



# Screening and Validation of Key Genes Associated with Non-alcoholic Fatty Liver Disease and Chronic Obstructive Pulmonary Disease

Yue Wu<sup>1</sup>, Xiaoli Lv<sup>2,3</sup>, Yuxia Hu<sup>2,3</sup>, Tuyai Bai<sup>2,3</sup>, Weizhong Huangfu<sup>1,\*</sup>, Mengdi Zhang<sup>2,3,\*\*</sup>

<sup>1</sup>The Affiliated Hospital of Inner Mongolia Medical University, Inner Mongolia Medical University, Huhhot, China

<sup>2</sup>School of Pharmacy, Inner Mongolia Medical University, Huhhot, China

<sup>3</sup>New Drug Screening Research Center of Inner Mongolia Autonomous Region, Huhhot, China

\*Corresponding Author: The Affiliated Hospital of Inner Mongolia Medical University, Inner Mongolia Medical University, Huhhot, China. Email: [huangfufuyuan@126.com](mailto:huangfufuyuan@126.com)

\*\*Corresponding Author: School of Pharmacy, Inner Mongolia Medical University, Huhhot, China. Email: [zhangmengdi108@163.com](mailto:zhangmengdi108@163.com)

Received: 27 November, 2024; Revised: 22 April, 2025; Accepted: 17 May, 2025

## Abstract

**Background:** The incidence of severe hepatic steatosis is notably elevated in individuals afflicted with chronic obstructive pulmonary disease (COPD) as opposed to individuals among the general public. Many individuals with non-alcoholic fatty liver disease (NAFLD) exhibit elevated blood glucose and lipid levels, alongside an elevated Body Mass Index (BMI). These are all recognized risk factors for COPD.

**Objectives:** This study aimed to identify shared key genes and biological pathways between NAFLD and COPD, given their common risk factors. We hypothesized these shared risk factors might indicate overlapping molecular mechanisms and sought to determine if common genetic and pathway dysregulation could explain the epidemiological link between the two diseases.

**Methods:** A cross-sectional study design was employed. Gene expression datasets GSE63067 (NAFLD) and GSE37768 (COPD) were obtained from GEO. Differentially expressed genes (DEGs) were identified and analyzed for commonality. Protein-protein interaction networks were constructed, and GO/KEGG enrichment analyses were performed. The diagnostic value of key genes was assessed using SPSS. HepG2 cells were exposed to 1 mmol/L free fatty acids (FFAs) for 24 hours to establish the NAFLD model. The COPD model used A549 cells, which were treated with 2.5 mL of 2.5% cigarette smoke extract (CSE). Respectively, key gene protein expression was validated by Western blotting.

**Results:** Thirty-four common genes, 41 GO items, and 10 KEGG pathways were obtained in NAFLD and COPD. Cycle protein-dependent kinase inhibitor 1A (*CDKN1A*), formyl peptide receptor 1 (FPR1), *ficolin-1* (*FCN1*), and *CLEC4D* were identified as key genes through protein PPI network analysis. Western blot analysis revealed a notable elevation in protein levels of *CDKN1A*, FPR1, *FCN1*, and *CLEC4D* when compared to the control group, with statistical significance ( $P < 0.05$ ).

**Conclusions:** These key genes, preliminarily validated through cellular experiments, play a key role in the pathogenesis of NAFLD and COPD, thereby providing a novel theoretical framework for understanding the pathogenesis of these diseases.

**Keywords:** Key Gene, Non-alcoholic Fatty Liver Disease, Chronic Obstructive Pulmonary Disease, Biomarkers

## 1. Background

Non-alcoholic fatty liver disease (NAFLD), often linked to metabolic syndrome and diabetes, can escalate to nonalcoholic steatohepatitis (NASH), leading to cirrhosis and potentially hepatocellular carcinoma, influenced by dietary habits, environmental, and genetic factors (1, 2). While the precise mechanisms

underlying its pathogenesis remain elusive, it is evident that inflammatory mediators play a substantial role (3). The rising prevalence of NAFLD is accompanied by an increase in patient numbers and healthcare costs, imposing a significant economic burden on society (4). In terms of clinical applications, *Bifidobacterium bifidum* has been shown to have a positive effect on the treatment of chronic diarrhea, antibiotic-associated

diarrhea, and constipation. In addition, bifidobacteria may also play a role in tumor prevention and treatment, liver protection, and prevention of hypertension and atherosclerosis.

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory lung disease characterized by irreversible progressive airflow limitation. It represents a significant challenge to global health, with the global Initiative for chronic obstructive lung disease (GOLD) emphasizing that COPD is one of the three leading causes of death worldwide (5, 6). The COPD risk factors include exposure to harmful particles or gases, predominantly affecting low-and-middle-income countries with aging populations (7). The disease's pathogenesis, while not fully elucidated, involves inflammatory factors that contribute to lung damage, exacerbating the global economic burden (8).

Patients with COPD have a higher prevalence of severe hepatic steatosis than the general population (9). Patients with NAFLD frequently exhibit comorbid hyperglycemia, dyslipidemia, and elevated Body Mass Index (BMI), which are established risk factors for COPD (9). Although NAFLD and COPD are linked, it appears that the subject of previous investigation, relevant genetic studies are still limited, and the relevant gene regulatory network has not been fully established, which requires further exploration.

## 2. Objectives

This project is based on a bioinformatics approach to identify the co-expressed genes of NAFLD and COPD, employing systems biology analytical methods to further elucidate their connections and preliminary experimental validation of the screened key genes by setting up NAFLD and COPD models in vitro. The central genes identified in this investigation may serve as a focal point for future research, and the molecular mechanisms and signaling pathways delineated here could provide valuable insights into the interplay between COPD and NAFLD.

## 3. Methods

### 3.1. Materials

HepG2 and BEAS-2B cell lines were sourced from the Institute of Basic Medical Sciences, Peking Union Medical College, China. Cell culture was performed in glucose-rich DMEM supplemented in accordance with

the manufacturer's instructions, fetal bovine serum (Gibco, USA), and rabbit monoclonal antibodies against COX-2, cycle protein-dependent kinase inhibitor 1A (CDKN1A), formyl peptide receptor 1 (FPR1), *ficolin-1* (FCN1), *CLEC4D*, and  $\beta$ -actin (Abcam, UK) were used.

### 3.2. Microarray Data Collection

Microarray datasets related to NAFLD, COPD, and healthy subjects were retrieved from the GEO database. Specifically, a genetic overview of 11 individuals with NAFLD and 7 control individuals with NAFLD controls are included within the GSE63067 dataset. In the GSE37768 dataset, the relevant gene expression data of 20 COPD patients and 18 controls were analyzed, of which 9 non-smoking healthy people were selected as controls.

### 3.3. Identifying Genes Variably Represented in Non-alcoholic Fatty Liver Disease and Chronic Obstructive Pulmonary Disease

Differentially expressed genes (DEGs) were identified using the limma package in the R software between patients with NAFLD and COPD and their respective controls (10, 11). The criteria for selection in NAFLD as well as COPD datasets were set at  $P < 0.05$  and  $|\log_2 FC| > 0.5$ . This dual-criteria approach balances sensitivity and specificity to capture biologically relevant changes while minimizing false negatives. The choice of  $|\log_2 FC| > 0.5$  is supported by prior studies in NAFLD and COPD, where this threshold has been shown to effectively identify genes with moderate but biologically significant expression changes (12-14). Additionally, the stringent P-value cutoff of  $< 0.05$  further ensures the robustness of our findings. After shortlisting, two sets of DEGs were established and analyzed using the Venn diagram tool available online to discover overlaps between the two diseases. Intersecting genes were obtained for follow-up analysis.

### 3.4. Common Differentially Expressed Gene Functional Enrichment Analysis

We analyzed the molecular features, biological processes, and cellular constituents of intersecting genes using GO and KEGG pathway enrichment analyses in the R software (15). We defined  $P < 0.05$  as statistically significant.

### 3.5. Identification of Key Genes and Construction of PPI

## Networks

To examine the interactions of the common genes identified, a PPI network was set up using the Interacting Gene Search Tool (STRING) release 11.0 (16). The lowest interaction threshold was set at medium confidence (0.400) to establish statistical significance, and the maximum number of tier 1 and 2 interactions was limited to 10. In the resulting network, nodes represent proteins, while edges indicate interactions between them.

### 3.6. Establishment of Non-alcoholic Fatty Liver Disease and Chronic Obstructive Pulmonary disease Cell Models

The NAFLD model was constructed by treating HepG2 cells with  $1 \text{ mmol} \cdot \text{L}^{-1}$  free fatty acid (FFA) for 24 hours (1). Intracellular triglyceride (TG) levels were measured according to the kit instructions. The COPD model was constructed by treating A549 cells with 2.5% cigarette smoke extract (CSE) for 2.5 mL (17). For routine cell culture, cells were grown in DMEM with 10% FBS at 37°C with 5% CO<sub>2</sub>, with growth monitored using an inverted microscope. When the cells reached 80 - 90% confluence, trypsin digestion was performed for passaging culture and subsequent experiments.

### 3.7. Oil Red O Staining for Cellular Lipid Accumulation Detection

Cells treated as described in "2.6" were washed three times with PBS. Cells were fixed for 30 minutes with 1 mL paraformaldehyde per well. They were then stained with 400 µL Oil Red O for 20 minutes in the dark, followed by washing. Finally, cells were rinsed with PBS and imaged under an optical microscope. A UV-Vis spectrophotometer was used to measure the absorbance of the extracted solution at a wavelength of 510 nm.

### 3.8. Detection of COX-2, Cycle Protein-Dependent Kinase Inhibitor 1A, Formyl Peptide Receptor 1, Ficolin-1, CLE4D Protein Expression by Western Blot

After the cells were treated for 24 hours, the protein lysate was used to lyse the cells to extract gross proteins from the cells of each group. The concentration of protein was then determined using the Bicinchoninic Acid Assay. Following quantification, the proteins were subjected to vertical SDS-PAGE, transferred to a moist PVDF membrane, and blocked for 20 minutes. The membranes were then incubated with primary

antibodies (1:5000) at 4°C overnight, followed by three rounds of TBST washing. Subsequently, they were incubated with a secondary antibody in the dark for 1 hour and detected using an infrared fluorescent scanning and imaging system. The results were analyzed using Image J software.

### 3.9. Statistical Analysis

Statistical analyses were performed using SPSS version 22.0. For comparisons between two independent groups, independent-samples *t*-tests were used. When data did not meet the assumptions of normality, the nonparametric Mann-Whitney U test was applied. For multiple group comparisons, one-way ANOVA was conducted, followed by Tukey's post-hoc test for pairwise comparisons to control for Type I error. Correlation analyses were performed using Pearson's correlation for normally distributed data and Spearman's rank correlation for non-normally distributed data. Receiver operating characteristic (ROC) curves were generated in GraphPad Prism to compare area under the curve (AUC) values. Multivariate regression models were used to address potential confounders. Results are presented as mean ± standard deviation ( $\bar{X} \pm \text{SD}$ ). A P-value of less than 0.05 was considered statistically significant, \* indicates a significant difference..

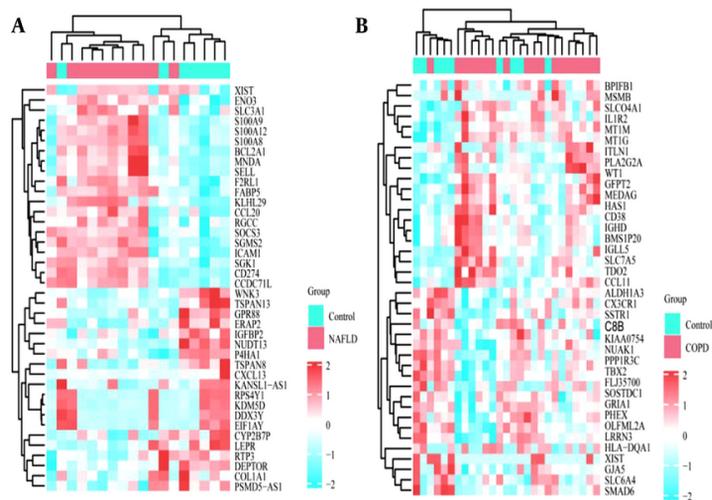
## 4. Results

### 4.1. Differential Expression Analysis of Differentially Expressed Genes in Non-alcoholic Fatty Liver Disease and Chronic Obstructive Pulmonary Disease

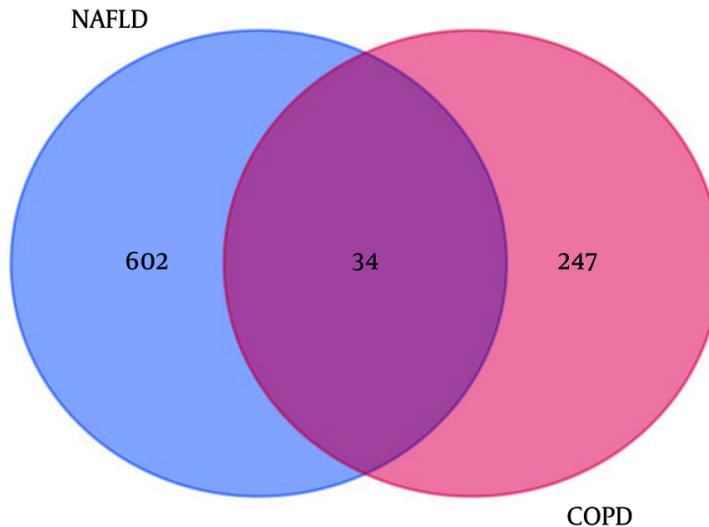
Heatmaps of the gene expression profiles of NAFLD and COPD are shown in Figure 1, with 636 and 281 DEGs obtained under the condition of  $P < 0.05$ . The 34 genes that overlap and are common to the two diseases were identified using the online Venn diagram tool (Figure 2). Detailed information on the common genes is presented in Appendix 1 in Supplementary File, while Figure 3 displays specific information regarding the analysis process.

### 4.2. GO and KEGG Enrichment Pathway Analysis

Enrichment analysis for functions and KEGG pathways of the 34 common genes in NAFLD and COPD was carried out using a P-value threshold of  $< 0.05$ . The main changes in the GO Biological Processes (BP) are as



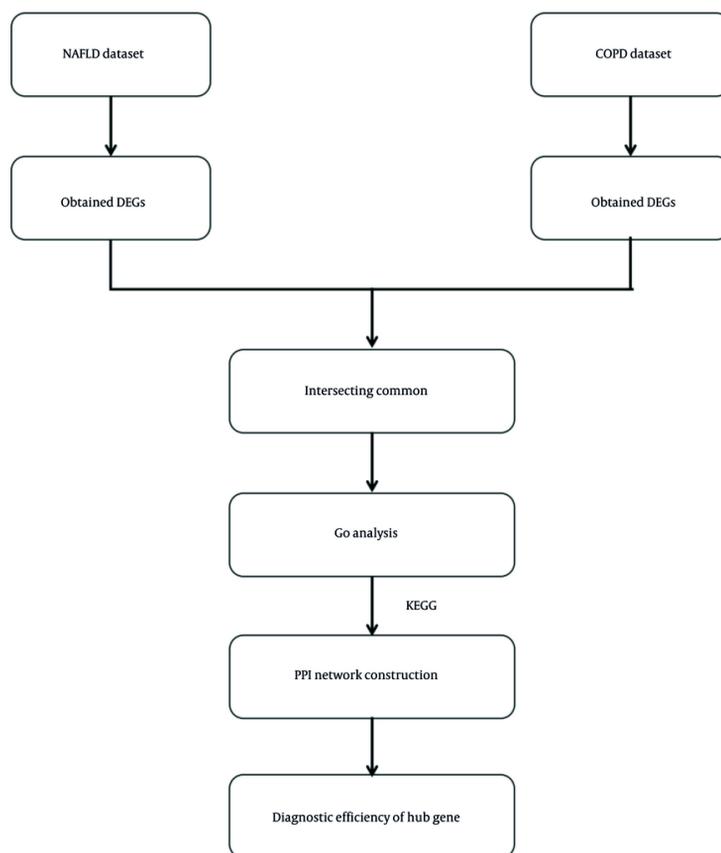
**Figure 1.** Differential genes heat map of diseases – A, differential gene of non-alcoholic fatty liver disease (NAFLD) compared to the control group; B, differential gene of chronic obstructive pulmonary disease (COPD) compared to the control



**Figure 2.** Venn diagram of non-alcoholic fatty liver disease (NAFLD) and chronic obstructive pulmonary disease (COPD) differential genes

follows: Activation of neutrophils involved in immune responses, neutrophil-driven immunity, and the regulation of innate immune response. Changes in Cellular Components (CC) were mainly focused on

fibronectin-rich granule body membranes, the inner lumen, and the lumen of the tertiary granule. In addition, in the molecular function (MF) category, significant changes were observed in glycosylation end



**Figure 3.** Analysis flow chart

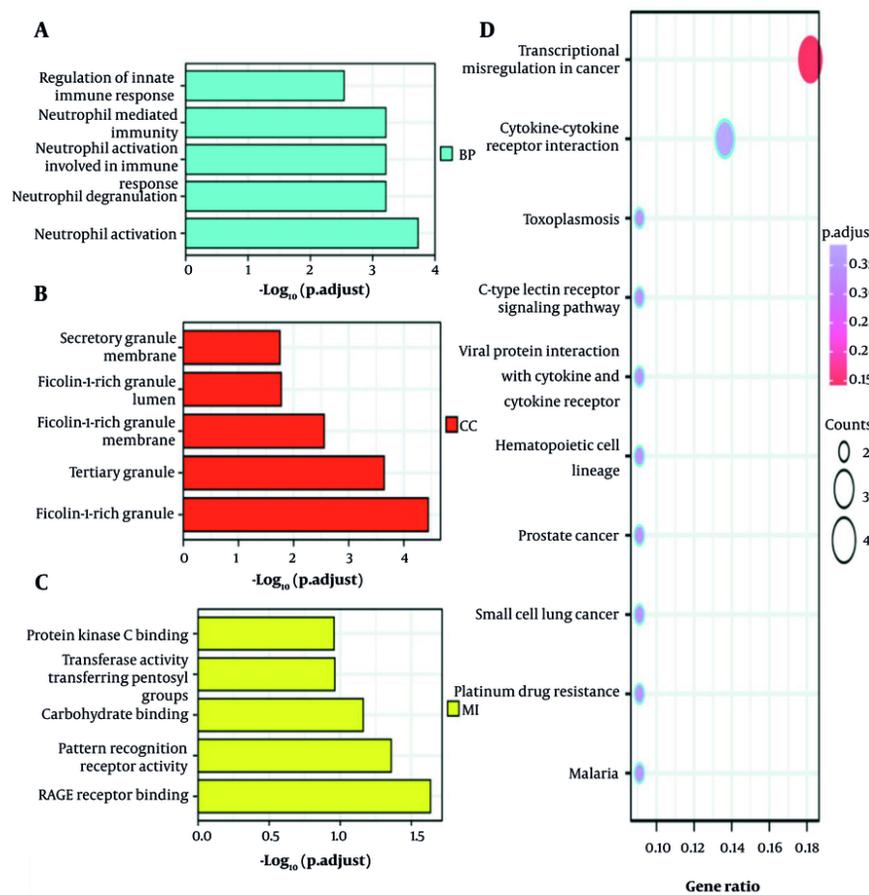
product receptor binding, pattern recognition receptor activity, and carbohydrate binding modules. Specifically, KEGG pathway changes focused on viral proteins interacting with cytokines and their corresponding receptors, the CLR-mediated signaling pathways, and cytokine-mediated cell-cell communication interactions (Figure 4).

#### 4.3. PPI Network Analysis and Core Gene Selection

To clarify the interrelationships of shared genes, we set up a PPI network. Proteins with interactions with other proteins were identified as more critical genes. These serve as the core nodes in the protein interaction network, with the precise degree of connectivity indicating the central gene's role, as shown in Figure 5. Less information was found about the interacting proteins, so information mining was used to expand the

network. Finally, the group selected proteins with interactions greater than 2 as key targets for further research.

To verify the predictive validity of the four central genes (*CDKN1A*, *FPR1*, *FCN1*, *CLEC4D*) based on the analyses, ROC curves were created using the results from the above analyses, and the AUC for transcriptional activity levels was determined in the NAFLD and COPD datasets. The outcomes of the ROC analysis for NAFLD and COPD are shown in Figure 6. The AUCs for *CDKN1A*, *FPR1*, *FCN1*, and *CLEC4D* among NAFLD patients and healthy controls were 0.767 (95% CI, 0.573 - 0.96;  $P < 0.05$ ), 0.7 (95% CI, 0.5 - 0.9;  $P = 0.079$ ), 0.656 (95% CI, 0.445 - 0.866;  $P = 0.72$ ), and 0.9 (95% CI, 0.787 - 1;  $P < 0.001$ ), respectively. The AUCs for the four central genes in COPD were 0.772 (95% CI, 0.584 - 0.959;  $P < 0.05$ ), 0.741 (95% CI, 0.549 - 0.933;  $P < 0.05$ ), 0.802 (95% CI, 0.623 - 0.982;  $P <$



**Figure 4.** Common gene function enrichment information - A, biological process; B, cell component; C, molecular function (MF); D, KEGG signaling pathways

0.05), and 0.796 (95% CI, 0.62 - 0.973;  $P < 0.05$ ), respectively.

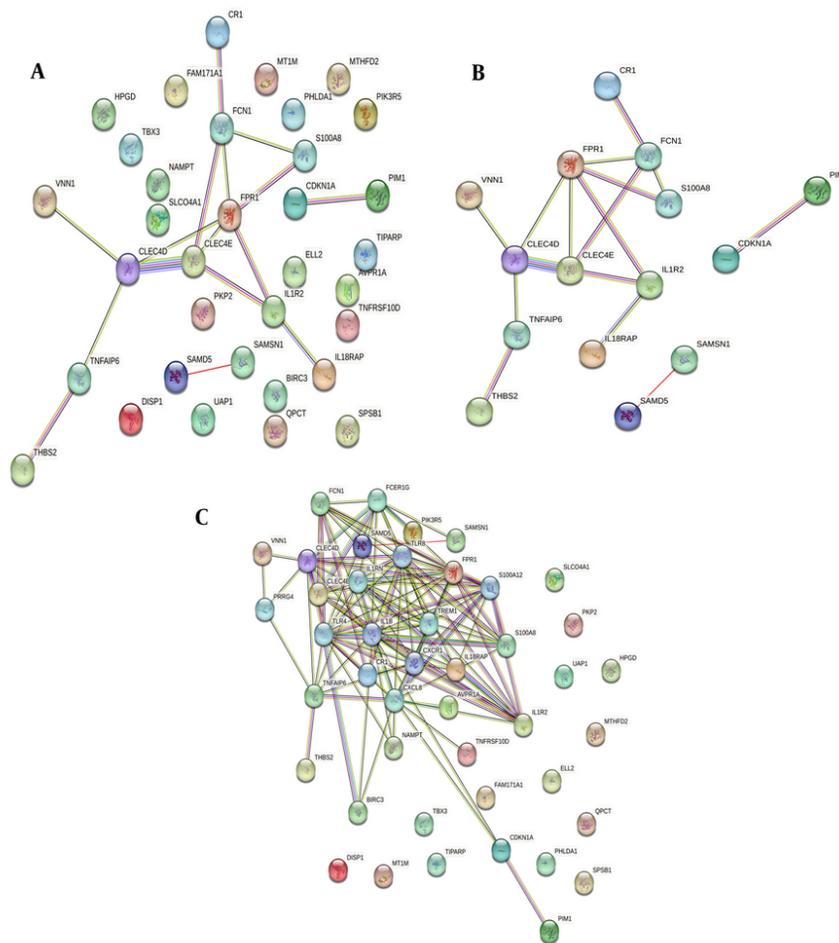
#### 4.4. Establishment of Non-alcoholic Fatty Liver Disease and Chronic Obstructive Pulmonary Disease Cell Models

Oil Red O, a lipophilic dye that can dissolve in fats, specifically stains neutral fats like triglycerides in cells red. Hematoxylin, a basic natural dye, can stain cell nuclei (18). As shown in Figure 7A and B, the lipid content in the NAFLD model group was markedly higher. The TG level in the NAFLD model group was significantly higher than in the control group (in Figure 7C). The COPD is accompanied by severe inflammatory responses. COX-2 can further transform arachidonic acid into eicosanoids, including prostaglandins.

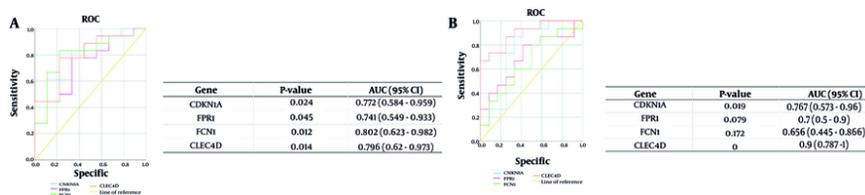
Research reveals that COX-2 expression is associated with the occurrence or amplification of inflammation (19). Therefore, COX-2 is chosen here to verify the successful establishment of the COPD model. Compared with the control group, COX-2 expression in A549 cells in the COPD model group was significantly increased ( $P < 0.001$ ), as shown in Figure 7D and E, indicating a marked inflammatory response in the cells.

#### 4.5. Detection of Cycle Protein-Dependent Kinase Inhibitor 1A, Formyl Peptide Receptor 1, Ficolin-1, and CLEC4D Protein Expression by Western Blot

The CDKN1A, FPR1, FCN1, and CLEC4D protein expression was detected by Western blot. The NAFLD model group indicated a pronounced enhancement in



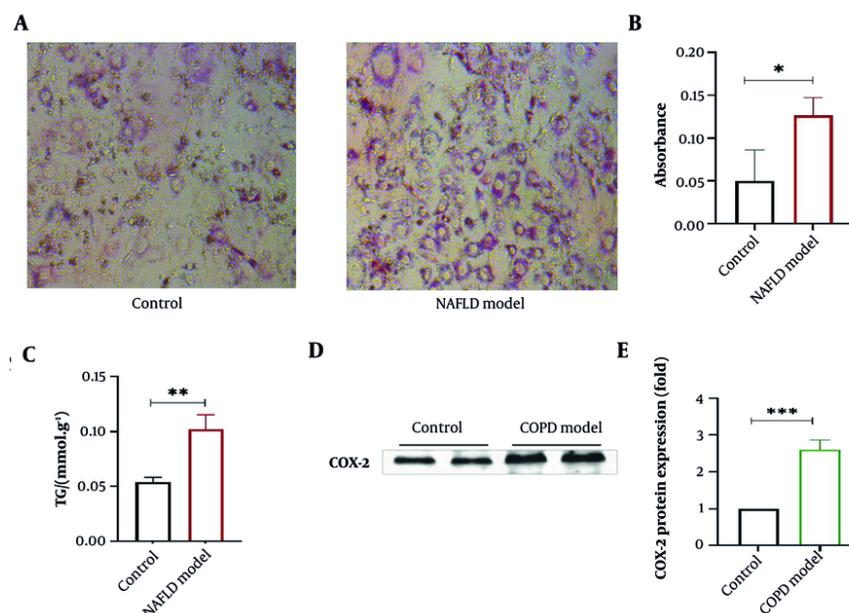
**Figure 5.** Inter-gene interactions – A, differential gene protein interaction network; B, differential protein with interaction; C, expanded interacting protein



**Figure 6.** Receiver operating characteristic (ROC) curve of hub genes – A, ROC information of hub genes in chronic obstructive pulmonary disease (COPD); B, ROC information of non-alcoholic fatty liver disease (NAFLD)

*CDKN1A*, *FPR1*, *FCN1*, and *CLEC4D* protein expression compared with the blank group (Figure 8A). Similarly,

the COPD model group exhibited a considerable enhancement in *CDKN1A*, *FPR1*, *FCN1*, and *CLEC4D* protein



**Figure 7.** Establishment of non-alcoholic fatty liver disease (NAFLD) and chronic obstructive pulmonary disease (COPD) cell models – A, oil Red O staining of cells in the NAFLD group; B, oil Red O staining absorbance comparison; C, comparison of triglyceride levels in normal and NAFLD cells; D, COX protein gel; E, western blot statistical chart – compared with control group, \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ,  $n=3$  (\* indicates a significant difference).

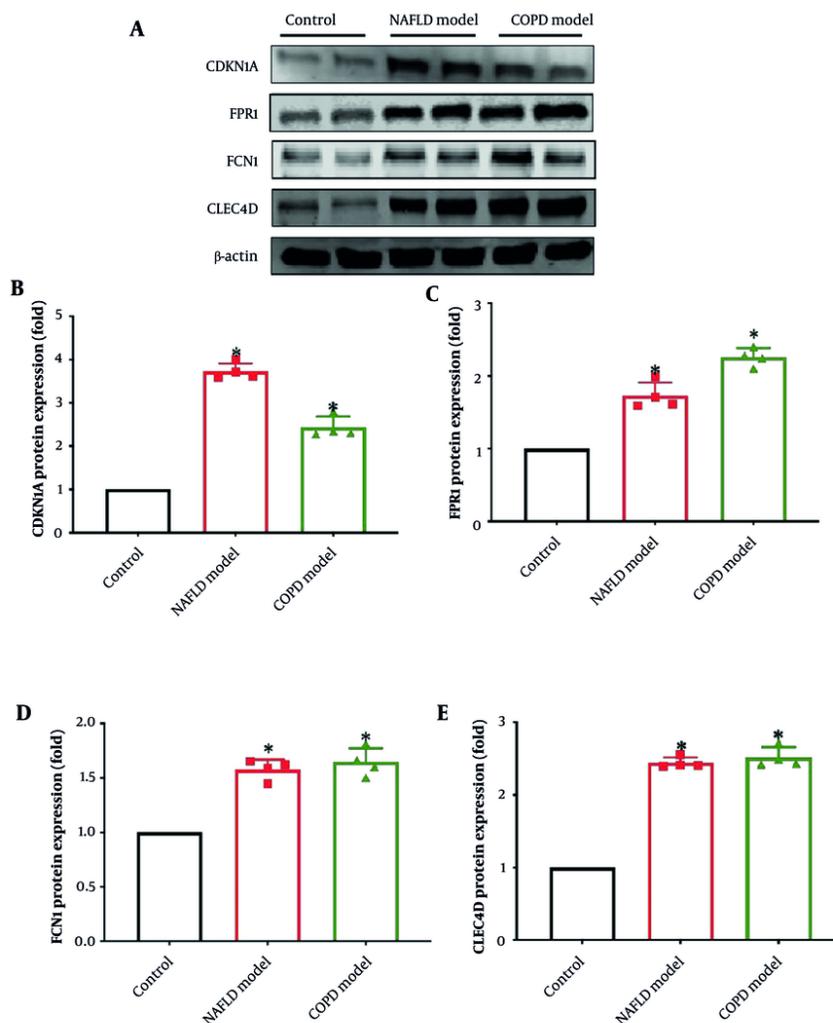
expression compared to the blank group, with *CLEC4D* protein expression also significantly increased (Figure 8B - E). All the differences were statistically significant ( $P < 0.05$ ). This suggests that *CDKN1A*, *FPR1*, *FCN1*, and *CLEC4D* are likely to play a crucial role in the initiation and progression of NAFLD and COPD.

## 5. Discussion

Both NAFLD and COPD are quite common around the world. Previous studies have shown that patients with COPD are more likely to have problems with NAFLD (20). The progressive increase in risk with longer disease duration necessitates further investigation. In this study, we explored NAFLD and COPD datasets from GEO and identified 34 shared DEGs. We conducted GO and KEGG pathway enrichment analyses and built a PPI network to pinpoint key genes within the common DEGs, shedding light on the molecular mechanisms and potential early targets for preventing disease progression in both conditions. We selected four key genes to evaluate their diagnostic potential in patients with NAFLD and COPD ( $P < 0.05$ ). It is possible that these

genes may serve as valuable predictors for the risk of developing COPD and NAFLD.

Cell *CDKN1A* is a biomarker of cell cycle arrest and premature senescence, and it has been found that *CDKN1A* expression is higher in the peripheral lungs of patients with COPD, which may be associated with muscle dysfunction (21). Skeletal muscle atrophy is a frequent complication in COPD patients (22). The *CDKN1A* was upregulated in the muscle tissue of COPD patients compared to healthy controls (23). The COPD patients show altered levels of *CDKN1A*, indicating its potential role in the development of COPD (19). The level of *CDKN1A* protein was also found to change significantly during the progression of NASH to the development of HCC (24), and a reasonable hypothesis has been put forward for the effect of *CDKN1A* in NAFLD, suggesting that the *CDKN1A* rs762623 variant may help protect hepatocytes from advanced fibrosis by mitigating their senescence (25). In COPD, *CDKN1A* is elevated in peripheral lungs and muscle tissue, potentially contributing to muscle dysfunction and skeletal muscle atrophy (26). In NAFLD, *CDKN1A* levels



**Figure 8.** The protein expression of cycle protein-dependent kinase inhibitor 1A (*CDKN1A*), formyl peptide receptor 1 (*FPR1*), *ficolin-1* (*FCN1*), *CLEC4D* – A, each protein strip graph; B, *CDKN1A* protein expression; C, *FPR1* protein expression; D, *FCN1* protein expression; E, *CLEC4D* protein expression – compared with control group, \*  $P < 0.05$ ,  $n = 3$

change during NASH progression to HCC, with a variant possibly protecting hepatocytes from fibrosis (27, 28).

The *FPR1* is a member of the GPCR superfamily. It is predominantly expressed in mammalian leukocytes and plays a crucial role in inflammatory responses, as well as in the regulation of brain homeostasis (29). When hepatic necrosis was induced in mice, a rapid accumulation of neutrophils was found, but when *FPR1* was inhibited, the migration of neutrophils to the area of hepatic necrosis was significantly reduced (30). The *FPR1*/*FPR2* heterodimerization triggers a delayed and

induced pro-inflammatory response to the JNK pathway with neutrophil apoptosis (31). The *FPR1* has been shown to be an important receptor in COPD, as the gene alteration contributes to protection against cigarette smoke-induced emphysema formation in a mouse model, consistent with other findings that *FPR1* expression is increased in neutrophils from COPD patients with high levels of dyspnea (32). The *FPR1*, a GPCR family member crucial for inflammatory responses, shows altered expression in COPD patients and mouse models, indicating its role in cigarette

smoke-induced emphysema and neutrophil regulation (32). In NAFLD, FPR1 inhibition reduces neutrophil migration to hepatic necrosis sites, suggesting its involvement in liver inflammation (33).

The FCN1, encoded by the *FCN1* gene, is synthesized in the bone marrow and is present in type II alveolar epithelial cells, granulocytes, and monocytes (34). This protein contains collagen-like domains that interact with mannose-binding lectin-associated serine protease (MASP), thereby triggering the complement lectin pathway (LP) cascade upon target binding (35). Notably, the interaction of M-fibrillar collagen with natural killer (NK) cells differs from that with activated T cell subsets. While NK cells interact directly with M-fibrillar collagen, T cell subsets engage via specific ligand binding sites in the FBG structural domain, bridging adaptive and innate immunity (36). Elevated serum levels of M-ficolin have been linked to exacerbated inflammation, potentially leading to poorer outcomes in pediatric pneumonia (37).

Recent multi-omics research has identified *FCN1* as a key gene in NAFLD, implicating it in disease development, particularly in regulating fat accumulation and liver inflammation (38). *FCN1* also appears to influence immune cell infiltration, possibly through ceRNA regulation, which is critical in atherosclerosis progression (39). In NAFLD, *FCN1* activates the complement system via the LP by recognizing and binding to specific carbohydrates and injured cells, such as apoptotic cells. This binding activates associated serine proteases like MASP, which then cleave C4 and C2 to form the C3 convertase. The subsequent cleavage of C3 into C3a and C3b initiates a cascade of complement activation reactions, contributing to inflammation and liver injury. For example, a study found that *FCN1* is involved in humoral immune responses, complement activation, and phagocytosis in PIBD (40). Our study further solidifies *FCN1*'s involvement in NAFLD-related fat accumulation, liver inflammation, and immune cell infiltration.

In COPD's pathologic process, macrophages mainly polarize to the M1 type. M1 macrophages secrete pro-inflammatory cytokines like *IL-6*, *TNF- $\alpha$* , and *IL-1 $\beta$* , recruiting other immune cells and worsening pulmonary inflammation (41). *CLEC4D* (Member 4 of the C-type lectin family), predominantly expressed in macrophages and myeloid cells, plays a role in the body's defense against mycobacterial infections (37). In

humans, *CLEC4D* gene polymorphisms with reduced expression have been associated with increased susceptibility to tuberculosis (42). *CLEC4D* also impacts chronic diseases, being implicated in causing type 1 diabetes mellitus (T1D) and contributing to adipose tissue fibrosis through the interaction between adipocytes and macrophages (41). Previous studies link *CLEC4D* polymorphisms to tuberculosis susceptibility and T1D development (42). Our research adds to this by showing its involvement in adipose tissue fibrosis and macrophage responses to viral infections.

*CLEC4D* likely regulates macrophage polarization in COPD through multiple mechanisms (43). It may recognize components in cigarette smoke or lung DAMPs, activating macrophage signaling pathways like NF- $\kappa$ B to promote M1 polarization (44). Additionally, *CLEC4D* may interact with other macrophage receptors or signaling molecules, modulating intracellular signaling and influencing polarization-related gene expression to drive the shift to the M1 phenotype (41). However, the exact molecular mechanisms and signaling transduction require further study.

In COPD's pathologic process, macrophages mainly polarize to the M1 type. M1 macrophages secrete pro-inflammatory cytokines like *IL-6*, *TNF- $\alpha$* , and *IL-1 $\beta$* , recruiting other immune cells and worsening pulmonary inflammation. *CLEC4D* likely regulates this polarization through multiple mechanisms. It may recognize components in cigarette smoke or lung DAMPs, activating macrophage signaling pathways like NF- $\kappa$ B to promote M1 polarization. Additionally, *CLEC4D* may interact with other macrophage receptors or signaling molecules, modulating intracellular signaling and influencing polarization-related gene expression to drive the shift to the M1 phenotype. However, the exact molecular mechanisms and signaling transduction require further study (45).

In this study, COPD and NAFLD datasets were extracted from GEO, and a comparative analysis between patient and normal samples identified 34 shared genes. The GO and KEGG enrichment analyses were then conducted to identify immune-related biological functions and pathways. Key genes are emerging as central to these mechanisms. The pivotal genes *CDKN1A*, *FPR1*, *FCN1*, and *CLEC4D* were selected by constructing a PPI network. We identified key genes, and their diagnostic potential was validated through data analysis with SPSS. The important roles of the four key

genes (*CDKN1A*, *FPR1*, *FCN1*, and *CLEC4D*) in the disease were preliminarily verified by in vitro establishment of NAFLD and COPD models. Our research outcomes might offer promising therapeutic targets for the management and prevention of NAFLD and COPD.

However, our study has certain limitations. First, the DEGs identified via bioinformatics could potentially serve as predictive biomarkers for NAFLD and COPD development. Yet, the lack of in vitro validation and clinical cohort studies restricts the confirmation and generalizability of these findings. Second, the analyzed samples come from diverse ethnic groups, so the applicability of these results may be limited to specific populations and may not necessarily apply to all demographic groups. Moreover, the sample sizes are relatively small, which may lead to biases and limit the representativeness of the broader population. Additionally, our study did not specifically examine potential biases related to age or gender differences, which could further influence the generalizability of our findings. Therefore, further studies are needed in other populations, with larger sample sizes, and more detailed consideration of ethnic, age, and gender factors to validate our results and ensure their broader applicability.

In our study, several limitations should be acknowledged. We did not account for batch effects in GEO datasets. Such batch effects, arising from variations in experimental conditions, processing times, or technical factors, might have influenced our analysis results. Additionally, the risk of cell line misidentification was not considered. Misidentified or cross-contaminated cell lines could lead to inaccurate findings and affect the reliability of our conclusions. Moreover, while HepG2 and A549 cells provided molecular insights into NAFLD and COPD, they do not fully mirror human pathophysiology. Future studies should use patient-derived samples like liver/lung biopsies to boost clinical relevance.

Also, the sample sizes in GSE63067 and GSE37768 were small. Although the high power and significant effect sizes suggest robust results, the limited samples may restrict generalizability. Despite these limitations, this study reveals key shared molecular mechanisms between NAFLD and COPD and pinpoints genes for further study. These insights enhance our understanding of disease pathogenesis and lay a foundation for future research. In future studies, we will

implement stricter quality control measures and validation steps to address the aforementioned issues and minimize their potential impact.

### 5.1. Conclusions

The results of this study show that *CDKN1A*, *FPR1*, *FCN1*, and *CLEC4D* are key genes associated with NAFLD and COPD. The key genes screened have been preliminarily verified in cell experiments and are likely to contribute significantly to the pathogenesis and progression of NAFLD and COPD, thus providing a new theoretical basis for the occurrence of these diseases.

### Supplementary Material

Supplementary material(s) is available [here](#) [To read supplementary materials, please refer to the journal website and open PDF/HTML].

### Footnotes

**Authors' Contribution:** Conceptualization: M. D. Z.; Data curation: M. D. Z.; Methodology: Y. W.; Project administration: Y. W.; Resources: Y. X. H.; Software: Y. W.; Supervision: X. L. L.; Validation: M. D. Z.; Visualization: T. Y. B.; Writing-original draft: W. Z. H. F.; Writing-review & editing: W. Z. H. F.

**Conflict of Interests Statement:** The authors declare no conflicts of interest.

**Data Availability:** The data presented in this study are uploaded during submission as a supplementary file and are openly available for readers upon request.

**Ethical Approval:** In the present study, the protocols were approved by the Animal Protection and Use Committee of Inner Mongolia Medical University (approval number: YKD202403097).

**Funding/Support:** This research was funded by "Natural Science Foundation of Inner Mongolia Autonomous Region, Grant No. 2023QN08016", "General Project of Inner Mongolia Medical University, Grant No. YKD2025MS049 、 YKD2021MS016", "Inner Mongolia Autonomous Region Key Technology Research and Development Programs, Grant No. 2021GG0173" and "National Natural Science Foundation of China, Grant No. 82260733".

## References

- Yu H, Wang Z, Zhao G. Chronic Liver Disease and Promising Therapeutic Strategy: A Concise Review. *Pharmacogn Mag.* 2024;**20**(4):1031-47. <https://doi.org/10.1177/09731296241228361>.
- Yan Y, Gan D, Zhang P, Zou H, Li M. A machine learning-based predictive model discriminates nonalcoholic steatohepatitis from nonalcoholic fatty liver disease. *Heliyon.* 2024;**10**(21): e38848. [PubMed ID: 39512464]. [PubMed Central ID: PMC11539579]. <https://doi.org/10.1016/j.heliyon.2024.e38848>.
- Luo Z, Dai X, Liu D, Li J, Wang H, Liang J, et al. Bi-level electricity and heat sharing strategy for cross-border integrated energy system based on Nash game. *Appl Therm Eng.* 2025;**258**. <https://doi.org/10.1016/j.applthermaleng.2024.124704>.
- de Celis Alonso B, Shumbayawonda E, Beyer C, Hidalgo-Tobon S, Lopez-Martinez B, Dies-Suarez P, et al. Liver magnetic resonance imaging, non-alcoholic fatty liver disease and metabolic syndrome risk in pre-pubertal Mexican boys. *Sci Rep.* 2024;**14**(1):26104. [PubMed ID: 39478096]. [PubMed Central ID: PMC11526175]. <https://doi.org/10.1038/s41598-024-77307-8>.
- Halpin DM, Criner GJ, Papi A, Singh D, Anzueto A, Martinez FJ, et al. Global initiative for the diagnosis, management, and prevention of chronic obstructive lung disease. The 2020 GOLD science committee report on COVID-19 and chronic obstructive pulmonary disease. *Am J Crit Care Med.* 2021;**203**(1):24-36.
- Halpin DMG, Celli BR, Criner GJ, Frith P, Lopez Varela MV, Salvi S, et al. The GOLD Summit on chronic obstructive pulmonary disease in low- and middle-income countries. *Int J Tuberc Lung Dis.* 2019;**23**(11):1131-41. [PubMed ID: 31718748]. <https://doi.org/10.5588/ijtld.19.0397>.
- Chen S, Kuhn M, Prettnner K, Yu F, Yang T, Barnighausen T, et al. The global economic burden of chronic obstructive pulmonary disease for 204 countries and territories in 2020-50: A health-augmented macroeconomic modelling study. *Lancet Glob Health.* 2023;**11**(8):e1183-93. [PubMed ID: 37474226]. [PubMed Central ID: PMC10369014]. [https://doi.org/10.1016/S2214-109X\(23\)00217-6](https://doi.org/10.1016/S2214-109X(23)00217-6).
- Du W, Tang H, Ma W, Li X, Li K, Wang Q, et al. Effect of noninvasive ventilation combined with pulmonary rehabilitation on diaphragm function and serum inflammatory factors in COPD patients with hypercapnia. *Minerva Surg.* 2024;**79**. <https://doi.org/10.23736/s2724-5691.24.10357-7>.
- Zheng D, Liu X, Zeng W, Zhou W, Zhou C. Association of hepatic steatosis and liver fibrosis with chronic obstructive pulmonary disease among adults. *Sci Rep.* 2024;**14**(1):10822. [PubMed ID: 38734742]. [PubMed Central ID: PMC11088642]. <https://doi.org/10.1038/s41598-024-61696-x>.
- Liu K, Kang M, Li J, Qin W, Wang R. Prognostic value of the mRNA expression of members of the HSP90 family in non-small cell lung cancer. *Exp Ther Med.* 2019;**17**(4):2657-65. <https://doi.org/10.3892/etm.2019.7228>.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;**43**(7): e47. [PubMed ID: 25605792]. [PubMed Central ID: PMC4402510]. <https://doi.org/10.1093/nar/gkv007>.
- Liu Y, Chen M. Histone Demethylation Profiles in Nonalcoholic Fatty Liver Disease and Prognostic Values in Hepatocellular Carcinoma: A Bioinformatic Analysis. *Curr Issues Mol Biol.* 2023;**45**(4):3640-57. [PubMed ID: 37185761]. [PubMed Central ID: PMC10136463]. <https://doi.org/10.3390/cimb45040237>.
- Dai W, Sun Y, Jiang Z, Du K, Xia N, Zhong G. Key genes associated with non-alcoholic fatty liver disease and acute myocardial infarction. *Med Sci Monit.* 2020;**26**. <https://doi.org/10.12659/msm.922492>.
- Lu C, Han M, Ma Q, Ying L, Zhang Y. Identification of biomarkers associated with coronary artery disease and non-alcoholic fatty liver disease by bioinformatics analysis and machine learning. *Sci Rep.* 2025;**15**(1):3557. [PubMed ID: 39875572]. [PubMed Central ID: PMC11775188]. <https://doi.org/10.1038/s41598-025-87923-7>.
- Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* 2006;**34**(Database issue):D322-6. [PubMed ID: 16381878]. [PubMed Central ID: PMC1347384]. <https://doi.org/10.1093/nar/gkj021>.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;**43**(Database issue):D447-52. [PubMed ID: 25352553]. [PubMed Central ID: PMC4383874]. <https://doi.org/10.1093/nar/gku1003>.
- Lopez-Campos JL, Soler-Cataluna JJ, Miravittles M. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2019 Report: Future Challenges. *Arch Bronconeumol (Engl Ed).* 2020;**56**(2):65-7. [PubMed ID: 31320191]. <https://doi.org/10.1016/j.arbres.2019.06.001>.
- Andres-Manzano MJ, Andres V, Dorado B. Oil Red O and Hematoxylin and Eosin Staining for Quantification of Atherosclerosis Burden in Mouse Aorta and Aortic Root. *Methods Mol Biol.* 2015;**1339**:85-99. [PubMed ID: 26445782]. [https://doi.org/10.1007/978-1-4939-2929-0\\_5](https://doi.org/10.1007/978-1-4939-2929-0_5).
- Boss AP, Freeborn RA, Duriancik DM, Kennedy RC, Gardner EM, Rockwell CE. The Nrf2 activator tBHQ inhibits the activation of primary murine natural killer cells. *Food Chem Toxicol.* 2018;**121**:231-6. [PubMed ID: 30171972]. [PubMed Central ID: PMC6287942]. <https://doi.org/10.1016/j.fct.2018.08.067>.
- Vigilino D, Jullian-Desayes I, Minoves M, Aron-Wisniewsky J, Leroy V, Zarski JP, et al. Nonalcoholic fatty liver disease in chronic obstructive pulmonary disease. *Eur Respir J.* 2017;**49**(6). [PubMed ID: 28596431]. <https://doi.org/10.1183/13993003.01923-2016>.
- Yang J, Zhang M, Du Y, Ji X, Qu Y. Identification and validation of CDKN1A and HDAC1 as senescence-related hub genes in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2022;**17**:1811-25.
- Lakhdar R, Drost EM, MacNee W, Bastos R, Rabinovich RA. 2D-DIGE proteomic analysis of vastus lateralis from COPD patients with low and normal fat free mass index and healthy controls. *Respir Res.* 2017;**18**(1):81. [PubMed ID: 28468631]. [PubMed Central ID: PMC5415759]. <https://doi.org/10.1186/s12931-017-0525-x>.
- Willis-Owen SAG, Thompson A, Kemp PR, Polkey MI, Cookson W, Moffatt MF, et al. COPD is accompanied by co-ordinated transcriptional perturbation in the quadriceps affecting the mitochondria and extracellular matrix. *Sci Rep.* 2018;**8**(1):12165. [PubMed ID: 30111857]. [PubMed Central ID: PMC6093887]. <https://doi.org/10.1038/s41598-018-29789-6>.
- Yang D, Yan Y, Hu F, Wang T. CYP1B1, VEGFA, BCL2, and CDKN1A Affect the Development of Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis.* 2020;**15**:167-75. [PubMed ID: 32158203]. [PubMed Central ID: PMC6986178]. <https://doi.org/10.2147/COPD.S220675>.
- de Conti A, Ortega JF, Tryndyak V, Dreval K, Moreno FS, Rusyn I, et al. MicroRNA deregulation in nonalcoholic steatohepatitis-associated

- liver carcinogenesis. *Oncotarget*. 2017;**8**(51):88517-28. [PubMed ID: 29179453]. [PubMed Central ID: PMC5687623]. <https://doi.org/10.18632/oncotarget.19774>.
26. McCluskey E, Velli SK, Kaminski R, Markward T, Leming H, Yu D, et al. HOXA1 Contributes to Bronchial Epithelial Cell Cycle Progression by Regulating p21/CDKN1A. *Int J Mol Sci*. 2025;**26**(5). [PubMed ID: 40076953]. [PubMed Central ID: PMC11899960]. <https://doi.org/10.3390/ijms26052332>.
  27. Dong C, Huoshen W, Bai Y, Liu J, Li B, Guan Y, et al. Uncovering the molecular networks of ferroptosis in the pathogenesis of type 2 diabetes and its complications: a multi-omics investigation. *Mol Med*. 2024;**30**(1):268. [PubMed ID: 39716081]. [PubMed Central ID: PMC11665190]. <https://doi.org/10.1186/s10020-024-01045-w>.
  28. Valenti L, Dongiovanni P. CDKN1A: A double-edged sword in fatty liver? *Cell Cycle*. 2014;**13**(9):1371-2. [PubMed ID: 24699009]. [PubMed Central ID: PMC4050133]. <https://doi.org/10.4161/cc.28754>.
  29. Wang Z, Wang Y, Yan Q, Cai C, Feng Y, Huang Q, et al. FPR1 signaling aberrantly regulates S100A8/A9 production by CD14(+)FCN1(hi) macrophages and aggravates pulmonary pathology in severe COVID-19. *Commun Biol*. 2024;**7**(1):1321. [PubMed ID: 39402337]. [PubMed Central ID: PMC11473795]. <https://doi.org/10.1038/s42003-024-07025-4>.
  30. Honda M, Takeichi T, Hashimoto S, Yoshii D, Isono K, Hayashida S, et al. Intravital Imaging of Neutrophil Recruitment Reveals the Efficacy of FPR1 Blockade in Hepatic Ischemia-Reperfusion Injury. *J Immunol*. 2017;**198**(4):1718-28. [PubMed ID: 28062700]. <https://doi.org/10.4049/jimmunol.1601773>.
  31. Dorward DA, Lucas CD, Chapman GB, Haslett C, Dhaliwal K, Rossi AG. The role of formylated peptides and formyl peptide receptor 1 in governing neutrophil function during acute inflammation. *Am J Pathol*. 2015;**185**(5):1172-84. [PubMed ID: 25791526]. [PubMed Central ID: PMC4419282]. <https://doi.org/10.1016/j.ajpath.2015.01.020>.
  32. Cardini S, Dalli J, Fineschi S, Perretti M, Lungarella G, Lucatelli M. Genetic ablation of the fpr1 gene confers protection from smoking-induced lung emphysema in mice. *Am J Respir Cell Mol Biol*. 2012;**47**(3):332-9. [PubMed ID: 22461430]. <https://doi.org/10.1165/rcmb.2012-0036OC>.
  33. Wang S, Lu K, Lin L, Li G, Han Y, Lin Z, et al. Exploring the mechanism of berberine treatment for atherosclerosis combined with non-alcoholic fatty liver disease based on bioinformatic and experimental study. *PLoS One*. 2024;**19**(12). e0314961. [PubMed ID: 39700090]. [PubMed Central ID: PMC11658604]. <https://doi.org/10.1371/journal.pone.0314961>.
  34. Chen YC, Lin MC, Lee CH, Liu SF, Wang CC, Fang WF, et al. Defective formyl peptide receptor 2/3 and annexin A1 expressions associated with M2a polarization of blood immune cells in patients with chronic obstructive pulmonary disease. *J Transl Med*. 2018;**16**(1):69. [PubMed ID: 29544524]. [PubMed Central ID: PMC5856198]. <https://doi.org/10.1186/s12967-018-1435-5>.
  35. Bhagwat SR, Choudhary K, Pandya N, Sharma S, Srivastava S, Kumar A, et al. Identification of substrates of MBL Associated Serine Protease-1 (MASP-1) from human plasma using N-terminomics strategy. *Mol Immunol*. 2022;**151**:114-25.
  36. Endo Y, Matsushita M, Fujita T. New insights into the role of ficolins in the lectin pathway of innate immunity. *Int Rev Cell Mol Biol*. 2015;**316**:49-110. [PubMed ID: 25805122]. <https://doi.org/10.1016/bs.ircmb.2015.01.003>.
  37. Genster N, Ma YJ, Munthe-Fog L, Garred P. The pattern recognition molecule ficolin-1 exhibits differential binding to lymphocyte subsets, providing a novel link between innate and adaptive immunity. *Mol Immunol*. 2014;**57**(2):181-90. [PubMed ID: 24161415]. <https://doi.org/10.1016/j.molimm.2013.09.006>.
  38. Elkoumi MA, Abdellatif SH, Mohamed FY, Sherif AH, Elashkar SSA, Saleh RM, et al. Ficolin-1 gene (FCN1) -144 C/A polymorphism is associated with adverse outcome of severe pneumonia in the under-five Egyptian children: A multicenter study. *Pediatr Pulmonol*. 2020;**55**(5):1175-83. [PubMed ID: 32142211]. <https://doi.org/10.1002/ppul.24719>.
  39. Lin J, Zhang R, Liu H, Zhu Y, Dong N, Qu Q, et al. Multi-omics analysis of the biological mechanism of the pathogenesis of non-alcoholic fatty liver disease. *Front Microbiol*. 2024;**15**:1379064. [PubMed ID: 39132138]. [PubMed Central ID: PMC11310135]. <https://doi.org/10.3389/fmicb.2024.1379064>.
  40. Chen X, Gao Y, Xie J, Hua H, Pan C, Huang J, et al. Identification of FCN1 as a novel macrophage infiltration-associated biomarker for diagnosis of pediatric inflammatory bowel diseases. *J Transl Med*. 2023;**21**(1):203. [PubMed ID: 36932401]. [PubMed Central ID: PMC10022188]. <https://doi.org/10.1186/s12967-023-04038-1>.
  41. Kim GD, Lim EY, Shin HS. Macrophage Polarization and Functions in Pathogenesis of Chronic Obstructive Pulmonary Disease. *Int J Mol Sci*. 2024;**25**(11). [PubMed ID: 38891820]. [PubMed Central ID: PMC1172060]. <https://doi.org/10.3390/ijms25115631>.
  42. Xia X, Wang M, Li J, Chen Q, Jin H, Liang X, et al. Identification of potential genes associated with immune cell infiltration in atherosclerosis. *Math Biosci Eng*. 2021;**18**(3):2230-42. [PubMed ID: 33892542]. <https://doi.org/10.3934/mbe.2021112>.
  43. Zhou W, Wu L, Song J, Jiao L, Zhou Y, Zhou J, et al. A case-control study on correlation between the single nucleotide polymorphism of CLEC4E and the susceptibility to tuberculosis among Han people in Western China. *BMC Infect Dis*. 2021;**21**(1):788. [PubMed ID: 34376176]. [PubMed Central ID: PMC8353747]. <https://doi.org/10.1186/s12879-021-06448-2>.
  44. Yip L, Fuhlbrigge R, Alkhataybeh R, Fathman CG. Gene Expression Analysis of the Pre-Diabetic Pancreas to Identify Pathogenic Mechanisms and Biomarkers of Type 1 Diabetes. *Front Endocrinol (Lausanne)*. 2020;**11**:609271. [PubMed ID: 33424774]. [PubMed Central ID: PMC7793767]. <https://doi.org/10.3389/fendo.2020.609271>.
  45. Jin S, Wu Z. Study on Immune-Related Genes and Clinical Validation of Acute Myocardial Infarction Based on Bioinformatics. *Biochem Genet*. 2025;**29**(31). <https://doi.org/10.1007/s10528-025-11029-y>.