subtypes, C1 and C2, based on the expression changes of four senescence-featured genes, notable differences in various regulatory pathways were observed, elucidating the biological disparities between these subtypes[83]. The C2 subtype exhibited pronounced expression in the tight junction and adhesion junction pathways, suggesting enhanced cell–cell interactions, adhesion, and maintenance of cellular barriers. This reflects a potential advantage of the C2 subtype over C1 in terms of lung organization and cellular connectivity[84]. Conversely, the lower expression observed in pathways like glycine, serine and threonine metabolism, purine metabolism, and taurine and hypotaurine metabolism in C2 indicates more conservative biological features compared to C1. These pathways are integral to cellular metabolism, energy homeostasis, antioxidant functions, and metabolic equilibrium, influencing drug metabolism and sensitivity in these subtypes[85,86]. This differential pathway expression between the two COPD subtypes underscores the complexity of the disease and provides a nuanced understanding of the molecular variations underlying COPD's pathogenesis.

Single-cell analysis emerges as a pivotal technique in COPD research, offering invaluable insights by delineating gene expression patterns at the individual cell level. This approach is instrumental in identifying cellular heterogeneity, predicting cellular differentiation trajectories, and elucidating intercellular communication, thereby enhancing our understanding of COPD pathogenesis and aiding the development of precise treatment strategies[87]. Concurrently, molecular docking technology is integral to drug discovery and development, especially in COPD, where it facilitates the simulation and analysis of interactions between drugs and their targets. This technology has shed light on the specific binding mechanism between Resveratrol, a polyphenol found in grape skins and red wine known for its antioxidant, anti-inflammatory, and anti-tumor properties, and NFE2L1, a transcription factor linked to intracellular redox homeostasis and the antioxidant response[88,89]. These findings suggest that COPD treatment strategies could be devised by targeting NFE2L1, impacting its role in antioxidant pathways, inflammation regulation, and apoptosis, thus opening new avenues for individualized medicine in COPD management.

In the realm of COPD diagnosis and treatment, the research presented here marks a significant stride in innovation. The employment of WGCNA coupled with three machine learning algorithms enabled a systematic screening and modeling of senescence-related genes that possess the highest diagnostic value. The neural network model developed as part of this research demonstrates remarkable stability and specificity, effectively distinguishing COPD patients from healthy individuals. Additionally, the application of molecular docking technology has facilitated the prediction of drug effects based on COPD-related senescent genes, unveiling potential sites for action on macromolecular proteins and thereby offering new avenues for intervention therapy in COPD. Despite these advancements, the study acknowledges certain limitations. The reliance on data from public databases may introduce some degree of bias in reflecting the actual clinical scenario. Moreover, while preliminary research has been conducted on the mechanism of the four senescent genes in the enrichment pathway, further experimental validation is necessary to enhance our understanding and confirm these findings.

Conclusion

In conclusion, this study represents a significant advancement in understanding and diagnosing COPD through innovative bioinformatics and molecular biology techniques. By employing WGCNA and machine learning algorithms, we have identified and modeled key senescence-related genes with high diagnostic value for COPD. The resulting neural network model, characterized by its exceptional stability and specificity, effectively distinguishes COPD patients from healthy individuals. Furthermore, molecular docking technology has provided insights into potential therapeutic interventions by predicting drug effects on specific senescent genes, opening new pathways for COPD treatment. While our findings offer promising directions, they are derived from public database analyses and require further experimental validation to confirm their applicability in clinical settings. This study not only sheds light on the complex pathogenesis of COPD but also paves the way for the development of personalized medicine approaches in its management.

Data availability

The datasets utilized in this study can be accessed at the GEO database (https://www.ncbi.nlm.nih.gov/geo/). For further information regarding the datasets used and/or analyzed during the study, interested parties may contact the corresponding author and make a reasonable request. All raw data are available at https://www.jianguoyun.com/p/DZ-wmWoQw9CUDBiH3KQFIAA.

References

1. Huang, Q. et al. Single-cell transcriptomics highlights immunological dysregulations of monocytes in the pathobiology of COPD. *Respir. Res.* **23**, 367 (2022).
2. Solomon, G. M. et al. Pilot evaluation of ivacaftor for chronic bronchitis. *Lancet Respir. Med.* **4**, e32-3 (2016).
3. Rabe, K. F. & Watz, H. Chronic obstructive pulmonary disease. *Lancet (Lond., Engl.)* **389**, 1931–1940 (2017).
4. Ornatowski, W. et al. Complex interplay between autophagy and oxidative stress in the development of pulmonary disease. *Redox Biol.* **36**, 101679 (2020).

5. Zhong, S. *et al.* Identification and validation of aging-related genes in COPD based on bioinformatics analysis. *Aging* **14**, 4336–4356 (2022).

6. Mannino, D. M. & Braman, S. The epidemiology and economics of chronic obstructive pulmonary disease. *Proc. Am. Thorac. Soc.* **4**, 502–6 (2007).

7. Cho, S. J. & Stout-Delgado, H. W. Aging and lung disease. *Annu. Rev. Physiol.* **82**, 433–459 (2020).

8. Zhang, F. *et al.* Tissue-specific landscape of metabolic dysregulation during ageing. *Biomolecules* **11**, 235 (2021).

9. Barnes, P. J., Baker, J. & Donnelly, L. E. Cellular senescence as a mechanism and target in chronic lung diseases. *Am. J. Respir. Crit. Care Med.* **200**, 556–564 (2019).

10. Yanagi, S. *et al.* The impacts of cellular senescence in elderly pneumonia and in age-related lung diseases that increase the risk of respiratory infections. *Int. J. Mol. Sci.* **18**, 503 (2017).

11. Kadota, T. *et al.* Emerging role of extracellular vesicles as a senescence-associated secretory phenotype: Insights into the pathophysiology of lung diseases. *Mol. Aspects Med.* **60**, 92–103 (2018).

12. Gershon, A. S. *et al.* Trends in chronic obstructive pulmonary disease prevalence, incidence, and health services use in younger adults in Ontario, Canada, 2006–2016. *Am. J. Respir. Crit. Care Med.* **203**, 1196–1199 (2021).

13. Barnes, P. J. Senescence in COPD and its comorbidities. *Annu. Rev. Physiol.* **79**, 517–539 (2017).

14. Araya, J. *et al.* PRKN-regulated mitophagy and cellular senescence during COPD pathogenesis. *Autophagy* **15**, 510–526 (2019).

15. Nyunoya, T. *et al.* Cigarette smoke induces cellular senescence via Werner's syndrome protein down-regulation. *Am. J. Respir. Crit. Care Med.* **179**, 279–87 (2009).

16. Tsuji, T., Aoshiba, K. & Nagai, A. Cigarette smoke induces senescence in alveolar epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **31**, 643–9 (2004).

17. Zhang, X. Y. *et al.* Roles of sirtuin family members in chronic obstructive pulmonary disease. *Respir. Res.* **23**, 66 (2022).

18. Chandran, A., Oliver, H. J. & Rochet, J. C. Role of NFE2L1 in the regulation of proteostasis: Implications for aging and neurodegenerative diseases. *Biology* **12**, 1169 (2023).

19. Marengo, M. *et al.* Role of the CD40-CD40 ligand pathway in cardiovascular events, neurological alterations, and other clinical complications of chronic hemodialysis patients: Protective role of adsorptive membranes. *Blood Purif.* **52**, 1–16 (2023).

20. Park, S. Y., Kim, S. H., Yoon, H. K., Yim, C. H. & Lim, S. K. The role of nuclear factor-E2-related factor 1 in the oxidative stress response in MC3T3-E1 osteoblastic cells. *Endocrinol. Metab. (Seoul, Korea)* **31**, 336–42 (2016).

21. He, N. *et al.* Resveratrol suppresses microglial activation and promotes functional recovery of traumatic spinal cord via improving intestinal microbiota. *Pharmacol. Res.* **183**, 106377 (2022).

22. Kisková, T. & Kassayová, M. Resveratrol action on lipid metabolism in cancer. *Int. J. Mol. Sci.* **20**, 2704 (2019).

23. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–8 (2008).

24. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–82 (2009).

25. Zhang, H. *et al.* A multitask clustering approach for single-cell RNA-seq analysis in recessive dystrophic epidermolysis bullosa. *PLoS Comput. Biol.* **14**, e1006053 (2018).

26. Wu, K. *et al.* Redefining tumor-associated macrophage subpopulations and functions in the tumor microenvironment. *Front. Immunol.* **11**, 1731 (2020).

27. Li, X., Liao, Z., Deng, Z., Chen, N. & Zhao, L. Combining bulk and single-cell RNA-sequencing data to reveal gene expression pattern of chondrocytes in the osteoarthritic knee. *Bioengineered* **12**, 997–1007 (2021).

28. Williams, J. W. *et al.* Single cell RNA sequencing in atherosclerosis research. *Circ. Res.* **126**, 1112–1126 (2020).

29. Mangiola, S., Doyle, M. A. & Papenfuss, A. T. Interfacing Seurat with the R tidy universe. *Bioinformatics (Oxf., Engl.)* **37**, 4100–4107 (2021).

30. Gonzalez Castro, L. N., Tirosh, I. & Suvà, M. L. Decoding cancer biology one cell at a time. *Cancer Discov.* **11**, 960–970 (2021).

31. Chi, H. *et al.* Deciphering a prognostic signature based on soluble mediators defines the immune landscape and predicts prognosis in HNSCC. *Front. Biosci. (Landmark Ed.)* **29**, 130 (2024).

32. Huang, X. *et al.* Disulfidptosis in head and neck squamous carcinoma: Integrative bioinformatic and in-vitro analysis. *Oral Diseases* https://doi.org/10.1111/odi.14977 (2024).

33. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

34. Ito, K. & Murphy, D. Application of ggplot2 to pharmacometric graphics. *CPT Pharmacomet. Syst. Pharmacol.* **2**, e79 (2013).

35. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass))* **2**, 100141 (2021).

36. Zhang, J. *et al.* CD8 + T-cell marker genes reveal different immune subtypes of oral lichen planus by integrating single-cell RNA-seq and bulk RNA-sequencing. *BMC Oral Health* **23**, 464 (2023).

37. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).

38. Xu, N., Guo, H., Li, X., Zhao, Q. & Li, J. A five-genes based diagnostic signature for sepsis-induced ARDS. *Pathol. Oncol. Res. POR* **27**, 580801 (2021).

39. Alderden, J. *et al.* Predicting pressure injury in critical care patients: A machine-learning model. *Am. J. Crit. Care Off. Publ. Am. Assoc. Crit. Care Nurses* **27**, 461–468 (2018).

40. Engebretsen, S. & Bohlin, J. Statistical predictions with glmnet. *Clin. Epigenet.* **11**, 123 (2019).

41. Beck, M. W. NeuralNetTools: Visualization and analysis tools for neural networks. *J. Stat. Softw.* **85**, 1–20 (2018).

42. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).

43. Liu, Z. *et al.* Identification of GLS as a cuproptosis-related diagnosis gene in acute myocardial infarction. *Front. Cardiovasc. Med.* **9**, 1016081 (2022).

44. Hu, X., Ni, S., Zhao, K., Qian, J. & Duan, Y. Bioinformatics-led discovery of osteoarthritis biomarkers and inflammatory infiltrates. *Front. Immunol.* **13**, 871008 (2022).

45. Shen-Orr, S. S. & Gaujoux, R. Computational deconvolution: Extracting cell type-specific information from heterogeneous samples. *Curr. Opin. Immunol.* **25**, 571–8 (2013).

46. Miller, M. R. *et al.* Standardisation of spirometry. *Eur. Respir. J.* **26**, 319–38 (2005).

47. Singh, D. *et al.* Altered gene expression in blood and sputum in COPD frequent exacerbators in the ECLIPSE cohort. *PLoS One* **9**, e107381 (2014).

48. Garudadri, S. & Woodruff, P. G. Targeting chronic obstructive pulmonary disease phenotypes, endotypes, and biomarkers. *Ann. Am. Thorac. Soc.* **15**, S234-s238 (2018).

49. Li, X. *et al.* Antioxidant stress and anticancer activity of peptide-chelated selenium in vitro. *Int. J. Mol. Med.* https://doi.org/10.3892/ijmm.2021.4986 (2021).

50. Tzeng, H. E. *et al.* Interleukin-6 induces vascular endothelial growth factor expression and promotes angiogenesis through apoptosis signal-regulating kinase 1 in human osteosarcoma. *Biochem. Pharmacol.* **85**, 531–40 (2013).

51. Wiegman, C. H. *et al.* Oxidative stress-induced mitochondrial dysfunction drives inflammation and airway smooth muscle remodeling in patients with chronic obstructive pulmonary disease. *J. Allergy Clin. Immunol.* **136**, 769–80 (2015).

52. Chen, L. *et al.* Identification of hub genes associated with COPD through integrated bioinformatics analysis. *Int. J. Chronic Obstr. Pulm. Dis.* **17**, 439–456 (2022).
53. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–7 (2015).
54. Fantus, D., Rogers, N. M., Grahammer, F., Huber, T. B. & Thomson, A. W. Roles of mTOR complexes in the kidney: Implications for renal disease and transplantation. *Nat. Rev. Nephrol.* **12**, 587–609 (2016).
55. McInturff, A. M. *et al.* Mammalian target of rapamycin regulates neutrophil extracellular trap formation via induction of hypoxia-inducible factor 1 α. *Blood* **120**, 3118–25 (2012).
56. Vitiello, D., Neagoe, P. E., Sirois, M. G. & White, M. Effect of everolimus on the immunomodulation of the human neutrophil inflammatory response and activation. *Cell. Mol. Immunol.* **12**, 40–52 (2015).
57. Wang, X. *et al.* Genetic variants in ADAM33 are associated with airway inflammation and lung function in COPD. *BMC Pulm. Med.* **14**, 173 (2014).
58. Liguori, I. *et al.* Oxidative stress, aging, and diseases. *Clin. Interv. Aging* **13**, 757–772 (2018).
59. Li, B. S., Jin, A. L., Zhou, Z., Seo, J. H. & Choi, B. M. DRG2 accelerates senescence via negative regulation of SIRT1 in human diploid fibroblasts. *Oxid. Med. Cell. Longev.* **2021**, 7301373 (2021).
60. Qin, H., Zhang, H., Zhang, S., Zhu, S. & Wang, H. Protective effect of Sirt1 against radiation-induced damage. *Radiat. Res.* **196**, 647–657 (2021).
61. Chen, Z., Chen, C., Li, L., Zhang, T. & Wang, X. Pan-cancer analysis reveals that E1A binding protein p300 mutations increase genome instability and antitumor immunity. *Front. Cell Dev. Biol.* **9**, 729927 (2021).
62. Li, D. *et al.* SAM protects against alveolar septal cell apoptosis in autoimmune emphysema rats. *Eur. J. Med. Res.* **28**, 460 (2023).
63. Dyakin, V. V., Dyakina-Fagnano, N. V., McIntire, L. B. & Uversky, V. N. Fundamental clock of biological aging: Convergence of molecular, neurodegenerative, cognitive and psychiatric pathways: Non-equilibrium thermodynamics meet psychology. *Int. J. Mol. Sci.* **23**, 285 (2021).
64. Dailah, H. G. Therapeutic potential of small molecules targeting oxidative stress in the treatment of chronic obstructive pulmonary disease (COPD): A comprehensive review. *Molecules (Basel, Switzerland)* **27**, 5542 (2022).
65. Zhang, M., Wang, Z., Yang, G., Han, L. & Wang, X. NFE2L1 restrains ferroptosis by transcriptionally regulating HJURP and participates in the progress of oral squamous cell carcinoma. *J. Bioenerget. Biomembr.* **55**, 467–478 (2023).
66. Dou, R., Wang, X. & Zhang, J. Prognostic value and immune infiltration analysis of nuclear factor erythroid-2 family members in ovarian cancer. *BioMed. Res. Int.* **2022**, 8672258 (2022).
67. Lee, Y. K. *et al.* Mitochondrial respiratory defect enhances hepatoma cell invasiveness via STAT3/NFE2L1/STX12 axis. *Cancers* **12**, 2632 (2020).
68. Liu, X., Xu, C., Xiao, W. & Yan, N. Unravelling the role of NFE2L1 in stress responses and related diseases. *Redox Biol.* **65**, 102819 (2023).
69. Li, Z. *et al.* Nuclear translocation of LDHA promotes the catabolism of BCAAs to sustain GBM cell proliferation through the TxN antioxidant pathway. *Int. J. Mol. Sci.* **24**, 9365 (2023).
70. Zhang, S. *et al.* Construction of a diagnostic model for hepatitis B-related hepatocellular carcinoma using machine learning and artificial neural networks and revealing the correlation by immunoassay. *Tumour Virus Res.* **16**, 200271 (2023).
71. Zhang, H. *et al.* Brown adipose tissue-derived exosomes delay fertility decline in aging mice. *Front. Endocrinol.* **14**, 1180104 (2023).
72. Halwani, R. *et al.* IL-17 enhances chemotaxis of primary human B cells during asthma. *PLoS One* **9**, e114604 (2014).
73. Lao, P. *et al.* Regulatory T cells in lung disease and transplantation. Biosci. Rep. https://doi.org/10.1042/BSR20231331 (2023).
74. Wu, Y. *et al.* Susceptibility to aspergillus infections in rats with chronic obstructive pulmonary disease via deficiency function of alveolar macrophages and impaired activation of TLR2. *Inflammation* **39**, 1310–8 (2016).
75. Benna, C. *et al.* Genetic variation of clock genes and cancer risk: A field synopsis and meta-analysis. *Oncotarget* **8**, 23978–23995 (2017).
76. Ambrocio-Ortiz, E. *et al.* SNPs sets in codifying genes for xenobiotics-processing enzymes are associated with COPD secondary to biomass-burning smoke. *Curr. Issues Mol. Biol.* **45**, 799–819 (2023).
77. Yang, M. *et al.* Proteomic profiling of lung immune cells reveals dysregulation of phagocytic pathways in female-dominated molecular COPD phenotype. *Respir. Res.* **19**, 39 (2018).
78. Nakata, Y. *et al.* Expression and function of nicotinic acetylcholine receptors in induced regulatory T cells. *Int. J. Mol. Sci.* **23**, 1779 (2022).
79. Wang, W. *et al.* The role of glucagon-like peptide-1 receptor agonists in chronic obstructive pulmonary disease. *Int. J. Chronic Obstr. Pulm. Dis.* **18**, 129–137 (2023).
80. Felton, J. M., Lucas, C. D., Rossi, A. G. & Dransfield, I. Eosinophils in the lung - modulating apoptosis and efferocytosis in airway inflammation. *Front. Immunol.* **5**, 302 (2014).
81. Zhang, S., Pang, K., Feng, X. & Zeng, Y. Transcriptomic data exploration of consensus genes and molecular mechanisms between chronic obstructive pulmonary disease and lung adenocarcinoma. *Sci. Rep.* **12**, 13214 (2022).
82. Chan, S. M. H., Selemidis, S., Bozinovski, S. & Vlahos, R. Pathobiological mechanisms underlying metabolic syndrome (MetS) in chronic obstructive pulmonary disease (COPD): Clinical significance and therapeutic strategies. *Pharmacol. Ther.* **198**, 160–188 (2019).
83. Chi, H. *et al.* Proposing new early detection indicators for pancreatic cancer: Combining machine learning and neural networks for serum miRNA-based diagnostic model. *Front. Oncol.* **13**, 1244578 (2023).
84. Wang, H. *et al.* Proteomic and phosphorylated proteomic landscape of injured lung in juvenile septic rats with therapeutic application of umbilical cord mesenchymal stem cells. *Front. Immunol.* **13**, 1034821 (2022).
85. Kelly, R. S. *et al.* Metabolomic differences in lung function metrics: Evidence from two cohorts. *Thorax* **77**, 919–928 (2022).
86. Kim, H. Y. *et al.* Comprehensive targeted metabolomic study in the lung, plasma, and urine of PPE/LPS-Induced COPD mice model. *Int. J. Mol. Sci.* **23**, 2748 (2022).
87. Handly, L. N., Yao, J. & Wollman, R. Signal transduction at the single-cell level: Approaches to study the dynamic nature of signaling networks. *J. Mol. Biol.* **428**, 3669–82 (2016).
88. Wang, G. & Zhu, W. Molecular docking for drug discovery and development: A widely used approach but far from perfect. *Future Med. Chem.* **8**, 1707–10 (2016).
89. Cucciolla, V. *et al.* Resveratrol: From basic science to the clinic. *Cell Cycle (Georgetown, Tex)* **6**, 2495–510 (2007).

Acknowledgements

Author contributions

XS, HC and HW conceived the study. XL, WZ, JT, WL, JY, HQC, LJ, XZ, JH, SZ, LD, XS, HC and HW drafted the manuscript. XL and WZ performed the literature search and collected the data. XL, JT analyzed and visualized the data. XS, HC and HW helped with the final revision of this manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-71773-w.

Correspondence and requests for materials should be addressed to X.S., H.C. or H.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.