scientific reports

OPEN

Integrated bioinformatics analysis screened the key genes and pathways of idiopathic pulmonary fibrosis

Juan Wu[1,2], Yangyang Wei[1,2], Hong Kang[1], Xiao Yu[1], Shuting Wei[1], Ting Xue[1] & Xiaomei Kong[1✉]

To investigate the molecular mechanisms underlying idiopathic pulmonary fibrosis (IPF), we analyzed the GSE173355 and GSE173356 datasets obtained from the NCBI-GEO database. We identified differentially expressed genes (DEGs) and differentially methylated sites. Functional enrichment analysis was conducted for both DEGs and differentially methylated sites. Functional enrichment analysis was performed for both DEGs and differentially methylated sites, alongside an examination of immune-related scores, proportions, and GSVA enrichment scores of immune cells in IPF versus control samples. An integrated gene-methylation association analysis revealed 8 genes with expression levels negatively influenced by methylation. The Rap1 pathway, Focal adhesion, and Axon guidance were significantly enriched among both DEGs and differentially methylated sites. Immune-related scores were notably lower in the IPF group compared to the control group, with marked differences in immune cell proportions and GSVA enrichment scores. Screening of DEGs identified 361 differentially expressed immune-related genes (IRGs). Protein-protein interaction (PPI) network analysis using the STRING database and Cytoscape software unveiled 10 key genes and 3 core subnetworks. RT-qPCR validation in bleomycin-induced IPF mouse model and A549 EMT model confirmed the reliability of most findings. These results provide new insights into IPF pathogenesis and potential therapeutic strategies, necessitating further functional validation.

Keywords Idiopathic pulmonary fibrosis, Integrated bioinformatics, Immune environment, Immune-related gene, Methylation

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, fibrotic interstitial pneumonia characterized by fibroblast hyperplasia, massive extracellular matrix deposition, and structural destruction of lung tissue, which is caused by abnormal repair of alveolar epithelium after repeated minor damage[1]. IPF is more common in the elderly, and is often accompanied by progressive dyspnea and deterioration of pulmonary function[2]. Without active treatment, the average life expectancy after diagnosis is only three to five years[2]. At present, the pathogenesis of IPF is still unclear, which limits the development of therapeutic methods. Therefore, it is of great clinical significance to screen out specific molecular biomarkers that are beneficial to the diagnosis and treatment of IPF and to explore their potential molecular mechanisms.

DNA methylation is a newly discovered biomarker used in diagnosis, prognosis and prediction of treatment[3]. It is also one of the most characteristic, earliest discovered and most important epigenetic modifications, which can provide new research directions and treatment strategies for various diseases[4]. In the development of IPF, DNA methylation changes are often captured and further exacerbate the disease[5]. Further explanation of the mechanism can provide new diagnostic methods and treatment strategies for pulmonary diseases. In addition, although the pathogenesis of IPF has not been fully elucidated, it is currently believed that it is mainly related to repeated damage of epithelial cells, fibrosis and accumulation of collagen caused by abnormal tissue damage repair[6]. Studies have shown that the interaction between inflammatory cells and fibroblasts can promote the occurrence and development of IPF[6,7]. Immune dysregulation has been identified as the driving factor of IPF[8], and improving the understanding of the role of the immune system in IPF and developing therapies targeting

[1]NHC Key Laboratory of Pneumoconiosis, Shanxi Key Laboratory of Respiratory Diseases, Department of Pulmonary and Critical Care Medicine, The First Hospital of Shanxi Medical University, Taiyuan, China. [2]Juan Wu and Yangyang Wei contributed equally to this work. ✉email: kongxm@sxmu.edu.cn

1

immune regulation can prevent or even reverse pulmonary fibrosis. Therefore, an in-depth investigation of the differential immune-related genes (IRGs) in IPF plays an important role in understanding the disease.

It is generally believed that new biomarkers can be discovered by using gene chips and gene testing technology. With the wide application of microarray, online public databases can provide a large amount of genomic data and valuable information for molecular cytology research. In this study, gene expression profile data and corresponding methylation data of IPF were downloaded from the NCBI-GEO database, and differentially expressed genes (DEGs) and differentially expressed methylation sites were screened. Systematic bioinformatics methods were used to analyze the biological functions and related pathways of DEGs and differentially expressed methylation sites, and to analyze the relationship between IPF and immunity, so as to clarify the possible pathogenesis of IPF and provide a research basis for IPF prognostic factors and potential therapeutic targets.

Material and methods

Dataset sources

GSE173355 and GSE173356 profile datasets were downloaded from the NCBI-GEO database (https://www.ncbi.nlm.nih.gov/geo/). GSE173355 is RNA expression profile data, including 23 samples from IPF and 14 samples from control, and the platform is GPL24676 Illumina NovaSeq 6000 (Homo sapiens). And GSE173356 is the corresponding methylation data, also including 23 IPF samples and 14 control samples, and the platform is GPL23976 Illumina Infinium HumanMethylation850 BeadChip. Platform and series matrix file(s) were downloaded as TXT files.

Expression analysis of DEGs and differentially methylated sites

The original data of GSE173355 and GSE173356 were transformed into an expression matrix, and then the expression of 37 samples was normalized. Firstly, the DESeq software package of R language (version 3.6.2) was used to perform Principal Components Analysis (PCA) on each sample in GSE173355 according to the mRNA expression. DEGs analysis between IPF and control cohorts was conducted utilizing both the limma and edgeR packages in the R language. Furthermore, differentially methylated sites between IPF and control was performed employing the t-test methodology implemented in the limma package[9]. Log fold change > 1 and FDR < 0.05 were used as screening criteria for differentially expressed transcripts, and FDR < 0.01 was used as screening criteria for differentially methylated sites. And the R package was used to draw the volcano map and heatmap of DEGs or differentially methylated sites, respectively. In addition, the DMRcate package of R language was used to identify differentially methylated regions, each of which contained at least two consecutive differentially methylated sites.

Functional enrichment analyses

To further explore the biological processes and signaling pathways that DEGs or differentially methylated sites may be involved in, we subsequently performed functional analyses. ClusterProfiler[10] in R language was used to analyze the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of DEGs, and missMethyl package[11] was used to analyze the GO and KEGG pathways of differentially methylated sites. Furthermore, the biological functions of DEGs were analyzed by GSVA package[12], using KEGG gene sets (c2.cp.kegg.v7.5.1.symbols.gmt) download from GSEA website (https://www.gsea-msigdb.org/gsea/index.jsp). Then, the differential expression of 185 KEGG pathways was analyzed by limma package, and the differential KEGG pathways were selected according to log FC > 0.1 and FDR < 0.05, and then the heatmap was drawn.

Association analysis of mRNA and methylation

The ELMER package[13] of R language was used to analyze the association between mRNA and methylation, and the genes whose expression levels were significantly affected by methylation were screened.

Analysis of immune cells

ESTIMATE algorithm[14] was used to calculate the stromal score, immune score and ESTIMATE score in the IPF and control sample. To account for potential confounders such as age, sex, and smoking history, we conducted a multivariate regression analysis to adjust for these factors in the immune-related score and immune cell composition analyses. And the composition of 22 immune cells in each sample was calculated by CIBERSORTx (https://cibersortx.stanford.edu/), the proportion map and heatmap were drawn, and differences between groups were analyzed. In addition, based on the gene annotation files related to 29 immune cells[15], the GSVA package of R was used to analyze the score of immune cells in IPF samples and control samples, and the boxplot was drawn.

Integration of protein-protein interaction networks (PPI)

Immune-related genes (IRGs) were obtained from the ImmPort database. The DEGs obtained by differential expression analysis between IPF and control samples were intersected with IRGs to obtain differentially expressed IRGs, and draw a Venn diagram. PPI of differentially expressed IRGs was constructed using STRING (https://cn.string-db.org/)[16], and interactions with combined scores > 0.9 were defined as statistically significant. The integrated regulatory network was visualized using Cytoscape software[17]. The hub nodes in the PPI network were analyzed according to Cytoscape plug-in cytohubba. The top10 genes were defined as hub genes according to the degree of nodes, and the interaction network of top10 hub nodes was drawn. The Cytoscape plug-in MCODE was used to further detect the core subnetwork in the PPI network.

Validation of DNA methylation-related and immune-associated gene expression in bleomycin-induced IPF mouse model and TGF-β-induced EMT cell model in A549 cells

Shanxi Medical University reviewed and granted approval for this study. All procedures involving animals were conducted in accordance with the guidelines established by the China FDA and other relevant ethical standards.

Additionally, the research adhered to the ARRIVE guidelines to enhance transparency and reproducibility. Every effort was made to minimize suffering and employ humane endpoints. The male C56BL/6 mice (6–8 weeks of age, average weight 20–23 g) randomly divided into two groups: the IPF group, and the control group. The mice were purchased from the Laboratory Animal Center of Shanxi Medical University (Shanxi, China). BLM solution for use was prepared by dissolving 10 mg BLM in 4mL sterile 0.9% NaCl. After anesthesia with isoflurane, the IPF animal model was established by intratracheal instillation of BLM (2.5 mg/kg, MedChemExpress, HY−17565, USA), while the control group received an equal amount of saline via intratracheal instillation. The bleomycin model is the best-characterized and the most extensively used animal model of IPF[18]. After 21 days, the animals were anesthetized under isoflurane, and euthanized by cervical dislocation, and tissue samples were collected for analyses, part of which was used for subsequent verification and the other part for HE and Masson staining.The A549 cell line, obtained from the Chinese Academy of Sciences Cell Bank (Shanghai, China), was maintained in RPMI-1640 supplemented with 10% fetal bovine serum at 37 °C with 5% CO_2. Cells were seeded in plate and serum-starved for 12 h, followed by 72 h of 10 ng/mL TGF-β1 treatment to establish the EMT model[19]. α-SMA (ET1607-53), Vimentin (ET1610-39), and E-cadherin (ET1607-75) (Hangzhou Huaan Biotechnology Co., Ltd.). RNA extracted from lung tissue and cell were collected for Quantitative Real-time PCR (qPCR). Relative gene expression was calculated with the Eq. $2^{-\Delta\Delta ct}$.

Results

Data information and identification of DEGs

The mRNA expression levels of 37 samples (23 IPF, 14 control) from the microarray dataset GSE173355 were normalized. The boxplot post-normalization is depicted in Fig. 1A. Subsequently, the DESeq software package in R was utilized for principal components analysis (PCA) based on expression levels, revealing significant differences between the IPF and control groups (Fig. 1B). To enhance the robustness of our study, multiple independent R packages were employed for differential expression analysis. Differential expression analysis of mRNA in the GSE173355 dataset was conducted using the limma package in R. A total of 5328 differentially expressed transcripts, corresponding to 4083 genes, were identified based on logFC > 1 and FDR < 0.05 (Supplementary Table 1). Subsequently, employing the same criteria, differential expression analysis of mRNA was performed using the edgeR package in R, resulting in 4876 differentially expressed transcripts, corresponding to 3712 genes (Supplementary Table 2). A total of 4012 transcripts were found to be differentially expressed by both R packages, corresponding to 3206 genes (Fig. 2A,B). The volcano plot of DGEs is presented in Figs. 1C and 2C, while the heatmap based on DEGs is displayed in Figs. 1D and 2D. These DEGs effectively distinguished IPF from control samples, as illustrated in the figures.

Functional enrichment analysis of DEGs

Comprehensive Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were systematically performed on the differentially expressed genes (DEGs) identified by the limma and edgeR packages to elucidate the most significantly associated biological functions and pathways. The results were highly consistent. The dot-plot illustrating GO enrichment analysis can be observed in Figs. 1E and 2E. Regarding biological processes (BP), DEGs exhibited significant enrichment in "cell junction assembly," "cell-substrate adhesion," and "extracellular matrix organization." In terms of cellular components (CC), DEGs were predominantly enriched in "cell-substrate junction," "focal adhesion," and "collagen-containing extracellular matrix." Concerning molecular functions (MF), DEGs were primarily enriched in "integrin binding," "growth factor binding," "protein tyrosine kinase activity," "transmembrane receptor protein kinase activity," and "actin binding," all closely linked to the progression of pulmonary fibrosis[20]. The KEGG pathway analysis revealed significant enrichment of DEGs in pathways such as "PI3K-Akt signaling pathway," "MAPK signaling pathway," "Focal adhesion," "Axon guidance," and "Rap1 signaling pathway"[21] (Figs. 1F and 2F). The MAPK and PI3K/Akt signaling pathways may collaboratively participate in the progression of fibrogenesis, potentially regulating extracellular matrix deposition synergistically[20]. The biological functions of DEGs were analyzed using the GSVA package in R language. The GSVA package (Bioconductor v2.0.5) was utilized to characterize the DEGs derived from limma and edgeR, resulting in 180 and 101 KEGG pathways, respectively, identified through GSEA portal analysis (Supplementary Table 3). A Venn diagram analysis showed a complete overlap (101 pathways) in KEGG pathway signatures (Fig. 2G). The limma package in R was employed to assess the differential expression of KEGG pathways. Differentially expressed KEGG pathways were selected based on criteria of FDR < 0.05 and logFC > 1. The heatmap presented in Figs. 1G and 2H illustrates clear separation between IPF and control samples, indicating distinct pathway differences between the two groups.

Methylation data analysis

Methylation is generally considered to play an important role in gene expression. We used the GSE173356 dataset to analyze the differences in methylation between IPF and controls. First, the methylation data were normalized, and the results before and after normalization were shown in Fig. 3A. Then, the differentially methylated sites of the normalized methylation data were analyzed by using the limma package, and a total of 4933 differentially methylated sites were selected according to FDR < 0.01. The volcano map of the differentially methylated sites was shown in Fig. 3B. A heatmap based on differentially methylated sites was shown in Fig. 3C. The DMRcate package was further used to identify differentially methylated regions, each containing at least two consecutive differentially methylated sites. The average methylation level difference between the IPF and control groups was > 0.1, and a total of 87 differentially methylated regions were identified, as shown in Supplementary Table 4. A schematic representation of the first and second differentially methylated regions was shown in Supplementary Fig. 1.
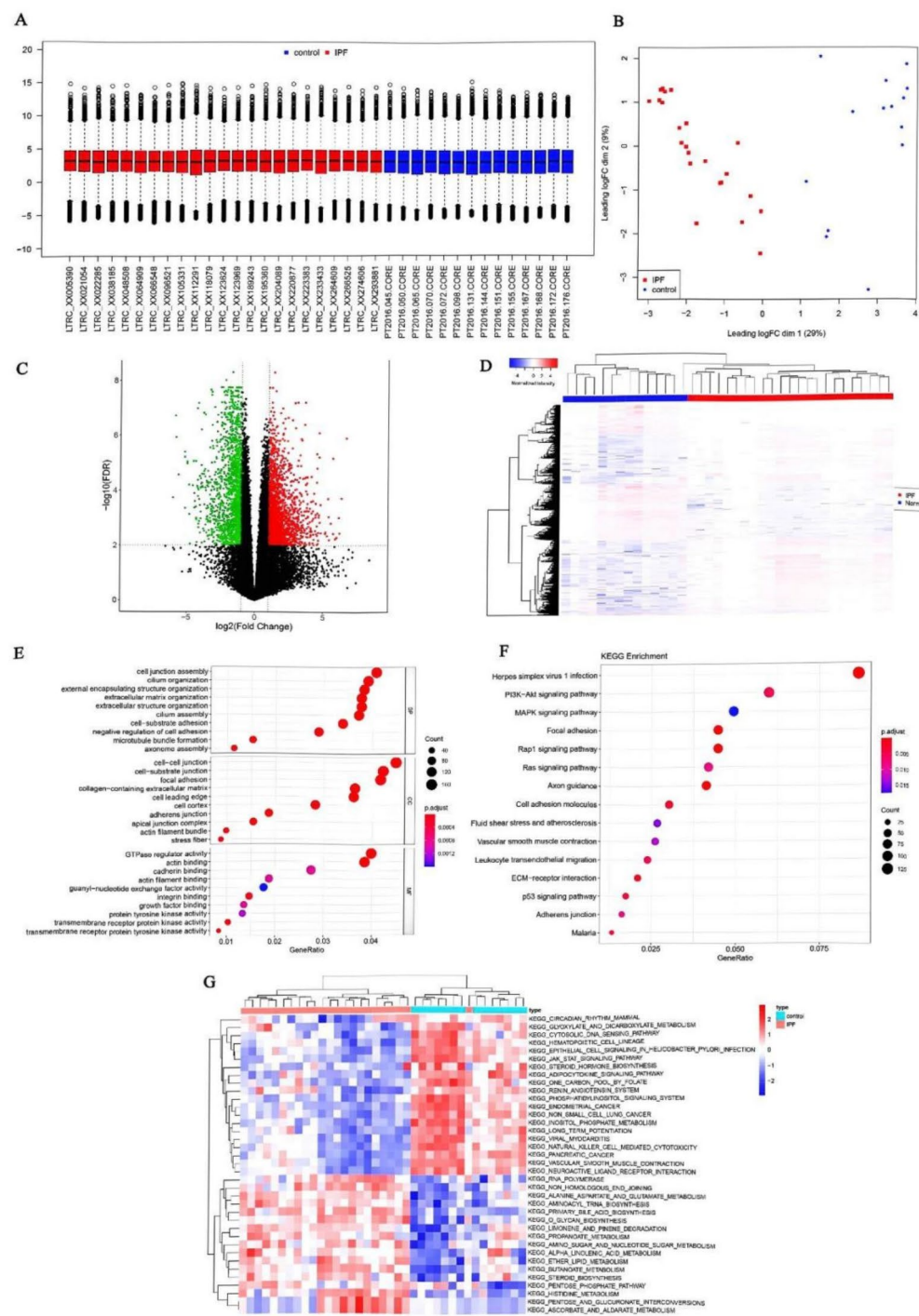
Fig. 1. Analysis of differentially expressed genes (DEGs) between IPF and control samples (**A**) Standardization of gene expression in GSE173355. (**B**) Principal component analysis (PCA) results between IPF and control group. (**C**) Volcano map of differentially expressed genes analysis. (**D**) Heatmap of differentially expressed genes between IPF and control group. (**E**) Dot-plot of GO enrichment analysis. (**F**) Significantly enriched KEGG pathways obtained from KEGG analysis. (**G**) Heatmap of 37 selected differential KEGG pathways. *GO* gene ontology, *BP* biological process, *CC* cellular component, *MF* molecular function, *KEGG* Kyoto Encyclopedia of Genes and Genomes, *IPF* Idiopathic pulmonary fibrosis.
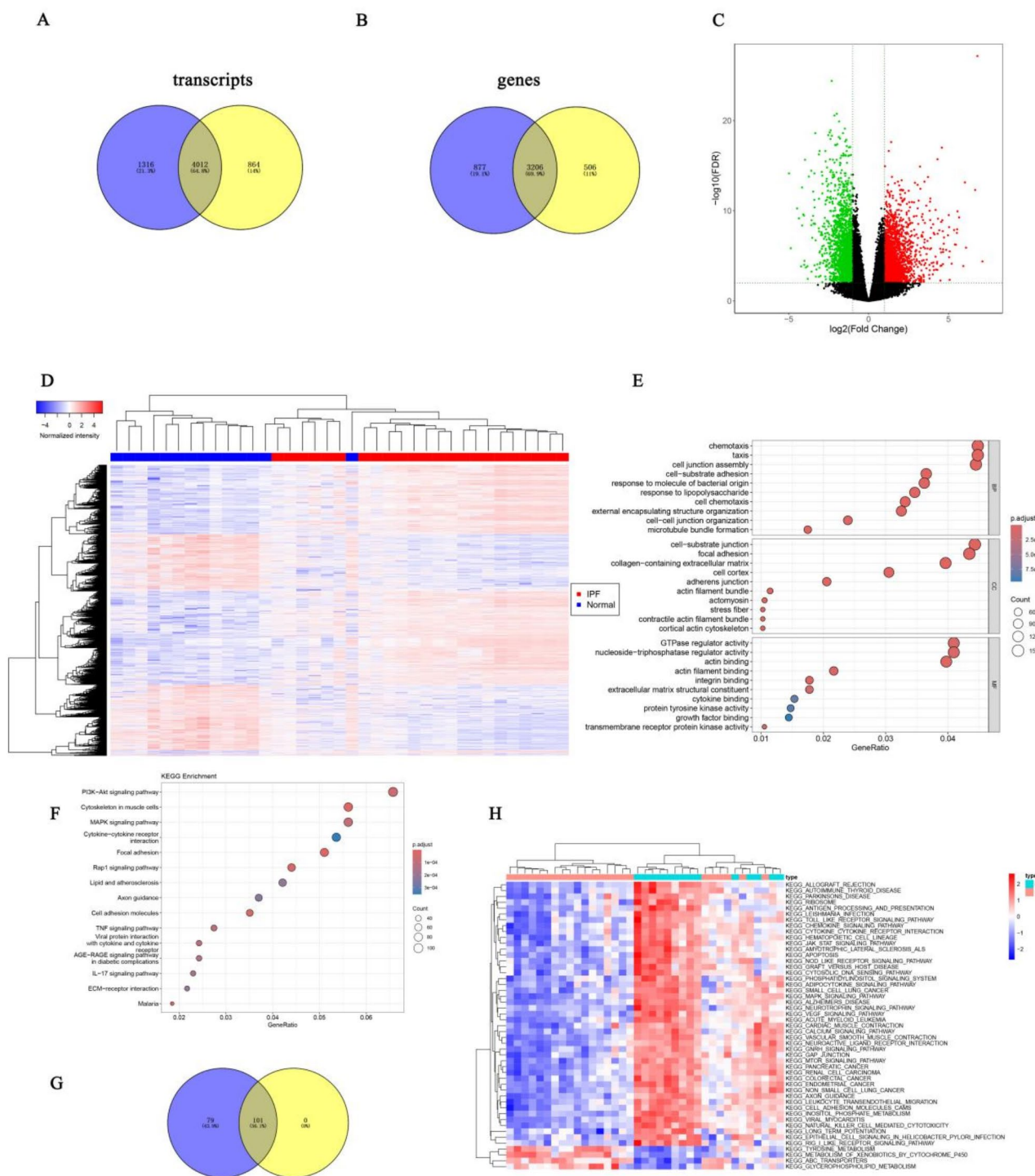
Fig. 2. Analysis of DEGs between IPF and control samples was conducted using edgeR Venn diagrams illustrate the overlapping differentially expressed transcripts (**A**) and genes (**B**) identified by the Limma and edgeR packages, respectively. (**C**) Volcano map of differentially expressed genes analysis. (**D**) Heatmap of differentially expressed genes between IPF and control group. (**E**) Dot-plot of GO enrichment analysis. (**F**) Significantly enriched KEGG pathways obtained from KEGG analysis. (**G**) Venn diagrams were utilized to identify the overlapping KEGG pathways derived from the DEGs analyzed using both the limma and edgeR packages.(**H**) Heatmap of selected differential KEGG pathways. *GO* gene ontology, *BP* biological process, *CC* cellular component, *MF* molecular function, *KEGG* Kyoto Encyclopedia of Genes and Genomes.
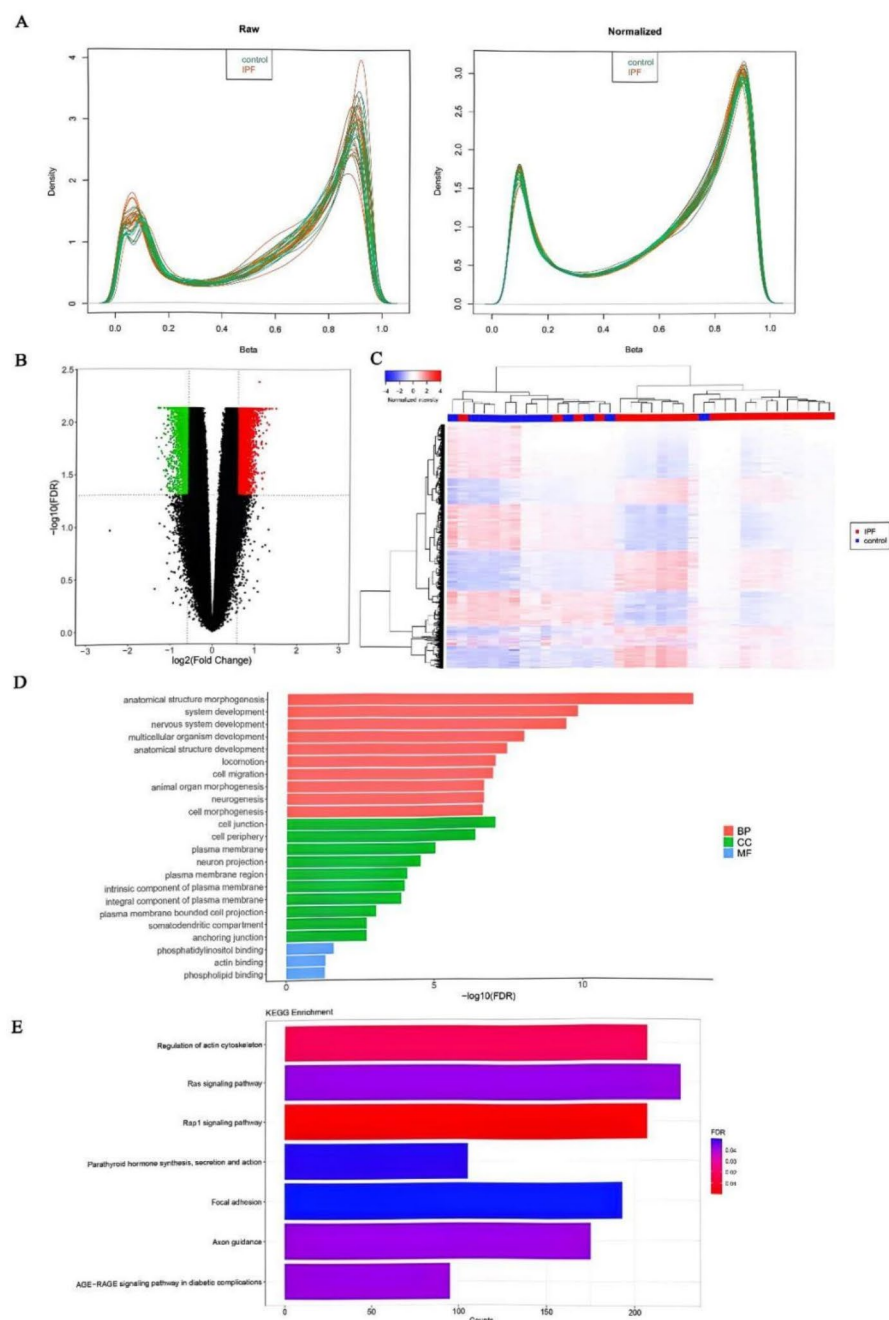
Fig. 3. Analysis of differentially methylated sites between IPF and control groups (**A**) Standardization of gene expression in GSE173356. (**B**) Volcano map of differentially methylated sites. (**C**) Heatmap of differentially methylated sites. (**D**) Column chart of GO enrichment analysis. (**E**) Significantly enriched KEGG terms obtained from KEGG analysis. *GO* gene ontology, *BP* biological process, *CC* cellular component, *MF* molecular function, *KEGG* Kyoto Encyclopedia of Genes and Genomes.

Function enrichment analysis of differentially methylated sites

The differentially methylated sites were analyzed by GO and KEGG enrichment using the missMethyl package. As shown in Fig. 3D, in BP, differentially methylated sites were significantly enriched in the cell migration, anatomical structure morphogenesis, and system development; in CC, they were enriched in cell junction, cell periphery and plasma membrane, and phosphatidylinositol binding; and in MF, they were enriched in actin binding and phospholipid binding. And in KEGG enrichment analysis (Fig. 3E), differentially methylated sites were significantly enriched in the Rap1 signaling pathway, focal adhesion, Axon guidance and Ras signaling pathway. It is worth noting that the functional enrichment results of DEGs and differentially methylated sites were similar.

Association analysis of mRNA and methylation

Then, the ELMER package was used to analyze the association between mRNA and methylation in IPF (Supplementary Fig. 2). The DEGs whose expression levels were significantly affected by methylation in IPF were screened, among which the CpG site CG11299543 in the promoter region of the PDCD1LG2 gene was hypermethylated, resulting in low expression of PDCD1LG2 in IPF (Fig. 4A). In addition, the CpG sites in the promoter regions of TTC34 (Fig. 4B), CNNM1 (Fig. 4C), ADAMTS16 (Fig. 4D), MKX (Fig. 4E), SLC22A3 (Fig. 4F), C1orf53 (Fig. 4G) and CP (Fig. 4H) genes were hypomethylated, resulting in high gene expression in IPF.

Immune microenvironment analysis of IPF

Immune dysregulation has been identified as a driver of IPF, so we further analyzed the immune environment in IPF. We use the ESTIMATE algorithm to calculate the immune-related scores in the IPF and control groups. As shown in Fig. 5A, stromal score, immune score and ESTIMATE score in the IPF group were significantly lower than those in the control group ($P < 0.01$), indicating that there was a large immune difference between the IPF and the control group. To account for potential confounders such as age, sex, and smoking history, we conducted a multivariate regression analysis to adjust for these factors in the immune-related score and immune
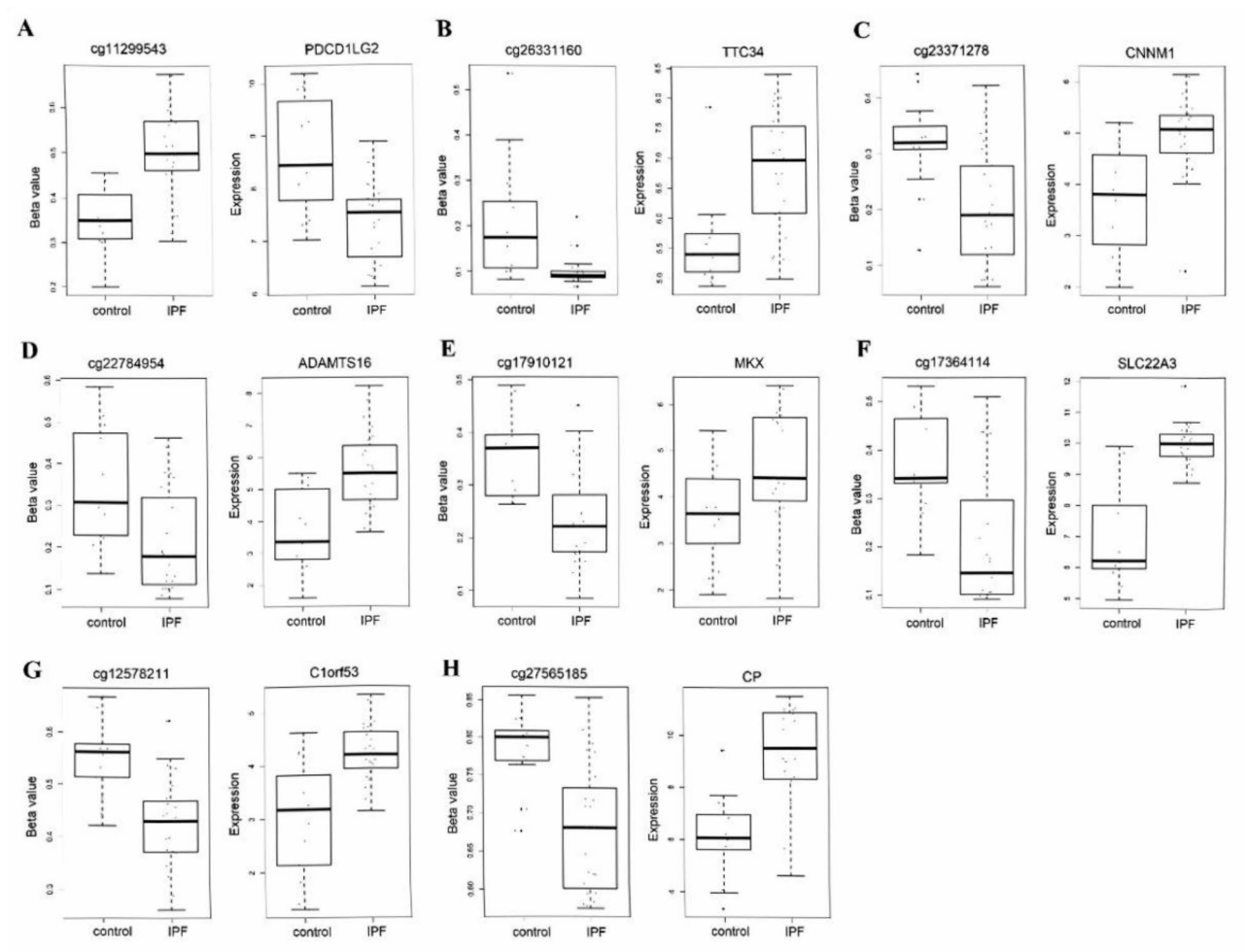

Fig. 4. Methylation levels of CpG sites and mRNA expression levels of genes. *IPF* Idiopathic pulmonary fibrosis.
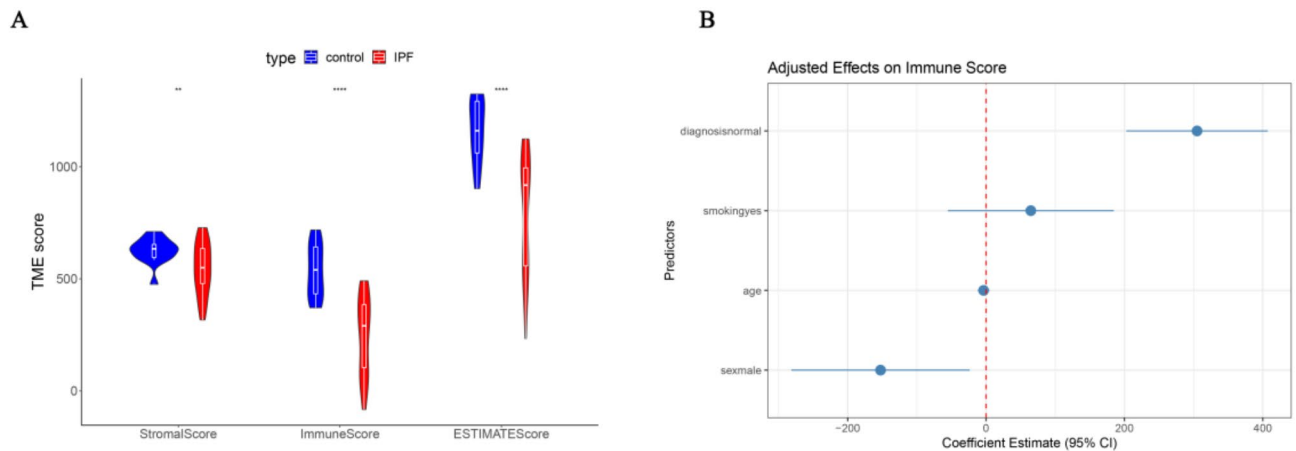
Fig. 5. Immune microenvironment analysis of IPF (**A**) Stromal score, immune score and ESTIMATE score in IPF and control groups ** $P < 0.01$, *** $P < 0.001$. (**B**) Multivariate regression analysis was conducted to adjust for the influence of these factors on immune-related scores and immune cell composition analysis.
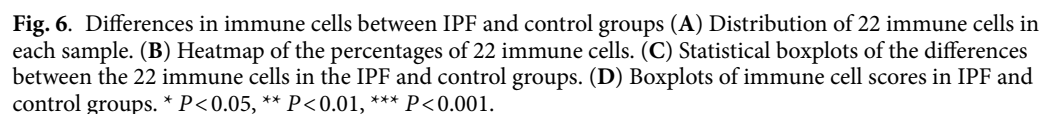
cell composition analyses. The normal group demonstrated significantly higher immune scores compared to the IPF group ($\beta = 304.97$, 95% CI = [206.8, 403.1], $p < 0.001$) (Fig. 5B). The confidence interval was entirely located to the right of the reference line ($\beta = 0$), indicating that the difference in immune scores between the normal and IPF groups remained highly statistically significant after adjusting for age, sex, and smoking. Males exhibited significantly lower immune scores than females ($\beta = -152.54$, 95% CI = [-277.5, -27.6], $p = 0.021$), establishing a statistically significant sex-based disparity. These sex differences may involve the regulation of immune function by sex hormones, suggesting the need to explore sex-specific immune intervention strategies. Although smokers exhibited a positive trend in immune scores, the confidence interval included zero, indicating that the independent association of smoking behavior did not reach statistical significance after adjusting for other variables. The age range of the population was 55–75 years, and as shown in the figure, the linear effect of age on immune scores may have been obscured by other variables.

The composition of 22 kinds of immune cells in each sample was further calculated using CIBERSORTx (https://cibersortx.stanford.edu/). The content of immune cells was shown in Fig. 6A and the heatmap of cells proportional was shown in Fig. 6B. The differences in the infiltration abundance of 12 immune cell types between IPF and normal control groups indicate significant alterations in the immune microenvironment of IPF patients. Among them, Dendritic cells resting, Macrophages M0, Macrophages M2, Mast cells resting, Plasma cells, T cells CD4 memory activated and T cells gamma delta in the IPF group were significantly higher than that in the control group, while Mast cells activated, Monocytes, Neutrophils, NK cells activated and T cells CD4 memory resting in the IPF group were significantly lower than that in the control group ($P < 0.05$, Fig. 6C), suggesting that these immune cells may be closely related to the formation of IPF.

In addition, according to the gene annotation files related to 29 kinds of immune cells[15] (Supplementary Table 5), the GSVA package of the R language was used to analyze the GSVA enrichment scores of immune cells in the IPF samples and control samples. As shown in Fig. 6D, a total of 19 immune cell enrichment scores were significantly different between IPF and control groups ($P < 0.05$). GSVA enrichment score of DCs, iDCs, macrophages, Mast cells and T cell co-stimulation in the IPF group was significantly higher than that in the control group ($P < 0.05$, Fig. 6D), which was roughly consistent with the previous results of CIBERSORTx results (Fig. 6C). All the results confirmed that different immune environments might be related to IPF.

Analysis of differentially expressed IRGs

In order to further clarify the role of immunity in the occurrence and development of IPF, 4083 DEGs obtained by differential expression analysis of IPF and control groups were crossed with IRGs (obtained from the ImmPort database), and 361 differentially IRGs were screened in total. The specific gene names were shown in Supplementary Table 6, and the Venn diagram between IRGs and DEGs was shown in Fig. 7A. The PPI of differentially expressed IRGs was further constructed using STRING (https://cn.string-db.org/) with combined scores > 0.9 (Supplementary Fig. 3). The Cytoscape was used to analyze the hub nodes in the PPI network diagram, and the top10 hub nodes (HRAS, MAPK3, FYN, JUN, RHOA, NFκB1, AKT1, B2M, FOS, IL6) were obtained by sorting the nodes according to their degree. The interaction network of top10 hub nodes was shown in Fig. 7B. The darker the color in the diagram, the higher the degree of the nodes. In addition, the MCODE plug-in of Cytoscape was used to analyze the core subnetworks in the PPI network, and a total of three core subnetworks, Cluster 1 (Score 11, Fig. 7C), Cluster 2 (Score 5.6, Fig. 7D) and Cluster 3 (Score 5.333, Fig. 7E), were obtained.

Fig. 6. Differences in immune cells between IPF and control groups (**A**) Distribution of 22 immune cells in each sample. (**B**) Heatmap of the percentages of 22 immune cells. (**C**) Statistical boxplots of the differences between the 22 immune cells in the IPF and control groups. (**D**) Boxplots of immune cell scores in IPF and control groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Validation of expression of DNA methylation and immune-related genes in a bleomycin-induced pulmonary fibrosis mouse model and an A549 EMT model

To validate the results of the bioinformatics analysis, we conducted an in vivo animal experiment. The BLM mouse model is the most established preclinical model for pulmonary fibrosis. We successfully constructed a
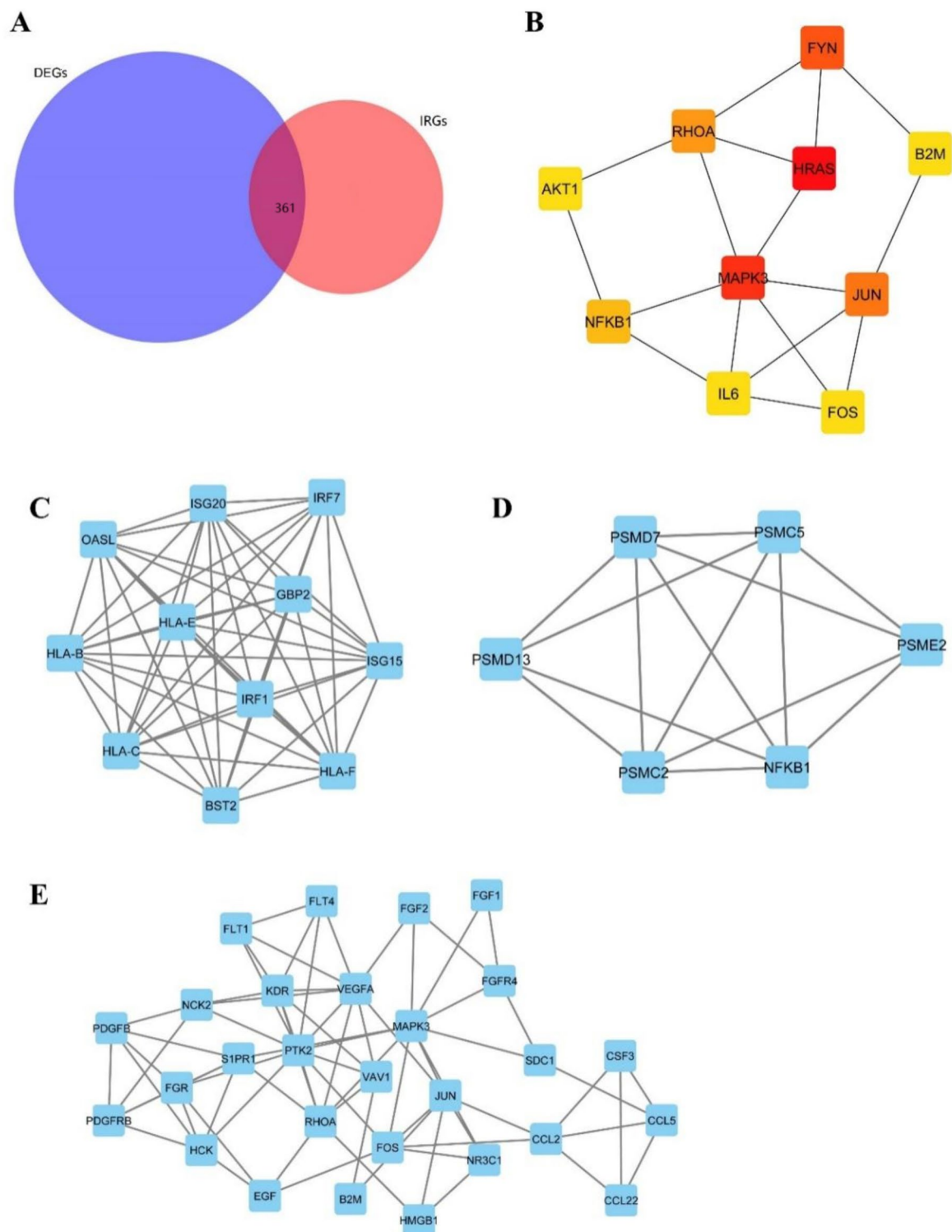
Fig. 7. Analysis of differential immune-related genes (IRGs) in the control and IPF groups (**A**) The intersection of DEGs in IPF and control samples and IRGs. (**B**) Interaction network diagram of the Top10 hub nodes in the PPI map of differentially expressed IRGs. (**C**) Core subnetwork of PPI network, Cluster 1 (Score 11). (**D**) Core subnetwork of PPI network, Cluster 2 (Score 5.6). (**E**) Core subnetwork of PPI network, Cluster 3 (Score 5.333). *DEG* differentially expressed genes, *IRGs* immune-related genes.

bleomycin-induced pulmonary fibrosis mouse model for in vivo experiments. The findings from HE and Masson staining showed that compared with the control group, the bleomycin group exhibited significantly increased lung tissue inflammation and collagen deposition (Fig 8A,B). The survival rates were lower in the BLM group compared to the normal group (Fig. 8C), and the lung coefficient was significantly higher in the BLM group than in the normal group (Fig. 8D). Penh was used as a pulmonary function test as it represents a good measure of airway resistance. Compare with the control group, Penh in the model group significantly increased, while the value of Vt decreased significantly at the same week (Fig. 8E,F). All the above data indicated the Successful construction of the BLM model.

The methylation of DNA affects mRNA expression. In most cases, the promoter area becomes hypermethylated, reducing gene expression, or the intra-gene region becomes hypomethylated, increasing expression. RT-PCR was used to detect the mRNA expression of methylation-related genes, the expression of ADAMTS16 and SLC22A3 showed the same trend as the above in our study (Fig. 8G). Then using National Center for Biotechnology Information (NCBI) site to find the ADAMTS16 and SLC22A3 promoter sequence of upstream of 2000 bp. Methprimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) was adopted to find the distribution of CpG island in promoter of ADAMTS16 and SLC22A3, we found tat there were CpG inslands in the prompter regin (Fig. 8H,I). The CpG islands methylation in the promoter regions of ADAMTS16 (Fig. 8H) and SLC22A3 (Fig. 8I) may altered gene expression in IPF.

The expression of key immune regulatory genes in BLM animal samples was confirmed. The qRT-PCR validation results in BLM samples aligned with the bioinformatics findings, indicating the precision and significance of the bioinformatics analysis. Consistent with both the bioinformatics analysis and qRT-PCR results, the mRNA levels of seven genes (HRAS, MAPK3, FYN, JUN, NF-κB, B2M, and IL−6) were down regulated in the IPF mice model (Fig. 8J). Furthermore, A549 cells were exposed to TGF-β to induce an epithelial-to-mesenchymal transition (EMT) in vitro. Post-TGF-β treatment, there was a notable increase in the expression of mesenchymal markers α-SMA and Vimentin (Fig. 8K,M), original blots are presented in Supplementary Fig. 4. In the A549 EMT model, the mRNA levels of HRAS, MAPK3, FYN, JUN, NF-κB, B2M, and IL−6 were decreased compared to baseline conditions, consistent with the trends observed in the bioinformatics analysis and in vivo animal models (Fig. 8L).

Discussion

IPF is a chronic progressive interstitial lung disease characterized by complex and multifaceted pathogenesis and progression. Currently, there is no effective treatment, and the median survival time of patients is only 3 years[2]. Its incidence increases with age, and with the progress of population aging, IPF will become one of the increasingly serious public health problems.

In this study, we downloaded GSE datasets from the GEO database to identify DEGs and methylation sites between IPF and normal lung tissues. Subsequently, we performed GO and KEGG analyses to explore the biological functions of DEGs and differential methylation in IPF. Our findings demonstrate that there is a significant consistency of DEGs analyzed using two different R packages. Despite fundamental methodological differences, with limma employing a linear modeling framework on variance-stabilized log-CPM values through voom transformation and edgeR utilizing a negative binomial generalized linear model on raw counts with TMM normalization, the overlap of differentially expressed transcripts was substantial. Specifically, 75.3% of transcripts were shared, with 4012 out of 5328 being limma-derived and 3712 out of 4876 being edgeR-derived. This methodological agreement extended to functional annotation analyses, where there was over 90% concordance in Gene Ontology (GO) terms related to extracellular matrix remodeling and kinase activity (Fig. 2E). Furthermore, the analysis of differentially expressed KEGG pathways demonstrated complete overlap in pathway signatures, with all pathways identified by edgeR being encompassed within limma-derived results (Fig. 2G). These findings suggest that when analyzing high-quality datasets using different R packages, the identification of differential genes and gene functions was highly consistent, thus validating the reliability of our research outcomes.

Meanwhile, there were 4933 differentially methylated sites in IPF, which were significantly enriched in similar pathways. And a total of eight DEGs that were negatively regulated by methylation were screened. In addition, the immune microenvironment in IPF was significantly changed, and 361differentially expressed IRGs were obtained by screening DEGs. The differentially expressed IRGs were used to construct the PPI network, and 10 key node genes and 3 core subnetworks were obtained. This study suggested that the pathogenesis of IPF is closely related to DNA methylation and immune dysregulation, and the selected 8 DEGs regulated by methylation and 10 immune-related key node genes may be the key to studying IPF.

The multi-layered pathway analysis delineates a cohesive molecular architecture underlying IPF pathogenesis, wherein the dysregulated extracellular matrix (ECM) and transductive signaling converge to perpetuate fibrotic progression. The pronounced enrichment of "extracellular matrix organization" (GO-BP) and "collagen-containing ECM" (GO-CC) directly correlates with the pathological hallmark of IPF—excessive collagen deposition. In addition, the DEGs and differentially methylated sites had similar KEGG enrichment results, including Rap1 signaling pathway, Axon guidance and Focal adhesion pathway. Rap1 is a novel positive regulator of NO release and endothelial function, which has been shown to prevent excessive cytokine receptor signaling and pro-inflammatory NF-κB activation[22]. Rap1 participates in the regulation of diverse biological processes, including cell proliferation, differentiation, and apoptosis. In pulmonary fibrosis, the Rap1 signaling pathway is closely associated with the proliferation of lung fibroblasts and the advancement of fibrosis. Studies have discovered that suppressing the Rap1 signaling pathway can reduce the proliferation of lung fibroblasts, thereby decelerating the progress of fibrosis[6]. By regulating the Rap1 signaling pathway, for instance, by using drugs such as Asiaticoside, new approaches for the treatment of pulmonary fibrosis might be offered. Axon guidance molecules are thought to play multifaceted roles in regulating tissue inflammation and metabolic disorders[23],
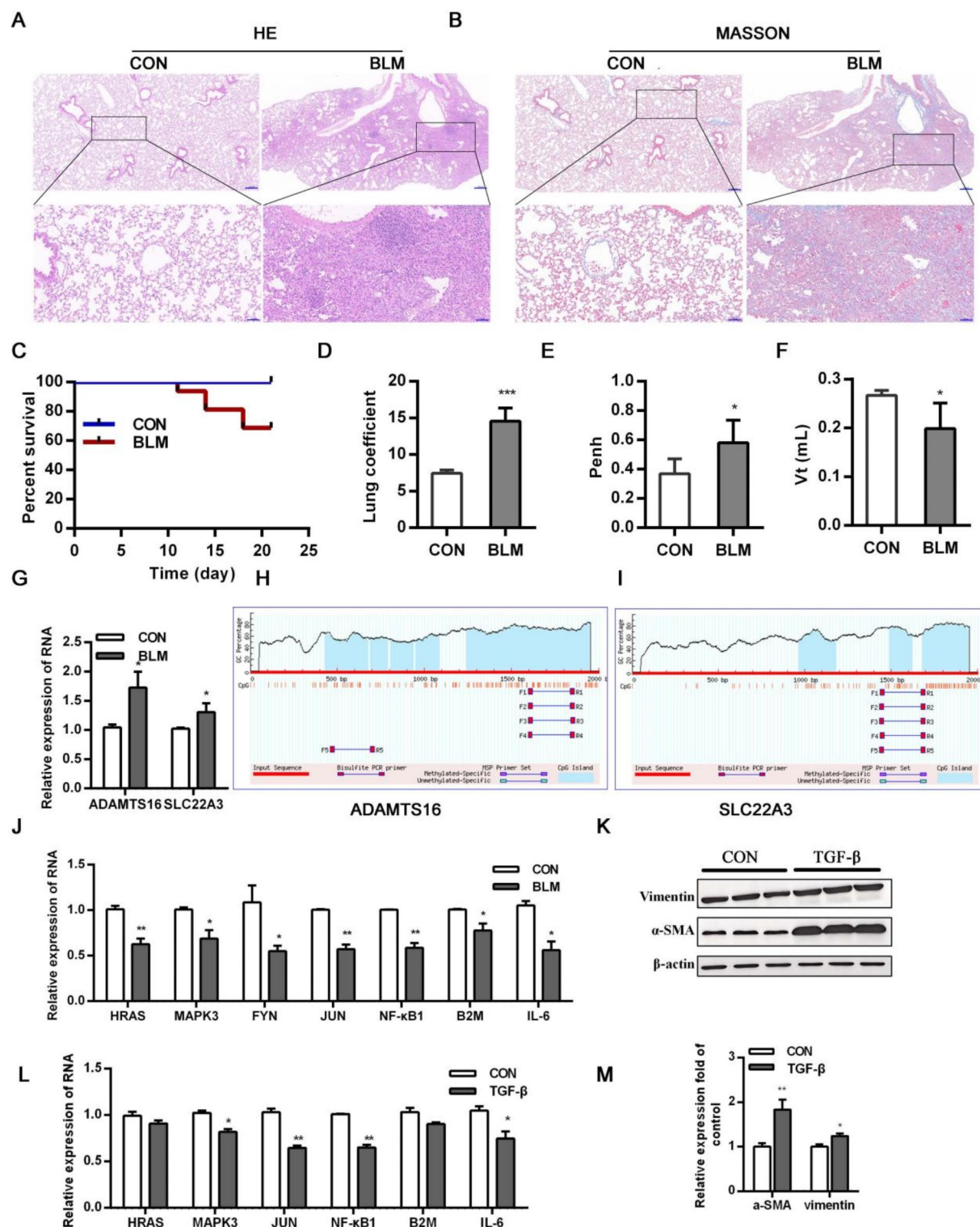
Fig. 8. Animal and cell model validation. (**A**) HE-stained section. (**B**) Masson trichrome-stained section. (**C**) Survival curves of mice. (**D**) Lung coefficient of lung tissues in mice. Comparison of lung function indexes Penh (**E**) and Vt (**F**) of mice in each group. (**G**) ADAMTS16 and SLC22A3 expression level was examined by RT-qPCR. Diagram of CpG sites and CpG island in the promoter region of ADAMTS16 (**H**) and SLC22A3 (**I**). (**J**) Differential immune-related genes (IRGs) in the control and IPF groups. (**K**) The protein expression of α-SMA and Vimentin in A549 cells were measured by western blotting. (**L**) Statistical analysis of relative expression levels of proteins in K. (**M**) The mRNA expression of IRGs in A549 cells were assessed by qRT-PCR.* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. *IPF* Idiopathic pulmonary fibrosis.

and developmental differences in focal adhesion kinase expression modulate pulmonary endothelial barrier function in response to inflammation[24]. Rap1, Axon guidance and Focal adhesion pathways were proved to be key pathways of IPF, and were also related to immune inflammatory response, suggesting that there might be an immune imbalance in IPF. The KEGG analysis of DEGs identified by different R packages consistently showed significant enrichment in the PI3K-Akt and MAPK signaling pathways. The PI3K-Akt pathway regulates mTORC1 to promote the expression of EMT-related transcription factors and the secretion of pro-fibrotic mediators[25]. Concurrently, MAPK signaling plays a pivotal role in IPF pathogenesis by driving fibroblast migration and collagen synthesis. Researchers found that the activation of the Ras/Raf/MEK/ERK signaling cascade can indirectly activate the PI3K/Akt/mTOR signaling pathway. Furthermore, crosstalk between the PI3K-Akt and MAPK signaling pathways amplifies fibroblast migration and proliferation, culminating in excessive collagen production[26]. Therapeutic strategies simultaneously targeting both the PI3K-Akt and MAPK signaling pathways may offer new avenues for IPF treatment, potentially yielding more effective interventions. The combined therapeutic strategy targeting the PI3K-Akt and MAPK pathways holds promise as a potential breakthrough for IPF.

It is generally believed that the process of IPF is affected by both genetic and environmental factors[27]. Environmental exposure and genetic variation may lead to changes in the expression of key genes involved in IPF by changing epigenetic marks, thus promoting the progression of IPF[28]. Epigenetics refers to the heritable and reversible changes in gene expression that do not change in DNA sequence[29]. Among them, DNA methylation occupies a key position, and abnormal methylation of promoter CpG can change chromosome structure and inhibit gene expression, which has been confirmed to be closely related to a variety of biological processes[30]. More and more studies have shown that abnormal DNA methylation patterns could promote IPF[31], and the study of DNA methylation can create a new research direction for the diagnosis and targeted therapy of IPF. Therefore, we analyzed IPF gene expression profile and methylation data simultaneously, and a total of 8 DEGs (PDCD1LG2, TTC34, CNNM1, ADAMTS16, MKX, SLC22A3, C1orf53 and CP) negatively regulated by methylation were screened, among which PDCD1LG2, CNNM1, ADAMTS16 and SLC22A3 were common oncogenes. Expression analysis revealed two genes, including ADAMTS16 and SLC22A3, which showed a consistent expression trend as that detected by prior bioinformatics analysis (Fig. 9I). ADAMTS16, a matrisome-associated protease, targets fibronectin to inhibit ECM assembly[31], further suggesting that ECM organisation is dysregulated in IPF tissue. ADAMTS16 was validated a novel putative extracellular markers of NSCLC, for discrimination of malignant and non-malignant lung tissue[32]. SLC22A3 expression was previously associated with progression of several cancer[33]. IPF patients are generally considered to have a higher risk of lung cancer than the general population[34]. The incidence of lung cancer in IPF patients ranges from 3 to 22%, and in some cases exceeds 50%[35], possibly due to the high number of oncogenes involved in IPF DEGs.

The theory that fibroblasts participate in IPF has been widely accepted. Due to the poor efficacy of anti-inflammatory drugs and immunosuppressants in the treatment of IPF, the inflammatory immune mechanism is considered to be not the direct mechanism leading to IPF. However, some researchers have found that immune cells affect the formation and development of IPF by interacting with fibroblasts[7]. Type 2 helper T lymphocytes (Th2) in the lungs of patients with IPF produce more cytokines (such as IL−4, IL−5 and IL13), among which IL−4 and IL−13 can promote macrophage activation and induce the expression of Fra−2 in macrophage[36], which is considered to be one of the pathogenic factors of human pulmonary fibrosis[37]. Similarly, inflammatory factors can stimulate the proliferation of fibroblasts and collagen synthesis, thereby accelerating the process of IPF[38]. The effect of immune cells and lung fibroblasts is bidirectional, activated fibroblasts can secrete proinflammatory factors and chemokines to stimulate immune cells and mediate immune responses[39]. In this study, bioinformatics analysis showed that Stromal score, immune score and ESTIMATE score were significantly decreased in IPF samples, which further confirmed the existence of immune imbalance in IPF. Subsequent analytical assessments employing multiple linear regression coupled with forest plot graphical representation demonstrated that, following adjustments for age, sex, and smoking history, both diagnostic classification (IPF versus normal) and sex emerged as statistically independent predictors of immune scores ($p < 0.05$). The immune scores of IPF patients were significantly lower than those of the normal group, suggesting that intervention strategies targeting etiology-specific immune suppression pathways may have higher potential therapeutic efficacy. However, we acknowledge that other unmeasured confounders (e.g., comorbidities, medication use) may still influence the results. Future studies with larger sample sizes and more comprehensive clinical data are needed to further validate these findings and explore the impact of additional confounders on immune dysregulation in IPF.

Similarly, 12 types of immune cells in IPF were significantly different, and the GASA enrichment scores of 19 types of immune cells were also different, and the different immune cells were mainly concentrated in macrophages, neutrophils, dendritic cells and mast cells. Macrophages, which exist in all tissues with highly plastic and diverse, are the main effector cells mediating innate immunity and are involved in multiple processes in fibrosis. Neutrophils are an important part of the body's immune response to infection and can effectively fight against invading pathogens through phagocytosis and degranulation, but their powerful virulence mediators may also cause extensive tissue damage[40]. The uncoordinated neutrophils function of the body is considered to be a potential factor to promote the formation of IPF[41]. Dendritic cells, as antigen-presenting cells, are the bridge between innate immunity and adaptive immunity, and play a role in immune surveillance in the lung epithelium and mesenchyme[42]. These results further verified that the immune microenvironment involved in the above immune cells was closely related to IPF. The differentially expressed IRGs were further screened out, and PPI analysis was performed to obtain 10 key node genes (HRAS, MAPK3, FYN, JUN, RHOA, NFKB1, AKT1, B2M, FOS, IL6) and 3 core subnetworks. Moreover, the expression levels of 10 key node were validated for qPCR. The expression trend of HRAS, MAPK3, JUN, NFKB1, B2M, and IL6 was basically consistent with the results of bioinformatic analysis. HRAS, MAPK3, and AKT1 regulate cell proliferation, survival, and fibrosis, all of which are implicated in the pathogenesis of IPF. JUN is a component of the transcription factor AP1

family, which coordinates the transcriptional regulation of numerous genes critical for various cellular processes, including differentiation, proliferation, and apoptosis[37]. In IPF, upregulation of JUN is associated with fibroblast activation and proliferation, and its overexpression is linked to increased collagen synthesis, thereby promoting pulmonary tissue fibrosis. FYN, a tyrosine kinase, plays a critical role in various cellular signaling pathways and immune responses. Its activity may influence the immune microenvironment by modulating macrophage and lymphocyte activity, potentially impacting IPF progression. B2M (β2-microglobulin), a crucial component of the immune system, has been found elevated in the serum and lung tissues of IPF patients, potentially correlating with disease severity. Altered B2M levels may also contribute to the imbalance between apoptosis and proliferation during fibrotic processes[43]. These genes play distinct and significant roles in the initiation and progression of IPF, offering both new insights into IPF biology and potential therapeutic targets. Future research may focus on validating the therapeutic potential of these hub genes, including in vitro and in vivo experiments, as well as the development of targeted therapies.

While the bleomycin-induced mouse model employed in this study effectively replicates key pathological hallmarks such as pulmonary fibrosis and functional impairment, it may not fully capture the chronic and progressive nature of IPF. Human IPF exhibits significant heterogeneity in lung tissue architecture and complex intercellular interactions, which are often incompletely reproduced in many mice models. Spatial transcriptomic sequencing has revealed substantial interspecies differences in immune responses and tissue repair mechanisms between mice and humans[44]. Furthermore, the regenerative capacity of mice differs from that of humans, which may influence disease progression and response. The use of the bleomycin-induced mouse model for qPCR validation in this study has certain limitations. Although this study performed mechanistic validation through in vivo and in vitro experimental models, it should be noted that the current model systems cannot fully replicate the complete pathological features of idiopathic pulmonary fibrosis (IPF), highlighting the limitations of the models. To further validate the research findings, follow-up studies plan to utilize human tissue samples for multi-dimensional validation experiments, aiming to deeply investigate the spatial heterogeneity of key genes and their regulatory signaling pathways identified in this study within the fibrotic lesion microenvironment.

Conclusion

In conclusion, this study provided a comprehensive bioinformatics method to screen DEGs and differentially expressed methylation sites in IPF, and identified eight genes that were significantly negatively regulated by methylation. Moreover, the analysis showed that IPF was closely related to the immune environment, and 10 key node genes and 3 core subnetworks might be the key to immune regulation in IPF. Through the development of this study, we hope to provide precise direction and a strong theoretical basis for subsequent research on the molecular mechanism of IPF, and provide clues for the discovery of new diagnostic biomarkers and treatment strategies.

Data availability

The datasets generated during the current study are available in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE173355 and GSE173356. All analysis scripts, including those for RNA-seq data preprocessing, differential expression analysis, and pathway enrichment, have been uploaded to GitHub and are available at the following link: https://github.com/wujuan948725/RNAseq-methylation-analysis-2025.

References

1. Benegas Urteaga, M., Ramírez Ruz, J. & Sánchez González, M. Idiopathic pulmonary fibrosis. *Radiologia* **3**, 227–239. https://doi.org/10.1016/j.rxeng.2022.10.009 (2022).
2. Wakwaya, Y. & Brown, K. K. Idiopathic pulmonary fibrosis: epidemiology, diagnosis and outcomes. *Am. J. Med. Sci.* **357**, 359–369. https://doi.org/10.1016/j.amjms.2019.02.013 (2019).
3. Meng, H. et al. DNA methylation, its mediators and genome integrity. *Int. J. Biol. Sci.* **11**, 604–617. https://doi.org/10.7150/ijbs.11218 (2015).
4. Mattei, A. L., Bailly, N. & Meissner, A. DNA methylation: a historical perspective. *Trends Genet.* **38**, 676–707. https://doi.org/10.1016/j.tig.2022.03.010 (2022).
5. Sanders, Y. Y. et al. Altered DNA methylation profile in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **186**, 525–535. https://doi.org/10.1164/rccm.201201 (2012).
6. Luo, J. et al. Asiaticoside might attenuate bleomycin-induced pulmonary fibrosis by activating cAMP and Rap1 signalling pathway assisted by A2AR. *J. Cell. Mol. Med.* **24**, 8248–8261. https://doi.org/10.1111/jcmm.15505 (2020).
7. Heukels, P., Moor, C. C., von der Thüsen, J. H., Wijsenbeek, M. S. & Kool, M. Inflammation and immunity in IPF pathogenesis and treatment. *Respir. Med.* **147**, 79–91. https://doi.org/10.1016/j.rmed.2018.12.015 (2019).
8. Shenderov, K., Collins, S. L., Powell, J. D. & Horton, M. R. Immune dysregulation as a driver of idiopathic pulmonary fibrosis. *J. Clin. Investig.* **131**, e143226. https://doi.org/10.1172/JCI143226 (2021).
9. 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).
10. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287. https://doi.org/10.1089/omi.2011.0118 (2012).
11. Phipson, B., Maksimovic, J & Oshlack, A. MissMethyl: an R package for analyzing data from Illumina's human Methylation450 platform. *Bioinformatics* **32**, 286–288. (2016).
12. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7. (2013).
13. Silva, T. C. et al. ELMER V.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* **35**, 1974–1977. https://doi.org/10.1093/bioinformatics/bty902 (2019).

14. Meng, Z. et al. Using ESTIMATE algorithm to establish an 8-mRNA signature prognosis prediction system and identify immunocyte infiltration-related genes in pancreatic adenocarcinoma. *Aging* **12**, 5048–5070. https://doi.org/10.18632/aging.102931 (2020).
15. 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).
16. Szklarczyk, D. et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568. https://doi.org/10.1093/nar/gkq973 (2011).
17. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432. https://doi.org/10.1093/bioinformatics/btq675 (2011).
18. Liu, T., De Los Santos, F. G. & Phan, S. H. The bleomycin model of pulmonary fibrosis. *Methods Mol. Biol.* **1627**, 27–42. (2017).
19. Qi, J. Y. et al. Fibroblast growth factor 21 alleviates idiopathic pulmonary fibrosis by inhibiting PI3K-AKT-mTOR signaling and stimulating autophagy. *Int. J. Biol. Macromol.* **273**, 132896. https://doi.org/10.1016/j.ijbiomac (2024).
20. Wang, M. et al. Association between MAPK and PI3K/Akt signaling pathway-related gene polymorphisms and migraine. *Mol. Genet. Genom. Med.* **12**, e2503. https://doi.org/10.1002/mgg3.2503 (2024).
21. Kanehisa, M. & Goto, S. K. E. G. G. Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
22. Singh, B. et al. Endothelial Rap1 (Ras-Association proximate 1) restricts inflammatory signaling to protect from the progression of atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **41**, 638–650. https://doi.org/10.1161/ATVBAHA.120.315401 (2021).
23. Nakanishi, Y., Kang, S. & Kumanogoh, A. Axon guidance molecules in immunometabolic diseases. *Inflamm. Regen.* **42**, 5. (2022).
24. Ying, L., Alvira, C. M. & Cornfield, D. N. Developmental differences in focal adhesion kinase expression modulate pulmonary endothelial barrier function in response to inflammation. American journal of physiology. *Lung Cell. Mol. Physiol.* **315**, L66–L77. https://doi.org/10.1152/ajplung.00363.2017 (2018).
25. Wang, Z. Y. et al. Targeting the PI3K/mTOR pathway in idiopathic pulmonary fibrosis: advances and therapeutic potential. *Bioorg. Med. Chem.* **115**, 0968–0896. https://doi.org/10.1016/j.bmc (2024).
26. Zhou, Y., Ling, T. & Shi, W. Current state of signaling pathways associated with the pathogenesis of idiopathic pulmonary fibrosis. *Respir. Res.* **25**, 245. https://doi.org/10.1186/s12931-024-02878-z (2024).
27. Chanda, D. et al. Developmental pathways in the pathogenesis of lung fibrosis. *Mol. Aspects Med.* **65**, 56–69. https://doi.org/10.1016/j.mam.2018.08.004 (2019).
28. Phan, T. H. G. et al. Emerging cellular and molecular determinants of idiopathic pulmonary fibrosis. *Cell. Mol. Life Sci. CMLS* 78, 2031– (2057). (2021).
29. Peixoto, P., Cartron, P. F., Serandour, A. A. & Hervouet, E. From 1957 to nowadays: A brief history of epigenetics. *Int. J. Mol. Sci.* **21**, 7571. (2020).
30. Law, P. P. & Holland, M. L. DNA methylation at the crossroads of gene and environment interactions. *Essays Biochem.* **63**, 717–726. https://doi.org/10.1042/EBC20190031 (2019).
31. Luo, Q. K., Zhang, H. & Li, L. Research advances on DNA methylation in idiopathic pulmonary fibrosis. *Adv. Exp. Med. Biol.* **1255**, 73–81. (2020).
32. Schnellmann, R. et al. A selective extracellular matrix proteomics approach identifies fibronectin proteolysis by A Disintegrin-like and metalloprotease domain with thrombospondin type 1 motifs (ADAMTS16) and its impact on spheroid morphogenesis. *Mol. Cell. Proteom.* **17**, 1410–1425. (2018).
33. Titmarsh, H. F. et al. Quantitative proteomics identifies tumour matrisome signatures in patients with non-small cell lung cancer. *Front. Oncol.* **13**, 1194515. https://doi.org/10.3389/fonc.2023.1194515 (2023).
34. Cervenkova, L. et al. Protein expression of ABCC2 and SLC22A3 associates with prognosis of pancreatic adenocarcinoma. *Sci. Rep.* **9**, 19782. https://doi.org/10.1038/s41598-019-56059-w (2019).
35. Kinoshita, T. & Goto, T. Molecular mechanisms of pulmonary fibrogenesis and its progression to lung cancer: A review. *Int. J. Mol. Sci.* **20**, 1461. https://doi.org/10.3390/ijms20061461 (2019).
36. Ucero, A. C. et al. Fra–2-expressing macrophages promote lung fibrosis in mice. *J. Clin. Investig.* **129**, 3293–3309. https://doi.org/10.1172/JCI125366 (2019).
37. Eferl, R. et al. Development of pulmonary fibrosis through a pathway involving the transcription factor Fra–2/AP–1. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10525–10530. (2008).
38. Ayed, K., Beji, S., Jameleddine, S. & R., & Idiopathic pulmonary fibrosis:pathophysiological data. *Tunis. Med.* **95**, 756–766 (2017).
39. Fernando, M. R., Giembycz, M. A. & McKay, D. M. Bidirectional crosstalk via IL–6, PGE2 and PGD2 between murine myofibroblasts and alternatively activated macrophages enhances anti-inflammatory phenotype in both cells. *Br. J. Pharmacol.* **173**, 899–912. https://doi.org/10.1111/bph.13409 (2016).
40. Liew, P. X. & Kubes, P. The neutrophil's role during health and disease. *Physiol. Rev.* **99**, 1223–1248. https://doi.org/10.1152/physrev.00012.2018 (2019).
41. Kolahian, S., Fernandez, I. E., Eickelberg, O. & Hartl, D. Immune mechanisms in pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **55**, 309–322. (2016).
42. Gardner, A., de Mingo Pulido, Á. & Ruffell, B. Dendritic cells and their role in immunotherapy. *Front. Immunol.* **11**, 924. (2020).
43. Prasse, A. et al. BAL cell gene expression is indicative of outcome and airway basal cell involvement in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 622–630. (2019).
44. Franzén, L. et al. Mapping spatially resolved transcriptomes in human and mouse pulmonary fibrosis. *Nat. Genet.* **56**, 1725–1736. (2024).

Author contributions

This study was conceived and designed by Xiaomei Kong. The experiments were finished by Juan Wu, Yangyang Wei, Hong Kang, Shuting Wei, and Xiao Yu. All data were analyzed by Juan Wu, Ting Xue, and Shuting Wei. The manuscript was written by Juan Wu and Yangyang Wei and revised by Xiaomei Kong.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement
The animal study was approved by the Animal Ethics Committee of Shanxi Medical University, No. SYDL2023023.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-97037-9.

Correspondence and requests for materials should be addressed to X.K.

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

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