scientific reports

Check for updates

OPEN

Prediction model of mitochondrial energy metabolism related genes in idiopathic pulmonary fibrosis and its correlation with immune microenvironment

Linlin Yao[2,3,4], Baoyan Liu[2,3,4] & Yong Wang[1]✉

Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease. Recent evidence suggests that the pathogenesis of IPF may involve abnormalities in mitochondrial energy metabolism. This study aimed to identify mitochondrial energy metabolism related differentially expressed genes (MEMRDEGs) and to elucidate their potential mechanistic involvement in IPF. We employed a multistep bioinformatics approach, including data extraction from the Gene Expression Omnibus database, removal of batch effects, and normalization and differential gene expression analyses. We then conducted Gene Ontology, Kyoto Encyclopedia of Genes and Genomes enrichment, and gene set enrichment analyses. A protein-protein interaction network was constructed from the STRING database, and hub genes were identified. Receiver operating characteristic curve analysis was performed to evaluate immune infiltration. Our integrated analysis of IPF datasets identified 25 MEMRDEGs. Nine hub genes emerged as central to mitochondrial energy metabolism in IPF. *COX5A*, *EHHADH*, and *SDHB* are potential biomarkers for diagnosing IPF with high accuracy. Single-sample gene set enrichment analysis revealed significant differences in the abundances of specertainfic immune cell types between IPF samples and controls. In conclusion, *COX5A*, *EHHADH*, and *SDHB* are potential biomarkers for the high-accuracy diagnosis of IPF. These findings pave the way for further investigations into the molecular mechanisms underlying IPF.

Keywords Idiopathic pulmonary fibrosis, Immune microenvironment, Bioinformatics analysis, Mitochondrial energy metabolism, Genes

Abbreviations

EHHADH	Enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase
ACAT1	Acetyl-CoA acetyltransferase 1
ECI2	Enoyl-CoA delta isomerase 2
ACADS	Acyl-CoA dehydrogenase short chain
CS	Citrate synthase
SDHB	Succinate dehydrogenase complex iron sulfur subunit B
COX5A	Cytochrome C oxidase subunit 5 A
ETFA	Electron transfer flavoprotein subunit alpha
ACADL	Acyl-CoA dehydrogenase long chain

Idiopathic pulmonary fibrosis (IPF) is an interstitial lung disease with unknown causes that mainly manifests as a chronic progressive development and primarily affects older people. The global incidence of IPF is estimated to be approximately 10.7 per 100,000 person-years with a median survival time of 3–5 years[1]. Despite recent

[1] Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences, Jinan 250117, Shandong Province, China. [2]Shandong First Medical University affiliated occupational disease Hospital (Shandong Occupational Disease Hospital), Jinan 250062, Shandong Province, China. [3]Shandong Academy of Occupational Health and Occupational Medicine, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, China. [4]Linlin Yao and Baoyan Liu contributed equally to this work. ✉email: doctorwy@163.com

advances in therapeutic interventions, options remain limited and are often associated with significant side effects and only modest efficacy in slowing disease progression[2]. Lung transplantation remains the only curative option; however, it is available for only a small subset of patients, owing to stringent eligibility criteria and organ availability[3]. Early diagnosis is conducive to timely intervention, delaying the progression of the disease.The need for improved diagnostic and therapeutic strategies underscores the urgency of research into the molecular mechanisms underlying IPF. Our primary research objective is to provide a basis for improving early diagnosis and guiding treatment strategies by identifying diagnostic biomarkers in IPF.

Mitochondrial dysfunction has been implicated in various pulmonary diseases, including IPF, and alterations in energy metabolism may contribute to its pathophysiology[4]. Studies have shown that mitochondrial energy metabolism related differentially expressed genes (MEMRDEGs) play pivotal roles in other fibrotic diseases, suggesting common therapeutic pathways[5]. Moreover, mitochondrial dynamics and bioenergetics have been linked to cellular senescence and myofibroblast differentiation, which are key features of IPF pathology[6]. These findings highlight the potential of MEMRDEGs as diagnostic biomarkers and therapeutic targets in IPF.

In recent years, studies have shown that the immune microenvironment plays a crucial role in the pathogenesis of IPF. Immune cells, incluchding macrophages and T cells, promote the fibrotic process by releasing pro-inflammatory and profibrotic cytokines (e.g., TGF-β, IL-6)[7,8]. Meanwhile, mitochondrial dysfunction is closely related to the immune response, which may further exacerbate IPF through oxidative stress and metabolic[9].

Our study aimed to elucidate the role of MEMRDEGs in IPF by integrating multi-platform datasets and employing a series of bioinformatic analyses. We conducted differential expression analysis, functional enrichment studies, protein-protein interaction (PPI) network construction, regulatory network mapping, diagnostic receiver operating characteristic (ROC) curve evaluation, and immune infiltration evaluation. Gene network analysis and complex network topology are integral to understanding these processes and have been effectively utilized in recent bioinformatics research[10,11].This comprehensive approach enabled the identification of the key genes and pathways involved in IPF pathogenesis.

Our study provides new insights into the mitochondrial energy metabolism signature in IPF and offers promising avenues for future diagnostic and therapeutic development. Identifying the crucial MEMRDEGs and their regulatory networks will enhance our understanding of the molecular mechanisms underlying this debilitating disease.

Materials and methods

Data download

We obtained the IPF datasets GSE24206[12,13] and GSE110147[13,14] from the Gene Expression Omnibus (GEO) database[15] using the R package GEOquery[16] (https://www.ncbi.nlm.nih.gov/geo/). The samples from the GSE24206 and GSE110147 datasets were all from *Homo sapiens*, and the tissue source was the lungs. The chip platform for dataset GSE24206 was GPL570, while the chip platform for dataset GSE110147 was GPL6244. Specific information is shown in Table 1.The dataset GSE24206 contained 17 IPF and 6 control samples. This study included both IPF and control samples. Dataset GSE110147 contained 22 IPF, 10 non-specific interstitial pneumonia, 5 IPF-non-specific interstitial pneumonia and 11 control samples, additionally 22 IPF and 11 control samples were included in this study.

We collected Mitochondrial Energy Metabolism related genes(MEMRGs) from the GeneCards database[17] (https://www.genecards.org/) and relevant literature. We used the term "Mitochondrial Energy Metabolism" as the keyword, and 47 MEMRGs were obtained after keeping only genes with "Protein Coding" and "Relevance Score > 1." Additionally, "Mitochondrial Energy Metabolism" was used as a keyword on the PubMed website (https://pubmed.ncbi.nlm.nih.gov/) to find published literature[18] on the Mitochondrial Energy Metabolism related genes set. A total of 28 MEMRGs were included in this study. A total of 74 MEMRGs were obtained after combined duplication removal; detailed information is shown in Table S1.

The R package sva[19] was used to debatch GSE24206 and GSE110147 to obtain the combined GEO datasets. The combined datasets included 39 IPF and 17 control samples. Finally, we used the R package limma[20] to *orm*ard*l*ize the combined GEO dataset, annotate probes, and standardize them. The expression matrices before and after removing the batch effect were subjected to principal component analysis (PCA)[21] to verify the efficacy of removing the batch effect. Data were transformed into low-dimensional data and displayed as two- or three-dimensional graphs.

	GSE24206	GSE110147
Platform	GPL570	GPL6244
Species	Homo sapiens	Homo sapiens
Tissue	Lung	Lung
Samples in IPF group	17	22
Samples in Control group	6	11
Reference	PMID: 21,974,901 PMID: 30,506,760	PMID: 30,111,332 PMID: 30,506,760

Table 1. GEO microarray chip information. *GEO* gene expression omnibus, *IPF* idiopathic pulmonary fibrosis.

Differentially expressed genes (DEGs) related to mitochondrial energy metabolism in IPF

The samples were divided into IPF and control groups according to the sample grouping of the combined GEO datasets. Differential analysis of genes between the IPF and control samples was performed using the R package limma. To incorporate as many differentially expressed genes with potential biological significance as possible, and to comprehensively identify potential important biomarkers, | logFC | > 0.25 and adj. $p < 0.05$ were set as the thresholds for the DEGs with logFC < – 0.25 and adj. $p < 0.05$ were downregulated. The Benjamini-Hochberg (BH) test was used as the p-value correction method. A volcano plot was constructed using the R package ggplot2 to demonstrate the differential analysis results.

To obtain MEMRDEGs related to IPF, all DEGs were obtained through differential analysis in the integrated GEO dataset with | logFC | > 0.25 and adj. $p < 0.05$, and intersected with MEMRGs; a Venn plot was drawn to obtain MEMRDEGs through the R package pheatmap and using the R package RCircos[22] to map chromosomal location.

Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses

GO analysis[23] included biological processes (BP), cell*ular* components (CC), and molecular functions (MF). KEGG(www.kegg.jp/kegg/kegg1.html)[24] is a widely used database that stores information on genomes, biological pathways, diseases, and drugs. GO and KEGG enrichment analyses were performed on the MEMRDEGs using the R software package clusterProfiler[25]. A p-value < 0.05 and a false discovery rate value (q value) < 0.25 were considered statistically significant, and the p-value correction method was B*enjamini*-H*ochberg (BH)*.

Gene set enrichment analysis (GSEA)

Genes from the combined GEO datasets were first ranked by log *fold change (log*FC*)* values, and the R package clusterProfiler was used to perform GSEA[26] for all genes in the integrated GEO datasets. The parameters used in GSEA were as follows: The seed was 2020, the number of computations was 1000, the minimum number of genes contained in each gene set was 10, and the maximum number of genes contained in each gene set was 500. Using the Molecular Signatures Database[27] (https://www.gsea-msigdb.org/gsea/msigdb), the c2.cp.all.v2022.1.Hs.symbols.gmt [All Canonical Pathways](3050) was used to perform GSEA, and the screening criterion for GSEA was set at $p < 0.05$.

Protein-protein interaction (PPI) network and hub gene screening

In this study, we aimed to identify the hub genes related to mitochondrial energy metabolism. To this end, the PPI data related to MEMRDEGs were first obtained from the STRING database[28] (https://string-db.org/), and the minimum interaction coefficient was set to > 0.700 to ensure the reliability of the selected data. This threshold was based on existing literature and the results of data analysis, which could effectively filter out possible false positives, thus focusing on stronger interactions.

Subsequently, the obtained interaction data were imported into the Cytoscape software to generate the PPI network. In this network, the interactions between genes were represented in the form of nodes and edges, where nodes represented genes and edges reflected the interaction intensity between them. To evaluate the importance of each gene the PPI network, we utilized the CytoHubba[29] plug-in in Cytoscape[30], applying five algorithms including Maximum Clique Centrality (MCC), Degree, Maximum Neighborhood Composition (MNC), Edge Penetration Composition (EPC), and Closeness[31]. By calculating the MEMRDEG score, we selected the top ten MEMRDEGs with the highest scores. Finally, we integrated the genes using Venn diagrams based on different algorithms to identify the intersection genes, which were confirmed as hub genes involved in mitochondrial energy metabolism. This series of analyses also constructed the PPI network of MEMRDEGs and laid a solid foundation for subsequent biological research.

Construction of a regulatory network

Transcription factors (TFs) control gene expression during the post-transcriptional stage by interacting with hub genes. The ChIPBase database[32] (http://rna.sysu.edu.cn/chipbase/) was used to retrieve the TFs. Only TFs with the sum of "Number of samples found (upstream)" and "Number of samples found (downstream)" greater than 8 were retained. We analyzed the regulatory role of transcription factors on hub genes and visualized the regulatory network of mRNA-TF using Cytoscape software.

Additionally, miRNAs play important regulatory roles in biological development and evolution. To analyze the hub genes and the relationship between miRNAs, using the TarBase[33] database (https://rnasysu.com/encori) and the hub gene-associated microRNAs, only miRNAs with "pancancerNum > 8" were retained, and the regulatory network of mRNA-miRNA was visualized using Cytoscape software.

Differential expression verification and ROC curve analysis

To further explore the differences in the expression of hub genes between IPF and control samples in the combined GEO datasets, a group comparison map was drawn based on the expression levels of hub genes. The R package pROC was used to plot the ROC curves of the hub genes and calculate the area under the curve (AUC). Hub gene expression levels were evaluated to diagnose IPF. The AUC of the ROC curves was generally between 0.5 and 1. The closer the AUC was to 1, the better the diagnostic performance.

Immune infiltration analysis

Single-sample gene set enrichment analysis (ssGSEA)[34] was used to quantify the relative abundance of each immune cell infiltrate. First, each infiltrating immune cell type was labeled. Second, we used the enrichment fraction calculated using ssGSEA to represent the relative abundance of each immune cell infiltration in

each sample to obtain the immune cell infiltration matrix. Two groups of immune cells showing significant differences were selected for subsequent analyses. Correlations between immune cells were calculated using the Spearman algorithm. The R package pheatmap was used to draw a correlation heatmap to display the correlation analysis results for immune cells. The correlation between hub genes and immune cells was calculated using the Spearman algorithm. The R software package ggplot2 was used to draw a correlation bubble chart to display the correlation analysis results between hub genes and immune cells.

Statistical analysis

All data processing and analyses were performed using R software (version 4.3.0). To compare continuous variables between the two groups, the independent Student's *t*-test was used to analyze the differences between non-normally distributed variables. The Mann–Whitney U-test (Wilcoxon rank-sum test) was used to analyze the differences between non-normally distributed variables. The Kruskal–Wallis test was used to compare three or more groups. Spearman's correlation analysis was used to calculate the correlation coefficients between different molecules. Unless otherwise specified, all statistical p-values were two-tailed, and statistical significance was set at $p < 0.05$.

Results

The technical roadmap of this study

This study first obtained sample data from two datasets, where GSE24206 contains 17 IPF samples and 6 control samples, and GSE110147 contains 22 IPF samples and 11 control samples. We consolidated and normalized these data to obtain a combined dataset (39 IPF samples and 17 control samples). Subsequently, based on differential expression analysis, DEGs were identified and screened for the MEMRDEGs. Enrichment analyses of MEMRDEGs, including GO and KEGG analyses, were performed to understand gene functions and pathways. Next, a PPI network based on MEMRDEGs was constructed, and Hub genes were identified. Further, an ExpDiff analysis and ROC curve evaluation were conducted to examine the diagnostic efficacy of Hub genes, and the ssGSEA was applied to evaluate immune infiltration, and the relationship between Hub genes and immune cell infiltration was explored. Finally, we constructed a regulatory network of Hub genes that includes their interactions with transcription factors (mRNA-TF) and miRNAs (mRNA-miRNA) (Fig. 1). This network provides a systematic perspective on the relationship between mitochondrial energy metabolism and immune response.

Merging of IPF datasets

First, batch effects due to technical variation may occur between different experimental batches, which can mask real biological differences and reduce the reliability of the analysis results. We used the R package SVA to remove batch effects from the IPF datasets GSE24206 and G110147 to obtain a more reliable and combined GEO dataset.

Subsequently, we compared the expression values of the dataset before and after removing batch effects using a distribution box plot (Fig. 2a–b). *Additionally*, we compared the distribution of low-dimensional features before and after removing batch effects using PCA (Fig. 2c–d). The results confirmed that the batch effects in the IPF dataset samples were successfully eliminated after removal.

DEGs related to mitochondrial energy metabolism in IPF

First, we divided the data from the combined GEO dataset into IPF and control samples and analyzed the differences in gene expression values between these samples using the R software package limma. DEGs were identified between the two sets of data. A total of 4044 upregulated genes (logFC > 0.25 and adj. $p < 0.05$) and 4557 downregulated genes (logFC < − 0.25 and adj. $p < 0.05$) were identified, and a volcano map was drawn based on the difference analysis results of this dataset (Fig. 3a).

To obtain MEMRDEGs, the intersection of all DEGs and MEMRGs was obtained with | logFC | > 0.25 and adj. $p < 0.05$, and a Venn plot was constructed (Fig. 3b). We identified 25 MEMRDEGs. Specific information is provided in Table S2. Based on the cross results, the expression differences of the top 10 positive and negative logFC MEMRDEGs between different sample groups were analyzed. The R package pheatmap was used to draw a heatmap displaying the analysis results (Fig. 3c). Finally, the R package RCircos was used to analyze the locations of the 25 MEMRDEGs on human chromosomes and draw a chromosome localization map (Fig. 3d). Chromosomal mapping showed that most MEMR DEGs were located on chromosomes 1 and 11.

GO and KEGG enrichment analyses

The 25 MEMRDEGs were used for GO and KEGG enrichment analyses; the results are shown in Table 2. The results show that the 25 MEMRDEGs were mainly enriched in fatty acid beta-oxidation, fatty acid oxidation, fatty acid catabolic process, lipid oxidation and monocarboxylic acid catabolic process, as well as other BP; CC including mitochondrial matrix, microbody, peroxisome, mitochondrial intermembrane space, and organelle envelope lumen; and molecular functions such as acyl − CoA dehydrogenase activity, oxidoreductase activity, acting on the CH − CH group of donors, flavin adenine dinucleotide binding, electron transfer activity, and intramolecular oxidoreductase activity, transposing C = C bonds, along with other MF. Fatty acid degradation, fatty acid metabolism, carbon metabolism, valine, leucine, and isoleucine degradation, and peroxisome proliferator-activated receptor (PPAR) signaling pathways were also enriched in the KEGG pathway. The GO and KEGG enrichment analysis results are visualized using bar (Fig. 4a) and bubble (Fig. 4b) plots.

Network diagrams of BP, CC, MF, and KEGG were constructed (Fig. 4c–f). The results showed that more genes were enriched in the mitochondrial matrix of CC.

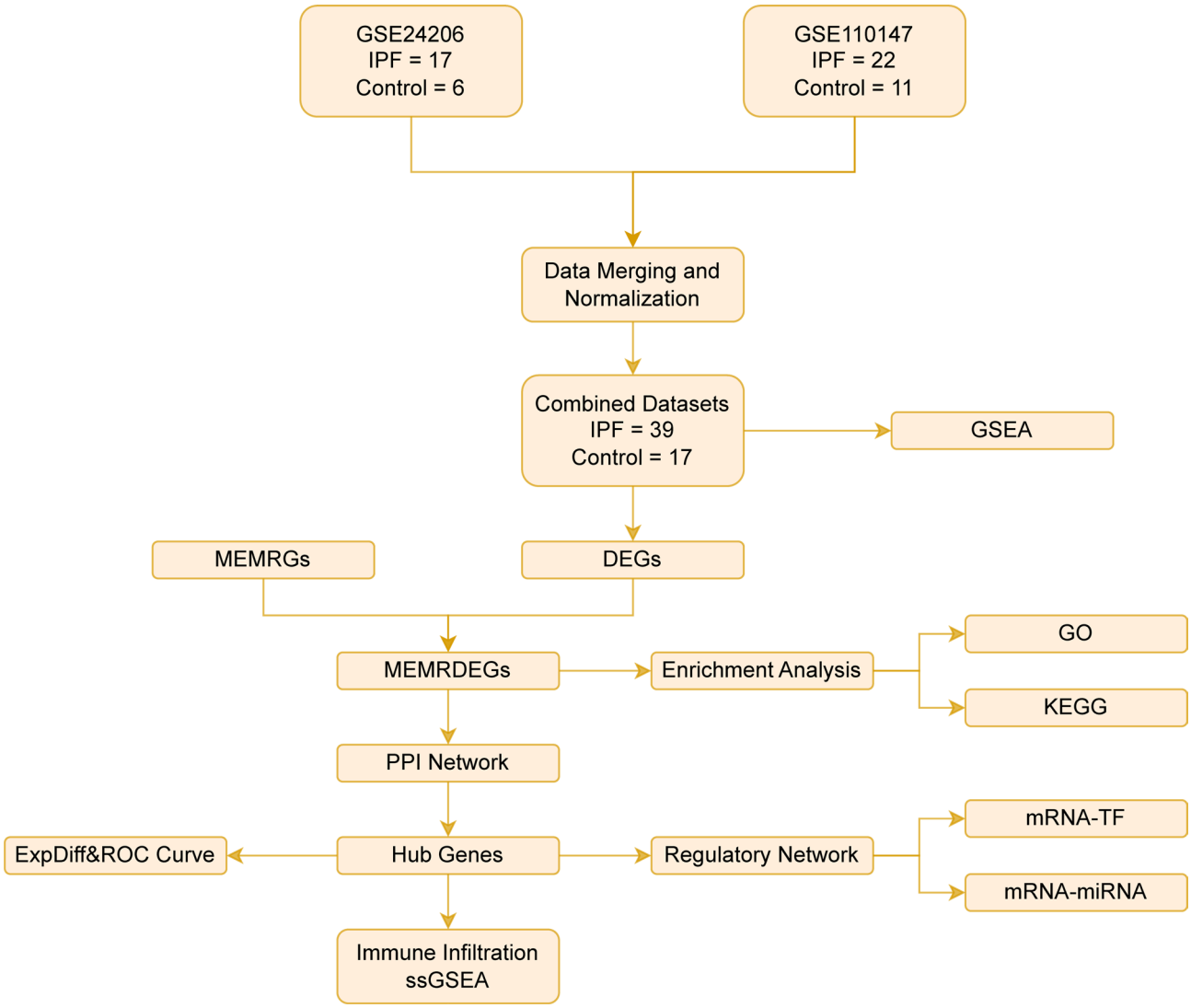

Fig. 1. Flowchart for the comprehensive analysis of MEMRDEGs. *IPF* idiopathic pulmonary fibrosis, *GSEA* gene set enrichment analysis, *DEGs* differentially expressed genes, *MEMRGs* mitochondrial energy metabolism related genes, *GO* gene ontology, *KEGG* kyoto encyclopedia of genes and genomes, *MEMRDEGs* mitochondrial energy metabolism related differentially expressed genes, *PPI* protein-protein interaction, *ROC* receiver operating characteristic, *TF* transcription factor, *ssGSEA* single-sample gene set enrichment analysis.

GSEA for IPF

GSEA was used to investigate the relationship between the expression of all genes in the integrated GEO datasets and the BP involved, the cellular components affected, and the MF played (Fig. 5a). The detailed results are presented in Table 3. The results showed that all genes in the combined GEO datasets were significantly ($p < 0.05$) enriched in the IL7 pathway (Fig. 5b), regulation of TP53 activity through phosphorylation (Fig. 5c), regulation of Wnt beta-catenin signaling by small molecule compounds (Fig. 5d), Tgf Beta receptor signaling pathway (Fig. 5e), Hedgehog Gli pathway (Fig. 5f), and other biologically related functions and signaling pathways.

Construction of PPI network and screening of hub genes

First, PPI analysis was performed, and a PPI network of 25 MEMRDEGs was constructed using the STRING database (Fig. 6a). The PPI network results showed that 15 MEMRDEGs were related: *COX5A*, *SDHB*, *CYCS*, *CS*, *DLAT*, *PPARGC1B*, *ACAT1*, *ACADS*, *ETFA*, *EHHADH*, *ACADM*, *ACADL*, *ECI2*, *EIF4EBP1*, and *RPS6KB1*. Subsequently, the five algorithms of the CytoHubba plug in of Cytoscape software were used to calculate the scores of the 15 MEMRDEGs. Based on the scores, the five algorithms were MCC, degree, MNC, EPC, and closeness. The top 10 MEMRDEGs from these five algorithms were used to construct the PPI network as follows: MCC (Fig. 6b), degree (Fig. 6c), MNC (Fig. 6d), EPC (Fig. 6e), and closeness (Fig. 6f). The colors of the circles, from red to yellow, represent ratings from high to low. Finally, the intersection of the genes from the five algorithms was obtained, and a Venn diagram (Fig. 6g) was drawn for analysis. The genes that intersected

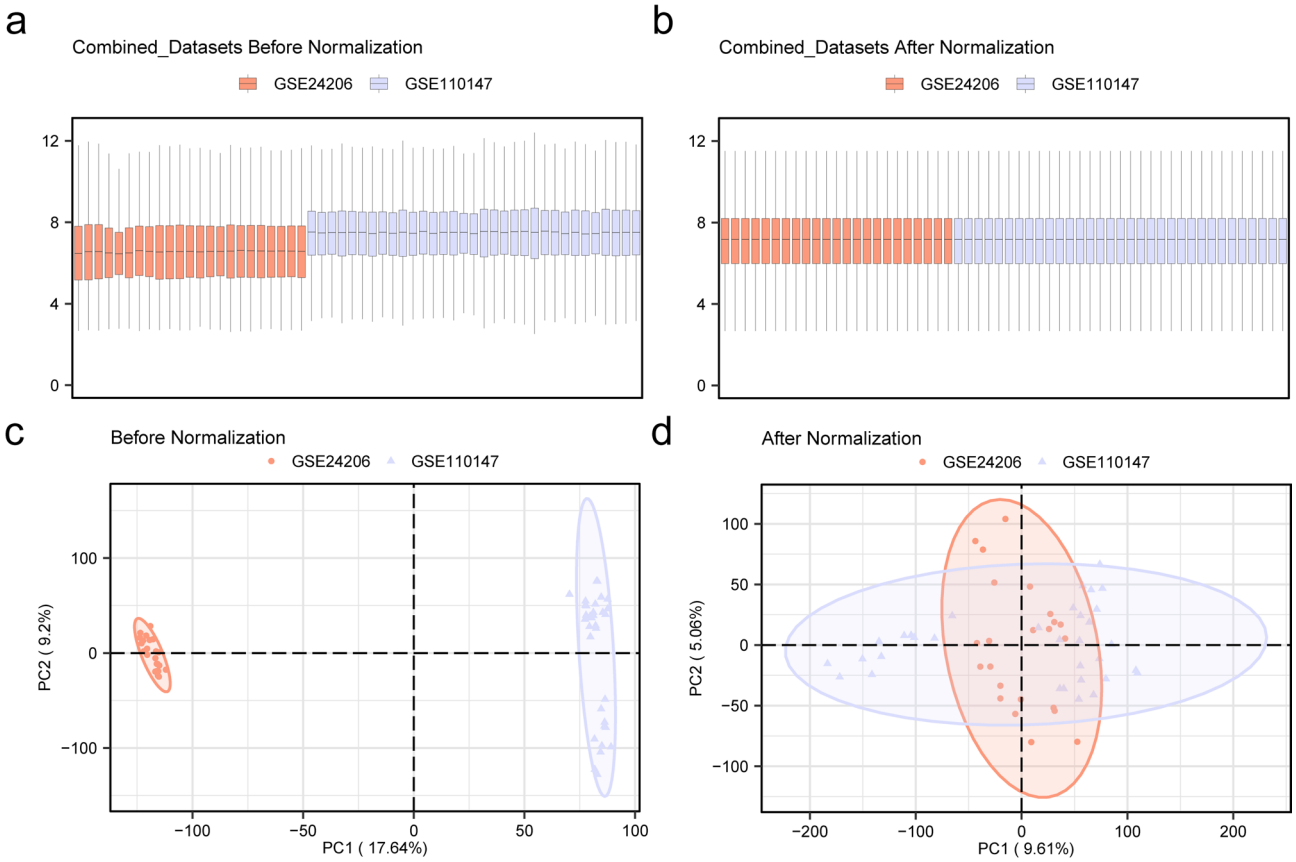

Fig. 2. Batch effects removal of GSE24206 and GSE110147. (**a**) Box plots of GEO combined with the dataset distribution before removing batch effects. (**b**) Postbatch integrated GEO dataset distribution boxplots. (**c**) 2D PCA plot of integrated GEO datasets before being debatched. (**d**) 2D PCA plot of the integrated GEO datasets after debatching. IPF datasets GSE24206 and GSE110147 are shown in orange and blue, respectively. *GEO* gene expression omnibus, *PCA* principal component analysis, *IPF* idiopathic pulmonary fibrosis.

between the algorithms were hub genes related to mitochondrial energy metabolism; the nine hub genes were *EHHADH*, *ACAT1*, *ECI2*, *ACADS*, *CS*, *SDHB*, *COX5A*, *ETFA*, and *ACADL*.

Construction of regulatory networks

First, we obtained TFs that bind to hub genes from the ChIPBase database and constructed and visualized an mRNA-TF regulatory network using the Cytoscape software (Fig. 7a). Among them, there were 7 hub genes and 38 TFs ; specific information is shown in Table S3.

miRNAs related to hub genes were obtained from the TarBase database, and an mRNA-miRNA regulatory network was constructed and visualized using Cytoscape software (Fig. 7b). 7 hub genes and 31 miRNAs were identified, and specific information is shown in Table S4.

Differential expression verification and ROC curve analysis

Group comparison (Fig. 8a) showed the differential expression of the nine hub genes in the IPF and control samples. The results showed that the expression levels of the 9 hub genes were significantly different ($p < 0.01$). The expression levels of 3 hub genes (*ACADL*, *ACAT1*, and *CS*) in the IPF and control samples were significantly different ($p < 0.01$), whereas the expression levels of 6 hub genes (*ACADL*, *COX5A*, *ECI2*, *EHHADH*, *ETFA*, and *SDHB*) were also significantly different ($p < 0.01$). Next, we used the R package pROC to draw ROC curves based on the expression levels of hub genes (Fig. 8b–j). The ROC curve showed that the expression levels of the 3 hub genes (*COX5A*, *EHHADH*, and *SDHB*) in the IPF samples were highly accurate among the different groups (AUC > 0.9). The expression levels of 6 hub genes (*ACADL*, *ACADS*, *ACAT1*, *CS*, *ECI2*, and *ETFA*) in the IPF samples showed a certain accuracy in the different groups (0.7 < AUC < 0.9).

Immune infiltration analysis of IPF

The expression matrix of the combined dataset was applied to the ssGSEA algorithm to calculate the immune infiltration abundance of the 28 immune cells. First, immune cells with p-value < 0.05 were screened using a group comparison plot, and the expression differences of immune cell infiltration abundance in different groups are shown. The group comparison diagram (Fig. 9a) shows that four types of immune cells–activated CD8 T cells, activated dendritic cells, central memory CD4 + T cells, and macrophages–were significantly different between

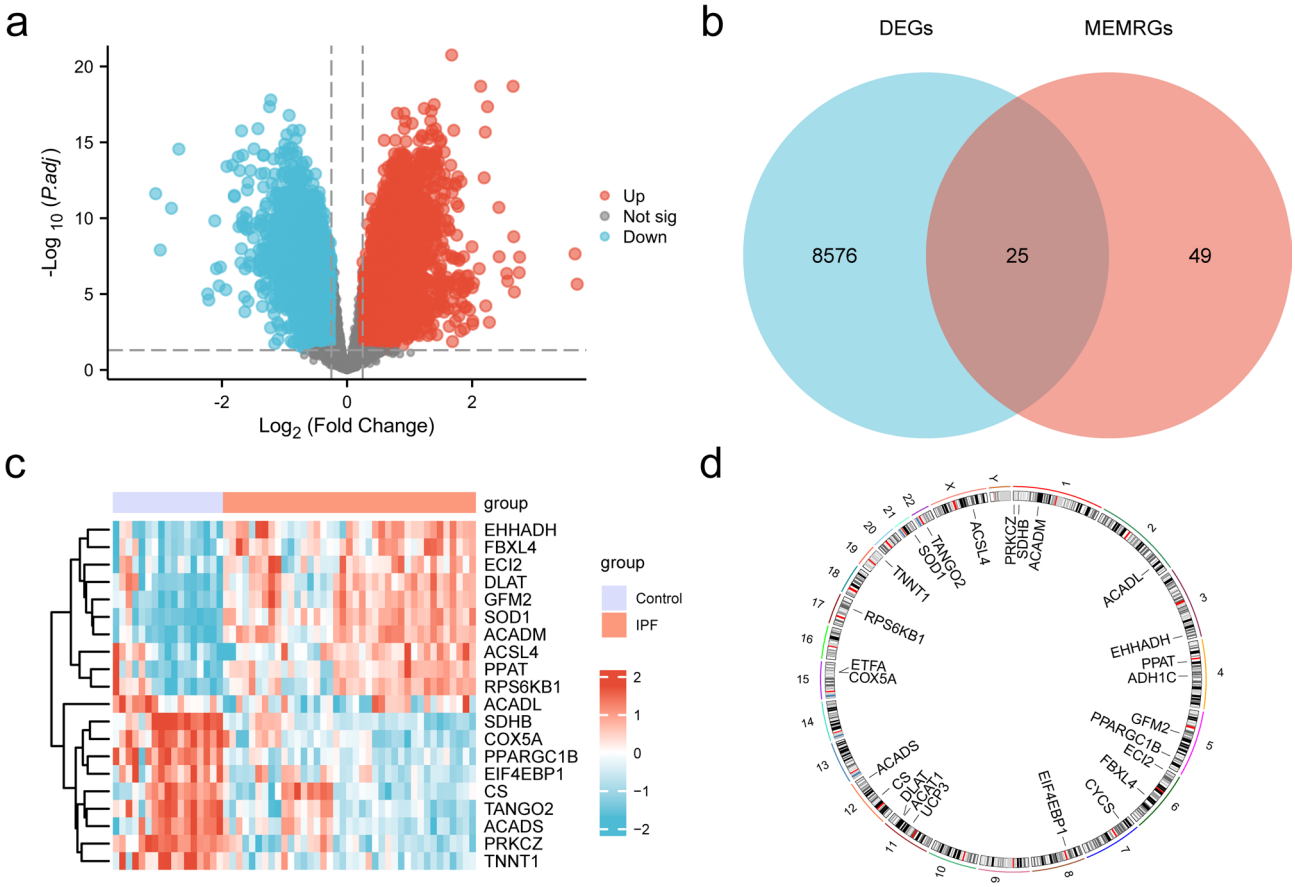

Fig. 3. Differential gene expression analysis.(**a**) Volcano plot of differentially expressed genes analysis in IPF and control samples in the combined GEO datasets. (**b**) DEG and MEMRG Venn diagrams of the integrated GEO datasets. (**c**) Heatmap of the top 10 positive and negative logFC MEMRDEGs in the integrated GEO datasets. (**d**) Chromosomal mapping of MEMRDEGs. In the heatmap group, orange represents the IPF sample, and blue represents the control sample. In the heatmap, red represents high expression, and blue represents low expression. *IPF* idiopathic pulmonary fibrosis, *DEGs* differentially expressed genes, *MEMRGs* mitochondrial energy metabolism related genes, *MEMRDEGs* mitochondrial energy metabolism related differentially expressed genes.

the IPF samples and controls ($p < 0.05$). Four types of immune cells–central memory CD8 T cells, gamma delta T cells, immature dendritic cells, and T follicular helper cells–examined significantly differed between the IPF and control samples ($p < 0.01$). Twelve immune cell types–CD56dim natural killer cells, effector memory CD4 T cells, eosinophils, mast cells, myeloid-derived suppressor cells, monocytes, natural killer cells, neutrophils, regulatory T cells, type 1 T helper cells, type 17 T helper cells, and type 2 T helper cells–were significantly different between IPF and control samples ($p < 0.001$). A correlation heatmap was used to show the correlation results of the abundance of the 20 immune cell infiltrations in the immune cell infiltration analysis (Fig. 9b). The results showed that most immune cells were positively correlated. The correlation between nine hub genes and 20 immune cells was analyzed and displayed using a correlation bubble plot (Fig. 9c), and the results *indicat*ed that the strongest positive correlation was found between hub gene *ACADS* and immune cells CD56dim natural killer cells ($r = 0.86$, $p < 0.05$). There was a strong negative correlation between the hub gene *ACAT1* and immune CD56dim natural killer cells ($r = -0.77$, $p < 0.05$).

Discussion

IPF is a chronic, progressive, and fatal interstitial lung disease characterized by scarring of the lung tissue, which leads to a relentless decline in respiratory function[35]. The etiology of IPF remains elusive and poses a significant threat to public health, with an estimated median survival of 3–5 years after diagnosis[36]. Patients with IPF often experience debilitating symptoms, such as dyspnea and cough, which severely affect their quality of life and daily activities[37]. The complexity and heterogeneity of this disease underscore the urgent need for comprehensive research on the underlying molecular mechanisms to improve diagnostic methods and therapeutic strategies.

Recent studies have highlighted the potential role of mitochondrial energy metabolism in IPF pathogenesis, suggesting that alterations in this fundamental cellular process contribute to fibroblast activation and aberrant tissue repair[38].

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	*p*.adjust	qvalue
BP	GO:0006635	fatty acid beta-oxidation	7/25	76/18,800	6.03 e-12	4.79 e-09	3.17 e-09
BP	GO:0009062	fatty acid catabolic process	7/25	102/18,800	4.98 e-11	1.73 e-08	1.15 e-08
BP	GO:0019395	fatty acid oxidation	7/25	106/18,800	6.55 e-11	1.73 e-08	1.15 e-08
BP	GO:0034440	lipid oxidation	7/25	112/18,800	9.69 e-11	1.92 e-08	1.28 e-08
BP	GO:0072329	monocarboxylic acid catabolic process	7/25	124/18,800	1.99 e-10	3.17 e-08	2.10 e-08
CC	GO:0005759	mitochondrial matrix	9/25	473/19,594	3.74 e-09	2.58 e-07	1.61 e-07
CC	GO:0005777	peroxisome	4/25	141/19,594	2.89 e-05	6.65 e-04	4.16 e-04
CC	GO:0042579	microbody	4/25	141/19,594	2.89 e-05	6.65 e-04	4.16 e-04
CC	GO:0005758	mitochondrial intermembrane space	3/25	83/19,594	1.58 e-04	2.72 e-03	1.70 e-03
CC	GO:0031970	organelle envelope lumen	3/25	93/19,594	2.21 e-04	3.05 e-03	1.91 e-03
MF	GO:0003995	acyl-CoA dehydrogenase activity	3/24	12/18,410	4.25 e-07	4.04 e-05	1.83 e-05
MF	GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	4/24	59/18,410	9.64 e-07	4.58 e-05	2.08 e-05
MF	GO:0050660	flavin adenine dinucleotide binding	4/24	85/18,410	4.19 e-06	1.33 e-04	6.03 e-05
MF	GO:0009055	electron transfer activity	4/24	125/18,410	1.94 e-05	4.60 e-04	2.09 e-04
MF	GO:0016863	intramolecular oxidoreductase activity, transposing C = C bonds	2/24	14/18,410	1.47 e-04	2.79 e-03	1.27 e-03
KEGG	hsa00071	Fatty acid degradation	8/19	43/8164	2.15 e-14	2.22 e-12	1.41 e-12
KEGG	hsa01212	Fatty acid metabolism	6/19	57/8164	2.24 e-09	1.15 e-07	7.30 e-08
KEGG	hsa00280	Valine, leucine and isoleucine degradation	4/19	48/8164	3.82 e-06	1.30 e-04	8.24 e-05
KEGG	hsa01200	Carbon metabolism	5/19	115/8164	5.05 e-06	1.30 e-04	8.24 e-05
KEGG	hsa03320	PPAR signaling pathway	4/19	75/8164	2.29 e-05	4.73 e-04	2.99 e-04

Table 2. Result of GO and KEGG enrichment analysis for MEMRDEGs. *GO* gene ontology, *BP* biological process, *CC* cellular component, *MF* molecular function, *KEGG* kyoto encyclopedia of genes and genomes, *MEMRDEGss* Mitochondrial energy metabolism related differentially expressed genes.

This study aimed to provide novel insights into the molecular landscape of IPF by focusing on MEMRDEGs. This study first obtained sample data from two datasets, of which GSE24206 contains 17 IPF samples and 6 control samples, and GSE110147 contains 22 IPF samples and 11 control samples. We merged and normalized these data to obtain a combined dataset (39 IPF samples and 17 control samples). Subsequently, based on differential expression analysis, DEGs were identified and MEMRDEGs were screened. The datasets we selected (GSE24206 and GSE110147) have important biological and technical background. First, both datasets are derived from human samples, which makes the results of our analyses biologically comparable. In addition, these datasets focus on lung tissue, and analyzing gene expression in lung tissue can more directly reveal the molecular mechanisms of IPF.

This approach holds promise for identifying potential biomarkers for early diagnosis and targets for therapeutic interventions that could significantly enhance patient outcomes and alleviate the burden of this devastating disease[39].

Identification of 9 hub genes (*EHHADH*, *ACAT1*, *ECI2*, *ACADS*, *CS*, *SDHB*, *COX5A*, *ETFA*, and *ACADL*) underscores their central roles in mitochondrial energy metabolism. These hub genes exhibited high connectivity within the constructed PPI network and were consistently rated significant across multiple algorithmic evaluations. Their importance was further supported by ROC curve analyses, indicating high diagnostic accuracy for distinguishing between IPF samples and controls, particularly for *COX5A*, *EHHADH*, and *SDHB*, with AUC values exceeding 0.9 or demonstrating moderate accuracy ($0.7 < AUC < 0.9$).

COX5A plays an essential roles in electron transport and ATP synthesis[40]. The expression of *COX5A* can increase cellular glucose uptake, enhance the activity of mitochondrial cytochrome c oxidase (COX), increase ATP, and improve mitochondrial energy metabolism[41].Its high diagnostic accuracy ($AUC > 0.9$) highlights its potential as an IPF biomarker and emphasizes the importance of mitochondrial dysfunction in the context of this disease.Targeting *COX5A* and its metabolic pathways may lead to new treatments for IPF.

EHHADH is a type of enzyme involved in fatty acid peroxisomes β dual functional enzymes involved in oxidative pathways[42]. It catalyzes the second and third steps of peroxisomal fatty acid oxidation[43].Its role in lipid metabolism suggests a potential connection with IPF, as disruptions in lipid homeostasis have been implicated in the pathogenesis of this disease[44]. The identification of *EHHADH* as a hub gene with high diagnostic accuracy ($AUC > 0.9$) underscores its significance in IPF and raises the possibility that it could be a biomarker for disease detection or progression.Modulation of *EHHADH* may help to improve the metabolic imbalance associated with IPF.

SDHB is part of the succinate dehydrogenase complex that links the Krebs cycle and the electron transport chain. Decreased *SDHB* enhances hypoxia-inducible factor (HIF)-1α activity, induces mitochondrial injury, bioenergetic changes, and cytokine release[45].Mutations affecting *SDHB* can lead to dysregulated cellular energetics and increased oxidative stress[46], both of which have been implicated in IPF pathogenesis. The high diagnostic accuracy ($AUC > 0.9$) of *SDHB* further supports its relevance in IPF.Modulation of *SDHB* activity and its interconnected metabolic networks could offer potential therapeutic avenues in managing IPF.

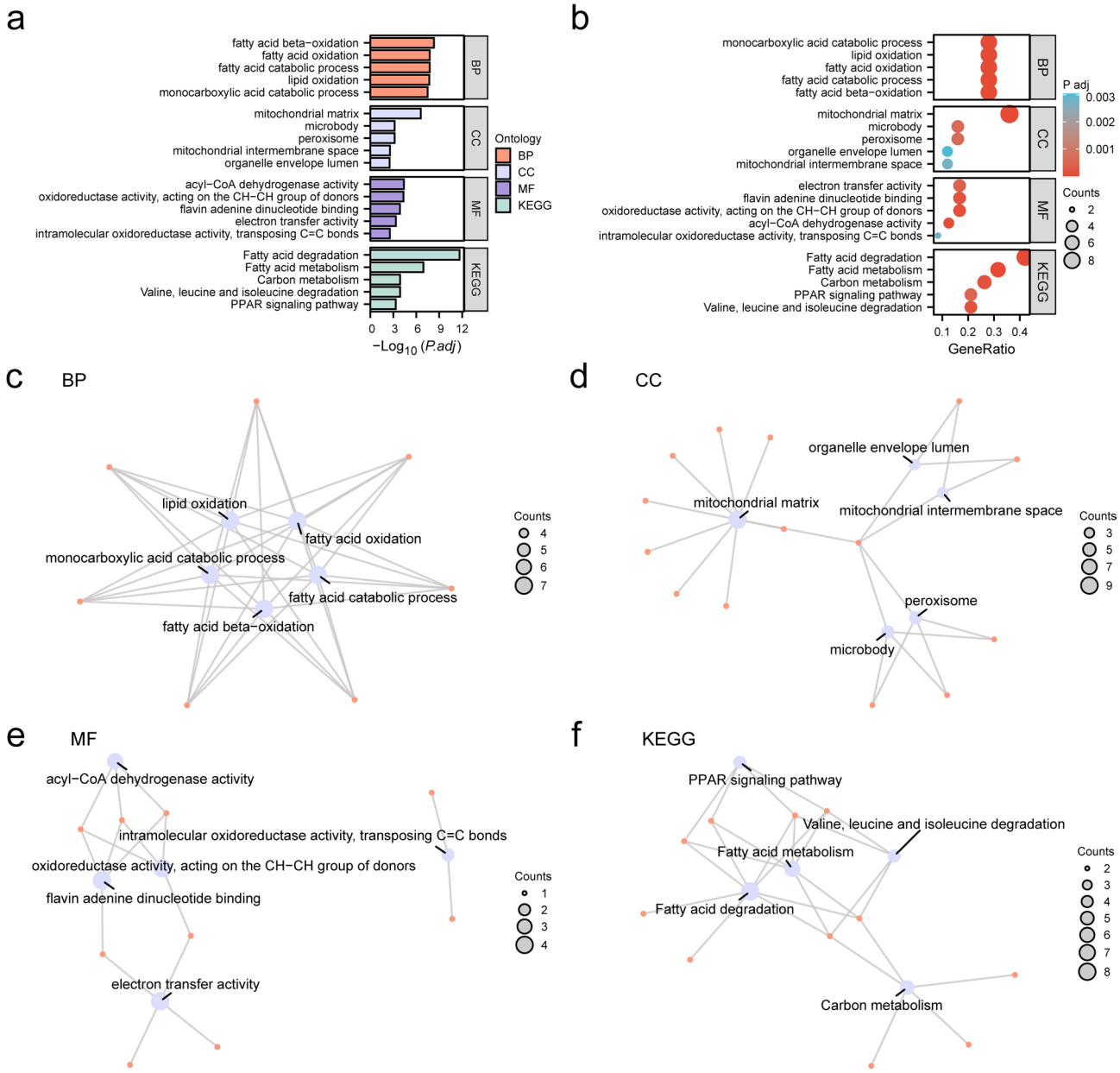

Fig. 4. GO and KEGG enrichment analysis for MEMRDEGs. (**a**, **b**). GO and KEGG enrichment analysis results for MEMRDEGs. The bar graph (**a**) and bubble plot (**b**) show BP, CC, MF, and KEGG. GO and KEGG terms are shown on the ordinate. (**c**–**f**). GO and KEGG enrichment analysis results for MEMRDEGs: BP (**c**), CC (**d**), MF (**e**), and KEGG (**f**). Blue nodes represent items, orange nodes represent molecules, and lines represent the relationships between items and molecules. The screening criteria for GO and KEGG enrichment analyses were adjusted to $p < 0.05$, FDR value (q value) < 0.25, and BH as the p-value correction method. The use of KEGG software has been licensed by the Kanehisa laboratory.*MEMRDEGs* mitochondrial energy metabolism related differentially expressed genes, *GO* gene ontology, *KEGG* kyoto encyclopedia of genes and genomes, *BP* biological process, *CC* cellular component, *MF* molecular function.

ACAT1 is critical for ketone body synthesis and cholesterol metabolism[47]. Since metabolic reprogramming is a hallmark of fibrotic processes, *ACAT1*'s differential expression in IPF may reflect alterations in cellular energy utilization that contribute to disease pathology[48]. The association between *ACAT1* expression and IPF indicates its potential utility as a diagnostic panel or therapeutic target. *ECI2* functions in the mitochondrial fatty acid beta-oxidation pathway. Alterations in mitochondrial function are increasingly recognized as central to the development of fibrotic diseases such as IPF[49]. As one of the identified hub genes, *ECI2's* involvement in mitochondrial energy metabolism is integral to the pathophysiology of IPF. *ACADS* catalyzes the initial steps in mitochondrial fatty acid oxidation. Mutations and dysregulation of *ACADS* have been linked to metabolic disorders[50]; however, their roles in lung diseases remain underexplored. The identification of *ACADS* among the hub genes with moderate diagnostic accuracy (0.7 < AUC < 0.9) suggests that *ACADS* may contribute to the

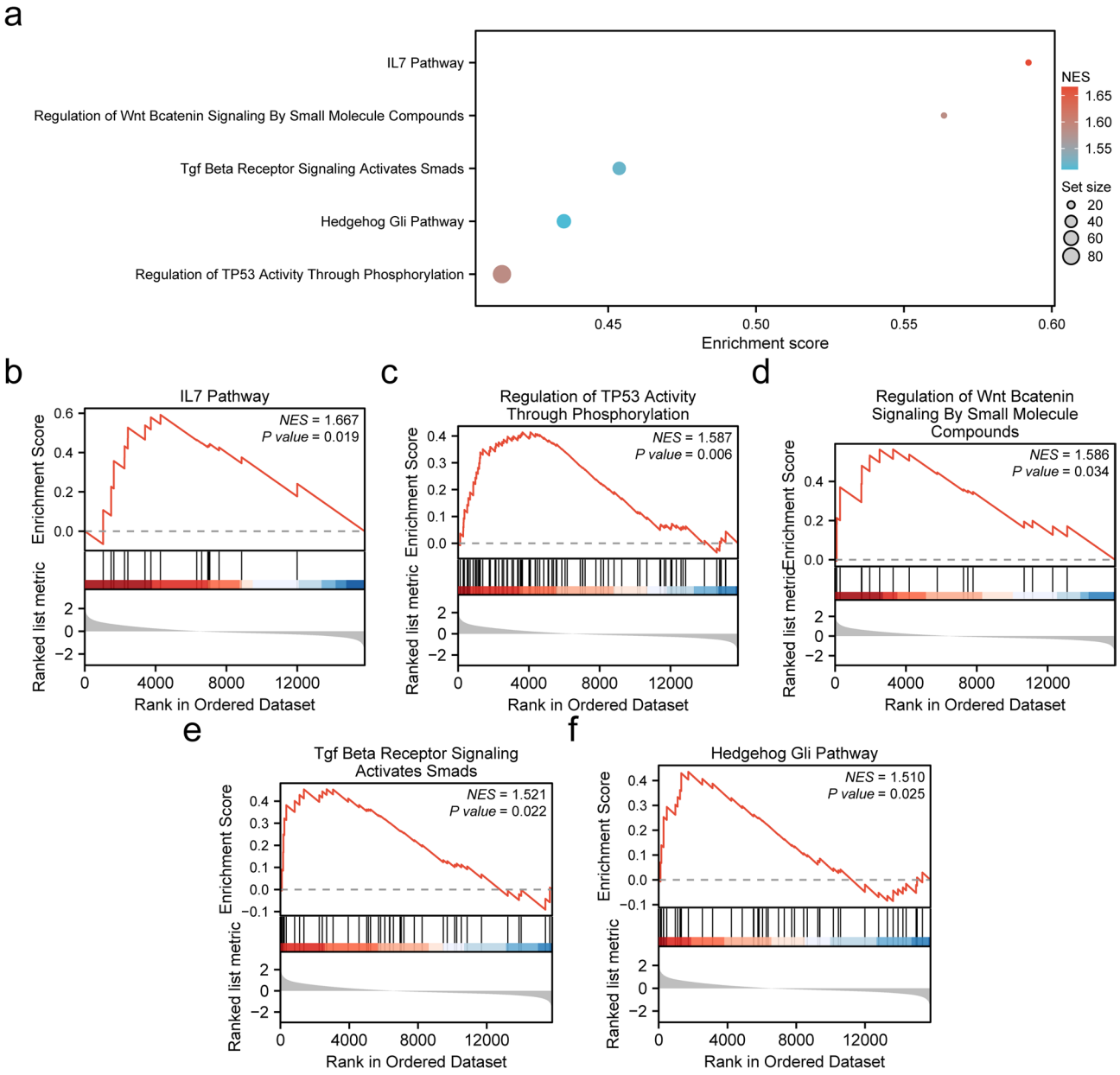

Fig. 5. GSEA for combined datasets. (**a**). GSEA biological function bubble plot of integrated GEO datasets. (**b**–**f**). GSEA showed that all genes were significantly enriched in the IL7 pathway (**b**), Regulation of TP53 activity through phosphorylation (**c**), regulation of Wnt beta-catenin signaling by small molecule compounds (**d**), Tgf Beta receptor signaling molecules (**e**), and the Hedgehog Gli pathway (**f**). The bubble size represents the number of enriched genes, and the color of the bubble represents the size of the NES value. The redder the color, the higher the value, while the bluer the color indicates a lower value. The screening criterion for GSEA was set at $p < 0.05$. *GSEA* gene set enrichment analysis.

metabolic disturbances observed in IPF. *CS* is a key enzyme in the Krebs cycle and is essential for aerobic energy production within the mitochondria[51]. Its altered expression may indicate shifts toward glycolytic pathways, a phenomenon known as the Warburg effect, which has been observed in various fibrotic diseases, including IPF[52].

The *ETFA* subunit is involved in electron transfer from fatty acid oxidation to the mitochondrial respiratory chain[53]. Disruption of *ETFA* can impair ATP production and exacerbate oxidative stress, which are key features associated with lung fibrosis.

Finally, *ACADL* is involved in fatty acid beta-oxidation within the mitochondria. Variants affecting *ACADL* function have been associated with metabolic syndromes[54], suggesting that its dysregulation may similarly affect energy metabolism pathways relevant to IPF pathology.

These 9 hub genes may influence the initiation and progression of IPF through processes including oxidative phosphorylation, fatty acid oxidation, and metabolic reprogramming. Therapeutic agents or small-molecule

ID	Set size	Enrichment score	NES	*p* value
WP_GENES_RELATED_TO_PRIMARY_CILIUM_DEVELOPMENT_BASED_ON_CRISPR	86	6.91 e-01	2.66 e + 00	1.38 e-03
WP_CILIOPATHIES	152	6.12 e-01	2.53 e + 00	1.26 e-03
REACTOME_CILIUM_ASSEMBLY	173	5.79 e-01	2.45 e + 00	1.23 e-03
REACTOME_COLLAGEN_DEGRADATION	63	6.53 e-01	2.38 e + 00	1.45 e-03
REACTOME_MITOTIC_PROMETAPHASE	176	5.48 e-01	2.31 e + 00	1.23 e-03
PID_SYNDECAN_1_PATHWAY	46	6.59 e-01	2.28 e + 00	1.49 e-03
REACTOME_MITOTIC_SPINDLE_CHECKPOINT	100	5.73 e-01	2.26 e + 00	1.34 e-03
WP_BARDETBIEDL_SYNDROME	75	5.94 e-01	2.25 e + 00	1.40 e-03
REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES	61	6.16 e-01	2.24 e + 00	1.45 e-03
WP_DNA_IRDAMAGE_AND_CELLULAR_RESPONSE_VIA_ATR	72	5.88 e-01	2.21 e + 00	1.41 e-03
REACTOME_RESOLUTION_OF_SISTER_CHROMATID_COHESION	109	5.56 e-01	2.21 e + 00	1.34 e-03
NABA_COLLAGENS	44	6.36 e-01	2.18 e + 00	1.52 e-03
REACTOME_COLLAGEN_CHAIN_TRIMERIZATION	44	6.36 e-01	2.18 e + 00	1.52 e-03
REACTOME_SYNDECAN_INTERACTIONS	24	7.16 e-01	2.18 e + 00	1.62 e-03
REACTOME_CELL_CYCLE_CHECKPOINTS	230	4.96 e-01	2.16 e + 00	1.18 e-03
BIOCARTA_IL7_PATHWAY	16	5.92 e-01	1.67 e + 00	1.94 e-02
REACTOME_REGULATION_OF_TP53_ACTIVITY_THROUGH_PHOSPHORYLATION	84	4.14 e-01	1.59 e + 00	5.53 e-03
WP_REGULATION_OF_WNT_BCATENIN_SIGNALING_BY_SMALL_MOLECULE_COMPOUNDS	16	5.64 e-01	1.59 e + 00	3.40 e-02
REACTOME_TGF_BETA_RECEPTOR_SIGNALING_ACTIVATES_SMADS	40	4.54 e-01	1.52 e + 00	2.19 e-02
PID_HEDGEHOG_GLI_PATHWAY	47	4.35 e-01	1.51 e + 00	2.53 e-02

Table 3. Results of GSEA for combined datasets.

modulators targeting these genes and their associated signaling pathways could potentially provide novel strategies for *the treatment of* IPF.

Immune infiltration analysis using ssGSEA revealed significant differences in the abundance of certain immune cell types between the IPF and control groups, indicating a potential role of the immune system in the pathogenesis of IPF. Macrophages, T cells, and B cells showed altered levels of infiltration. Macrophages contribute to tissue remodeling and fibrosis by secreting pro-fibrotic cytokines such as transforming growth factor-beta[55]. The dysregulation of T cells has been implicated in chronic inflammatory responses that may lead to fibrotic processes[56]. Similarly, B cells influence fibrosis by producing antibodies and modulating T cell function[57].

Moreover, the observed correlation between specific hub genes identified in our MEMRDEG analysis and particular immune cell subtypes underscores their potential involvement in the modulation of immune responses in the IPF lung environment. Mitochondria are not only the main energy supply center of cells, but also play a key role in the regulation of oxidative stress, cell death, immune cell function, and their abnormalities may affect the immune response of IPF through various mechanisms.For instance, *COX5A* is involved in mitochondrial respiratory chain function and can affect cellular metabolism within immune cells[58], whereas *EHHADH* plays a role in peroxisomal fatty acid beta-oxidation, which may affect lipid signaling pathways relevant to inflammation. The association between *SDHB* and mitochondrial complex II suggests a link between energy metabolism and redox balance, which can influence immune cell activity. *COX5A*, *EHHADH*, and *SDHB* have promising research prospects in immune-related fields. Ali et al.[59,60] developed detection methods for Gene Ontology and critical metabolic pathways, which strengthened the discussion on hub gene screening and the KEGG pathway analysis. This advancement will facilitate our understanding of these genes' roles in IPF.

Integrative analysis of GEO datasets identified a subset of 25 MEMRDEGs that were significantly associated with IPF. These MEMRDEGs are predominantly located on chromosomes 1 and 11, suggesting a chromosomal clustering phenomenon related to the pathophysiology of IPF. Genes such as *PRKCZ*, *SDHB*, and *ACADM* on chromosome 1 and *DLAT*, *ACAT1*, and *UCP3* on chromosome 11 have been implicated in mitochondrial dysfunction and energy metabolism[61].

Enrichment analyses via GO and KEGG pathways revealed that these MEMRDEGs were closely linked to BP, such as fatty acid metabolism, oxidative processes, and the PPAR signaling pathway. Fatty acid metabolism is crucial for maintaining cellular energy homeostasis and is reportedly dysregulated in IPF[62]. Oxidative stress is a well-known contributor to the pathogenesis of IPF and induces fibroblast proliferation and extracellular matrix deposition[63]. The involvement of the PPAR pathway suggests a potential regulatory mechanism in lipid homeostasis and inflammation in lung tissue[64].

- These findings highlight the importance of mitochondrial energy metabolism in cellular bioenergetics and its broader implications for immune regulation in IPF. Given that IPF is a progressive disease, the identified MEMRDEGs may be closely related to the severity and progression of the disease. Specifically, the regulation of some MEMRDEGs in energy metabolism, oxidative stress, and apoptosis may affect the degree of fibrosis, the rate of lung function decline and clinical outcomes[62,63].

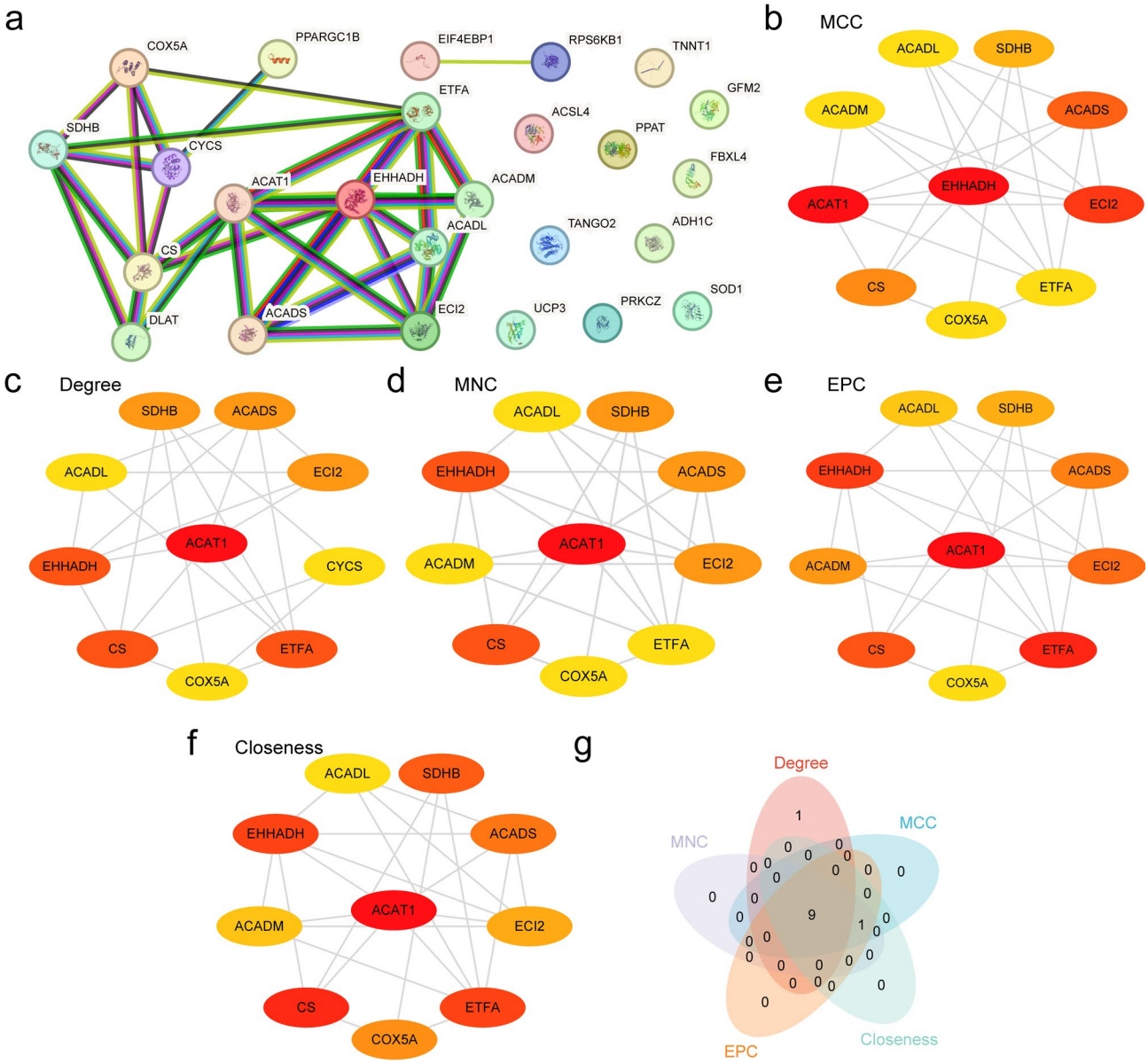

Fig. 6. PPI network and hub gene analysis. a. PPI network of MEMRDEGs calculated from the STRING database. (**b**–**f**). The PPI network of the top 10 MEMRDEGs calculated by five algorithms of the CytoHubba plug-in, including MCC (**b**), Degree (**c**), MNC(**d**), EPC (**e**), and closeness (**f**). g. Venn diagram of the top 10 MEMRDEGs for the five algorithms of the CytoHubba plug-in. *PPI network* protein-protein interaction network, *MEMRDEGs* mitochondrial energy metabolism related differentially expressed genes, *MCC* maximal clique centrality, *MNC* maximum neighborhood component, *EPC* edge-percolated component.

The significant correlations between hub genes related to mitochondrial function and specific immune cell populations suggest that targeting these metabolic pathways could modulate the inflammatory milieu and potentially ameliorate disease progression. Furthermore, given their diagnostic potential, as reflected by ROC curve analyses (AUC > 0.9; *COX5A*, *EHHADH*, and *SDHB*), these hub genes may serve as biomarkers for identifying patients with aberrant immune profiles who might benefit from therapies aimed at correcting metabolic dysfunctions.

Our findings highlight the significance of mitochondrial energy metabolism in IPF pathogenesis MEMRDEGs serve as biomarkers, and their expression levels may be related to the clinical manifestations and prognosis of patients, which may be beneficial for the early diagnosis and disease monitoring of IPF.These MEMRDEGs have the potential to serve as new therapeutic targets to improve patients' disease progression through the development of targeted drugs or other interventions. Developing individualized treatment plans based on the expression profiles of different MEMRDEGs in patients may improve the effectiveness of treatment and provide more precise care for IPF patients.

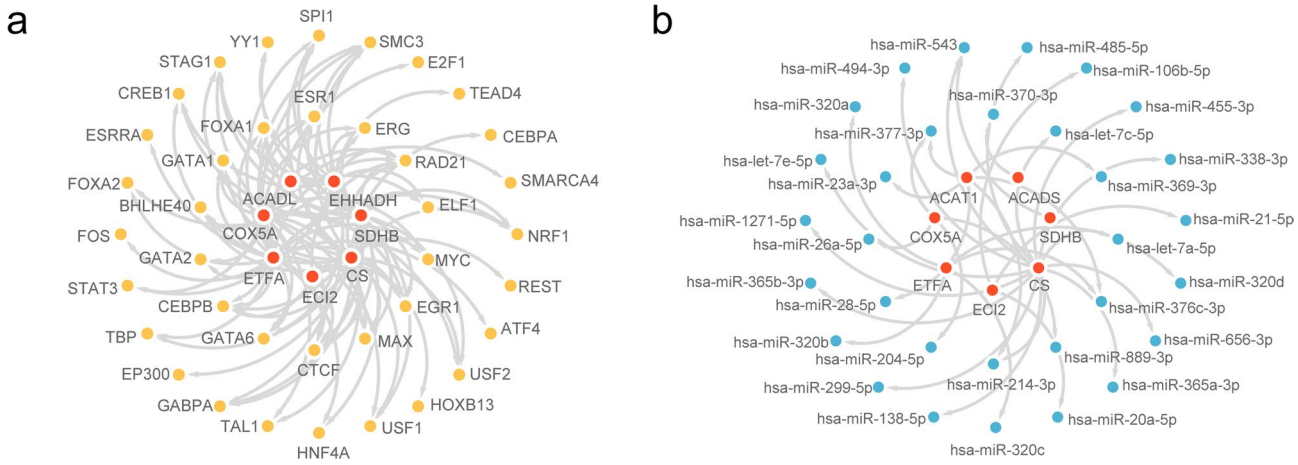

Fig. 7. Regulatory network of hub genes. (**a**) The mRNA-TF regulatory network of hub genes. (**b**) mRNA-miRNA regulatory network of hub genes. *TF* transcription factor. mRNAs are shown in red, TFs in yellow, and miRNAs in blue.

Although our study demonstrated good predictive performance on an independent validation set, with results of high reliability and stability, it is important to acknowledge its limitations. First, we recognize that further experimental validation, such as Western blotting or immunofluorescence staining, would be more comprehensive to confirm the expression of these genes at the protein level. We plan to consider these additional experiments in future research to enhance the credibility and comprehensiveness of the study. Regarding the specific roles of *COX5A*, *EHHADH*, and *SDHB* in the pathogenesis of IPF, we also plan to conduct more in-depth mechanistic studies and data analysis to explore how these genes promote the occurrence and development of IPF by affecting mitochondrial energy metabolism and the immune microenvironment. Second, due to the small clinical sample size, it may affect the universality of the results. We plan to introduce a larger clinical sample for verification in our research. At the same time, the dataset used will be expanded and more in-depth comparison and statistical analysis will be carried out. Third, we understand the importance of adopting innovative bioinformatics technologies, and we plan to introduce new bioinformatics methods, such as machine learning algorithms and complex network analysis in our future research to enhance the reliability and innovation of the research results.

Conclusion

Our study utilized the IPF dataset from the GEO database to identify key differentially expressed genes (MEMRDEGs) related to mitochondrial energy through bioinformatics methods. The results showed that these genes were closely related to the pathological process of IPF and might affect the immune microenvironment. Future research will focus on exploring the specific mechanisms of these genes, especially their roles in the pathogenesis of IPF. In addition, we plan to conduct experimental verification to support these bioinformatics findings and expand the scale of clinical samples to improve the general applicability of the study.

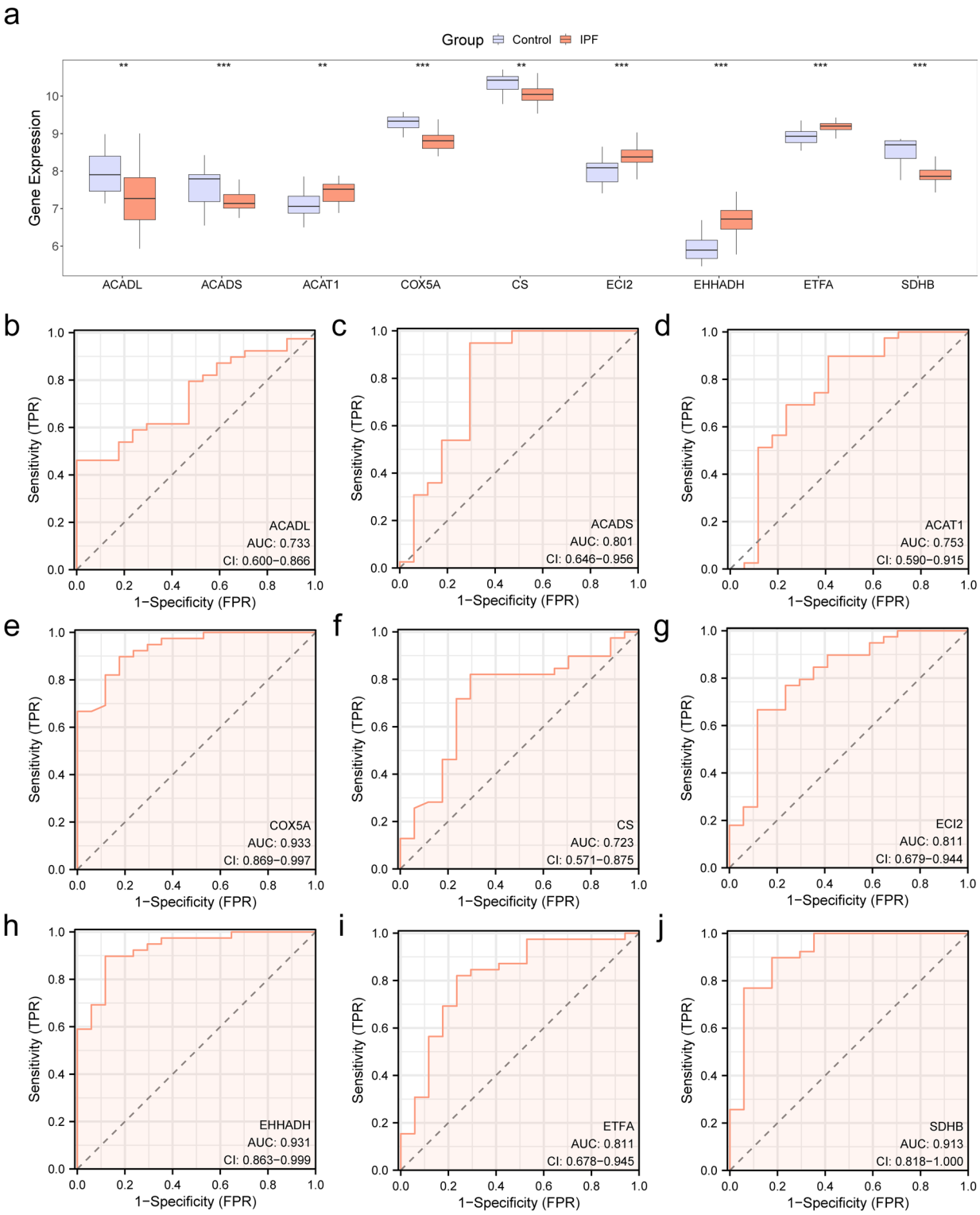

Fig. 8. Differential expression validation and ROC curve analysis.(**a**). Group comparison plots of hub genes in IPF and control samples from combined GEO datasets. In the group comparison plots, the IPF samples are orange, and the control samples are blue. (**b**–**j**). ROC curves of hub genes in the integrated GEO datasets .The x-axis represents the specificity of gene diagnosis, and the y-axis represents the sensitivity of diagnosis, with larger values indicating greater specificity or sensitivity. AUC > 0.9 indicates high accuracy, and AUC between 0.7 and 0.9 indicates moderate accuracy. $p < 0.01$, highly statistically significant; $p < 0.001$, very highly statistically significant. *IPF* idiopathic pulmonary fibrosis, *ROC* receiver operating characteristic, *AUC* area under the curve, *TPR* true positive rate, *FPR* false positive rate.

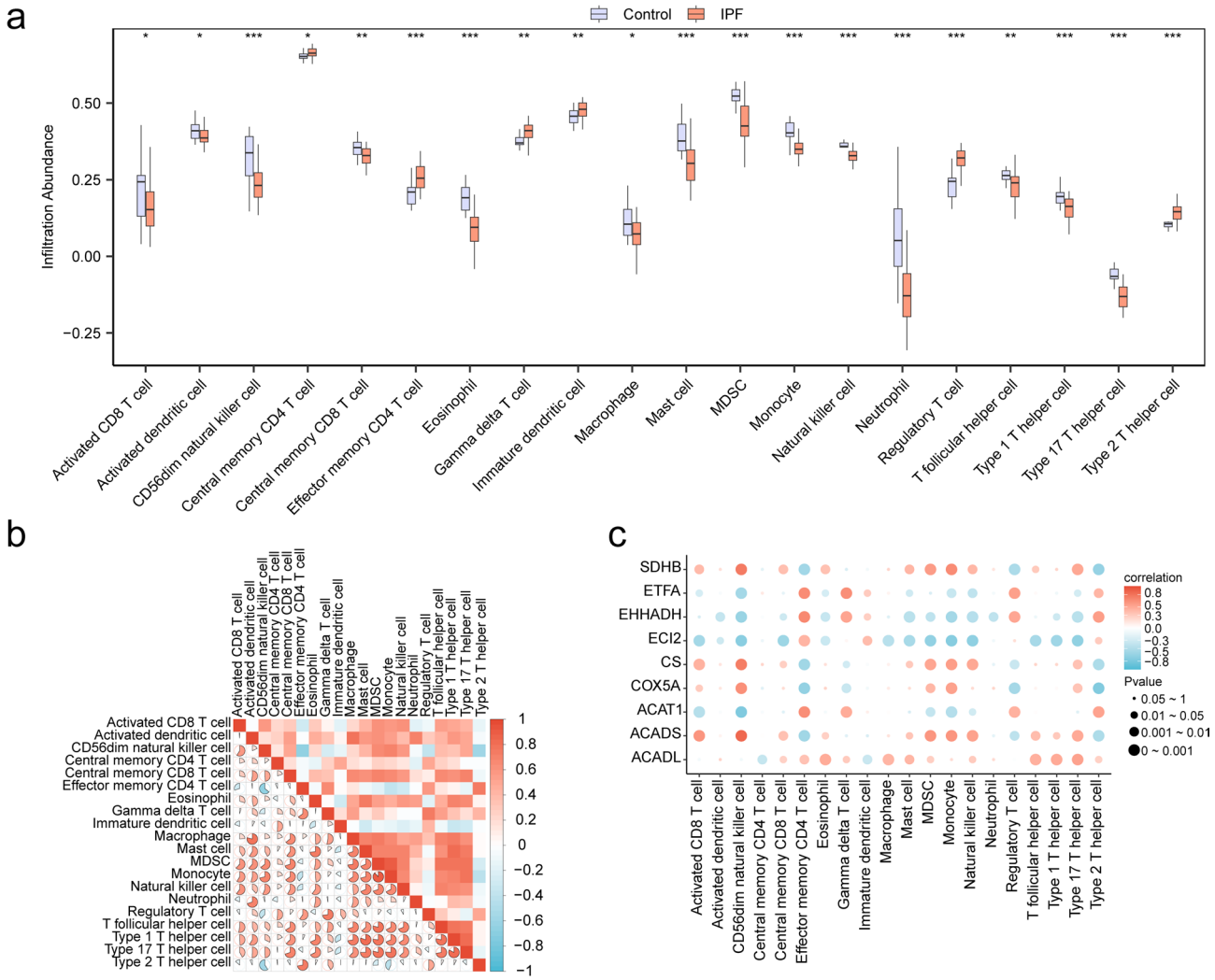

Fig. 9. Immune infiltration analysis using the SsGSEA algorithm.(**a**) Group comparison plot of immune cells in IPF and control samples from combined GEO datasets. (**b**) Heat map of immune cell infiltration abundance in the integrated GEO datasets. (**c**) Bubble plot of the correlation between hub genes and immune cell infiltration abundance in the integrated GEO datasets. ssGSEA, single-sample gene set enrichment analysis. * represents $p < 0.05$, indicating statistical significance; ** represents $p < 0.01$, indicating high statistical significance; *** represents $p < 0.001$, indicating very high statistical significance. In the group comparison plots, the IPF samples are depicted in orange, and the control samples are depicted in blue. The absolute value of the correlation coefficient (r-value) below 0.3 indicates weak or no correlation, 0.3 to 0.5 indicates a weak correlation, 0.5 to 0.8 indicates a moderate correlation, and above 0.8 indicates a strong correlation. In the correlation heatmap, red and blue represent positive and negative correlations, respectively. The depth of the color indicates the strength of the correlation.

Data availability

The datasets employed in our study can be acquired in the GEO repository (https://www.ncbi.nlm.nih.gov/geo/). The accession numbers are GSE24206 and GSE110147.We collected MEMRGs from the GeneCards database (https://www.genecards.org/) and the related literature (https://pubmed.ncbi.nlm.nih.gov/).

Received: 23 November 2024; Accepted: 8 May 2025
Published online: 14 May 2025

References

1. Richeldi, L., Collard, H. R. & Jones, M. G. Idiopathic pulmonary fibrosis. *Lancet* **389**, 1941–1952 (2017).
2. King, T. E. Jr et al. A phase 3 trial of Pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl. J. Med.* **370**, 2083–2092 (2014).
3. Le Pavec, J. et al. Lung transplantation for idiopathic pulmonary fibrosis. *Presse Med.* **49**, 104026. https://doi.org/10.1016/j.lpm.2020.104026 (2020).

4. Liu, J. et al. Mitochondrial quality control in lung diseases: current research and future directions. *Front. Physiol.* **14**, 1236651. https://doi.org/10.3389/fphys.2023.1236651 (2023).
5. Piao, L., Marsboom, G. & Archer, S. L. Mitochondrial metabolic adaptation in right ventricular hypertrophy and failure. *J. Mol. Med.* **88**, 1011–1020. https://doi.org/10.1007/s00109-010-0679-1 (2010).
6. Yue, Y. L. et al. The role of autophagy in idiopathic pulmonary fibrosis: from mechanisms to therapies. *Ther. Adv. Respir Dis.* **16**, 17534666221140972. https://doi.org/10.1177/17534666221140972 (2022).
7. Chen, H. et al. The effect of Tocilizumab treatment for skin fibrosis by inhibiting CD38 + macrophages in systemic sclerosis. *Cell. Immunol.* **408**, 104914. https://doi.org/10.1016/j.cellimm.2024.104914 (2024).
8. Dutta, A. et al. Tissue fibrosis in cardiorenal syndrome: crosstalk between heart and kidneys. *Nephrol. Dial Transpl.* https://doi.org/10.1093/ndt/gfaf009 (2025). gfaf009.
9. Tiwari, P. et al. Immune cells crosstalk pathways, and metabolic alterations in idiopathic pulmonary fibrosis. *Int. Immunopharmacol.* **135**, 112269. https://doi.org/10.1016/j.intimp.2024.112269 (2024).
10. Göös, H. et al. Human transcription factor protein interaction networks. *Nat. Commun.* **13**, 766. https://doi.org/10.1038/s41467-022-28341-5 (2022).
11. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508. https://doi.org/10.1038/s41586-022-04434-5 (2022).
12. Meltzer, E. B. et al. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. *BMC Med. Genomics*. **4**, 70. https://doi.org/10.1186/1755-8794-4-70 (2011).
13. Wang, H. et al. Integrative analyses of genes associated with idiopathic pulmonary fibrosis. *J. Cell. Biochem.* **120**, 8648–8660. https://doi.org/10.1002/jcb.28153 (2019).
14. Cecchini, M. J. et al. Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. *Respir Res.* **19**, 153. https://doi.org/10.1186/s12931-018-0857-1 (2018).
15. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995. https://doi.org/10.1093/nar/gks1193 (2013).
16. Davis, S. & Meltzer, P. S. GEOquery: a Bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* **23**, 1846–1847. https://doi.org/10.1093/bioinformatics/btm254 (2007).
17. Stelzer, G. et al. The genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.* **54** (30.1–1), 1. https://doi.org/10.1002/cpbi.5 (2016).
18. Cao, Z. et al. An integrated bioinformatic investigation of mitochondrial energy metabolism genes in colon adenocarcinoma followed by preliminary validation of CPT2 in tumor immune infiltration. *Front. Immunol.* **13**, 959967. https://doi.org/10.3389/fimmu.2022.959967 (2022).
19. Leek, J. T. et al. The Sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883. https://doi.org/10.1093/bioinformatics/bts034 (2012).
20. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. https://doi.org/10.1093/nar/gkv007 (2015).
21. Ben Salem, K. Ben Abdelaziz, A. Principal component analysis (PCA). *Tunis Med.* **99**, 383–389 (2021).
22. Zhang, H. et al. RCircos: an R package for circos 2D track plots. *BMC Bioinform.* **14** https://doi.org/10.1186/1471-2105-14-244 (2013).
23. Mi, H. et al. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426. https://doi.org/10.1093/nar/gky1038 (2019).
24. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
25. Yu, G. et al. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287. https://doi.org/10.1089/omi.2011.0118 (2012).
26. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A.* **102**, 15545–15550. https://doi.org/10.1073/pnas.0506580102 (2005).
27. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260 (2011).
28. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613. https://doi.org/10.1093/nar/gky1131 (2019).
29. Dang, Y. et al. FTH1- and SAT1-induced astrocytic ferroptosis is involved in Alzheimer's disease: evidence from single-cell transcriptomic analysis. *Pharmaceuticals* **15**, 1177. https://doi.org/10.3390/ph15101177 (2022).
30. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. https://doi.org/10.1101/gr.1239303 (2003).
31. Yang, X. et al. Study on the multitarget mechanism and key active ingredients of herba siegesbeckiae and volatile oil against rheumatoid arthritis based on network Pharmacology. *Evid. Based Complement. Alternat Med.* **2019** (8957245). https://doi.org/10.1155/2019/8957245 (2019).
32. Zhou, K. R. et al. ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.* **45**, D43–D50. https://doi.org/10.1093/nar/gkw965 (2017).
33. Vlachos, I. S. et al. Diana-TarBase v7.0: indexing more than half a million experimentally supported MiRNA:mRNA interactions. *Nucleic Acids Res.* **43**, D153–D159. https://doi.org/10.1093/nar/gku1215 (2015).
34. Xiao, B. et al. Identification and verification of immune-related gene prognostic signature based on SsGSEA for osteosarcoma. *Front. Oncol.* **10**, 607622. https://doi.org/10.3389/fonc.2020.607622 (2020).
35. Raghu, G. et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am. J. Respir Crit. Care Med.* **183**, 788–824. https://doi.org/10.1164/rccm.2009-040GL (2011).
36. Ley, B. et al. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am. J. Respir Crit. Care Med.* **183**, 431–440. https://doi.org/10.1164/rccm.201006-0894CI (2011).
37. King, T. E. Jr et al. Idiopathic pulmonary fibrosis. *Lancet* **378**, 1949–1961. https://doi.org/10.1016/S0140-6736(11)60052-4 (2011).
38. Sosulski, M. L. et al. Deregulation of selective autophagy during aging and pulmonary fibrosis: the role of TGFβ1. *Aging Cell.* **14**, 774–783. https://doi.org/10.1111/acel.12357 (2015).
39. Mora, A. L. et al. Mitochondria in the spotlight of aging and idiopathic pulmonary fibrosis. *J. Clin. Invest.* **127**, 405–414. https://doi.org/10.1172/JCI87440 (2017).
40. Zeng, J. et al. miR-204/COX5A axis contributes to invasion and chemotherapy resistance in Estrogen receptor-positive breast cancers. *Cancer Lett.* **492**, 185–196. https://doi.org/10.1016/j.canlet.2020.07.027 (2020).
41. Zhang, P. et al. COX5A alleviates doxorubicin-induced cardiotoxicity by suppressing oxidative stress, mitochondrial dysfunction and cardiomyocyte apoptosis. *Int. J. Mol. Sci.* **24**, 10400. https://doi.org/10.3390/ijms241210400 (2023).
42. Wanders, R. J. A. et al. The physiological functions of human peroxisomes. *Physiol. Rev.* **103**, 957–1024. https://doi.org/10.1152/physrev.00051.2021 (2023).
43. Ranea-Robles, P. et al. Peroxisomal L-bifunctional protein (EHHADH) deficiency causes male-specific kidney hypertrophy and proximal tubular injury in mice. Kidney **360** 2, 1441–1454 (2021). https://doi.org/10.34067/KID.0003772021

44. Parker, A. R. et al. Lipid mediators in aspirin-exacerbated respiratory disease. *Immunol. Allergy Clin. North. Am.* **36**, 749–763. https://doi.org/10.1016/j.iac.2016.06.009 (2016).
45. Zhang, H. et al. SDHB reduction promotes oral lichen planus by impairing mitochondrial respiratory function. *Ann. Transl Med.* **10**, 1367. https://doi.org/10.21037/atm-22-5999 (2022).
46. Burnichon, N. et al. The succinate dehydrogenase genetic testing in a large prospective series of patients with paragangliomas. *J. Clin. Endocrinol. Metab.* **94**, 2817–2827. (2009).
47. Goudarzi, A. The recent insights into the function of ACAT1: a possible anti-cancer therapeutic target. *Life Sci.* **232**, 116592. https://doi.org/10.1016/j.lfs.2019.116592 (2019).
48. Li, J. et al. Metabolic reprogramming of pulmonary fibrosis. *Front. Pharmacol.* **13**, 1031890. https://doi.org/10.3389/fphar.2022.1031890 (2022).
49. Sharma, A. et al. Mitochondrial dynamics and mitophagy in lung disorders. *Life Sci.* **284**, 119876. https://doi.org/10.1016/j.lfs.2021.119876 (2021).
50. Chen, Y. & Su, Z. Reveal genes functionally associated with ACADS by a network study. *Gene* **569**, 294–302. https://doi.org/10.1016/j.gene.2015.05.069 (2015).
51. Chhimpa, N. et al. The novel role of mitochondrial citrate synthase and citrate in the pathophysiology of Alzheimer's disease. *J. Alzheimers Dis.* **94**, S453–S472. https://doi.org/10.3233/JAD-220514 (2023).
52. Chen, Z. et al. Involvement of the Warburg effect in non-tumor diseases processes. *J. Cell. Physiol.* **233**, 2839–2849. https://doi.org/10.1002/jcp.25998 (2018).
53. Yan, Y. et al. Electron transfer Flavoprotein (ETF) α controls blood vessel development by regulating endothelial mitochondrial bioenergetics and oxygen consumption. *Oxid. Med. Cell. Longev.* (7969916). https://doi.org/10.1155/2022/7969916 (2022).
54. Zhao, X. et al. ACADL plays a tumor-suppressor role by targeting Hippo/YAP signaling in hepatocellular carcinoma. *Npj Precis Oncol.* **4** (7). https://doi.org/10.1038/s41698-020-0111-4 (2020).
55. Wynn, T. A. Cellular and molecular mechanisms of fibrosis. *J. Pathol.* **214**, 199–210. https://doi.org/10.1002/path.2277 (2008).
56. Luzina, I. G. et al. Roles of T lymphocytes in pulmonary fibrosis. *J. Leukoc. Biol.* **83**, 237–244. https://doi.org/10.1189/jlb.0707504 (2008).
57. Fireman, E. et al. Predictive value of response to treatment of T-lymphocyte subpopulations in idiopathic pulmonary fibrosis. *Eur. Respir J.* **11**, 706–711. https://doi.org/10.1183/09031936.98.11030706 (1998).
58. Hüttemann, M. et al. Regulation of mitochondrial oxidative phosphorylation through cell signaling. *Biochim. Biophys. Acta.* **1773**, 1701–1720. https://doi.org/10.1016/j.bbamcr.2007.10.001 (2007).
59. Ali, A., Ajil, A., Meenakshi Sundaram, A. & Joseph, N. Detection of gene ontology clusters using biclustering algorithms. *SN Comput. Sci.* **4**, 217. https://doi.org/10.1007/s42979-022-01624-w (2023).
60. Ali, A., Mohan, J., Nadaf, T. A. A., Ravishankar, H. & Deepa, K. R. Bioinformatics-Driven discovery of signaling pathways and genes influencing cervical Cancer. *SN Comput. Sci.* **5**, 1–7. https://doi.org/10.1007/s42979-024-03347-6 (2024).
61. Piantadosi, C. A. & Suliman, H. B. Mitochondrial dysfunction in lung pathogenesis. *Annu. Rev. Physiol.* **79**, 495–515. https://doi.org/10.1146/annurev-physiol-022516-034322 (2017).
62. Geng, J. et al. Fatty acid metabolism and idiopathic pulmonary fibrosis. *Front. Physiol.* **12**, 794629. https://doi.org/10.3389/fphys.2021.794629 (2022).
63. Otoupalova, E. et al. Oxidative stress in pulmonary fibrosis. *Compr. Physiol.* **10**, 509–547. https://doi.org/10.1002/cphy.c190017 (2020).
64. Hart, C. M. et al. PPARgamma: a novel molecular target in lung disease. *J. Investig Med.* **56**, 515–517. https://doi.org/10.2310/JIM.0b013e318165e89d (2008).

Acknowledgements

We thank the grants from the Shandong Postdoctoral Innovation Program. We thank the GEO database for its platforms.

Author contributions

YL: Software, Data curation, Visualization, Writing- original draft LB: Writing - Review and Editing WY: Conceptualization and Methodology.

Funding

This work was supported by the Shandong Postdoctoral Innovation Program (grant SDCX-ZG-202302011).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01759-9.

Correspondence and requests for materials should be addressed to Y.W.