Construction of an artificial neural network diagnostic model and investigation of immune cell infiltration characteristics for idiopathic pulmonary fibrosis

Huizhe Zhang[1†], Haibing Hua[2†], Cong Wang[3,4], Chenjing Zhu[3,4], Qingqing Xia[3,4], Weilong Jiang[3,4*], Xiaodong Hu[3,4*] and Yufeng Zhang[3,4*]

Abstract

Background Idiopathic pulmonary fibrosis (IPF) is a severe lung condition, and finding better ways to diagnose and treat the disease is crucial for improving patient outcomes. Our study sought to develop an artificial neural network (ANN) model for IPF and determine the immune cell types that differed between the IPF and control groups.

Methods From the Gene Expression Omnibus (GEO) database, we first obtained IPF microarray datasets. To conduct protein-protein interaction (PPI) networks and enrichment analyses, differentially expressed genes (DEGs) were screened between tissues of patients with IPF and tissues of controls. Afterward, we identified the important feature genes associated with IPF using random forest (RF) analysis, and then constructed and validated a prediction ANN mode. In addition, the proportions of immune cells were quantified using cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) analysis, which was performed on microarray datasets based on gene expression profiling.

Results A total of 11 downregulated and 36 upregulated DEGs were identified. PPI networks and enrichment analyses were carried out; the immune system and extracellular matrix were the subjects of the enrichments. Using RF analysis, the significant feature genes *LRRC17*, *COMP*, *ASPN*, *CRTAC1*, *POSTN*, *COL3A1*, *PEBP4*, *IL13RA2*, and *CA4* were identified. The nine feature gene scores were integrated into the ANN to develop a diagnostic prediction model. The receiver operating characteristic (ROC) curves demonstrated the strong diagnostic ability of the ANN in predicting IPF in the training and testing sets. An analysis of IPF tissues in comparison to normal tissues revealed a reduction in the

---

†Huizhe Zhang and Haibing Hua contributed equally.

*Correspondence:
Weilong Jiang
jwljytcm@163.com
Xiaodong Hu
huxido2000@sina.com
Yufeng Zhang
yufengzhang@njucm.edu.cn

Full list of author information is available at the end of the article

infiltration of natural killer cells resting, monocytes, macrophages M0, and neutrophils; conversely, the infiltration of T cells CD4 memory resting, mast cells, and macrophages M0 increased.

Conclusion *LRRC17, COMP, ASPN, CRTAC1, POSTN, COL3A1, PEBP4, IL13RA2,* and *CA4* were determined as key feature genes for IPF. The nine feature genes in the ANN model will be extremely important for diagnosing IPF. It may be possible to use differentiated immune cells from IPF samples in comparison to normal samples as targets for immunotherapy in patients with IPF.

Keywords Artificial neural network, Immune cell infiltration, Idiopathic pulmonary fibrosis, Random forest analysis, Feature gene

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial lung disease. The etiology of IPF is unknown, and its high-resolution computed tomography (HRCT) or pathological manifestation is usual interstitial pneumonia (UIP) [1, 2]. In Europe and North America, the incidence of IPF is between 2.8 and 9.3 per 100,000 people, making it a rare disease. The epidemiological data about IPF are scarce in China, but its incidence has significantly increased in recent years [3, 4]. IPF progresses slowly at the early stage, and it will gradually cause diffuse fibrosis of the lungs, eventually leading to respiratory failure and death [5]. IPF has developed into a severe, potentially fatal condition as a result of a lack of early management and comprehensive understanding of the disease's pathophysiology [6]. Patients with IPF continue to have a dismal prognosis, with a median survival of about three years [7]. It is critical to identify novel targets for the diagnosis and treatment of IPF to enhance the prognosis of affected patients.

IPF is an intricate and multifactorial disease that arises from the interplay between genetic and environmental elements. Genetic factors have been demonstrated to be crucial in the pathogenesis of IPF [8, 9]. An array of characteristic genes that serve as references for the clinical diagnosis of IPF have been linked to its occurrence and progression [10–12]. However, these genes remain inadequate for the early detection of IPF. At present, the diagnosis of IPF is still based on whether HRCT or histological manifestation of the lung is UIP, the application of genomics has had some help in the diagnosis of IPF [1, 2]. Thus, further investigation is required to identify novel approaches that can identify feature genes and establish diagnostic models.

As a chronic lung disease, inflammation and fibrosis are involved in the pathogenesis of IPF. It is mainly due to aberrant wound healing response following repetitive epithelial cell injury. Inflammatory cytokines released by immune cells may activate fibroblasts and connective tissue cell proliferation [13]. Immune dysregulation is involved in the occurrence and development of IPF [14]. Research from animal modeling and human research indicates that innate and adaptive immune mechanisms can orchestrate existing fibrotic responses [15].

Artificial intelligence and artificial neural networks (ANNs) have been progressively introduced into the medical field to assist physicians in managing vast volumes of data and implementing precision medicine more easily. ANN is a type of computing mode, which was inspired by the human brain [16]. The learning and trial-and-error methods form the foundation of the ANN algorithm. The prognosis and prediction of tumors were the primary focus of earlier ANN research [17, 18]. Recently, one research constructed an ANN model that demonstrated robust performance across multiple cohorts, but it was not analyzed from the perspective of immune infiltration [19].

Thus, our work aimed to develop an ANN model for IPF using candidate gene weight and compare immune cell types in IPF and control groups. As a first step in this investigation, we gathered IPF microarray datasets from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) between tissues of patients with IPF and tissues of controls were screened to perform enrichment analyses and protein-protein interaction (PPI) network. Afterwards, we identified the important feature genes associated with IPF using random forest (RF) analysis, and then constructed and validated a prediction ANN mode. The prediction power of these crucial feature genes was screened using receiver operating characteristic (ROC) curves. Furthermore, based on the gene expression profiling of microarray datasets, cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) analysis was used to quantify the proportions of immune cells.

Methods

Data acquisition

The GSE110147, GSE21369, and GSE24206 series of matrix files were acquired from the GEO database of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/). The Affymetrix Human Gene 1.0 ST Array's GPL6244 platform serves as the foundation for GSE110147 [20]. The GPL570 platform, which is part of the Affymetrix Human Genome

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:458

Page 3 of 15

U133 Plus 2.0 Array, was used to create both GSE21369 and GSE24206 [21, 22]. The GSE110147 dataset contained 11 samples of normal lung tissue obtained from tissue flanking lung cancer resections and 22 samples collected from the organs of those with IPF (Supplementary File 1A). Eleven samples from patients who had been diagnosed with IPF and six normal samples serving as controls comprised the GSE21369 dataset (Supplementary File 1B). The GSE24206 dataset comprised six control specimens retrieved from healthy donor lungs and 17 samples from patients with IPF (Supplementary File 1C).

Probe annotation files were utilized to convert probes in each dataset into gene symbols. Gene expression values were calculated using the probe with the highest expression level where multiple probes had the same gene symbol.

For further integration analysis, the matrix files of multiple datasets were merged into a merged dataset cohort due to their shared platform and the importance of incorporating large sample size data from various datasets. The "SVA" package's combat function was utilized to preprocess and eliminate batch effects after the three datasets were merged into a single dataset cohort (Supplementary File 1D).

Lung tissue samples from 50 healthy controls and 119 patients with IPF were included in the testing cohort. The GSE32537 dataset, which was based on the Affymetrix Human Gene 1.0 ST Array GPL6244 platform, was used for the study (Supplementary File 1E) [23].

Screening DEGs in dataset between IPF and control samples

The "linear models for microarray data (limma)" package was used to standardize presentation data and identify DEGs [24]. The DEG threshold values were established as follows: |log2 fold change (FC)| > 2 between the IPF and control samples, and adjusted (adj) P value < 0.05. The "ggplot2" and "pheatmap" packages in R plotted volcano plots and heatmaps.

Enrichment analyses of DEGs

Using Metascape (http://metascape.org/), we performed various bioinformatics analyses to get more biological insights into the DEGs [25]. The ontology categories DisGeNET, Pattern Gene Database (PaGenBase), and Transcription Regulatory Relationships Unravelled Sentence-based Text mining (TRRUST) all showed gene list enrichments. A discovery platform called DisGeNET (https://www.disgenet.org/) houses one of the most publicly accessible libraries of genes and variations linked to human diseases [26]. A free database called PaGenBase (https://bioinf.xmu.edu.cn/PaGenBase/) contains information on the pattern genes of eleven model organisms that have been discovered using serial gene expression

profiles under various physiological conditions [27]. A manually maintained library of transcriptional regulatory networks in humans and mice is called TRRUST (https://www.grnpedia.org/trrust/) [28]. The enrichment background comprised all of the genome's genes. Terms that met the following criteria were gathered and clustered: membership similarities, P value < 0.01, minimum count of 3, and enrichment factor (the ratio between the observed counts and the counts expected by chance) > 1.5.

The "org.Hs.eg.db" and "clusterProfiler" packages in R were used to perform the gene ontology (GO) functional enrichment analyses and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for the DEGs [29, 30]. GO functional enrichments comprised molecular function (MF), cellular component (CC), and biological process (BP). Enrichment was statistically significant at a q value < 0.05. The outcomes of these enrichment analyses were visualized using R's ggplot2 package.

Establishment of a PPI network

To develop a PPI network, the DEGs were incorporated into the STRING database (https://string-db.org/). STRING contains known and projected PPIs. The interactions are a combination of direct and indirect linkages that come from the sharing of knowledge between organisms, computational prediction, and the compilation of interactions from other databases [31]. The PPI network was constructed with "homo sapiens" as the study species and a minimum interaction value of 0.4.

Identification of important feature genes and construction of an ANN model

The "randomForest" package was then utilized to perform an RF analysis with the parameter (number of decision trees) set to 500. We then filtered the DEGs to determine which nodes had the lowest cross-validation errors, which we then selected as the parameter for the final model. Genes with importance scores > 1.0 were considered IPF key feature genes, and a subset of significant genes were found to have importance scores. The "pheatmap" package was utilized to visualize significant feature genes and group the data based on their expression levels.

We scored the DEGs according to their expression concerning the median value to remove batch effects between cohorts. Genes that were upregulated were given a score of 1 if their levels were higher than the median. Otherwise, they received a score of 0. The opposite trend was seen in the score when this gene was down-regulated. Using gene scores, we developed an ANN model to diagnose IPF. Three layers make up the ANN: an output, a hidden, and an input layer. In this

stage, the R packages "neuralnet" and "NeuralNetTools" were utilized [32, 33].

Evaluation of the ANN model
The gene cohort was tested and validated using the same methodology, which was also utilized to assess the IPF model's diagnostic accuracy. Using the "pROC" package, we created ROC curves for each of the two cohorts to assess the effectiveness of the ANN model. The true positive rate, or "Sensitivity," is represented by the vertical scale in the ROC curve, whereas the horizontal axis represents the false positive rate, or "1-Specificity." The area under the curve (AUC) showed how accurate the model was.

Discovery of immune cell infiltration characteristics
To quantify the relative proportions of infiltrating immune cells from the gene expression profiles in IPF, a bioinformatics algorithm called CIBERSORT (https://cibersortx.stanford.edu/) was used to calculate immune cell infiltration characteristics. CIBERSORTx is an analytical tool from the Alizadeh Lab and Newman Lab to impute gene expression profiles and provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data [34, 35]. Based on a reference set of 22 immune cell subtypes (download the LM22 Signature Matrix file from CIBERSORTx), 1,000 permutations were used to calculate immune cell abundance.

Distribution and correlation analyses of 22 different types of invading immune cells were performed using the R "corrplot" package. To illustrate how the immune cell infiltration of the IPF and control samples differed, plots were generated using the R package.

Statistical analysis
We used RGui 4.2.3 for all statistical analyses. DEGs were compared between IPF and control samples using an adj P value < 0.05 and |log2FC| > 2. We collected terms having a P value < 0.01, a minimum count of 3, and an enrichment factor > 1.5 from DisGeNET, PaGenBase, and TRUST ontologies. For GO functional enrichment and KEGG pathway enrichment, a q value < 0.05 indicated statistical significance. The last interaction value in the

PPI network was set as 0.4. The feature genes' diagnostic efficacy was assessed using ROC curve analysis and AUC value. In continuous variable group comparisons, the Student's t-test was used for normally distributed data and the Mann-Whitney U for abnormally distributed variables. $P < 0.05$ was considered significant for all two-sided statistical analyses.

Results

Identification of DEGs in merged dataset cohort
Following the merge of three datasets (GSE110147, GSE21369, and GSE24206), batch effects were preprocessed and eliminated using the "SVA" package's combat function to produce a merged dataset cohort. Using the "limma" package, the DEGs of the merged dataset were tested. Using adj P value < 0.05 and |log2FC| > 2.0 thresholds, 47 DEGs were identified, with 11 downregulated and 36 upregulated (Table 1, Supplementary File 2). Figure 1A illustrates the heatmap depicting the expression levels of the eleven downregulated DEGs and thirty-six upregulated DEGs. Additionally, Fig. 1B illustrates the volcano plot of these DEGs.

Prediction of the disease spectrum and function of DEGs
The DisGeNET enrichment analysis summary showed that IPF was linked to lung diseases (interstitial), lung diseases, and connective tissue diseases (Fig. 2A). Summary of enrichment analysis in PaGenBase showed tissues and cells were related to IPF such as lung, bronchial epithelial cells, and trachea (Fig. 2B). The summary of enrichment analysis in TRRUST showed IPF-related transcription factors, including SP1, STAT3, TFAP2A, BRCA1, REAL, NFKB1, and JUN (Fig. 2C).

GO functional and KEGG pathway enrichment analyses
The GO BP enrichment analysis revealed that the DEGs were remarkably enriched in various biological processes including extracellular matrix (ECM) organization, extracellular structure organization, external encapsulating structure organization, collagen fibril organization, response to nutrient, antimicrobial humoral immune response mediated by antimicrobial peptide, humoral immune response, collagen metabolic process, organ or tissue specific immune response, and blood coagulation.

Table 1 47 DEGs in merged dataset cohort

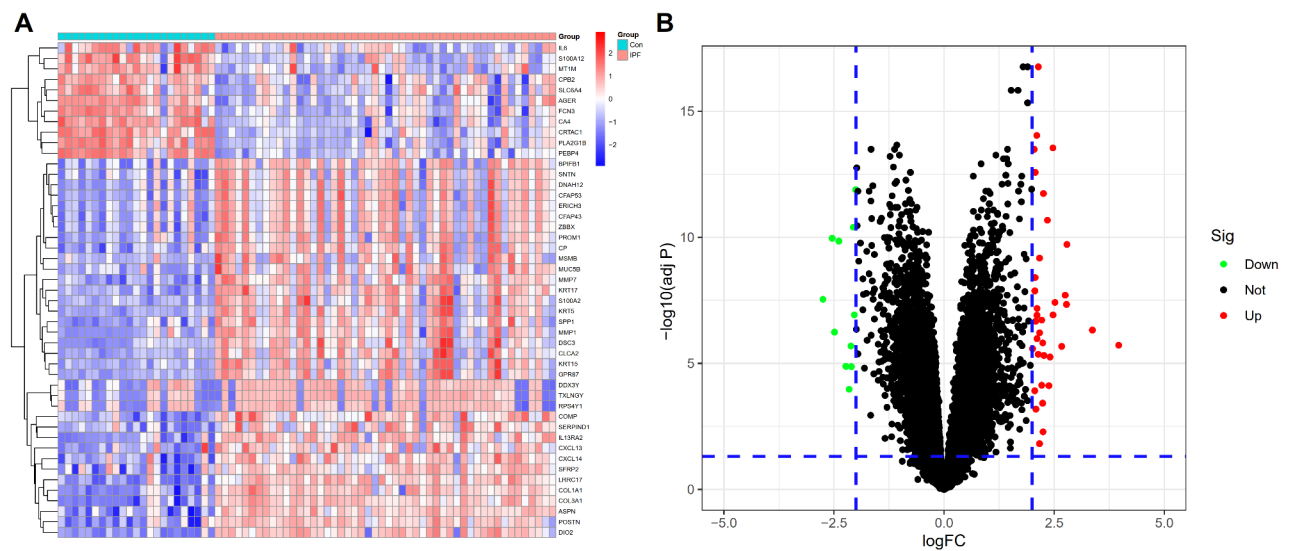| Downregulated DEGs	*S100A12*	*PLA2G1B*	*FCN3*	*CA4*	*IL6*	*SLC6A4*
	MT1M	*CPB2*	*PEBP4*	*AGER*	*CRTAC1*	
Upregulated DEGs	*CFAP53*	*ASPN*	*ERICH3*	*CXCL14*	*KRT17*	*COL1A1*
	SERPIND1	*TXLNGY*	*LRRC17*	*SFRP2*	*KRT15*	*DSC3*
	CFAP43	*COL3A1*	*RPS4Y1*	*COMP*	*S100A2*	*SNTN*
	CLCA2	*MSMB*	*CXCL13*	*DDX3Y*	*IL13RA2*	*MUC5B*
	POSTN	*PROM1*	*ZBBX*	*DIO2*	*CP*	*GPR87*
	DNAH12	*KRT5*	*SPP1*	*MMP7*	*MMP1*	*BPIFB1*

Fig. 1 DEGs in merged dataset. (**A**) The expression levels of the 11 downregulated DEGs and 36 upregulated DEGs in the merged dataset. Control samples (Con) and IPF samples (IPF) showed varied expression levels. Blue denotes low expression, whereas red denotes high expression. (**B**) The volcano plot presents 11 downregulated DEGs and 36 upregulated DEGs in the merged dataset. The thresholds were established at |log2FC| > 2.0 and adj $P < 0.05$; the genes upregulated and downregulated in the IPF samples are shown by the red (Up) and green (Down) dots respectively; genes that do not exhibit a difference in expression between the IPF and normal samples are represented by the black dots (Not)
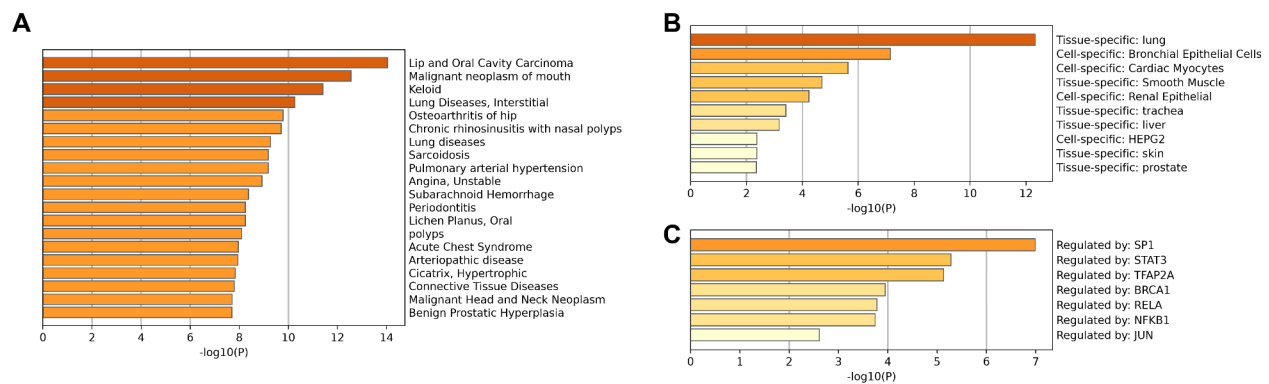
Fig. 2 Enrichment analyses using Metascape. (**A**) Summary of enrichment analysis in DisGeNET. (**B**) Summary of enrichment analysis in PaGenBase. (**C**) Summary of enrichment analysis in TRRUST. Terms that met the following criteria were gathered and clustered: membership similarities, *P* value < 0.01, minimum count of 3, and enrichment factor > 1.5

The DEGs were considerably abundant in collagen-containing ECM, endoplasmic reticulum lumen, fibrillar collagen trimer, banded collagen fibril, collagen trimer, and complex of collagen trimers, according to the GO CC enrichment analysis. The results of the GO MF enrichment analysis demonstrated that the DEGs exhibited a significant enrichment in the following functional domains: ECM structural constituent, platelet-derived growth factor binding, integrin binding, heparin binding, calcium-dependent protein binding, metallopeptidase activity, metalloendopeptidase activity, cytokine activity, glycosaminoglycan binding, growth factor binding, and other functions (Supplementary File 3A). The top 10 GO

functional enrichments ranked by q value are shown in Fig. 3A.

The analysis of the KEGG pathway enrichment revealed that the DEGs exhibited a high enrichment in advanced glycation end products (AGE)-receptor for AGE (RAGE) signaling pathway in diabetic complications signaling pathway, ECM−receptor interaction, interleukin 17 (IL-17) signaling pathway, viral protein interaction with cytokine and cytokine receptor, pancreatic secretion, amoebiasis, protein digestion and absorption(Supplementary File 3B). The seven KEGG pathway enrichments ranked by q value are shown in Fig. 3B.
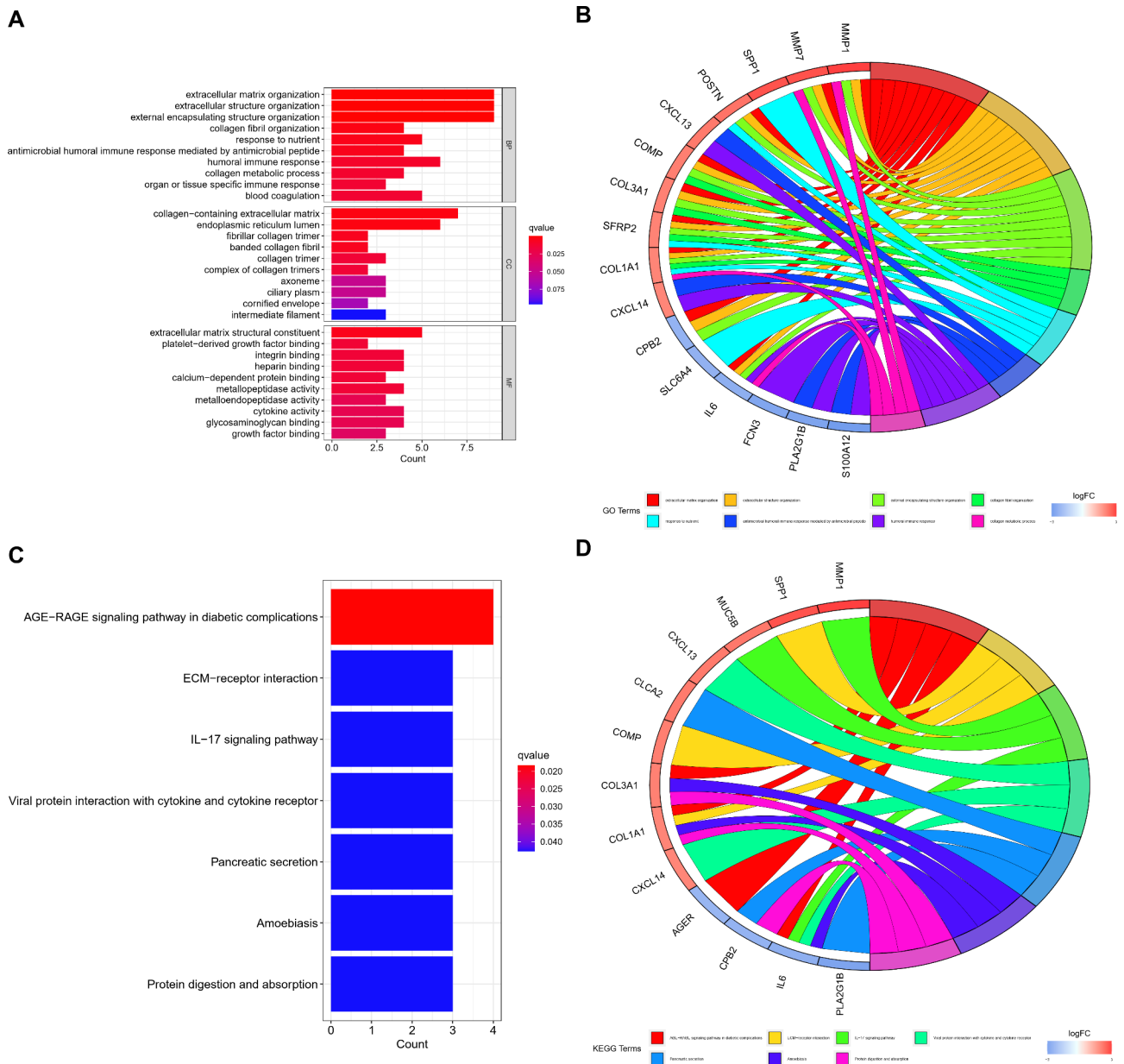
Fig. 3 GO functional and KEGG pathway enrichment analyses. (**A**) Top 10 GO functional enrichments ranked by q value. BP: biological process, CC: cellular component, MF: molecular function. (**B**) Chord plot of GO BP. The top eight GO BP functional enrichments are represented by the GO terms, and the enriched genes are indicated by the gene names with the relationship. (**C**) The nine KEGG pathway enrichments ranked by q value. (**D**) Chord plot of KEGG. The top eight KEGG pathway enrichments are shown by the KEGG terms, and the enriched genes are indicated by the gene names with the connection

PPI network construction

Using the STRING database, we built a PPI network to examine the interactions between the 47 DEGs in more detail. The network has 46 nodes for target proteins and 83 edges for protein interactions when the lowest interaction score was 0.40 (Supplementary file 4, Fig. 4).

The network's 46 targets and 83 edges showed target interactions when setting the lowest interaction score to 0.40.The increase in the degree value is directly related to the extent of connections.

Selection of important genes using RF analysis

To identify key feature genes on 47 DEGs, RF analysis was performed. The number of decision trees was determined using cross-validation error. It was determined that the cross-validation error was minimized at 39 decision trees. As the final model parameter, 39 decision trees were subsequently selected (Fig. 5A). Following this, a subset of significant genes was identified and assigned importance scores; the 30 most important genes, arranged in ascending order of importance scores, are displayed in Fig. 5B. Among them, leucine-rich repeat
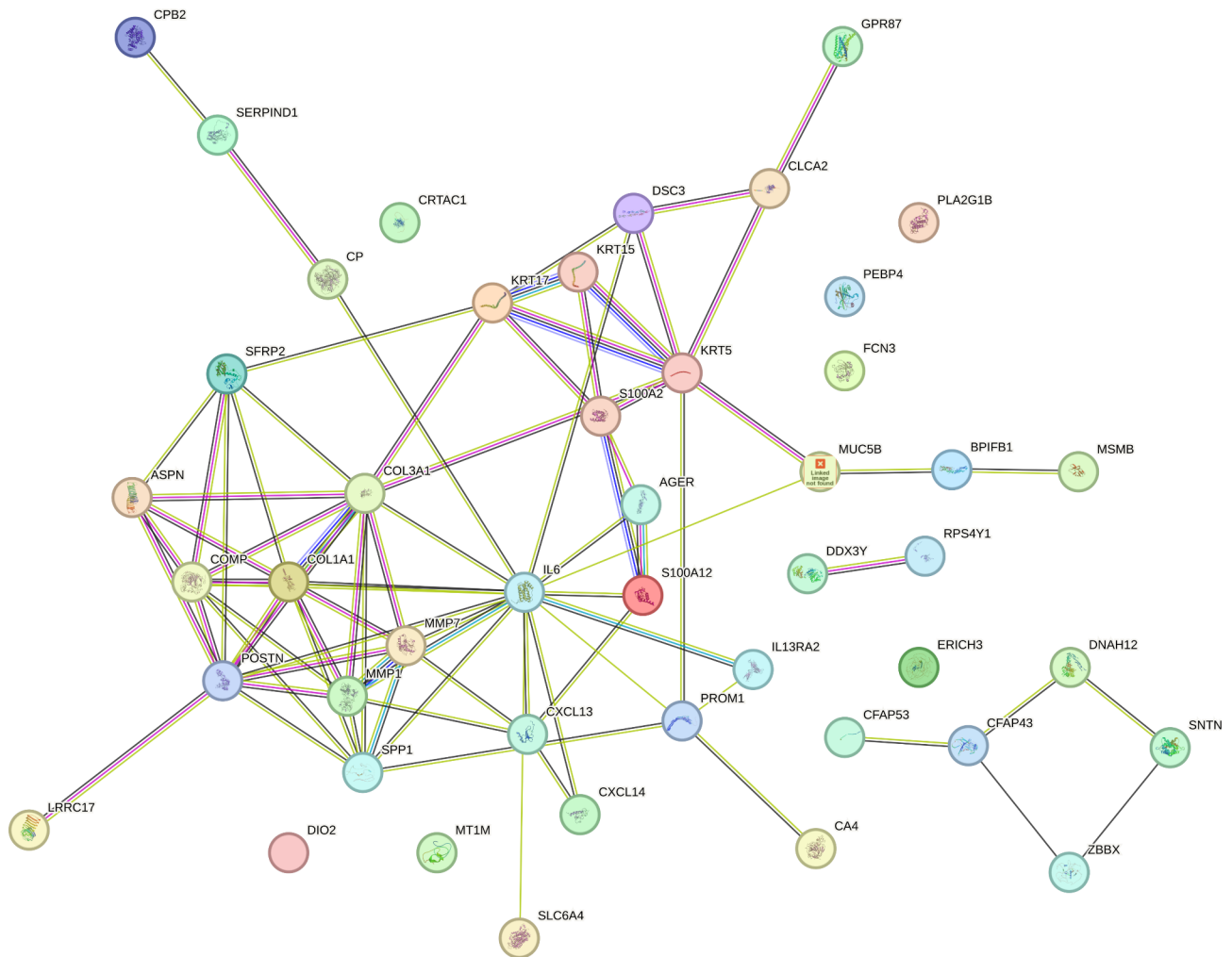
Fig. 4 PPI network

containing 17 (LRRC17), cartilage oligomeric matrix protein (COMP), asporin (ASPN), cartilage acidic protein 1 (CRTAC1), collagen type III alpha 1 chain (COL3A1), periostin (POSTN), phosphatidylethanolamine binding protein 4 (PEBP4), interleukin 13 receptor subunit alpha 2 (IL13RA2), and carbonic anhydrase 4 (CA4) with importance scores>1.0 were identified as feature genes for subsequent analysis. The heatmap presenting nine important feature genes is visualized in Figure S1.

Construction of an ANN model for IPF

Our score for the nine feature genes was their expression relative to the median. ANN was used to develop a diagnostic prediction model with three layers: input, hidden, and output, using the nine feature gene scores (Supplementary file 5A). To develop the ANN model, a deep machine-learning algorithm was performed using the feature gene weight. ANN model output data showed that the training method was repeated 114 times (the number of iterations), which was automatically selected

by the ANN algorithm (Figure S2). The ANN model based on gene scores is constructed as shown in Fig. 6A, where the hidden layer displaying genes relevant to IPF was connected to the input layer containing genes for several groups depending on the scores and weights that were obtained. Five nodes were found to be present in the hidden layer. Based on these five nodes and their respective weights, we obtained the output layer, which was the attribute of the sample.

The accuracy of the ANN model in predicting IPF is detailed in Tables 2 and 3, respectively, for the training and testing sets. Figure 6B shows the predictive model's AUC was 1.000 [95% confidence interval (CI) 1.000–1.000]. This value signifies that the model demonstrated a remarkable ability to predict IPF. The ANN model was utilized to detect feature genes in the assessment set that were identical to those found in the training set (Supplementary file 5B). The testing set AUC was 0.936 (95% CI 0.894–0.971), showing the ANN model's reliability and stability (Fig. 6C). The heatmap presenting nine
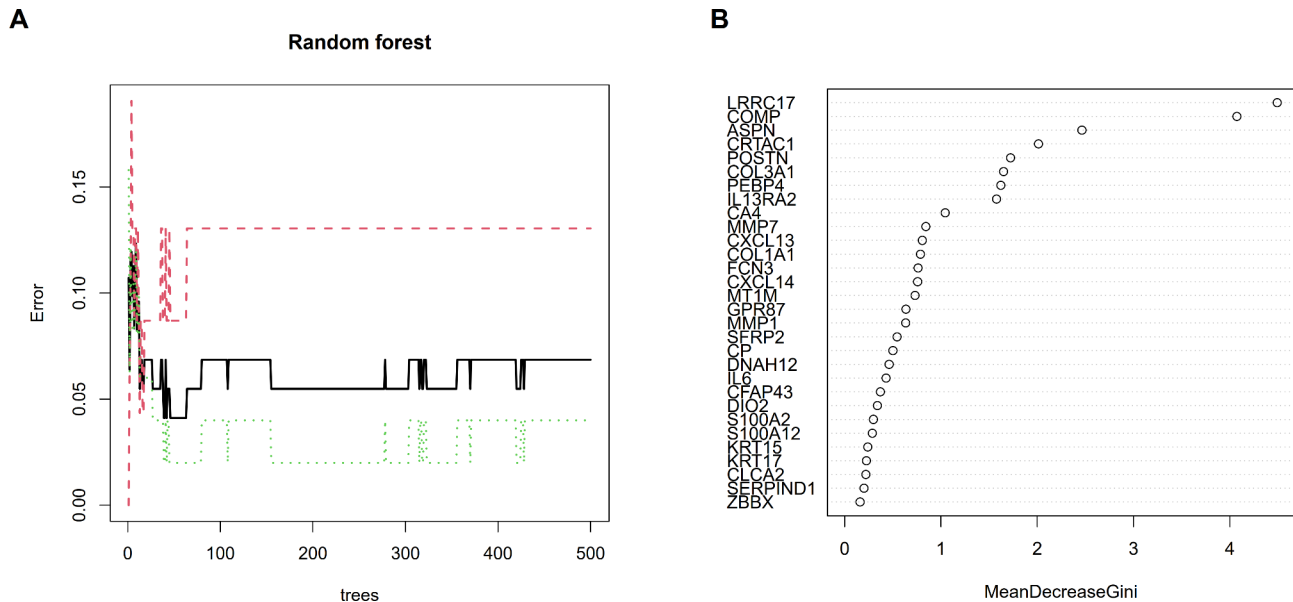
A

Random forest

B

Fig. 5 Identification of candidate important genes by RF analysis. (**A**) Effect on the error rate of the quantity of decision trees. The number of decision trees (trees) is denoted along the x-axis, whereas the error rate (Error) is represented along the y-axis. The black lines indicate the error values for all samples. (**B**) The 30 most significant genes as determined using RF analysis. Critical feature genes were identified in compliance with the specifications of the RF algorithm. MeanDecreaseGini represents the mean Gini index decrease value. A larger value indicates the more important of the variable

important feature genes in the testing set is visualized in Fig. 7A and the expression of nine important feature genes between IPF tissues and normal control tissuesin the testing set is visualized in Fig. 7B. These results were consistent with those of differential expression analysis in the metadata cohort.

Immune cell infiltration

The CIBERSORT bioinformatics algorithm was utilized to assess immune cell abundance using the LM22 signature matrix file with 1,000 permutations after downloading it (Supplementary File 6A). The results of CIBERSORT are presented in Supplementary File 6B.

Figure 8A shows the findings of the distribution analysis of 22 immune cell types in the IPF and control groups. Figure S3 shows immune cell correlation. Next, we investigated the immune cells that differed between IPF tissues and normal control tissues. IPF tissues had significantly decreased levels of T cells CD8, monocytes ($P=0.009$), natural killer (NK) cells resting ($P<0.001$), macrophages M1 ($P=0.010$), and neutrophils ($P=0.028$) compared to normal tissues. However, IPF tissues had significantly greater proportions of T cells CD4 memory resting ($P=0.020$), macrophages M0 ($P<0.001$), and mast cells resting ($P=0.028$) compared to normal tissues (Fig. 8B).

Discussion

IPF is an interstitial disease in which UIP is its primary pathological manifestation. IPF remains incurable and has a dismal prognosis at this time. The precise

mechanism by which IPF occurs and progresses remains poorly understood, despite the publication of numerous studies in the field [36]. The onset and progression of IPF may be influenced by epithelial-mesenchymal transition, ECM deposition, and pulmonary remodeling [37–39].

Patients frequently miss their best chance for treatment since there are no early diagnostic markers for IPF, which causes the disease to progress more quickly. It is essential to delve into the molecular mechanisms of IPF onset and progression, along with pinpointing the treatment target for the disease. Recent studies suggest that immune cell infiltration may play a major role in the development and progression of IPF and have the ability to eradicate aged alveolar epithelial cells [40, 41].

However, studies into the immune infiltration and abnormally expressed genes that distinguish IPF from normal tissues are limited. Initially, we employed microarray technology to gather three analogous cohorts from the GEO datasets. Subsequently, we conducted a merged dataset cohort comprising 23 control samples and 50 IPF samples. In total, 47 DEGs were found, 11 downregulated and 36 upregulated, which was consistent with the previous differential gene analyses [12]. The enrichment analyses showed that they were linked to IPF-related transcription factors, cells and tissues, and illnesses. The PPI network showed the interaction between these DEGs. The primary GO functional enrichments were associated with ECM, suggesting that these DEGs contribute to the formation of IPF and are intimately related to ECM [36–38]. Significant KEGG pathway enrichments
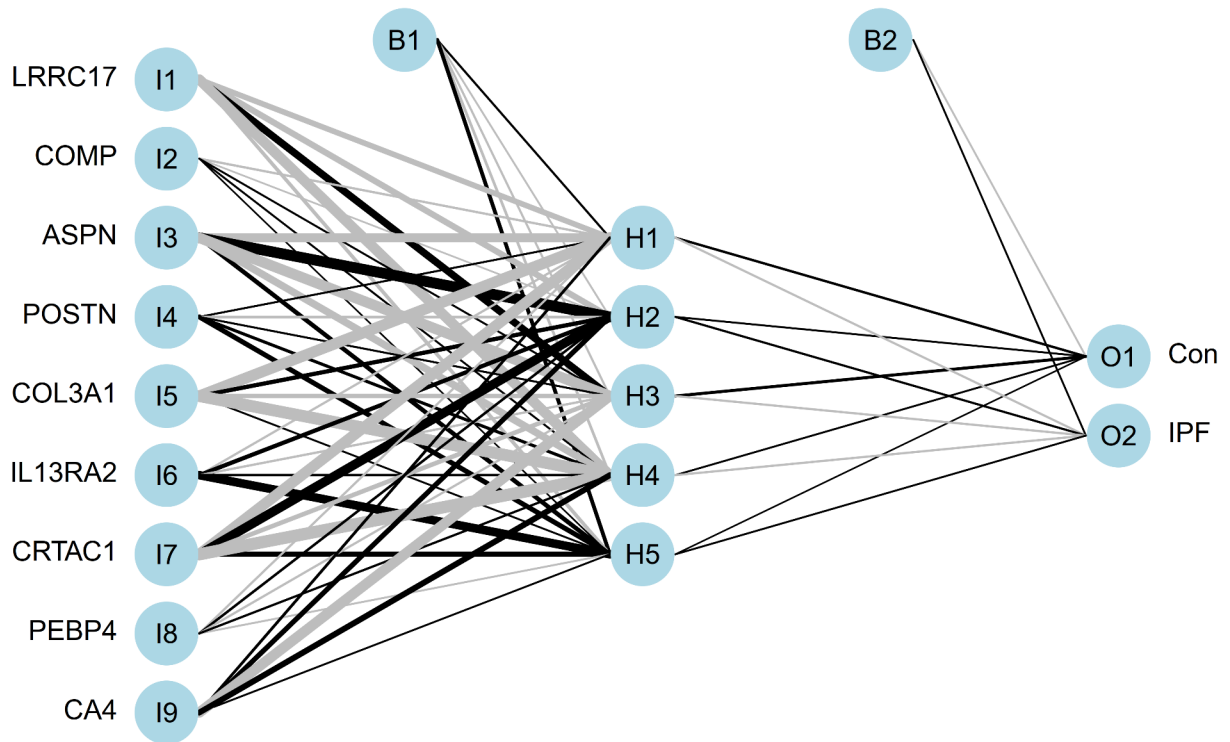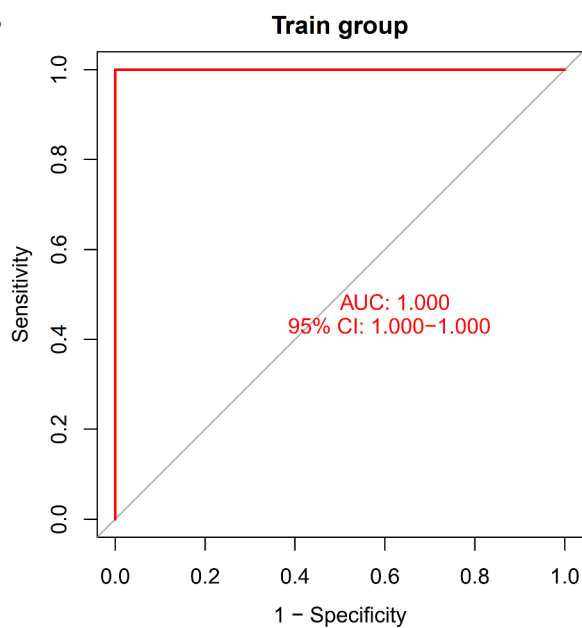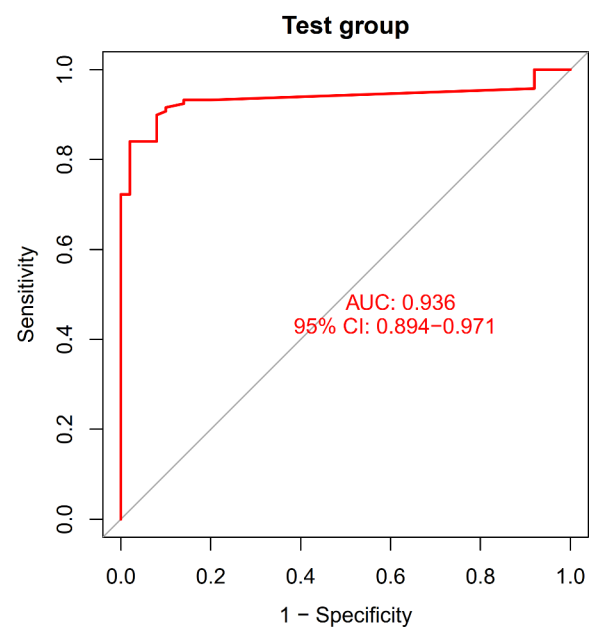
Fig. 6 The ANN model of the nine important genes for IPF. (**A**) Gene score-based ANN model generation. Three layers make up the ANN: an output (O1,O2), a hidden (H1-H5), and an input (I1-I9) layer. (**B**) The predictive model (Train group) AUC was 1.000 (95% CI 1.000–1.000). (**C**) Testing set (Test group) AUC was 0.936 (95% CI 0.894–0.971)

Table 2 IPF prediction accuracy of the ANN model in the training set

	Control	IPF	Total
Control	23	0	23
IPF	0	50	50

Table 3 IPF prediction accuracy of the ANN model in the testing set

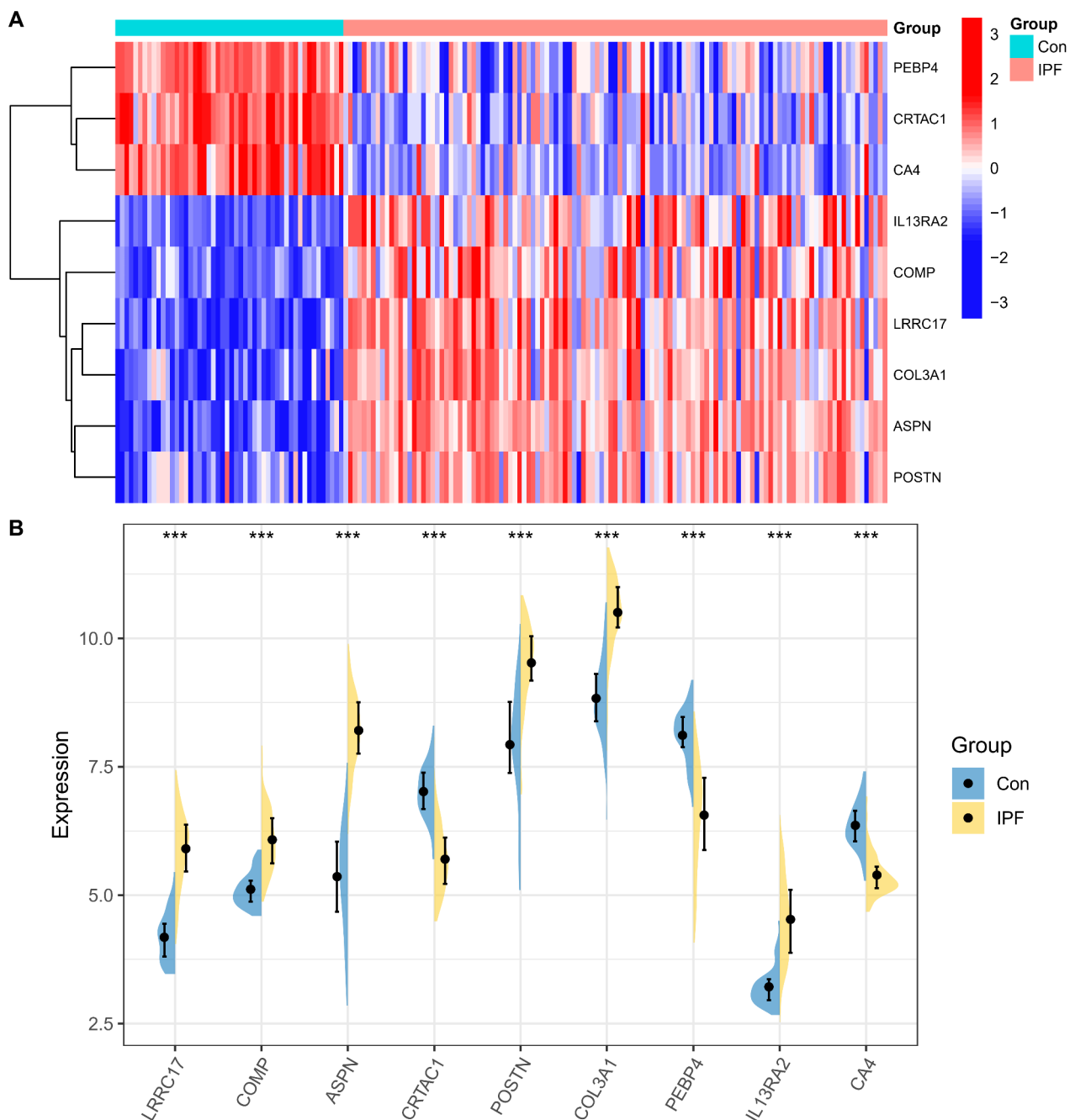|	Control	IPF	Total
Control	46	4	50
IPF	13	106	119

Fig. 7 Validation of the expression of the nine important genes in the GSE32537 dataset. (**A**) The heatmap presenting nine important feature genes in the testing set. Control samples (Con) and IPF samples (IPF) showed varied expression levels. Blue denotes low expression, whereas red denotes high expression. (**B**) The expression of nine important feature genes between IPF tissues and normal control tissues in the testing set. Control (Con) and IPF samples (IPF) are represented by blue and yellow colors correspondingly. *** $P < 0.05$

were observed in the following domains: IL-17 signaling pathway, AGE-RAGE signaling pathway, ECM-receptor interaction, pancreatic secretion, amoebiasis, viral protein interaction with cytokine and cytokine receptor, and protein digestion and absorption. These major pathways were also related to ECM and immune response,

including the most important pathways that are highly relevant and enriched in IPF such as transforming growth factor β (TGF-β), mitogen-activated protein kinase (MAPK), phosphatidylinositol 3 kinase (PI3K)-protein kinase B (Akt), and nuclear factor κB (NF-κB) signaling pathways.
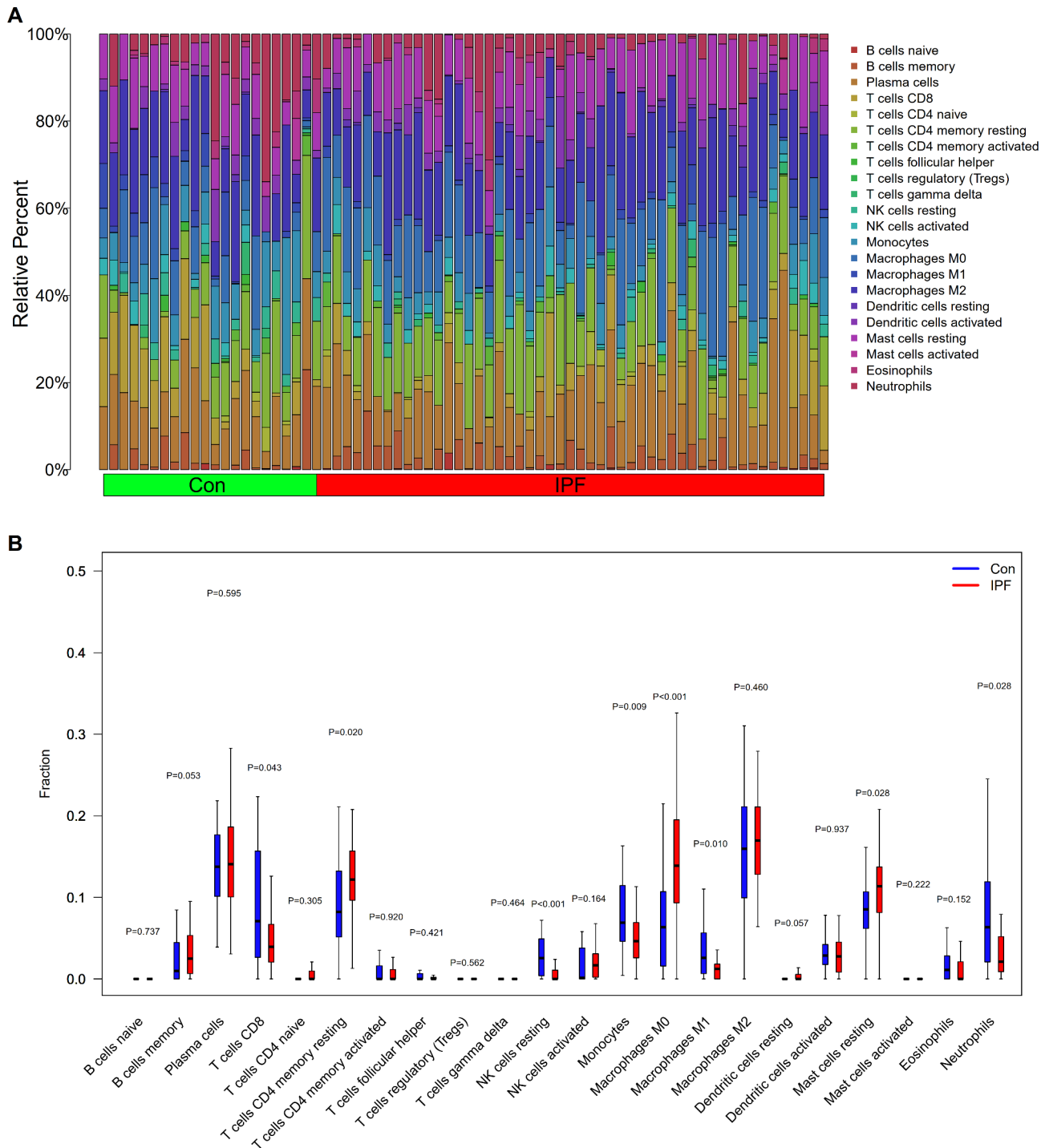
Fig. 8 Distribution and difference of immune cell infiltration. (**A**) The distribution analysis of 22 immune cell types in IPF samples (IPF) and control samples (Con). (**B**) The differential immune cells in IPF tissues comparing normal control tissues. Control (Con) and IPF samples (IPF) are represented by blue and red colors, correspondingly

Then, with the rapid development of science and technology, RF analysis and ANN model were used to identify important feature genes and establish a diagnostic model. The CIBERSORT instrument was utilized to investigate the involvement of immune cell infiltration features in IPF.

Using RF analysis, nine important feature genes were identified. Six upregulated genes were *LRRC17, COMP, ASPN, POSTN, COL3A1,* and *IL13RA2,* and three

downregulated genes were *CRTAC1*, *PEBP4*, and *CA4*. Therefore, the nine genes were constructed and validated as a prediction ANN mode. The results obtained from conducting the ROC and AUC analyses suggested that all nine genes possessed a significant potential in disease diagnosis.

It is anticipated that LRRC17 contributes to the development of bone marrow, negatively regulates osteoclast differentiation, and is active in ECM and extracellular space [42, 43]. *COMP* encodes a noncollagenous ECM protein [44]. The most intriguing clinical application of *COMP* is its utilization as a biomarker for IPF. *COMP* is a large pentameric glycoprotein that interacts with numerous ECM proteins in cartilage and other tissues [45, 46]. *ASPN* encodes a small leucine-rich proteoglycan cartilage extracellular protein [47]. Tissue regeneration and development are facilitated by a secreted ECM protein encoded by *POSTN* [48]. *ASPN* and *POSTN* may act as hub genes regulating pulmonary fibrosis [49]. *ASPN* promotes the differentiation of lung myofibroblasts induced by TGF-β by facilitating the recycling of TβRI, which is dependent on Rab11 [50]. Periostin is a useful biomarker for type 2 inflammation and pulmonary fibrosis [51]. In extensible connective tissues, *COL3A1* encodes type III collagen pro-alpha1 chains [52]. Dysregulated expression of COL3A1 might impact the development of IPF through modulating IPF-related biological processes and the expression level of COL3A1 is correlated with IPF prognosis [53]. *COL3A1* could serve as a biomarker for IPF and non-small cell lung cancer progression [54]. The protein encoded by IL13RA2, which is closely linked to IL13RA1, binds IL13 with high affinity and helps internalize it [55]. The induction of fibrotic markers by IL-13 in vitro is impeded by the overexpression of IL-13Ralpha2, which also prevents bleomycin-induced pulmonary fibrosis [56]. *CRTAC1* is responsible for producing a glycosylated ECM protein located in the interterritorial matrix of articular deep zone cartilage [57]. *CRTAC1* serves as a biomarker for the health status of alveolar type-2 epithelial cells in lavage fluid and plasma [58]. Protidylethanolamine-binding proteins, which comprise PEBP4, are a family of proteins that have undergone significant evolutionary conservation. These proteins play critical biological roles, including lipid binding and serine protease inhibition [59]. The glycosylphosphatidyl-inositol-anchored membrane isozyme CA4 is encoded by *CA4*. This isozyme is expressed on the proximal renal tubules and luminal surfaces of pulmonary capillaries [60]. While there are currently no IPF-related genes deserving further inquiry, these genes are linked to the disease and should be thoroughly investigated.

After the nine feature genes were included in the ANN, a diagnostic prediction model was developed, which exhibited outstanding IPF prediction performance. It has the potential to accurately differentiate IPF samples from normal samples, which will be crucial for the IPF diagnosis.

We utilized CIBERSORT to analyze immune cell infiltration in normal and IPF samples. Consequently, it was discovered that certain immune cell subtypes were intimately connected to significant BPs of IPF. It was found that there was an increase in mast cells, macrophages M0, and T cells CD4 memory resting in IPF tissues in comparison to normal tissues, and a decrease in the infiltration of monocytes, neutrophils, NK cells resting, and T cells CD8. These processes may be linked to the onset and progression of IPF. There are similar differences in other chronic lung diseases, our next research is to further analyze feature genes in order to find immune cell gene targets specific to IPF.

Indeed, it has been demonstrated previously that immunological and inflammatory cells are crucial to the development of IPF. A few of the findings line up with earlier research. The pathological result of suboptimal wound healing after a lung injury is IPF. M1 macrophages repair wounds after alveolar epithelial injury, while M2 macrophages resolve lung inflammation [61]. NF-κB exacerbates M1 macrophage polarization by promoting the release of proinflammatory cytokines [62]. According to research, polarized M1 macrophages cultured in a distinct polarizing medium can redifferentiate into a different cell phenotype or revert to M0 macrophages after 12 days in a cytokine-deficient medium [63]. NK cell resting percentage was lower in IPF tissue samples than in controls [64]. The interest in immunological dysregulation in IPF has been rekindled by recent publications emphasizing the prognostic and mechanistic roles of monocytes and monocyte-derived alveolar macrophages [65]. BLT1 mediates bleomycin-induced lung fibrosis independently of neutrophils and CD4+T Cells [66]. It may be possible to use these differentiated immune cells as targets for immunotherapy in patients with IPF.

A genomic classifier was developed with machine learning and whole transcriptome RNA sequencing using lung tissue obtained by biopsy. It was introduced and validated for lung tissue obtained by transbronchial forceps biopsy. Genetic testing of lung tissue can increase the multidisciplinary discussion of confidence in distinguishing diagnostic IPF from non-IPF. However, because there are few studies on genetic testing of lung tissue biopsy, the sensitivity of genetic testing is low, and it is prone to false negatives, more clinical studies are needed to further evaluate its sensitivity and specificity [1, 2].

Given the above results, we can detect the nine feature genes and increase confidence in IPF early diagnose .The detection of the nine feature genes before and after treatment in patients with a definite diagnosis of IPF to further validate our model. The efficacy after treatment and

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:458

Page 13 of 15

expression changes in the nine feature genes, combined with immune cell infiltration, provide a basis for further investigation of treatment-related mechanisms.

The study has limitations, despite our best efforts to conduct it properly. These should be noted as well. Even though we merged the three datasets to acquire as many samples as feasible, the metadata cohort requires more samples. Second, the validation cohort sample size must be raised. Ultimately, the roles of immune cell infiltration and nine feature genes in IPF were inferred from bioinformatics analysis. However, additional experimental study is required to validate these findings.

Conclusion

In conclusion, it was determined that key IPF feature genes included *LRRC17, COMP, ASPN, CRTAC*1, *POSTN, COL3A1, PEBP4, IL13RA2,* and *CA4.* The ability to accurately identify between IPF samples and normal samples is made possible by the nine feature genes ANN model's superiority, and this will be crucial for the diagnosis of IPF. Immune cells that differ between IPF and normal samples may have a role in the onset of the disease and may one day be the focus of immunotherapy for patients with IPF.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12890-024-03249-6.

Supplementary file 1A. Gene expression levels of 22 IPF and 11 control samples from GSE110147

Supplementary file 1B. Gene expression levels of 11 IPF and six control samples from GSE21369

Supplementary file 1C. Gene expression levels of 17 IPF and six control samples from GSE24206

Supplementary file 1D. Gene expression levels of 50 IPF and 23 control samples from the merged dataset

Supplementary file 1E. Gene expression levels of 119 IPF and 50 control samples from GSE32537

Supplementary file 2. 47 DEGs obtained in a merged dataset cohort

Supplementary file 3A. GO functional enrichment analyses

Supplementary file 3B. KEGG pathway enrichment analysis

Supplementary file 4. Interactions in PPI network

Supplementary file 5A. Feature gene scores of the training set

Supplementary file 5B. Feature gene scores of the testing set

Supplementary file 6A. LM22 signature matrix file

Supplementary file 6B. CIBERSORT results

Supplementary Material 14

Author contributions
H.Z.Z., H.B.H., W.L.J., X.D.H. and Y.F.Z. designed the study; H.Z.Z., C.W., C.J.Z., Q.Q.X., W.L.J., X.D.H. and Y.F.Z. analyzed the data and performed the research; H.Z.Z. H.B.H., W.L.J., X.D.H. and Y.F.Z. drafted the manuscript; W.L.J., X.D.H. and Y.F.Z. provided supervision and managed the project. All authors have read and approved the final manuscript.

Data availability
Publicly available datasets were analyzed in this study. Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) (Accessions: GSE110147, GSE21369, GSE24206 and GSE32537). The data generated and/or analyzed during the current study are available from the corresponding author (Yufeng Zhang, Email: yufengzhang@njucm.edu.cn) upon a reasonable request. Main data from this study are also included in this published article (and its Supplementary Information files).

Declarations

Ethics approval and consent to participate
GEO belongs to public databases. The patients involved in the database have obtained ethical approve. Users can download data for free for research and publish relevant articles.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Respiratory Medicine, Yancheng Hospital of Traditional Chinese Medicine; Yancheng TCM Hospital Affiliated to Nanjing University of Chinese Medicine, Yancheng, Jiangsu 224005, China
[2]Department of Gastroenterology, Jiangyin Hospital of Traditional Chinese Medicine; Jiangyin Hospital Affiliated to Nanjing University of Chinese Medicine, Jiangyin, Jiangsu 214400, China
[3]Department of Pulmonary and Critical Care Medicine, Jiangyin Hospital of Traditional Chinese Medicine; Jiangyin Hospital Affiliated to Nanjing University of Chinese Medicine, Jiangyin, Jiangsu 214400, China
[4]Research Institute of Respiratory Diseases, Jiangsu Province Clinical Academy of Traditional Chinese Medicine (Jiangyin Branch), Jiangyin, Jiangsu 214400, China

References
1. Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2022;205(9):e18–47. https://doi.org/10.1164/rccm.202202-0399ST.
2. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2018;198(5):e44–68. https://doi.org/10.1164/rccm.201807-1255ST.

Zhang *et al. BMC Pulmonary Medicine*   (2024) 24:458

Page 14 of 15

3. Zhang Y, Gu L, Xia Q, Tian L, Qi J, Cao M. Radix Astragali and Radix Angelicae sinensis in the treatment of idiopathic pulmonary fibrosis: a systematic review and Meta-analysis. FRONT PHARMACOL. 2020;11:415. https://doi.org/10.3389/fphar.2020.00415.

4. Zhang H, Wang C, Zhang Y. Progress of Radix Astragali and Radix Angelicae sinensis in the treatment of idiopathic pulmonary fibrosis. TMR Integr Med. 2022;6:e22001–36. https://doi.org/10.53388/TMRIM202206024.

5. Enomoto N, Naoi H, Aono Y, et al. Acute Exacerbation of unclassifiable idiopathic interstitial pneumonia: comparison with idiopathic pulmonary fibrosis. THER ADV RESPIR DIS. 2020;14:1022296482. https://doi.org/10.1177/1753466620935774.

6. Biondini D, Balestro E, Sverzellati N, Cocconcelli E, Bernardinello N, Ryerson CJ, Spagnolo P. Acute exacerbations of idiopathic pulmonary fibrosis (AE-IPF): an overview of current and future therapeutic strategies. Expert Rev Respir Med. 2020;14(4):405–14. https://doi.org/10.1080/17476348.2020.1724096.

7. Podolanczuk AJ, Thomson CC, Remy-Jardin M, Richeldi L, Martinez FJ, Kolb M, Raghu G. Idiopathic pulmonary fibrosis: state of the art for 2023. EUR RESPIR J. 2023;61(4). https://doi.org/10.1183/13993003.00957-2022.

8. Kaur A, Mathai SK, Schwartz DA. Genetics in Idiopathic Pulmonary Fibrosis Pathogenesis, Prognosis, and treatment. Front Med (Lausanne). 2017;4:154. https://doi.org/10.3389/fmed.2017.00154.

9. Stainer A, Faverio P, Busnelli S, Catalano M, Della ZM, Marruchella A, Pesci A, Luppi F. Molecular biomarkers in idiopathic pulmonary fibrosis: state of the art and future directions. INT J MOL SCI. 2021;22(12). https://doi.org/10.3390/ijms22126255.

10. Huang G, Xu X, Ju C, Zhong N, He J, Tang XX. Identification and validation of autophagy-related gene expression for Predicting Prognosis in patients with idiopathic pulmonary fibrosis. FRONT IMMUNOL. 2022;13:997138. https://doi.org/10.3389/fimmu.2022.997138.

11. He J, Li X. Identification and validation of aging-related genes in idiopathic pulmonary fibrosis. FRONT GENET. 2022;13:780010. https://doi.org/10.3389/fgene.2022.780010.

12. Zhang Y, Wang C, Xia Q, Jiang W, Zhang H, Amiri-Ardekani E, Hua H, Cheng Y. Machine learning-based prediction of candidate gene biomarkers correlated with Immune Infiltration in patients with idiopathic pulmonary fibrosis. Front Med (Lausanne). 2023;10:1001813. https://doi.org/10.3389/fmed.2023.1001813.

13. Jee AS, Sahhar J, Youssef P, Bleasel J, Adelstein S, Nguyen M, Corte TJ, Review. Serum biomarkers in idiopathic pulmonary fibrosis and systemic sclerosis Associated interstitial lung Disease - Frontiers and Horizons. Pharmacol Ther. 2019;202:40–52. https://doi.org/10.1016/j.pharmthera.2019.05.014.

14. Harrell CR, Sadikot R, Pascual J, Fellabaum C, Jankovic MG, Jovicic N, Djonov V, Arsenijevic N, Volarevic V. Mesenchymal Stem Cell-Based Therapy of Inflammatory Lung Diseases: Current Understanding and Future Perspectives. STEM CELLS INT. 2019; 2019:4236973. https://doi.org/10.1155/2019/4236973

15. Desai O, Winkler J, Minasyan M, Herzog EL. The role of Immune and Inflammatory cells in idiopathic pulmonary fibrosis. Front Med (Lausanne). 2018;5:43. https://doi.org/10.3389/fmed.2018.00043.

16. Dey P. Artificial neural network in Diagnostic Cytology. CYTOJOURNAL. 2022;19:27. https://doi.org/10.25259/Cytojournal_33_2021.

17. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in Cancer Prognosis and Prediction. Comput Struct Biotechnol J. 2015;13:8–17. https://doi.org/10.1016/j.csbj.2014.11.005.

18. Albaradei S, Thafar M, Alsaedi A, Van Neste C, Gojobori T, Essack M, Gao X. Comput Struct Biotechnol J. 2021;19:5008–18. https://doi.org/10.1016/j.csbj.2021.09.001. Machine Learning and Deep Learning Methods that Use Omics Data for Metastasis Prediction.

19. Li Z, Wang S, Zhao H, Yan P, Yuan H, Zhao M, Wan R, Yu G, Wang L. Artificial neural network identified the significant genes to Distinguish Idiopathic Pulmonary Fibrosis. Sci Rep. 2023;13(1):1225. https://doi.org/10.1038/s41598-023-28536-w.

20. Cecchini MJ, Hosein K, Howlett CJ, Joseph M, Mura M. Comprehensive Gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. Respir Res. 2018;19(1):153. https://doi.org/10.1186/s12931-018-0857-1.

21. Cho JH, Gelinas R, Wang K, et al. Systems Biology of interstitial Lung diseases: integration of mRNA and microRNA expression changes. BMC MED GENOMICS. 2011;4:8. https://doi.org/10.1186/1755-8794-4-8.

22. Meltzer EB, Barry WT, D'Amico TA, et al. Bayesian probit regression model for the diagnosis of Pulmonary Fibrosis: Proof-Of-Principle. BMC MED GENOMICS. 2011;4:70. https://doi.org/10.1186/1755-8794-4-70.

23. Yang IV, Coldren CD, Leach SM, et al. Expression of Cilium-Associated genes defines Novel Molecular subtypes of Idiopathic Pulmonary Fibrosis. Thorax. 2013;68(12):1114–21. https://doi.org/10.1136/thoraxjnl-2012-202943.

24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma Powers Differential expression analyses for RNA-sequencing and microarray studies. NUCLEIC ACIDS RES. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.

25. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-Level datasets. NAT COMMUN. 2019;10(1):1523. https://doi.org/10.1038/s41467-019-09234-6.

26. Pinero J, Sauch J, Sanz F, Furlong LI. Comput Struct Biotechnol J. 2021;19:2960–7. https://doi.org/10.1016/j.csbj.2021.05.015. The DisGeNET Cytoscape App: Exploring and Visualizing Disease Genomics Data.

27. Pan JB, Hu SC, Shi D, Cai MC, Li YB, Zou Q, Ji ZL. PaGenBase: a pattern gene database for the Global and Dynamic understanding of gene function. PLoS ONE. 2013;8(12):e80747. https://doi.org/10.1371/journal.pone.0080747.

28. Han H, Cho JW, Lee S, et al. TRRUST V2: an expanded reference database of Human and Mouse Transcriptional Regulatory Interactions. NUCLEIC ACIDS RES. 2018;46(D1):D380–6. https://doi.org/10.1093/nar/gkx1013.

29. Wu T, Hu E, Xu S, et al. Innov (Camb). 2021;2(3):100141. https://doi.org/10.1016/j.xinn.2021.100141. ClusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data.

30. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R Package for comparing Biological themes among Gene clusters. OMICS. 2012;16(5):284–7. https://doi.org/10.1089/omi.2011.0118.

31. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: Protein-Protein Association Networks and Functional Enrichment analyses for any sequenced genome of interest. NUCLEIC ACIDS RES. 2023;51(D1):D638–46. https://doi.org/10.1093/nar/gkac1000.

32. Chen Y, Xue J, Yan X, Fang DG, Li F, Tian X, Yan P, Feng Z. Identification of crucial genes related to heart failure based on GEO database. BMC Cardiovasc Disord. 2023;23(1):376. https://doi.org/10.1186/s12872-023-03400-x.

33. Yang Y, Xu L, Qiao Y, Wang T, Zheng Q. Construction of a neural Network Diagnostic Model and Investigation of Immune infiltration characteristics for Crohn's Disease. FRONT GENET. 2022;13:976578. https://doi.org/10.3389/fgene.2022.976578.

34. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. NAT METHODS. 2015;12(5):453–7. https://doi.org/10.1038/nmeth.3337.

35. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with Digital Cytometry. NAT BIOTECHNOL. 2019;37(7):773–82. https://doi.org/10.1038/s41587-019-0114-2.

36. Thomson CC, Duggal A, Bice T, Lederer DJ, Wilson KC, Raghu G. 2018 Clinical Practice Guideline Summary for Clinicians: Diagnosis of Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc. 2019; 16(3):285–290. https://doi.org/10.1513/AnnalsATS.201809-604CME

37. James DS, Jambor AN, Chang HY, Alden Z, Tilbury KB, Sandbo NK, Campagnola PJ. Probing ECM remodeling in Idiopathic Pulmonary Fibrosis Via Second Harmonic Generation Microscopy Analysis of Macro/Supramolecular Collagen Structure. J BIOMED OPT. 2019;25(1):1–13. https://doi.org/10.1117/1.JBO.25.1.014505.

38. Siekacz K, Piotrowski WJ, Iwanski MA, Gorski P, Bialas AJ. The Role of Interaction between Mitochondria and the Extracellular Matrix in the development of idiopathic pulmonary fibrosis. OXID MED CELL LONGEV. 2021;2021(9932442). https://doi.org/10.1155/2021/9932442.

39. Tomos IP, Tzouvelekis A, Aidinis V, Manali ED, Bouros E, Bouros D, Papiris SA. Extracellular matrix remodeling in idiopathic pulmonary fibrosis. It is the 'Bed' that counts and not 'The sleepers'. Expert Rev Respir Med. 2017;11(4):299–309. https://doi.org/10.1080/17476348.2017.1300533.

40. Serezani A, Pascoalino BD, Bazzano J, et al. Multiplatform single-cell analysis identifies Immune cell types enhanced in Pulmonary Fibrosis. Am J Respir Cell Mol Biol. 2022;67(1):50–60. https://doi.org/10.1165/rcmb.2021-0418OC.

41. Waters DW, Blokland KS, Pathinayake PS, Burgess JK, Mutsaers SE, Prele CM, Schuliga M, Grainge CL, Knight DA. Fibroblast senescence in the Pathology of Idiopathic Pulmonary Fibrosis. Am J Physiol Lung Cell Mol Physiol. 2018;315(2):L162–72. https://doi.org/10.1152/ajplung.00037.2018.

42. Kim T, Kim K, Lee SH, So HS, Lee J, Kim N, Choi Y. Identification of LRRc17 as a negative Regulator of receptor activator of NF-kappaB ligand (RANKL)-induced osteoclast differentiation. J BIOL CHEM. 2009;284(22):15308–16. https://doi.org/10.1074/jbc.M807722200.

43. Kim D, LaQuaglia MP, Yang SY. A cDNA encoding a putative 37 kDa leucine-rich repeat (LRR) protein, p37NB, isolated from S-type Neuroblastoma Cell has a Differential tissue distribution. Biochim Biophys Acta. 1996;1309(3):183–8. https://doi.org/10.1016/s0167-4781(96)00158-3.

44. Newton G, Weremowicz S, Morton CC, Copeland NG, Gilbert DJ, Jenkins NA, Lawler J. Characterization of human and mouse cartilage oligomeric matrix protein. Genomics. 1994;24(3):435–9. https://doi.org/10.1006/geno.1994.1649.

45. Posey KL, Coustry F, Hecht JT. Cartilage oligomeric matrix protein: COM-Popathies and Beyond. MATRIX BIOL. 2018;71–72:161–73. https://doi.org/10.1016/j.matbio.2018.02.023.

46. Neighbors M, Cabanski CR, Ramalingam TR, et al. Prognostic and predictive biomarkers for patients with idiopathic pulmonary fibrosis treated with Pirfenidone: Post-hoc Assessment of the CAPACITY and ASCEND trials. Lancet Respir Med. 2018;6(8):615–26. https://doi.org/10.1016/S2213-2600(18)30185-1.

47. Lorenzo P, Aspberg A, Onnerfjord P, Bayliss MT, Neame PJ, Heinegard D. Identification and characterization of Asporin. A Novel Member of the leucine-rich repeat protein family closely related to Decorin and Biglycan. J BIOL CHEM. 2001;276(15):12201–11. https://doi.org/10.1074/jbc.M010932200.

48. Gillan L, Matei D, Fishman DA, Gerbin CS, Karlan BY, Chang DD. Periostin secreted by epithelial ovarian carcinoma is a Ligand for alpha(V)beta(3) and alpha(V)beta(5) Integrins and promotes cell motility. CANCER RES. 2002;62(18):5358–64.

49. Zhang Y, Liang J, Cao N, Gao J, Song L, Tang X. Coal Dust nanoparticles Induced Pulmonary Fibrosis by promoting inflammation and epithelial-mesenchymal transition Via the NF-kappaB/NLRP3 pathway driven by IGF1/ROS-mediated AKT/GSK3beta signals. Cell Death Discov. 2022;8(1):500. https://doi.org/10.1038/s41420-022-01291-z.

50. Huang S, Lai X, Yang L, et al. Asporin promotes TGF-beta-induced lung myofibroblast differentiation by facilitating Rab11-Dependent recycling of TbetaRI. Am J Respir Cell Mol Biol. 2022;66(2):158–70. https://doi.org/10.1165/rcmb.2021-0257OC.

51. Ono J, Takai M, Kamei A, Azuma Y, Izuhara K. Pathological roles and clinical usefulness of Periostin in type 2 inflammation and pulmonary fibrosis. Biomolecules. 2021;11(8). https://doi.org/10.3390/biom11081084.

52. Janeczko RA, Ramirez F. Nucleotide and amino acid sequences of the entire human alpha 1 (III) collagen. NUCLEIC ACIDS RES. 1989;17(16):6742. https://doi.org/10.1093/nar/17.16.6742.

53. Wan H, Huang X, Cong P, et al. Identification of hub genes and pathways Associated with Idiopathic Pulmonary Fibrosis via Bioinformatics Analysis. Front Mol Biosci. 2021;8:711239. https://doi.org/10.3389/fmolb.2021.711239.

54. Yao Y, Li Z, Gao W. Identification of hub genes in idiopathic pulmonary fibrosis and NSCLC progression:evidence from Bioinformatics Analysis. FRONT GENET. 2022;13:855789. https://doi.org/10.3389/fgene.2022.855789.

55. Zhang JG, Hilton DJ, Willson TA, et al. Identification, purification, and characterization of a Soluble Interleukin (IL)-13-binding protein. Evidence that it is distinct from the cloned Il-13 receptor and Il-4 receptor alpha-chains. J BIOL CHEM. 1997;272(14):9474–80. https://doi.org/10.1074/jbc.272.14.9474.

56. Lumsden RV, Worrell JC, Boylan D, et al. Modulation of Pulmonary Fibrosis by IL-13Ralpha2. Am J Physiol Lung Cell Mol Physiol. 2015;308(7):L710–8. https://doi.org/10.1152/ajplung.00120.2014.

57. Steck E, Benz K, Lorenz H, Loew M, Gress T, Richter W. Chondrocyte expressed Protein-68 (CEP-68), a Novel human marker gene for cultured chondrocytes. BIOCHEM J. 2001;353(Pt 2):169–74. https://doi.org/10.1042/0264-6021:3530169.

58. Mayr CH, Simon LM, Leuschner G, et al. Integrative Analysis of Cell State Changes in Lung Fibrosis with Peripheral protein biomarkers. EMBO MOL MED. 2021;13(4):e12871. https://doi.org/10.15252/emmm.202012871.

59. Li P, Wang X, Li N, Kong H, Guo Z, Liu S, Cao X. Anti-apoptotic hPEBP4 silencing promotes TRAIL-induced apoptosis of human ovarian Cancer cells by activating ERK and JNK pathways. INT J MOL MED. 2006;18(3):505–10.

60. Okuyama T, Batanian JR, Sly WS. Genomic Organization and localization of gene for human carbonic anhydrase IV to chromosome 17Q. Genomics. 1993;16(3):678–84. https://doi.org/10.1006/geno.1993.1247.

61. Zhang L, Wang Y, Wu G, Xiong W, Gu W, Wang CY. Macrophages: friend or foe in idiopathic pulmonary fibrosis? Respir Res. 2018;19(1):170. https://doi.org/10.1186/s12931-018-0864-2.

62. Mills CD, Ley K. M1 and M2 macrophages: the Chicken and the egg of immunity. J INNATE IMMUN. 2014;6(6):716–26. https://doi.org/10.1159/000364945.

63. Tarique AA, Logan J, Thomas E, Holt PG, Sly PD, Fantino E. Phenotypic, functional, and plasticity features of classical and alternatively activated human macrophages. Am J Respir Cell Mol Biol. 2015;53(5):676–88. https://doi.org/10.1165/rcmb.2015-0012OC.

64. Wang Z, Qu S, Zhu J, Chen F, Ma L. Comprehensive Analysis of lncRNA-associated competing endogenous RNA network and Immune Infiltration in Idiopathic Pulmonary Fibrosis. J THORAC DIS. 2020;12(5):1856–65. https://doi.org/10.21037/jtd-19-2842.

65. Perrot CY, Karampitsakos T, Herazo-Maya JD. Monocytes and macrophages: emerging mechanisms and novel therapeutic targets in Pulmonary Fibrosis. Am J Physiol Cell Physiol. 2023;325(4):C1046–57. https://doi.org/10.1152/ajpcell.00302.2023.

66. Lv J, Xiong Y, Li W, Yang W, Zhao L, He R. BLT1 mediates Bleomycin-Induced Lung Fibrosis independently of neutrophils and CD4＋T cells. J IMMUNOL. 2017;198(4):1673–84. https://doi.org/10.4049/jimmunol.1600465.

Publisher's note