a Open Access Full Text Article

ORIGINAL RESEARCH

Neutrophil-Related Gene Expression Signatures in Idiopathic Pulmonary Fibrosis: Implications for Disease Characteristic and Identification of Diagnostic Hub Genes

Yingying Lin[1,]*, Xiaofan Lai[1,]*, Tianxiang Lei[2,]*, Yuan Qiu[3], Qiwen Deng[1], Qi Liu[1], Zhongxing Wang [1], Wenqi Huang[1]

[1]Department of Anesthesiology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, People's Republic of China; [2]Department of Gastrointestinal Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, People's Republic of China; [3]Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-Sen University, Guangzhou, People's Republic of China

*These authors contributed equally to this work

Correspondence: Wenqi Huang; Zhongxing Wang, Email huangwq@mail.sysu.edu.cn; wzhxing@mail.sysu.edu.cn

Background: Idiopathic pulmonary fibrosis (IPF) is a disease with unclear etiology and a poor prognosis. Although the involvement of neutrophils in IPF pathogenesis has been suggested, the exact nature of this relationship remains unclear.

Methods: We analyzed data from the Gene Expression Omnibus (GEO) using immune infiltration analysis, weighted gene co-expression network analysis (WGCNA), and consensus cluster analysis. Neutrophil-related genes and hub genes related to neutrophils were identified and differentially expressed between IPF patients and healthy controls. We also validated the expression differences of hub genes in a bleomycin-induced mice model.

Results: Immune infiltration analysis revealed a significantly decreased percentage of neutrophils in the lung tissue of IPF patients compared with healthy controls (P<0.001) in both the train and validation sets. Neutrophil-related genes in IPF were identified by WGCNA, and functional enrichment analysis showed that these genes were mainly involved in the cytokine-cytokine receptor interaction pathway and correlated with lung disease, consistent with DEGs between IPF and healthy controls. Eight hub genes related to neutrophils were identified, including *MMP16, ARG1, IL1R2, PROK2, MS4A2, PIR*, and *ZNF436*. Consensus cluster analysis revealed a low neutrophil-infiltrating cluster that was correlated with IPF (P<0.001), and a principal component analysis-generated score could distinguish IPF patients from healthy controls, with an area under the curve of 0.930 in the train set and 0.768 in the validation set. We also constructed a diagnostic model using hub genes related to neutrophils, which showed a reliable diagnostic value with an area under the curve of 0.955 in the train set and 0.995 in the validation set.

Conclusion: Our findings provide evidence of a low neutrophil-infiltrating characteristic in the IPF microenvironment and identify hub genes related to neutrophils that may serve as diagnostic biomarkers for the disease.

Keywords: neutrophils, idiopathic pulmonary fibrosis, immune infiltration analysis, weighted gene co-expression network analysis, consensus cluster analysis, diagnostic model, biomarkers

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic and progressive interstitial lung disease characterized by the formation of scar tissue and remodeling of the alveolar spaces of the lung tissues.[1,2] The disease can ultimately lead to respiratory symptoms such as shortness of breath, persistent dry cough, fatigue, and loss of appetite.[3] IPF is a rare disease with an incidence of 8.2 cases per 100,000 individuals, and its prevalence increases with age.[4,5] It is typically diagnosed in individuals over the age of 50[6,7] and is more common in men than in women.[8,9] Unfortunately, the median survival of

2503

IPF is just 2–4 years since diagnosis,[6,10] indicating a worse prognosis than that of lung cancer.[11] However, the exact cause and pathogenesis of IPF are still unclear,[12] but genetic risk factors and immune dysregulation are believed to play significant roles.[13–15] A recent study evaluated the correlations between genetic variants and high-resolution computed tomography (HRCT) patterns in IPF and found that genetic factors played a pivotal role in the risk of developing IPF.[16] In another study, a correlation between genetic risk factors for IPF and immunopathogenesis were reviewed, and it suggested that genetic variants might influence the pathogenesis of IPF through modulation of innate immune processes.[17]

Among innate immune cells, neutrophils have been identified as potential modulators of fibrogenesis in the development of IPF.[18,19] Neutrophil extracellular trap (NET) and neutrophil elastase were reported involved in inducing the activation of fibroblasts and myofibroblasts differentiation, leading to fibrogenesis.[20–22] Increased levels of neutrophils in bronchoalveolar lavage (BAL) have been associated with poor progression of IPF.[23] However, it has been suggested that neutrophils may play a protective role in IPF by alleviating epithelial cell injury through the transfer of miR-223.[24] Furthermore, neutrophils have strong antibacterial properties, which may protect IPF lungs from the invading of pathogens.[25] The exact role of neutrophils in IPF is still controversial, and further research is needed to establish a clear correlation between neutrophils and IPF.[14,26]

To shed light on this issue, we investigated the relationship between neutrophils and IPF by exploring the immune microenvironment in the lung tissues. We used a combination of immune infiltration analysis and weighted gene co-expression network analysis (WGCNA) to identify neutrophil-related genes and explore their potential role in the pathogenic mechanisms of IPF. We found that several neutrophil-related genes were significantly associated with IPF. Furthermore, we employed consensus cluster analysis to identify the characteristic features associated with neutrophils in IPF. Additionally, we constructed a diagnostic model for IPF using hub genes related to neutrophils to highlight their potential role in the disease.

While the precise role of neutrophils in IPF remains a subject of controversy, our findings firstly offer insights into the distinctive low neutrophil infiltration observed in the lung tissue of IPF, shedding light on the potential underlying mechanisms.

Methods

Retrieval and Processing of Data

Our analyses were based on the integration of multiple gene expression profiling datasets obtained from the Gene Expression Omnibus (GEO). The workflow of our study is presented in Figure 1. We retrieved five expression profiling datasets generated by microarray technology, including GSE53945, GSE32537, GSE17978, GSE110147, and GSE10667, all of which comprised more than 30 lung tissue samples from *Homo sapiens* and including both IPF patients and healthy controls. These five cohorts were then combined to create a meta-GEO cohort, which served as the training set of our study. The meta-GEO cohort included 104 lung samples from healthy controls and 255 lung samples from IPF patients. Furthermore, a dataset generated by high-throughput sequencing, namely GSE213001, was downloaded as the validation set of our study. Additionally, we acquired an expression profiling dataset, GSE33566, consisting of peripheral blood samples from 30 healthy controls and 93 IPF patients, to further investigate the expression of these genes in circulating blood. To ensure consistency in the data, all merged datasets were background calibrated and normalized using the sva R package. We used the Limma R package to identify the differentially expressed genes (DEGs). And the DEGs were defined as those with a |log2-fold change| > 1 and an adjusted P value < 0.05.

Immune Infiltration Analysis

In this study, we employed the Cell Type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) algorithm as a computational tool to discern the proportions of various cell types based on gene expression data. Using this approach, we analyzed the relative abundance of 22 different immune cell types in each lung sample after normalizing the data. We then compared the proportions of immune cells that infiltrated the lung tissue in samples. We considered a significant difference to exist when the CIBERSORT output P value was less than 0.05.
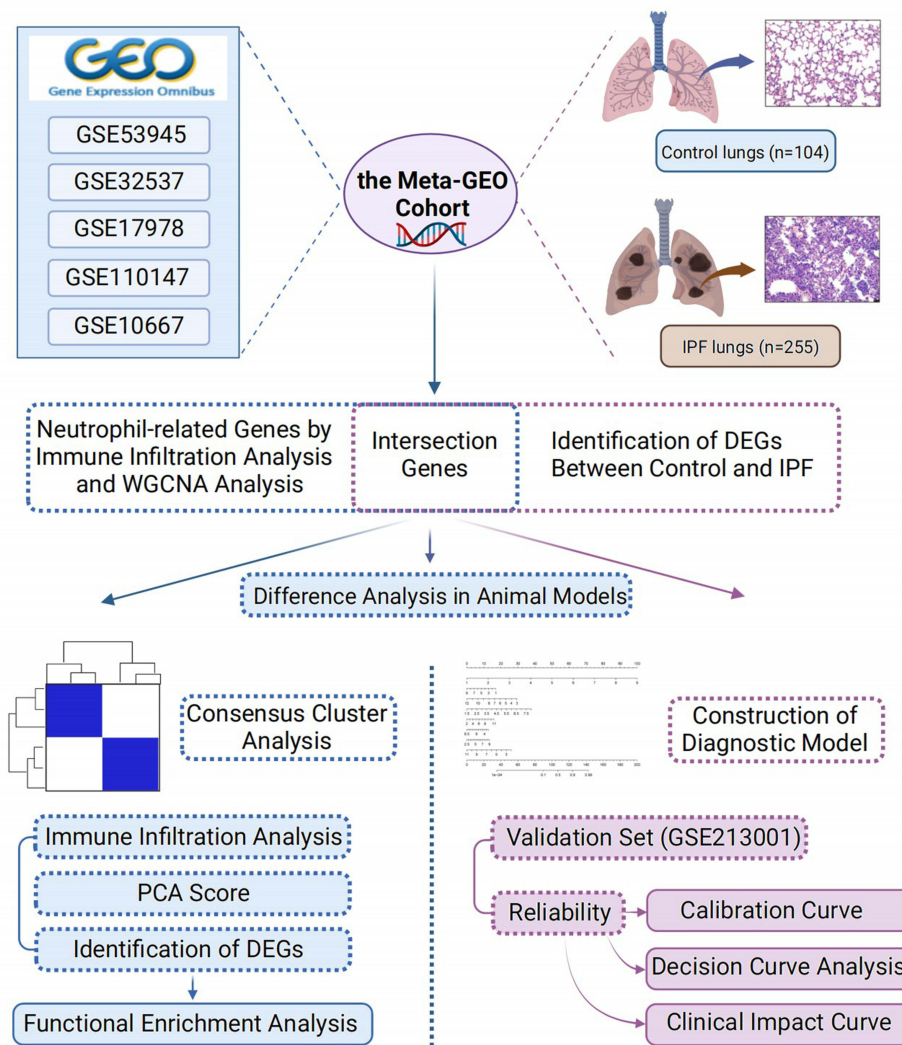
Figure 1 Workflow outlining the experimental methodology utilized in this study. Initially, five Gene Expression Omnibus (GEO) datasets were integrated to create a Meta-GEO cohort consisting of control and IPF lung samples. Through the implementation of WGCNA and the identification of DEGs, we identified hub genes related to neutrophils in IPF. Subsequently, a consensus cluster analysis was performed, followed by the development of a diagnostic model employing these hub genes.
Abbreviations: GEO, Gene Expression Omnibus; IPF, idiopathic pulmonary fibrosis; WGCNA, Weighted Correlation Network Analysis; DEGs, Differentially Expressed Genes; PCA, Principal Component Analysis.

To complement our findings, we also utilized the Single-Sample Gene Set Enrichment Analysis (ssGSEA) method to assess immune infiltration in lung samples. To do this, we analyzed immune infiltration in a lung sample based on the gene set associated with immune cells from the Molecular Signatures Database (MSigDB) (http://software.broadinstitute.org/gsea/msigdb), calculating enrichment scores for each gene set, which reflect the level of immune cell infiltration in the sample. We then compared the enrichment scores to identify differences in immune infiltration. A P value less than 0.05 was deemed to indicate significant difference. Furthermore, by utilizing these two methods, we analyzed the relationship between hub genes and infiltrating immune cells.

Weighted Gene Co-Expression Network Analysis

WGCNA is a well-established data-driven methodology that allows for the identification of gene modules displaying similar expression patterns and the investigation of gene relationships. In this study, we applied WGCNA to examine the correlation between gene modules and neutrophils, with the aim of uncovering putative neutrophil-related genes in IPF.

The WGCNA approach consisted of the following steps: preprocessing of gene expression data, computation of a similarity matrix, construction of a weighted gene co-expression network, module detection, characterization of gene modules via their module eigengene and correlation analysis between gene modules and infiltrating immune cells. And in this study, a Scale-free topological overlap (SFTO) threshold of 0.9 was utilized to identify the optimal soft-thresholding in the network construction.

Gene Set Enrichment Analysis

To investigate the function of the gene set in our study, we conducted Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Disease Ontology (DO) enrichment analyses. GO enrichment analysis is a widely used method that identifies overrepresented biological processes (BP), molecular functions (MF), and cellular components (CC) in a set of genes compared to a reference set, allowing the properties of genes and their products to be described. KEGG enrichment analysis employs the KEGG database, which includes metabolic pathways, regulatory pathways, signaling pathways, and other biological pathways, to determine which pathways are significantly enriched according to our gene set. We used the R package "clusterProfiler" in this study to obtain the results of GO and KEGG enrichment analyses. Through DO analysis, we determined the specific diseases mainly associated with our gene set, which can provide further understanding of the underlying mechanisms of IPF and suggest new directions for our research. The P values of the enriched output were considered significant if they were less than 0.05.

Consensus Clustering Analysis

Consensus clustering analysis is an unsupervised learning method that combines multiple clustering solutions based on different algorithms or parameter settings to improve the stability and robustness of cluster analysis results. In this study, we employed consensus clustering analysis to underscore the importance of neutrophil-related genes in IPF based on the expression of eight hub genes related to neutrophils. Our analysis employed the following parameters: item resampling (pItem) was set to 0.8, gene resampling (pFeature) was set to 1, and resamplings (reps) were set to 50. We obtained distinct neutrophil-related clusters, and the optimal number of clusters was determined based on the results of the consensus cumulative distribution function (CDF) plot, consensus matrix plot, and principal component analysis (PCA). Additionally, we developed a PCA score through PCA to quantify neutrophil-related clusters, and through ROC curve analysis, we estimated its diagnostic value in IPF in the train and validation sets.

Construction and Validation of a Diagnostic Model

Based on the expression data from hub genes related to neutrophils, we developed a diagnostic model using multiple logistic analysis and presented the results as a nomogram. To validate the reliability of the diagnostic model, we employed various statistical methods, including calibration curve, clinical impact curve, decision curve analysis, and ROC curve analysis. The calibration curve provides a measure of the goodness-of-fit of the model by plotting the predicted probabilities against the observed frequencies. The clinical impact curve is a plot that shows the relationship between sensitivity and specificity of the diagnostic model, indicating the proportion of individuals who receive correct diagnoses, thereby allowing us to estimate the clinical impact of the model. In addition, decision curve analysis was performed to evaluate the clinical utility of the diagnostic model by comparing the net benefit of different decision strategies across a range of threshold probabilities. Furthermore, we used ROC curve analysis to determine the diagnostic value of our model in both the train and validation sets. To evaluate the importance score of each variable in the diagnostic model, we employed the random forest algorithm to calculate the feature importance attribute of each hub gene.

Construction of a Bleomycin-Induced Pulmonary Fibrosis Model

Our animal experiments were conducted at the Sun Yat-sen University Animal Center, and approved by the Ethics Committee of Sun Yat-sen University, in accordance with the ethical principles and guidelines for the care and use of laboratory animals. C57BL/6 mice were housed in a colony room with causally access to food and water, and maintained under a 12:12 hour light/dark cycle. To induce pulmonary fibrosis in mice, 8-week-old mice were anesthetized with

isoflurane and intratracheally injected with bleomycin (Teva Pharmaceutical) at a dose of 3 U/kg, while control mice were intratracheally injected with an equivalent volume of PBS. At day 21, all mice were humanely euthanized and lung samples were collected for subsequent research.

Histopathology Analysis

Following euthanasia, lung samples were meticulously excised from mice and promptly fixed in 4% paraformaldehyde. The lung tissues were then subjected to 30% sucrose saturation for 24 hours to enable cryosections or embedded in paraffin. Paraffin-embedded lung samples underwent hematoxylin and eosin (H&E), Masson trichrome and Sirius red staining to evaluate the extent of collagen deposition and fibrosis in the lungs.

RNA Extraction and Quantitative PCR (qPCR)

Total RNA was extracted from the lung tissue samples of mice using TRIzol reagent (Molecular Research Center, Inc.). RNA concentration and purity were determined with a NanoDrop 8000 spectrophotometer. Subsequently, cDNA was synthesized from the extracted RNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, K1622). qPCR analysis was performed using the LightCycler480 SYBR Green I Master Mix (Roche, 4887352001–1), with β-actin serving as the reference gene. Primer sequences for the genes of interest are as follows: mouse *Il13ra2* forward 5'-TCCGAGAAGACGATATTTCCTGG-3', and reverse 5'-CCATTCGCTCCAAATTCCATCA-3'; mouse *Arg1* forward 5'-TGTCCCTAATGACAGCTCCTT-3', and reverse 5'-GCATCCACCCAAATGACACAT-3'; mouse *Il1r2* forward 5'-CAACGCTGATGGAAAGATACAGT-3', and reverse 5'-CAATAGGCGTGTGGGGTCTC-3'; mouse *Mmp16* forward 5'-TGTACCAGACCAGACAAGAGG-3', and reverse 5'-CATCAAAGGCACGGCGAATTG-3'; mouse *Ms4a2* forward 5'-GTCTGCTCCGTACTCTATGTTTC-3', and reverse 5'-GCCTCTCACCAGATACAATGTGT-3'; mouse *Prok2* forward 5'-CCTGAGGGGCTTGAAATTCAC-3', and reverse 5'-GTTGGTCTACTTGGAGGAACG-3'; mouse *Znf436* forward 5'-CGGAAATGTTGTCTCACTGGA-3', and reverse 5'-ATCGTCACTAAACTCCTGTTTGG-3'; mouse *Pir* forward 5'-GTCGAAGGTTTACACTCGCAC-3', and reverse 5'-AATAGGCTGGGAATGCTTTGC-3';

Statistical Analysis

All statistical analyses in this study were performed using R software version 4.1.3 (http://www.r-project.org). For continuous variables that satisfied the assumptions of normality and homoscedasticity, independent-sample t-tests were conducted to assess differences between two groups. In cases where the continuous variables were not normally distributed, Wilcoxon rank-sum tests were employed. Correlations between continuous variables were analyzed using both Pearson and Spearman coefficients, depending on the nature of the data. Categorical variables were analyzed for differences using chi-square tests or Fisher's exact tests based on the nature of the data. In the case of the PCA score, we defined values greater than 0 as high and values less than 0 as low, considering the relationship between PCA score, disease state, and cluster. Statistical significance was defined as a P-value less than 0.05.

Results

Identification of Neutrophil-Related Genes in IPF

Firstly, we investigated the difference in the immune microenvironment between the lung samples of IPF patients and healthy controls. Through utilizing the CIBERSORT algorithm for infiltrating immune analysis in the training set (the meta-GEO cohort), we discovered a noteworthy distinction in the distribution of infiltrating immune cells between IPF patients and healthy controls (Figure 2A and B). Our results showed a marked decrease in infiltrating neutrophils in IPF patients (P<0.001), which was validated in the validation set (P<0.001) (GSE213001) as shown in the Supplementary Figure 1A and B.

Following this, we employed WGCNA on the training set to identify genes related to neutrophils in IPF. The sample clustering tree revealed that a single outlier was excluded from the analysis (Supplementary Figure 1C). The optimal power of soft-threshold was determined to be 7 (Figure 2C) and two gene modules were identified (Supplementary Figure 1D and
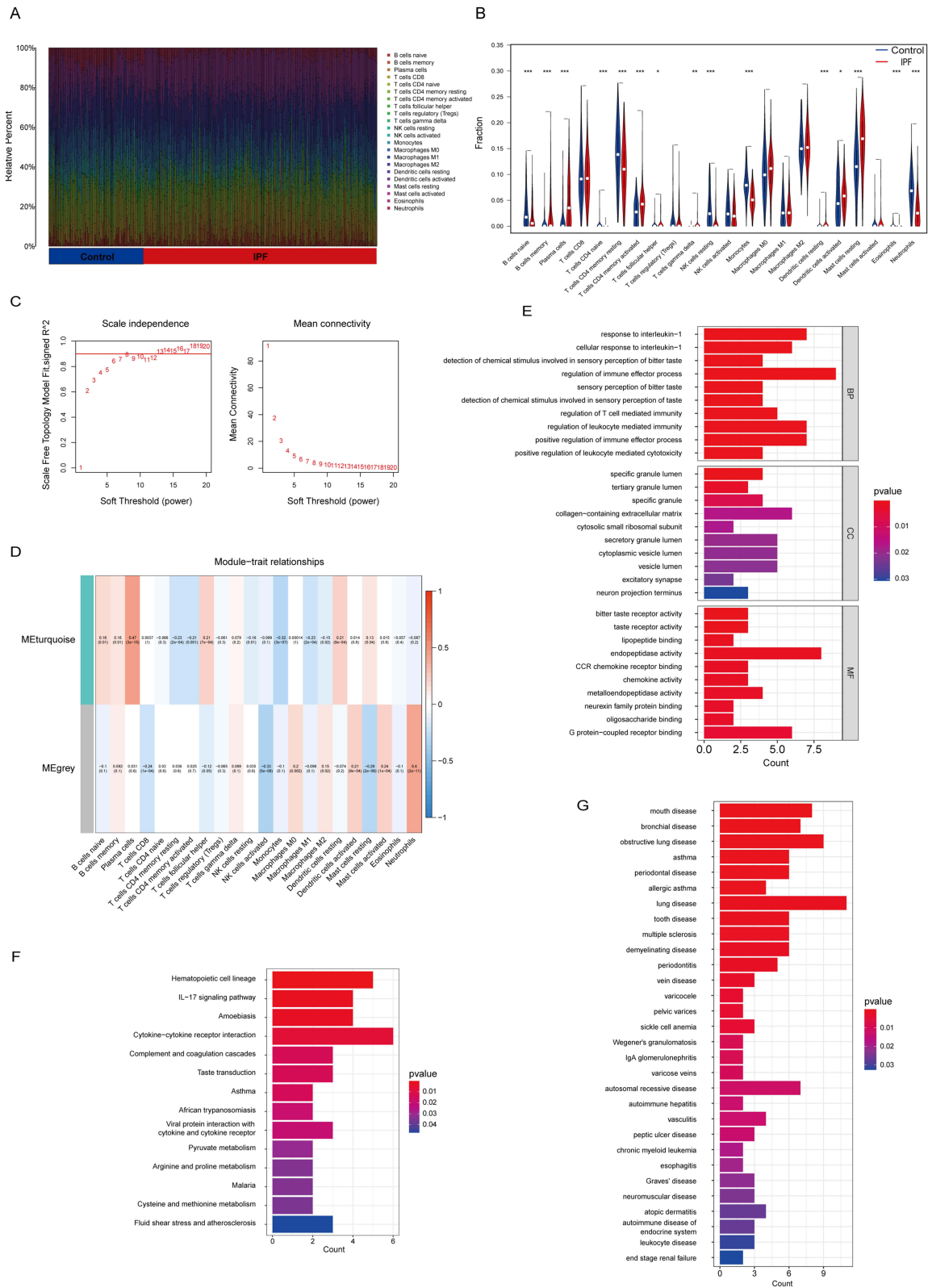
Figure 2 Identification of neutrophil-related genes in IPF. (**A**) The fraction of different infiltrating immune cells in IPF patients or healthy controls lung samples in the meta-GEO cohort. (**B**) Differential analysis of the distribution of different infiltrating immune cells between IPF patients and healthy controls lung samples in the meta-GEO cohort. (**C**) Identification of optimal soft threshold power of WGCNA. (**D**) Correlation between WGCNA module and different infiltrating immune cells in IPF in the meta-GEO cohort. The data are presented as correlation coefficient (P value). (**E**) GO functional enrichment analysis of neutrophil-related genes. (**F**) KEGG functional enrichment analysis of neutrophil-related genes. (**G**) DO functional enrichment analysis of neutrophil-related genes. (*P <0.05, **P < 0.01, and ***P < 0.001).
Abbreviations: BP, biological processes; CC, cellular components; MF, molecular functions.

Supplementary Table 1). The result suggested that the genes in grey module was significantly correlated with infiltrating neutrophils in IPF (Correlation Coefficient = 0.4, P<0.001) (Figure 2D).

To further understand the biological functions of these genes related to neutrophils, we conducted GO functional enrichment analysis. The results indicated that these genes were mainly involved in the regulation of immune effector process (P<0.001), response to interleukin-1 (P<0.001) and regulation of leukocyte mediated immunity (P<0.001) in the BP category. Furthermore, our analysis revealed that these genes play a role in collagen-containing extracellular matrix (P=0.016) in the CC category and endopeptidase activity (P=0.001) in the MF category (Figure 2E). Additionally, a pathway analysis using the KEGG database elucidated that cytokine-cytokine receptor interaction was the predominant pathway of these genes (P=0.007) (Figure 2F). And DO analysis revealed a significant correlation between these genes and lung diseases (P=0.001) (Figure 2G). In conclusion, these genes were related to the function of neutrophils and might play a role in the pathological process of IPF.

Identification of Hub Genes Related to Neutrophils in IPF

In the training set, a comprehensive analysis of DEGs revealed 138 genes between the lung samples of IPF patients and healthy controls, of which 48 were downregulated and 90 were upregulated (Figure 3A). Subsequent functional enrichment analysis indicated that these DEGs were mainly enriched in extracellular matrix organization, extracellular structure organization and external encapsulating structure organization in BP category, collagen-containing extracellular matrix and collagen trimer in CC category, and receptor ligand activity and signaling receptor activator activity in MF category (Figure 3B). Additionally, KEGG analysis revealed that IPF was strongly associated with cytokine-cytokine receptor interaction and PI3K-Akt signaling pathway (Figure 3C). DO analysis showed that these DEGs were enriched in lung disease and pulmonary fibrosis (Figure 3D), indicating their reliability for further research.

Next, we conducted a venn analysis to identify the intersected genes between the neutrophil-related genes and DEGs, and eight hub genes were identified, including *IL13RA2, MS4A2, MMP16, ZNF436, PIR, PROK2, ARG1* and *IL1R2* (Figure 3E–F). Among these hub genes, *IL13RA2, MS4A2, MMP16* and *ZNF436* were upregulated, while *PIR, PROK2, ARG1* and *IL1R2* were downregulated in IPF patients in the train set (Supplementary Figure 2A). And the results were confirmed by the validation set (Supplementary Figure 2B). Subsequent GO functional analysis demonstrated that upregulated genes were mainly enriched in negative regulation of mast cell activation involved in immune response (Supplementary Figure 2C), and downregulated genes, on the other hand were mainly enriched in negative regulation of cytokine−mediated signaling pathway (Supplementary Figure 2D). Furthermore, ROC curve analysis showed that these hub genes had a high ability to distinguish IPF patients from healthy controls (Supplementary Figure 3A–H), and the results were validated by the validation set (Supplementary Figure 3I–P). Additionally, we observed that *PROK2, ARG1*, and *IL1R2* were the only genes expressed in peripheral blood samples of IPF patients, and they exhibited significant differences between IPF patients and healthy controls (Supplementary Figure 4A). Notably, these genes also showed a strong correlation with neutrophils (Supplementary Figure 4B–D). ROC curve analysis revealed that all three genes retained their discriminatory ability to distinguish IPF patients from healthy controls (Supplementary Figure 4E–G).

Construction of Neutrophil-Related Clusters Based on Hub Genes Using Unsupervised Clustering Analysis

The expression patterns of the eight hub genes were identified in the previous section. Through CIBERSORT algorithm and ssGSEA immune infiltration analysis, it was showed a positive correlation between infiltrating neutrophils in IPF and the expression of AR*G1, IL1R2, PROK2* and *PIR* and a negative correlation with the expression of *IL13RA2, MMP16, MS4A2* and *ZNF436* (Figure 4A and B).

To further investigate the expression patterns of these hub genes, consensus clustering analysis was performed on the training set using the expression data. The results showed that the optimal number of clusters was two, as determined by the minimum descending slope of the cumulative distribution function (CDF) when k = 2 and by the agreement in the relative consensus matrix plot (Figure 4C and D). Furthermore, PCA demonstrated that cluster A was obviously distinguished from cluster B (Figure 4E), and the expression levels of the hub genes differed significantly between the
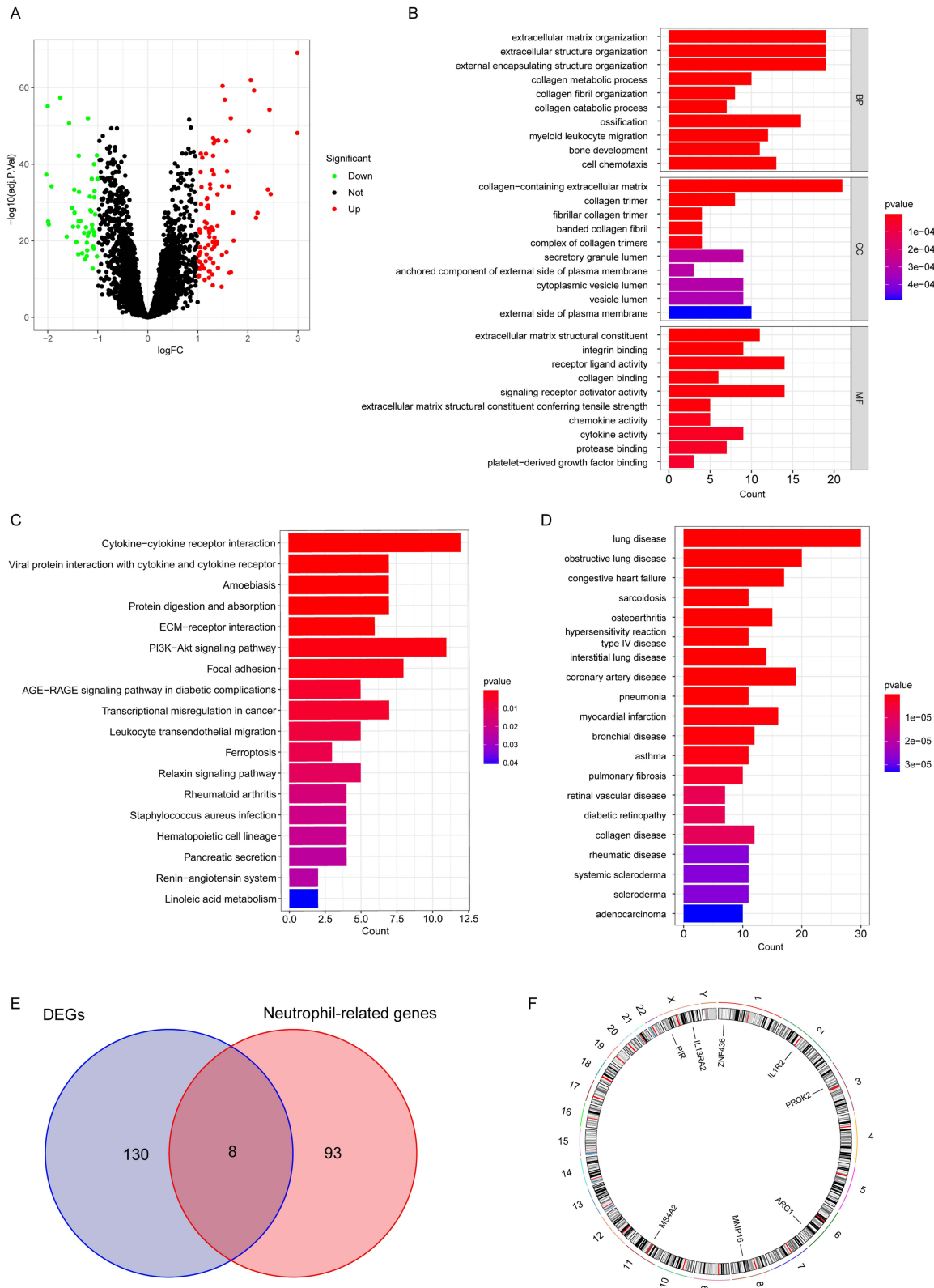
Figure 3 Identification of hub genes related to neutrophils in IPF. (**A**) Volcano plot of the DEGs between IPF and healthy controls lung samples in the meta-GEO cohort. (**B**) GO functional enrichment analysis of the DEGs. (**C**) KEGG functional enrichment analysis of the DEGs. (**D**) DO functional enrichment analysis of the DEGs. (**E**) Venn diagram of the intersection genes between the DEGs and neutrophil-related genes. (**F**) Circos diagram depicting the chromosomal locations of hub genes.
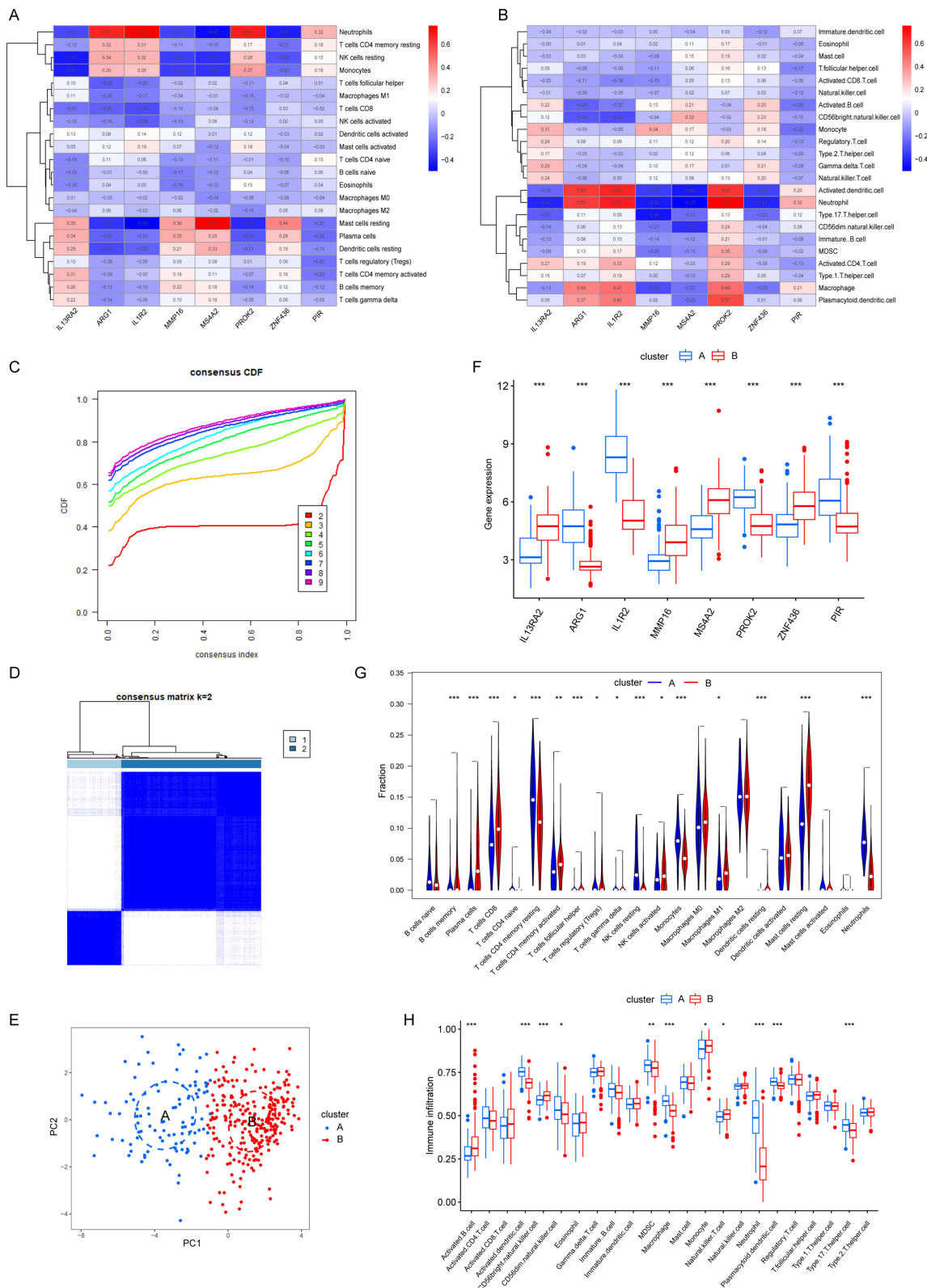
Figure 4 Construction of neutrophil-related clusters based on hub genes using unsupervised clustering analysis. (**A** and **B**) Correlation between hub genes and different infiltrating immune cells in the meta-GEO cohort using CIBERSORT algorithm (**A**) and ssGSEA (**B**). (**C**) CDF diagram of the meta-GEO cohort for k=2–9. (**D**) Consensus matrices of the meta-GEO cohort when k=2. (**E**) PCA scatter plot of two neutrophil-related clusters. (**F**) The differences in the expression of eight hub genes between two clusters. (**G** and **H**) Differential analysis in the distribution of different infiltrating immune cells between the two clusters using CIBERSORT algorithm (**G**) and ssGSEA (**H**). ($*P < 0.05$, $**P < 0.01$, and $***P < 0.001$).
Abbreviation: CDF, cumulative distribution function.

two clusters. Specifically, *IL13RA2, MS4A2, MMP16* and *ZNF436* were upregulated in cluster B, while *PIR, PROK2, ARG1* and *IL1R2* were downregulated (Figure 4F), which was consistent with the variation tendency of IPF. Additionally, immune infiltration analysis showed a significant difference in the proportion of infiltrating neutrophils between the two clusters, with a lower proportion in cluster B compared to cluster A (Figure 4G and H). These results led us to classify cluster B as a less neutrophil-infiltrating cluster and cluster A as a neutrophil-infiltrating cluster.

Characterization of the Gene Signature in the Low Neutrophil-Infiltrating Was Similar to IPF

We identified 90 DEGs between the neutrophil-infiltrating cluster A and the low neutrophil-infiltrating cluster B (Figure 5A). Out of these DEGs, 44 were upregulated and 46 were downregulated in cluster B. Interestingly, further functional enrichment analysis showed that the gene signature of neutrophil-related clusters was similar to IPF. It was depicted that the gene signature was mainly enriched in extracellular matrix organization, extracellular structure organization and external encapsulating structure organization in BP category, collagen-containing extracellular matrix in CC category, and cytokine binding, immune receptor activity and extracellular matrix structural constituent in MF category (Figure 5B). KEGG analysis revealed that the most relevant pathways was cytokine−cytokine receptor interaction (Figure 5C). While the gene signature was most strongly correlated with lung disease in DO analysis (Figure 5D).

We then performed PCA on the gene signature of neutrophil-related clusters and found that the PCA score had diagnostic value for IPF, with an AUC of 0.930 in the train set (Supplementary Figure 5A) and 0.768 in the validation set (Supplementary Figure 5B). As showed in Figure 5E, cluster B had a higher PCA score, consistent with the pattern observed in IPF (Figure 5F). Furthermore, a Sankey diagram revealed that cluster B was primarily grouped with the disease state of IPF (P<0.001) (Figure 5G and Supplementary Figure 5C). Overall, our results suggest that the low neutrophil-infiltrating cluster B has a gene signature similar to IPF and is characterized by a higher PCA score.

Construction of a Diagnostic Model Using Hub Genes Related to Neutrophils

Using multivariate logistic regression analysis and nomogram, we constructed a diagnostic model incorporating above eight hub genes (*IL13RA2, MS4A2, MMP16, ZNF436, PIR, PROK2, ARG1* and *IL1R2*) related to neutrophils (Figure 6A). And a risk score was obtained based on the expression of these hub genes, as represented by the following equation: risk score = −2.7571 + (the expression level of *IL13RA2* × 1.3788) + (the expression level of *ARG1* × −0.1996) + (the expression level of *IL1R2* × −0.3534) + (the expression level of *MMP16* × 0.6252) + (the expression level of *MS4A2* × 0.1868) + (the expression level of *PROK2* × −0.2646) + (the expression level of *ZNF436* × 0.1633) + (the expression level of *PIR* × −0.2862).

The results of the study showed that the diagnostic model was well-calibrated, as demonstrated by the calibration curve diagram (Figure 6B). The predicted number of high-risk patients was also very close to the actual number of IPF patients in the clinical impact curve (Figure 6C). In addition, decision curve analysis showed that the diagnostic model was effective at threshold probabilities ranging from 0 to 1, with the greatest net benefit obtained when the threshold probability of IPF was set at 0.8 (Figure 6D). Also, the performance of the diagnostic model was further evaluated using ROC curve analysis, which showed that the AUC was 0.955 for the train set and 0.995 for the validation set (Figure 6E and F), indicating that the diagnostic model had a remarkable ability to distinguish IPF patients from healthy controls. Through random forest algorithm, it was revealed that *IL13RA2* had the highest importance score, followed by *MMP16, ARG1, IL1R2, PROK2, MS4A2, PIR*, and *ZNF436* (Figure 6G and H).

The Expression of Hub Genes Exhibited a Significant Difference Between the Bleomycin-Induced Models and Controls

To confirm the differential expression of eight hub genes, a mouse model of bleomycin-induced pulmonary fibrosis was constructed. Intratracheal injection of bleomycin or PBS was administered on day 0, and lung tissue samples were collected until day 21, as indicated in Figure 7A. Significant collagen deposition and fibrosis were observed in the lungs of the bleomycin-induced mice, as demonstrated by H&E, Masson trichrome, and Sirius Red staining, when compared to
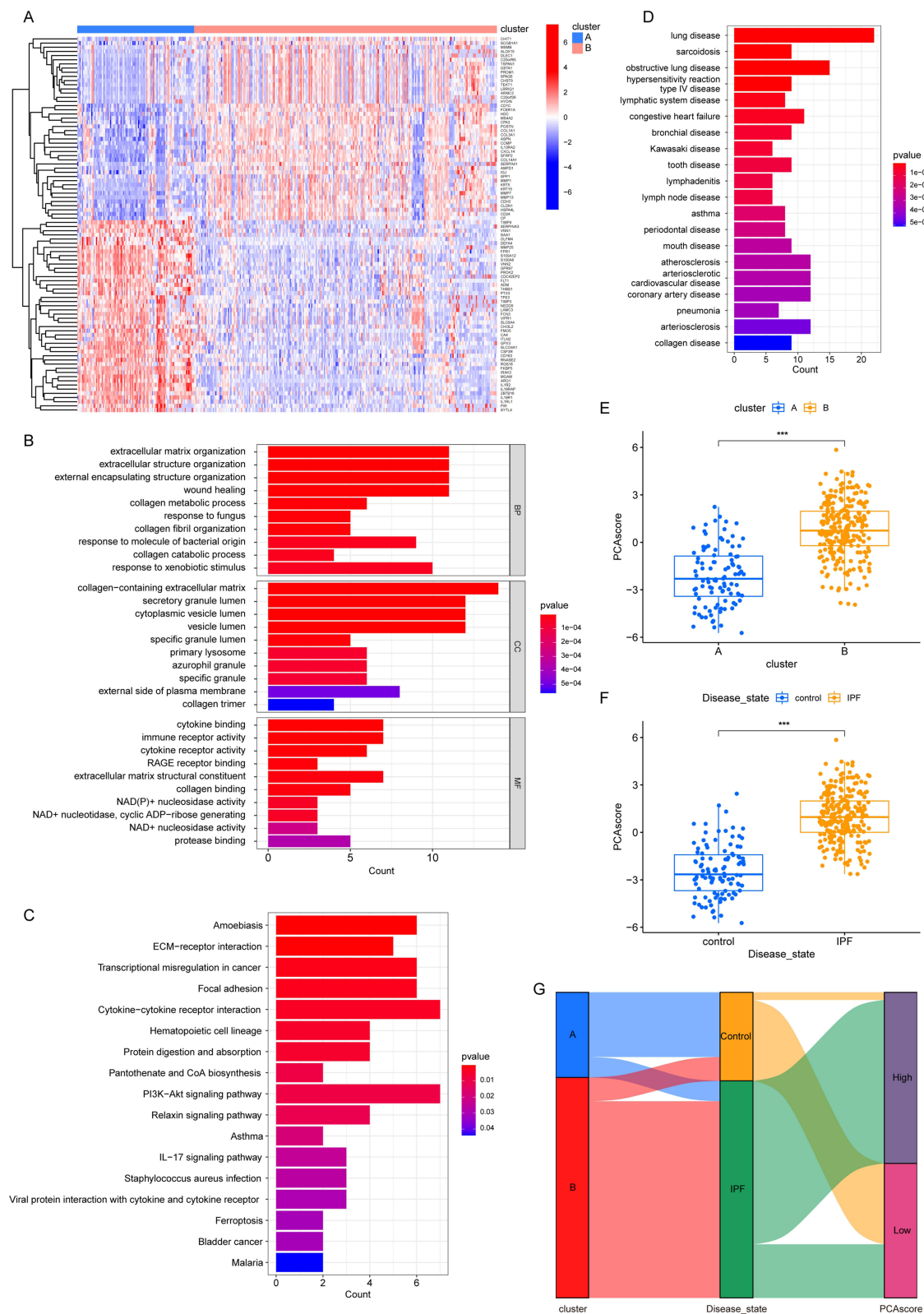
Figure 5 The less neutrophil-infiltrating cluster showed a similar gene signature to IPF. (**A**) Heatmap of DEGs between two clusters. (**B**) GO functional enrichment analysis of DEGs between two clusters. (**C**) KEGG functional enrichment analysis of DEGs between two clusters. (**D**) DO functional enrichment analysis of DEGs between two clusters. (**E**) Differential analysis of PCA score between two clusters. (**F**) Differential analysis of PCA score between IPF and healthy controls lung samples. (**G**) Sankey diagram of the correlations among neutrophil-related clusters, disease state and PCA score. (***P < 0.001).
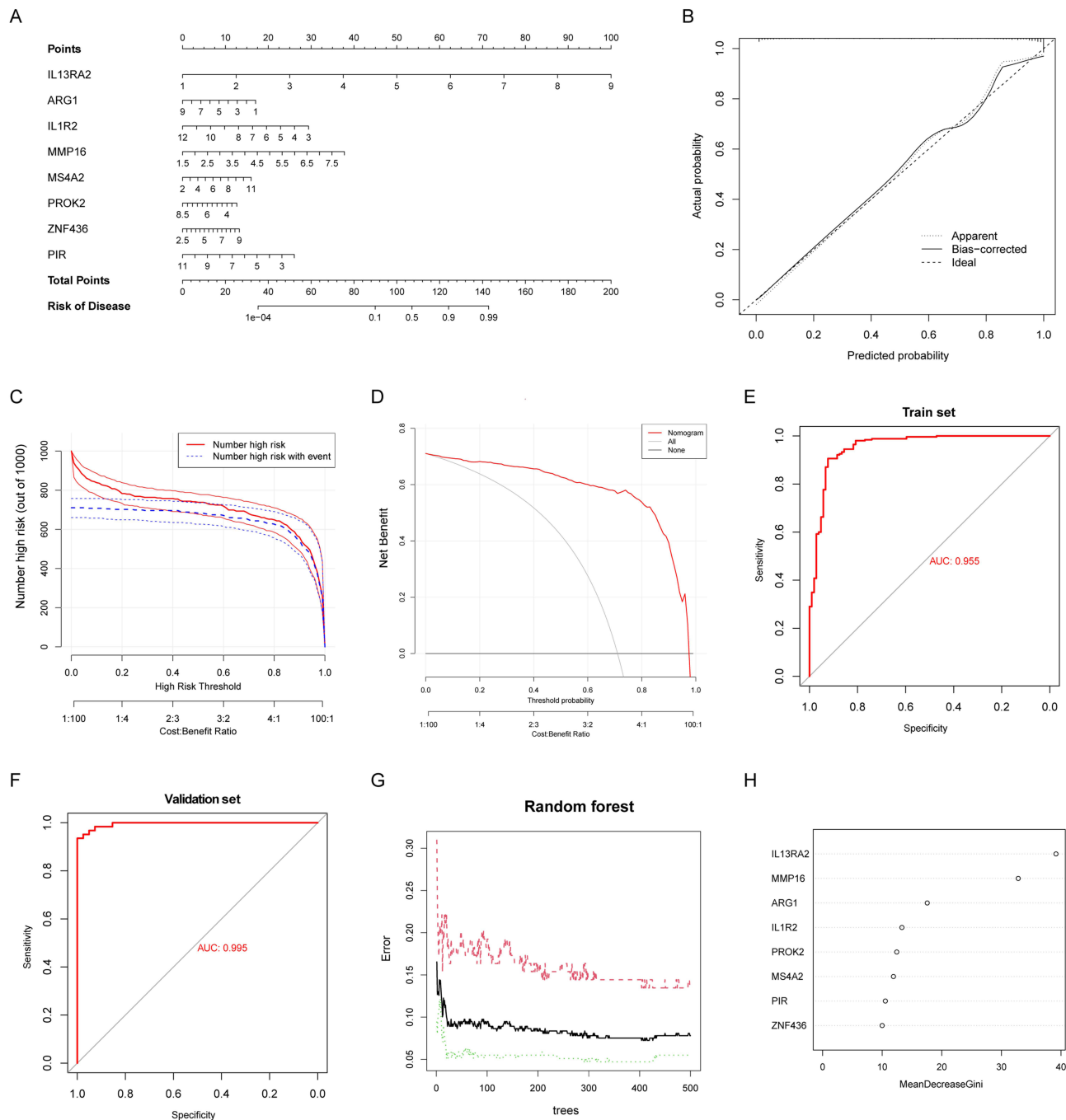
Figure 6 Construction of a diagnostic model using hub genes related to neutrophils. (**A**) Nomogram presented the diagnostic model constructed using hub genes. (**B**) Calibration curve of the diagnostic model. (**C**) Clinical impact curve for the nomogram. (**D**) Decision curve analysis for the nomogram. (**E**) ROC curve for the meta-GEO cohort. (**F**) ROC curve for the validation set (GSE213001). (**G**) Random forest for the relationship between cross-validated error and the numbers of random trees. (**H**) The importance score of hub genes using random forest algorithm.

the controls (Figure 7B). Additionally, qPCR analysis revealed that the expression of eight hub genes differed significantly between the bleomycin-induced models and controls. Specifically, *IL13RA2, MMP16, MS4A2*, and *ZNF436* displayed higher expression levels in bleomycin-induced models compared to controls, while *PIR, PROK2, ARG1*, and *IL1R2* demonstrated a decreasing trend in expression (Figure 7C–J).

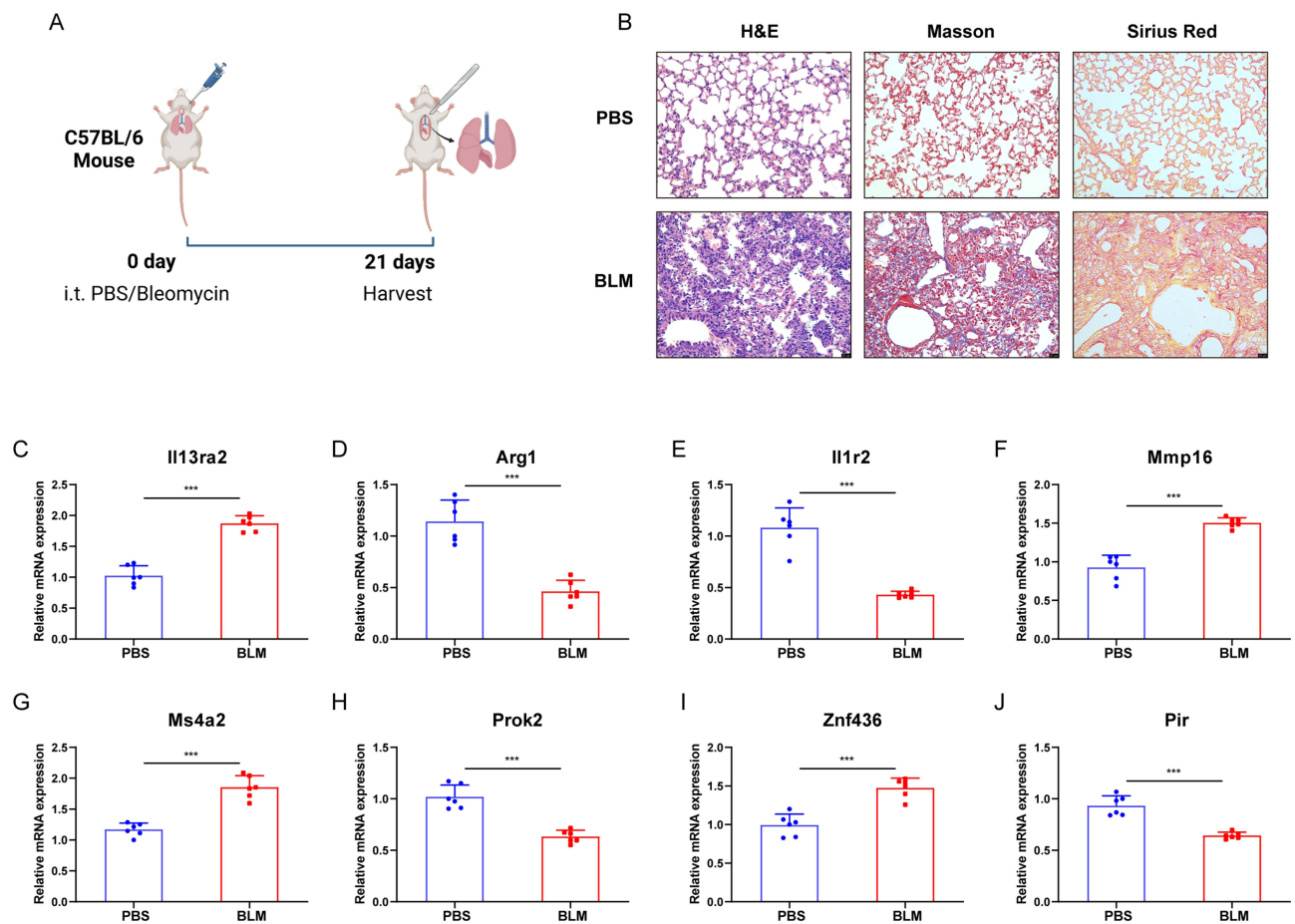**Figure 7** The expression of hub genes exhibited a significant difference between the bleomycin-induced models and controls. (**A**) Construction of the animal models. (**B**) Representative images of hematoxylin and eosin (H&E), Masson's trichrome and Sirius Red staining from the lung samples of bleomycin-induced mice models or controls. Scale bars, 50 μm. (**C–J**) qPCR analysis of relative mRNA expression level in the lung samples of bleomycin-induced mice models compared with controls (*IL13RA2, MMP16, ARG1, IL1R2, PROK2, MS4A2, PIR*, and *ZNF436*). (***P < 0.001).

Discussion

Our study provided evidence of a distinct low neutrophil-infiltrating characteristic within the lung tissue microenvironment of IPF. Additionally, we observed that neutrophil-related genes in IPF were mainly enriched in the cytokine-cytokine receptor interaction pathway. Moreover, we successfully developed a diagnostic model utilizing eight hub genes (*MMP16, ARG1, IL1R2, PROK2, MS4A2, PIR, IL13RA2* and *ZNF436*) related to neutrophils, demonstrating a substantial diagnostic value for IPF.

As a key defense against invading pathogens, neutrophils play a crucial role in protecting the body from infections by rapidly recruiting to the site of infection and phagocytosing the pathogens.[27,28] Additionally, neutrophils modulate both innate and adaptive immune responses,[29] involved in many pathophysiological processes. Although several studies have suggested that neutrophils might be involved in the pathogenesis of IPF,[18,19,30] there is still lack of direct evidence showing the correlation between neutrophils and human IPF, and the role of neutrophils in IPF remains controversial.

In this study, we found that compared with healthy controls, there was a significant decrease in infiltrating neutrophils in the lung tissues of IPF patients. This result was further supported by a low neutrophil-infiltrating cluster B, obtained from consensus cluster analysis based on the expression data of eight hub genes related to neutrophils, which showed a similar gene signature in IPF. Our results provided evidence of a low neutrophil-infiltrating characteristic in the IPF microenvironment. IPF was used to be characterized by a disrupting inflammatory response causing dysregulated wound repair.[31,32] However, subsequent clinical trials have indicated that anti-inflammatory therapy was ineffective for IPF and even caused prolonged hospitalization and increased mortality.[33] It has also been demonstrated that IPF contained fewer

inflammatory cells and little ground-glass opacities on chest imaging compared with other Interstitial Lung Disease (ILDs).[8,34] Neutrophils are among the most common and important inflammatory cells in the inflammatory response.[25] Our results suggested that a low neutrophil-infiltrating microenvironment might explain why anti-inflammatory therapy was ineffective for IPF and even be harmful to patients. IPF patients have been shown to have an increased bacterial burden compared with healthy controls, resulting in declining lung function and a higher risk of death.[35,36] And an abundance of streptococcal and staphylococcal organisms has been observed among IPF patients, and this has been associated with their poor progression.[37] Neutrophils are the body's main defender against bacteria,[38,39] and a low neutrophil-infiltrating microenvironment in IPF may result in a bacterial burden among these patients. However, further studies are needed to explore this hypothesis.

Moreover, our results suggested that a decreased presence of neutrophils might be implicated in the pathogenesis of IPF. Kinder et al discovered that a higher percentage of BAL neutrophils was related to increased death in a cohort of 156 patients with IPF.[23] It has also been shown that neutrophil elastase contributed to the activation of fibroblasts and the deposition of extracellular matrix,[22,40] and that Sivelestat, a neutrophil elastase inhibitor, alleviated pulmonary fibrosis in the bleomycin mouse model.[41] Neutrophils can also release pro-fibrotic cytokines and chemokines, such as TGF-β1, IL-8 and IL-17.[42–44] In this sense, the role of neutrophils likes a "foe", intending to prompt the progression of IPF. However, the lung tissue of IPF actually contained few neutrophils. An alternative explanation is that neutrophils play a protective rather than detrimental role and their recruitment is intended to counteract the pathogenic process of IPF, maintaining the body's homeostasis.[14,26] Tabuena et al failed to link the BAL neutrophilia to the increased mortality in IPF.[45] And in some bleomycin-induced pulmonary fibrosis models, depletion of neutrophils did not reduce pulmonary fibrosis.[46,47] Indeed, in other tissue models, neutrophils have been reported to have an antifibrogenic effect.[48] Our results suggested that neutrophils might play a protective role against the pathogenic mechanisms of IPF, as indicated by the lower percentage of infiltrating neutrophils in IPF patients. Further investigation is required to determine whether targeting to increase the appropriate percentage of neutrophils can help to restrain the progression of IPF.

Notably, in the lung tissues of patients with IPF, we observed that neutrophil-related genes were predominantly enriched in the cytokine-cytokine receptor interaction pathway, and displayed a significant correlation with lung disease. This finding is consistent with the DEGs observed between IPF patients and healthy controls, suggesting that neutrophils might contribute to the pathogenesis of IPF through their interaction with cytokine-cytokine receptors.

In order to further investigate the importance of neutrophils in IPF, we identified hub genes related to neutrophils by intersecting genes that are both associated with neutrophils in IPF and DEGs between IPF patients and healthy controls. Subsequently, we constructed a neutrophil-related cluster using the expression data of eight hub genes. We found that DEGs between low neutrophil-infiltrating cluster and neutrophil-infiltrating cluster were involved in similar functions and pathways as those between IPF patients and healthy controls. Notably, a higher score on the PCA was able to differentiate IPF patients from healthy controls, underscoring the potential diagnostic value of our cluster. Furthermore, we developed a diagnostic model using the hub genes, which exhibited reliable diagnostic value in both the train and validation sets. Our results suggest that neutrophils may play a crucial role in IPF, and our cluster and diagnostic model can aid in the diagnosis and management of this disease.

Despite the promising findings, our study still has several limitations. We were unable to fully elucidate the role of neutrophils, and the underlying mechanisms through which neutrophils contribute to IPF pathogenesis remain unclear. Our results were solely based on data obtained from the Gene Expression Omnibus (GEO) and thus lack comprehensive experimental and clinical data. Nevertheless, we believe that our study can serve as a valuable starting point for further research, as we hope to inspire additional investigations to explore the role of neutrophils in IPF, ultimately leading to the development of new therapies that can better serve the needs of IPF patients.

Conclusion

This study aimed to investigate the correlation between neutrophils and IPF using gene expression datasets from the GEO. And through immune infiltration analysis, we revealed a decrease in the percentage of neutrophils in the lung tissues of IPF patients compared to healthy controls. Then we identified neutrophil-related genes in IPF through WGCNA and analyzed their function by enrichment analysis. Interestingly, these genes were mainly enriched in the cytokine-cytokine receptor

interaction pathway and showed a significant correlation with lung disease, consistent with DEGs between IPF and healthy controls. By taking the intersection genes between these neutrophil-related genes and DEGs between IPF and healthy controls, we identified eight hub genes related to neutrophils in IPF, including *IL13RA2, MMP16, ARG1, IL1R2, PROK2, MS4A2, PIR*, and *ZNF436*. And their expression exited significantly different between bleomycin-induced pulmonary models and controls. Using consensus cluster analysis based on the expression of these hub genes, we found that the gene signature in a low neutrophil-infiltrating cluster was similar to that in IPF, and both had a higher PCA score, further indicating a lower percentage of neutrophils in the lung tissue of IPF. Notably, through PCA, we found that a higher PCA score was able to distinguish IPF patients from healthy controls, indicating the potential diagnostic value of our cluster. Furthermore, we successfully constructed a diagnostic model for IPF using hub genes related to neutrophils, which showed good performance in terms of calibration curve, clinical impact, decision curve analysis, and ability to distinguish IPF patients from healthy controls in both the train and validation sets, providing valuable insights into the potential use of hub genes related to neutrophils as biomarkers for the diagnosis of IPF.

Overall, the findings of this study offer significant insights into the correlation between neutrophils and IPF, thereby potentially contributing to the detection and diagnosis of IPF. Furthermore, the identification of hub genes related to neutrophils presents a promising avenue for the development of potential biomarkers for IPF. We hope that through understanding the underlying mechanisms involving neutrophils in IPF, researchers and clinicians can develop targeted treatments that aim to modulate neutrophil activity and potentially halt or slow disease progression.

Ethics Approval and Consent to Participate

This study involving animals was carried out in accordance with the recommendations of the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Ethics Committee of Sun Yat-sen University. All efforts were made to minimize animal suffering and to reduce the number of animals used.

Acknowledgments

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors report no conflicts of interest in this work.

References

1. Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. *N Engl J Med*. 2018;378(19):1811–1823. doi:10.1056/NEJMra1705751
2. Moss BJ, Ryter SW, Rosas IO. Pathogenic mechanisms underlying idiopathic pulmonary fibrosis. *Annu Rev Pathol*. 2022;17:515–546. doi:10.1146/annurev-pathol-042320-030240

3. Behr J, Kreuter M, Hoeper MM, et al. Management of patients with idiopathic pulmonary fibrosis in clinical practice: the INSIGHTS-IPF registry. *Eur Respir J.* 2015;46(1):186–196. doi:10.1183/09031936.00217614

4. Olson AL, Gifford AH, Inase N, Fernández Pérez ER, Suda T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. *Eur Respir Rev.* 2018;27(150):180077. doi:10.1183/16000617.0077-2018

5. Duchemann B, Annesi-Maesano I, Jacobe de Naurois C, et al. Prevalence and incidence of interstitial lung diseases in a multi-ethnic county of Greater Paris. *Eur Respir J.* 2017;50(2):1602419. doi:10.1183/13993003.02419-2016

6. King TE Jr, Pardo A, Selman M. Idiopathic pulmonary fibrosis. *Lancet.* 2011;378(9807):1949–1961. doi:10.1016/S0140-6736(11)60052-4

7. Raghu G, Chen SY, Hou Q, Yeh WS, Collard HR. Incidence and prevalence of idiopathic pulmonary fibrosis in US adults 18–64 years old. *Eur Respir J.* 2016;48(1):179–186. doi:10.1183/13993003.01653-2015

8. Raghu G, Collard HR, Egan JJ, 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.* 2011;183(6):788–824. doi:10.1164/rccm.2009-040GL

9. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med.* 2000;161(2 Pt 1):646–664. doi:10.1164/ajrccm.161.2.ats3-00

10. Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. *Annu Rev Pathol.* 2014;9:157–179. doi:10.1146/annurev-pathol-012513-104706

11. Tzouvelekis A, Kaminski N. Epigenetics in idiopathic pulmonary fibrosis. *Biochem Cell Biol.* 2015;93(2):159–170. doi:10.1139/bcb-2014-0126

12. Ruaro B, Matucci Cerinic M, Salton F, Baratella E, Confalonieri M, Hughes M. Editorial: pulmonary fibrosis: one manifestation, various diseases. *Front Pharmacol.* 2022;13:1027332. doi:10.3389/fphar.2022.1027332

13. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. *Eur Respir J.* 2015;45(6):1717–1727. doi:10.1183/09031936.00163814

14. Shenderov K, Collins SL, Powell JD, Horton MR. Immune dysregulation as a driver of idiopathic pulmonary fibrosis. *J Clin Invest.* 2021;131(2). doi:10.1172/JCI143226

15. O'Dwyer DN, Armstrong ME, Trujillo G, et al. The Toll-like receptor 3 L412F polymorphism and disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2013;188(12):1442–1450. doi:10.1164/rccm.201304-0760OC

16. Baratella E, Ruaro B, Giudici F, et al. Evaluation of correlations between genetic variants and high-resolution computed tomography patterns in idiopathic pulmonary fibrosis. *Diagnostics.* 2021;11(5):762. doi:10.3390/diagnostics11050762

17. Michalski JE, Schwartz DA. Genetic risk factors for idiopathic pulmonary fibrosis: insights into immunopathogenesis. *J Inflamm Res.* 2020;13:1305–1318. doi:10.2147/JIR.S280958

18. Ding L, Yang J, Zhang C, Zhang X, Gao P. Neutrophils modulate fibrogenesis in chronic pulmonary diseases. *Front Med.* 2021;8:616200. doi:10.3389/fmed.2021.616200

19. Jegal Y. The role of neutrophils in the pathogenesis of IPF. *Korean J Intern Med.* 2022;37(5):945–946. doi:10.3904/kjim.2022.267

20. Chrysanthopoulou A, Mitroulis I, Apostolidou E, et al. Neutrophil extracellular traps promote differentiation and function of fibroblasts. *J Pathol.* 2014;233(3):294–307. doi:10.1002/path.4359

21. Zhang S, Jia X, Zhang Q, et al. Neutrophil extracellular traps activate lung fibroblast to induce polymyositis-related interstitial lung diseases via TLR9-miR-7-Smad2 pathway. *J Cell Mol Med.* 2020;24(2):1658–1669. doi:10.1111/jcmm.14858

22. Gregory AD, Kliment CR, Metz HE, et al. Neutrophil elastase promotes myofibroblast differentiation in lung fibrosis. *J Leukoc Biol.* 2015;98(2):143–152. doi:10.1189/jlb.3HI1014-493R

23. Kinder BW, Brown KK, Schwarz MI, Ix JH, Kervitsky A, King TE Jr. Baseline BAL neutrophilia predicts early mortality in idiopathic pulmonary fibrosis. *Chest.* 2008;133(1):226–232. doi:10.1378/chest.07-1948

24. Neudecker V, Brodsky KS, Clambey ET, et al. Neutrophil transfer of miR-223 to lung epithelial cells dampens acute lung injury in mice. *Sci Transl Med.* 2017;9(408). doi:10.1126/scitranslmed.aah5360

25. Nathan C. Neutrophils and immunity: challenges and opportunities. *Nat Rev Immunol.* 2006;6(3):173–182. doi:10.1038/nri1785

26. Ishikawa G, Liu A, Herzog EL. Evolving perspectives on innate immune mechanisms of IPF. *Front Mol Biosci.* 2021;8:676569. doi:10.3389/fmolb.2021.676569

27. Silvestre-Roig C, Fridlender ZG, Glogauer M, Scapini P. Neutrophil diversity in health and disease. *Trends Immunol.* 2019;40(7):565–583. doi:10.1016/j.it.2019.04.012

28. Metzemaekers M, Gouwy M, Proost P. Neutrophil chemoattractant receptors in health and disease: double-edged swords. *Cell Mol Immunol.* 2020;17(5):433–450. doi:10.1038/s41423-020-0412-0

29. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol.* 2013;13(3):159–175. doi:10.1038/nri3399

30. Kolahian S, Fernandez IE, Eickelberg O, Hartl D. Immune mechanisms in pulmonary fibrosis. *Am J Respir Cell Mol Biol.* 2016;55(3):309–322. doi:10.1165/rcmb.2016-0121TR

31. Parra ER, Kairalla RA, Ribeiro de Carvalho CR, Eher E, Capelozzi VL. Inflammatory cell phenotyping of the pulmonary interstitium in idiopathic interstitial pneumonia. *Respiration.* 2007;74(2):159–169. doi:10.1159/000097133

32. Carré PC, Mortenson RL, King TE Jr, Noble PW, Sable CL, Riches DW. Increased expression of the interleukin-8 gene by alveolar macrophages in idiopathic pulmonary fibrosis. A potential mechanism for the recruitment and activation of neutrophils in lung fibrosis. *J Clin Invest.* 1991;88(6):1802–1810. doi:10.1172/JCI115501

33. Raghu G, Anstrom KJ, King TE Jr, Lasky JA, Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med.* 2012;366(21):1968–1977.

34. Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med.* 2013;188(6):733–748. doi:10.1164/rccm.201308-1483ST

35. Molyneaux PL, Cox MJ, Willis-Owen SA, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2014;190(8):906–913. doi:10.1164/rccm.201403-0541OC

36. Spagnolo P, Molyneaux PL, Bernardinello N, et al. The role of the lung's microbiome in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Int J Mol Sci.* 2019;20(22):5618. doi:10.3390/ijms20225618

37. Han MK, Zhou Y, Murray S, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med*. 2014;2(7):548–556. doi:10.1016/S2213-2600(14)70069-4

38. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol*. 2008;6(4):288–301. doi:10.1038/nrmicro1871

39. Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A. Neutrophil function: from mechanisms to disease. *Annu Rev Immunol*. 2012;30:459–489. doi:10.1146/annurev-immunol-020711-074942

40. Chua F, Dunsmore SE, Clingen PH, et al. Mice lacking neutrophil elastase are resistant to bleomycin-induced pulmonary fibrosis. *Am J Pathol*. 2007;170(1):65–74. doi:10.2353/ajpath.2007.060352

41. Takemasa A, Ishii Y, Fukuda T. A neutrophil elastase inhibitor prevents bleomycin-induced pulmonary fibrosis in mice. *Eur Respir J*. 2012;40 (6):1475–1482. doi:10.1183/09031936.00127011

42. Mócsai A. Diverse novel functions of neutrophils in immunity, inflammation, and beyond. *J Exp Med*. 2013;210(7):1283–1299. doi:10.1084/jem.20122220

43. Scapini P, Cassatella MA. Social networking of human neutrophils within the immune system. *Blood*. 2014;124(5):710–719. doi:10.1182/blood-2014-03-453217

44. Kruger P, Saffarzadeh M, Weber AN, et al. Neutrophils: between host defence, immune modulation, and tissue injury. *PLoS Pathog*. 2015;11(3): e1004651. doi:10.1371/journal.ppat.1004651

45. Tabuena RP, Nagai S, Tsutsumi T, et al. Cell profiles of bronchoalveolar lavage fluid as prognosticators of idiopathic pulmonary fibrosis/usual interstitial pneumonia among Japanese Patients. *Respiration*. 2005;72(5):490–498. doi:10.1159/000087673

46. Thrall RS, Phan SH, McCormick JR, Ward PA. The development of bleomycin-induced pulmonary fibrosis in neutrophil-depleted and complement-depleted rats. *Am J Pathol*. 1981;105(1):76–81.

47. Manoury B, Nénan S, Guénon I, Lagente V, Boichot E. Influence of early neutrophil depletion on MMPs/TIMP-1 balance in bleomycin-induced lung fibrosis. *Int Immunopharmacol*. 2007;7(7):900–911. doi:10.1016/j.intimp.2007.02.009

48. Calvente CJ, Tameda M, Johnson CD, et al. Neutrophils contribute to spontaneous resolution of liver inflammation and fibrosis via microRNA-223. *J Clin Invest*. 2019;129(10):4091–4109. doi:10.1172/JCI122258