Genes related to N6-methyladenosine in the diagnosis and prognosis of idiopathic pulmonary fibrosis

Jingcheng Zhang[1], Ying Zhang[2], Ziyuan Wang[3], Jiachao Zhao[4], Zhenyu Li[3], Keju Wang[3], Lin Tian[2], Baojin Yao[1], Qibiao Wu[5,6,7]*, Tan Wang[2]* and Jing Wang[1,2]*

[1]Northeast Asia Research Institute of Traditional Chinese Medicine, Changchun University of Chinese Medicine, Changchun, China, [2]Department of Respiratory, The Affiliated Hospital to Changchun University of Chinese Medicine, Changchun, China, [3]College of Traditional Chinese Medicine, Changchun University of Chinese Medicine, Changchun, China, [4]College of Integrated Traditional Chinese and Western Medicine, Changchun University of Chinese Medicine, Changchun, China, [5]State Key Laboratory of Quality Research in Chinese Medicines, Faculty of Chinese Medicine, Macau University of Science and Technology, Macau, China, [6]Guangdong-Hong Kong-Macao Joint Laboratory for Contaminants Exposure and Health, Guangdong University of Technology, Guangzhou, China, [7]Zhuhai MUST Science and Technology Research Institute, Zhuhai, China

Introduction: Idiopathic pulmonary fibrosis (IPF) is a chronic progressive pulmonary fibrotic disease with unknown etiology and poor outcomes. It severely affects the quality of life. In this study, we comprehensively analyzed the expression of N6-methyladenosine (m6A) RNA methylation regulators using gene expression data from various tissue sources in IPF patients and healthy volunteers.

Methods: The gene expression matrix and clinical characteristics of IPF patients were retrieved from the Gene Expression Omnibus database. A random forest model was used to construct diagnosis signature m6A regulators. Regression analysis and correlation analysis were used to identify prognosis m6A regulators. Consensus cluster analysis was used to construct different m6A prognosis risk groups, then functional enrichment, immune infiltration and drug sensitivity analysis were performed.

Result: Five candidate m6A genes from lung tissue were used to predict the incidence, and the incidence was validated using datasets from bronchoalveolar lavage fluid (BALF) and peripheral blood mononuclear cells. Subsequently, the BALF dataset containing outcomes data was used for the prognosis analysis of m6A regulators. *METTL14*, *G3BP2*, and *ZC3H13* were independent protective

---

Abbreviations: IPF, idiopathic pulmonary fibrosis; m6A, N6-methyladenosine; GEO, Gene Expression Omnibus; BLAF, bronchoalveolar lavage fluid; PBMC, peripheral blood mononuclear cell; PCA, principal component analysis; RF, random forest; SVM, support vector machine; DCA, decision curve analysis; ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval; LASSO, least absolute shrinkage and selection operator; GAP, global alignment and proportion; DLCO, diffusing capacity of the lungs for carbon monoxide; CDF, cumulative distribution function; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological processes; CC, cellular components; MF, molecular function; ssGSEA, single-sample gene set enrichment analysis.

factors. Using correlation analysis with lung function in the lung tissue-derived dataset, *METTL14* was a protective factor in IPF. Based on *METTL14* and *G3BP2*, a consensus cluster analysis was applied to distinguish the prognostic m6A regulation patterns. The low-risk group's prognosis was significantly better than the high-risk group. Biological processes regulated by various risk groups included fibrogenesis and cell adhesion. Analysis of immune cell infiltration showed upregulation of neutrophils in the m6A high-risk group. Subsequently, five m6A high-risk group sensitive drugs and one m6A low-risk group sensitive drug were identified.

Discussion: These findings suggest that m6A regulators are involved in the diagnosis and prognosis of IPF, and m6A patterns are a method to identify IPF outcomes.

KEYWORDS

idiopathic pulmonary fibrosis, N6-methyladenosine, *METTL14*, diagnosis, prognosis

1 Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive fibrotic lung disease with unknown causes. The incidence of IPF is rising, and the prognosis are extremely poor, severely affecting the quality of life (Mei et al., 2021). Cardinal symptoms of IPF include progressive dyspnea and dry cough (Hochhegger et al., 2019; Janowiak et al., 2022). Currently, there is no effective treatment for IPF, and the median survival time is 3 to 5 years. The overall survival is poor, and there are individual differences. Despite significant progress in understanding its pathogenesis, it remains difficult reliably to predict the disease course and individual patient responses to treatment. Serum biomarkers such as osteonectin, matrix metallopeptidase-7, intercellular adhesion molecule-1, and periostin are somewhat predictive for diagnosis, prognosis, and treatment responses (Drakopanagiotakis et al., 2018; Clynick et al., 2022). Nevertheless, there remains a need for biomarkers that predict IPF onset and prognosis.

N6-methyladenosine (m6A) RNA modification involves methyltransferase catalyzing the methylation of adenine at the 6N position. The methyltransferase complex catalyzes the formation of m6A (Bokar et al., 1994), including three modifications. The m6A readers include *EIF3A*, *FMR1*, *G3BP1/2*, *HNRNPA2B1*, *HNRNPC*, *HNRNPG*, *LRPPRC*, *IGF2BP1/2/3*, *PRRC2A*, *RBMX*, *YTHDC1/2*, *YTHDF1/2/3*, and *SND1* (Lan et al., 2019; Lan et al., 2021; Zhang et al., 2021; Li et al., 2022); The m6A writers include *CBLL1*, *KIAA1429*, *METTL3*, *METTL5*, *METTL14*, *METTL16*, *PCIF1*, *RBM15*, *RBM15B*, *VIRMA*, *WTAP*, *ZC3H13*, *ZCCHC4*, and *HAKAI* (Wang et al., 2015; Xiao et al., 2016; Wu et al., 2018; Lan et al., 2019; Lan et al., 2021; Zhang et al., 2021; Li et al., 2022; Qu et al., 2022); m6A erasers include *FTO*, *ALKBH3* and *ALKBH5* (Lan et al., 2019; Gao et al., 2021; Lan et al., 2021; Li et al., 2022). Writers and erasers are in the nucleus, where m6A binds a specific nuclear reader, affecting mRNA shearing and other nuclear processes.

After export to the cytoplasm, m6A binds to a specific cytoplasmic reader, affecting mRNA stability, translation, and localization (Zaccara et al., 2019; Jiang et al., 2021).

Regulators of m6A RNA methylation modify the course of lung cancer and pneumonia. The m6A regulators that aid the diagnosis and prognosis of IPF haves not yet been discovered. Therefore, we performed a preliminary analysis of m6A-regulated genes based on high-throughput sequencing data of clinical samples. RNA methylation is the most abundant modification in eukaryotes; m6A regulators may be involved in IPF pathogenesis and progression and influence outcomes by regulating biological processes and immune infiltration. In this study, we identified the m6A regulators for IPF diagnosis and outcomes and explored the biological processes regulated by m6A regulators.

2 Materials and methods

2.1 Collection of clinical information and data sets

The gene expression matrix and clinical characteristics of IPF patients were retrieved from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), and a total of 505 IPF patients and 147 healthy volunteers were included. The bronchoalveolar lavage fluid (BALF) came from the GSE70866 series, including IPF patients ($n = 176$) and healthy volunteers ($n = 20$), which were detected using the GPL14550 and GPL17077 platforms, respectively (Prasse et al., 2019). The lung tissue came from the GSE47460 series, including interstitial lung disease patients ($n = 254$) and healthy volunteers ($n = 108$), which were tested using the GPL6480 and GPL14550 platforms, respectively (Kim et al., 2015; Peng et al., 2016; Anathy et al., 2018). The source of peripheral blood mononuclear cells (PBMCs) was the

GSE28221 series, including IPF patients (n = 75) and healthy volunteers (n = 19), which were detected using the GPL6480 platform (Herazo-Maya et al., 2013; Huang et al., 2015). Due to the different detecting platforms of samples, the batch effect detected by the platforms was eliminated using the ComBat method in the "SVA" R package for data from the same histologic origin. Principal component analysis (PCA) was used to determine the goodness of batch correction.

2.2 m6A identification

Based on previous studies, 36 m6A regulators were considered, including 19 readers, 14 writers, and three erasers. The m6A regulators described above were extracted using the "limma" package in R software. The Wilcoxon test was used to determine whether there were significant changes between IPF patients and healthy volunteers. The "ggpubr" package was employed to draw expression boxplots, and the "pheatmap" package was used to draw a visual m6A expression heatmap.

2.3 Random forest (RF) and support vector machine (SVM) model testing and development of nomogram models

We constructed machine learning classifiers of RF and SVM to predict the reliability of IPF occurrence (Noble, 2006; Altmann et al., 2010). A hyperplane was identified to distinguish IPF patients from healthy volunteers. Then, we established a nomogram to predict the occurrence of IPF in different GEO series based on five selected m6A regulators. Statistical analysis was performed using the R packages "rms" and "rmda." To assess model accuracy and validity, plot calibration curves, decision curve analysis (DCA), and receiver operating characteristic (ROC) curves. Finally, the nomogram formula of the GSE47460 series was brought into GSE70866 and GSE28221 series for verification. ROC curves were drawn to calculate the area under the curve (AUC) and 95% confidence interval (CI).

2.4 Cox proportional hazards regression model, least absolute shrinkage and selection operator (LASSO) regression model

The 31 m6A-related genes of the BALF-derived GSE70866 series were analyzed using univariate Cox regression analysis. Multi-variables faced the risk of over-fitting. LASSO regression was applied to analyze the influence of the survival status of these three variables, and the parameter λ was used to adjust the model complexity to obtain a less variable and more representative combination. Prognostic high-risk and low-risk clusters were obtained according to the risk score using LASSO regression. The 33 m6A-related genes of the PBMC-derived GSE28221 series were analyzed using univariate Cox regression analysis.

The risk score of IPF patients can be used as an independent prognostic factor. Among the clinical data, age, gender, and global alignment and proportion (GAP) score were included as independent variables in univariate and multivariate Cox regression analyses to clarify independent prognostic factors for IPF patients. To balance the statistical data, we used 1000× risk scores. To perform these analyses, we used the "forestplot," "survival," and "glmnet" packages.

2.5 Spearman correlation analysis

Spearman correlation analysis was performed on the lung function using the diffusing capacity of the lungs for carbon monoxide (DLCO) and m6A regulators ($|R| > 0.2$, $p < 0.001$) in the GSE47460 series using the "limma" package. Spearman correlation analysis was performed on *METTL14*, *G3BP2*, and *ZC3H13* with other m6A regulators ($|R| > 0.4$, $p < 0.001$) in the GSE70866 series. The "ggplot2," "ggpubr," and "ggExtra" packages were used for visual scatter plot drawing.

2.6 Consensus cluster analysis, PCA and survival analysis

To determine whether the outcomes-related m6A regulators *METTL14*, *G3BP2*, and *ZC3H13* are associated with IPF, series from GSE70866 were divided into groups according to the consensus levels of m6A regulators using the "ConsensusClusterPlus" package. Outputs included consensus cumulative distribution function (CDF) plots, the relative change in area under the CDF curve, and the consensus matrix. PCA was used to determine the fitness of classification. Kaplan-Meier survival analysis was used to evaluate the difference in overall survival between different m6A types using the "survival" package.

2.7 Differentially expressed genes (DEGs) analysis and Gene Ontology (GO) functional annotation and kyoto encyclopedia of genes and genomes (KEGG) pathway analysis of m6A regulators

The "limma" package and the "VennDiagram" package were used to identify DEGs associated with different m6A
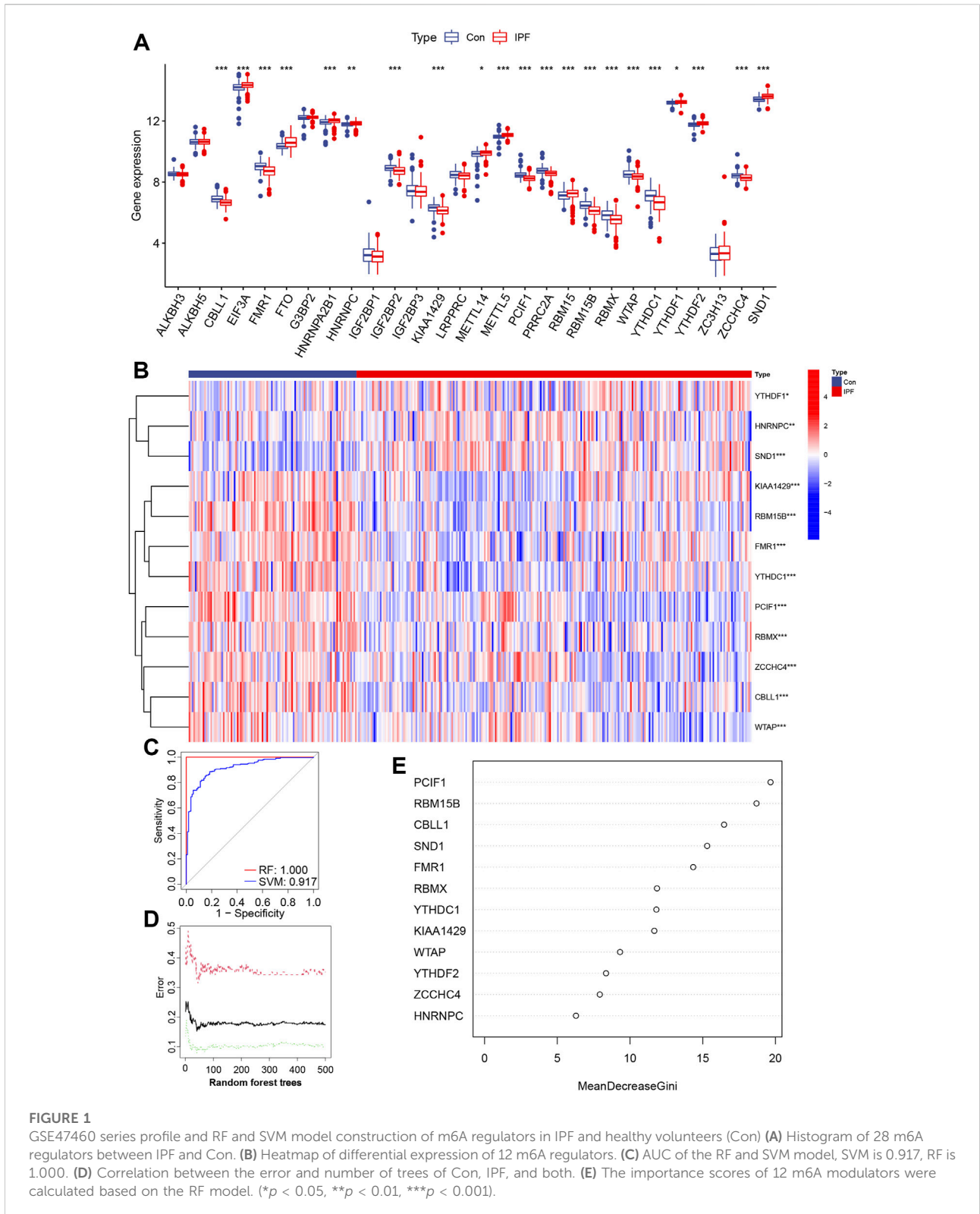
FIGURE 1
GSE47460 series profile and RF and SVM model construction of m6A regulators in IPF and healthy volunteers (Con) **(A)** Histogram of 28 m6A regulators between IPF and Con. **(B)** Heatmap of differential expression of 12 m6A regulators. **(C)** AUC of the RF and SVM model, SVM is 0.917, RF is 1.000. **(D)** Correlation between the error and number of trees of Con, IPF, and both. **(E)** The importance scores of 12 m6A modulators were calculated based on the RF model. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

modifications. Adjusted *p*-values <0.05 and |log2FC| > 1 were considered as DEGs, and volcano plots were drawn for all genes. The "clusterProfiler" package was used to analyze GO functional annotation, including the biological process (BP), celluar component (CC), molecular function (MF) terms, and KEGG pathway analysis. The significance criterion was set at *p*-value <0.05. The results were plotted using a bioinformatics tool (https://www.bioinformatics.com.cn) based on the R language (version 4.03) (Ashburner et al., 2000; Kanehisa and Goto, 2000; Gene Ontology, 2021).

2.8 Immune infiltration assay

Initially, we used single-sample gene set enrichment analysis (ssGSEA) to quantify the relative abundance of 23 immune cell types in the immune microenvironment across different m6A clusters. Panels of specific functional genes are used to mark each immune cell type. The relative abundance of each immune cell type was calculated from the enrichment fraction of gene set expression in transcriptome sequencing in ssGSEA analysis and normalized to a uniform distribution. Next, based on the results of the immune-related analysis, a boxplot of differentially expressed immune cells of different m6A clusters was drawn, and an association heatmap was drawn by using m6A regulators and immune cell expression abundance for correlation analysis.

2.9 Potential target drug analysis of different m6A clusters

The "pRRophetic" package was used to predict potential drugs sensitive to IPF patients with different m6A risk groups. Drug susceptibility was assessed by the half-maximal inhibitory concentration.

3 Results

3.1 Identification of signature m6A regulators genes in the diagnosis of IPF

To investigate the differentially expressed m6A gene between IPF patients and healthy people, according to the detection ratio of the groups, we selected the GSE47460 series derived from lung tissue for differentially expressed m6A gene analysis. After batch correction, the two data sets with the same sample source were merged, and the m6A genes were extracted (Supplementary Figures S1A–H). Due to the small number of healthy volunteers, only differentially expressed m6A genes with opposite expression trends in the GSE70866 and GSE47460 series were excluded. We extracted 21 differentially expressed m6A genes from the
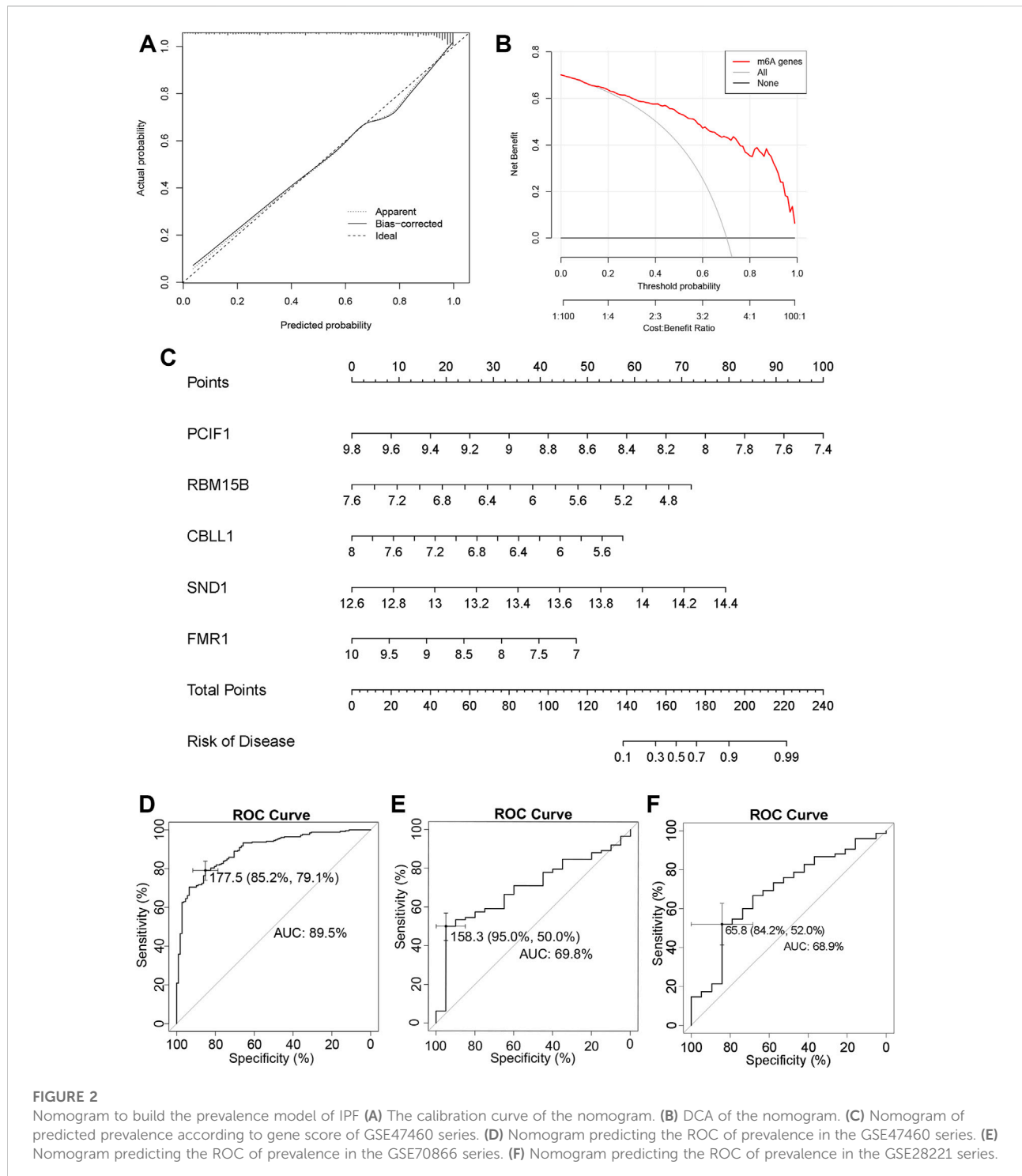
GSE47460 series, eight DEGs with the opposite expression trend, and one undetected m6A gene of GSE47460 was removed (Figure 1A). We obtained 12 reliable and significantly differentially expressed m6A regulators, including six writers and six readers, and presented the heatmap of differentially expressed m6A regulators (Figure 1B).

Modeling of IPF disease-associated m6A regulators was performed using differentially expressed m6A regulators. Simultaneous RF and SVM models were established in the GSE47460 series to identify m6A regulators characterized by IPF. "ROC curve," "residual box plot," and "residual reverse cumulative distribution" suggested that the RF model was more suitable than the SVM model for constructing m6A genes related to pathogenesis (Figure 1C and Supplementary Figures S1I,J). The point with the smallest cross-validation error constructed by RF trees was used as the best model, and the importance scores of 12 m6A regulators were calculated (Figure 1D). The top five m6A regulators (*PCIF1*, *RBM15B*, *CBLL1*, *SND1*, and *FMR1*) were used to build a nomogram to predict prevalence in healthy volunteers and IPF patients (Figure 1E).

The calibration curve can be seen that the "apparent" dotted line corresponding to the entire cohort and the bootstrap corrected solid line are close to the "Ideal" dotted line, indicating that the model has predictive performance (Figure 2A). The DCA method was used to draw a "decision curve" to evaluate the relationship between the nomogram and the gene score to predict the benefits and risks of various cut points in the IPF prevalence model. Figure 2B shows that the selected m6A genes have better benefits at various threshold probabilities.

An m6A gene-associated nomogram of the GSE47460 derived from lung tissue series was then drawn (Figure 2C). The ROC curve indicated the sensitivity and specificity of the diagnosis. The AUC of the SE47460 series was 89.5% (95% CI: 86.19%–92.89%), suggesting a high prediction accuracy (Figure 2D). Then we brought the five m6A regulators into the GSE70866 and GSE28221 to establish the diagnosis-related nomogram construction. The AUC value of the ROC curve was 78.5% (Supplementary Figures S1K,L) and 80.8% (Supplementary Figures S1M,N).

To validate the nomogram constructed based on the lung tissue-derived GSE47460 series (Figure 2C), we used 196 patients from the BALF-derived GSE70866 series and 94 patients from the PBMC-derived GSE28221 series to perform IPF incidence prediction validation and draw ROC curves. The AUCs for the GSE70866 and GSE28221 series were 69.8% and 68.9%, respectively (Figures 2E,F). The source of BALF is lung tissue, and the predictive accuracy was better. There were few samples in the GSE70866 and GSE28221 series; a larger sample size is needed. These findings suggest that *PCIF1*, *RBM15B*, *CBLL1*, *SND1*, and *FMR1* can be used as the characteristic m6A regulators of IPF, and they have diagnostic significance in lung tissue and PBMC.

FIGURE 2
Nomogram to build the prevalence model of IPF **(A)** The calibration curve of the nomogram. **(B)** DCA of the nomogram. **(C)** Nomogram of predicted prevalence according to gene score of GSE47460 series. **(D)** Nomogram predicting the ROC of prevalence in the GSE47460 series. **(E)** Nomogram predicting the ROC of prevalence in the GSE70866 series. **(F)** Nomogram predicting the ROC of prevalence in the GSE28221 series.

3.2 Identification of signature genes for m6A regulators on outcomes in IPF

To clarify the impact of m6A-related genes on the outcomes of IPF, the clinical information provided by the BALF-derived GSE70866 series was used to perform univariate Cox regression analysis on the 31 m6A regulators extracted from the data set (Figure 3A). Three protective m6A regulators ($p < 0.05$) (*METTL14*, *G3BP2*, and *ZC3H13*) were identified. LASSO regression was employed to identify these regulators, and risk scores were calculated (Figures 3B,C). We used univariate Cox regression models to compare the effects of age, gender, GAP,
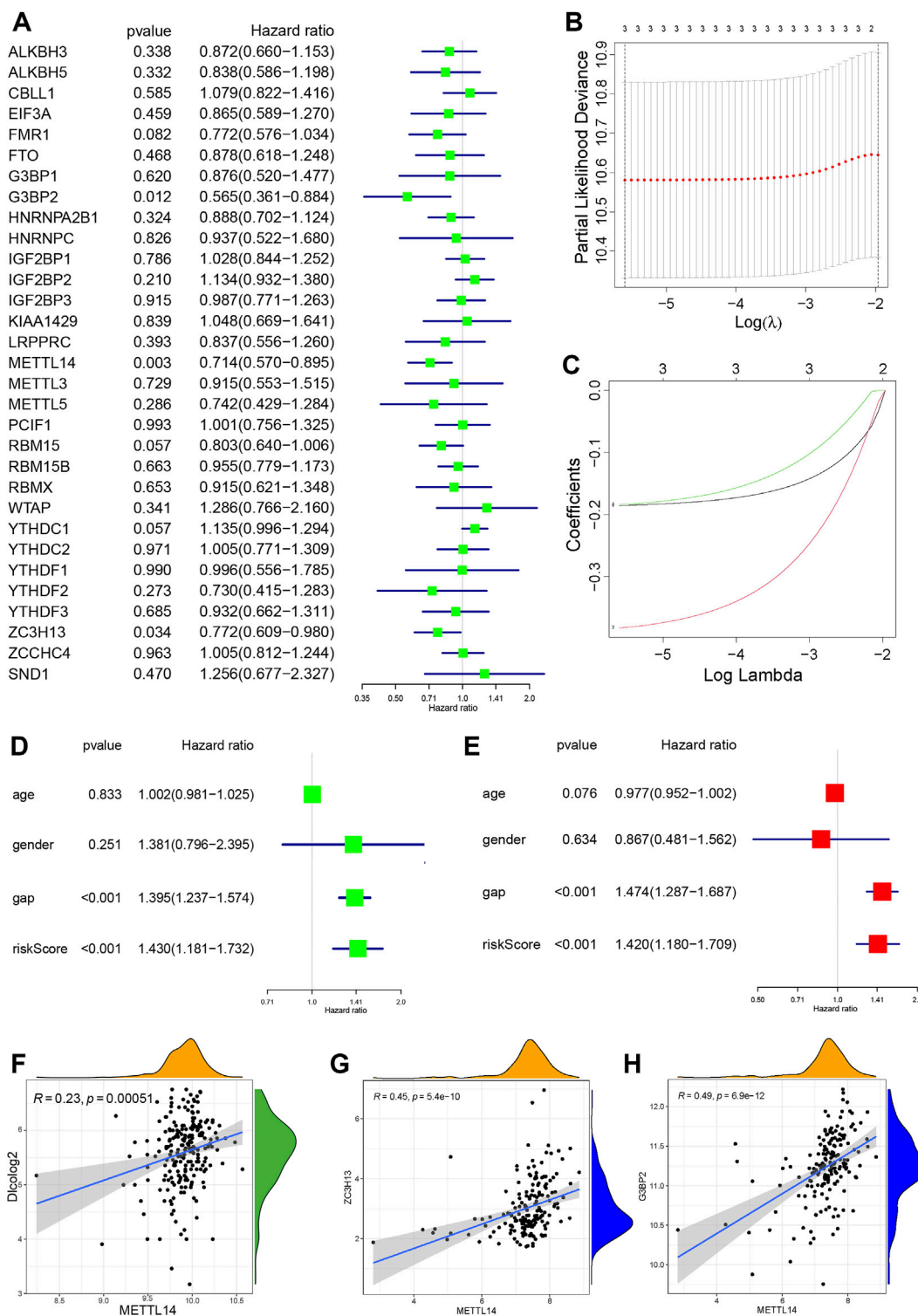
FIGURE 3
Identification of characteristic genes related to IPF outcomes **(A)** Forest plot of three m6A regulators with $p < 0.05$ by univariate Cox regression. **(B)** Selection of the optimal parameter (lambda) in the LASSO regression model. **(C)** LASSO coefficient profiles of the three m6A regulators in the GSE70866 series. **(D)** Forest plot of m6A risk score and three clinical factors by univariate Cox regression. **(E)** Forest plot of m6A risk score and three clinical factors by multivariate Cox regression. **(F)** Correlation between *METTL14* and DLCO ($|R| > 0.2$ and $p < 0.001$). **(G)** Correlation analysis of *METTL14* with *ZC3H13*. **(H)** Correlation analysis of *METTL14* with *G3BP2* ($|R| > 0.4$ and $p < 0.001$).

and m6A regulator risk score on patient survival. Figure 3D shows that GAP (HR 1.395, 95% CI: 1.237–1.574, $p < 0.001$) and the m6A regulator risk score (HR 1.430, 95% CI: 1.181–1.732, $p < 0.001$) predict IPF prognosis.

Multivariate Cox regression analysis was performed to control for the influence of confounding factors and to analyze the influencing factors with independent effects. Figure 3E shows that the GAP score (HR 1.474, 95% CI: 1.287–1.687, $p < 0.001$) and m6A regulator risk score (HR 1.420, 95% CI: 1.180–1.709, $p < 0.001$) remained IPF predictors. The BALF-derived GSE28221 series was used to perform univariate Cox regression analysis on the 33 m6A regulators extracted from the data set (Supplementary Figures S2A). YTHDC1 was a protective m6A regulator ($p = 0.048$). Considering the different sources of PBMC and BALF, PBMC may be affected by several organs and tissues through blood circulation, and the sample size of GSE28221 is small. The evidence for identifying IPF outcomes related m6A regulators in PBMC is insufficient, and more clinical data support is needed.

To identify the optimal IPF prognosis-related m6A modulators, we used lung function data from the GSE47460 series. Lower DLCO indicates worse lung function and is associated with poor outcomes. Spearman correlation analysis was used to analyze the correlation between DLCO and m6A regulators ($|R| > 0.2$, $p < 0.01$), and we found six genes in the GSE47460 series that were weakly correlated with DLCO. KIAA1429, FMR1, PCIF1, YTHDC1, and METTL14 were positively correlated with DLCO and considered potential protective factors for prognosis. YTHCF1 negatively correlated with DLCO and was a potential risk factor for poor prognosis (Supplementary Figures S3A–E). The expression of METTL14 was positively correlated with DLCO ($R = 0.23$) (Figure 3F) and was identified as a critical variable for outcomes in the GSE70866 and GSE47460 series. The expression of YTHDC1 was positively correlated with DLCO, consistent with the PMBC-derived GSE28221 series.

We determined whether METTL14, G3BP2, and ZC3H13 levels correlate with other m6A genes. METTL14, G3BP2, and ZC3H13 are protective factors for prognosis, and METTL14 is positive correlations with ZC3H13 and G3BP2 (Figures 3G,H). Besides, five m6A genes were moderately correlated with the METTL14 expression level, including three readers (IGF2BP3, LRPPRC, and YTHDC2) and the two writers (KIAA1429 and WTAP) (Supplementary Figures S3F–J). G3BP2 found correlated with LRPPRC, YTHDC2, YTHDF2, YTHDF3, KIAA1429, and WTAP by Spearman correlation analysis ($|R| > 0.4$, $p < 0.001$) (Supplementary Figures S3K–P). In summary, for the three outcomes-related m6A regulators, we believe that METTL14 has the strongest regulation on outcomes, including the regulation of lung function, and positively correlates with m6A writers and readers to promote RNA methylation, followed by G3BP2, which is also positively correlated to m6A writers and readers.

3.3 Two m6A patterns identified by prognosis-related m6A regulators

We applied the "consensus clustering method" to explore the impact of different m6A patterns on prognosis based on three m6A prognostic regulators. Compared with clusters 1 to 9, the growth rate in cluster two was flat in the CDF plot (Supplementary Figure S4A). Kaplan-Meier survival analysis of IPF patients with two different m6A clusters showed that the prognosis of clusters A and B were not significant ($p = 0.123$) (Supplementary Figure S4B). We believe that this m6A model does not correlate well with IPF outcomes. According to these conclusions, we selected METTL14 and G3BP2 with strong prognostic correlations as variables to distinguish m6A patterns. Compared with clusters 1 to 9, the growth rate in cluster three was flat in the CDF plot (Figure 4A). Figure 4B shows that the relative change in area under the CDF curve increased insignificantly for cluster 3. In the consistency matrix of cluster 3, the intra-group correlation was higher, and the inter-group correlation was low (Figure 4C and Supplementary Figures S4D–J). In summary, we identified three m6A patterns. Kaplan-Meier survival analysis of IPF patients showed significant differences in outcomes among the three clusters ($p = 0.005$) (Figure 4D). Clusters B (89 cases) and C (17 cases) had poor prognosis; therefore, we believe that cluster A (70 cases) is the m6A low-risk group (containing 70 cases); clusters B and C are the m6A high-risk group (containing 106 cases) for Kaplan-Meier survival analysis. The low-risk group had significantly better outcomes than the high-risk group ($p = 0.008$) (Figure 4E). The PCA scatter diagram was obtained for visualization, which distinguished between m6A low-risk and high-risk (Figure 4F). Boxplots and heatmaps of m6A-related genes were drawn based on different m6A risk groups to demonstrate the differential expression levels of 31 m6A regulators (Figures 4G,H); 18 m6A regulators were significantly differentially expressed.

3.4 Identification of DEGs and enrichment of molecular mechanisms and biological functions in different m6A risk groups

A clinical heatmap was drawn for the five prognostic m6A genes, gender, age (older than 65 years old), survival status, and GAP score data of patients in the m6A high- and low-risk groups (Figure 5A). The chi-square test was performed on the clinical information, and there was no significant difference in gender and age among patients between risk groups, while the survival status ($p < 0.001$) and GAP scores ($p < 0.05$) were significantly different. Using the "limma" package to analyze the DEGs of the m6A high-risk versus low-risk group, 195 DEGs were obtained (Supplementary Table S1), including 27 downregulated genes and 168 upregulated genes. Supplementary Figure S4K is a volcano plot of various m6A clusters, and Figure 5B is an enhanced volcano plot of various m6A risk groups, showing
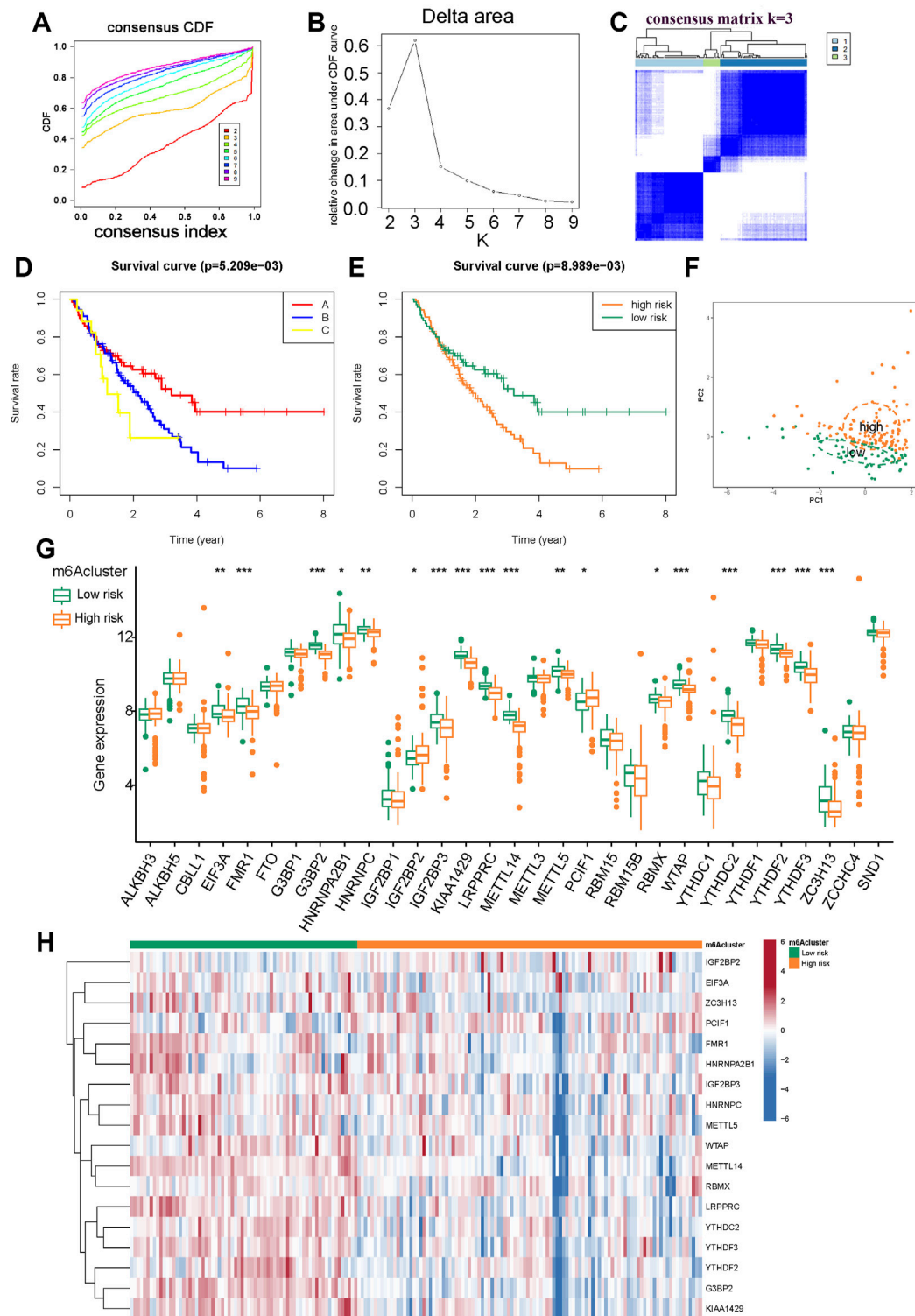
FIGURE 4
Consensus clustering of prognosis-related m6A modulators in IPF. **(A)** The CDF plot. **(B)** The relative change in area under the CDF curve. **(C)** Consensus matrixes of the two significant m6A regulators (*METTL14* and *G3BP2*) for k = 3. **(D)** Kaplan-Meier plot of overall survival in different m6A clusters. **(E)** Kaplan-Meier plot of overall survival in m6A low-risk (cluster A) and m6A high-risk (cluster B and C). **(F)** PCA for m6A low-risk and high-risk. **(G)** Histogram of 31 m6A regulators in m6A low-risk and m6A high-risk. **(H)** Expression heatmap of 18 differentially expressed m6A regulators in m6A low-risk and high-risk.
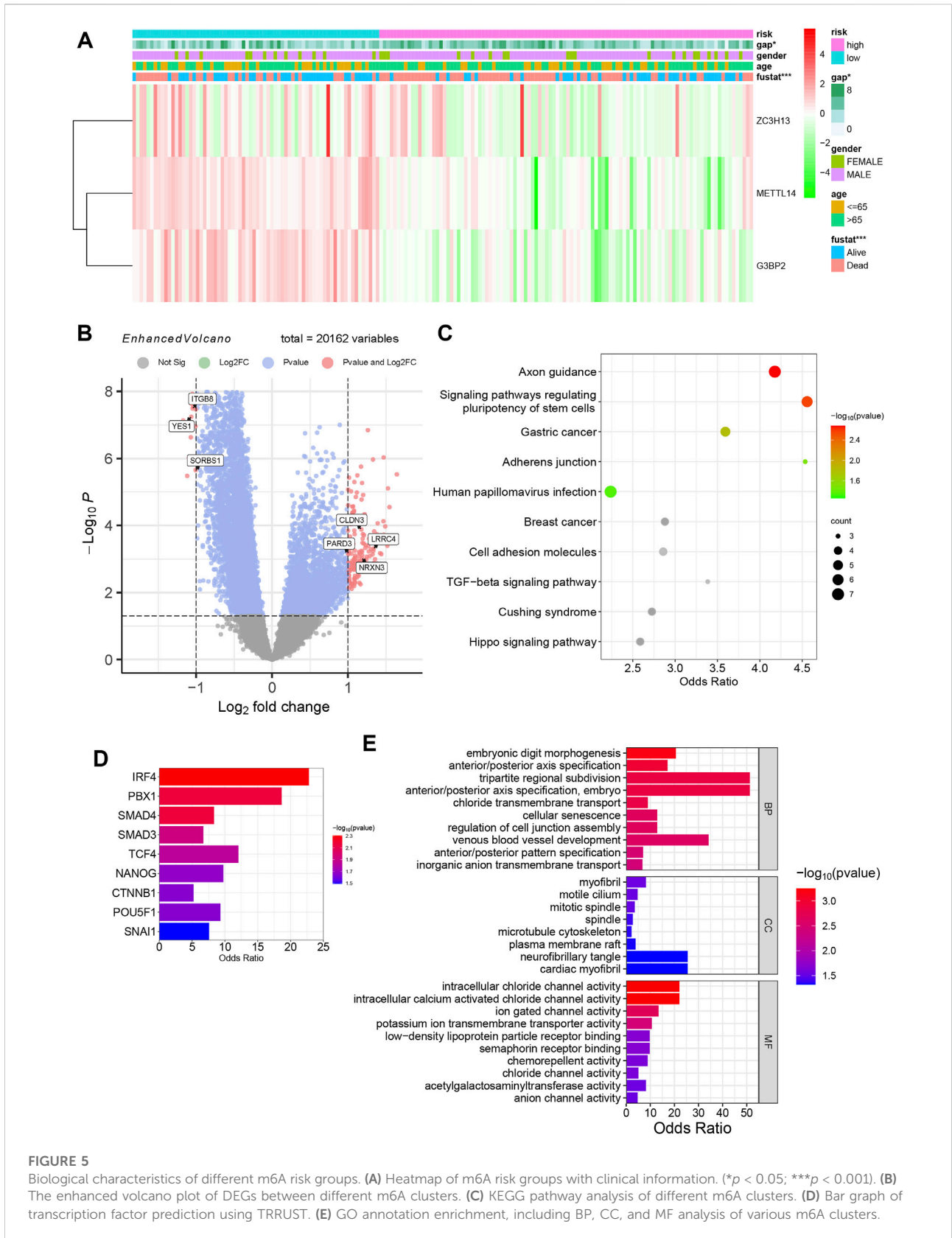
FIGURE 5
Biological characteristics of different m6A risk groups. **(A)** Heatmap of m6A risk groups with clinical information. (*$p < 0.05$; ***$p < 0.001$). **(B)** The enhanced volcano plot of DEGs between different m6A clusters. **(C)** KEGG pathway analysis of different m6A clusters. **(D)** Bar graph of transcription factor prediction using TRRUST. **(E)** GO annotation enrichment, including BP, CC, and MF analysis of various m6A clusters.
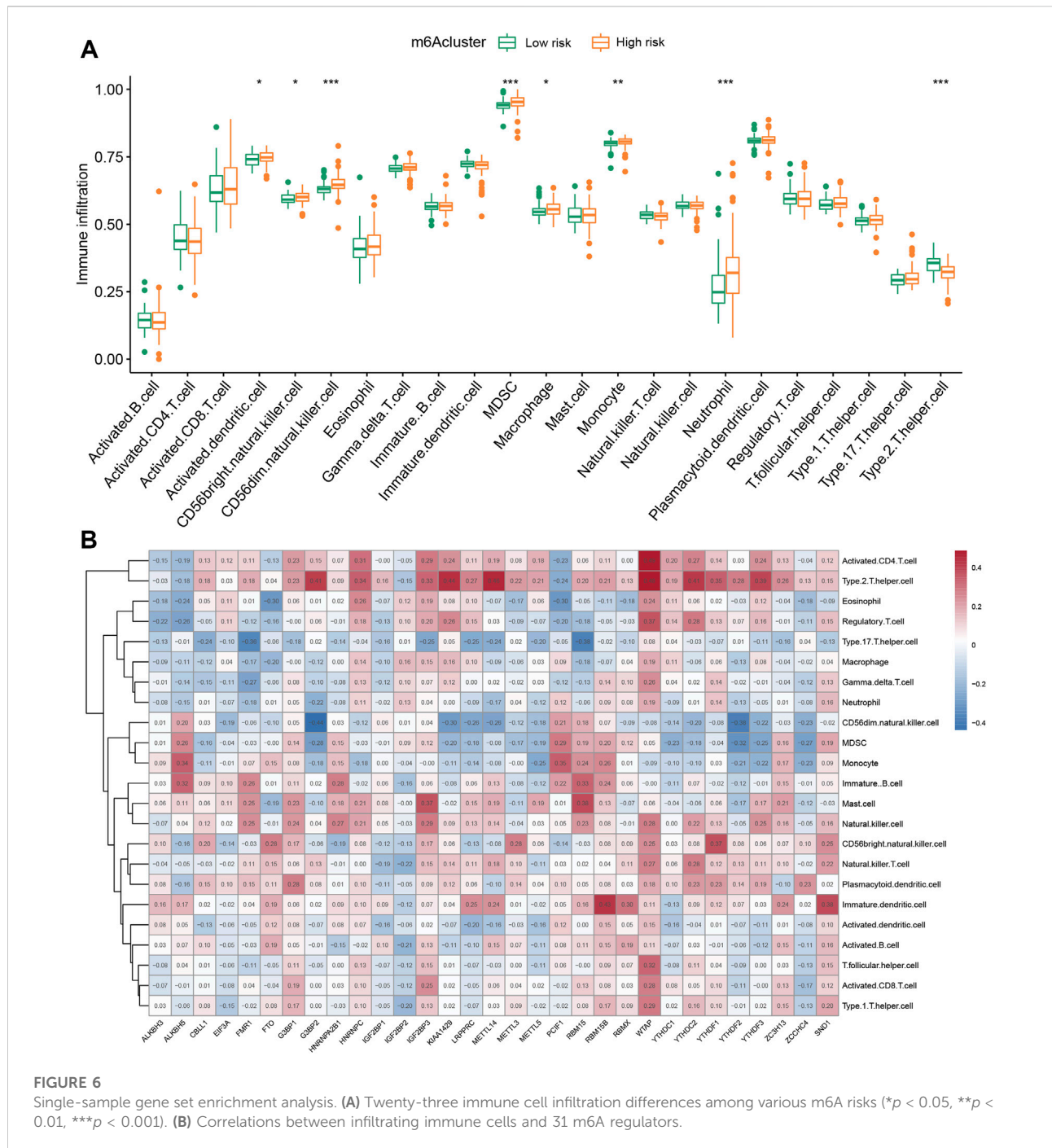
FIGURE 6
Single-sample gene set enrichment analysis. **(A)** Twenty-three immune cell infiltration differences among various m6A risks (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). **(B)** Correlations between infiltrating immune cells and 31 m6A regulators.

the primary DEGs and marking the genes related to cell adhesion. The KEGG pathway analysis enriched adherens junction and cell adhesion molecules (Figure 5C). *YES1*, *ITGB8*, *SORBS1*, *CLDN3*, *LRRC4*, *NRXN3*, and *PARD3* were DEGs related to adherens junction and cell adhesion molecules. Three downregulated genes and four upregulated genes in the high-risk group are shown in Figure 5B. Functions related to cell adhesion were observed in the GO annotation. Regulation of cell junction

assembly was the affected BP (Figure 5E). These findings suggest that cell adhesion and junction functions might be regulated by m6A regulators, mediating IPF progression. Using TRRUST prediction, SMAD4 and SMAD3 were the main transcription factors of the TGF-β/SMAD pathway and are considered the essential pathway for IPF progression (Figure 5D). Myofibril and microtubule cytoskeleton were the affected CCs associated with fibroblasts.
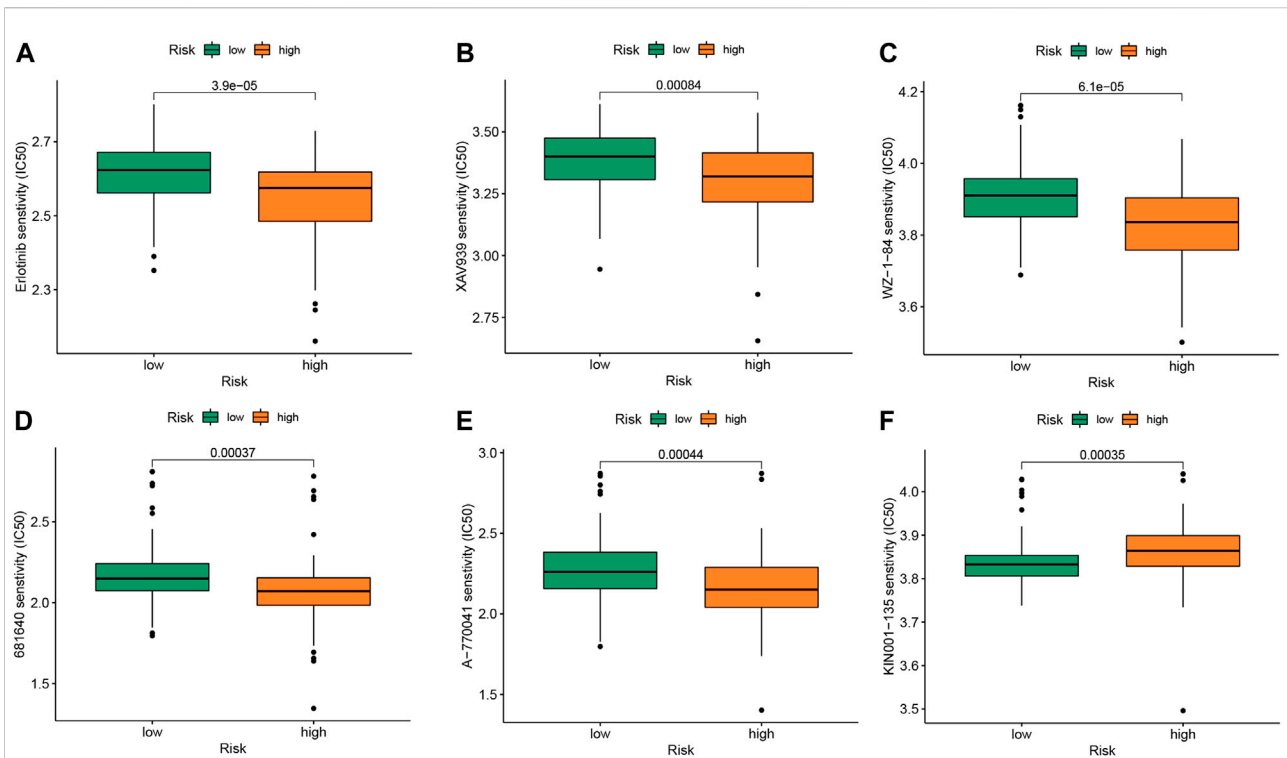
FIGURE 7
Drug sensitivity analysis of IPF patients in different m6A risk group. **(A)** Erlotinib **(B)** XAV939 **(C)** WZ-1-84 **(D)** 681640 **(E)** A-770041 **(F)** KIN001-135. The comparison of the drug sensitivity between m6A low- and high-risk groups. Lower IC_{50} indicates more sensitivity.
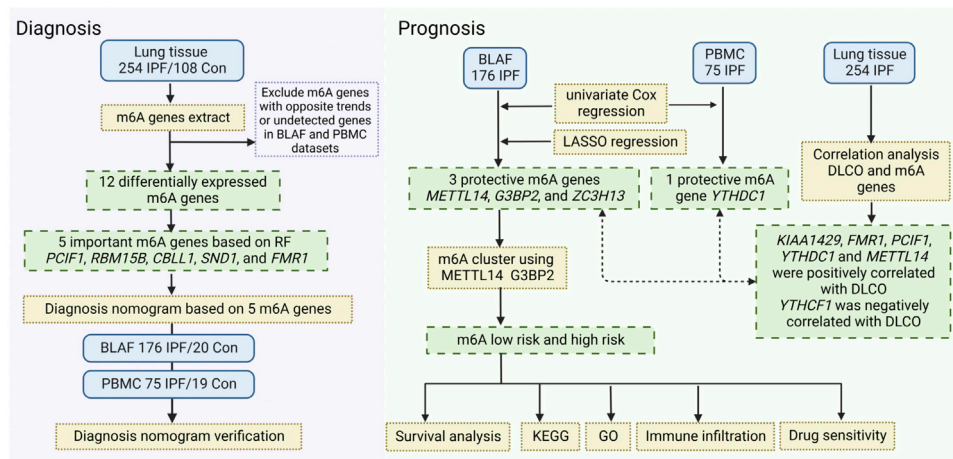

FIGURE 8
The flow chart of the contents of the project.

3.5 Immune infiltration analysis of different m6A patterns

To explore the influence of immune cells on IPF patients with different m6A types, we applied ssGSEA to visualize the abundance of immune cells in IPF samples. We found that neutrophils in the high-risk group were significantly upregulated compared with the low-risk group. Clinical study also showed that the upregulation of neutrophils signals poor IPF outcomes (Jegal, 2022). The high-risk group was markedly enriched in activated CD56dim natural killer

cells, myeloid-derived suppressor cells, and type 2 T helper cells (Figure 6A). We applied Spearman correlation analysis to assess the relationships between m6A regulators and immune cell infiltration (Figure 6B). We evaluated the correlation between three m6A regulators regulating immune cells and found that *G3BP2* and *METTL14* were m6A regulators with negative correlations with neutrophil expression. We also explored differential immune cell infiltration between patients with high and low *G3BP2* or *METTL14* expression (Supplementary Figures S5A,B). *WTAP* is an m6A methyltransferase, and various immune cells positively correlate with *WTAP*. *WTAP* upregulates T cells, including activated CD4 T cell, regulatory T cell, T follicular helper cell, and type 2 T helper cell (Supplementary Figures S5C).

3.6 Potential drug targets for m6A clusters

The "pRRophetic" package is a computational model that predicts chemotherapy responses based on tumor gene expression data (Altmann et al., 2010). IPF drugs are associated with the development of anti-tumor drugs such as nintedanib, an oral small molecule tyrosine kinase inhibitor initially developed for lung cancer and approved for the treatment of IPF. Pirfenidone has a sensitizing effect on the combination of paclitaxel and carboplatin used in clinical practice (Branco et al., 2022). We believe that the identification of related anti-tumor drugs has potential significance for IPF treatment. We obtained six drugs potentially associated with m6A based on a two-sided *p*-value <0.01. Drugs sensitive to the m6A high-risk group included erlotinib, XAV939, WZ-1-84, 681640, and A-770041 (Figures 7A–E). KIN001-135 was more sensitive to the m6A low-risk group (Figures 7F). The m6A low-risk group had better prognosis, while the m6A high-risk group was associated with more potentially sensitive drugs.

4 Discussion

We explored the significance of m6A regulators in IPF diagnosis and prognosis. First, we explored the models by which different m6A modification patterns are associated with IPF diagnosis and identified five diagnostic m6A regulators from a lung tissue-derived dataset, including *PCIF1*, *RBM15B*, *CBLL1*, *SND1*, and *FMR1*. The diagnostic accuracy was subsequently validated on BALF and PBMC-derived datasets. We believe that the five diagnostical m6A regulators have diagnostic significance, and it is possible to collect PBMCs from whole blood for diagnosis; nevertheless, more clinical samples are needed for verification.

We then explored the m6A regulators associated with IPF outcomes and identified *METTL14*, *G3BP2*, and *ZC3H13* as m6A-associated prognostic protective factors. The risk score

calculated by the LASSO regression model and the GAP score were identified as independent m6A prognostic factors. Next, to predict the regulatory patterns of prognosis-related m6A, we performed consistent clustering using *METTL14* and *G3BP2* and identified three subgroups with significantly different prognosis, among which subgroups B and C had poor outcomes; therefore, we considered subgroup A as the m6A low-risk group, and subgroups B and C as the m6A high-risk groups.

Fibrogenesis aggravates pulmonary fibrosis, and TGF-β/ SMAD is a critical pathway for fibrosis (Chanda et al., 2019; Kadota et al., 2021). SMAD3 and SMAD4 were predicted to be significant differential transcription factors across risk groups. Enhanced tight junction function can maintain the integrity and plasticity of alveolar epithelial cells (Varma et al., 2014). Tight junction and adherens junction proteins are upregulated in regenerated alveolar epithelial cells after pulmonary fibrosis, suggesting that enhanced cell adhesion function is beneficial to the prognosis of fibrosis (Lappi-Blanco et al., 2013). Tight junctions provide a physical barrier to epithelial cells and regulate the flow between cells (Zou et al., 2020). METTL14 is an m6A methyltransferase complex that stabilizes the structure and recognizes target RNA (Liu et al., 2018). We speculate that the upregulation of cell adhesion-related proteins in the low-risk group is related to the stabilization of cell adhesion-related RNA by METTL14.

We then analyzed the expression of immune cells in different m6A risk groups and found that neutrophils were significantly upregulated in the high-risk group. Studies linked neutrophils to pulmonary fibrosis disease. Clinical studies showed that blood neutrophil levels positively correlated with IPF progression (Achaiah et al., 2021; Nathan et al., 2021), and the ratio of neutrophils in the blood of IPF patients was inversely proportional to forced vital capacity and forced expiratory volume in one second (Nathan et al., 2021; D'Alessandro et al., 2022). A cohort study of 156 IPF patients identified a high proportion of neutrophils in BALF as an independent predictor of early death (Jegal, 2022). Therefore, the upregulation of neutrophils can be considered a reliable prognostic risk factor in IPF patients, associated with shortened overall survival (Chen et al., 2022). A flowchart outlines the contents of the project (Figure 8).

5 Conclusion

We assessed m6A modification patterns in 505 IPF patients and 147 healthy volunteers based on m6A modulators and identified five diagnostic m6A modulators. Based on clinical data, we identified prognosis-related m6A regulators in IPF. *METTL14* and *G3BP2* were the critical m6A regulators for prognosis that divided IPF patients into m6A types with different risks and revealed the biological mechanisms behind different m6A modification patterns. m6A modification may play a role in stabilizing tight

junction-associated mRNA. There was upregulation of neutrophil expression in the m6A high-risk group, which indicated poor outcomes. Finally, we tested the drug sensitivity of various risk types. These findings provide a basis for the study of IPF-related m6A regulation.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70866; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47460; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28221.

Author contributions

JZhang: data collection and analysis, investigation, visualization, and writing—original draft. YZ, ZW, JZhao, ZL, KW, LT, and BY: review and technical support. QW, TW, and JW: review, technical support, and funding acquisition. QW, TW, and JW: conceptualization, writing—review and editing, supervision, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1102422/full#supplementary-material

References

Achaiah, A., Rathnapala, A., Pereira, A., Bothwell, H., Dwivedi, K., Barker, R., et al. (2021). Monocyte and neutrophil levels are potentially linked to progression to IPF for patients with indeterminate UIP CT pattern. *BMJ Open Respir. Res.* 8 (1), e000899. Epub 2021/11/21PubMed PMID: 34799353; PubMed Central PMCID: PMCPMC8606778. doi:10.1136/bmjresp-2021-000899

Altmann, A., Tolosi, L., Sander, O., and Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics* 26 (10), 1340–1347. Epub 2010/04/14PubMed PMID: 20385727. doi:10.1093/bioinformatics/btq134

Anathy, V., Lahue, K. G., Chapman, D. G., Chia, S. B., Casey, D. T., Aboushousha, R., et al. (2018). Reducing protein oxidation reverses lung fibrosis. *Nat. Med.* 24 (8), 1128–1135. Epub 2018/07/11PubMed PMID: 29988126; PubMed Central PMCID: PMCPMC6204256. doi:10.1038/s41591-018-0090-y

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. The gene Ontology consortium. *Nat. Genet.* 25 (1), 25–29. Epub 2000/05/10PubMed PMID: 10802651; PubMed Central PMCID: PMCPMC3037419. doi:10.1038/75556

Bokar, J. A., Rath-Shambaugh, M. E., Ludwiczak, R., Narayan, P., and Rottman, F. (1994). Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *J. Biol. Chem.* 269 (26), 17697–17704. Epub 1994/07/01. PubMed PMID: 8021282. doi:10.1016/s0021-9258(17)32497-3

Branco, H., Oliveira, J., Antunes, C., Santos, L. L., Vasconcelos, M. H., and Xavier, C. P. R. (2022). Pirfenidone sensitizes NCI-H460 non-small cell lung cancer cells to paclitaxel and to a combination of paclitaxel with carboplatin. *Int. J. Mol. Sci.* 23 (7), 3631. Epub 2022/04/13PubMed PMID: 35408988; PubMed Central PMCID: PMCPMC8998757. doi:10.3390/ijms23073631

Chanda, D., Otoupalova, E., Smith, S. R., Volckaert, T., De Langhe, S. P., and Thannickal, V. J. (2019). Developmental pathways in the pathogenesis of lung fibrosis. *Mol. Asp. Med.* 65, 56–69. Epub 2018/08/22PubMed PMID: 30130563; PubMed Central PMCID: PMCPMC6374163. doi:10.1016/j.mam.2018.08.004

Chen, Y., Cai, J., Zhang, M., and Yan, X. (2022). Prognostic role of NLR, PLR and MHR in patients with idiopathic pulmonary fibrosis. *Front. Immunol.* 13, 882217. Epub 2022/05/17PubMed PMID: 35572564; PubMed Central PMCID: PMCPMC9096781. doi:10.3389/fimmu.2022.882217

Clynick, B., Corte, T. J., Jo, H. E., Stewart, I., Glaspole, I. N., Grainge, C., et al. (2022). Circulating RNA differences between patients with stable and progressive idiopathic pulmonary fibrosis. *Eur. Respir. J.* 59 (3), 1902058. Epub 2021/10/23PubMed PMID: 34675050. doi:10.1183/13993003.02058-2019

D'Alessandro, M., Bergantini, L., Carleo, A., Cameli, P., Perrone, A., Fossi, A., et al. (2022). Neutrophil-to-lymphocyte ratio in bronchoalveolar lavage from IPF patients: A novel prognostic biomarker? *Minerva Med.* 113 (3), 526–531. Epub 2020/05/15PubMed PMID: 32407050. doi:10.23736/S0026-4806.20.06614-8

Drakopanagiotakis, F., Wujak, L., Wygrecka, M., and Markart, P. (2018). Biomarkers in idiopathic pulmonary fibrosis. *Matrix Biol.* 68-69, 404–421. Epub 2018/02/07PubMed PMID: 29408012. doi:10.1016/j.matbio.2018.01.023

Gao, R., Ye, M., Liu, B., Wei, M., Ma, D., and Dong, K. (2021). m6A modification: A double-edged sword in tumor development. *Front. Oncol.* 11, 679367. Epub 2021/08/13PubMed PMID: 34381710; PubMed Central PMCID: PMCPMC8350482. doi:10.3389/fonc.2021.679367

Gene Ontology, C. (2021). The gene ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. Epub 2020/12/09PubMed PMID: 33290552; PubMed Central PMCID: PMCPMC7779012. doi:10.1093/nar/gkaa1113

Herazo-Maya, J. D., Noth, I., Duncan, S. R., Kim, S., Ma, S. F., Tseng, G. C., et al. (2013). Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci. Transl. Med.* 5, 205ra136. 205ra136. Epub 2013/10/04PubMed PMID: 24089408; PubMed Central PMCID: PMCPMC4175518. doi:10.1126/scitranslmed.3005964

Hochhegger, B., Marchiori, E., Zanon, M., Rubin, A. S., Fragomeni, R., Altmayer, S., et al. (2019). Imaging in idiopathic pulmonary fibrosis: Diagnosis and mimics. *Clin. (Sao Paulo)* 74, e225. Epub 2019/02/07PubMed PMID: 30726312; PubMed Central PMCID: PMCPMC6384526. doi:10.6061/clinics/2019/e225

Huang, Y., Ma, S. F., Vij, R., Oldham, J. M., Herazo-Maya, J., Broderick, S. M., et al. (2015). A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. *BMC Pulm. Med.* 15, 147. Epub 2015/11/22PubMed PMID: 26589497; PubMed Central PMCID: PMCPMC4654815. doi:10.1186/s12890-015-0142-8

Janowiak, P., Szymanowska-Narloch, A., and Sieminska, A. (2022). IPF respiratory symptoms management - current evidence. *Front. Med. (Lausanne)* 9, 917973. Epub 2022/08/16PubMed PMID: 35966835; PubMed Central PMCID: PMCPMC9368785. doi:10.3389/fmed.2022.917973

Jegal, Y. (2022). The role of neutrophils in the pathogenesis of IPF. *Korean J. Intern Med.* 37 (5), 945–946. Epub 2022/09/08PubMed PMID: 36068714; PubMed Central PMCID: PMCPMC9449193. doi:10.3904/kjim.2022.267

Jiang, X., Liu, B., Nie, Z., Duan, L., Xiong, Q., Jin, Z., et al. (2021). The role of m6A modification in the biological functions and diseases. *Signal Transduct. Target. Ther.* 6 (1), 74. Epub 2021/02/22PubMed PMID: 33611339; PubMed Central PMCID: PMCPMC7897327. doi:10.1038/s41392-020-00450-x

Kadota, T., Fujita, Y., Araya, J., Watanabe, N., Fujimoto, S., Kawamoto, H., et al. (2021). Human bronchial epithelial cell-derived extracellular vesicle therapy for pulmonary fibrosis via inhibition of TGF-beta-WNT crosstalk. *J. Extracell. Vesicles* 10 (10), e12124. Epub 2021/08/12PubMed PMID: 34377373; PubMed Central PMCID: PMCPMC8329991. doi:10.1002/jev2.12124

Kanehisa, M., and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. Epub 1999/12/11PubMed PMID: 10592173; PubMed Central PMCID: PMCPMC102409. doi:10.1093/nar/28.1.27

Kim, S., Herazo-Maya, J. D., Kang, D. D., Juan-Guardela, B. M., Tedrow, J., Martinez, F. J., et al. (2015). Integrative phenotyping framework (iPF): Integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* 16, 924. Epub 2015/11/13PubMed PMID: 26560100; PubMed Central PMCID: PMCPMC4642618. doi:10.1186/s12864-015-2170-4

Lan, Q., Liu, P. Y., Haase, J., Bell, J. L., Huttelmaier, S., and Liu, T. (2019). The critical role of RNA m(6)A methylation in cancer. *Cancer Res.* 79 (7), 1285–1292. Epub 2019/03/22PubMed PMID: 30894375. doi:10.1158/0008-5472.CAN-18-2965

Lan, Q., Liu, P. Y., Bell, J. L., Wang, J. Y., Huttelmaier, S., Zhang, X. D., et al. (2021). The emerging roles of RNA m(6)A methylation and demethylation as critical regulators of tumorigenesis, drug sensitivity, and resistance. *Cancer Res.* 81 (13), 3431–3440. Epub 2021/07/07PubMed PMID: 34228629. doi:10.1158/0008-5472.CAN-20-4107

Lappi-Blanco, E., Lehtonen, S. T., Sormunen, R., Merikallio, H. M., Soini, Y., and Kaarteenaho, R. L. (2013). Divergence of tight and adherens junction factors in alveolar epithelium in pulmonary fibrosis. *Hum. Pathol.* 44 (5), 895–907. Epub 2012/12/21PubMed PMID: 23253490. doi:10.1016/j.humpath.2012.08.016

Li, W., Hao, Y., Zhang, X., Xu, S., and Pang, D. (2022). Targeting RNA N(6)-methyladenosine modification: A precise weapon in overcoming tumor immune escape. *Mol. Cancer* 21 (1), 176. Epub 2022/09/08PubMed PMID: 36071523; PubMed Central PMCID: PMCPMC9454167. doi:10.1186/s12943-022-01652-3

Liu, J., Eckert, M. A., Harada, B. T., Liu, S. M., Lu, Z., Yu, K., et al. (2018). m(6 A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nat. Cell Biol.* 20 (9), 1074–1083. Epub 2018/08/30PubMed PMID: 30154548; PubMed Central PMCID: PMCPMC6245953. doi:10.1038/s41556-018-0174-4

Mei, Q., Liu, Z., Zuo, H., Yang, Z., and Qu, J. (2021). Idiopathic pulmonary fibrosis: An update on pathogenesis. *Front. Pharmacol.* 12, 797292. Epub 2022/02/08PubMed PMID: 35126134; PubMed Central PMCID: PMCPMC8807692. doi:10.3389/fphar.2021.797292

Nathan, S. D., Mehta, J., Stauffer, J., Morgenthien, E., Yang, M., Limb, S. L., et al. (2021). Changes in neutrophil-lymphocyte or platelet-lymphocyte ratios and their associations with clinical outcomes in idiopathic pulmonary fibrosis. *J. Clin. Med.* 10 (7), 1427. Epub 2021/05/01PubMed PMID: 33916077; PubMed Central PMCID: PMCPMC8037167. doi:10.3390/jcm10071427

Noble, W. S. (2006). What is a support vector machine? *Nat. Biotechnol.* 24 (12), 1565–1567. Epub 2006/12/13PubMed PMID: 17160063. doi:10.1038/nbt1206-1565

Peng, X., Moore, M., Mathur, A., Zhou, Y., Sun, H., Gan, Y., et al. (2016). Plexin C1 deficiency permits synaptotagmin 7-mediated macrophage migration and enhances mammalian lung fibrosis. *FASEB J.* 30 (12), 4056–4070. Epub 2016/09/10PubMed PMID: 27609773; PubMed Central PMCID: PMCPMC5102121. doi:10.1096/fj.201600373R

Prasse, A., Binder, H., Schupp, J. C., Kayser, G., Bargagli, E., Jaeger, B., et al. (2019). BAL cell gene expression is indicative of outcome and airway basal cell involvement in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 199 (5), 622–630. Epub 2018/08/25PubMed PMID: 30141961; PubMed Central PMCID: PMCPMC6396865. doi:10.1164/rccm.201712-2551OC

Qu, X., Zhang, L., Li, S., Li, T., Zhao, X., Wang, N., et al. (2022). m(6 A-related angiogenic genes to construct prognostic signature, reveal immune and oxidative stress landscape, and screen drugs in hepatocellular carcinoma. *Oxid. Med. Cell Longev.* 2022, 8301888. Epub 2022/10/18PubMed PMID: 36246403; PubMed Central PMCID: PMCPMC9554665. doi:10.1155/2022/8301888

Varma, S., Mahavadi, P., Sasikumar, S., Cushing, L., Hyland, T., Rosser, A. E., et al. (2014). Grainyhead-like 2 (GRHL2) distribution reveals novel pathophysiological differences between human idiopathic pulmonary fibrosis and mouse models of pulmonary fibrosis. *Am. J. Physiol. Lung Cell Mol. Physiol.* 306 (5), L405–L419. Epub 2014/01/01PubMed PMID: 24375798; PubMed Central PMCID: PMCPMC3949053. doi:10.1152/ajplung.00143.2013

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161 (6), 1388–1399. Epub 2015/06/06PubMed PMID: 26046440; PubMed Central PMCID: PMCPMC4825696. doi:10.1016/j.cell.2015.05.014

Wu, B., Su, S., Patil, D. P., Liu, H., Gan, J., Jaffrey, S. R., et al. (2018). Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. *Nat. Commun.* 9 (1), 420. Epub 2018/01/31PubMed PMID: 29379020; PubMed Central PMCID: PMCPMC5789076. doi:10.1038/s41467-017-02770-z

Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., et al. (2016). Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* 61 (4), 507–519. Epub 2016/02/16PubMed PMID: 26876937. doi:10.1016/j.molcel.2016.01.012

Zaccara, S., Ries, R. J., and Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* 20 (10), 608–624. Epub 2019/09/15PubMed PMID: 31520073. doi:10.1038/s41580-019-0168-5

Zhang, Z., Zhang, C., Yang, Z., Zhang, G., Wu, P., Luo, Y., et al. (2021). m(6 A regulators as predictive biomarkers for chemotherapy benefit and potential therapeutic targets for overcoming chemotherapy resistance in small-cell lung cancer. *J. Hematol. Oncol.* 14 (1), 190. Epub 2021/11/12PubMed PMID: 34758859; PubMed Central PMCID: PMCPMC8579518. doi:10.1186/s13045-021-01173-4

Zou, J., Li, Y., Yu, J., Dong, L., Husain, A. N., Shen, L., et al. (2020). Idiopathic pulmonary fibrosis is associated with tight junction protein alterations. *Biochim. Biophys. Acta Biomembr.* 1862 (5), 183205. Epub 2020/02/01PubMed PMID: 32001212. doi:10.1016/j.bbamem.2020.183205