

# Identification of Oxidative Stress-Associated Biomarkers in Chronic Obstructive Pulmonary Disease: An Integrated Bioinformatics Analysis

Xianwei Jiang<sup>1,2</sup>, Minghang Wang<sup>1-3</sup>, Huiru Li<sup>1,2</sup>, Yuanyuan Liu<sup>1,2</sup>, Xiaosheng Dong<sup>1,2</sup>

<sup>1</sup>National Regional TCM (Lung Disease) Diagnostic and Treatment Center, The First Affiliated Hospital of Henan University of CM, Zhengzhou, People's Republic of China; <sup>2</sup>First Clinical Medical College, Henan University of Chinese Medicine, Zhengzhou, People's Republic of China; <sup>3</sup>Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Henan University of CM, Zhengzhou, People's Republic of China

Correspondence: Minghang Wang, The First Affiliated Hospital of Henan University of CM, National Regional TCM (Lung Disease) Diagnostic and Treatment Center, Renmin Road, Zhengzhou, People's Republic of China, Email wmh107hn@163.com

**Purpose:** Chronic obstructive pulmonary disease (COPD) is among the three leading causes of death worldwide, with its prevalence, morbidity, and mortality rates increasing annually. Oxidative stress (OS) is a key mechanism in COPD development, making the identification of OS-related biomarkers beneficial for improving its diagnosis and treatment.

**Methods:** The genetic data from patients with COPD and controls were obtained from the Gene Expression Omnibus database to identify OS-related genes (OSRGs). Functional enrichment analysis was conducted using the Kyoto encyclopedia of genes and genomes signaling pathway and gene ontology (GO). Protein-protein interaction networks were constructed to identify the core genes, which were further evaluated using receiver operating characteristic (ROC) curves. Diagnostic models were developed based on the core genes. Besides, the correlation between the expression of the core genes and the immune cells was analyzed using single-sample gene set enrichment analysis. Drug-gene interactions were explored to predict target drugs, and related microribonucleic acid (miRNA) and transcription factors (TFs) were identified using miRNet.

**Results:** In this study, we identified 299 differential genes, including 16 OSRGs. Among these, five core genes—heat shock protein family A (Hsp70) member 1A (HSPA1A), glutamate-cysteine ligase modifier subunit, interleukin-1 beta (IL-1 $\beta$ ), intercellular adhesion molecule 1 (ICAM1), and glutamate-cysteine ligase catalytic subunit (GCLC)—were screened and validated using ROC curve analysis. The results of GO enrichment analysis were mainly focused on the OS response, the negative regulation of the exogenous apoptosis signaling pathway, and the regulation of the apoptosis signaling pathway. Additionally, 33 target drugs were predicted, including ofloxacin, cisplatin, and pegolimumab, among others. Meanwhile, the regulatory networks comprising 33 miRNAs related to the core genes and 38 TFs associated with HSPA1A, IL-1 $\beta$ , ICAM1, and GCLC were constructed. A diagnostic model based on the five genes was constructed and validated with an area under the curve of 0.981 (95% confidence interval: 0.941–1.000).

**Conclusion:** This study identifies potential biomarkers for diagnosing COPD, new potential targets, and new directions for drug development and treatment.

**Keywords:** oxidative stress, COPD, biomarkers, bioinformatics, diagnostic model

## Introduction

Chronic obstructive pulmonary disease (COPD) is a heterogeneous ailment characterized by progressively worsening airflow limitation, manifesting clinically through symptoms such as dyspnea, cough, and expectoration.<sup>1</sup> Due to its protracted, recurrent, and evolving nature, COPD has become a prevalent and refractory condition in clinical practice. According to the World Health Organization, the prevalence of COPD is projected to increase significantly over the next four decades, driven by increasing smoking rates in developing nations and aging populations in high-income countries. By 2060, deaths related to COPD and associated disorders are expected to exceed 5.4 million annually.<sup>1,2</sup> This increasing trajectory is intrinsically linked to continuous exposure to COPD risk factors and demographic senescence.<sup>3</sup> COPD

presents a formidable public health challenge but remains amenable to early preventive and therapeutic interventions. The research elucidates that the etiology of COPD is multifaceted, correlated with smoking, age, utilization of biomass fuels, and genetic predispositions. These risk elements contribute to airway inflammation, oxidative stress (OS) responses, cellular apoptosis, and a protease-antiprotease imbalance, all of which play a critical role in the disease's pathogenesis. The OS denotes a dynamic disequilibrium between oxidants and antioxidants precipitated by intrinsic or extrinsic stimuli, resulting in damage to DNA, lipids, and proteins, thereby inducing cellular apoptosis and tissue injury.<sup>4</sup> Chronic inflammation affecting the airways, lung parenchyma, and pulmonary vasculature constitutes characteristic alterations in COPD, where an upsurge in pulmonary reactive oxygen species (ROS) exacerbates pro-inflammatory cytokines, driving the recruitment of inflammatory cells, such as neutrophils and macrophages, augments ROS levels, perpetuating a malignant cycle of OS and chronic pulmonary inflammation. Moreover, OS significantly impacts cellular apoptosis, protease-antiprotease disequilibrium, and immune responses.<sup>5-7</sup> It catalyzes the release of chemokines such as C-C motif chemokine ligand 2, mobilizing dendritic cells and lymphocytes, which partake in immune reactions, thereby promoting autoantibody production and culminating in pulmonary damage. Advancing research has identified several OS-related biomarkers, including glutathione,<sup>8</sup> protein sulfhydryls,<sup>9</sup> malondialdehyde,<sup>10</sup> and 8-hydroxydeoxyguanosine.<sup>11</sup> While these biomarkers are primarily used in scientific research, they are yet to be widely adopted in clinical practice. Current therapeutic strategies for COPD focus on the administration of glucocorticoids and bronchodilators. The OS-targeted therapies, such as mucolytic agents, including N-acetylcysteine, have demonstrated effectiveness in mitigating acute exacerbations, enhancing patient quality of life, and prolonging survival.<sup>12</sup> Given the critical role of OS in COPD evolution and progression, further exploration into its mechanisms is imperative. This study exploits gene expression data related to COPD from the Gene Expression Omnibus (GEO) database to identify OS-related biomarkers. Using various databases, it performs a comprehensive analysis, establishes and validates a diagnostic model, and predicts relevant microribonucleic acid (miRNA) and targeted therapeutics. These findings aimed to offer potential targets and methods for diagnosing and treating COPD.

## Materials and Methods

### Data Sources

Gene expression datasets related to COPD were retrieved from the GEO database (<https://www.ncbi.nlm.nih.gov/>)<sup>13</sup> using "COPD" as the search term. This search yielded three datasets: GSE130928, GSE11906, and GSE11952. The GSE130928 was designated as the training set, comprising gene expression profiles from alveolar macrophages in bronchoalveolar lavage fluid from 22 patients with COPD and 24 with healthy controls. The GSE11906, serving as the validation set, contained gene expression profiles from the airway epithelium of 33 patients with COPD and 72 with healthy controls. Given that smoking is the primary risk factor for COPD and is closely associated with OS,<sup>14</sup> the GSE11952 was also utilized as a validation set, including gene expression profiles from small airway epithelia of 38 non-smokers (healthy controls) and 45 smokers. The OS-related genes (OSRGs) were acquired from the Human Gene Database (<https://www.genecards.org/>), totaling 878 genes ([Supplementary Table 1](#)). The above data were obtained from public databases, and their download and use were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Henan University of CM (Opinion No. 2024HL-536).

### Construction of Weighted Gene Co-Expression Network and Module Identification

The weighted gene co-expression network analysis (WGCNA) package (version 1.72-5) was used to conduct a WGCNA on the GSE130928 dataset to identify gene modules associated with COPD. Initial steps included clustering of samples to remove outliers and setting the optimal threshold to ensure the WGCNA gene network adhered to a scale-free topology assumption. Based on the similarity of gene expression, genes were categorized into modules, with free genes (genes with low or no change in expression) being discarded. Subsequent association analyses between clinical traits and these modules yielded genesignificance (GS) scores and modulemembership (MM) values for each gene. Finally, modules with a correlation ( $|Cor|$ ) > 0.4 were selected as key modules for further analysis.

## Identification and Functional Enrichment Analysis of Candidate Genes Associated with OS in COPD

Intersections were conducted among differentially expressed genes (DEGs), OS genes, and genes from key modules to identify OSRGs using the Venn diagram package (version 1.7.3). Subsequently, the cluster profiler package (version 4.10.1) was used for gene ontology (GO) biological function enrichment and Kyoto encyclopedia of genes and genomes (KEGG) pathway analyses of the candidate genes. A significance threshold of  $P < 0.05$  was set to filter and visualize enriched biological functions or pathways. Moreover, the OmicCircos package (version 1.32.0) was utilized to depict the genomic locational distribution and expression levels of the candidate genes.

### Core Gene Selection and Identification

The STRING database (<https://cn.string-db.org/>) was applied to establish a protein-protein interaction network, and the network data were imported into the Cytoscape software (version 3.8.2). The cytoHubba plugin was used to analyze the network, determining node degrees to identify core genes. These genes were then validated through receiver operating characteristic (ROC) curve analysis to assess their discriminative power between COPD and control groups within the GSE130928 dataset, with genes achieving an area under the curve (AUC)  $> 0.7$  being designated as core genes. Relative expression levels of these genes in the test and control groups of the GSE130928 dataset were visualized using the ggpubr package (version 0.6.0).

### Immune Infiltration and Gene Set Enrichment Analysis (GSEA) of Core Genes

We conducted GSEA enrichment analysis using molecular feature databases To explore the functions of these genes (<https://www.gsea-msigdb.org/gsea/msigdb>). The downloaded c2. cp. kegg. Hs. symbols. gmt gene set was used as a reference for enrichment analysis of high and low expression samples, with  $P < 0.05$  as the screening condition to determine significantly enriched pathways. The first three pathways with normalized enrichment scores  $> 0$  and  $< 0$  were visualized.

Patients with COPD often have immune dysfunction.<sup>15</sup> We used single-sample GSEA (ssGSEA) to perform immune infiltration analysis on core genes and evaluate the correlation between core genes and 28 types of immune cells.

### Construction of Core Gene Regulatory Networks and Drug Prediction

The miRNET database<sup>16</sup> (<https://www.mirnet.ca/>) was used to predict microRNAs (miRNAs) and transcription factors (TFs) regulating core genes and elucidate the regulatory networks involving core genes and related signal transduction. Furthermore, potential therapeutic drugs for core genes were predicted using the Drug-gene interaction database<sup>17</sup> (<https://dgidb.genome.wustl.edu/>). Results were visualized using Cytoscape software.

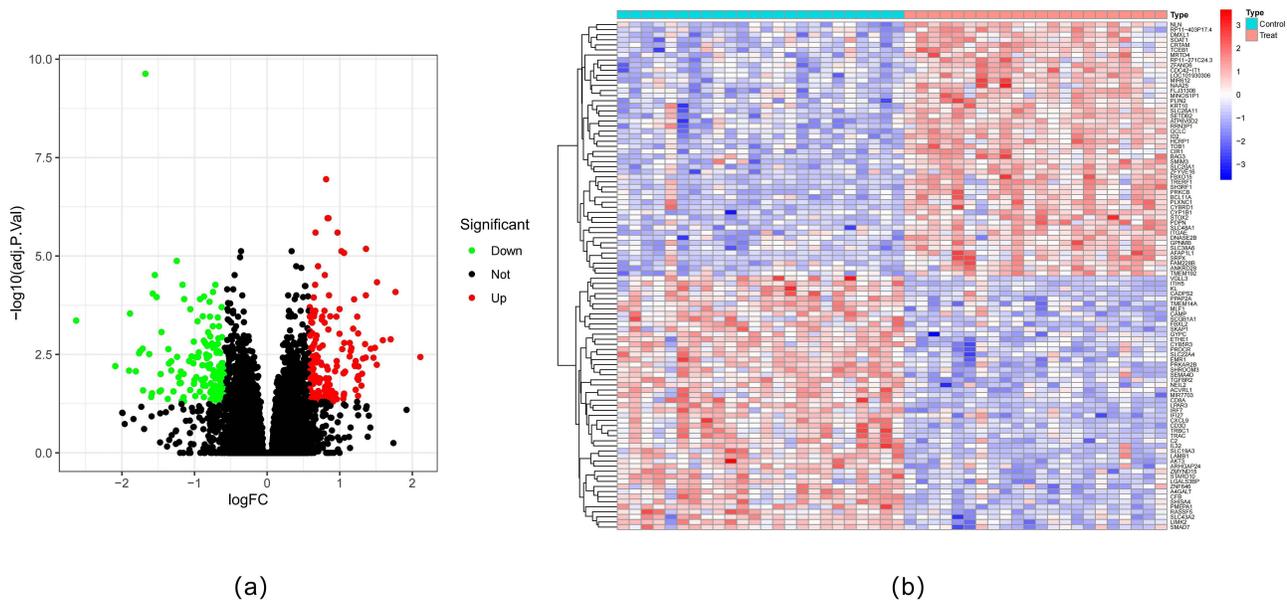
### Diagnostic Model Construction and Validation

Constructing a lasso regression diagnostic model using core genes related to OSRG. Gene expression data and gene lists were read, and genes included in the lists were selected from the expression data. The glmnet package (version 4.1-8) was used to apply the least absolute shrinkage and selection operator (LASSO) regression analysis to derive model construction coefficients. To validate the model's effectiveness, the GSE130928 dataset served as the training set to differentiate between disease and control groups using the model, evaluated using confusion matrix and ROC curve analyses. GSE11906 and GSE11952 datasets were used as validation sets to repeat the analysis and assess the model's validity.

## Results

### Identification of DEGs and WGCNA Network Modules

A total of 299 DEGs were identified in the examination of expression data from disease and control groups within the GSE130928 dataset, with 156 upregulated and 143 downregulated genes (Figure 1). To construct the WGCNA network, gene data from all samples were incorporated into a dendrogram, and outliers were excluded. This facilitated the creation

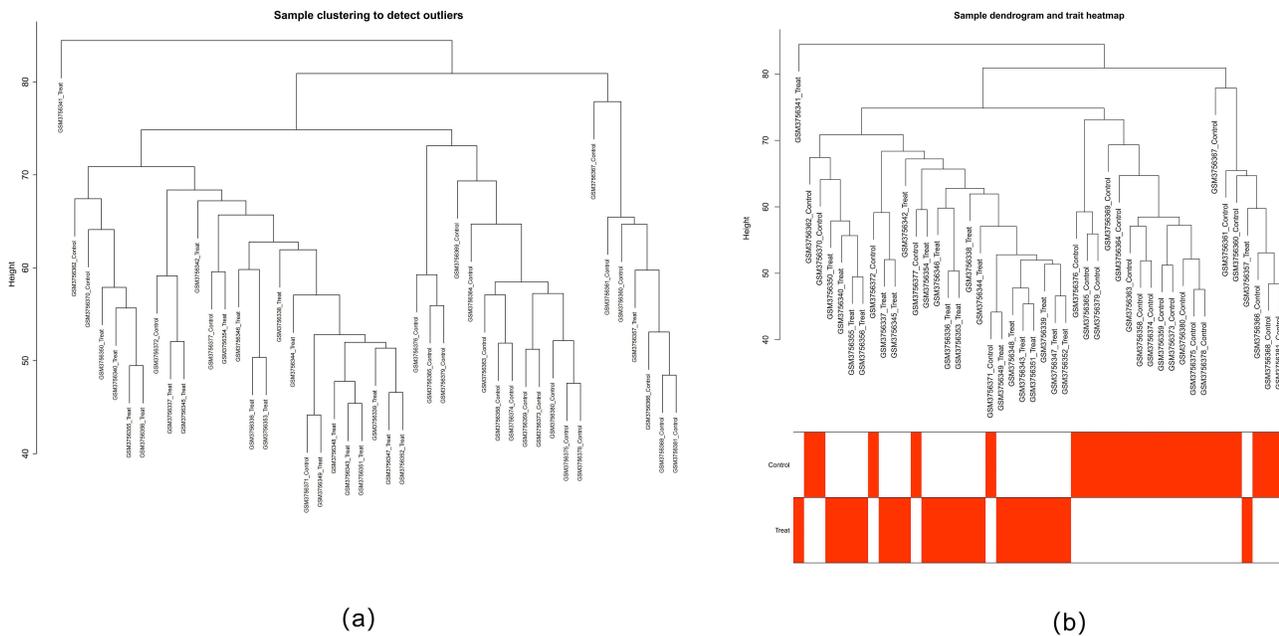


**Figure 1** (a) Differential gene volcano map; (b) Differential gene heatmap.

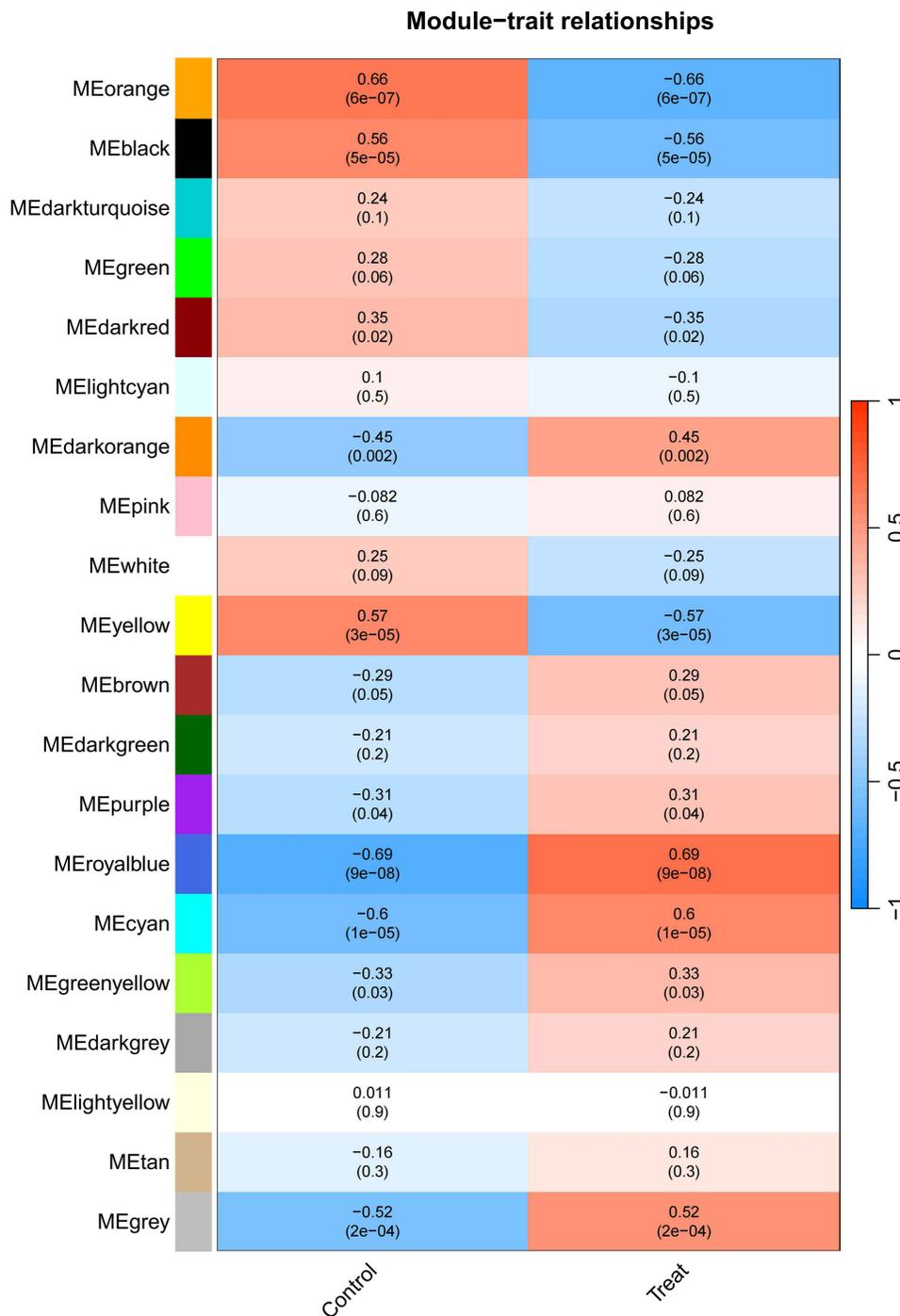
of a scale-free network ([Supplementary Figure 1](#)), followed by the division of genes into various modules based on expression similarity ([Figure 2](#)). A total of 20 modules were identified ([Supplementary Figure 2](#)). Subsequent correlation analysis with clinical expressions led to the selection of modules with an absolute correlation coefficient ( $|Cor|$ ) > 0.4. These modules were labeled with the colors orange, black, dark orange, yellow, sapphire, cyan, and grey ([Figure 3](#)).

### Screening and Functional Enrichment Analysis of OS-Related Candidate Genes

By intersecting 299 DEGs with 12,286 module genes and 878 OS genes, 16 OSRGs were identified ([Supplementary Table 2](#)). The GO enrichment analysis performed on these OSRGs yielded 734 results, including 645 biological processes (BPs) entries primarily related to OS responses, negative regulation of extrinsic apoptotic signaling pathway, and mediation of apoptotic signaling pathway; a total of 35 cellular component entries mainly involving aggresomes, stress

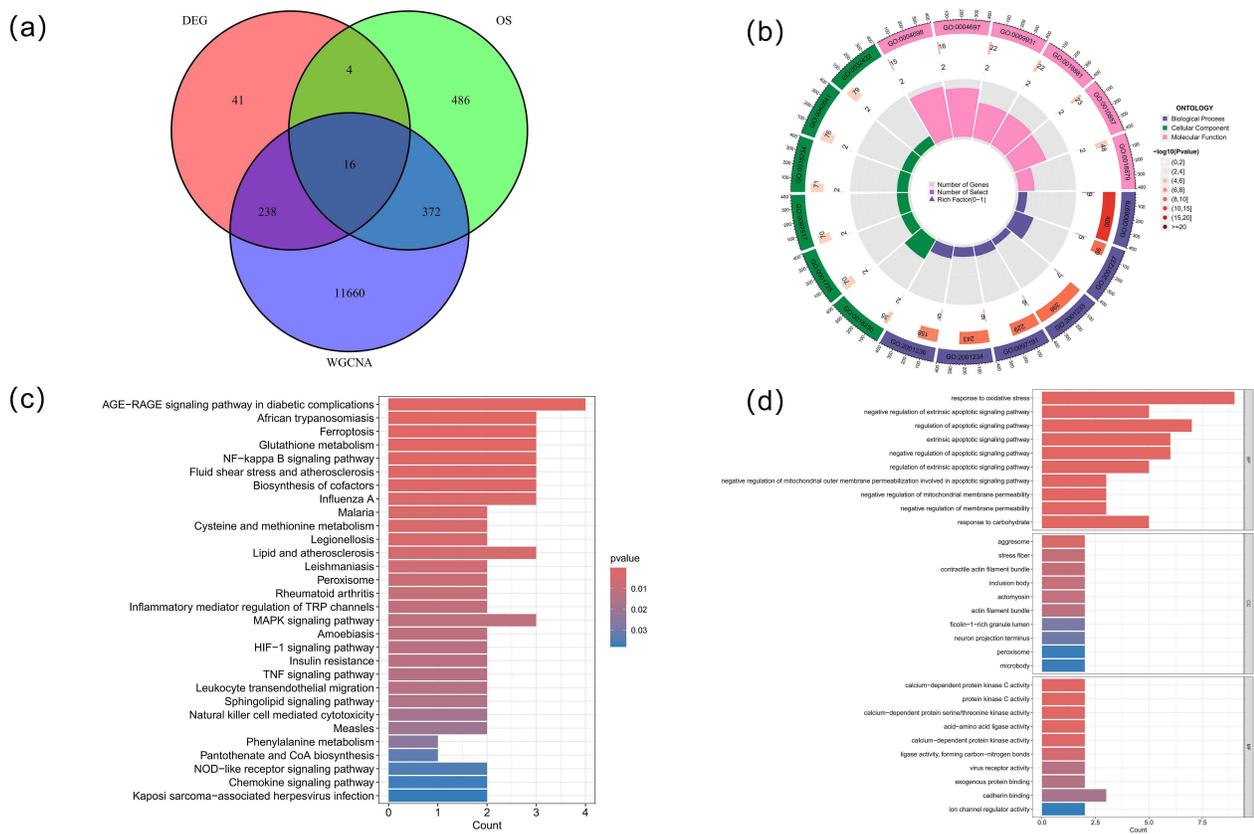


**Figure 2** (a) Gene trees of all samples after removing outliers; (b) Cluster Tree.

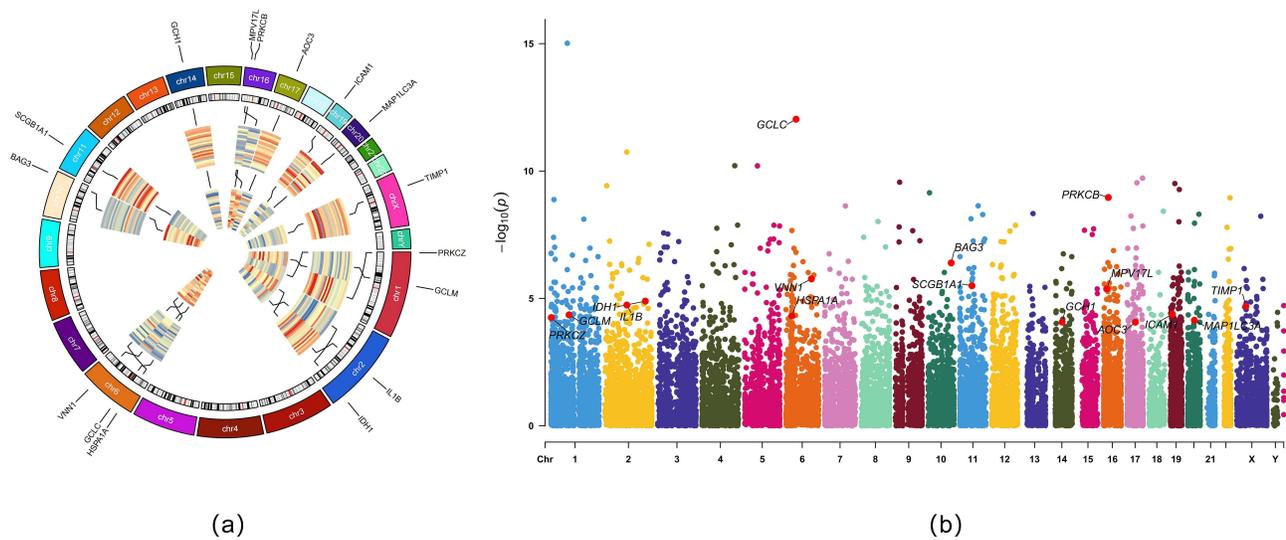


**Figure 3** Module and Disease Correlation.

fibers, inclusion bodies, and myosin complexes; and 54 molecular functions (MFs) entries primarily concerning calcium-dependent protein kinase C activity, protein kinase C activity, and calcium-dependent phosphotransferase activity. The KEGG pathway enrichment analysis was used to highlight 38 pathways, involving the AGE-RAGE signaling pathway in diabetic complications, ferroptosis, glutathione metabolism, and the nuclear factor kappa B (NF-κB) signaling pathway (Figure 4). Recent studies have demonstrated that gene loci are closely associated with COPD development and progression.<sup>18,19</sup> Establishing the chromosomal positions of relevant genes helps in understanding whether candidate genes influence COPD through genetic factors. The results indicated that the candidate genes are distributed across various chromosomes (Figure 5), with the most substantial number present on chromosome 6.



**Figure 4** (a) Venn diagram of intersected genes; (b) GO enrichment analysis pie chart; (c) KEGG pathway enrichment analysis diagram; (d) GO enrichment analysis bar chart.



**Figure 5** (a) Chromosomal positions of candidate genes. **Notes:** The expression levels of candidate genes in the GSE130928 dataset are represented in the internal circular heatmap. Red represents upregulation, blue represents downregulation, the outer circle of the heatmap represents the control group, and the inner circle represents the disease group; The outermost circle represents chromosomes, and lines from each gene point to their specific chromosomal positions. (b) Manhattan map of candidate genes.

### Core Gene Selection and Identification

The 16 OSRGs were identified through the screening. To screen for core genes, a protein interaction network of candidate genes was constructed using the STRING website, and network topology analysis was performed using cytoHubba

(Figure 6). Five core genes were identified using the degree value (Supplementary Figure 3): heat shock protein family A (Hsp70) member 1A (HSPA1A; located on chromosome 6), glutamate-cysteine ligase modifier subunit (GCLM; located on chromosome 1), interleukin-1 beta (IL-1 $\beta$ ; located on chromosome 2), intercellular adhesion molecule 1 (ICAM1; located on chromosome 19), and glutamate-cysteine ligase catalytic subunit (GCLC; located on chromosome 6).

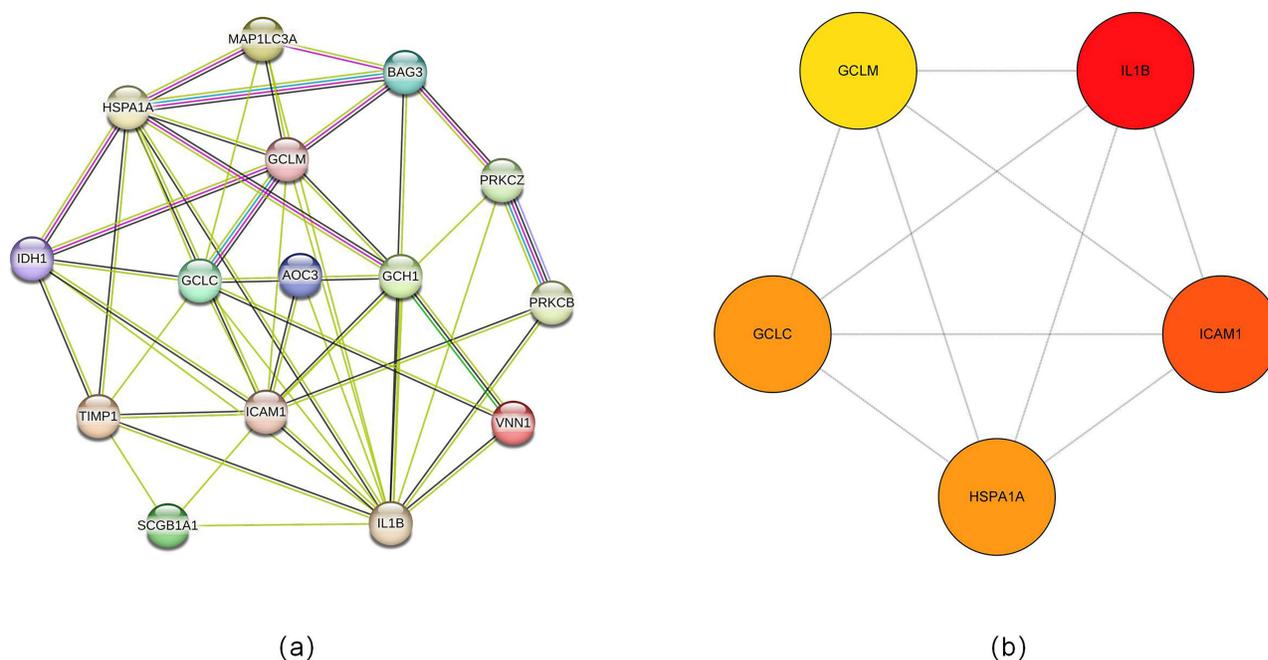
Subsequent expression analysis of these core genes within the training set indicated elevated expression levels in the COPD group for all five genes (Supplementary Figure 4). The ROC curve analysis yielded AUC values substantiating their diagnostic potential: HSPA1A (AUC = 0.843), GCLM (AUC = 0.841), IL-1 $\beta$  (AUC = 0.860), ICAM1 (AUC = 0.826), and GCLC (AUC = 0.979), all surpassing the threshold of 0.7, thus qualifying for further analytical scrutiny. Co-expression analysis demonstrated a positive correlation among all core genes, with particularly strong correlations between IL-1 $\beta$  and ICAM1, as well as between HSPA1A and GCLM (Figure 7).

## GSEA

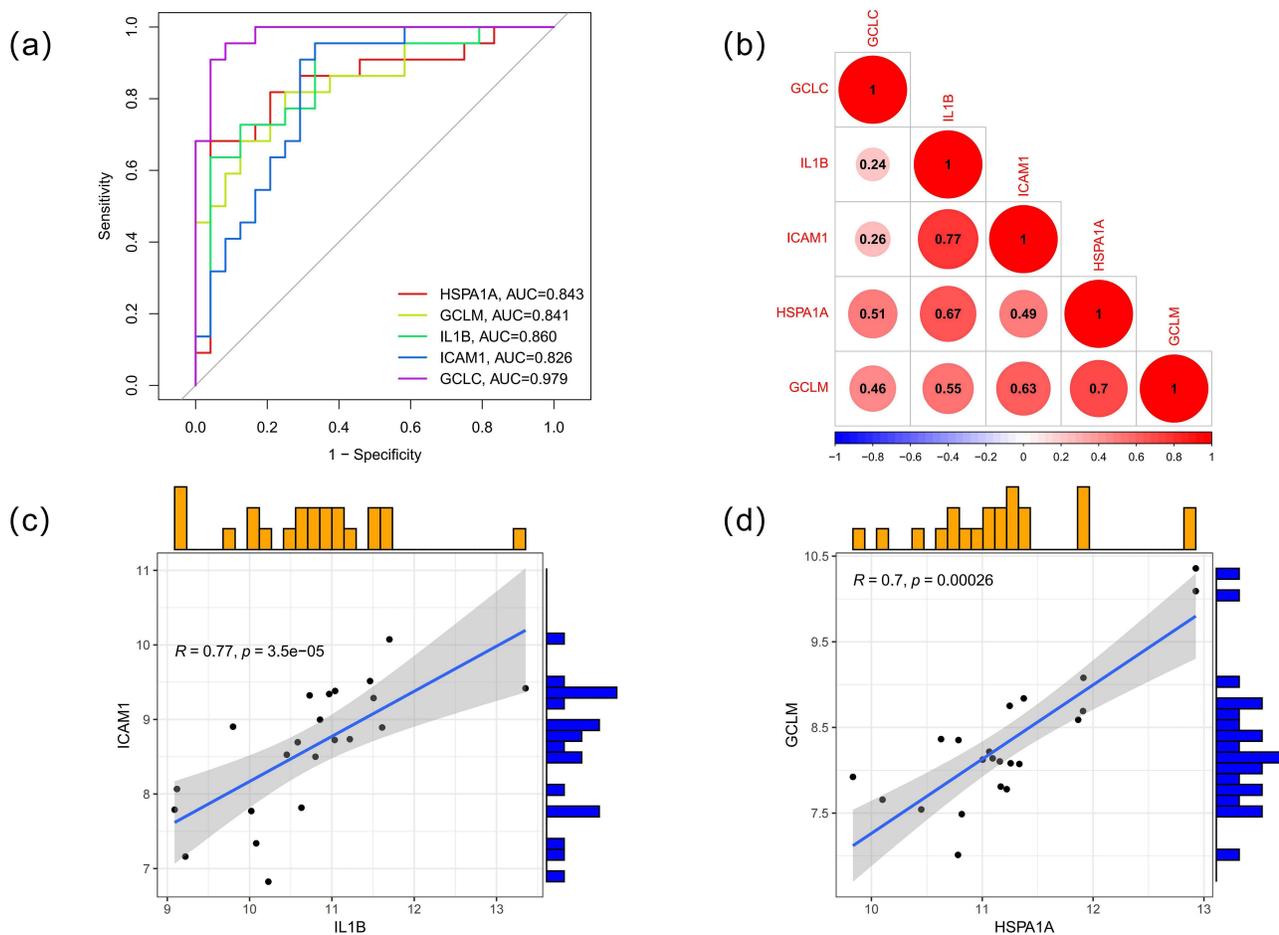
GSEA was conducted on the core genes to further explore their MFs and associated pathways. All five core genes were enriched in the “cytokine-cytokine receptor interaction” pathway with an upregulated expression profile. Additionally, GCLM, ICAM1, and IL-1 $\beta$  were upregulated in the “chemokine signaling pathway”. The “mitogen-activated protein kinase (MAPK) signaling pathway” indicated upregulation for GCLM, ICAM1, and IL-1 $\beta$ , whereas GCLC was downregulated in this pathway. In the “Parkinson’s disease pathway”, GCLC, GCLM, and IL-1 $\beta$  were all downregulated. Furthermore, low expression levels of HSPA1A, ICAM1, and IL-1 $\beta$  were associated with the “ribosome” pathway (Figure 8).

## Immune Cell Infiltration Scoring and Core Gene Correlation Analysis

A delineation of associations between pivotal genes and diverse immune cellular phenotypes was established using immunocyte infiltration scores. The HSPA1A was used to manifest significant positive associations with adipocyte-like mast cells, nascent B cells, plasmacytoid dendritic cells, central memory CD4<sup>+</sup> T cells, activated CD4<sup>+</sup> T cells, type 2 helper T cells, and gamma-delta T cells while exhibiting a notable negative correlation with activated B cells. The GCLM displayed pronounced positive correlations with adipocyte-like mast cells, activated CD4<sup>+</sup> T cells, and



**Figure 6** (a) Network diagram of candidate gene protein interactions; (b) Core gene network diagram.

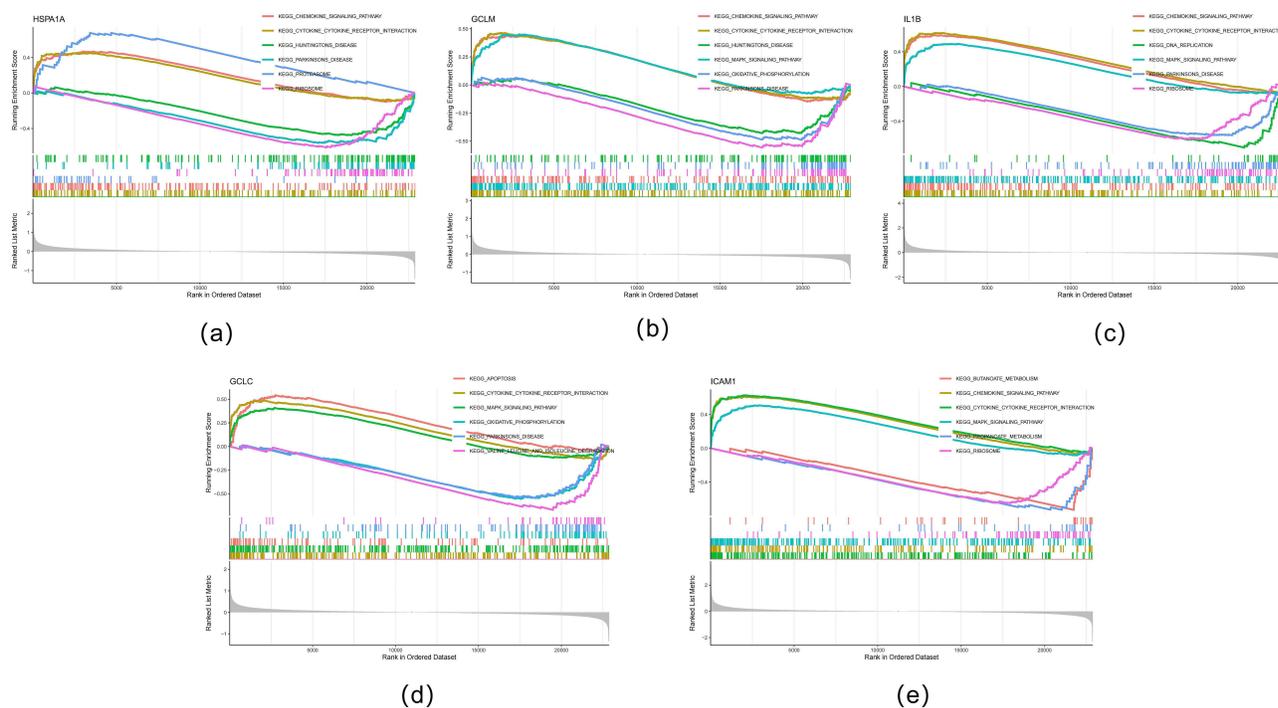


**Figure 7** (a) ROC curve analysis of core genes; (b) The correlation between core genes and expression levels; (c) IL-1 $\beta$  Correlation with ICAM1 expression; (d) Correlation between HSPA1A and GCLM expression.

plasmacytoid dendritic cells and was inversely related to regulatory T cells (Tregs), T follicular helper cells, activated B cells, macrophages, and myeloid-derived suppressor cells. The IL-1 $\beta$  was positively correlated with adipocyte-like mast cells, cytotoxic natural killer T cells, activated CD4+ T cells, natural killer cells, nascent B cells, plasmacytoid dendritic cells, and type 17 helper T cells (Th17), and negatively with regulatory T cells. The ICAM1 correlated positively with adipocyte-like mast cells, monocytes, plasmacytoid dendritic cells, activated CD4+ T cells, natural killer cells, activated dendritic cells, central memory CD4+ T cells, and Th17 cells. It demonstrated negative correlations with regulatory T cells, activated B cells, and immature dendritic cells (Figure 9). The GCLC did not exhibit notable correlations with any immune cells. These findings underscored the significant relationships between core genes and a spectrum of T cells, B cells, dendritic cells, and mast cells, corroborating previous research on their roles in COPD pathogenesis.<sup>20</sup>

### Core Gene Predictive Drug, miRNAs, and TFs Analysis

The Dgidb database was used to facilitate the prognostication of 33 pharmacological agents targeting these core genes (Supplementary Table 3). Specifically, 20 agents were predicted for IL-1 $\beta$ , including erythromycin, aspirin, hydrocortisone, ofloxacin, cephalexin, and gevokizumab; 5 agents including pegolimab and BI-505, were predicted for ICAM1; 4 agents including cisplatin and sulfamethoxazole were foreseen for GCLC; and predictions for GCLM and HSPA1A involved 3 and 1 agents respectively. The miRNet database was employed to predict 33 miRNAs (Supplementary Table 4) and 37 TFs (Supplementary Table 5) impacting these core genes, excluding GCLM (Figure 10). For instance,



**Figure 8** Gene set enrichment analysis (a) HSPA1A; (b) GCLM; (c) IL-1 $\beta$ ; (d) ICAM1; (e) GCLC.

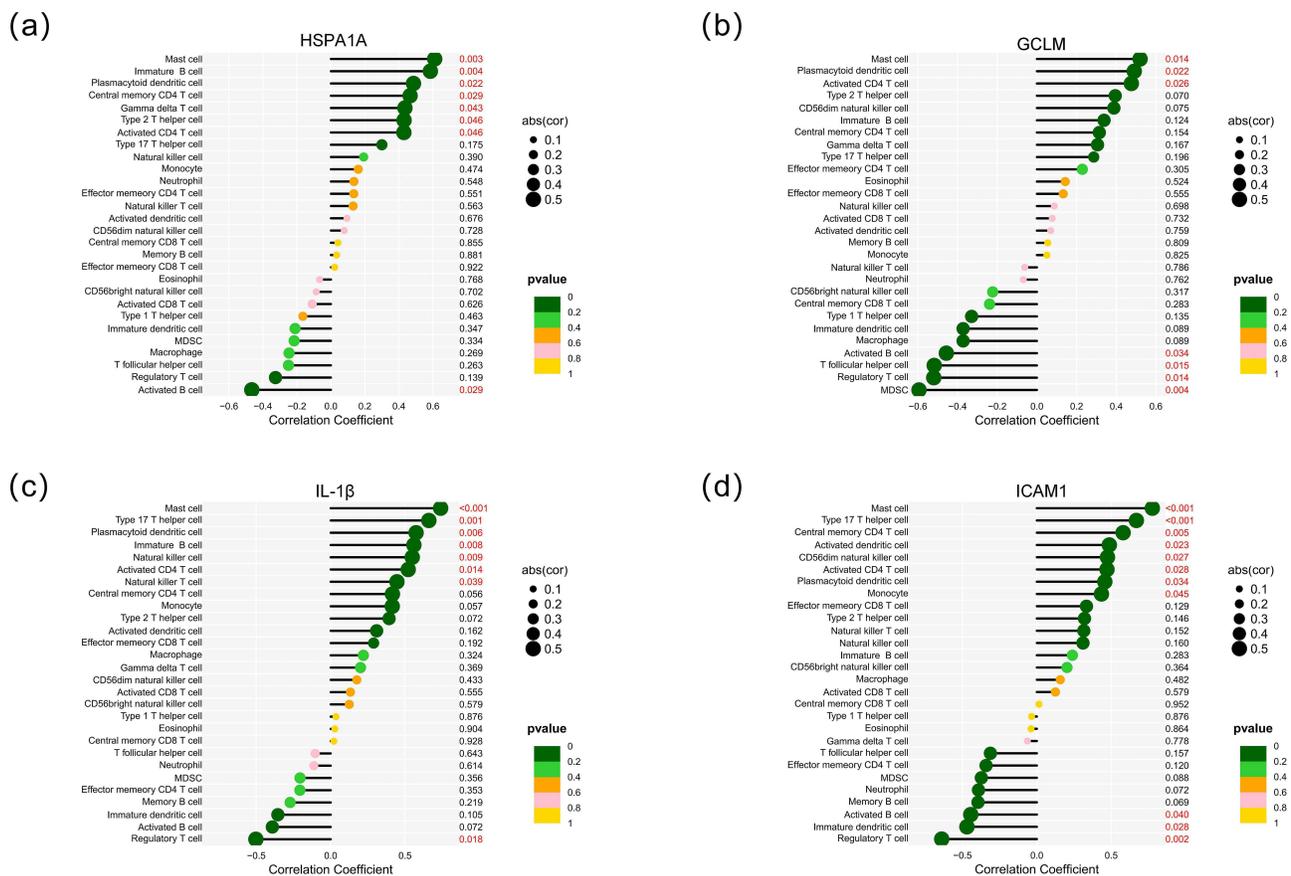
miR-335-5p was identified as a regulator for HSPA1A, ICAM1, and GCLC; TFs such as NF- $\kappa$ B1 and RelA were shared regulators for GCLC, ICAM1, and IL-1 $\beta$ .

## Diagnostic Model Establishment and Validation for Core Genes

A diagnostic model predicated on these five core genes was constructed using LASSO regression and ROC analysis, with the trajectory of regression coefficients depicted. Subsequent validation through the ROC curve evidenced an AUC of 0.981, affirming the model's diagnostic precision. Besides, a confusion matrix demonstrated an accuracy of 0.93, precision of 0.96, recall rate of 0.92, and specificity of 0.95 (Figure 11). To corroborate this model, GSE11906 served as an external validation set for disease, indicating an AUC of 0.949, with the confusion matrix reporting accuracy of 0.94, a precision of 0.92, a recall rate of 1.00, and a specificity of 0.81. The GSE11952, used as an external validation set for risk factors, indicated an AUC of 0.906, with the confusion matrix revealing an accuracy of 0.87, precision of 0.85, recall rate of 0.87, and a specificity of 0.87 (Figure 12). These metrics underscore the model's general applicability and potential as a pre-diagnostic tool for diseases.

## Discussion

COPD is a common respiratory system disease, with OS as its primary pathogenic mechanism. Currently, clinical interventions targeting OS are limited, and there is a scarcity of relevant diagnostic biomarkers. Early diagnosis or screening of COPD remains a challenging issue in clinical settings. With the recent escalation in environmental pollution, including smoke from cigarettes and cooking oils, and PM<sub>2.5</sub>, there has been an increase in both exogenous ROS caused by these pollutants and endogenous ROS produced by lung inflammation and structural cells. This increase results in damage to the airway epithelial cells, induces the proliferation of smooth muscle cells, leads to a decline in lung function, and makes the OS response increasingly significant in COPD development and progression.<sup>21</sup> As a result, selecting biomarkers related to OS and the development of drugs targeting these markers might become hotspots in future COPD diagnosis and treatment research. In this study, we identified 16 biomarkers related to OS through the screening of DEGs in the GSE130928 dataset. The GO enrichment analysis indicated that these OSRGs are mainly involved in BPs such as OS responses, regulation of apoptotic signaling pathways, and protein kinase activity. Previous research has revealed that

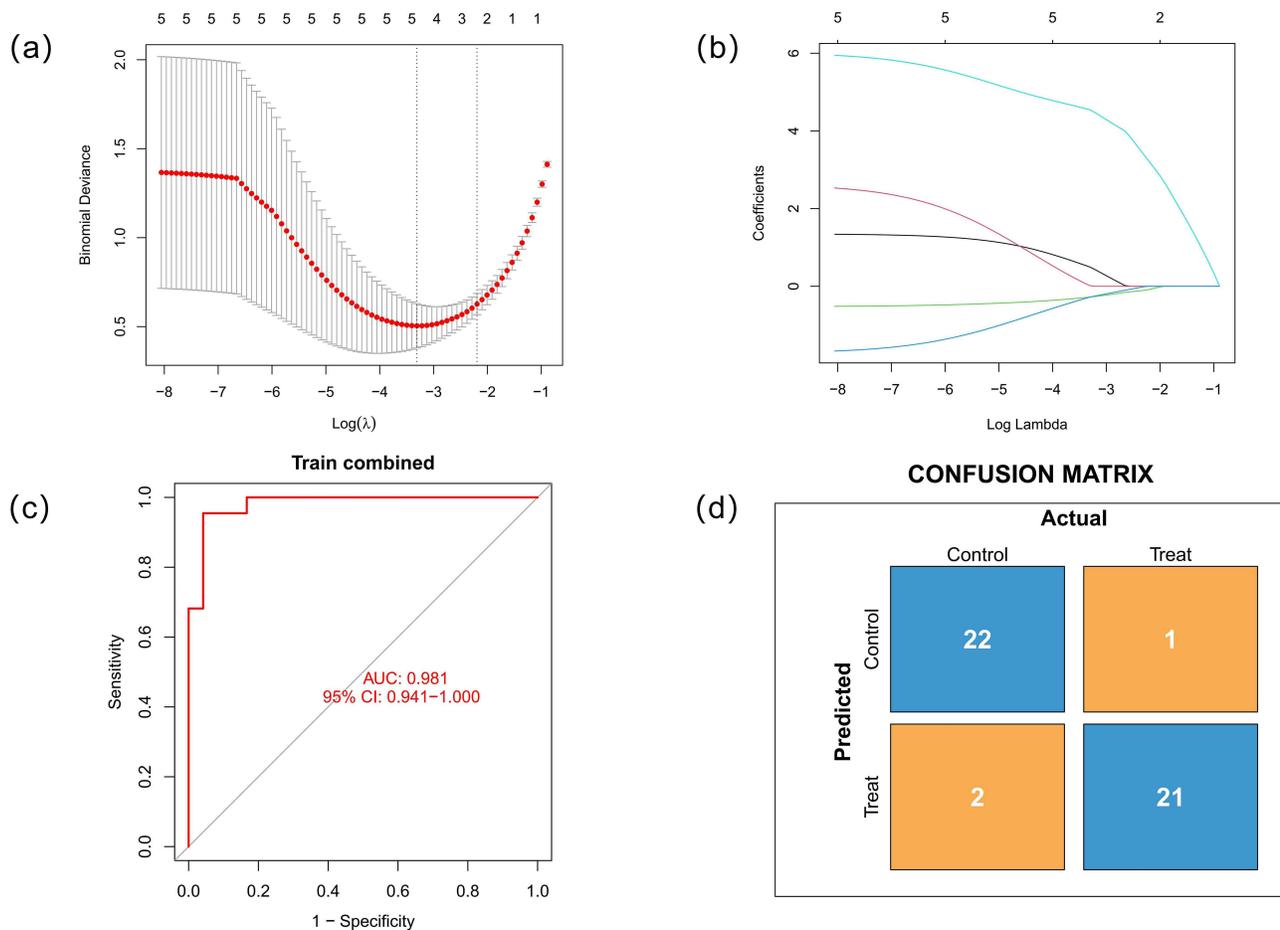


**Figure 9** Correlation between immune cell infiltration score and core genes (a) HSPA1A; (b) GCLM; (c) IL-1β; (d) ICAM1. **Note:** The highlighted red part on the vertical axis in the figure represents the correlation between genes and this type of immune cell (P<0.05).

an enhanced OS response intensifies the process of cellular apoptosis.<sup>22</sup> The KEGG pathway enrichment analysis was used to demonstrate that OSRGs are primarily involved in pathways, including the AGE-RAGE signaling pathway related to complications in diabetes, ferroptosis, glutathione metabolism, and the NF-κB signaling pathway. These pathways are highly associated with OS, such as the activation of the NF-κB pathway, which induces the expression of various pro-inflammatory cytokines such as tumor necrosis factor-alpha, IL-6, and IL-1β, thereby promoting apoptosis and the OS process.<sup>23</sup> To further analyze the functions of OSRGs, we determined their chromosomal positions and observed that OSRGs are related to multiple chromosomes. Relevant studies indicate that the X chromosome is associated with COPD-related phenotypes, and a variant near TMSB4X, rs5979771, reveals genome-wide significance with lung function.<sup>19</sup> Our results indicate that most genes are located on chromosome 6, and previous studies have also confirmed that certain single nucleotide polymorphisms (SNPs) on chromosome 6 are consistently associated with early susceptibility to COPD.<sup>24</sup> These findings suggest that these chromosomes may be significantly related to COPD, and identifying COPD-related SNPs also has certain clinical significance.

Through further selection, we identified five core OSRGs: HSPA1A, GCLM, IL-1β, ICAM1, and GCLC. Based on these genes, we explored their potential MFs through GESA, immune infiltration scoring, miRNAs, and TF predictions while developing a diagnostic model predicated on these core OSRGs, characterized by high sensitivity and selectivity. HSPA1A, GCLM, and GCLC consistently exhibited elevated expression levels in COPD samples. The HSPs form a superfamily whose proteins may increase in response to cellular stress associated with pollutants. Previously reported SNPs within HSP genes are linked to the risk and severity of COPD, and intracellular HSP levels may vary with different external exposures.<sup>25</sup> Moreover, a study involving coal miners indicated that elevated plasma levels of HSPA1A could be associated with an increased risk of COPD among these workers.<sup>26</sup> GCLM and GCLC are part of the glutathione



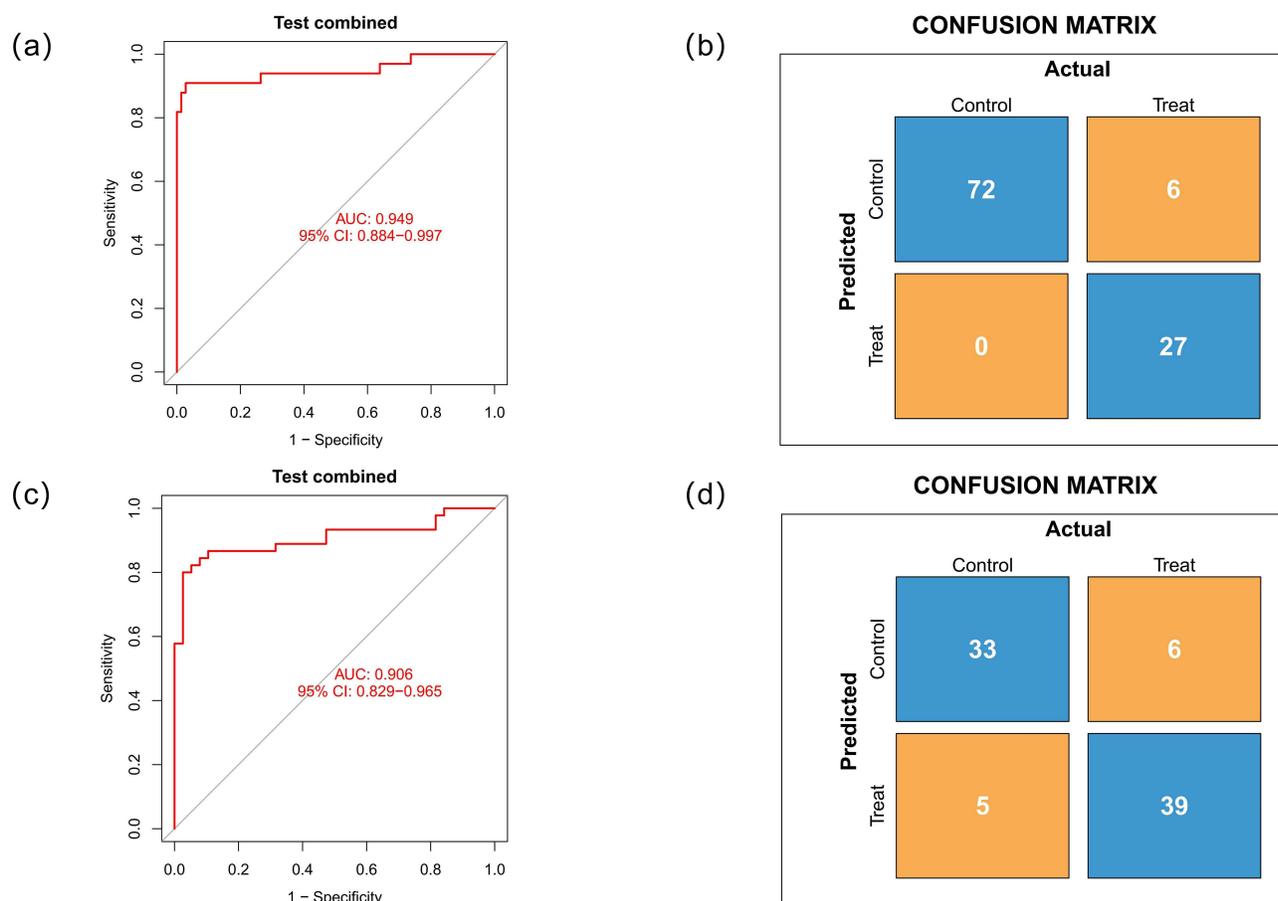


**Figure 11** Establishment and validation of LASSO regression model; (a) 10 fold cross validation chart; (b) Regression coefficient path diagram; (c) Training set ROC curve; (d) Training set confusion matrix.

mitigation of OS still requires further investigation.<sup>37,38</sup> Consequently, it is necessary to develop new antioxidants or targeted drugs to provide new options for the clinical treatment of COPD. We predicted 33 miRNAs related to COPD, such as miR-203a-3p, which is highly expressed in patients with COPD and smokers and is associated with basal cell proliferation.<sup>39</sup> The miR-221-3p can alleviate cell apoptosis and inflammatory responses in COPD.<sup>40</sup> Among TFs and NF-κB, a dimeric TF involved in inflammation, immune response, and cell proliferation belongs to the Rel protein family. Another member of this family, RelA, was also predicted, and both are implicated in regulating genes involved in the COPD process, a mechanism that has been confirmed in past studies.<sup>41,42</sup> The diagnostic model constructed based on the core genes demonstrated high accuracy and sensitivity in distinguishing patients from controls in both training and validation datasets. Additionally, attempts to use this model to differentiate between smokers and non-smokers among healthy individuals demonstrated high accuracy, which might be linked to the OS induced by smoking. Therefore, this model may have clinical value for diagnosing COPD and the potential for early screening of the disease.

## Conclusion

In this study, we utilized bioinformatics tools to identify five genes highly associated with OS during COPD progression. We comprehensively unveiled the relationship between these genes and the disease, along with their potential for clinical application translation through subsequent in-depth analysis of MFs, BPs, immune infiltration, drug predictions, miRNAs, and TFs, as well as the construction of a diagnostic model. However, further experimental validation is still required to clarify the OS functions related to COPD, including the mechanisms of many upstream regulatory miRNAs that remain undefined. Moreover, the function of this diagnostic model to differentiate between smokers and non-smokers still needs further confirmation and



**Figure 12** (a) ROC curve of disease validation set; (b) Disease validation set confusion matrix; (c) Risk factor validation set ROC curve; (d) Risk factor validation set confusion matrix.

validation. Overall, while the steps in this study were relatively comprehensive, there are shortcomings, such as the small sample size of the dataset and the limited number of OSRGs identified. Nonetheless, the high sensitivity and accuracy of the model provide new references for COPD clinical diagnosis and offer direction for subsequent research and the development of related pharmaceuticals.

## Data Sharing Statement

All relevant data are within the manuscript and its Additional files.

## Ethical Approval

GEO database (<https://www.ncbi.nlm.nih.gov/>) belong to public databases. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. Our study is based on open source data, their download and use were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Henan University of CM (Opinion No. 2024HL-536), so there are no ethical issues.

## Funding

This work was supported by the [National Key Research and Development Program of China] (2023YFC3502602/2023YFC3502600); the [Henan Provincial University Science and Technology Innovation Team] (23IRTSTHN027); the [National Clinical Research Base of Traditional Chinese Medicine Research Special Project of China] (2022JDZX046); the [Henan Province Science and Technology Research Project] (232102310472); and the [Henan Province Traditional Chinese Medicine Science Research Special Project] (2022ZY1047).

## Disclosure

The authors declare no competing interests in this work.

## References

1. Global initiative for chronic obstructive lung disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: 2024 REPORT. 2023. Available from: <https://goldcopd.org/2024-gold-report/>. Accessed March 25, 2025.
2. Sin DD, Doiron D, Agustí A, et al. Air pollution and COPD: GOLD 2023 committee report. *Eur Respir J*. 2023;31(5):2202469. doi:10.1183/13993003.02469-2022
3. Safiri S, Carson-Chahhoud K, Noori M, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990–2019: results from the Global Burden of Disease Study 2019. *BMJ*. 2022;378:e069679. doi:10.1136/bmj-2021-069679
4. Magallón M, Navarro-García MM, Dasi F. Oxidative Stress in COPD. *J Clin Med*. 2019;8(11):1953. doi:10.3390/jcm8111953
5. Lan X, Lederman R, Eng JM, et al. Nicotine induces podocyte apoptosis through increasing oxidative stress. *PLoS One*. 2016;11(12):e0167071. doi:10.1371/journal.pone.0167071
6. Tuleta I, Stöckigt F, Juergens UR, et al. Intermittent hypoxia contributes to the lung damage by increased oxidative stress, inflammation, and disbalance in protease/antiprotease system. *Lung*. 2016;194(6):1015–1020. doi:10.1007/s00408-016-9946-4
7. Upadhyay S, Vaish S, Dhiman M. Hydrogen peroxide-induced oxidative stress and its impact on innate immune responses in lung carcinoma A549 cells. *Mol Cell Biochem*. 2019;450(1–2):135–147. doi:10.1007/s11010-018-3380-2
8. Ben Anes A, Ben Nasr H, Garrouche A, et al. The Cu/Zn superoxide dismutase +35A/C (rs2234694) variant correlates with altered levels of protein carbonyls and glutathione and associates with severity of COPD in a Tunisian population. *Free Radic Res*. 2019;53(3):293–303. doi:10.1080/10715762.2019.1572888
9. Zinellu E, Zinellu A, Fois AG, Carru C, Pirina P. Circulating biomarkers of oxidative stress in chronic obstructive pulmonary disease: a systematic review. *Respir Res*. 2016;17(1):150. doi:10.1186/s12931-016-0471-z
10. Sarangi R, Varadhan N, Bahinipati J, Dhinakaran A, Anandaraj, Ravichandran K. Serum uric acid in chronic obstructive pulmonary disease: a hospital based case control study. *J Clin Diagn Res*. 2017;11(9):BC09–BC13. doi:10.7860/JCDR/2017/29300.10605
11. Liu X, Deng K, Chen S, et al. 8-Hydroxy-2'-deoxyguanosine as a biomarker of oxidative stress in acute exacerbation of chronic obstructive pulmonary disease. *Turk J Med Sci*. 2019;49(1):93–100. doi:10.3906/sag-1807-106
12. Kolarov V, Kotur Steviljević J, Ilić M, et al. Factorial analysis of N-acetylcysteine and propolis treatment effects on symptoms, life quality and exacerbations in patients with Chronic Obstructive Pulmonary Disease (COPD): a randomized, double-blind, placebo-controlled trial. *Eur Rev Med Pharmacol Sci*. 2022;26(9):3192–3199. doi:10.26355/eurev\_202205\_28737
13. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 2013;41(Database issue):D991–D995. doi:10.1093/nar/gks1193
14. Buculei I, Dobrin ME, Matei D, et al. Polycyclic aromatic hydrocarbons induced by smoking and air pollution: correlation with oxidative stress in chronic obstructive pulmonary disease patients. *Toxics*. 2022;10(11):681. doi:10.3390/toxics10110681
15. Bu T, Wang LF, Yin YQ. How do innate immune cells contribute to airway remodeling in COPD progression? *Int J Chron Obstruct Pulmon Dis*. 2020;15:107–116. doi:10.2147/COPD.S235054
16. Chang L, Zhou G, Soufan O, Xia J. miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res*. 2020;48(W1):W244–W251. doi:10.1093/nar/gkaa467
17. Cotto KC, Wagner AH, Feng YY, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res*. 2018;46(D1):D1068–D1073. doi:10.1093/nar/gkx1143
18. Nedeljkovic I, Carnero-Montoro E, Lahousse L, et al. Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics. *Eur J Hum Genet*. 2018;26(5):709–722. doi:10.1038/s41431-017-0089-8
19. Hayden LP, Hobbs BD, Busch R, et al. X chromosome associations with chronic obstructive pulmonary disease and related phenotypes: an X chromosome-wide association study. *Respir Res*. 2023;24(1):38. doi:10.1186/s12931-023-02337-1
20. Wang D, Chen B, Bai S, Zhao L. Screening and identification of tissue-infiltrating immune cells and genes for patients with emphysema phenotype of COPD. *Front Immunol*. 2022;13:967357. doi:10.3389/fimmu.2022.967357
21. Kume H, Yamada R, Sato Y, Togawa R. Airway smooth muscle regulated by oxidative stress in COPD. *Antioxidants*. 2023;12(1):142. doi:10.3390/antiox12010142
22. Guan R, Yao H, Li Z, et al. Sodium tanshinone IIA sulfonate attenuates cigarette smoke extract-induced mitochondrial dysfunction, oxidative stress, and apoptosis in alveolar epithelial cells by enhancing SIRT1 pathway. *Toxicol Sci*. 2021;183(2):352–362. doi:10.1093/toxsci/kfab087
23. Li Z, Li L, Lv X, Hu Y, Cui K. Ginseng saponin Rb1 attenuates cigarette smoke exposure-induced inflammation, apoptosis and oxidative stress via activating Nrf2 and inhibiting NF-κB signaling pathways. *Int J Chron Obstruct Pulmon Dis*. 2023;18:1883–1897. doi:10.2147/COPD.S418421
24. Lee YJ, Choi S, Kwon SY, et al. A genome-wide association study in early COPD: identification of one major susceptibility loci. *Int J Chron Obstruct Pulmon Dis*. 2020;15:2967–2975. doi:10.2147/COPD.S269263
25. Ambrocio-Ortiz E, Pérez-Rubio G, Ramírez-Venegas A, et al. Effect of SNPs in HSP family genes, variation in the mRNA and intracellular Hsp levels in COPD secondary to tobacco smoking and biomass-burning smoke. *Front Genet*. 2020;10:1307.
26. Cui X, Xing J, Liu Y, et al. COPD and levels of Hsp70 (HSPA1A) and Hsp27 (HSPB1) in plasma and lymphocytes among coal workers: a case-control study. *Cell Stress Chaperones*. 2015;20(3):473–481. doi:10.1007/s12192-015-0572-5
27. Oit-Wiscombe I, Soomets U, Altraja A. Antioxidant glutathione analogues UPF1 and UPF17 modulate the expression of enzymes involved in the pathophysiology of chronic obstructive pulmonary disease. *Curr Issues Mol Biol*. 2024;46(3):2343–2354. doi:10.3390/cimb46030149
28. Cheng L, Liu J, Li B, Liu S, Li X, Tu H. Cigarette smoke-induced hypermethylation of the GCLC gene is associated with COPD. *Chest*. 2016;149(2):474–482. doi:10.1378/chest.14-2309
29. Bonacini M, Rossi A, Ferrigno I, et al. miR-146a and miR-146b regulate the expression of ICAM-1 in giant cell arteritis. *J Autoimmun*. 2024;144:103186. doi:10.1016/j.jaut.2024.103186

30. Zou Y, Chen X, Liu J, et al. Serum IL-1 $\beta$  and IL-17 levels in patients with COPD: associations with clinical parameters. *Int J Chron Obstruct Pulmon Dis.* 2017;12:1247–1254. doi:10.2147/COPD.S131877
31. Kastelein TE, Duffield R, Marino FE. Acute immune-inflammatory responses to a single bout of aerobic exercise in smokers; the effect of smoking history and status. *Front Immunol.* 2015;6:634. doi:10.3389/fimmu.2015.00634
32. Pelaia C, Vatrella A, Sciacqua A, Terracciano R, Pelaia G. Role of p38-mitogen-activated protein kinase in COPD: pathobiological implications and therapeutic perspectives. *Expert Rev Respir Med.* 2020;14(5):485–491. doi:10.1080/17476348.2020.1732821
33. Higham A, Singh D. Inhaled corticosteroid responses in COPD: do mast cells hold the answer? *Thorax.* 2023;78(4):323–324. doi:10.1136/thorax-2022-219534
34. Risso K, Guillouet-de-Salvador F, Valerio L, et al. COPD in HIV-infected patients: CD4 cell count highly correlated. *PLoS One.* 2017;12(1):e0169359. doi:10.1371/journal.pone.0169359
35. Uzeloto JS, de Toledo-Arruda AC, Silva BSA, et al. Systemic cytokine profiles of CD4+ T lymphocytes correlate with clinical features and functional status in stable COPD. *Int J Chron Obstruct Pulmon Dis.* 2020;15:2931–2940. doi:10.2147/COPD.S268955
36. Paplinska-Goryca M, Misiukiewicz-Stepien P, Nejman-Gryz P, et al. Epithelial-macrophage-dendritic cell interactions impact alarmins expression in asthma and COPD. *Clin Immunol.* 2020;215:108421. doi:10.1016/j.clim.2020.108421
37. Barnes PJ. Oxidative stress-based therapeutics in COPD. *Redox Biol.* 2020;33:101544. doi:10.1016/j.redox.2020.101544
38. Barnes PJ. Oxidative stress in chronic obstructive pulmonary disease. *Antioxidants.* 2022;11(5):965. doi:10.3390/antiox11050965
39. van Nijnatten J, Brandsma CA, Steiling K, et al. High miR203a-3p and miR-375 expression in the airways of smokers with and without COPD. *Sci Rep.* 2022;12(1):5610. doi:10.1038/s41598-022-09093-0
40. Yang H, Zhang L, Wang Q. MicroRNA-221-3p alleviates cell apoptosis and inflammatory response by targeting cyclin dependent kinase inhibitor 1B in chronic obstructive pulmonary disease. *Bioengineered.* 2021;12(1):5705–5715. doi:10.1080/21655979.2021.1967837
41. Alharbi KS, Fuloria NK, Fuloria S, et al. Nuclear factor-kappa B and its role in inflammatory lung disease. *Chem Biol Interact.* 2021;345:109568. doi:10.1016/j.cbi.2021.109568
42. Günes Günsel G, Conlon TM, Jeridi A, et al. The arginine methyltransferase PRMT7 promotes extravasation of monocytes resulting in tissue injury in COPD. *Nat Commun.* 2022;13(1):1303. doi:10.1038/s41467-022-28809-4

International Journal of Chronic Obstructive Pulmonary Disease

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>

**Dovepress**  
Taylor & Francis Group