



OPEN ACCESS

EDITED BY

Paul Higgins,
Atlantic Technological University, Ireland

REVIEWED BY

Mark Z. Kos,
The University of Texas Rio Grande Valley,
United States
Wanjun Gu,
University of California, San Francisco,
United States

*CORRESPONDENCE

Jinxia Li,
✉ lijinxia1569@163.com

†These authors have contributed equally to
this work

RECEIVED 22 September 2024

ACCEPTED 31 January 2025

PUBLISHED 05 March 2025

CITATION

Li J, Liu X and Liu Y (2025) Identification of four
key genes related to the diagnosis of chronic
obstructive pulmonary disease using
bioinformatics analysis.
Front. Genet. 16:1499996.
doi: 10.3389/fgene.2025.1499996

COPYRIGHT

© 2025 Li, Liu and Liu. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Identification of four key genes related to the diagnosis of chronic obstructive pulmonary disease using bioinformatics analysis

Jinxia Li^{*†}, Xiuming Liu[†] and Yonghu Liu

Department of Respiratory and Critical Care Medicine, General Hospital of Ningxia Medical University, Yinchuan, China

Introduction: Chronic obstructive pulmonary disease (COPD) is projected to become the third leading cause of death worldwide. Despite extensive research over the past few decades, effective treatments remain elusive, making disease prevention and control a global challenge.

Methods: This study aimed to identify diagnostic key genes for COPD. We utilized the Gene Expression Omnibus database to obtain gene expression data specific to COPD. Differentially expressed genes (DEGs) were identified and analyzed through Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and Gene Set Enrichment Analysis. Integrated weighted gene co-expression network analysis was employed to examine related gene modules. To pinpoint key genes, we used SVM-RFE, RF, and LASSO.

Results: A total of 1782 DEGs were discovered, many of which were enriched in various biological pathways and activities. Four key genes—*MRC1*, *BCL2A1*, *GYPC*, and *SLC2A3*—were identified. We observed a significant difference in immune infiltration between COPD and normal groups, indicating potential interactions between immune cells and these genes. The identified key genes were further validated using external datasets.

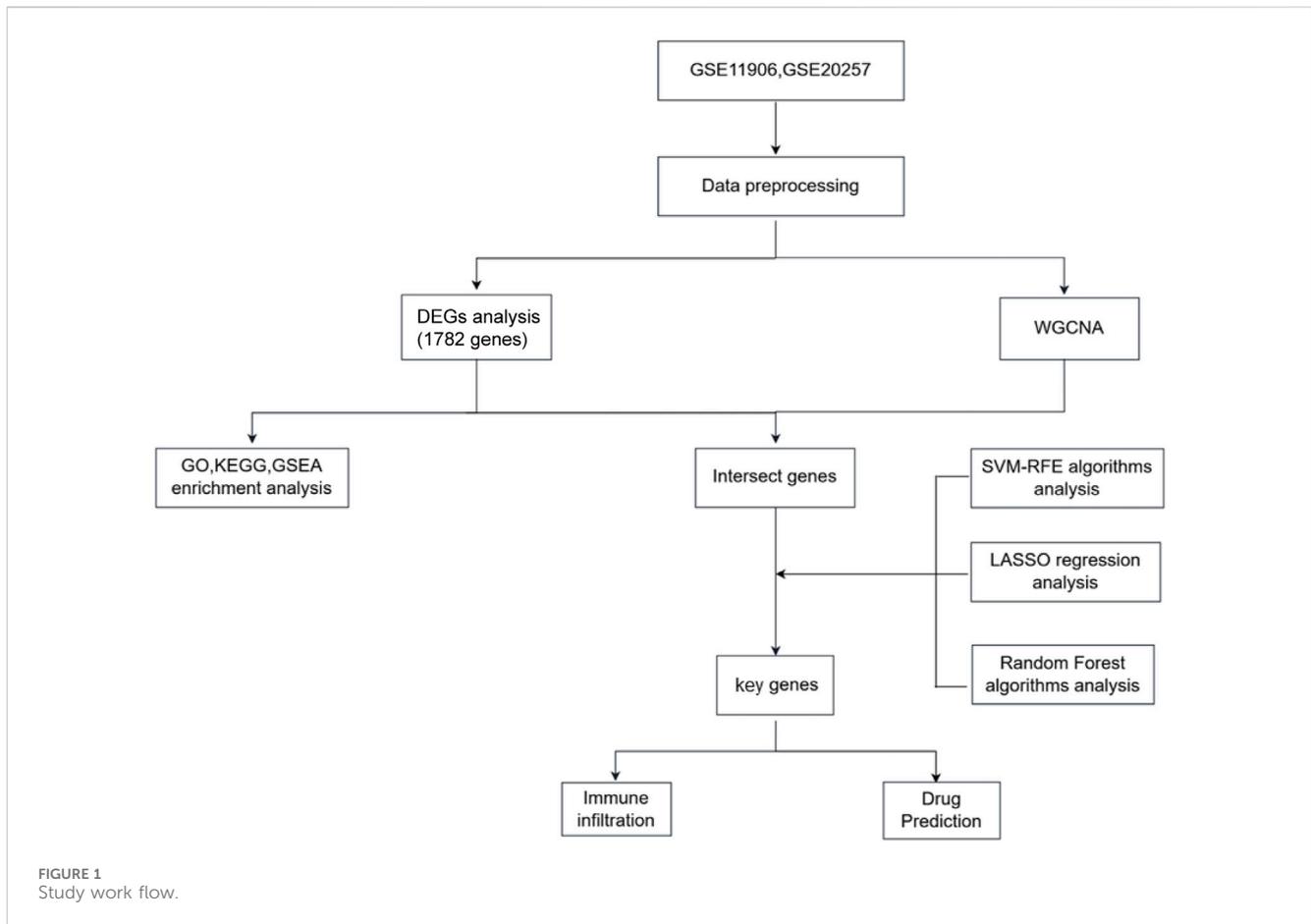
Discussion: Our findings suggest that *MRC1*, *BCL2A1*, *GYPC*, and *SLC2A3* are potential biomarkers for COPD. Targeting these diagnostic genes with specific drugs may potentially offer new avenues for COPD management; however, this hypothesis remains preliminary and requires further investigation, as the study does not directly assess therapeutic interventions.

KEYWORDS

COPD, enrichment analysis, machine learning, immune infiltration analysis, drug prediction

1 Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive lung condition marked by airflow limitation and chronic inflammation (McDonough et al., 2011; Vestbo et al., 2013). It results from a combination of genetic factors, such as α 1-antitrypsin deficiency, and environmental factors, particularly smoking (Leap et al., 2021). COPD is common and has high rates of disability and mortality, creating a significant economic burden worldwide (Iheanacho et al., 2020). Early diagnosis and treatment are crucial for slowing lung function decline and improving long-term outcomes. However, current diagnostic methods, such as



pulmonary function tests and imaging, are insufficient for detecting early-stage COPD, making accurate diagnosis challenging. This highlights the need to understand genetic differences between COPD patients and healthy individuals, identify high-risk markers, and find effective treatment targets.

In recent years, high-throughput sequencing and bioinformatics have become key tools in COPD research, helping identify disease-related genes and potential molecular targets for precision therapy. For example, genes like *HIF1A*, *CDKN1A*, *BAG3*, *ERBB2*, and *ATG16L1* influence COPD through autophagy regulation (Sun et al., 2021). However, the lack of objective diagnostic methods continues to make COPD diagnosis and treatment selection difficult. Therefore, developing reliable biomarkers for COPD is essential for improving treatment outcomes.

In this study, we analyzed gene expression data from four RNA-seq datasets (GSE11906, GSE20257, GSE5058, and GSE8545) containing airway epithelial cells from COPD patients and healthy individuals. Our goal was to identify gene expression changes involved in COPD and discover potential diagnostic biomarkers. We identified 1782 differentially expressed genes (DEGs) and key COPD-related modules through analysis of two Gene Expression Omnibus (GEO) datasets. Using algorithms like SVM-RFE, random forest (RF), and LASSO, we pinpointed four key genes—*MRC1*, *BCL2A1*, *GYPC*, and *SLC2A3*—that could improve COPD diagnosis in high-risk patients. Targeting

these genes with specific drugs may also enhance clinical management of COPD.

2 Materials and methods

2.1 Raw data acquisition

Datasets for four COPD airway tissues [GSE11906 (Raman et al., 2009), GSE20257 (Shaykhiev et al., 2011), GSE5058 (Carolan et al., 2006), and GSE8545 (Ammous et al., 2008)] were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). All datasets are gene expression arrays generated using the GPL570 (HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array.

GSE11906 and GSE20257 were used as the training set for airway tissue, the set contains 90 healthy and 28 COPD samples; While GSE5058 and GSE8545 were used as the validation set, the set contains 19 healthy and 21 COPD samples. The `normalizeBetweenArrays` function in the `limma` package (version 3.58.1) and `sva` (version 3.50.0) were applied for data combination and normalization. Probes not matching any known gene were eliminated, and if multiple probes matched a single gene, their average expression was calculated. The Perl programming language was used to remove lncRNA profiles and identify mRNA matrix files. The R package `ggplot2` (version 3.2.1) was employed to normalize the processed data. Detailed information about the

datasets is provided in [Supplementary Table 1](#), and the study's flow diagram is shown in [Figure 1](#).

2.2 Differentially expressed genes identification

Principal Coordinates Analysis (PCoA), a multivariate statistical method used to assess the similarity and dissimilarity between samples, was performed based on the Bray-Curtis distance metric. PCoA was performed to confirm that the genes could effectively differentiate between healthy individuals and COPD patients. A total of 22,836 genes were tested for differential expressions, from which 1,782 were identified as significantly differentially expressed genes (DEGs) using the limma R package. The cutoff criteria for DEGs were set to an adjusted P-value < 0.05 and $|\log \text{fold change (FC)}| > 0.5$. Heatmaps and volcano plots were generated using the ggplot2 package to visualize the results.

2.3 Enrichment analysis

To elucidate the biological implications of the identified genes and their functions, differentially expressed genes (DEGs) were subjected to both Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA).

For ORA, enrichment analyses were performed using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. This analysis was conducted on the 1,782 DEGs identified after correcting the $\log_2\text{FC}$ calculation error. Fisher's Exact Test was applied for statistical analysis, and the False Discovery Rate (FDR) method was used to control the false positive rate. The analysis was performed using the clusterProfiler R package, with a significant cutoff at a P-value of less than 0.05. The terms "Molecular Function" (MF), "Biological Process" (BP), and "Cellular Component" (CC) refer to categories within the Gene Ontology classification system.

For GSEA, enrichment of predefined gene sets was determined using the reference gene set "c2.cp.kegg.v6.2.symbols.gmt" from the Molecular Signature Database (MSigDB). Enrichment sets containing fewer than 10 or more than 200 genes were excluded from the analysis. Pathways with a normalized enrichment score (NES) greater than zero were considered upregulated, while those with an NES less than zero were considered downregulated. The five most significant pathways were identified with an FDR threshold of < 0.05 . The weighted Kolmogorov-Smirnov statistics were employed to calculate the enrichment score (ES), with genes ranked based on log fold change ($\log\text{FC}$) values.

2.4 Weighted gene co-expression network analysis

Data from GSE11906 and GSE20257 were combined and batch processed. Weighted gene co-expression network analysis (WGCNA) was used to identify trait-related modules. A

topological overlap matrix was constructed from the expression profiles, with a soft-thresholding power of 18 and a minimum module size of 30 to identify core modules. A height limit of 0.25 was set for module merging. Pearson's correlation test was then used to evaluate the modules, with a significance threshold of $P < 0.05$.

2.5 Support vector machine, random forest, and least absolute shrinkage and selection operator model construction

Candidate genes were identified by intersecting DEGs with genes from the WGCNA hub module. Hub genes were then classified by overlapping genes from the SVM-RFE method using the e1071 package (Noble, 2006), the RF algorithm using the randomForest R package (Paul et al., 2018), and the LASSO using the glmnet package (Vasquez et al., 2016). For Random Forest (RF), we set $n\text{tree} = 1,000$ and selected features with an importance score greater than 2. In LASSO, we used 10-fold cross-validation ($n\text{folds} = 10$) and set the regularization parameter $\alpha = 1$. For SVM-RFE, we applied 5-fold cross-validation ($k = 5$). These settings ensure the robustness and consistency of our results across different algorithms.

2.6 Immune infiltration analysis

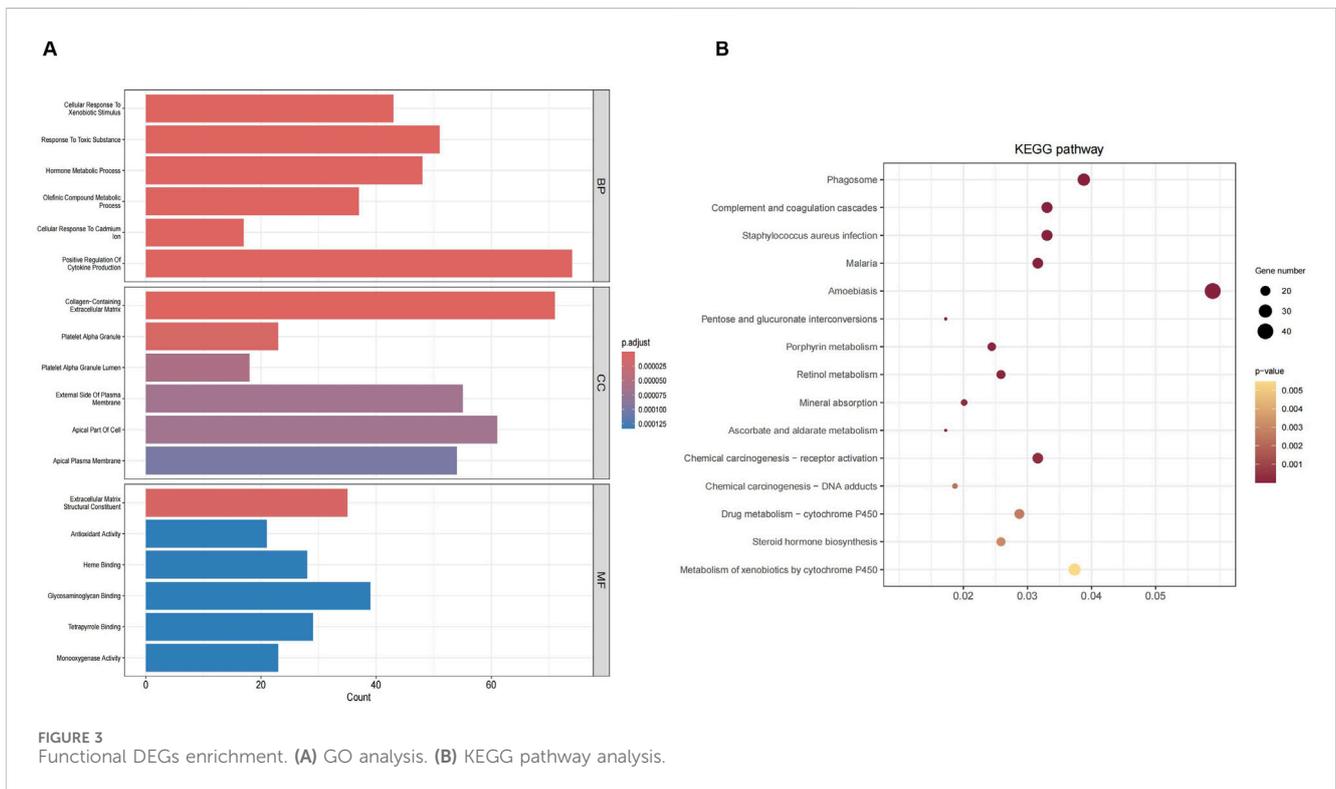
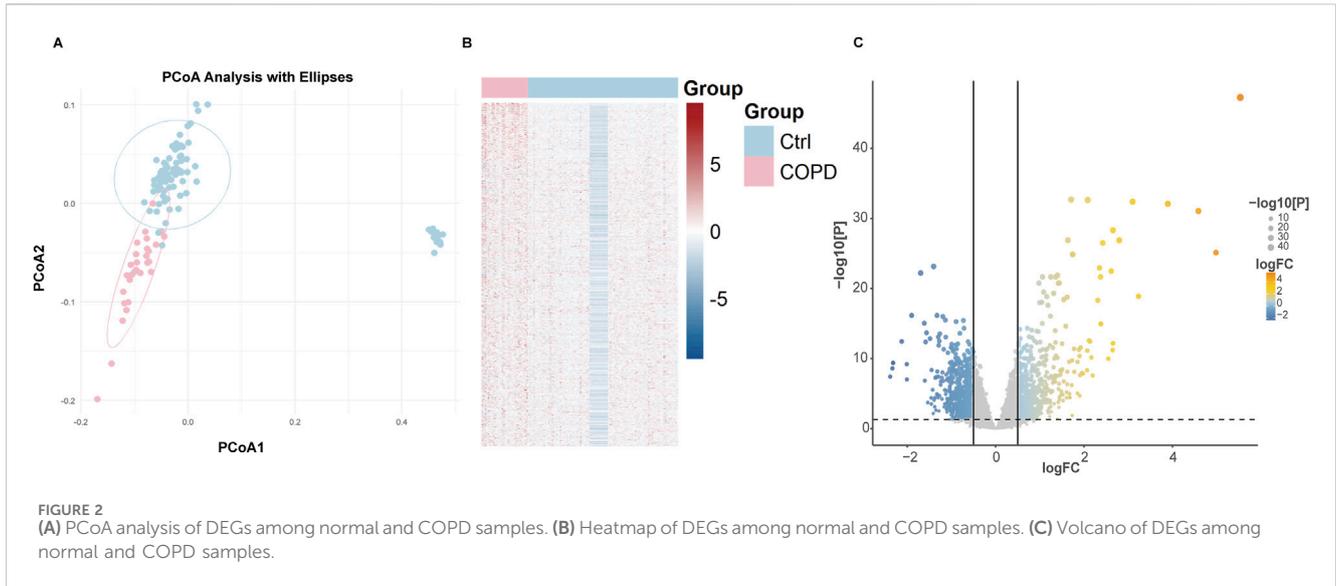
To verify the association of identified genes with disease immune infiltration, the CIBERSORT algorithm was used to evaluate the proportion of 22 immune cell types in normal and COPD samples based on transcriptome data. The correlation between the identified genes and the 22 types of immune cells was subsequently analyzed.

2.7 Prediction of drug-gene interactions

The Drug-Gene Interaction Database (DGIdb, <http://www.dgidb.org/>) aggregates drug-gene interaction data from various sources, including DrugBank, PharmGKB, ChEMBL, clinical trial databases, and PubMed literature. Information on over 40,000 genes and 10,000 drugs, involving over 100,000 drug-gene interactions, was collected and organized. Key genes identified as potential pharmaceutical targets for COPD treatment were imported into DGIdb to explore existing drugs or small organic compounds. The reliability of each drug-gene interaction was evaluated based on evidence from relevant drug databases such as DrugBank. Potential therapeutic drugs for COPD were selected based on the interaction score. Results were visualized using the "ggplot2 (3.2.1)" and "ggalluvial (0.11.1)" R packages.

2.8 Statistical analysis

All data analyses were performed using R software (version 4.4.0). The Wilcoxon test was used for group comparisons, with $P < 0.05$ considered statistically significant.



3 Results

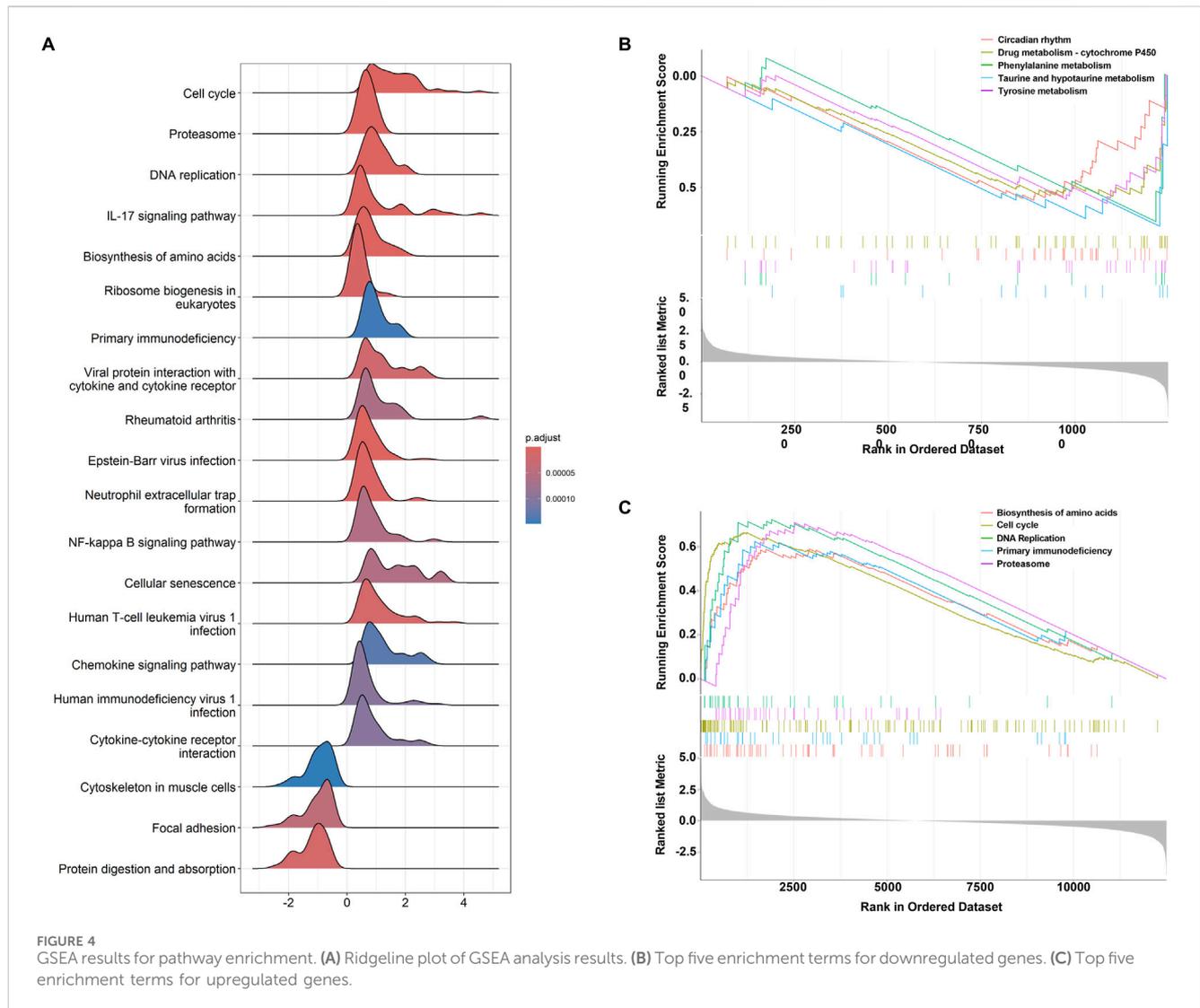
3.1 Differentially expressed genes identification in COPD and healthy control groups

In this study, two airway datasets (GSE11906 and GSE20257) were used to analyze differential gene expressions. The expression matrix is presented in [Supplementary Table 1](#). To verify the stability and consistency of clustering in classifying COPD patients, Principal Coordinates Analysis (PCoA) was employed, with results displayed

in [Figure 2A](#). The integrated expression matrix revealed 1782 DEGs, of which 920 were upregulated and 862 were downregulated, as shown in [Figure 2B](#). The volcano plot highlights DEGs with significant changes in expression levels in [Figure 2C](#). The differentially expressed genes are detailed in [Supplementary Table 2](#).

3.2 Functional analysis

Gene Ontology (GO) analysis identified 673 biological processes (BP), 30 cellular components (CC), and 61 molecular functions



(MF), as detailed in [Supplementary Table 3](#). The top six GO items are listed in [Figure 3A](#). The DEGs were significantly enriched in processes such as responses to xenobiotics, toxic substances, and cytokine production, as well as metabolic and hormonal regulation. They were also associated with the extracellular matrix, platelet granules, and plasma membrane components, with functions including antioxidant activity, enzyme binding, and structural roles. According to the KEGG analysis, the DEGs were enriched in various pathways, as shown in [Figure 3B](#).

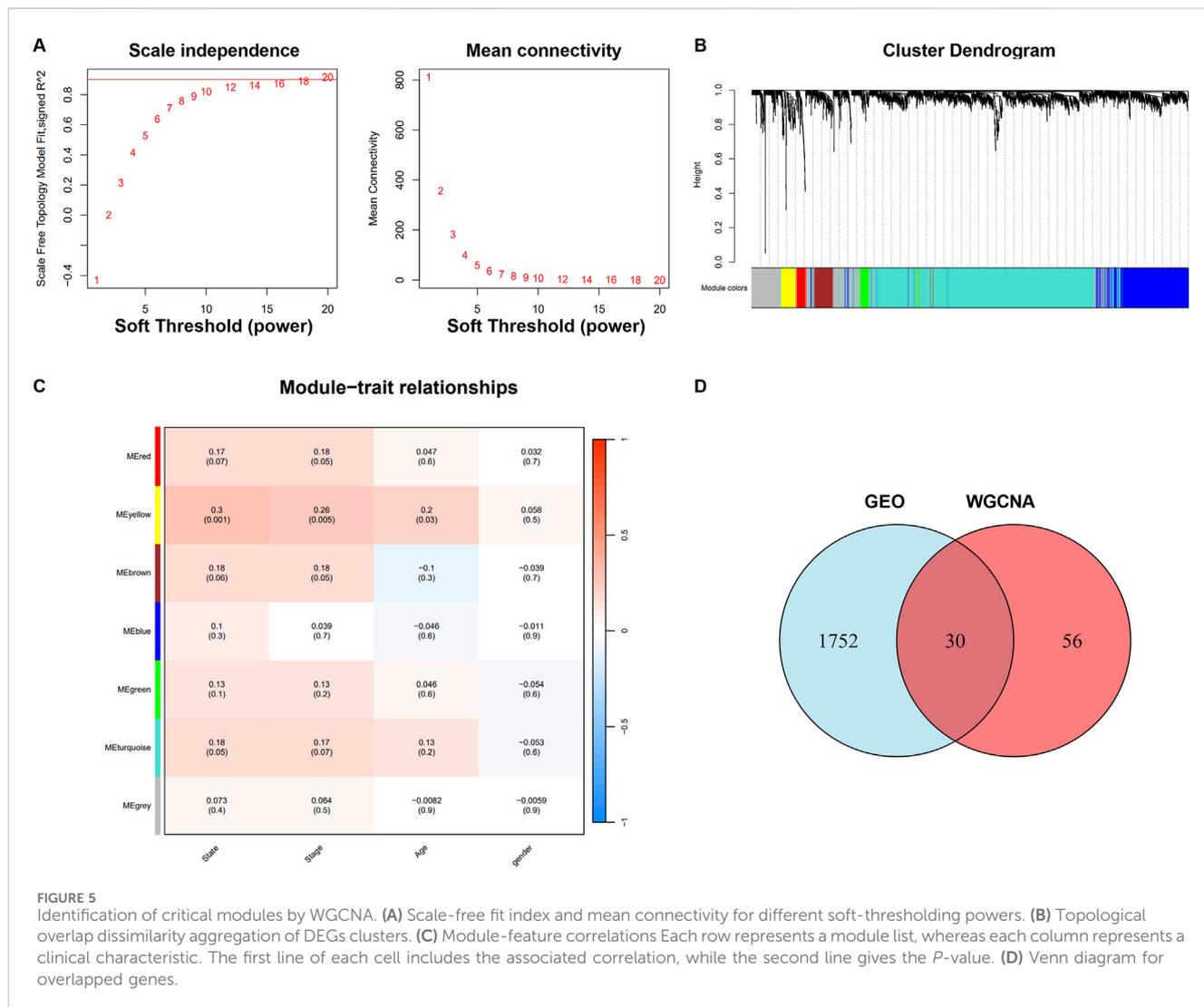
GSEA analysis ([Supplementary Table 4](#)) revealed distinct pathway enrichment patterns for upregulated and downregulated genes. [Figure 4A](#) shows the ridge plot of GSEA results, highlighting pathways such as the cell cycle, proteasome, DNA replication, and IL-17 signaling. Downregulated genes were enriched in circadian rhythm, drug metabolism-cytochrome P450, phenylalanine metabolism, and taurine and hypotaurine metabolism ([Figure 4B](#)). In contrast, upregulated genes were associated with amino acid biosynthesis, cell cycle, proteasome, primary immunodeficiency, and DNA replication ([Figure 4C](#)). These findings emphasize the critical roles of metabolic and immune-related pathways in the studied biological processes.

3.3 Overlap between COPD-Related module genes and differentially expressed genes

A scale-free network with a soft threshold of 18 ($R^2 = 0.9$) was constructed, as shown in [Figure 5A](#). We then computed module eigengenes, representing the overall gene expression level of each module, and grouped them based on their associations. Seven modules were identified, as depicted in [Figure 5B](#). The yellow module was found to be correlated with COPD ($\text{cor} = 0.3$, $P = 0.001$). This module contained 86 COPD-related genes, which were retained for further investigation, as shown in [Figure 5C](#). Ultimately, 30 genes were identified as overlapping between the DEGs and the selected module genes, as illustrated in [Figure 5D](#).

3.4 Key gene identification

To identify gene signatures, the 30 candidate genes were analyzed using SVM-RFE, RF, and LASSO methods. Using SVM-RFE, we identified a 7-gene signature with a precision of 0.897, as shown in [Figures 6A, B](#). LASSO analysis identified an 8-gene



signature, as depicted in Figures 6C, D. RF analysis identified a 6-gene signature, as shown in Figure 6E. To establish a robust gene signature for COPD, we determined the overlapping genes from these methods, resulting in the identification of four key genes: *MRC1*, *BCL2A1*, *GYPC* and *SLC2A3*, as illustrated in Figure 6F. These four genes were significantly upregulated in COPD samples compared to controls, as shown in Figure 7A. External validation using the GSE5058 and GSE8545 datasets confirmed this trend, as shown in Figure 7B.

3.5 Correlation of key genes and immune cell infiltration

Chronic inflammation of the airways, lung parenchyma, and pulmonary vasculature is a hallmark of COPD, involving inflammatory cells such as neutrophils, macrophages, and T-lymphocytes in the disease's pathogenesis. We examined the pattern of immune cell infiltration and found that the abundance of resting mast cells, M0 macrophages, and memory B cells was

significantly higher in COPD samples compared to normal samples. In contrast, native B cells, activated memory CD4 T cells, follicular helper T cells, and resting NK cells were significantly reduced, as shown in Figure 8A.

Additionally, we calculated the correlation between key gene expression and infiltrating immune cells. The results indicated that most immune cells had a positive correlation with key gene expressions, as shown in Figure 8B. These findings suggest that inflammatory components play a crucial role in the development of COPD, and that key genes may have a novel regulatory role in immune function.

3.6 Potential drugs targeting the diagnostic genes

To identify potential drugs for COPD therapy, we searched for drugs targeting the biomarkers using the DGIdb database. As shown in Figure 9, six drugs targeting *BCL2A1* and three drugs targeting *GYPC* were identified.

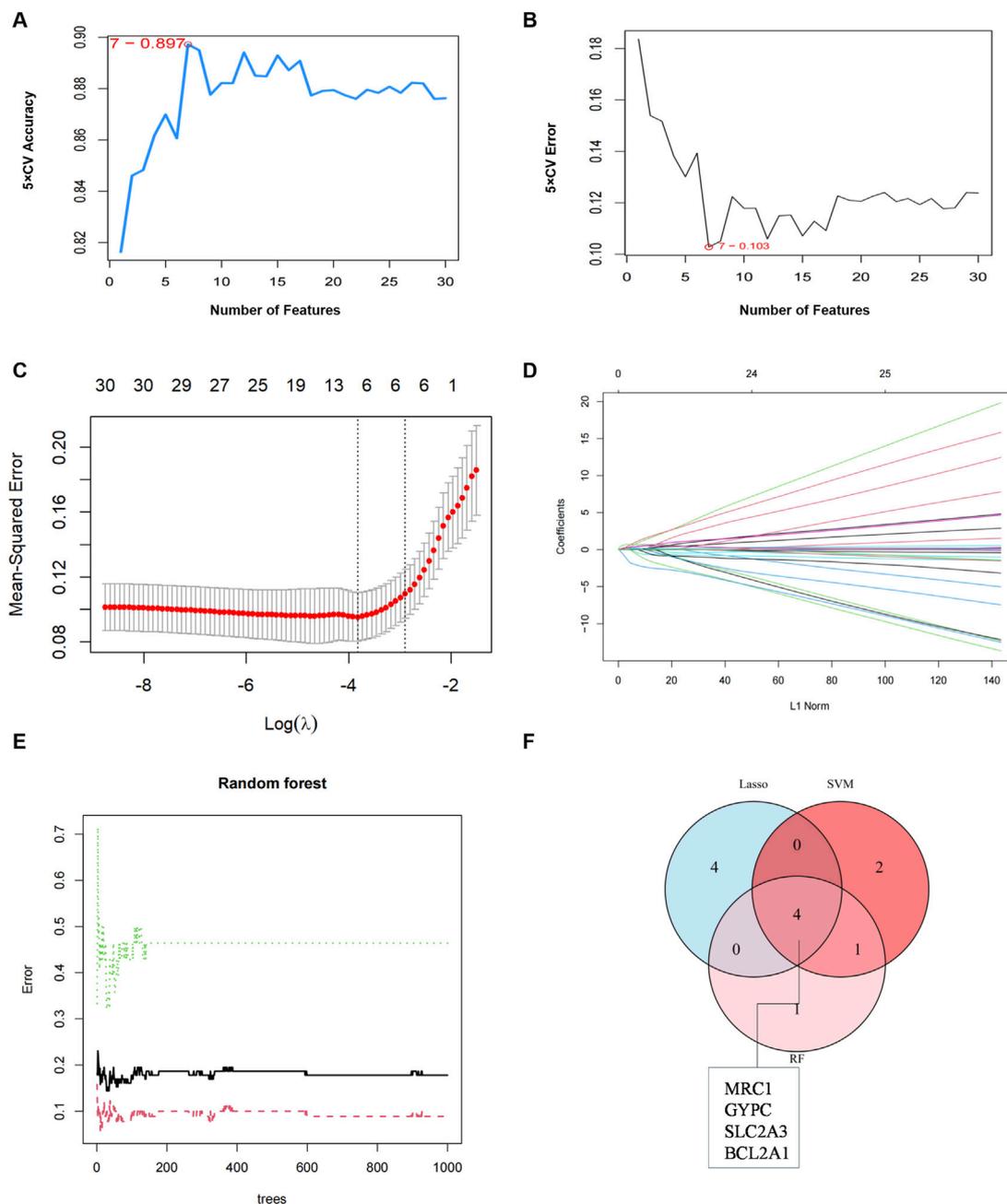


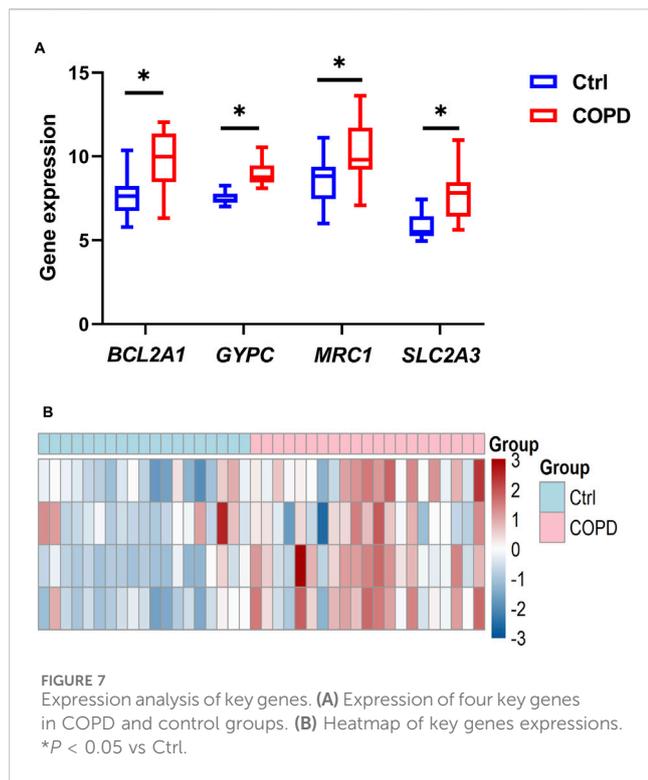
FIGURE 6 Key gene identification. **(A)** 7 gene signatures were identified by SVM-RFE analysis with an accuracy of 0.897. **(B)** Error of 0.103. **(C)** Cross-validation to select the optimal tuning parameter $\log(\lambda)$ in LASSO analysis. **(D)** LASSO coefficient profiles of candidate genes. **(E)** RF analyses identified six gene signatures **(F)** Venn diagram of four key genes shared by the SVM-RFE, RF, and LASSO algorithms.

4 Discussion

COPD is a leading cause of morbidity and mortality worldwide, with approximately 70%–80% of adults with COPD being undiagnosed (Soriano et al., 2009; Lamprecht et al., 2015; Martinez et al., 2015; Casas Herrera et al., 2016; Echazarreta et al., 2018; Gershon et al., 2018; Soriano et al., 2021). Undiagnosed patients are at increased risk of poor outcomes and a worsened quality of life, making early detection crucial for mitigating the impact of COPD and reducing the burden on

healthcare systems (Larsson et al., 2019; Kostikas et al., 2020). Over the past decade, there has been growing interest in developing effective strategies and instruments for COPD detection (Lin et al., 2023). Understanding critical pathways and gene signatures in COPD could aid in risk assessment, pathogenesis elucidation, and personalized therapy development.

In this study, the top three differentially expressed genes (DEGs) identified were ME1 (Malic Enzyme 1), NQO1 (NAD(P)H Quinone Dehydrogenase 1), and CYP1B1 (Cytochrome P450 Family 1 Subfamily B Member 1), all of which have well-established



roles in COPD pathogenesis. ME1 is a key enzyme involved in cellular metabolism and oxidative stress, both critical factors in the progression of COPD (Ryan et al., 2023). NQO1, an important antioxidant enzyme, plays a pivotal role in regulating oxidative stress, a hallmark feature of COPD (Li et al., 2022a). CYP1B1, on the other hand, is implicated in the metabolism of environmental toxins and xenobiotics, making it particularly relevant to COPD (Yang et al., 2020). These genes are not only highly differentially expressed but are also enriched in biological pathways central to COPD pathology, including the oxidative stress response and xenobiotic metabolism. Together, these findings highlight the potential of these genes as biomarkers or therapeutic targets in COPD research.

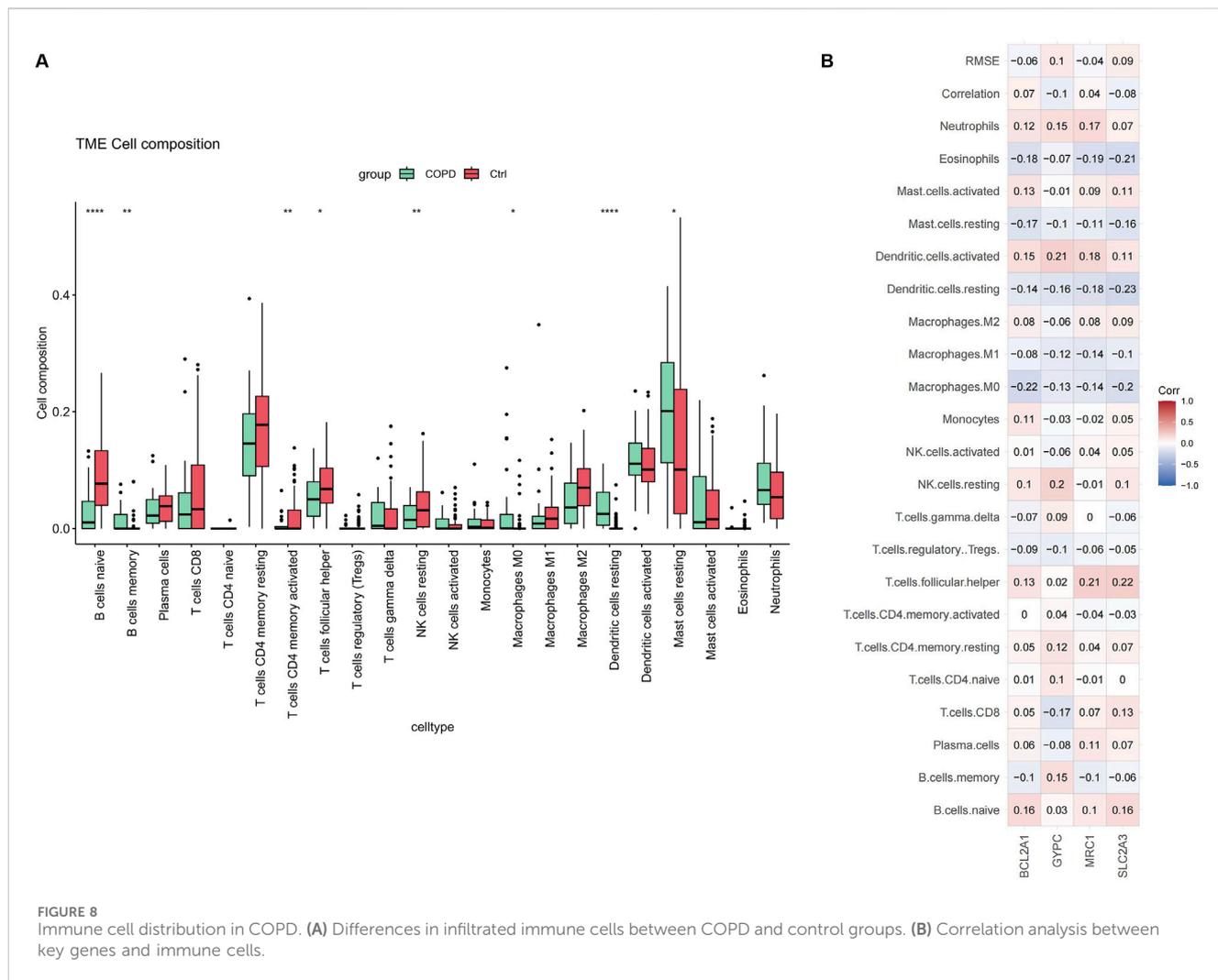
Advancements in bioinformatics have significantly enhanced our ability to use microarray data to uncover key genes, interaction networks, and pathways involved in COPD. In this study, both ORA and GSEA were applied to explore the biological processes influencing COPD progression. Enrichment analysis highlighted several key biological processes, including responses to xenobiotics and toxic substances, cytokine production, as well as metabolic and hormonal regulation, all of which are highly relevant to COPD pathogenesis. The response to xenobiotics and toxic substances reflects the lungs' defense mechanisms against environmental pollutants, cigarette smoke, and other harmful exposures, all of which trigger oxidative stress and inflammation—hallmarks of COPD. Previous studies have established the importance of these responses in exacerbating the disease (Christenson et al., 2022). The cytokine production pathway, crucial in amplifying the inflammatory response, also emerged as a significant factor in COPD. This process contributes to tissue damage and airway remodeling, which are central features of the disease (Barnes, 2009). Furthermore, metabolic and hormonal regulation emphasizes the systemic nature of COPD, suggesting

that metabolic dysregulation and hormonal imbalances may exacerbate disease progression. Recent research supports targeted reprogramming of metabolism as a promising therapeutic approach for respiratory diseases like COPD (Gan et al., 2024). Together, these findings corroborate previous studies and underscore the importance of these biological processes as potential diagnostic, prognostic, and therapeutic targets in COPD.

In our study, GSEA provided a deeper insight into the specific biological pathways enriched among DEGs. Notably, GSEA revealed that genes were primarily enriched in the *IL-17* signaling pathway, circadian rhythm, and drug metabolism-cytochrome P450. *IL-17* plays a crucial role in lung lymphoid neogenesis in COPD, contributing to airway inflammation, remodeling, and mucus hypersecretion (Kramer and Gaffen, 2007; Xiong et al., 2020; Henen et al., 2023). Preclinical studies have shown that anti-*IL-17* antibodies can reduce airway inflammation and remodeling in COPD models, supporting *IL-17* as a potential therapeutic target (Yousuf et al., 2019). Additionally, the circadian rhythm pathway emerged as significant in COPD pathogenesis. Disruption of circadian rhythms has been linked to various lung diseases, and the circadian clock gene *Clock-Bmal1* has been shown to regulate cellular responses to inflammation and immune activation in the lungs. This pathway may hold therapeutic potential for improving COPD outcomes by restoring circadian regulation (Li et al., 2022b). Although both ORA and GSEA identified pathways related to inflammation and immune response, their approaches provided complementary perspectives. ORA helped pinpoint over-represented functional categories among the most significantly differentially expressed genes, while GSEA offered a broader view by analyzing the entire ranked gene list. This allowed GSEA to identify pathways enriched at both ends of the gene expression spectrum, capturing subtle shifts in pathway activation that ORA might have missed. For example, GSEA highlighted pathways like the *IL-17* signaling pathway and circadian rhythm, which, while not dominated by a small number of highly differentially expressed genes, represent important, biologically significant alterations in COPD. These insights underscore the value of using both enrichment methods in combination to gain a more comprehensive understanding of the molecular mechanisms driving COPD.

Recent research has confirmed that innate and adaptive immune mechanisms play essential roles in COPD progression (Caramori et al., 2016; Bu et al., 2020). In this study, resting mast cells, M0 macrophages, and memory B cells were found to be upregulated in COPD samples. Macrophages and B cells are critical immune cells in COPD pathogenesis (Seys et al., 2015; Lee et al., 2016; Kapellos et al., 2018; Sullivan et al., 2019), and mast cells may also play an important role. Increased reticular basement membrane and lamina propria mast cells, as well as perivascular mast cells involved in angiogenesis, have been observed in COPD patients (Soltani et al., 2012). Understanding biology, heterogeneity, activation mechanisms, and signaling cascades of immune cells could lead to novel therapies for COPD.

In our study, four key genes were identified as being related to COPD. Mannose receptor C-type 1 (*MRC1*) is a critical regulator in macrophage-mediated immune responses (van der Zande et al., 2021). This receptor plays a significant role in several biological processes, including the regulation of circulating reproductive hormones, homeostasis, innate immunity, and infection responses (Cummings, 2022). Recent studies have highlighted the



role of *MRC1* in macrophage activation (Gantzel et al., 2020), a process crucial for chronic inflammation and tissue remodeling in COPD. Our findings suggest that *MRC1* may serve as a potential biomarker for COPD progression, particularly in immune regulation and the inflammatory pathways associated with the disease.

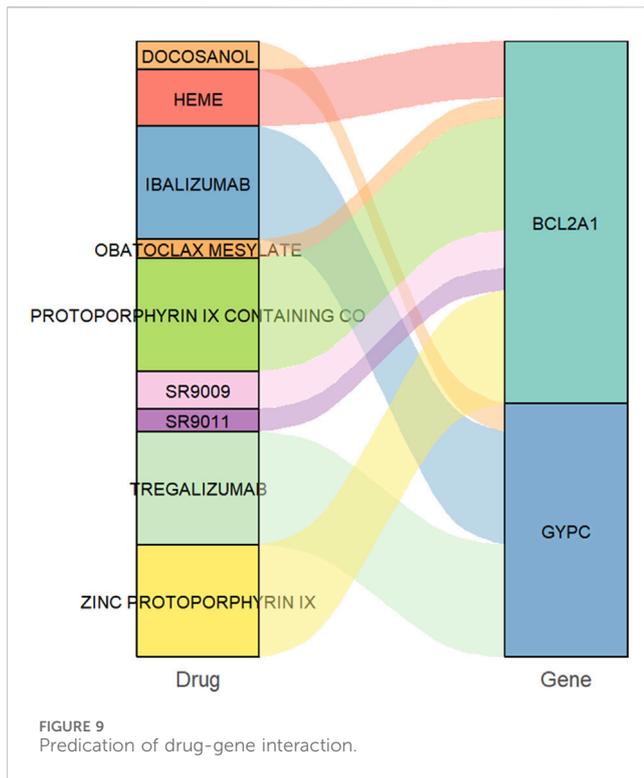
B-cell lymphoma 2-related protein A1 (*BCL2A1*), a highly regulated *NF-κB* target gene, is known for its pro-survival roles in the hematopoietic system and is overexpressed in various cancers, contributing to tumor progression (Vogler, 2012; Yue et al., 2021; Gao et al., 2023). *BCL2A1* has also been implicated in protecting against acute lung injury (Ren et al., 2024), although its direct role in COPD remains underexplored. Our study reveals that *BCL2A1* is highly expressed in the airway epithelial cells of COPD patients, suggesting that it may play an important role in the pathogenesis of COPD and could serve as a potential therapeutic target.

Glycophorin C (*GYPC*) is a membrane protein primarily expressed in red blood cells, where it is involved in cell adhesion and maintaining structural integrity. Although its role in pulmonary diseases is not well understood, previous studies have proposed the red blood cell as a biosensor for monitoring oxidative stress and imbalance in COPD (Lucantoni et al., 2006). In our study, *GYPC* expression was

significantly upregulated in COPD patients, indicating its potential involvement in the altered immune landscape in COPD and its promise as a biomarker for disease progression.

Solute carrier family 2 member 3 (*SLC2A3*), also known as *GLUT3*, is a high-affinity glucose transporter involved in cellular energy metabolism. Overexpression of *SLC2A3* has been shown to promote cell survival and growth in cancer (Yao et al., 2020; Yan et al., 2023). Our analysis, which focused on the immune microenvironment of COPD patients, revealed that *SLC2A3* was expressed in macrophages from COPD patients and was upregulated in THP-M cells and lung tissues of these patients (Zhang et al., 2023). In COPD, *SLC2A3* appears to play a crucial role in maintaining energy homeostasis under conditions of chronic inflammation and hypoxic stress. These findings suggest that *SLC2A3* could be a promising biomarker for COPD diagnosis and therapy, particularly in the context of metabolic reprogramming during disease progression.

To uncover diagnostic indications for COPD, we applied SVM-RFE, LASSO, and RF algorithms, and used CIBERSORT to examine immune cell infiltration. This study identified *MRC1*, *BCL2A1*, *GYPC* and *SLC2A3* as COPD diagnostic indicators. However, studying has several limitations. Firstly, the key genes should be validated by qPCR, and their localization and distribution should be verified.



Secondly, the study scope did not include detailed *in vivo* and *in vitro* validation. Finally, our findings were derived from bioinformatics analysis, and the specific mechanisms by which key genes affect COPD prognosis need further experimental confirmation.

One limitation of this study is the relatively small sample size, with the validation set comprising only 21 COPD patients and 19 controls. This may limit the statistical power and generalizability of the findings. However, despite the small sample size, we ensured the robustness of our results by validating the identified hub genes and pathways across multiple independent datasets. These datasets consistently supported our findings, which enhances the reliability of our conclusions and suggests that the observed gene expression patterns may be applicable to other cohorts.

Another limitation is the use of older datasets, with one microarray dataset being nearly 20 years old. Although these datasets are still widely cited, advances in sequencing technologies and metadata standards may impact their generalizability. Therefore, future studies should incorporate updated datasets and experimental validation to further confirm our findings.

To address the sample size limitation, we emphasize the need for future studies to utilize larger validation cohorts. A larger sample size would not only improve statistical power but also increase the generalizability of our findings across different patient populations. We believe these efforts will provide a more solid foundation for confirming the clinical relevance of the identified genes and pathways.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the Weill Medical College of Cornell University NIH General Clinical Research Center. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JL: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. XL: Data curation, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Writing—original draft, Writing—review and editing. YL: Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We appreciate the GEO database for providing the original data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1499996/full#supplementary-material>

References

- Ammous, Z., Hackett, N. R., Butler, M. W., Raman, T., Dolgalev, I., O'Connor, T. P., et al. (2008). Variability in small airway epithelial gene expression among normal smokers. *Chest* 133 (6), 1344–1353. doi:10.1378/chest.07-2245
- Barnes, P. J. (2009). The cytokine network in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* 41 (6), 631–638. doi:10.1165/rcmb.2009-0220TR
- Bu, T., Wang, L. F., and Yin, Y. Q. (2020). How do innate immune cells contribute to airway remodeling in COPD progression? *Int. J. Chron. Obstruct Pulmon Dis.* 15, 107–116. doi:10.2147/copd.S235054
- Caramori, G., Casolari, P., Barczyk, A., Durham, A. L., Di Stefano, A., and Adcock, I. (2016). COPD immunopathology. *Semin. Immunopathol.* 38 (4), 497–515. doi:10.1007/s00281-016-0561-5
- Carolan, B. J., Heguy, A., Harvey, B. G., Leopold, P. L., Ferris, B., and Crystal, R. G. (2006). Up-regulation of expression of the ubiquitin carboxyl-terminal hydrolase L1 gene in human airway epithelium of cigarette smokers. *Cancer Res.* 66 (22), 10729–10740. doi:10.1158/0008-5472.Can-06-2224
- Casas Herrera, A., Montes de Oca, M., López Varela, M. V., Aguirre, C., Schiavi, E., Jardim, J. R., et al. (2016). COPD underdiagnosis and misdiagnosis in a high-risk primary care population in four Latin American countries. A key to enhance disease diagnosis: the puma study. *PLoS One* 11 (4), e0152266. doi:10.1371/journal.pone.0152266
- Christenson, S. A., Smith, B. M., Bafadhel, M., and Putcha, N. (2022). Chronic obstructive pulmonary disease. *Lancet* 399 (10342), 2227–2242. doi:10.1016/s0140-6736(22)00470-6
- Cummings, R. D. (2022). The mannose receptor ligands and the macrophage glycome. *Current opinion in structural biology* 75, 102394. doi:10.1016/j.sbi.2022.102394
- Echazarreta, A. L., Arias, S. J., Del Olmo, R., Giugno, E. R., Colodenco, F. D., Arce, S. C., et al. (2018). Prevalence of COPD in 6 urban clusters in Argentina: the EPOC.AR study. *Arch. Bronconeumol (Engl Ed)* 54 (5), 260–269. doi:10.1016/j.arbres.2017.09.018
- Gan, P. X. L., Zhang, S., and Fred Wong, W. S. (2024). Targeting reprogrammed metabolism as a therapeutic approach for respiratory diseases. *Biochem. Pharmacol.* 228, 116187. doi:10.1016/j.bcp.2024.116187
- Gantzel, R. H., Kjær, M. B., Laursen, T. L., Kazankov, K., George, J., Møller, H. J., et al. (2020). Macrophage activation markers, soluble CD163 and mannose receptor, in liver fibrosis. *Front. Med. (Lausanne)* 7, 615599. doi:10.3389/fmed.2020.615599
- Gao, L., Ye, Z., Peng, S., Lei, P., Song, P., Li, Z., et al. (2023). BCL2A1 is associated with tumor-associated macrophages and unfavorable prognosis in human gliomas. *Aging (Albany NY)* 15 (20), 11611–11638. doi:10.18632/aging.205149
- Gershon, A. S., Thiruchelvam, D., Chapman, K. R., Aaron, S. D., Stanbrook, M. B., Bourbeau, J., et al. (2018). Health services burden of undiagnosed and overdiagnosed COPD. *Chest* 153 (6), 1336–1346. doi:10.1016/j.chest.2018.01.038
- Henen, C., Johnson, E. A., and Wiesel, S. (2023). Unleashing the power of IL-17: a promising frontier in chronic obstructive pulmonary disease (COPD) treatment. *Cureus* 15 (7), e41977. doi:10.7759/cureus.41977
- Iheanacho, I., Zhang, S., King, D., Rizzo, M., and Ismaila, A. S. (2020). Economic burden of chronic obstructive pulmonary disease (COPD): a systematic literature review. *Int. J. Chron. Obstruct Pulmon Dis.* 15, 439–460. doi:10.2147/copd.S234942
- Kapellos, T. S., Bassler, K., Aschenbrenner, A. C., Fujii, W., and Schultze, J. L. (2018). Dysregulated functions of lung macrophage populations in COPD. *J. Immunol. Res.* 2018, 2349045. doi:10.1155/2018/2349045
- Kostikas, K., Price, D., Gutzwiller, F. S., Jones, B., Loeffroth, E., Clemens, A., et al. (2020). Clinical impact and healthcare resource utilization associated with early versus late COPD diagnosis in patients from UK CPRD database. *Int. J. Chron. Obstruct Pulmon Dis.* 15, 1729–1738. doi:10.2147/copd.S255414
- Kramer, J. M., and Gaffen, S. L. (2007). Interleukin-17: a new paradigm in inflammation, autoimmunity, and therapy. *J. Periodontol.* 78 (6), 1083–1093. doi:10.1902/jop.2007.060392
- Lamprecht, B., Soriano, J. B., Studnicka, M., Kaiser, B., Vanfleteren, L. E., Gnatiuc, L., et al. (2015). Determinants of underdiagnosis of COPD in national and international surveys. *Chest* 148 (4), 971–985. doi:10.1378/chest.14-2535
- Larsson, K., Janson, C., Ställberg, B., Lisspers, K., Olsson, P., Kostikas, K., et al. (2019). Impact of COPD diagnosis timing on clinical and economic outcomes: the ARCTIC observational cohort study. *Int. J. Chron. Obstruct Pulmon Dis.* 14, 995–1008. doi:10.2147/copd.S195382
- Leap, J., Arshad, O., Cheema, T., and Balaan, M. (2021). Pathophysiology of COPD. *Crit. Care Nurs. Q.* 44 (1), 2–8. doi:10.1097/cnq.0000000000000334
- Lee, J., Machin, M., Russell, K. E., Pavlidis, S., Zhu, J., Barnes, P. J., et al. (2016). Corticosteroid modulation of immunoglobulin expression and B-cell function in COPD. *FASEB J.* 30 (5), 2014–2026. doi:10.1096/fj.20150135
- Li, J., Baker, J., Higham, A., Shah, R., Montero-Fernandez, A., Murray, C., et al. (2022a). COPD lung studies of Nrf2 expression and the effects of Nrf2 activators. *Inflammopharmacology* 30 (4), 1431–1443. doi:10.1007/s10787-022-00967-3
- Li, L., Zhang, M., Zhao, C., Cheng, Y., Liu, C., and Shi, M. (2022b). Circadian clock gene Clock-Bmal1 regulates cellular senescence in Chronic obstructive pulmonary disease. *BMC Pulm. Med.* 22 (1), 435. doi:10.1186/s12890-022-02237-y
- Lin, C. H., Cheng, S. L., Chen, C. Z., Chen, C. H., Lin, S. H., and Wang, H. C. (2023). Current progress of COPD early detection: key points and novel strategies. *Int. J. Chron. Obstruct Pulmon Dis.* 18, 1511–1524. doi:10.2147/copd.S413969
- Lucantoni, G., Pietraforte, D., Matarrese, P., Gambardella, L., Metere, A., Paone, G., et al. (2006). The red blood cell as a biosensor for monitoring oxidative imbalance in chronic obstructive pulmonary disease: an *ex vivo* and *in vitro* study. *Antioxid. Redox Signal* 8 (7–8), 1171–1182. doi:10.1089/ars.2006.8.1171
- Martinez, C. H., Mannino, D. M., Jaimes, F. A., Curtis, J. L., Han, M. K., Hansel, N. N., et al. (2015). Undiagnosed obstructive lung disease in the United States. Associated targets and mechanisms of sivelestat in protecting acute lung injury. *Comput. Biol. Med.* 170, 108080. doi:10.1016/j.compbiomed.2024.108080
- McDonough, J. E., Yuan, R., Suzuki, M., Seyednejad, N., Elliott, W. M., Sanchez, P. G., et al. (2011). Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *N. Engl. J. Med.* 365 (17), 1567–1575. doi:10.1056/NEJMoa1106955
- Noble, W. S. (2006). What is a support vector machine? *Nat. Biotechnol.* 24 (12), 1565–1567. doi:10.1038/nbt1206-1565
- Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintia, A. R., and Kundu, S. (2018). Improved random forest for classification. *IEEE Trans. Image Process* 27 (8), 4012–4024. doi:10.1109/tip.2018.2834830
- Raman, T., O'Connor, T. P., Hackett, N. R., Wang, W., Harvey, B. G., Attiyeh, M. A., et al. (2009). Quality control in microarray assessment of gene expression in human airway epithelium. *BMC Genomics* 10, 493. doi:10.1186/1471-2164-10-493
- Ren, J., Deng, G., Li, R., Jin, X., Liu, J., Li, J., et al. (2024). Possible pharmacological targets and mechanisms of sivelestat in protecting acute lung injury. *Comput. Biol. Med.* 170, 108080. doi:10.1016/j.compbiomed.2024.108080
- Ryan, E. M., Sadiku, P., Coelho, P., Watts, E. R., Zhang, A., Howden, A. J. M., et al. (2023). NRF2 activation reprograms defects in oxidative metabolism to restore macrophage function in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 207 (8), 998–1011. doi:10.1164/rccm.202203-0482OC
- Seys, L. J., Verhamme, F. M., Schinwald, A., Hammad, H., Cunoosamy, D. M., Bantsimba-Malanda, C., et al. (2015). Role of B Cell-Activating factor in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 192 (6), 706–718. doi:10.1164/rccm.201501-0103OC
- Shaykhiyev, R., Otaki, F., Bonsu, P., Dang, D. T., Teater, M., Strulovici-Barel, Y., et al. (2011). Cigarette smoking reprograms apical junctional complex molecular architecture in the human airway epithelium *in vivo*. *Cell Mol. Life Sci.* 68 (5), 877–892. doi:10.1007/s00018-010-0500-x
- Soltani, A., Ewe, Y. P., Lim, Z. S., Sohal, S. S., Reid, D., Weston, S., et al. (2012). Mast cells in COPD airways: relationship to bronchodilator responsiveness and angiogenesis. *Eur. Respir. J.* 39 (6), 1361–1367. doi:10.1183/09031936.00084411
- Soriano, J. B., Alfageme, I., Miravittles, M., de Lucas, P., Soler-Cataluña, J. J., García-Río, F., et al. (2021). Prevalence and determinants of COPD in Spain: EPISCAN II. *Arch. Bronconeumol (Engl Ed)* 57 (1), 61–69. doi:10.1016/j.arbres.2020.07.024
- Soriano, J. B., Zielinski, J., and Price, D. (2009). Screening for and early detection of chronic obstructive pulmonary disease. *Lancet* 374 (9691), 721–732. doi:10.1016/s0140-6736(09)61290-3
- Sullivan, J. L., Bagevalu, B., Glass, C., Sholl, L., Kraft, M., Martinez, F. D., et al. (2019). B cell-adaptive immune profile in emphysema-predominant chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 200 (11), 1434–1439. doi:10.1164/rccm.201903-0632LE
- Sun, S., Shen, Y., Wang, J., Li, J., Cao, J., and Zhang, J. (2021). Identification and validation of autophagy-related genes in chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct Pulmon Dis.* 16, 67–78. doi:10.2147/copd.S288428
- van der Zande, H. J. P., Nitsche, D., Schlautmann, L., Guigas, B., and Burgdorf, S. (2021). The Mannose Receptor: From Endocytic Receptor and Biomarker to Regulator of (Meta)Inflammation. *Frontiers in Immunology* 12, 765034.
- Vasquez, M. M., Hu, C., Roe, D. J., Chen, Z., Halonen, M., and Guerra, S. (2016). Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. *BMC Med. Res. Methodol.* 16 (1), 154. doi:10.1186/s12874-016-0254-8
- Vestbo, J., Hurd, S. S., Agustí, A. G., Jones, P. W., Vogelmeier, C., Anzueto, A., et al. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am. J. Respir. Crit. Care Med.* 187 (4), 347–365. doi:10.1164/rccm.201204-0596PP
- Vogler, M. (2012). BCL2A1: the underdog in the BCL2 family. *Cell Death Differ.* 19 (1), 67–74. doi:10.1038/cdd.2011.158
- Xiong, J., Zhou, L., Tian, J., Yang, X., Li, Y., Jin, R., et al. (2020). Cigarette smoke-induced lymphoid neogenesis in COPD involves IL-17/RANKL pathway. *Front. Immunol.* 11, 588522. doi:10.3389/fimmu.2020.588522

Yan, B., Li, X., Peng, M., Zuo, Y., Wang, Y., Liu, P., et al. (2023). The YTHDC1/GLUT3/RNF183 axis forms a positive feedback loop that modulates glucose metabolism and bladder cancer progression. *Exp. Mol. Med.* 55 (6), 1145–1158. doi:10.1038/s12276-023-00997-z

Yang, D., Yan, Y., Hu, F., and Wang, T. (2020). CYP1B1, VEGFA, BCL2, and CDKN1A affect the development of chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct Pulmon Dis.* 15, 167–175. doi:10.2147/copd.S220675

Yao, X., He, Z., Qin, C., Deng, X., Bai, L., Li, G., et al. (2020). SLC2A3 promotes macrophage infiltration by glycolysis reprogramming in gastric cancer. *Cancer Cell Int.* 20, 503. doi:10.1186/s12935-020-01599-9

Yousuf, A., Ibrahim, W., Greening, N. J., and Brightling, C. E. (2019). T2 biologics for chronic obstructive pulmonary disease. *J. Allergy Clin. Immunol. Pract.* 7 (5), 1405–1416. doi:10.1016/j.jaip.2019.01.036

Yue, T., Liu, X., Zuo, S., Zhu, J., Li, J., Liu, Y., et al. (2021). BCL2A1 and CCL18 are predictive biomarkers of cisplatin chemotherapy and immunotherapy in colon cancer patients. *Front. Cell Dev. Biol.* 9, 799278. doi:10.3389/fcell.2021.799278

Zhang, Z., Yu, H., Wang, Q., Ding, Y., Wang, Z., Zhao, S., et al. (2023). A macrophage-related gene signature for identifying COPD based on bioinformatics and *ex vivo* experiments. *J. Inflamm. Res.* 16, 5647–5665. doi:10.2147/jir.S438308