Investigation of a Hypoxia-Immune-Related Microenvironment Gene Signature and Prediction Model for Idiopathic Pulmonary Fibrosis

Xinyu Li1,2†, Haozheng Cai1†, Yufeng Cai3†, Quyan Zhang1,2, Yinghe Ding1,2 and Quan Zhuang1,4*

1 Transplantation Center, The 3rd Xiangya Hospital, Central South University, Changsha, China, 2 Xiangya School of Medicine, Central South University, Changsha, China, 3 School of Life Science, Central South University, Changsha, China, 4 Research Center of National Health Ministry on Transplantation Medicine, Changsha, China

Background: There is growing evidence found that the role of hypoxia and immune status in idiopathic pulmonary fibrosis (IPF). However, there are few studies about the role of hypoxia and immune status in the lung milieu in the prognosis of IPF. This study aimed to develop a hypoxia-immune-related prediction model for the prognosis of IPF.

Methods: Hypoxia and immune status were estimated with microarray data of a discovery cohort from the GEO database using UMAP and ESTIMATE algorithms respectively. The Cox regression model with the LASSO method was used for identifying prognostic genes and developing hypoxia-immune-related genes. Cibersort was used to evaluate the difference of 22 kinds of immune cell infiltration. Three independent validation cohorts from GEO database were used for external validation. Peripheral blood mononuclear cell (PBMC) and bronchoalveolar lavage fluid (BALF) were collected to be tested by Quantitative reverse transcriptase-PCR (qRT-PCR) and flow cytometry from 22 clinical samples, including 13 healthy controls, six patients with non-fibrotic pneumonia and three patients with pulmonary fibrosis.

Results: Hypoxia and immune status were significantly associated with the prognosis of IPF patients. High hypoxia and high immune status were identified as risk factors for overall survival. CD8+ T cell, activated CD4+ memory T cell, NK cell, activated mast cell, M1 and M0 macrophages were identified as key immune cells in hypoxia-immune-related microenvironment. A prediction model for IPF prognosis was established based on the hypoxia-immune-related one protective and nine risk DEGs. In the independent validation cohorts, the prognostic prediction model performed the significant applicability in peripheral whole blood, peripheral blood mononuclear cell, and lung tissue of IPF patients. The preliminary clinical specimen validation suggested the reliability of most conclusions.
INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial lung disease. The prognosis of patients with IPF is poor, with a median survival of 3 to 5 years (1). Several prognostic staging systems for IPF have been established by clinical and physiologic variables (2–5). Biomarkers in peripheral blood are also evaluated as a tool for prognosis (6–8). A recent study also revealed the role of genetic variability in the survival of IPF (9). In addition, molecular markers of IPF patients could also be identified by bronchoalveolar lavage (BAL) cells, and the collection of BAL cells is non-invasive compared with lung biopsy (10). However, very little is known whether molecular events in the lung milieu are predictive of outcome in IPF. Considering accurate diagnosis and personalized treatment, there is still a critical need for a way to predict the progression of IPF.

As a chronic lung disorder, the central processes in IPF are inflammation and fibrosis (11). Immune dysregulation has been considered as a promoting factor in the development of IPF, including several biomarkers associated with the prognosis of IPF (12). Inflammatory cytokines released by immune cells may activate fibroblasts, connective tissue cell proliferation, angiogenesis (11). Furthermore, hypoxia is common in the process of fibrosis in many diseases (13, 14). Excessive collagen synthesized by fibroblasts deteriorate oxygen supply and accelerate the pathological process. Studies showed the relationship between immune response and hypoxia and lung function (15). However, the underlying mechanisms have not been discussed.

With a series of genetic and bioinformatics analyses, we associated immune status with hypoxia and explore its value for the prognosis of patients with IPF. Here, we developed a hypoxia-immune-related prediction model for the prognosis of IPF, intended to provide novel ideas for accurate diagnosis and treatment at the gene level. Better knowledge of the oxygen balance control and the immune regulation involved is important to advance the development of IPF.

MATERIALS AND METHODS

Patient Cohort and Data Preparation

The discovery cohort of the study contained 176 IPF patients from the Gene Expression Omnibus (GEO, available at: https://www.ncbi.nlm.nih.gov/geo/) database (GSE70866). The tissue source of sequencing samples is the patients’ BAL cells. The microarray data of GSE70866 was based on GPL14550 Platform (Agilent-028004 SurePrint G3 Human GE v2 8 × 60K Microarray, Agilent Technologies), including 176 IPF patients’ BAL cells. Three validation cohorts were used for external validations (GSE93606, GSE28221, and GSE32537) to examine the predictive effect of the prediction method. The microarray data of GSE93606 was based on GPL11532 Platforms (Affymetrix Human Gene 1.1 ST Array, Affymetrix, Santa Clara, CA, USA), including 57 IPF patients’ peripheral whole blood. The microarray data of GSE28221 was based on GPL5175 Platforms (Affymetrix Human Exon 1.0 ST Array, Affymetrix, Santa Clara, CA, USA) and GPL6480 Platforms (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F, Agilent Technologies), including 120 IPF patients’ peripheral blood mononuclear cell. The microarray data of GSE32537 based on GPL6244 (Affymetrix Human Gene 1.0 ST Array, Affymetrix, Santa Clara, CA, USA) included 119 lung tissues with IPF. The batch effect was eliminated by sva package, which contains functions for identifying and building surrogate variables for high-dimensional data sets.

All procedures of this study complied with the protocol. For analyses of data from a public database, approval and informed consent from the local ethics committee were not required.

Identification of Hypoxia Status and Hypoxia-Related DEGs

To identify the hypoxia status, a non-linear dimensionality reduction algorithm of Uniform Manifold Approximation and Projection (UMAP) was applied, which could divide or condense a group of patients into a series of distinct clusters, according to the given hallmarks or signatures. The hallmark gene sets of hypoxia include 200 genes and were downloaded from the Molecular Signatures Database (MSigDB version 6.0). Based on the clusters, two groups including “hypoxia\textsuperscript{low}” and “hypoxia\textsuperscript{high}” were identified to identify the hypoxia status. The limma algorithm was applied to identify differentially expressed genes (DEGs) between the two groups (16). Genes with a false discovery rate (FDR) adjusted p-value <0.0001 and an absolute value of log2 (fold change) >1 were considered as hypoxia-related DEGs.

Identification of Immune Status and Immune-Related DEGs

To identify the immune status, the Estimation of Stromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) algorithm was applied to identify the infiltration degree of immune cells and predict the immune status (17). Based on the immune status, patients were classified into two groups. Maximally selected rank statistics was applied by using an R package “survival”, and “survminer” to identify the optimal cutting point to divide patients. Based on the optimal cutting
point, patients with high immune scores were attributed to “immune-high” group and “immune-low” group. The limma algorithm was applied to identify DEGs between the two groups. Genes with a FDR adjusted p-value <0.0001 and an absolute value of log2 (fold change) >1 were considered as immune-related DEGs.

To further identify the abundance of 22 immune cells, CIBERSORT is a deconvolution algorithm based on the gene expression data to resolve immune cell composition (18). Those with p <0.05 were included.

Identification of Hypoxia-Immune-Related Prognostic DEGs
According to the above hypoxia and immune grouping, patients were divided into three groups, i.e., hypoxia-high/immune-high, hypoxia-low/immune-low, and mix groups. The limma algorithm was applied to identify DEGs between “hypoxia-high/immune-high” group and “hypoxia-low/immune-low” group. Genes with a FDR adjusted p-value <0.0001 and an absolute value of log2 (fold change) >1 were considered as hypoxia-immune-related DEGs. DEGs were then divided into protective and risk DEGs. The risk DEGs contained all EDGs highly expressed in hypoxia-high/immune-high group and the rest were protective DEGs. To obtain hypoxia-immune-related prognostic DEGs, univariate Cox analyses were further performed to screen all protective and risk DEGs. Those with a p <0.001 were considered as significant.

Prognosis Prediction Model of IPF Based On Hypoxia-Immune-Related DEGs
The Least Absolute Shrinkage and Selection Operator (LASSO) is a kind of linear regression using shrinkage, which is applied to survival analysis with high-dimensional data (19). In this study, the LASSO Cox regression model was applied to select the optimal variables from all identified hypoxia-immune-related prognostic DEGs in the discovery cohort. Three-fold cross-validation and 1,000 iterations were conducted to reduce the potential instability of the results. The optimal tuning parameter λ was identified via 1-SE (standard error) criterion. Then we create the prognosis prediction model of IPF using the selected prognostic gene signature. For each patient, the risk score was the sum of the expression of the characteristic DEGs and the corresponding coefficients derived from the multivariate Cox regression model. According to the risk scores, the optimal cutting point was identified using the maximally selected rank method, and the prognosis prediction model of IPF was formed.

Functional and Pathway Enrichment Analysis
Database for Annotation, Visualization and Integrated Discovery was used for Gene Ontology (GO) enrichment analysis (20). The risk DEGs of IPF patients were screened for functional enrichment. GO analysis was used to evaluate the degree of enrichment of the DEGs in biological processes, cellular components, and molecular functions. Those with p-value <0.05 and count (the number of enriched genes) ≥3 were considered as the cutoff criterion.

Preliminary Validation of Clinical Specimens
Peripheral blood mononuclear cell (PBMC) and bronchoalveolar lavage fluid (BALF) were collected from 22 clinical samples, including 13 healthy controls, six with non-fibrotic pneumonia and three with pulmonary fibrosis. The clinical information such as age, gender, alveolar-arterial oxygen gradient (A-aDO2), and hospital day was shown in Supplementary Materials. The study was reviewed and approved by the institutional review board (Ethics Committee) of the 3rd Xiangya Hospital, Central South University (No. 21028).

Quantitative reverse transcriptase-PCR (qRT-PCR) was used to quantify expression of key DEGs. Total RNA was extracted from the tissues using TRIzol Reagent (Thermo Fisher Scientific). PCR was performed using an Thermo Scientific PikoReal PCR cycler. The cycle threshold (CT) data were determined, and the mean CT was determined from triplicate PCRs. Relative gene expression was calculated with the equation ΔΔCT.

Flow cytometry analysis was used to determine the proportion of immune cells. The cell suspension was counted and mixed with ACK Lysis Buffer (Thermo Fisher Scientific) to remove red blood cells. Then 1 × 10^6 cells were resuspended in 100 µl staining buffer and incubated with monoclonal antibodies in dark for 15 min at 4°C. Our flow cytometric staining strategy consisted of the following fluorochrome-conjugated monoclonal antibodies: anti-CD3-Alexa-Flour700 (Biolegend), anti-CD4-eFlour450 (eBioscience), anti-CD45RA-APC-eFlour 780 (eBioscience), anti-CCR7-PerCP-eFlour 710 (eBioscience), anti-CD16-eFlour506 (eBioscience), anti-CD56-PE (eBioscience), anti-CD206-PE-Cyanine7 (eBioscience), anti-CD68-FIITC (eBioscience), anti-CD107a-eFlour660 (eBioscience). CD3 and CD4 were used to identify T cells. CD3^+CD4^+CD45RA^−CCR7^− cells were defined as activated CD4^+ memory T cells. CD56 and CD16 were used to identify NK cells. CD107a^- NK cells were defined as activated NK cells. CD68^+CD16^+CD206^- cells were defined as M1-like macrophages and CD68^-CD16^-CD206^+ cells were defined as M0-like macrophages. After washing and resuspending, samples were detected using BD FACSDivia software and performed using BD FACS Canto II.

Statistical Analysis
All analyses were performed with R version 4.0.2 (www.r-project.org/) and the corresponding packages. UMAP algorithm was performed by using R package “umapr” for non-linear dimensionality reduction. Immune score was performed by using R package “estimate”. The Lasso Cox regression model was performed by using R package “glmnet”. Data were analyzed with standard statistical tests as appropriate. Multiple testing was adjusted by the FDR method. Multivariate Cox regression analysis was performed to identify optimal signatures. Flowjo V 10.6.2 was used to analyzed flow cytometric data. Original data from PCR and flow cytometry were presented as the mean ± standard deviation (SD) and were compared using Student’s t-test, Welch’s t-test or the Mann–Whitney U test, where appropriate. GraphPad Prism 7.0 (GraphPad Software Inc., La Jolla, California) was used for Gene Ontology (GO) enrichment analysis (20). The Enrichment Analysis Functional and Pathway component was used to evaluate the degree of enrichment of the DEGs in biological processes, cellular components, and molecular functions. Those with p-value <0.05 and count (the number of enriched genes) ≥3 were considered as the cutoff criterion.
Jolla, CA, USA) was used to perform the statistical analyses. Values of p <0.05 were considered statistically significant.

The general idea and methodologies used in this study were drawn as a flow chart (Figure 1).

RESULTS

Hypoxia Status and Hypoxia-Related DEGs in IPF

The discovery cohort contained 176 IPF patients from the GEO databases. The batch effect was eliminated by sva package (Figures 2A, B). Patient clinical information is shown in Table 1. With the expression matrix constructed by 200 hypoxia marker genes from MSigDB, the non-linear dimensionality reduction algorithm UMAP was used to determine two clusters, and each patient is assigned to the nearest clusters (Figure 2C). Cluster 1 and Cluster 2 contained 95 and 81 patients respectively. Expression profiles were compared between the two clusters, and 239 DEGs related to hypoxia were obtained. Enrichment analysis showed overexpressed genes in Cluster 2 were enriched in “oxygen transport (GO:0015671)” and “response to hypoxia (GO:0001666)” (Figure 2D). This indicated that the level of hypoxia in Cluster 2 was at high status. Thereby, the patients in Cluster 1 and Cluster 2 were determined as hypoxia-low and hypoxia-high groups. In addition, overexpressed genes in Cluster 2 were also enriched in positive regulation of GTPase activity and cell adhesion. Patients’ clinical information of each cluster is shown in Table 2. The survival status of patients in different groups was further analyzed (Figure 2E). There was a significant difference in survival between two clusters (log rank test, p < 0.0001), and the prognosis of patients with a high level of hypoxia is worse. Among 239 DEGs, 232 DEGs were overexpressed in the hypoxia-high cluster, which were regarded as hypoxia-associated risk DEGs. The other seven genes overexpressed in the hypoxia-low cluster, which are regarded as hypoxia-associated protective DEGs. In a word, most of the hypoxia-related DEGs are regarded as risk factors.

Immune Status and Immune-Related DEGs in IPF

The immune score was calculated by ESTIMATE to identify the infiltration degree of immune cells. The optimal cutting point “2,959.22” was determined based on maximally selected rank statistics (Figures 2F, G). Then the immune-high group and immune-low group were divided, containing 108 and 68 patients respectively. Patients’ clinical information of each cluster is shown in Table 3. Further survival analysis showed a significant difference between two groups (log rank test, p < 0.05), and the survival of patients with a high level of immune infiltration is worse (Figure 2I). Therefore, high immune infiltration is also a risk factor for bad prognosis. This conclusion is also supported by the highly enriched immune-related pathways in the hypoxia-high group with a poor prognosis. Expression profiles were compared between the two groups, and 196 DEGs related to immune status were obtained (Figure 2H). Among them, 191 genes were overexpressed in the immune-high group.
FIGURE 2  |  (A, B) Eliminating the batch effect between different sequencing platforms. (A) The PCA plot before elimination of batch effect, and (B) is the PCA plot after elimination. The distance of sample point clusters indicates that they come from different batches and sequencing platforms. While in B, after eliminating the batch effect, the difference in distance between batches is reduced. (C) UMAP clustering plot based on marker gene set of hypoxia. (D) Biological process functional enrichment analysis of differentially expressed genes between hypoxia\textsuperscript{high} and hypoxia\textsuperscript{low} groups. (E) Kaplan–Meier plot of overall survival in two clusters. (F) Histogram based on maximally selected rank grouping. (G) The cut-off point with the maximum standard log-rank statistic was marked with a vertical dashed line. (H) The differential gene expression profiles between hypoxia\textsuperscript{high} and hypoxia\textsuperscript{low} groups were visualized in heatmap. (I) Kaplan–Meier plot of overall survival between immune\textsuperscript{high} and immune\textsuperscript{low} groups.
cluster, which were regarded as immune-associated risk DEGs. The other five genes overexpressed in the immune low cluster, which are regarded as immune-associated protective DEGs.

### Hypoxia–Immune-Related DEGs in IPF

According to the above hypoxia and immune grouping, we further combined to form three groups: hypoxia\textsuperscript{high}/immune\textsuperscript{high}, hypoxia\textsuperscript{low}/immune\textsuperscript{low}, and mix groups. The survival status of patients in different groups was further analyzed (Figure 3A). There was a significant difference in survival among three groups (log rank test, \( p < 0.0001 \)). Survival in mix group is at an intermediate level. As we expected, a high level of hypoxia and immune activity is the most dangerous factor while patients in group hypoxia\textsuperscript{low}/immune\textsuperscript{low} have the best prognosis.

The differential gene expression profiles between hypoxia\textsuperscript{high}/immune\textsuperscript{high} and hypoxia\textsuperscript{low}/immune\textsuperscript{low} groups were visualized in heatmap (Figure 3B). We further intersect the hypoxia-related DEGs and the immune-related DEGs to identify the hypoxia–immune-related DEGs in IPF. We obtained a total of 62 DEGs, of which 61 were highly expressed in hypoxia\textsuperscript{high} and immune\textsuperscript{high} groups, so they were defined as hypoxia-immune-related risk DEGs (Figure 3C). Correspondingly, the remaining DEG is defined as hypoxia-immune-related protective DEG (Figure 3D). The GO enrichment analysis showed that “immune response”, “inflammatory response”, and “positive regulation of ERK1/2 cascade” are main biological process (Figure 3E).

### Prognosis Prediction Model of IPF Based on Hypoxia-Immune-Related DEGs

To further determine the DEGs significantly related to the prognosis, we used univariate Cox analysis for screening and 29 DEGs with \( p < 0.001 \) were retained (Figure 4A). Among them, one protective and 28 risk DEGs were included.
Using lasso regression method, nine optimal variables were obtained from the above 29 hypoxia-immune-prognostic-related DEGs (Figures 4B, C). Then we use the expression levels of nine characteristic DEGs and the corresponding coefficients derived from the multivariate Cox regression model to estimate the risk score for each patient: risk score = −0.13307 × expression of NALCN + 0.09893 × expression of IL1R2 + 0.06226 × expression of S100A12 + 0.06890 × expression of PROK2 + 0.04883 × expression of CCL8 + 0.05654 × expression of RAB15 + 0.10671 × expression of MARCKSL1 + 0.09986 × expression of TPCN1 + 0.05696 × expression of HS3ST3B1.

By calculating the risk score of each patient, we divided the patients into two groups through the maximally selected rank method: high-risk group and low-risk group (Figure 4D). The
nine optimal DEG expression profiles between high-risk and low-risk groups were visualized in heatmap (Figure 4E). Survival analysis showed that there was a significant difference between the two groups (log rank test, p < 0.0001). Compared with the low-risk group, the prognosis of the high-risk group is significantly worse (Figures 4F–H).
Hypoxia-Immune Related Immunocyte Infiltration Pattern

In addition, CIBERSORT was used to estimate the infiltration of 22 kinds of immune cells in the samples. Correlation analysis showed a general association in different immune cells (Figure 5A). Among them, the infiltration of six specific immune cells was significantly different in hypoxia high/immune high and hypoxia low/immune low group, that is, CD8+ T cell, activated CD4+ memory T cell, activated natural killer (NK) cell, activated mast cell, M0 macrophage, M1 macrophage (Figure 5B). Further correlation analysis showed the relationship between six specific immune cells and risk score. Among them, most infiltration degree is positively correlated with risk DEGs expression and risk score, but the infiltration of M0 cells was negatively correlated with the risk DEGs expression and risk score (Figure 5C). Among them, M0 macrophages and NK cells had the most significant correlation with key DEGs and risk score.

Validation of the Prognostic Prediction Model in External Independent Cohorts

The ROC curve showed that the AUCs within 1–5 years were all greater than 0.75 in discovery cohort (Figure 6A). This suggested the evaluation model had a good predictive value for the prognosis of IPF patients. We also developed a nomogram for 1–5 years overall survival prediction based on Cox model (Figure 6B).

We further verify the above prediction method in external data sets “GSE93606”, “GSE28221”, and “GSE32537”. Patient clinical information is shown in Table 4. In each independent validation cohort, we divided the IPF patients into high-risk and low-risk groups based on the risk score. In GSE28221 and GSE93606 IPF cohorts, survival comparison showed that low-risk group had significantly better prognosis outcomes than high-risk group (Figures 6C, D). In addition, we focused on the clinicopathologic features of IPF in GSE32537 cohort. The patients in low-risk group generally had higher forced vital capacity (FVC) and carbon monoxide diffusing capacity.
(DLCO) (P < 0.05), which meant better lung function (Figures 6E, F). While the patients in high-risk group had higher St. George’s total score (P < 0.05), which suggested worse lung function and quality of life (Figure 6G).

In a word, these results suggest that our prognostic prediction model is also of great significance based on peripheral whole blood, peripheral blood mononuclear cell, and lung tissue.

**Preliminary Validation of Clinical Specimens**

The results of qRT-PCR in PBMC suggested that DEGs of CCL8, IL1R2, NALCN, S100A12, and PROK2 were significantly different between the patients and healthy controls (Figure 7A). Among the patients, CCL8, IL1R2, and PROK2 were significantly up-expressed in fibrotic samples than non-

---

**FIGURE 6** | (A) ROC curve evaluated the predictive value of the model for the prognosis of patients in discovery cohort within 1–5 years. (B) The nomogram for 1–5 years overall survival based on Cox model. (C) Kaplan–Meier plot of overall survival between high-risk and low-risk groups in GSE28221 validation cohort. (D) Kaplan–Meier plot of overall survival between high-risk and low-risk groups in GSE93606 validation cohort. (E–G) The box plot showed the difference of FVC, DLCO, and George’s score between low-risk group and high-risk groups in GSE32537 validation cohort.
fibrotic samples (Figure 7B). The PBMC results of flow cytometry showed that the proportion of NK cells was increased significantly in the patients than healthy controls (Figure 7C), and among the patients’ BALF, the proportion of NK cells was also increased significantly in fibrotic samples than in non-fibrotic ones (Figure 7D). The proportion of activated NK cells in BALF samples was significantly higher than those in PBMC samples, and it had the following characteristics: the increasing trend of peripheral blood to pulmonary bronchus (Figure 7E). The proportion of NK or activated NK cells’ infiltration in BALF was positively correlated with patients A-aDO2 and hospital day, suggesting that high NK infiltration is a risk factor for poor prognosis (Figures 7F, G). In addition, the flow cytometry showed that the increasing CD4+ T cells in peripheral blood might promote the macrophage infiltration in pulmonary bronchus and promoted their polarization to M1-like phenotype (Figure 7H). The proportion of CD4+ T, activated CD4+ memory T cells, and M1-like macrophages infiltration in BALF or PBMC was positively correlated with patients’ A-aDO2 (Figures 7I–K). Figure 7L showed the relationship between the results of BALF and PBMC. The detailed data of sample information and experimental results of BALF/PBMC were shown in the Supplementary Materials.

**DISCUSSION**

Since the prognosis of patients with IPF is poor, the importance of building a prognostic staging system for personalized treatment is self-evident. The prognostic staging system could divide the patients into several groups according to the given markers. In this study, we used the transcriptional profiles of the bronchoalveolar lavage fluid (BALF) to analyze the relationship between the level of biomarkers and the prognosis of patients. We found that both hypoxia and immune status were related to the survival and even respiratory function of patients with IPF. Furthermore, we established a new prognostic classifier including nine-gene signature for patients with IPF. It is effective in the prognosis of patients with IPF in the discovery cohort and three independent validation cohorts. These findings provide a new insight to the relationship between biomarkers in the lung milieu and the prognosis and stratification of patients with IPF.

Several articles reported the role of immune and hypoxia microenvironment in lung diseases (15). On one hand, immune dysregulation and inflammation are regarded as the basis of chronic lung diseases, including IPF (11). Bioinformatics analysis of RNA network and immune infiltration showed that immune cells were associated with the severity of IPF (21). Both innate and adaptive immunity were activated in fibrogenesis (22). On the other hand, hypoxia is common in lung disease. Hypoxia-inducible factor-1α (HIF-1α) is a key regulating factor in cell response to hypoxia, which has been found to participate in many lung diseases (23–25). In hypoxia, the activation of HIF-1α mediates glycolysis modification, angiogenesis, and other adaptive mechanism (26, 27). Hypoxia promoted the epithelia–mesenchymal transition (EMT) of alveolar epithelial cells (AECs) in IPF, and transforming growth factor β (TGF-β) also promoted EMT with increased lactic acid produced by metabolic modification (24, 28). Also, HIF-1α was found to be active in fibroblasts from IPF patients and induced myofibroblast differentiation with the existence of TGF-β (24, 26, 29, 30). It is worth mentioning that hypoxia facilitated proliferation and the secretion of proinflammatory cytokines in mast cells, and thus influenced fibrogenesis (31). In IPF, alveolar macrophages showed a perturbation of mitochondria homeostasis, including increased mitochondria reactive oxygen species (mtROS), in which HIF-1α may have participated (32). These findings were in accordance with the results. The result of infiltration of different
types of immune cells showed that both innate immunity and adaptive immunity were activated in hypoxia\textsuperscript{high}/immune\textsuperscript{high} group, presenting poorer prognosis. In a word, hypoxia, as the inducement of immune activation, mediates chronic airway inflammation and leads to fibrosis. Then, cell and organ dysfunction caused by fibrosis aggravates the formation of hypoxic inflammatory microenvironment, which forms a feedback loop.

The result of CIBERSORT showed that the infiltration of M0 macrophage in hypoxia\textsuperscript{low}/immune\textsuperscript{low} group was higher than that in hypoxia\textsuperscript{high}/immune\textsuperscript{high} group. As inactive and naive macrophages, the low proportion of M0 macrophages in the low-risk group suggested a lower level of inflammatory activation (33). At the same time, the infiltration of CD8\textsuperscript{+} T cell, activated CD4\textsuperscript{+} memory T cell, activated NK cell, activated mast cell, and M1 macrophage in hypoxia\textsuperscript{high}/immune\textsuperscript{high} group was higher than that in hypoxia\textsuperscript{low}/immune\textsuperscript{low} group, presenting a higher level of inflammation. As was discussed above, the role of CD4\textsuperscript{+} T cell, mast cell, and M1 macrophage in fibrosis was widely reported. It is worth noting that M0 macrophage and NK cell had the most significant correlation with key hypoxia-immune DEGs and risk score. However, the role of CD8\textsuperscript{+} T cell, NK cell, and M0 macrophage in fibrosis has not been fully verified. These may provide a new idea to understand the characteristic immune cells in fibrotic infiltration.

Also, we found that hypoxia-immune DEGs were mainly risk DEGs, taking part in the integral component of membrane and immune response, and the protective DEG was a gene encoding voltage-independent, non-selective cation channel. The roles of these predictive signature genes have been reported previously in

![Image with diagrams and text](https://example.com/diagram.png)
lung diseases. Interleukin-1 receptor 2 (IL-1R2) is an anti-inflammatory cytokine. The increase level of IL-1R2 has been reported to be associated with poor prognosis in lung cancer (34, 35), and the increase level of IL-1 was associated with the development of chronic obstructive pulmonary disease (COPD) (36, 37). Another risk gene S100A12 is a novel inflammatory disease biomarker in acute respiratory distress syndrome (ARDS) (38), interstitial lung disease (ILD) (39), and COPD (40). Moreover, C-C Motif Chemokine Ligand 8 (CCL8) is a kind of monocyte chemoattractant, regulating group 2 ILCs in lung inflammation (41). These results were in accordance with the results in this article that the overexpression of IL-1R2, S100A12, and ILC2s may be predictive for poor diagnosis of IPF patients. NALCN gene encodes a voltage-independent, non-selective cation channel. Other signature genes are rarely reported in lung diseases. Interleukin-1 receptor 2 (IL-1R2) is an anti-inflammatory cytokine. The increase level of IL-1R2 has been reported to be associated with poor prognosis in lung cancer (34, 35), and the increase level of IL-1 was associated with the development of chronic obstructive pulmonary disease (COPD) (36, 37). Another risk gene S100A12 is a novel inflammatory disease biomarker in acute respiratory distress syndrome (ARDS) (38), interstitial lung disease (ILD) (39), and COPD (40). Moreover, C-C Motif Chemokine Ligand 8 (CCL8) is a kind of monocyte chemoattractant, regulating group 2 ILCs in lung inflammation (41). These results were in accordance with the results in this article that the overexpression of IL-1R2, S100A12, and ILC2s may be predictive for poor diagnosis of IPF patients. NALCN gene encodes a voltage-independent, non-selective cation channel. Other signature genes are rarely reported in lung diseases. In this study, combing hypoxia and immune status, we identified these signature genes to provide new insights into the prognosis of IPF.

In particular, in the independent external validations of this study, the prognostic prediction model was performed in peripheral whole blood, peripheral blood mononuclear cell, and lung tissue of IPF patients, and the outcomes were significant. The preliminary clinical specimen validation suggested the reliability of most conclusions, but there are still limitations, such as insufficient sample size. And because of the individual differences and other confounding factors, the results based on the existing database must have some deviation from the reality. Although these results provided more possibilities and a wider application of this predictive model in clinical setting, a well-designed and multi-center study is needed for further exploration.

CONCLUSIONS

The immune and hypoxia status in alveolar molecular environment is associated with the prognosis of patients with IPF. The prognostic model based on several signature genes raised a new way to predict the progression and prognosis of IPF.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Gene Expression Omnibus (GEO, available at: https://www.ncbi.nlm.nih.gov/geo/) database (GSE70866).

REFERENCES

1. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic Pulmonary Fibrosis. Nat Rev Dis Primers (2017) 3:17074. doi: 10.1038/nrdp.2017.74
2. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Ascertainment of Individual Risk of Mortality for Patients With Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med (2011) 184 (4):459–66. doi: 10.1164/rccm.201111-1790OC
3. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomasetti S, Lee JS, et al. A Multidimensional Index and Staging System for Idiopathic Pulmonary Fibrosis. Ann Intern Med (2012) 156(10):684–91. doi: 10.7326/0003-4819-156-10-201205150-00004
4. Ley B, Bradford WZ, Weycker D, Vittinghoff E, du Bois RM, Collard HR. Unified Baseline and Longitudinal Mortality Prediction in Idiopathic Pulmonary Fibrosis. Eur Respir J (2015) 45(5):1374–81. doi: 10.1183/09031936.00416314
5. Kamiya H, Panaquio OM. Systematic Review and Meta-Analysis of Prognostic Factors of Acute Exacerbation of Idiopathic Pulmonary Fibrosis. BMJ Open (2020) 10(6):e035420. doi: 10.1136/bmjopen-2019-035420
6. Richards TJ, Kaminski N, Barbaud F, Flavin S, Bredmerkel C, Horowitz D, et al. Peripheral Blood Proteins Predict Mortality in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med (2012) 185(1):67–76. doi: 10.1164/rccm.201101-0058OC

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board (Ethics Committee) of the 3rd Xiangya Hospital, Central South University (No. 21028). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The study was conceived and designed by QZhu. Statistical analyses were performed by XL and YC. Software package was prepared by QYZ and YD. Flow cytometry was performed by XL and HC. Manuscript was written by XL, HC, and QZhu. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (81700658) and the Hunan Provincial Natural Science Foundation-Outstanding Youth Foundation (2020JJ3058).

ACKNOWLEDGMENTS

We thanked Ms. Hanying Liu from Department of Pulmonary Disease. the 3rd Xiangya Hospital of Central South University for supplying clinical samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.629854/full#supplementary-material

Supplementary Figure 1 | Kaplan–Meier plot of overall survival between hypoxia<sub>low</sub>/immune<sub>high</sub>, hypoxia<sub>high</sub>/immune<sub>low</sub>, hypoxia<sub>low</sub>/immune<sub>low</sub> and hypoxia<sub>high</sub>/immune<sub>high</sub> groups.
7. Li X, Zhou Y, Zou R, Chen H, Liu X, Qiu X, et al. Associations of Serological Biomarkers of Sars-cov-2. J. Int Med Res (2020) 21(7):4534. doi: 10.1136/jimr.2020.054772
8. Rodriguez-Reyna TS, Ferrara A, Gennari D, et al. Detection of CD4 T cells in lung tissue from patients with idiopathic pulmonary fibrosis. J Thorac Dis (2020) 12(4):1311. doi: 10.21037/jtd.2019.12.31
9. Wang Z, Qu S, Yu M, et al. Changes in pulmonary fibrosis and fibrotic diseases. J Am Soc Nephrol (2021) 32(2):361. doi: 10.1681/ASN.2020081726
10. Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, et al. Bal Cell Pneumonias. J Pathol (2019) 248(1):79. doi: 10.1002/path.5164
11. Zhang M, Zhang S. T Cells in Fibrosis and Fibrotic Diseases. J Appl Physiol (2021) 121(1):1108–19. doi: 10.1152/japplphysiol.00363.2020
12. Steel M, Lunden LG, Coldren CD, Murphy E, et al. Hypoxia-Related Microenvironment in IPF. J Immunol (2020) 165:4534. doi: 10.4049/jimmunol.1900717
13. Romero-Ragó M, Fernández-Sánchez F, Rodríguez-Reyna TS, Ruendia-Roldán I, Guitrón-Castillo G, Núñez-Alvarez CA, et al. The Transcription Factor SCX is a Potential Serum Biomarker of Fibrotic Diseases. J Med Sci (2020) 21(4):5012. doi: 10.3390/jms21145012
14. Wiertz LA, Moll SA, Seeliger B, Barlo NP, van der Vis JJ, Korthagen NM, et al. Genetic Variation in CCL18 Gene Influences CCL18 Expression and Correlates With Survival in Idiopathic Pulmonary Fibrosis. Part a. J Clin Med (2020) 9 (6):1940. doi: 10.3390/jcm9061940
15. Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, et al. Bal Cell Pneumonias. J Pathol (2019) 248(1):79. doi: 10.1002/path.5164
16. 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. doi: 10.21037/jtd-19-2842
17. Zhang M, Zhang S. T Cells in Fibrosis and Fibrotic Diseases. Front Immunol (2021) 12:1142. doi: 10.3389/fimmu.2020.01142
18. Shukla SD, Walters EH, Simpson JL, Keeley S, Wark PA, O'Toole RF, et al. Hypoxia-Inducible Factor and Bacterial Infections in Chronic Obstructive Pulmonary Disease. Respir Res (Carlton Vic) (2020) 21(5):53–63. doi: 10.1186/s12931-020-11372
19. Chen X, Liu M, Li L, Chen L. Involvement of the Warburg Effect in non-Tumor Diseases Processes. J Cell Physiol (2018) 233(4):2839–49. doi: 10.1002/jcp.25998
20. Tzouvelekis A, Harokopos V, Paparountas T, Oikonomou N, Chatziioannou A, Vilalar G, et al. Comparative Expression Profiling in Pulmonary Fibrosis Suggests a Role of Hypoxia-Inducible factor-1alpha in Disease Pathogenesis. Am J Respir Crit Care Med (2007) 176(1):1108–19. doi: 10.1154/jcrc.200701-6805C
21. Contraseras-Lopez R, Elizondo-Vega R, Paredes MJ, Luque-Campos N, Torres Tejedor G, et al. Hypoxia-Dependent Metabolic Reprogramming Governs Mesenchymal Stem/Stromal Cell Immunoregulatory Functions. FASEB J (2020) 34(6):8250–64. doi: 10.1096/fj.20190232R
22. Delbré E, Uzunhan Y, Somare A, Gille T, Marchant D, Planès C, et al. Er Stress Is Involved in Epithelial-To-Mesenchymal Transition of Alveolar Epithelial Cells Exposed to a Hypoxic Microenvironment. Int J Mol Sci (2019) 20(6):1299. doi: 10.3390/ijms20061299
23. Zhao X, Kwan JY, Yin K, Liu PP, Liu FF. Targeting Metabolic Dysregulation for Fibrosis Therapy. Nat Rev Drug Discov (2020) 19(1):57–75. doi: 10.1038/s41573-019-0040-5
24. Bagolgi E, Refini RM, d'Alessandro M, Bergantini L, Cameli P, Vantaggiato L, et al. Metabolic Dysregulation in Idiopathic Pulmonary Fibrosis. J Int Med Res (2020) 48(4):1027–39. doi: 10.1177/0300060519887841
25. Baines KJ, Fu JJ, McDonald VM, Gibson PG. Airway Gene Expression of IL-1 Pathway Mediators Predicts Exacerbation Risk in Obstructive Airway Disease. Int J Chronic Obstructive Pulmonary Dis (2017) 12:541–50. doi: 10.2147/copd.S119443
26. Zhang M, Han N, Shen Y. S100A12 Promotes Inflammation and Cell Apoptosis in Sepsis-Induced ARDS Via Activation of NLRP3 Inflammasome Signaling. Mol Immunol (2019) 123:38–48. doi: 10.1016/j.molimm.2019.03.022
27. Lou Y, Zheng Y, Fan B, Zhang L, Zhu F, Wang X, et al. Serum S100A12 Levels Are Correlated With Clinical Severity in Patients With Dermatomyositis-Associated Interstitial Lung Disease. J Int Med Res (2019) 47(4):1187. doi: 10.1177/0300060518887841
28. Dai H, Huang YJ, He X, Zhao M, Wang X, Liu ZS, et al. Acetylation Blocks Cgas Activity and Inhibits Self-DNA-Induced Autoimmunity. Cell (2017) 176 (6):1477–80. doi: 10.1016/j.cell.2019.01.016. e14
29. Mindt BC, Fritz JH. Next Stop: Perivasculature! ILC2s Hitch a Ride on the CLEC8B Express. Sci Immunol (2019) 4(18):130. doi: 10.1126/sciimmunol.aax4583