scientific reports

Check for updates

OPEN

ITGAV, a specific biomarker associated with the pathogenesis of idiopathic pulmonary fibrosis

Zhangyang Bi[1], Xiaodong Wang[2], Xue Liu[2], Xue Zhu[2], Guodong Zang[2], Li Tian[2] & Wei Zhang[2]✉

Objective Idiopathic pulmonary fibrosis (IPF), which radiologically/pathologically manifests mainly as usual interstitial pneumonia (UIP), is easily confused with chronic hypersensitivity pneumonitis (CHP) and collagenous vascular disease in clinical settings, affecting the physician's diagnosis and treatment. Accurate identification of IPF from various diseases presenting as UIP is essential for effective diagnosis and therapy. Methods Gene expression data of CHP, IPF, and rheumatoid arthritis-UIP samples were downloaded from the GEO database, and specific biomarkers were identified to differentiate idiopathic UIP/IPF from secondary UIP. We compared gene expression of specific biomarkers between control, secondary UIP, and IPF groups. The mechanism of specific biomarkers in PF was explored by immunohistochemistry, quantitative polymerase chain reaction, immunofluorescence, and flow cytometry. Results We identified integrin alpha V (ITGAV) as a specific biomarker for distinguishing IPF from secondary UIP. We observed a gradual increase in ITGAV expression across the control, secondary UIP, and IPF groups. Animal studies indicated that the elevated expression of ITGAV in various immune cells, particularly in monocyte-derived macrophages, contributed to the development of PF. Conclusion ITGAV is a specific biomarker linked to the pathogenesis of IPF. The identification of ITGAV provides new perspectives for clinical diagnosis, mechanistic studies and new drug development in IPF.

Keywords Machine learning, ITGAV, IPF, Immune cells, UIP

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial pneumonia of unknown origin, which predominantly presents radiologically and pathologically as usual interstitial pneumonia (UIP). However, various conditions and diseases, including asbestosis, connective tissue diseases (CTD), chronic hypersensitivity pneumonitis (CHP), and medication toxicity, can also lead to the development of UIP. In terms of treatment options and natural progression, IPF differs significantly from other diseases that present as UIP. For instance, in contrast to IPF, CHP treatment focuses on modifying the environment or modifying the host immune response[1]. Environmental control[2], hormonal therapies and immunosuppressants have stabilized or improved the condition of CHP patients. Although immunosuppressive treatments are associated with a higher mortality rate in IPF, they can be beneficial for patients with CTD-interstitial lung disease (ILD) by effectively halting the progression of ILD lesions[3]. The current state of IPF treatment presents a more significant challenge than other secondary UIP disorders, which have more diverse treatment options and more reliable prognoses. Antifibrotic medications, such as nintedanib and pirfenidone, are the primary therapeutic agents available for IPF; however, their limited efficacy and numerous side effects often hinder patient adherence. In Asia-Pacific countries, the median survival following an IPF diagnosis ranges from three to five years[4].

Although experts have carefully differentiated the etiology, pathologic manifestations, therapeutic approaches, and natural course of diseases that exhibit UIP patterns, challenges remain in terms of clinical diagnostic accuracy. Clinically, IPF is particularly easy to confuse with CHP and CTD-ILD. Imaging for CHP or fibrotic HP (fHP) may reveal reticular shadows and honeycomb signs[5]. Due to the complexity and variability of the clinical presentation and natural course of the disease, the multiple irritants associated with CHP can be difficult for pulmonologists to identify. Some CTD-ILDs have an insidious onset and are not characterized by early immunologic features, particularly rheumatoid arthritis (RA)-associated UIP[6], which can be misdiagnosed as IPF in its early stages. Therefore, the incorporation of additional diagnostic tools facilitates the clarification

[1]Clinical Department of Integrated Traditional Chinese and Western Medicine, The First Clinical Medical College of Shandong, University of Traditional Chinese Medicine, Jinan, China. [2]Department of Respiratory and Critical Care Medicine, Affiliated Hospital of Shandong University of Chinese Medicine, Jinan, China. ✉email: huxizhijia@126.com

Dataset ID	Platform	IPF/UIP	CHP/FHP	Control
Train group				
GSE150910	GPL 24,676			
Transplant group		67	56	
Biopsy group		36	26	
Test group				
GSE124685	GPL 17,303	49		35
GSE184316	GPL 17,303	40	36	24

Table 1. Details for a chosen GEO dataset 1.

Dataset ID	Platform	RA-UIP	IPF-UIP	Non-UIP
GSE199152	GPL16791	3	20	4

Table 2. Details for a chosen GEO dataset 2.

of disease diagnosis, the smooth application of appropriate treatments, and the reasonable determination of fibrosis evolution and disease progression.

With the recent advancements in microarray technology within the field of bioinformatics, it is possible to discover diagnostic biomarkers to distinguish idiopathic UIP/IPF from secondary UIP. Definitive molecular diagnosis facilitates the early identification and prediction of IPF, allowing for the implementation of intervention strategies during the initial stages of the disease and enhancing survival rates. Additionally, it aids in distinguishing other secondary UIP, targeting various etiologies and increasing confidence in both diagnosis and treatment. Therefore, the purpose of this study was to precisely distinguish IPF from a wide range of diseases presenting as UIP, such as CHP and RA-UIP, and to investigate how particular biomarkers contribute to the pathophysiology of IPF. We identified key diagnostic genes for IPF and validated these biomarkers using machine learning algorithms. We validated IPF-specific genes through animal experiments, thereby establishing a foundation for future investigations into the pathogenesis of IPF and the development of new therapeutic agents.

Methods

Processing of microarray data

From the GEO database (https://www.ncbi.nlm.nih.gov/gds), we obtained datasets containing IPF, CHP, and CTD-ILD (Tables 1 and 2). We first identified the dataset GSE150910, which contains a large sample size of IPF and CHP, as the training group. To minimize batch differences and expand the number of samples, we chose the GSE124685 and GSE184316 datasets with the same sequencing platform as the test group. Since the training group was strongly influenced by the sampling method, we separated the samples into two distinct cohorts, both the biopsy and transplant groups[7]. The DESeq2 package in R software (http://www.bioconductor.org/) was used to adjust for sex, race, age, gender, and smoking history separately for both cohorts[8]. Background correction and normalization of the two datasets from the test group were performed using the limma package[9]. The SVA package was used to eliminate batch effects following the merger of the two datasets[10].

Identification and function of candidate diagnostic genes

Differentially expressed genes (DEGs) were identified using DESeq2, and genes were classified as DEGs if $|\log$ fold change (FC)$| > 1$ and corrected $P < 0.05$. The weighted gene co-expression network analysis (WGCNA) method is frequently used to examine the global co-expression of all samples in order to find gene clusters that have a strong correlation with disease[11]. To exclude outlier samples, hierarchical cluster analysis was carried out using the "hclust" function in the "stat" R package. The network's topology was examined using the "pickSoftThreshold" function in the "WGCNA" R package, and the appropriate soft threshold β was chosen for network construction. The "blockwiseModules" function was used for network construction and module detection, focusing on the two modules with the strongest correlation with disease diagnosis. A correlation heatmap of the gene modules and traits was drawn, the correlations and p values of the modules and genes were calculated, and the core genes of the modules were summarized. A Venn diagram was used to identify the genes that intersected in the differential gene expression (DGE) analysis and WGCNA as potential important genes for illness diagnosis (https://bioinformatics.psb.ugent.be/webtools/Venn/). We used Metescape (http://metascape.org/gp/index.html) to conduct Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses of the screened key genes in order to look into their roles[12].

Screening for specific biomarkers for the diagnosis of IPF

We applied three machine learning methods to improve the forecast accuracy. least absolute shrinkage and selection operator (LASSO) regression is a linear model regularization method that facilitates feature selection for the model. Support vector machine-recursive feature elimination (SVM-REF) is an effective and reliable supervised machine learning technique for resolving binary classification issues. Random forest is a simple and

intuitive supervised learning method for improving prediction accuracy by integrating multiple decision trees. Logistic regression was implemented by the glmnet package, support vector machines were constructed by the e1071 package, and random forests were fitted by the random forest package to screen the potential diagnostic biomarkers identifying IPF and CHP. Receiver operating characteristic (ROC) curves were plotted to evaluate the predictive power of the above biomarkers in both the training and test groups. An area under the ROC curve (AUC) greater than 70% indicates high diagnostic efficacy. A dataset that included both IPF-UIP and RA-UIP samples was presented. A ROC curve and box-and-line plots were created in order to further identify potential diagnostic biomarkers that distinguish IPF-UIP from RA-UIP. The ROC curve for specific diagnostic biomarker was calculated using the presence or absence of IPF as a status variable.

Analysis of immune infiltration

From CIBERSORT (http://cibersort.stanford.edu/), we obtained the 22 immune cells (LM22) matrix, which is a collection of genetic markers with interpretative characteristics. 22 immune cells that infiltrated the tissue were quantified using the CIBERSORT method. The 22 invading immune cells were correlated and shown using the R software's corrplot package. Correlation heatmaps of selected diagnostic biomarkers with infiltrating immune cell levels were plotted using the ggthemes package.

Analysis of specific biomarker expression in control, secondary UIP, and IPF groups

We used the GSE184316 and GSE199152 datasets to compare the expression differences of specific biomarkers among the control group, the secondary UIP group and IPF group.

Animal studies

A total of 12 SPF-grade healthy male Sprague–Dawley (SD) rats weighing 200 ± 20 g were obtained from Shandong Jinan Peonyue Laboratory Animal Breeding Co. We confirmed that all experiments were performed in accordance with the relevant guidelines and regulations of the Institutional Animal Care and Use Committee. This study was approved by the Experimental Animal Ethics Committee of the Affiliated Hospital of Shandong University of Traditional Chinese Medicine under the approval number SDSZYYAWE20240930002. This study followed 《The Animal Research: Reporting of in vivo Experiments [ARRIVE] guidelines》(The ARRIVE guidelines 2.0 | ARRIVE Guidelines). Using the random number table approach, twelve healthy male SD rats were split into two groups: the blank group and the model group. Bleomycin (Invitrogen, R25001) was prepared as a 5 mg/ml solution in saline and set aside. The rats were anesthetized with isoflurane inhalation, and bleomycin solution was injected via a 1 ml/kg tracheal drip to establish a rat model of pulmonary fibrosis (PF). After 28 days, rats were anesthetized with 6 ml/kg of 2.5% tribromoethanol solution, and alveolar lavage fluid and lung tissues were obtained to carry out the relevant tests.

Histopathology

Hematoxylin and eosin (HE) staining and Masson staining were used to observe lung tissue injury and fibrosis. Rat lung tissue was fixed with 4% paraformaldehyde for 48 h, dehydrated, embedded in paraffin and sectioned. Some sections were subjected to HE staining according to the procedure of HE staining kit and sealed with neutral gum. Other sections were stained with Masson staining according to the procedure of Masson Trichrome Staining Kit. The sections were examined using a microscope and analyzed by image acquisition.

Immunohistochemical (IHC) and Immunofluorescence (IF) analysis

The expression of ITGAV in rat lung tissues was determined by immunohistochemistry, and the average optical density (AOD) value of positive particles staining in the sections was calculated using Image-Pro Plus software. The co-expression of CD14 and ITGAV in rat lung tissues was determined by IF, and the correlation between CD14 and ITGAV was obtained by calculating the Pearson's coefficient using Image J software.

Quantitative real-time PCR (RT-qPCR)

Total RNA was extracted by RNA extraction kit and its content and purity were determined. Following the reverse transcription kit's instructions, the total RNA was then reverse transcribed into cDNA (reaction conditions: 25 °C, 5 min; 42 °C, 30 min; 85 °C, 5 s). Next, RT-qPCR was used to determine the ITGAV mRNA expression level (reaction conditions: first step: 95 ℃, 30 s, 1 cycle; second step: 95 ℃, 15 s; 60 ℃, 38 s, 40 cycles), and GAPDH was used as internal reference. The amplification and cleavage curves were plotted, and the relative expression of target gene mRNA was calculated by the 2-ΔΔCt method. All PCR primers were designed and synthesized by Wuhan Xavier Biotechnology Co. The primer sequences are shown in Table 3.

Gene	Primer sequence
ITGAV	F: CTGAGCAAGGAGGAAGAAATCA R: CAGGGAATAGGAATGGTTCTGGTT
GAPDH	F: CTGGAGAAACCTGCCAAGTATG R: GGTGGAAGAATGGGAGTTGCT

Table 3. Gene names and primer sequences.

Flow cytometry

Alveolar lavage fluid was centrifuged at 4500 rpm for 2 min at 4 °C and the supernatant was discarded. Add 100 µl of PBS to resuspend the cells. Add antibodies (CD45, CD11b, CD51) and incubate at room temperature for 20 min. add 1–2 mI PBS and centrifuge at 1000 rpm for 5 min, discard supernatant. Add 1 mL Foxp3 Fixation/ Breakage Working Solution to each tube and pulse vortex. incubate at 2–8 °C or room temperature for 40 min, protected from light. Add 2 mL of 1X Membrane Breaking Solution to each tube and centrifuge the samples at 500xg for 5 min at room temperature and discard the supernatant. Resuspend the precipitate in the remaining volume of 1X membrane-breaking solution, typically 100 L. Add antibody CD68 and incubate for 20 min at room temperature. add 2 mL of 1X membrane-breaking solution per tube and centrifuge samples at 500xg for 5 min at room temperature. Discard supernatant. Resuspend stained cells in appropriate volume of flow cytometry staining solution and analyze samples by flow cytometry.

Data analysis

We used GraphPad Prism 9 for the statistical analysis. We used Student's t test to compare the different groups, and $P < 0.05$ was considered to indicate a statistically significant difference.

ITGAV is a specific diagnostic biomarker for IPF

The research process

The process of specific diagnostic biomarker identification is shown in Fig. 1. We identified key diagnostic genes between the IPF and CHP groups in the training group by DGE analysis and WGCNA. Biological

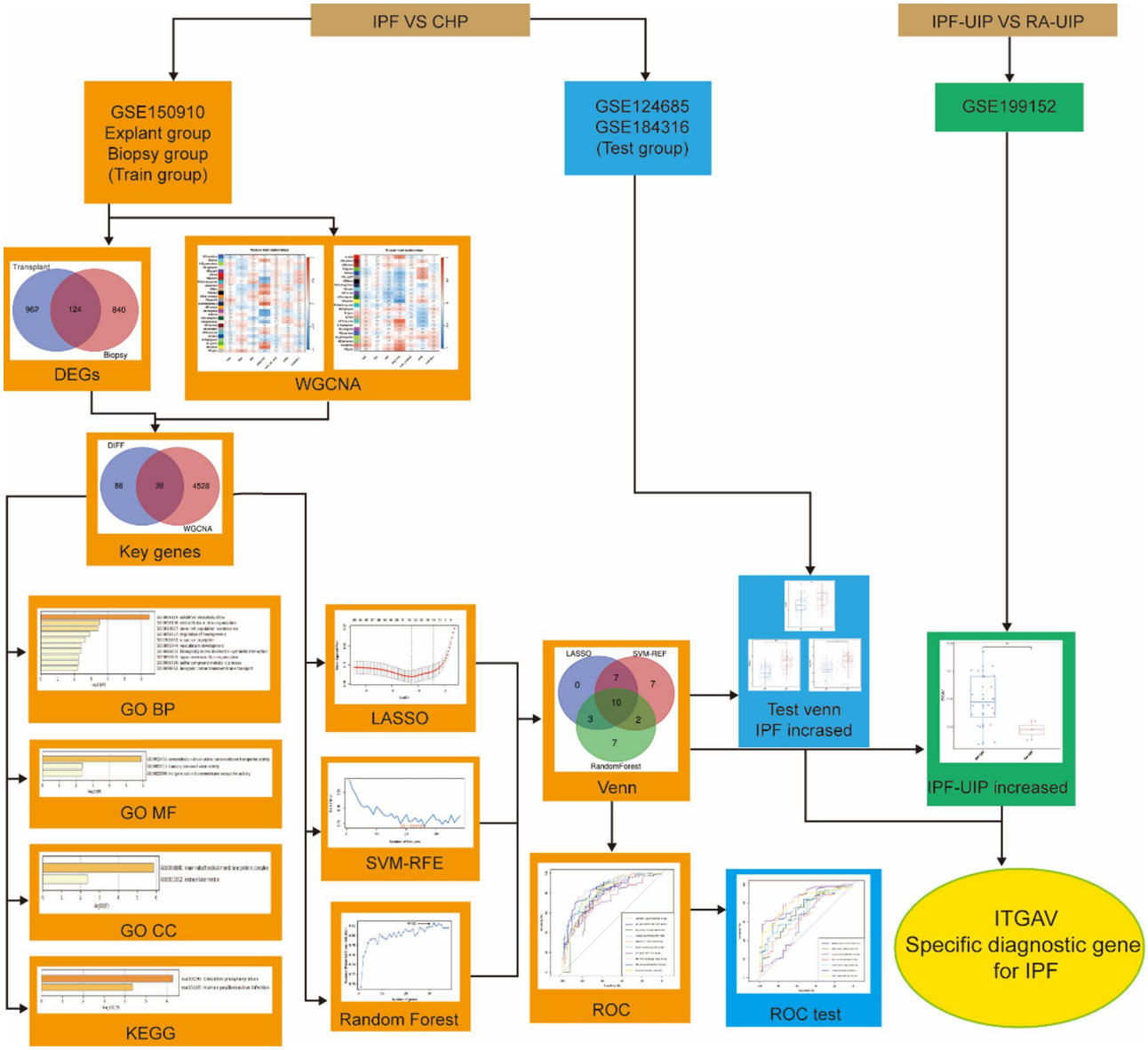

Fig. 1. This study's analysis flow diagram.

function analysis of key diagnostic genes was performed via GO and KEGG analyses. Diagnostic biomarkers of overlapping IPF and CHP were found using LASSO, SVM-RFE, and random forest screening. The predictive power of the biomarkers was evaluated using ROC curves. ROC curves and box plots were utilized in the test group to screen biomarkers that distinguish IPF from CHP. We also used datasets containing IPF- UIP and RA-UIP to further identify biomarkers that distinguish IPF from RA-UIP. We finally obtained a specific diagnostic biomarker to distinguish IPF from secondary UIP.

Identification of candidate diagnostic genes

We first used a dataset (GSE150910) to identify diagnostic key genes between IPF patients and CHP patients. Since the different sampling methods in this dataset have a large impact on gene expression, we divided this dataset into two subgroups, the transplant group and biopsy group, for DGE analysis and WGCNA, respectively. The transplant group included 67 IPF and 56 CHP samples, and the biopsy group included 36 IPF and 26 CHP samples. In the transplant group, we obtained 1086 DEGs, with 762 showing significant downregulation and 324 showing substantial upregulation (Fig. 2A). In the biopsy group, we obtained 964 DEGs (Fig. 2B), with 748 showing significant downregulation and 216 showing substantial upregulation. After taking the intersection, 124 common DEGs were obtained (Fig. 2C).

The soft threshold β for WGCNA in the transplant and biopsy groups was 10. WGCNA identified gene modules in the transplant and biopsy groups. Each module is indicated by a different color. Clustering dendrograms showed that the original and merged modules had similar expression patterns (Fig. 3A, E). Heatmaps were generated to visualize the relationships between each module and trait (Fig. 3B, F). In the transplant group, the turquoise ($r = 0.79$, $p < 1E\text{-}200$) and yellow ($r = 0.84$, $p < 1E\text{-}200$) modules, comprising 2,492 and 1,500 genes, respectively, were strongly correlated with disease diagnosis (Fig. 3C, D). In the biopsy group, the midnightblue ($r = 0.55$, $p = 2.3E\text{-}10$) and magenta ($r = 0.66$, $p = 1.8E\text{-}67$) modules, which included 114 and 529 genes, respectively, were strongly associated with disease diagnosis (Fig. 3G, H). A total of 38 overlapping key diagnostic genes were obtained after taking the intersection of the DEGs with the genes obtained from the WGCNA key disease diagnostic module analysis (Fig. 3I).

GO and KEGG analysis

Analysis using GO and KEGG showed that the critical diagnostically genes were associated with oxidative phosphorylation, extracellular matrix organization, stem cell population maintenance, vasculature development, biological processes involved in symbiotic interactions, supramolecular fiber organization, metabolic processes of sulfur compounds, and transmembrane transport of inorganic cations (Fig. 4)[13–15].

Diagnostic biomarker confirmation

Potential biomarkers were chosen using the LASSO, SVM-RFE, and random forest algorithms. LASSO regression was used to identify 20 important genes as diagnostic biomarkers (Fig. 5A). SVM-RFE and random forest were used to validate 26 and 32 key genes, respectively (Fig. 5B and C). When integrating these three algorithms, 10 overlapping diagnostic biomarkers were selected (Fig. 5D).

Validation of diagnostic biomarkers

The diagnostic performance of the 10 genes was verified using ROC curves (Fig. 6A). In addition, to generate more reliable and accurate gene sequences, we used the GSE124685 and GSE184316 datasets to confirm the expression levels of the 10 gene sequences. The expression levels of ITGAV, BHLHE41, COL5A3, IRF4, and LIF in the IPF samples were significantly greater than those in the CHP samples ($P < 0.05$); the expression levels of BUD13 and NKG7 were considerably lower than those in the CHP samples ($P < 0.05$); ADAMTS1 did not significantly vary between the CHP and IPF groups; and MT-ND2 and MT-ND2 were lost in the test group. We further analyzed the diagnostic performance of seven genes. The AUCs of COL5A3, BHLHE41 and ITGAV were 76.03%, 84.98% and 71.66%, respectively, indicating good diagnostic value for the early detection of IPF;

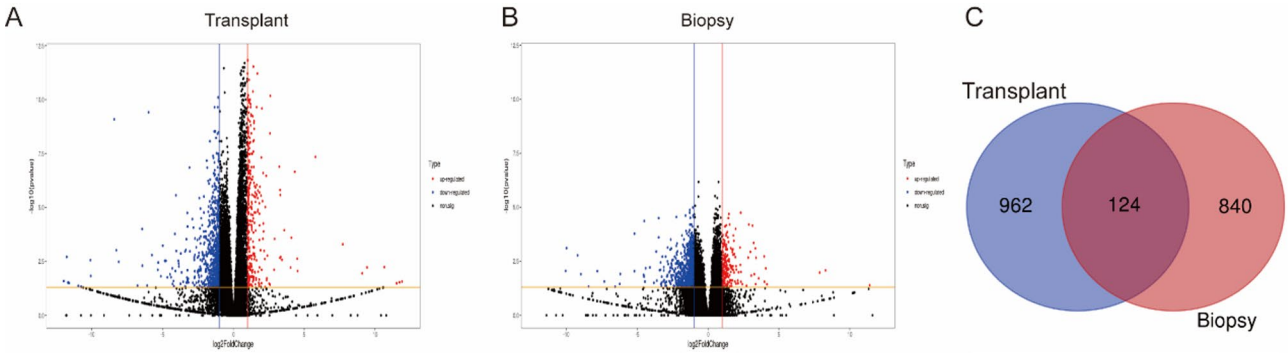

Fig. 2. Identification of DEGs. (A) IPF and CHP control sample volcano plots in the transplant group. (B) IPF and CHP control sample volcano plots in the biopsy group. In the volcano graphic, genes that were not differently expressed are represented by black dots, downregulated genes by blue dots, and upregulated genes by red dots. (C) Venn diagram showing the 124 DEGs shared by the transplant and biopsy groups.

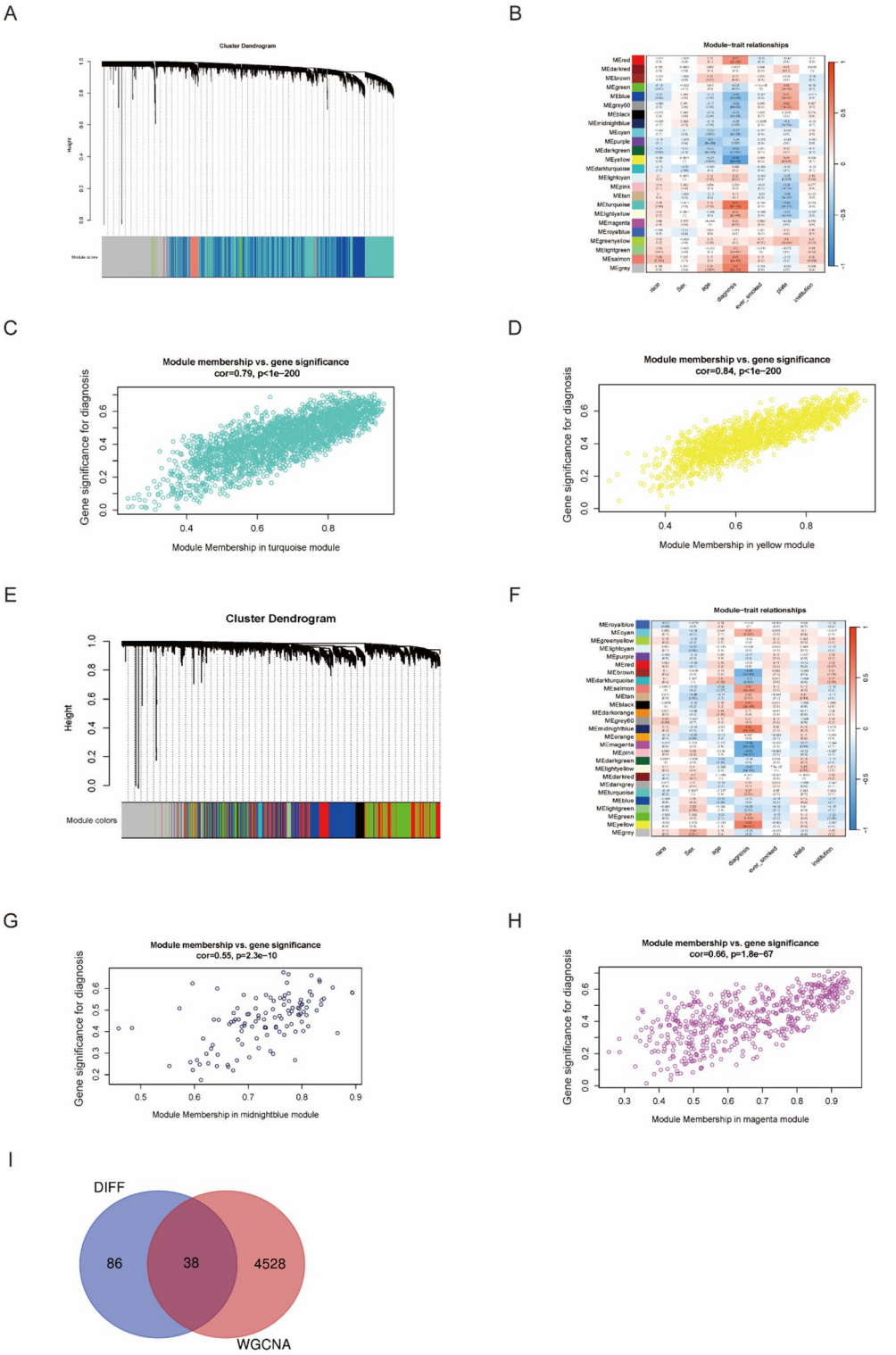

Fig. 3. WGCNA of transplant and biopsy groups. A Cluster dendrogram showing the transplant group's co-expressed genes and modules. B Module-trait connections within the transplant cohort. C Modules of strong positive correlations between genes and disease diagnoses. D Modules of strong negative correlations between genes and disease diagnoses. E cluster dendrogram of the biopsy group's co-expressed genes and modules. F Module-trait connections within the biopsy cohort. G Modules of genes strongly positively correlated with disease diagnosis. H Modules of genes strongly negatively correlated with disease diagnosis. I Venn diagram showing critical diagnostically genes identified by both differential expression analysis and WGCNA.

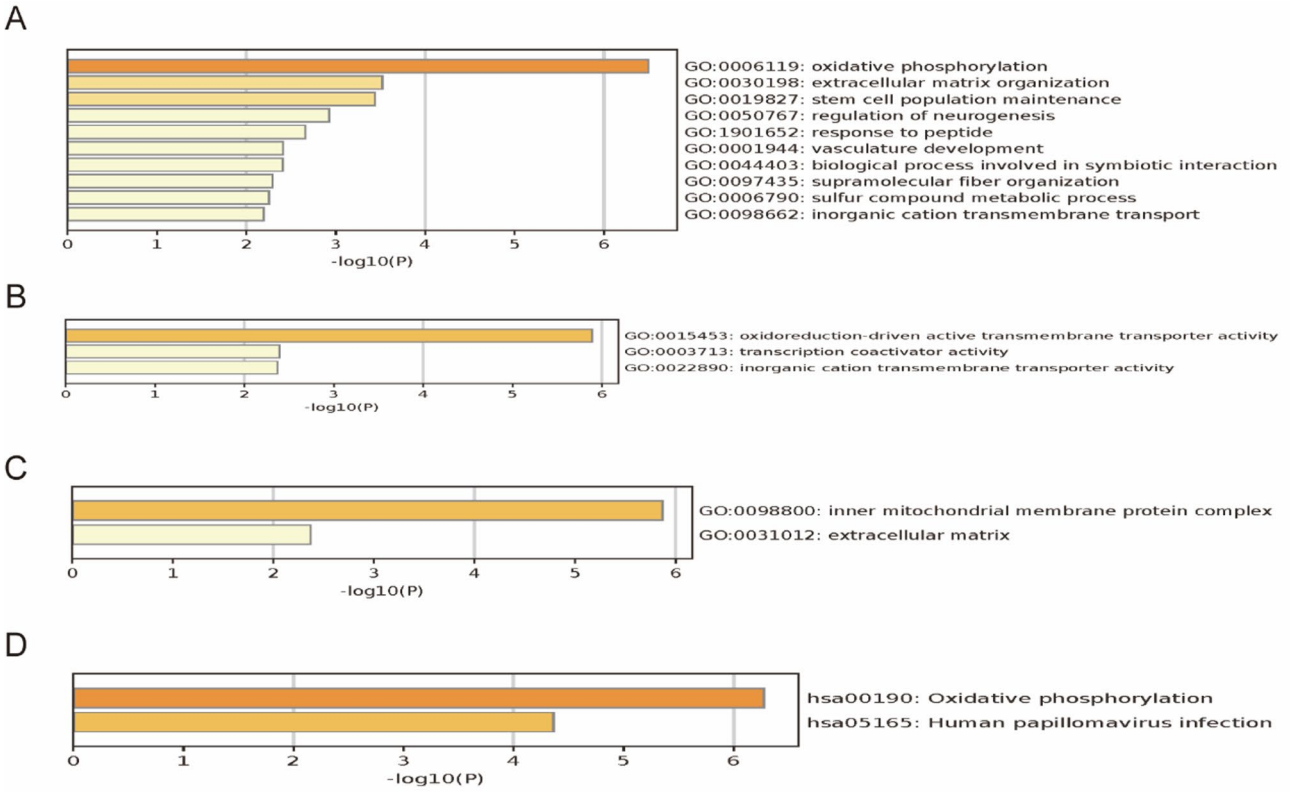

Fig. 4. Functional enrichment analysis identified potential biological processes. GO analysis categorized diagnostic key genes into three functional groups: (A) biological process (BP), (B) molecular function (MF), and (C) cellular composition (CC). (D) KEGG analysis of candidate diagnostic genes.

the AUCs of BUD13 and NKG7 were 71.44% and 80.77%, respectively, demonstrating strong diagnostic utility for CHP early detection (Fig. 6B). The gene expression levels with diagnostic value for IPF are shown in Fig. 6C.

Validity of diagnostic biomarkers in IPF

To better identify different UIP patterns, we used the GSE199152 dataset to further analyze the differences in the gene expression of COL5A3, BHLHE41 and ITGAV in IPF-UIP and RA-UIP. BHLHE41 and COL5A3 did not substantially differ between the IPF-UIP and RA-UIP groups, while the expression level of ITGAV in the IPF-UIP group was considerably higher than that in the RA-UIP group (Fig. 7A-C). We found that ITGAV, with an AUC of 85%, demonstrated a significant advantage in distinguishing IPF from RA-UIP (Fig. 7D). The ROC curve for ITGAV was calculated using the presence or absence of IPF as the outcome variable. The AUC was 72.15%-a significant value. The ITGAV cutoff value for IPF was 5.696, with sensitivity of 64.22% and specificity of 74.36% (Fig. 7E). The results indicate that the ITGAV gene, which encodes the integrin αv subunit protein, is a specific biomarker for distinguishing IPF from secondary UIP.

Immune infiltration analysis

Previous studies have demonstrated that the dynamic recruitment and tissue infiltration of immune cells are crucial aspects of the pathogenesis of IPF[16]. The integrin family, which consists of transmembrane receptors formed by α/β heterodimers, plays a crucial role in the migration and activation of immune cells by mediating cell-matrix adhesion, mechanistic signaling, and microenvironmental remodeling. To elucidate the potential regulatory role of the integrin αv subunit (encoded by the ITGAV gene) in the immune microenvironment of IPF, we employed the CIBERSORT method to evaluate the differences in immune cell infiltration between IPF and healthy samples in the GSE124685 and GSE184316 datasets. Additionally, we examined the relationship between ITGAV and infiltrating immune cells in IPF (Fig. 8). The results indicate that the pathogenesis and progression of IPF are potentially associated with various immune cells, including B cells, T cells, neutrophils, monocytes, and macrophages. Furthermore, alterations in the expression of ITGAV are implicated in the abnormal activities of these immune cells. Among them, the monocyte-macrophage system (including monocytes, M0 macrophages, M1 macrophages, and M2 macrophages) exhibited a significant change in infiltration in IPF, which was closely associated with ITGAV. In IPF, there was a decrease in monocyte infiltration and an increase in the infiltration of both M1 and M2 macrophages in lung tissue. Additionally, a trend toward elevated levels of M0 macrophages was observed. The ITGAV gene exhibited a negative correlation with monocytes and M1 and M2 macrophages, while showing a positive correlation with M0 macrophages.

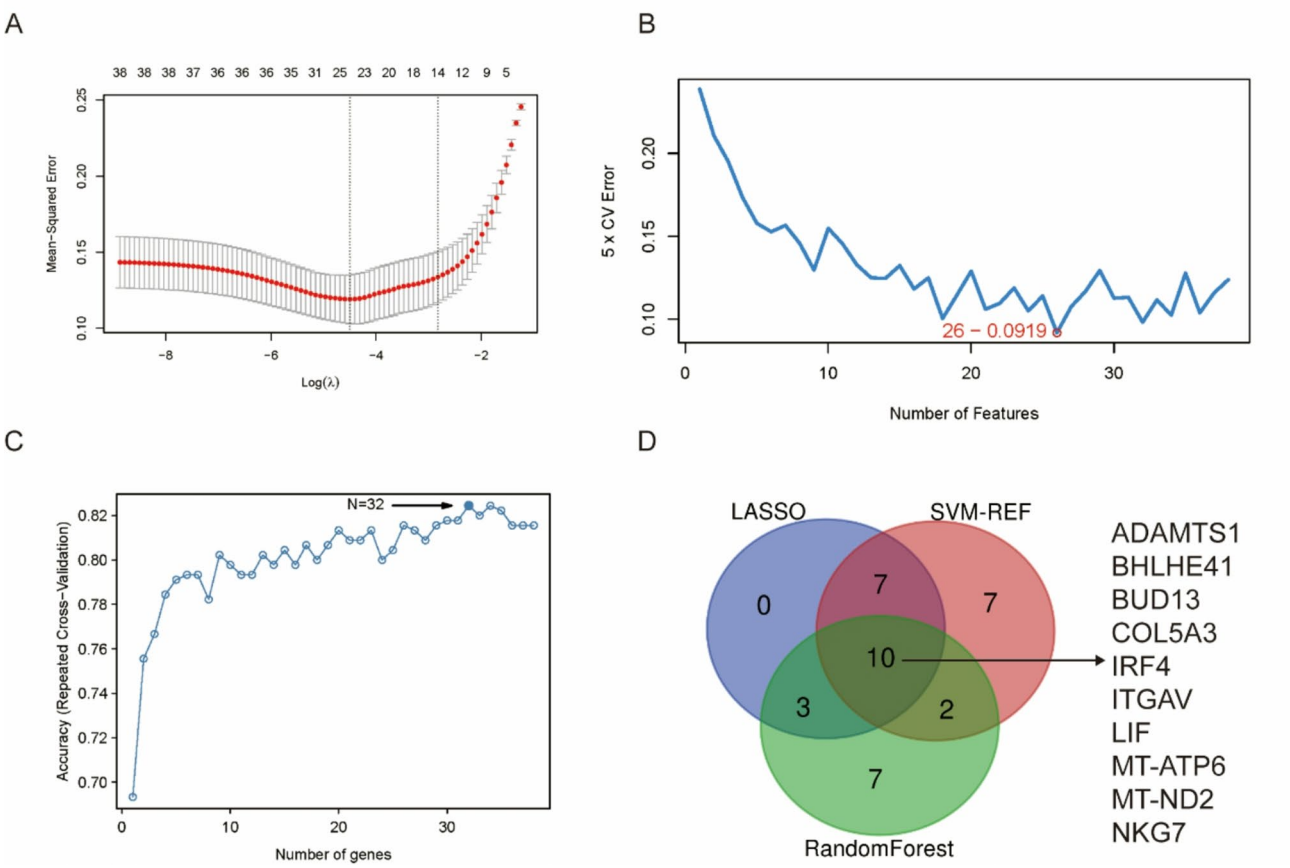

Fig. 5. Screening process for potential diagnostic biomarkers. (A) Feature selection tuning of the LASSO model. (B) SVM-RFE-based biomarker selection map. (C) Random forest algorithm-based predictor variable screening. (D) Venn diagram displaying the 10 diagnostic biomarkers common to the LASSO, SVM-RFE, and random forest algorithms.

ITGAV contributes to the onset of PF

ITGAV is variably elevated in diseases presenting as UIP

We compared the expression levels of ITGAV gene in the control group, the secondary UIP group, and the IPF group in the GSE184316 and GSE199152 datasets in order to investigate the role of ITGAV in the pathophysiology of diseases that display a UIP pattern (Fig. 9). In comparison to controls, ITGAV was increased to varied degrees in all diseases exhibiting a UIP pattern. Further, ITGAV was more significantly elevated in IPF compared to secondary UIP. The findings imply that ITGAV, a particular biomarker of IPF, plays a role in the onset of PF. Dramatic elevation of ITGAV may be one of the important pathogenic mechanisms of IPF.

ITGAV is highly expressed in PF rats

The results of HE and Masson staining revealed that the control group's lung tissue structure was mostly normal, with the alveolar septum displaying a tiny amount of blue collagen deposition. The establishment of PF was indicated in the model group by the disordered lung tissue structure, the substantial infiltration and aggregation of inflammatory cells, and the accumulation of blue-stained collagen fibers in the alveolar septum (Fig. 10A, B). We evaluated the expression of ITGAV in the lung tissues of bleomycin-induced PF rats by IHC and RT-qPCR (Fig. 10C, D, E). PF rats had higher levels of ITGAV expression than control rats ($P < 0.05$). Specifically, ITGAV was predominantly expressed on cell membranes and was highly expressed on a variety of immune cells, including alveolar macrophages, which matched the findings of earlier assessments of immune infiltration.

ITGAV is highly expressed on the membrane of monocytes/macrophages in rats with PF

Previous findings suggest that monocytes and their derived macrophages drive the development of PF[17]. We first verified the expression level of ITGAV on monocytes/macrophages by IF. Monocytes/macrophages (CD14+) were increased in lung tissues of rats with PF compared with controls and highly expressed the ITGAV. It was seen that ITGAV co-localized with monocytes/macrophages (Fig. 11A-C). We further examined the expression level of ITGAV on monocyte-derived alveolar macrophages (MO-AMs) by flow cytometry. The results indicated that the population of monocyte-derived macrophages (MO-Mφ) characterized by high expression of ITGAV (CD45 + CD11b + CD68 + ITGAV) was significantly elevated in the alveolar lavage fluid of rats with PF compared

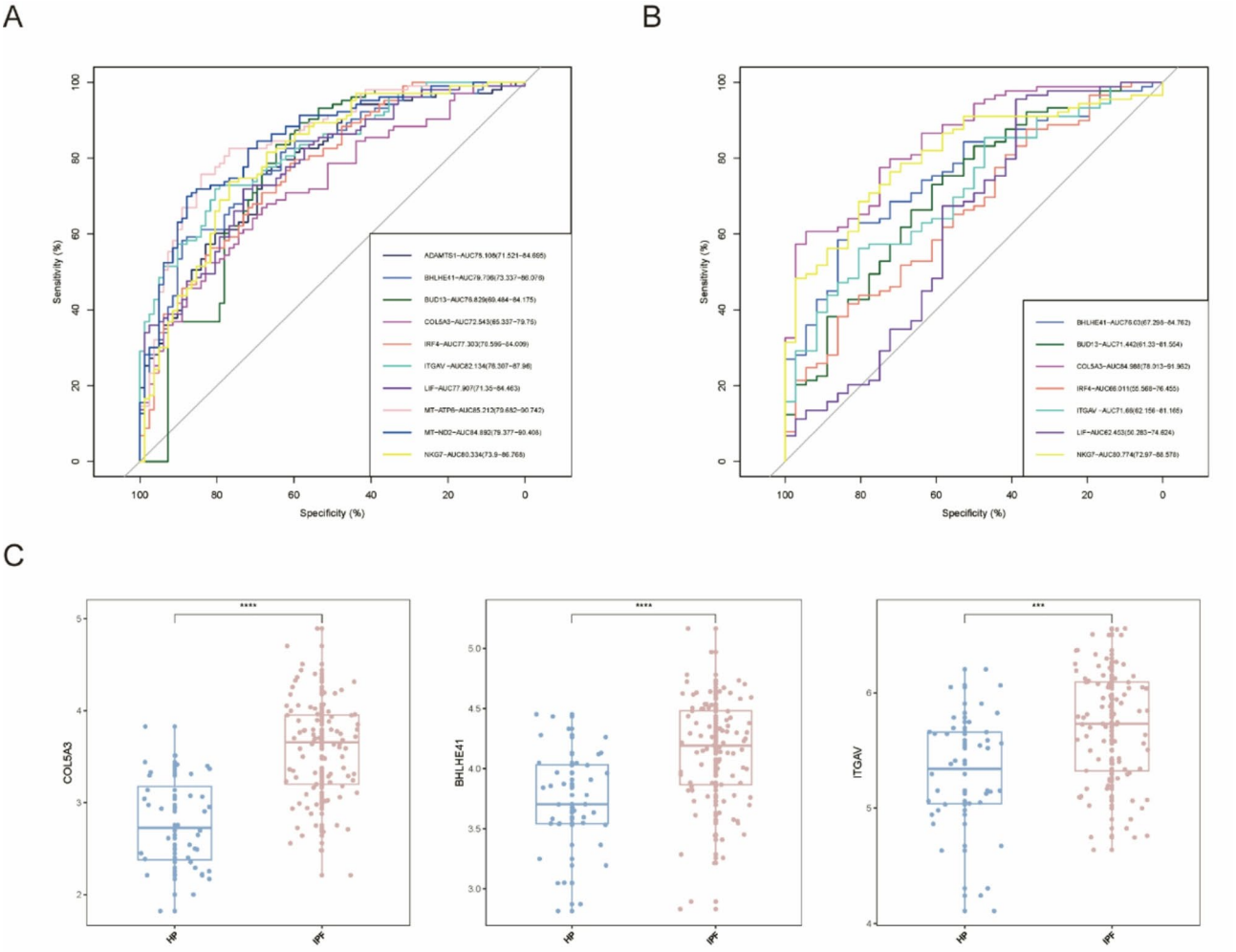

Fig. 6. Validation process for potential diagnostic biomarkers. (A) ROC curves for the 10 diagnostic markers' diagnostic efficacy. (B) ROC curves for indicators of diagnostic efficacy in the GSE124685 and GSE184316 datasets. (C) Expression levels of IPF diagnostic biomarkers in the GSE124685 and GSE184316 datasets.

to the control group (Fig. 11D, E). This finding implies that ITGAV-associated signaling pathway on MO-Mφ is associated with the onset and progression of PF.

Discussion

Accurate identification of IPF among various diseases that present as UIP is crucial for developing personalized treatment strategies and enhancing patient prognosis. In recent years, significant advancements have been made in the field of precision diagnosis for fibrotic ILD[18]. Machine learning applied to plasma proteomics can effectively differentiate CTD-ILD from IPF[19]. Specific microbial community patterns and immune response-related gene expression signatures in bronchoalveolar lavage can assist in distinguishing fHP from IPF[20]. The Envisia Genomic Classifier (Veracyte) is a commercially available genomic classifier that demonstrates high specificity for histopathological patterns of UIP[21]. Although the aforementioned studies have significantly enhanced the differentiation of IPF, the current diagnostic system still requires the integration of multiple sets of differential diagnostic tools. In this study, we facilitated the precise differentiation of IPF from secondary UIP conditions, such as CTD-ILD and CHP, by identifying biomarkers with dual diagnostic properties related to etiology and pathological patterns. In this study, we identified three diagnostic biomarkers—COL5A3, BHLHE41, and ITGAV—that effectively differentiate IPF from CHP. These biomarkers exhibited AUC of 76.03%, 84.98%, and 71.66%, respectively, indicating their strong diagnostic value. We further validated these biomarkers for distinguishing IPF from RA-UIP. The expression of ITGAV was significantly elevated in IPF compared to RA-UIP, yielding an AUC of 85%, which further underscores its strong diagnostic potential. We ultimately confirmed the diagnostic efficacy of ITGAV as a specific biomarker for differentiating IPF from secondary UIP. This conclusion is supported by a sensitivity of 64.22%, a specificity of 74.36%, and an AUC of 72.15%. There are limitations to this study. The dataset containing IPF and CHP did not differentiate between specific pathologies and disease severity, which can impact the specificity of diagnostic biomarkers. The number of patients with RA-UIP was insufficient, and the validation performance of the GSE199152 dataset was suboptimal. Although

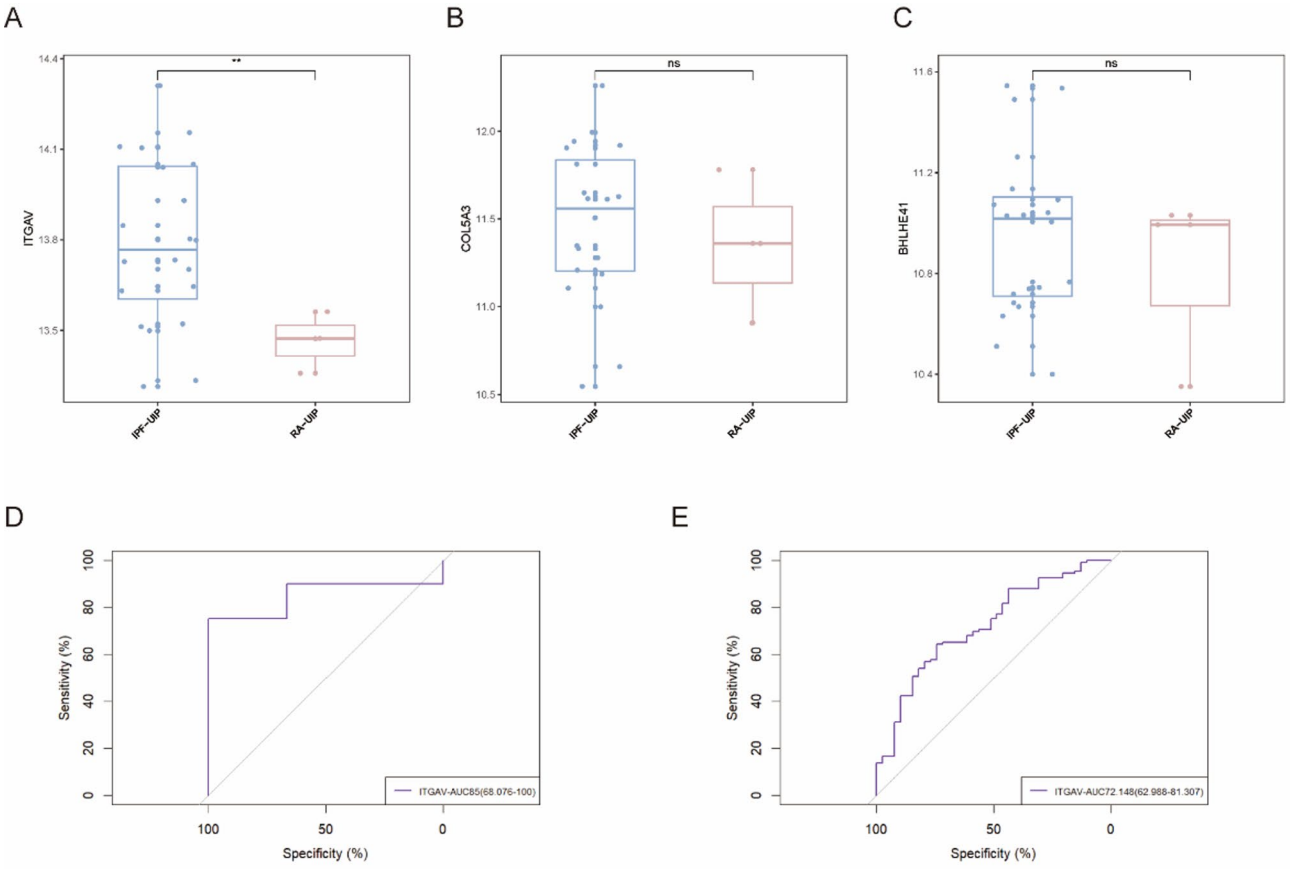

Fig. 7. Assessment of the utility of ITGAV as a diagnostic biomarker for IPF. Expression levels of (A) ITGAV, (B) COL5A3, and (C) BHLHE41 in IPF and RA-UIP samples. (D) ROC curve of ITGAV in distinguishing IPF from RA-UIP. (E) ROC curve of ITGAV in the diagnosis of IPF.

ITGAV demonstrated potential diagnostic value in this study, further efforts are required to effectively translate it into clinical applications. Establishing the study design and assay process for ITGAV as a clinical diagnostic biomarker presents a significant challenge. The accuracy and reliability of ITGAV diagnoses must be further validated through prospective studies involving larger sample sizes of patients with IPF. Standardized workflows should be developed, encompassing sample collection, isolation methods, sample processing, and downstream testing, while also addressing potential sources of bias in data analysis. Furthermore, the characterization of biomarkers in biological fluids has emerged as a powerful approach for disease diagnosis. In this study, the samples for ITGAV gene testing were lung tissues, which necessitate invasive procedures for collection. Exploring alternative sample types from biological fluids allows for rapid, non-invasive, and cost-effective disease diagnosis, as well as monitoring of disease state and treatment response. Finally, the expression of ITGAV may be influenced by various factors, including individual differences. Therefore, it is essential to integrate other diagnostic indicators in clinical practice to enhance diagnostic accuracy.

Our study found that ITGAV, a specific diagnostic biomarker for IPF, is closely associated with the development of PF. ITGAV exhibited a trend of gradual elevation among the control group, the secondary UIP group, and the IPF group. The integrin subunit αv, encoded by the ITGAV gene, is predominantly expressed on alveolar epithelial cells, pulmonary interstitial cells, T cells, and MO-Mφ. It can form heterodimers with β1, β3, β5, β6, or β8 subunits, which are central to the development of PF[22]. Multiple investigators have demonstrated in experimental models of PF that αv integrins activate latent TGF-β, a well-known pro-fibrotic cytokine, thereby promoting the progression of PF[23–25]. It has been shown that αvβ3 regulates the activity and localization of the Src family of kinases by interacting with Thy-1, a cell surface glycoprotein, thereby affecting the force perception and fibrotic response of fibroblasts[26]. αvβ6 and αvβ1 are the most effective targets that have been thoroughly validated, as they are highly expressed in alveolar epithelial cells and pulmonary interstitial cells, respectively. These integrins may play a role in alveolar injury, TGF-β activation, and extracellular matrix deposition[27]. It has been demonstrated that among CD4 + T cells, regulatory T cells are the primary producers of latent TGF-β1 and possess a unique ability to activate latent TGF-β1 through the expression of the αv integrin[28]. When co-cultured with primary lung fibroblasts, Jurkat T cells that overexpress the integrins αvβ3 or αvβ5 promoted collagen accumulation and facilitated the nuclear translocation of Smad2[29]. Experimental evidence indicates that the depletion of circulating monocytes and the absence of Mo-AM recruitment to the lungs significantly reduce fibrosis following lung injury[17]. Misharin et al. found that monocyte-derived interstitial macrophages and AMs

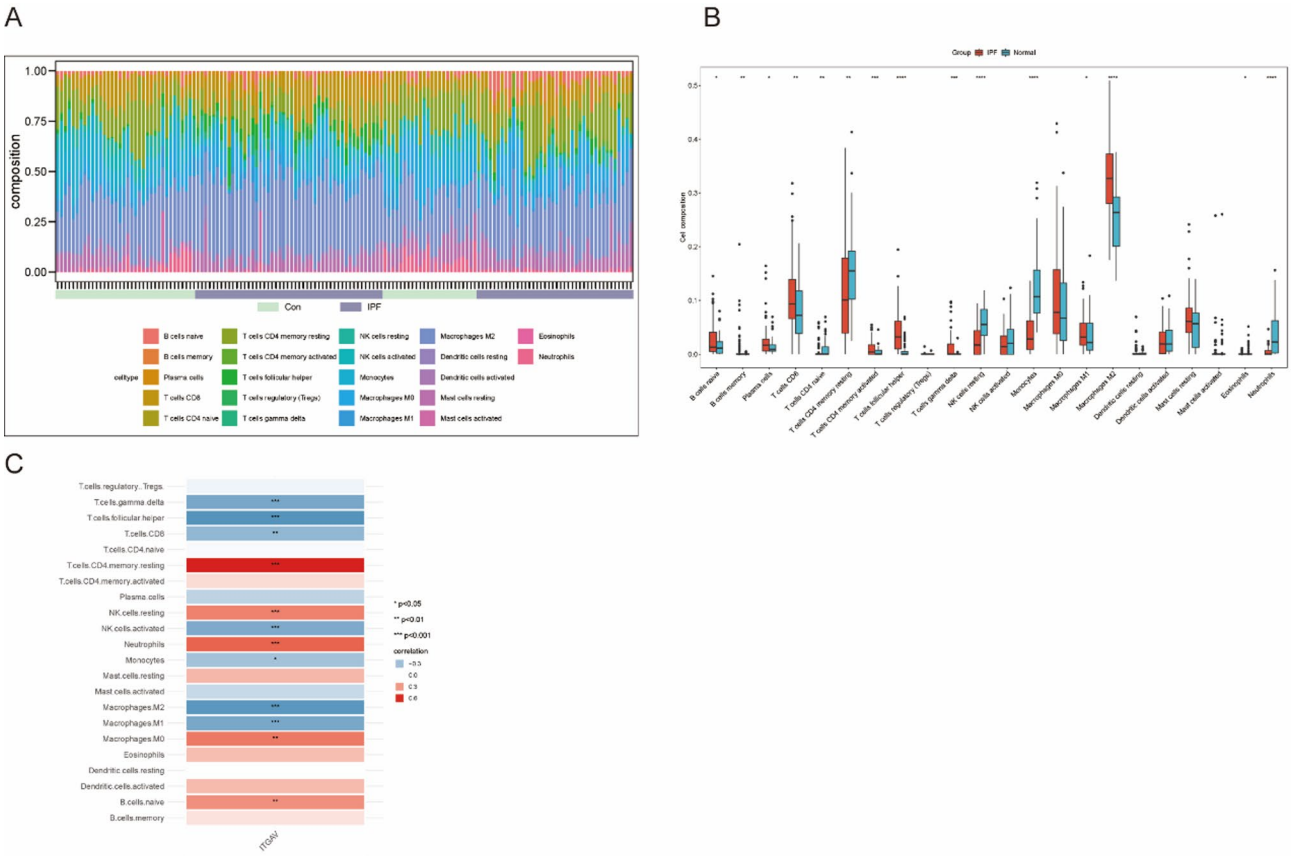

Fig. 8. Immune cell infiltration distribution, visualization and correlation analysis. (A) Bar graph displaying the ratio of 22 immune cell subtypes between IPF and healthy samples. (B) Box plots displaying the total distribution of immune cells in IPF and healthy samples. (C) The relationship between ITGAV and infiltrating immune cells in IPF.

highly expressed multiple fibrogenic genes, including the ITGAV gene, compared with tissue-resident AMs by DGE analysis in bleomycin-induced PF mice[30].

We found that the infiltration of the monocyte-macrophage system, which includes monocytes, M0 macrophages, M1 macrophages, and M2 macrophages, was significantly altered in IPF compared to other immune cells. Previous studies have shown that elevated monocyte counts are associated with an increased risk of IPF progression, hospitalization, and death and are a prognostic biomarker for IPF[31]. Using RNA microarray data, a comparative analysis of 122 IPF patients and 91 healthy controls revealed that the lung tissue of IPF patients had significantly lower levels of monocyte infiltration while M0- and M2-type macrophage infiltration was significantly higher[32]. M2 macrophage-associated markers sCD206 and sCD163 were significantly elevated in patients with IPF. Additionally, high serum levels of sCD206 were correlated with increased mortality in individuals with IPF[33]. The above results suggest that in IPF, ongoing injury and repair processes may lead to monocyte recruitment to the lungs, differentiation into M0-type macrophages and polarization into M2-type macrophages that drive fibrosis. The decrease of monocytes in lung tissue and the increase of monocytes in peripheral blood reflect their differentiation and compensatory proliferation process[17]. Additionally, the ITGAV gene was positively associated with M0 macrophages and negatively associated with monocytes, as well as M1 and M2 macrophages. Our experimental results showed that monocyte-derived M0 macrophages (CD45 + CD11b + CD68+) expressed more ITGAV in PF rats compared to controls. Here, we propose a research hypothesis that in the microenvironment of lung fibrosis, MO-Mφ may enhance their pro-fibrotic function by upregulating ITGAV expression. In the future, we will verify the effect of this gene deletion on pro-fibrotic function through conditional knockout of ITGAV in MO-Mφ. Due to the dynamic regulatory role of the monocyte-macrophage system in PF, we will enhance the time-series mapping of ITGAV expression through single-cell transcriptome sequencing in the future. This approach will allow for a detailed analysis of the relationship between ITGAV expression and the functional state transitions of monocytes and macrophages.

Because of the critical role of the αv subunit and its associated integrins in the development of IPF, recent preclinical studies and clinical trials have focused on the development of αv integrin inhibitors[34]. New inhibitors of integrins αvβ6 and αvβ1 are under clinical investigation for the treatment of IPF[35]. Besotegrast (PLN-74809) is an oral small-molecule inhibitor that targets the αvβ6 and αvβ1 integrins. It has demonstrated favorable safety and tolerability profiles and is regarded as the most promising drug candidate for success in clinical trials. The development of integrin inhibitors has provided a new direction in the search for more effective IPF treatments.

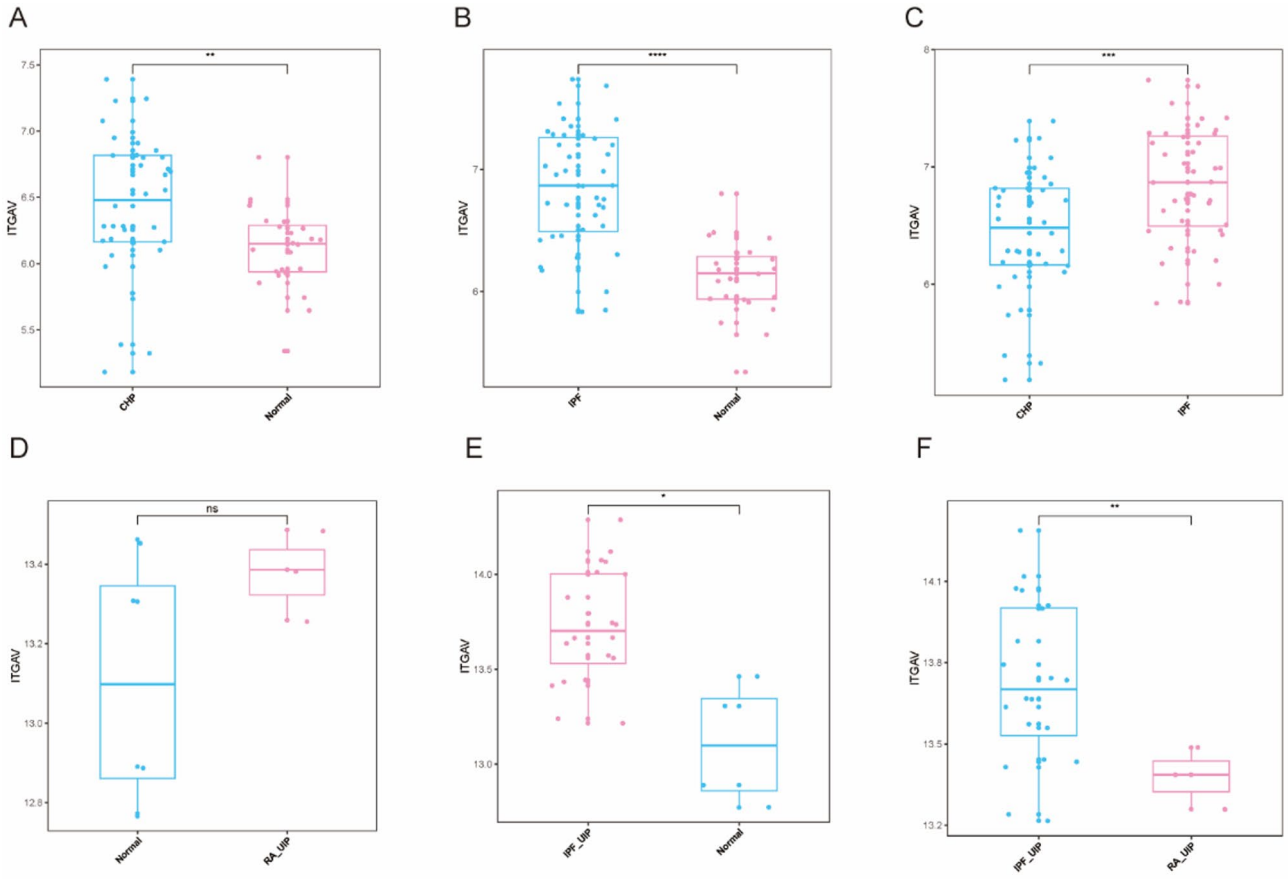

Fig. 9. Comparison of ITGAV expression in control, secondary UIP and IPF groups in GSE184316 and GSE199152 dataset.

The aforementioned studies confirmed that the integrin αv subunit is widely expressed in various lung tissue cell populations, including alveolar epithelial cells, interstitial cells, and immune cells. It plays a crucial role in the pathological processes of IPF by mediating intercellular and matrix interactions. Looking forward, it is feasible to further delineate the key pathological cell subpopulations exhibiting high ITGAV expression and to construct gene co-expression networks through single-cell transcriptome sequencing. This approach aims to develop targeted delivery systems with cell type specificity, thereby improving the precision and safety of targeted therapies. Additionally, it will facilitate the design of multi-pathway combination therapies that synergistically inhibit fibrotic signaling cascades, ultimately enhancing the efficacy of treatments.

Conclusion

According to the GEO database, ITGAV is the central gene that distinguishes IPF from secondary UIP. In PF rats, MO-Mφ highly express ITGAV, which may be related to the pathogenesis of IPF. These findings provide a molecular basis for future diagnosis and treatment of IPF.

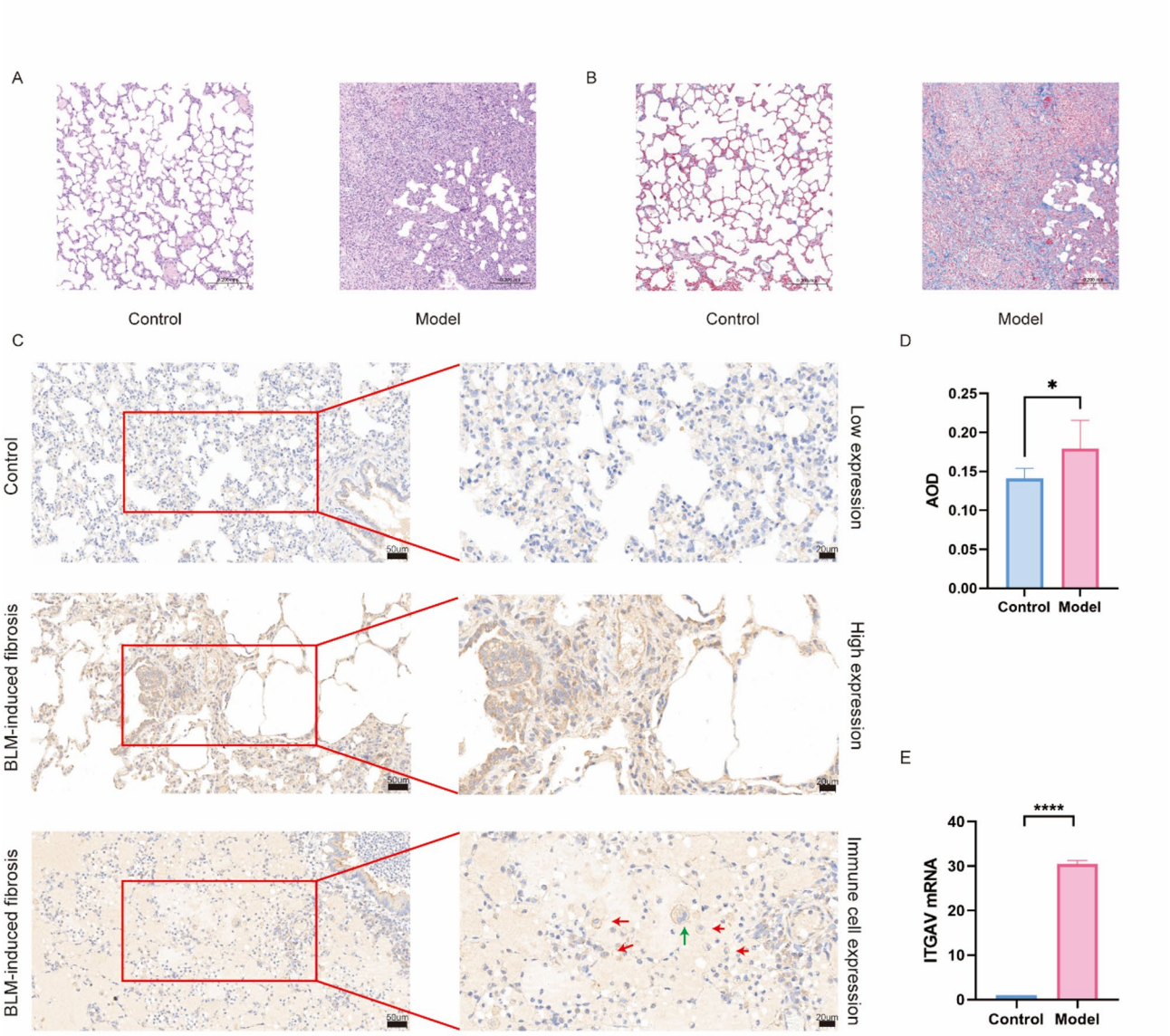

Fig. 10. ITGAV is highly expressed in PF rats. (A)HE staining of control and model groups. (B)Masson staining of control and model groups. (C) Representative images (×200 and ×400) of IHC staining of ITGAV in healthy and PF rats (low expression vs. high expression, $n = 6$). A variety of immune cells (indicated by red arrows) and alveolar macrophages (indicated by green arrows). (D) AOD levels of ITGAV in healthy and PF rats ($n = 6$). (E) mRNA expression levels of ITGAV in healthy and PF rats ($n = 6$). * $p < 0.05$, **** $p < 0.0001$.

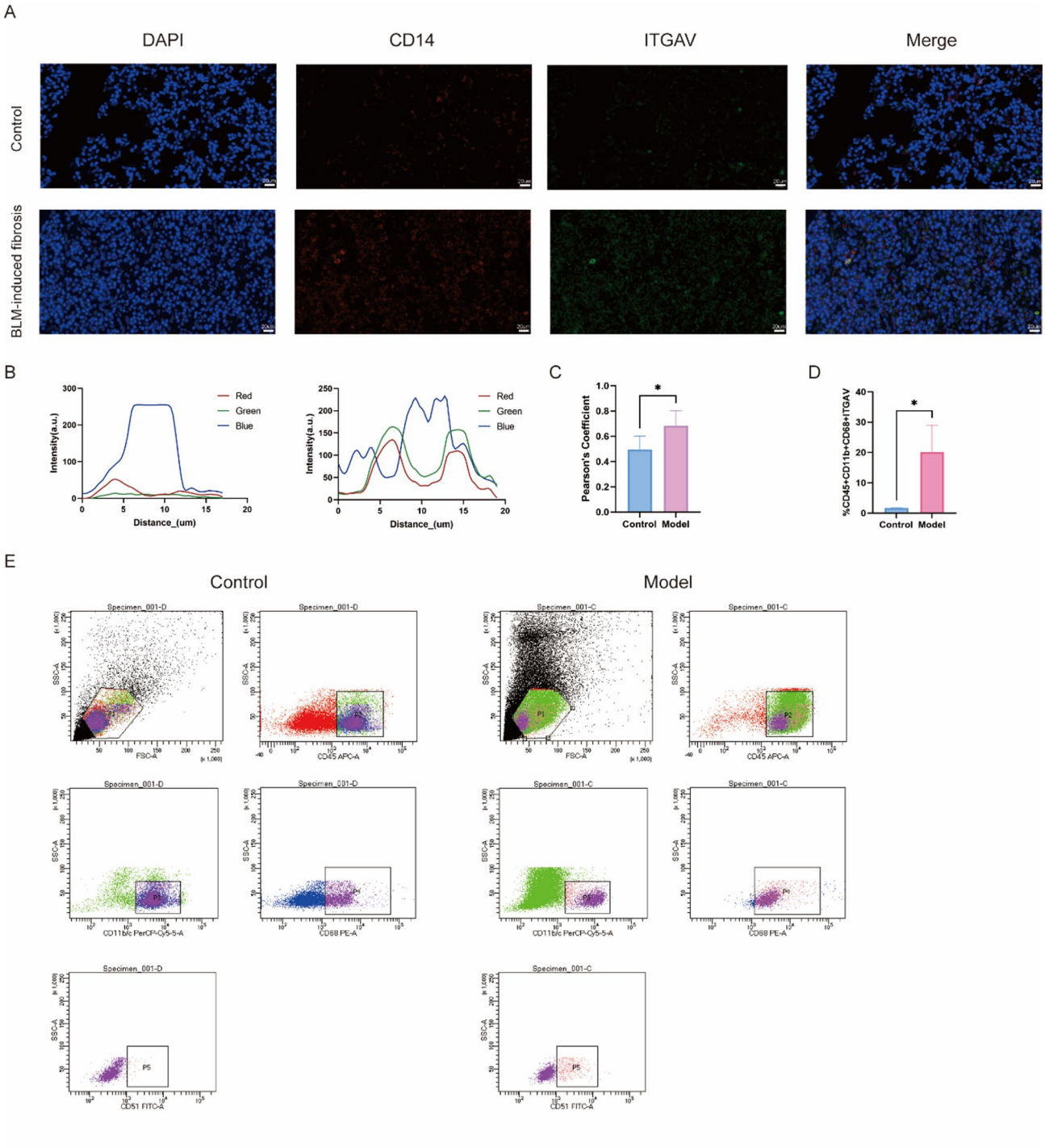

Fig. 11. ITGAV is highly expressed on the membrane of monocytes/macrophages in rats with PF. (A) CD14(red) and ITGAV (green) fluorescence double labeled monocytes/macrophages expressing integrin αv. (B)Fluorescence intensity analysis of nuclei, CD14, and ITGAV in the control and model groups. (C) Colocalization of CD14 and ITGAV in the control and model groups. (D) Proportion of MO-Mφ populations with high expression of ITGAV in alveolar lavage fluid of control and model groups. (E) Flow cytometric scatter plots of control and model groups (CD45 + CD11b + CD68 + ITGAV). * $p < 0.05$.

Data availability

The data that support the findings of this study are available in Mendeley Data, V1, doi: 10.17632/gdfbm83yhb.1.

Received: 16 February 2025; Accepted: 6 May 2025
Published online: 15 May 2025

References

1. Selman, M., Pardo, A. & King, T. E. Hypersensitivity pneumonitis: insights in diagnosis and pathobiology. *Am. J. Respir Crit. Care Med.* **186**, 314–324 (2012).
2. Girard, M., Israël-Assayag, E. & Cormier, Y. Pathogenesis of hypersensitivity pneumonitis. *Curr. Opin. Allergy Clin. Immunol.* **4**, 93–98 (2004).
3. Behr, J. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N. Engl. J. Med.* 367, 869; author reply 870-1 (2012).
4. 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 (2011).
5. Raghu, G. et al. Diagnosis of hypersensitivity pneumonitis in adults. An official ATS/JRS/ALAT clinical practice guideline. *Am. J. Respir Crit. Care Med.* **202**, e36–e69 (2020).
6. Chen, R. X. et al. [Clinical analysis of autoimmune diseases associated with interstitial lung diseases initially presented with idiopathic pulmonary fibrosis]. *Zhonghua Jie He He Hu Xi Za Zhi.* **45**, 775–782 (2022).
7. Furusawa, H. et al. Chronic hypersensitivity pneumonitis, an interstitial lung disease with distinct molecular signatures. *Am. J. Respir Crit. Care Med.* **202**, 1430–1444 (2020).
8. Love, M. I., Huber, W. & Anders, S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
9. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
10. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The Sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
11. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
12. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
13. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
14. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
15. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
16. 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 (2019).
17. Perrot, C. Y., Karampitsakos, T. & Herazo-Maya, J. D. Monocytes and macrophages: emerging mechanisms and novel therapeutic targets in pulmonary fibrosis. *Am. J. Physiol. - Cell. Physiol.* **325**, C1046–C1057 (2023).
18. Karampitsakos, T., Tourki, B. & Herazo-Maya, J. D. The dawn of precision medicine in fibrotic interstitial lung disease. *Chest* **167**, 1120–1132 (2025).
19. Tan, C. et al. Soluble Thy-1 reverses lung fibrosis via its integrin-binding motif. *JCI Insight*. **4**, 1–13 (2019).
20. Invernizzi, R. et al. The respiratory Microbiome in chronic hypersensitivity pneumonitis is distinct from that of idiopathic pulmonary fibrosis. *Am. J. Respir Crit. Care Med.* **203**, 339–347 (2021).
21. Richeldi, L. et al. Utility of a molecular classifier as a complement to High-Resolution computed tomography to identify usual interstitial pneumonia. *Am. J. Respir Crit. Care Med.* **203**, 211–220 (2021).
22. Conroy, K. P., Kitto, L. J. & Henderson, N. C. αv integrins: key regulators of tissue fibrosis. *Cell. Tissue Res.* **365**, 511–519 (2016).
23. Yi, M. et al. Inhibition of TGFβ1 activation prevents radiation-induced lung fibrosis. *Clin. Transl Med.* **14**, e1546 (2024).
24. Checa, M. et al. Cigarette smoke enhances the expression of profibrotic molecules in alveolar epithelial cells. *PLoS One*. **11**, e0150383 (2016).
25. Huang, X. et al. Molecular characterization of a precision-cut rat lung slice model for the evaluation of antifibrotic drugs. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **316**, L348–L357 (2019).
26. Fiore, V. F. et al. Conformational coupling of integrin and Thy-1 regulates Fyn priming and fibroblast mechanotransduction. *J. Cell. Biol.* **211**, 173–190 (2015).
27. Slack, R. J., Macdonald, S. J. F., Roper, J. A., Jenkins, R. G. & Hatley, R. J. D. Emerging therapeutic opportunities for integrin inhibitors. *Nat. Rev. Drug Discov.* **21**, 60–78 (2022).
28. Edwards, J. P., Thornton, A. M. & Shevach, E. M. Release of active TGF-β1 from the latent TGF-β1/GARP complex on T regulatory cells is mediated by integrin B8. *J. Immunol.* **193**, 2843–2849 (2014).
29. Luzina, I. G. et al. Regulation of pulmonary inflammation and fibrosis through expression of integrins alphaVbeta3 and alphaVbeta5 on pulmonary T lymphocytes. *Arthritis Rheum.* **60**, 1530–1539 (2009).
30. Misharin, A. V. et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J. Exp. Med.* **214**, 2387–2404 (2017).
31. Kreuter, M. et al. Monocyte count as a prognostic biomarker in patients with idiopathic pulmonary fibrosis. *Am. J. Respir Crit. Care Med.* **204**, 74–81 (2021).
32. Zhao, T. et al. Identification and validation of chemokine system-related genes in idiopathic pulmonary fibrosis. *Front. Immunol.* **14**, 1–15 (2023).
33. Zou, R. et al. Association of serum macrophage-mannose receptor CD206 with mortality in idiopathic pulmonary fibrosis. *Int. Immunopharmacol.* **86**, 106732 (2020).
34. Hatley, R. J. D. et al. An αv-RGD integrin inhibitor toolbox: drug discovery insight, challenges and opportunities. *Angew Chem. Int. Ed. Engl.* **57**, 3298–3321 (2018).
35. Decaris, M. L. et al. Dual Inhibition of Avβ6 and Avβ1 reduces fibrogenesis in lung tissue explants from patients with IPF. *Respir Res.* **22**, 265 (2021).

Author contributions

ZYB: Conceptualization, Writing – original draft. XDW: Conceptualization, Resources. XL: Conceptualization, Supervision. XZ: Data curation, Project administration. GDZ: Data curation, Resources. LT: Formal Analysis, Visualization, Writing – review & editing. WZ: Conceptualization, Funding acquisition, Resources, Supervision, Validation, Writing – review & editing.

Funding

Financial funding was obtained for the research, writing, and/or publication of this article. The National Natural Science Foundation of China (82104799, 81874442), the Natural Science Foundation of Shandong Province (ZR2021LZY031) and Shandong Province Taishan Scholar Project, (No: tsqn202306392) provided funding for this work.

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/s41598-025-01459-4.

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

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.