



OPEN

5-methyladenosine regulators play a crucial role in development of chronic hypersensitivity pneumonitis and idiopathic pulmonary fibrosis

Yiyi Zhou, Zhenli Hu, Qinying Sun & Yuchao Dong

5-methyladenosine (m5C) modification regulates gene expression and biological functions in oncologic areas. However, the effect of m5C modification in chronic hypersensitivity pneumonitis (CHP) and idiopathic pulmonary fibrosis (IPF) remains unknown. Expression data for 12 significant m5C regulators were obtained from the interstitial lung disease dataset. Five candidate m5C regulators, namely tet methylcytosine dioxygenase 2, NOP2/Sun RNA methyltransferase 5, Y-box binding protein 1, tRNA aspartic acid methyltransferase 1, and NOP2/Sun RNA methyltransferase 3 were screened using random forest and nomogram models to predict risks of pulmonary fibrosis. Next, we applied the consensus clustering method to stratify the samples with different m5C patterns into two groups (cluster A and B). Finally, we calculated immune cell infiltration scores via single-sample gene set enrichment analysis, then compared immune cell infiltration, related functions as well as the expression of programmed cell death 1 (PD-1, PDCD1) and programmed death protein ligand-1 (PD-L1, CD274) between the two clusters. Principal component analysis of m5C-related scores across the 288 samples revealed that cluster A had higher immune-related expression than B. Notably, T helper cell (Th) 2 type cytokines and Th1 signatures were more abundant in clusters A and B, respectively. Our results suggest that m5C is associated with and plays a crucial role in development of pulmonary fibrosis. These m5C patterns could be potential biomarkers for identification of CHP and IPF, and guide future development of immunotherapy or other new drugs strategies for pulmonary fibrosis.

CHP, one form of hypersensitivity pneumonitis (HP), is an inflammatory and/or fibrotic disease characterized by diverse histopathological and clinical features. Studies have shown that the disease, which is caused by immune-mediated responses to dominant or recessive inhaled antigens in susceptible individuals¹ and causes serious morbidities and mortalities², affects a wide range of humans, including older and younger individuals^{3,4}. Notably, CHP patients present with symptoms similar to those observed in IPF counterparts, including chronic flu-like symptoms, dry cough, mid-inspiratory squeaks, dyspnea, chest tightness, and weight loss^{4,5}. Approximately 11–65% of patients have radiographic findings of pulmonary fibrosis⁶. Pulmonary fibrosis can be easily confused with IPF^{7,8}. Since subtypes of hypersensitivity pneumonitis, namely acute, subacute, and chronic, are still vaguely defined, there is need for clinicians to obtain patients' clinical history, perform high-resolution computed tomography (HRCT), histopathology and multiple discipline diagnosis (MDD)^{9,10}. The histopathological pattern of CHP includes a type called usual interstitial pneumonia (UIP), which exhibits a course similar to IPF. Studies have shown that this subtype not only causes clinically indistinguishable disease, but also delays recognition and avoidance of CHP stimulating antigen¹¹. One study demonstrated that UIP/IPF is over-diagnosed and many of the cases so flagged were actually CHP with a UIP-like pattern¹². The pathogenesis of IPF is repeated subclinical epithelial cell injury superimposed with accelerated epithelial aging, a phenomenon that results in abnormal repair of damaged alveoli and myofibroblast-induced interstitial fibrotic deposition⁵. To date, only a handful of studies have described similarities and differences among molecular mechanisms underlying occurrence and

Department of Respiratory and Critical Care Medicine, Changhai Hospital, Shanghai, China. email: dongyc1020@aliyun.com

development of CHP and IPF. Understanding these mechanisms will greatly improve early screening of high-risk groups, identification of CHP and IPF, as well as development of targeted treatment therapies.

Modified nucleotides were discovered in large numbers of cellular RNAs as early as the 1960s¹³. Consequently, RNA modifications, including 5-methylcytosine (m5C), N1-methyladenosine (m1A), N6-methyladenosine (m6A), and uridylation (U-tail), have been shown to play a crucial role in modifying protein-coding messenger RNAs (mRNAs) and noncoding RNAs (ncRNAs)^{14–17}. One such modification, termed m5C, has been shown to play a crucial role in RNA translation, transcription, and splicing^{18,19}. RNA-modifying proteins (RMPs) that regulate m5C RNA methylation are called “writers”. These mainly include [members of the NOL1/NOP2/SUN domain (NSUNs) and DNA methyltransferase (DNMTs) families]^{20,21}, “readers” [Aly/REF export factor (ALYREF) and Y-box binding protein 1 (YBX1)]²², and “erasers” [the ten-eleven translocator (TETs) family and AlkB homolog 1, histone H2A dioxygenase (ALKBH1)]²⁰. Researchers have gradually focused their efforts on the importance of m5C modification in tumors, with evidence showing that high expression of NSUN2 is associated with poor prognosis in prostate cancer²³. Another study demonstrated that NSUN2-deficient hepatocellular carcinoma cells were less in G1 and S phases, compared to G0 and G2 phases, indicating that inhibition of the NSUN2 gene significantly suppressed proliferation and division of HepG2 cells²⁴. Moreover, DNMT1 was shown to be an epigenetic target for destruction and disruption metastatic and invasive phenotypes of TNBC cells²⁵. To date, however, the expression pattern of m5C regulators in non-neoplastic diseases remains unknown.

In the present study, we hypothesized that m5C regulators regulate the mechanism underlying occurrence and development of CHP and/or IPF. To this end, we evaluated the functions of m5C regulators in diagnosis and classification of CHP and IPF based on the GSE150910 dataset. Our results reveal m5C regulator-mediated RNA methylation modification patterns and immune microenvironment infiltration characterization, and indicate that this genetic factor has potential as an immunotherapeutic agent.

Methodology

Data collection and processing. We used RNA-seq data from the NCBI Gene Expression Synthesis Database (NCBI–GEO, <https://www.ncbi.nlm.nih.gov/geo/>), under accession number GSE150910. The tissues were isolated from surgical lung biopsy or lung transplants, including 103 unaffected controls, 103 and 82 patients IPF and CHP patients, respectively. We initially analyzed a total of 17 m5C regulators, including 11 writers (NOP2, TRDMT1, NSUN2-7, DNMT1, and DNMT3A-B), 2 readers (ALYREF, and YBX1) and 4 erasers (TET1-3 and ALKBH1). Apart from ALYREF, ALKBH1, NOP2, TET1 and TET3, we screened out 12 m5C regulators using differential expression analysis between patients (CHP, IPF) and their corresponding controls.

Construction of random forest (RF) and nomogram models. RF is a combination of tree predictors that mitigates individual bias by combining and weighted regression or classification^{26,27}. We systematically built a machine learning classifier, RF, to predict the occurrence of CHP and IPF, then used it to select candidate regulators from 12 m5C regulators. Furthermore, we examined the importance of the 22 m5C regulators and selected 5 with the highest scores for construction of a nomogram prognostic model for predicting prevalence of CHP and IPF.

Consensus clustering and principal component analyses. To explore the relationship between m5C regulators with CHP and IPF, we stratified the GEO cohort into different subgroups according to the consensus level of m5C regulators. Next, we employed the “ConsensusClusterPlus” package in R to analyze and process the data, then visualized the results using consensus cumulative distribution function (CDF) plots, delta area plots and heatmaps. We also employed “ggplot2” package to generate a PCA plot and visualize the grouped data.

Differential gene expression, and functional and pathway analyses. We employed the “limma” package in R to identify differentially expressed genes (DEGs) associated with m5C modifications, based on $P < 0.05$ and absolute value of \log_2FC (fold-change) $>$ as cutoff. The identified DEGs were subjected to Gene Ontology (GO) functional enrichment, as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis using the “clusterProfiler” package and P value < 0.05 as the threshold.

Analysis of immune cell infiltration and function. We performed single sample gene set enrichment analysis (ssGSEA), which calculates an enrichment score to reveal the absolute enrichment level of a metageneset in a given dataset in each sample and gene set²⁸, to count the abundance of immune cells in subgroups (cluster A and B). Immuno-correlation analysis was also performed based on ssGSEA and results visualized using heatmaps and boxplots. We then used the “estimate” package to perform ESTIMATE algorithm and generate a microenvironment score (immune score and stromal score), which was used to define the degree of immune cell infiltration and stromal cell infiltration, respectively.

Identification of candidate small molecule drugs. To identify potential CHP/IPF therapeutic drugs, we divided DEGs associated with m5C typing into either down-regulated and up-regulated groups, then uploaded them to the Connectivity Map (CMap) database platform (<https://clue.io/query>), a platform for uncovering functional links between small molecule compounds, genes, and disease states^{29,30}. We investigated the degree of similarity (which is represented by score and ranged from -100 to 100) between the uploaded gene list and 2429 sets of small molecule processing expression profile data. Scores closer to 100 represents genes that were more similar to the small molecule treatment record. Conversely, a value closer to -100 denotes that this small molecule inhibits these genes.

Statistical analysis. We employed the “biomaRt” package to filter the whole dataset, as well as delete duplicated and missing, then used the statistical procedure \log_2 (TPM) for data transformation. The correlation between writers and erasers was determined using linear regression, and the Wilcoxon-test applied to detect differences among several groups. Differences between CHP and IPF groups were compared using pairwise prop test. Model construction was achieved using “rms” and “rmda” packages in R. Model accuracy and validation were determined by plotting calibration, and clinical impact curves, as well as decision curve analysis (DCA). All data analysis procedures were performed using packages implemented R software versions 4.2.0 and R 4.2.1.

Results

Landscape of the m5C regulators in CHP and IPF. A summary of the study workflow is presented in Fig. 1, while profiles of differential expression among the 14 m5C regulators, between controls and tests (CHP and IPF), are illustrated in Fig. 2A. We screened out a total of 12 significant m5C regulators (DNMT1, DNMT3A-3B, NSUN2-7, TET2, TRDMT1, YBX1) (Fig. 2B), and found that all of them were significantly down-regulated in experimental tissues, relative to corresponding controls. Profiles of chromosomal positions of the m5C regulators, as detected by the “RCircos” package, are shown in Fig. 2C.

Erasers are correlated with writers. Genes cannot function in isolation, and there is evidence that m6A “writers”, “erasers”, and “readers” cooperate with each other in the context of cancer³¹. Therefore, the genetic expression correlations between m5C regulators were investigated. We used linear regression to explore the relationship among m5C “writers”, “readers” and “erasers”, and found a significant correlation between writers and erasers. Notably, the expression of 11 writers (DNMT1, DNMT3A-3B, NOP2, NSUN2-7, TRDMT1) was positively correlated with that of TET2 (Fig. 3). Correlation results are outlined in Table 1.

Construction of RF and nomogram models. We used the “randomForest” package construct a RF model comprising the top five m5C regulators (TET2, NSUN5, YBX1, TRDMT1 and NSUN3) to predict occurrence of CHP and IPF. Profiles of the M5C regulators, visualized after ranking these genes according to their importance, are presented in Figs. 4A,B. The constructed nomogram is presented in Fig. 4C. Notably, there is a short distance between the dashed and solid lines in calibration curve (Fig. 4D), as well as the clinical impact curve (Fig. 4F), indicating that the nomogram was highly accurate. The following red line in DCA curve staying above the grey and black lines reflected benefit to CHP and IPF (Fig. 4E).

Identification of the two m5C clusters. Next, we employed consensus clustering analysis to distinguish among m5C patterns. As illustrated in Fig. 5A, we divided the tests (IPF and CHP) into 2 clusters (m5C cluster A and cluster B) according to the discrepancy analysis result file gained from Fig. 2B. Discrepancies between analysis result file and clustering results are presented in Supplementary Tables S1, S2, respectively. Notably,

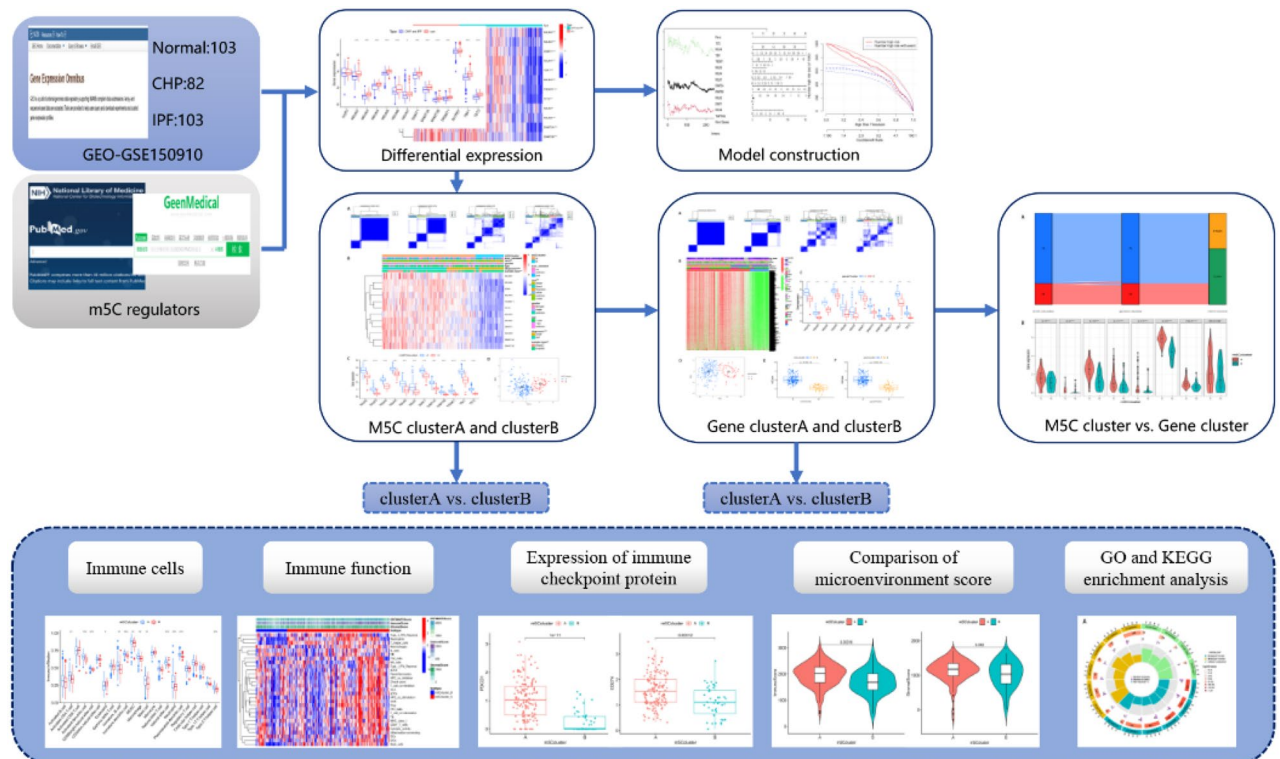


Figure 1. Flowchart of this study.

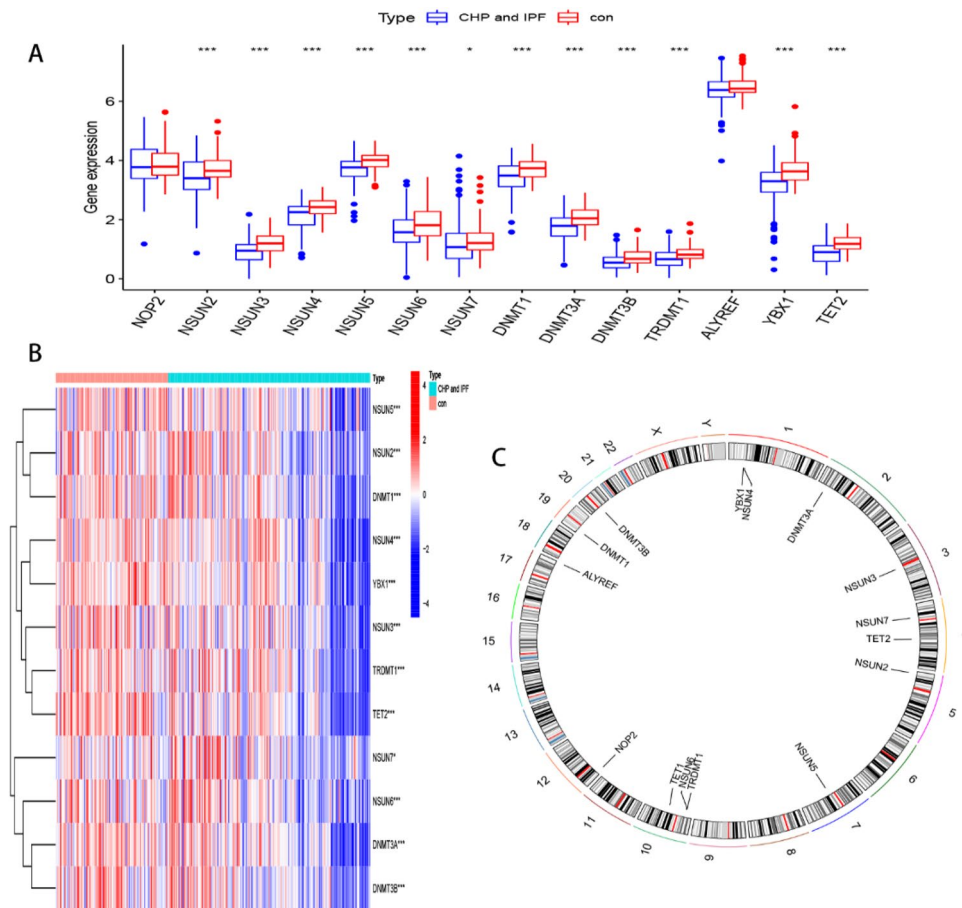


Figure 2. Landscape of the m5C regulators in CHP and IPF. **(A)** The boxplot of 14 m5C regulators expression in tests (CHP and IPF) and controls. **(B)** Expression of 12 significantly differentially expressed m5C regulators in tests and controls. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. **(C)** Chromosomal positions of those m5C regulators.

all the 12 m5C regulators (DNMT1, DNMT3A-3B, NSUN2-7, TET2, TRDMT1 and YBX1) had significantly higher expression levels in cluster A than B (Figs. 5B,C). We further visualized the correlation between clinical characteristics and m5C clusters using a heatmap. To this end, it was evident that patients in cluster A tended to be more diagnosed with IPF, and also had more lung biopsies and diverse races compared to their counterparts in cluster B. These results were verified using PCA (Fig. 5D).

Next, we calculated the abundance of immune cells in clusters A and B and found that cluster A was significantly associated with Th2 cells ($p < 0.001$) and mast cells (MCs) ($p < 0.01$) than B. On the other hand, cluster B was more closely associated with eosinophils ($p < 0.05$) (Fig. 6A). It can be seen from Fig. 6B that all the 12 m5C regulators (DNMT1, DNMT3A-3B, NSUN2-7, TET2, TRDMT1 and YBX1) had negative association with monocyte and positive association with CD56^{bright} natural killer cell, immature dendritic cell, plasmacytoid dendritic cell and natural killer cell. We also observed that NSUN4 was positively correlated with those immune cells (Fig. 6B), which inspired us to explore the relationship between NSUN4 expression and immune cell infiltration. Results showed that patients with high NSUN4 expression also had elevated immune cell infiltration (Fig. 6C).

Next, we employed the “GSVA” and “GSEABase” packages to determine immune-related functions in the two clusters, and presented the results using boxplots and heatmaps. Results showed that the 11 types of immune function had different profiles across the subtypes (Figs. 7A,B). Apart from dendritic cells (DCs), it was evident that cluster A had markedly higher immune activity than cluster B. The m5C cluster A got higher immune score ($p = 0.00016$) and stromal score ($p = 0.093$) than cluster B (Figs. 7C–E). Considering the relevance of immune checkpoints to immunotherapy, we studied the immune checkpoint expression and found that all the 47 immune checkpoint genes (ICGs) were significantly overexpressed in cluster A relative to cluster B (Fig. 8A). Previous studies on pulmonary fibrosis have shown that PD-1/PD-L1 promotes the development of pulmonary fibrosis through different pathways^{32–35}. Therefore, generated violin plots to visualize PD-1/PD-L1 and found that it was significantly upregulated in cluster A ($p < 0.001$) (Figs. 8B,C).

GO and KEGG analyses revealed that a total of 3346 DEGs between m5C cluster A and cluster B have relevance in small molecule drug screening (Supplementary Table S3). Next, we selected the top 11 genes, based on adj. P values as well as removal of invalid genes and valid but not used genes in query, were selected from each of the upregulated DGEs and downregulated DEGs ($|\log_{2}FC| > 2.4$ and adj. $P < 0.001$). Results of the top 11 genes

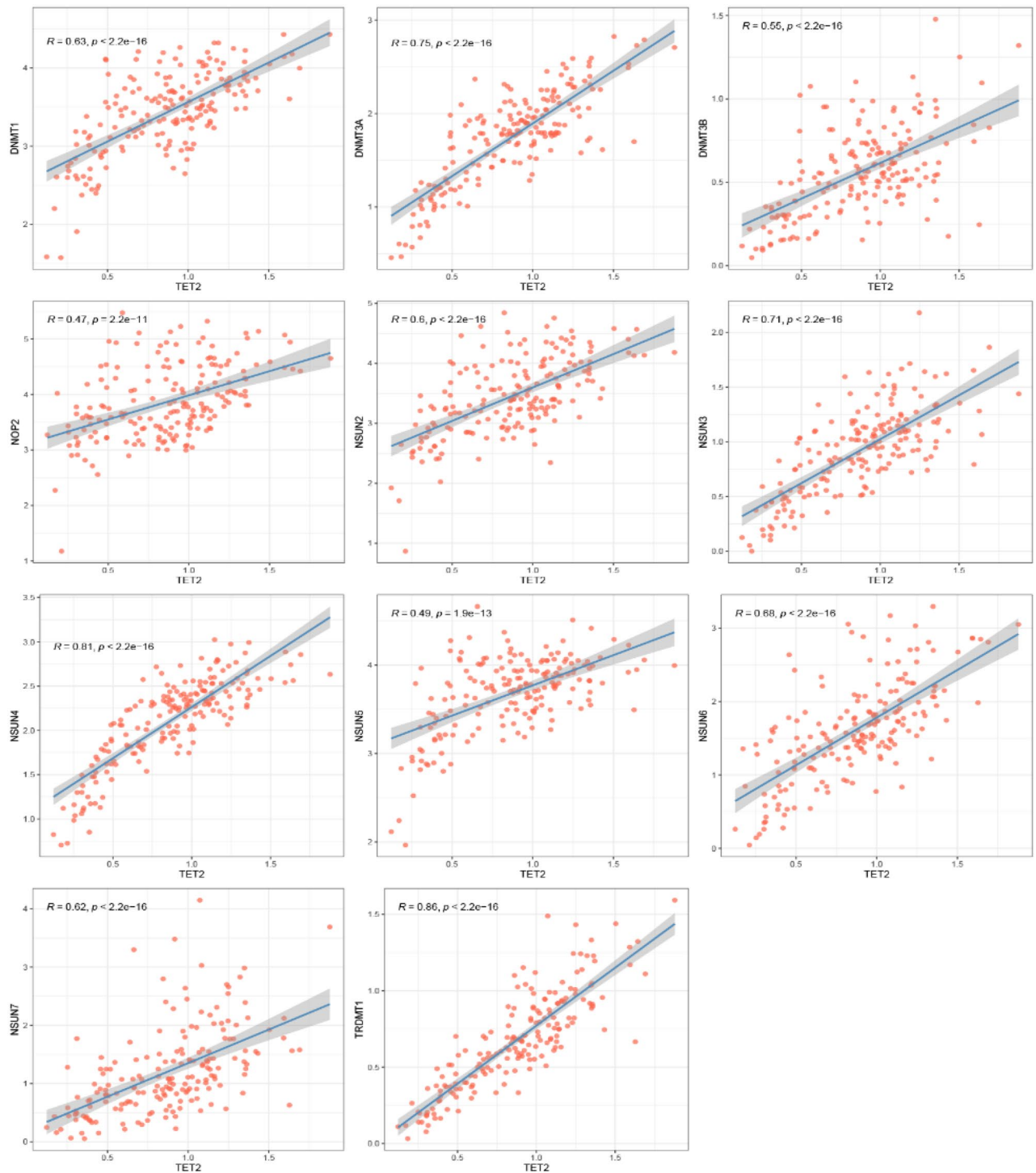


Figure 3. Correlation between m5C erasers and writers. Eraser gene: TET2. Writer genes: DNMT1, DNMT3A, DNMT3B, NOP2, NSUN2, NSUN3, NSUN4, NSUN5, NSUN6, NSUN7, TRDMT1. The threshold value: $|R| > 0.4$ and $P < 0.001$. R, correlation coefficient.

are shown in Supplementary Tables S4, S5, while the structure of ruxolitinib is presented in Fig. 8D,E. The top 10 strongest inhibitory small molecule drugs are outlined in Table 2.

GO analysis revealed that the DEGs were mainly enriched in RNA splicing (GO:0008380), RNA splicing via transesterification reactions (GO:0000375), mitochondrial inner membrane (GO:0005743), and electron transfer activity (GO:0009055) (Fig. 8F). On the other hand, KEGG pathway analysis showed that the DEGs were mainly involved in NOD—like receptor signaling (hsa04621) and FoxO signaling (hsa04068) pathways (Fig. 8G).

Gene1	Gene2	cor	<i>p</i> value
TET2	NOP2	0.468168	1.82E-11
TET2	NSUN2	0.604137	8.71E-20
TET2	NSUN3	0.711152	8.19E-30
TET2	NSUN4	0.814944	3.18E-45
TET2	NSUN5	0.485332	2.53E-12
TET2	NSUN6	0.679229	2.24E-26
TET2	NSUN7	0.624686	2.06E-21
TET2	DNMT1	0.632135	4.94E-22
TET2	DNMT3A	0.752238	5.43E-35
TET2	DNMT3B	0.546126	8.97E-16
TET2	TRDMT1	0.861982	6.93E-56

Table 1. The correlation results between erasers and writers.

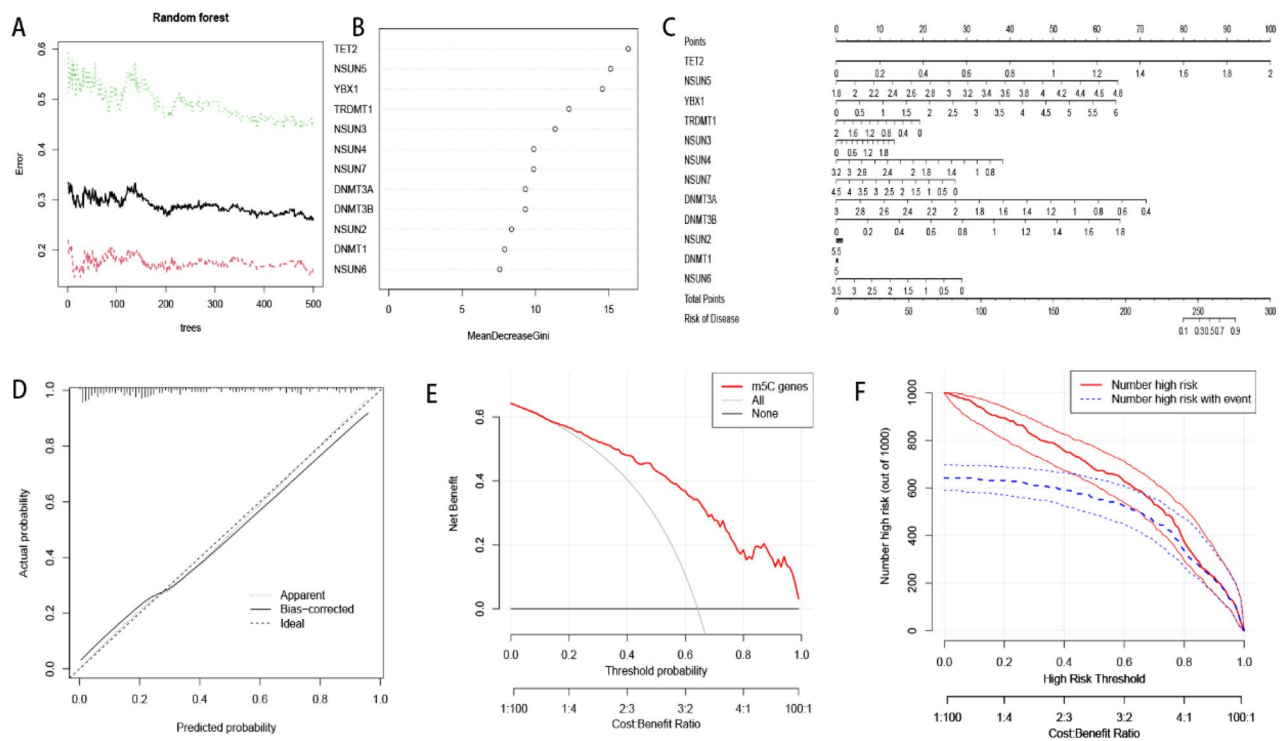


Figure 4. The RF and nomogram model construction. (A) The red line represented the error levels of treat groups, the green line represented control groups and the black line represented overall samples. (B) The importance of the m5C regulators were calculated based on the RF model. (C) The construction of nomogram model and gene score was used to predict prevalence. (D) The accuracy of nomogram model was assessed by calibration curve. (E) The decision curve might be benefit to the disease. (F) Clinical impact curve was applied for assessing clinical impact of the model.

Identification of the two gene clusters. Next, we used consensus cluster analysis to distinguish among DEGs patterns, and divided the DEG patterns into 2 clusters, 3 clusters, 4 clusters or 5 clusters, and identified 2 clusters (gene cluster A and cluster B) (Fig. 9A). Summarily, DEGs and 12 m5C regulators (DNMT1, DNMT3A-3B, NSUN2-7, TET2, TRDMT1 and YBX1) were significantly upregulated in cluster A than B (Figs. 9B,C). A similar pattern was observed with regards to m5C patterns, where patients in cluster A tended to be diagnosed more with IPF, exhibit more lung biopsies and diverse races compared to their counterparts in B. These findings were confirmed by PCA plots (Fig. 9D). In IPF addition, gene cluster A was associated with immune cell infiltration, and cluster A was significantly associated with Th2 cells and MCs than B ($p < 0.001$) (Fig. 10A). We also studied expression of immune checkpoints and found a significantly higher expression of all the 47 ICGs in cluster A relative to B (Fig. 10B). Differences across the subtypes are presented by boxplots and heatmaps in Figs. 11A,B. Apart from dendritic cells (DCs), it was evident that cluster A had higher immune activity, immune scores ($p = 1.3e-06$) and stromal scores ($p = 0.0083$) than cluster B (Figs. 11C-E).

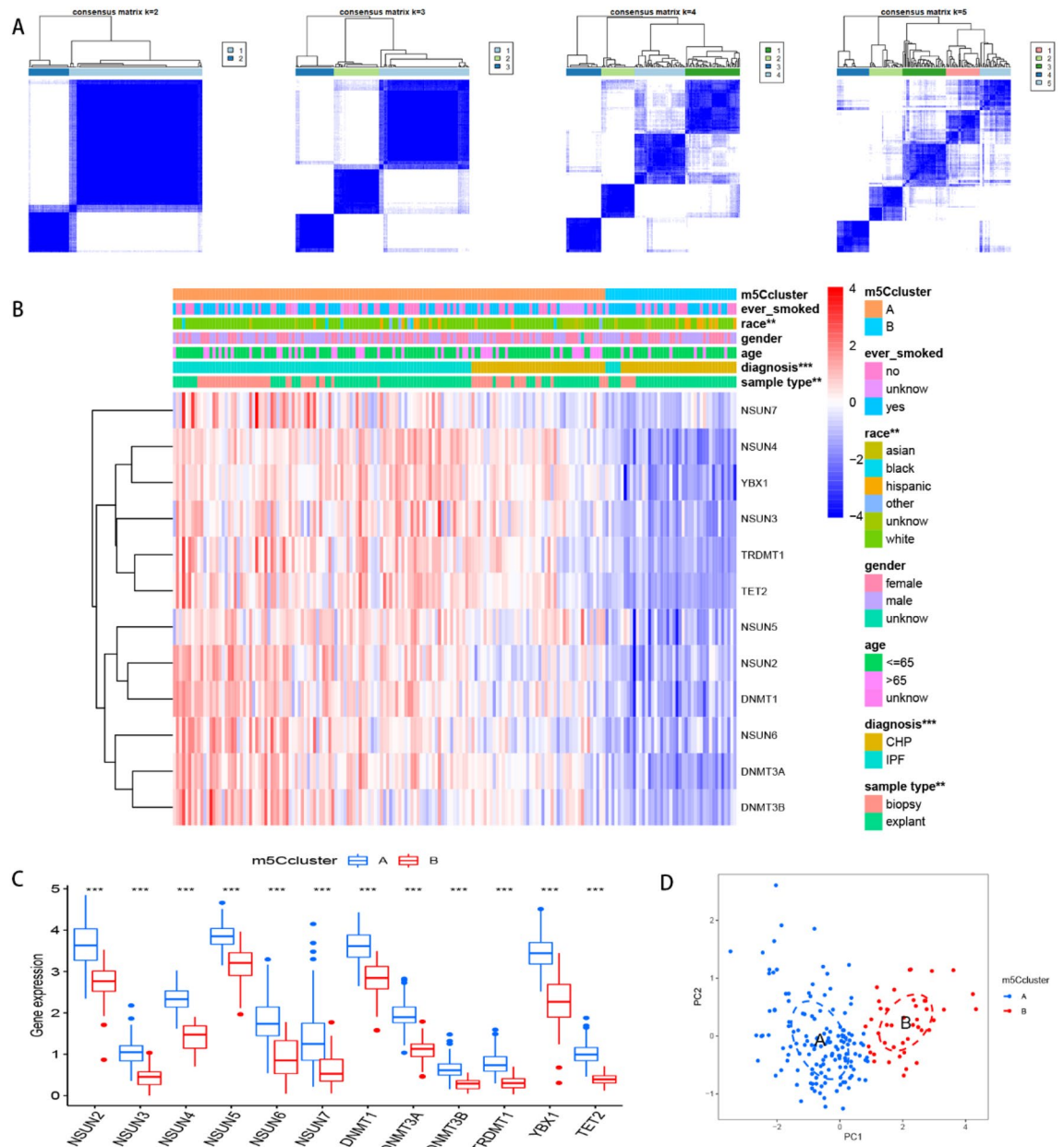


Figure 5. Consensus clustering of m5C regulators. **(A)** Consensus matrices of the 12 m5C regulators ($k=2-5$). **(B)** Expression of 12 significant m5C regulators in the heatmap. **(C)** Expression of 12 significant m5C regulators in the boxplot. **(D)** PCA showed the striking difference in modification cluster A (the blue pattern) and cluster B (the red pattern). * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Comparison in m5C and gene patterns. We generated a PCA plot to depict m5C-related scores across samples. Summarily, m5C and gene typing were analogous, with the sankey diagram distinctly indicating that the clusters were associated with m5C-related scores (Fig. 12A). Notably, cluster A exhibited significantly higher scores than cluster B (Figs. 12B,C). Previous studies have suggested that high mucin 5B (MUC5B) expression is associated with honeycombing and fibrosis in both CHP and IPF subjects^{11,36,37}, combining our earlier results suggested a strong link between m5C/gene cluster A and Th2 cells. Consequently, we analyzed the relation between m5C clusters and Th2 cytokines [Interleukin (IL) -4, IL-5, IL-10, IL-13, and thymic stromal lymphopoeitin (TSLP)], as well as MUC5B, and found a significant positive association (Figs. 12D,E). Notably, cluster A recorded higher expression levels than B, indicating that the former was highly linked to honeycombing and fibrosis, as well as lung fibrosis characterized by the Th2 immune response.

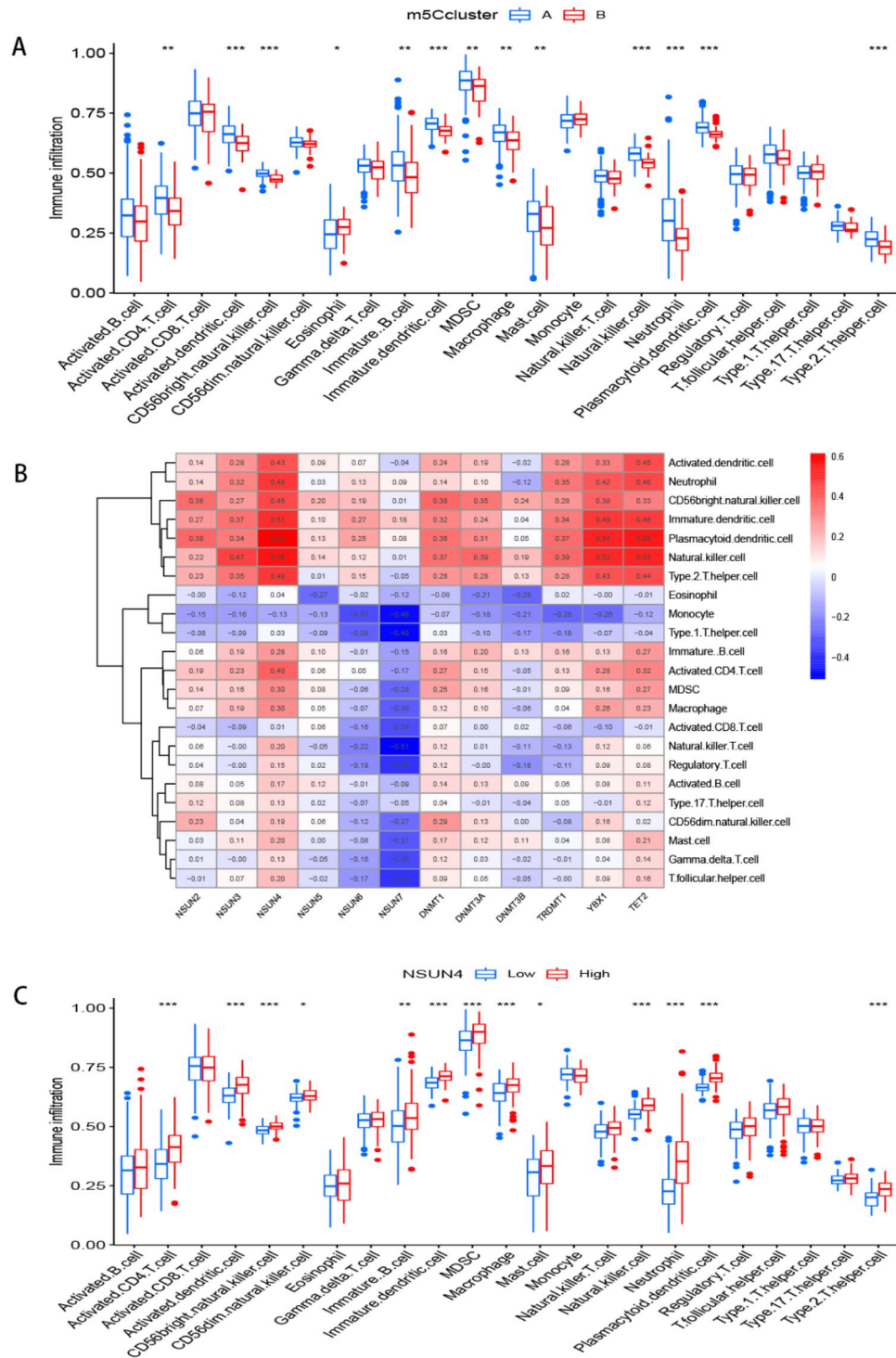


Figure 6. Single sample gene set enrichment analysis of immune cells infiltration. (A) The relationship between immune cells infiltration with two m5C patterns. (B) The heatmap of the 12 significant m5C regulators and infiltrating immune cells. (C) Immune cell infiltration between high NSUN4 expression pattern and low NSUN4 expression pattern. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Discussion

In our previous study, we demonstrated that two distinct m6A patterns were associated with fibrosis and might distinguish CHP and IPF³⁸. Inspired by these findings, we analyzed expression patterns of the major m5C regulators altered in GSE150910 and found that they played a key role in occurrence and development of CHP and IPF. Firstly, we explored the correlation between “writers”, “erasers” and “readers” and found a significant relationship between various writers and one eraser. Next, we used five candidate m5C regulators to construct

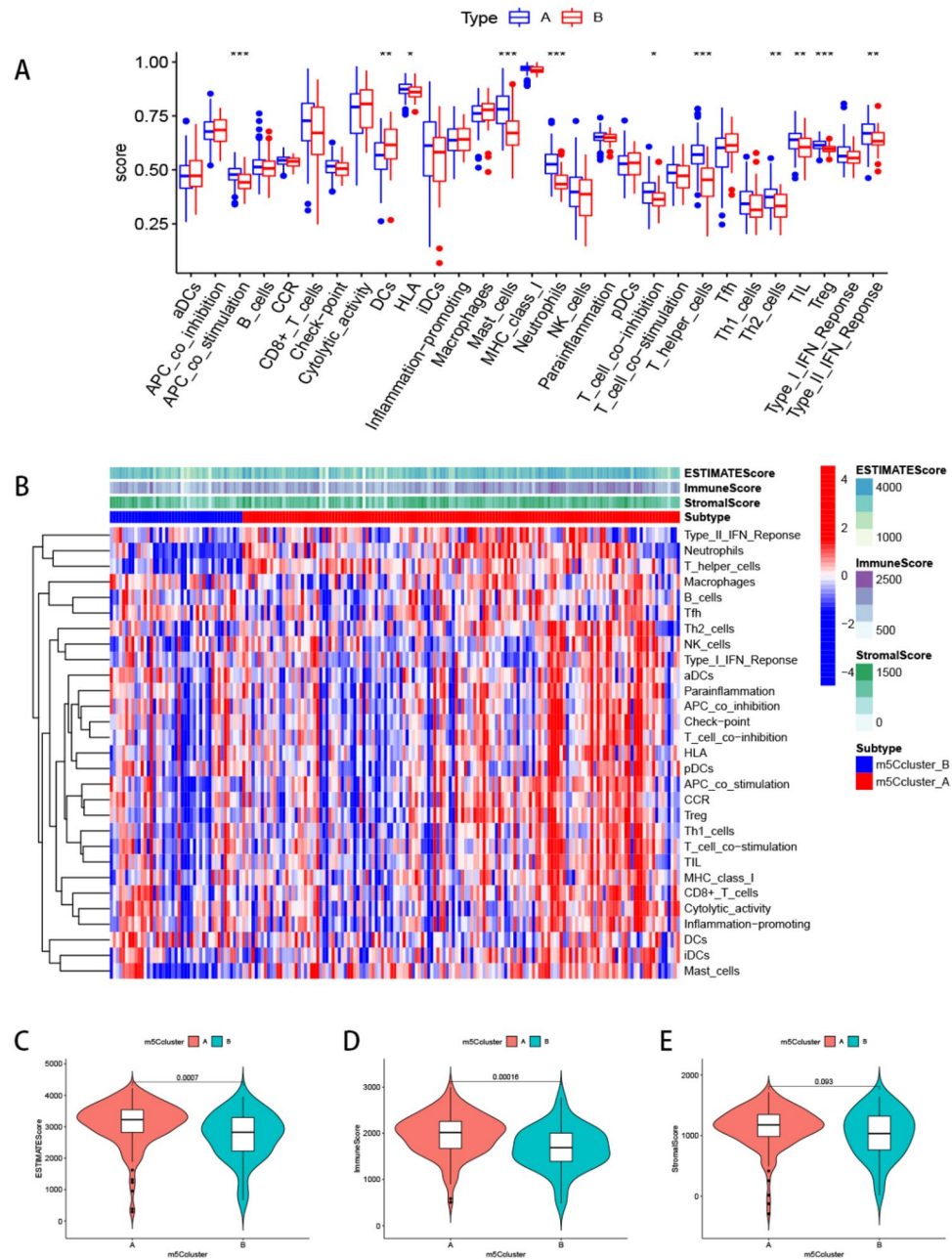


Figure 7. Single sample gene set enrichment analysis of the two m5Cclusters. **(A)** The boxplot of immune function in the two m5C patterns. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. **(B)** The heatmap of immune function in the two m5C patterns. Other characteristics including estimate score, immune score, and stromal score. Differential estimate score **(C)**, immune score **(D)**, and stromal score **(E)** between m5Ccluster A and B.

a nomogram model for predicting the prevalence of CHP or IPF patients. Results indicated that this model was effective in diagnosis and treatment of these patients. Finally, we studied various methylation modification patterns to explore two distinct m5C patterns and hypothesized the two patterns may associate with lung fibrosis and distinguish patients with CHP or IPF.

Analysis of clinical characteristics, immune microenvironment, and potential drug therapies were performed. Clinical characteristics in the two m5C patterns revealed more IPF patients in m5C cluster A than cluster B, implying that these two patterns might distinguish IPF from CHP. Previous studies have shown that HP is mediated by Th1 cytokine immunity³⁹ and Th2 is related to formation of pulmonary fibrosis^{40–43}. MCs cooperate with fibroblasts to promote pulmonary fibrosis⁴⁴, and accumulation of MCs has been found in HP⁴⁵, IPF^{44,46}, sarcoidosis⁴⁷ and silicosis⁴⁸. Moreover, Cha SI et al. found the MCs in IPF is more than CHP⁴⁶. In our study, more Th2 cells and MCs aggregation in cluster A further implied that CHP and IPF could be distinguished by the two patterns, and the mechanism of fibrosis might be fundamentally different between CHP and IPF. Accumulation

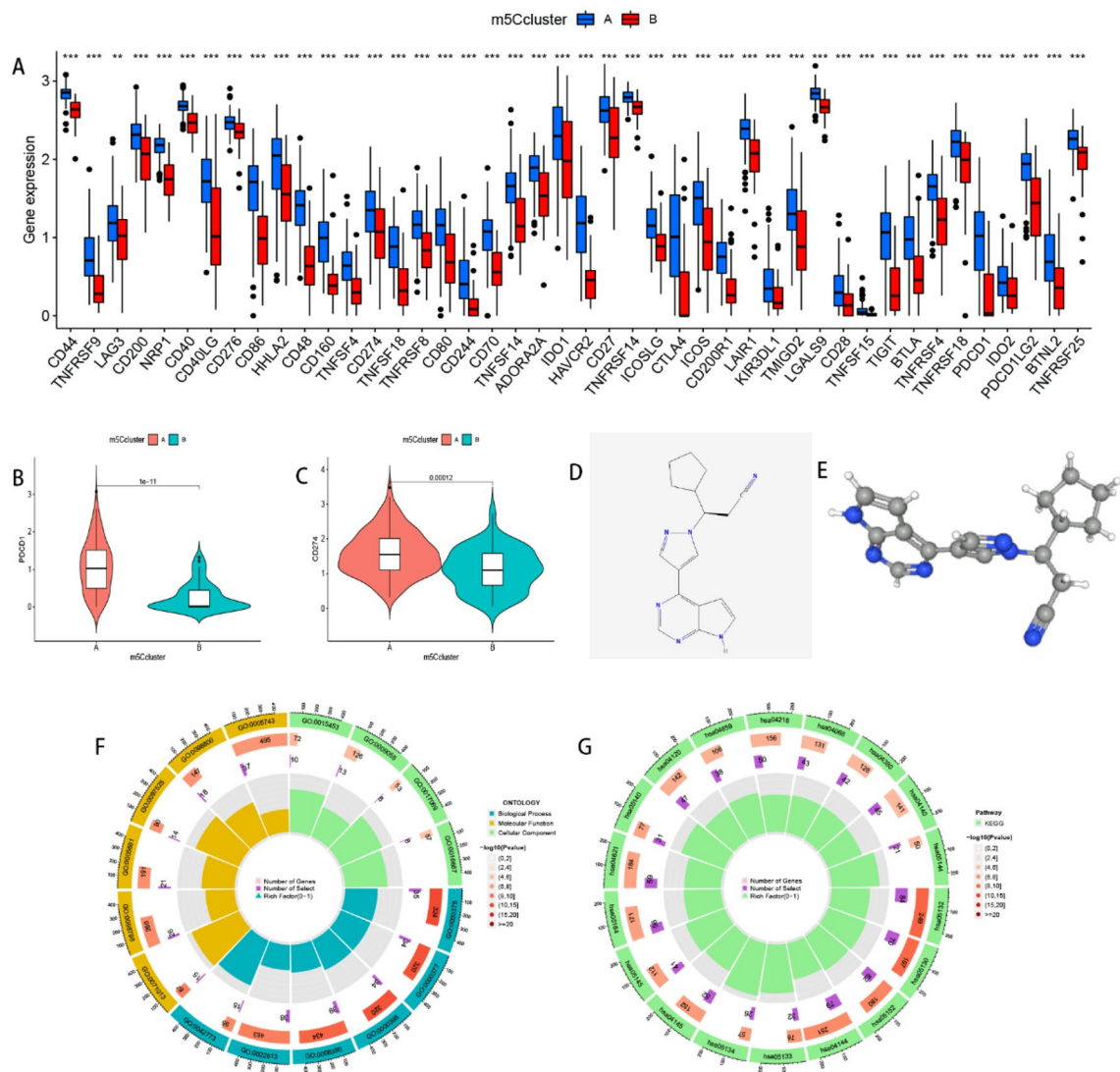


Figure 8. Analyses of immune checkpoints, small molecule drug therapy, GO, as well as KEGG Pathway in two m5Cclusters. (A) The boxplot of immune checkpoints in the two m5C patterns. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. (B) The expression of PDCD1 (PD-1) in the two m5C patterns. (C) The expression of CD274 (PD-L1) in the two m5C patterns. Two-dimensional molecular structure (D) and three-dimensional molecular structure (E) of ruxolitinib. The GO (F) and KEGG (G) enrichment analysis for the m5C-related differentially expressed genes (DEGs).

Name	ID	Score	Description
ruxolitinib	BRD-K53972329	- 98.91	JAK inhibitor
orantinib	BRD-K91696562	- 98.77	FGFR inhibitor
UNC-0321	BRD-K74236984	- 98.72	Histone lysine methyltransferase inhibitor
AR-A014418	BRD-K67860401	- 97.64	Glycogen synthase kinase inhibitor
doxercalciferol	BRD-K14550461	- 97.11	Vitamin D receptor agonist
PT-630	BRD-A73680854	- 96.76	Dipeptidyl peptidase inhibitor
BRL-50481	BRD-K84266862	- 96.71	Phosphodiesterase inhibitor
rolipram	BRD-A34255068	- 96.61	Phosphodiesterase inhibitor
actarit	BRD-K33483813	- 95.39	Interleukin receptor agonist
forskolin	BRD-A70449690	- 95.09	Adenylyl cyclase activator

Table 2. Top 10 strongest inhibitory small molecule drugs.

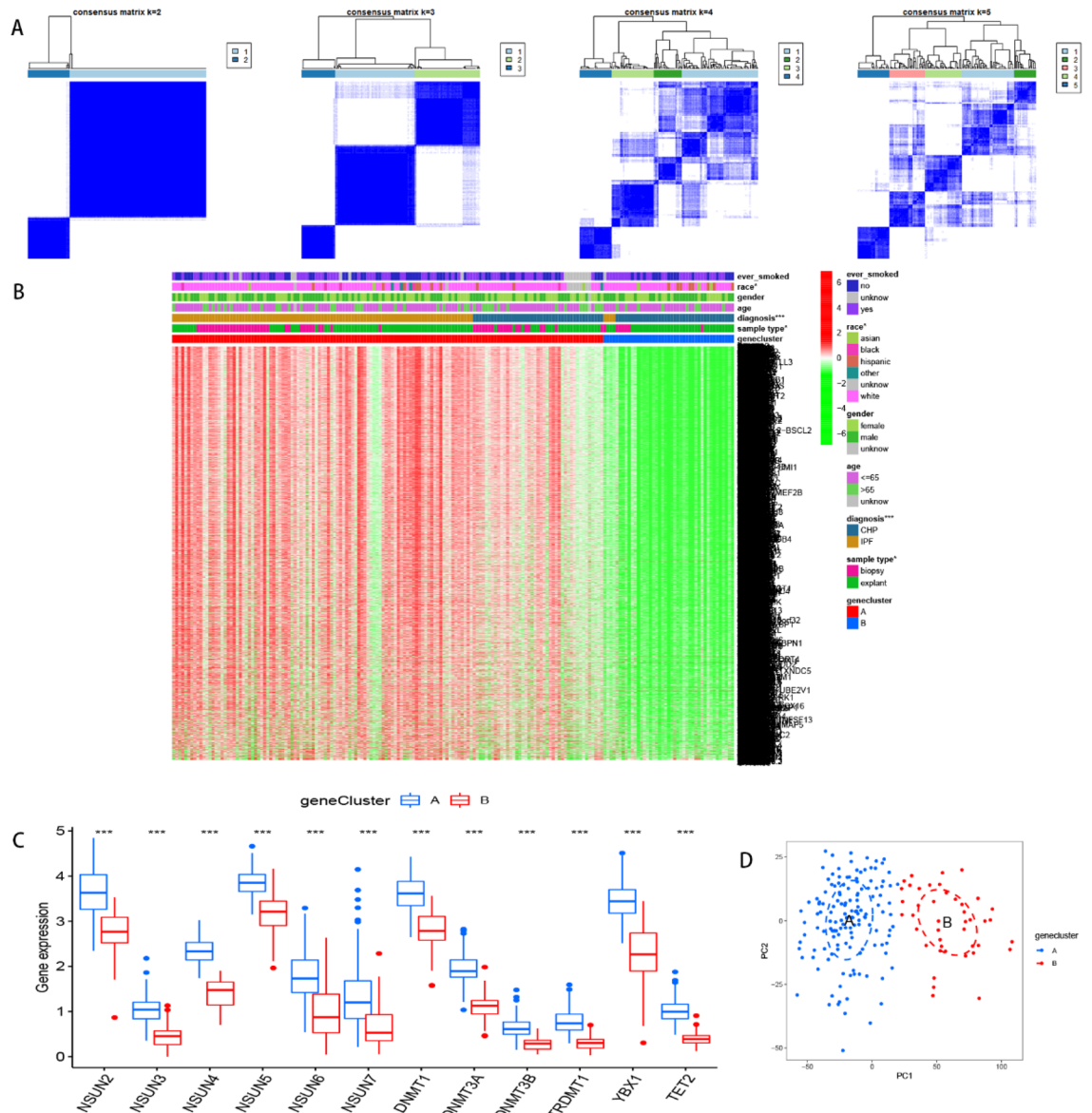


Figure 9. Consensus clustering of the 3346 m5C-related DEGs. **(A)** Consensus matrices of the DGEs ($k=2-5$). **(B)** Expression of the DEGs in gene cluster A and cluster B. **(C)** Expression of the 12 m5C regulators in gene cluster A and cluster B. **(D)** Principal component analysis for the expression profiles of gene subtypes, also showing a remarkable difference between different modification patterns. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

of eosinophils in cluster B was consistent with the deduction^{49,50}. Next, we compared expression levels of Th2-related cytokines between the clusters and found a higher expression in cluster A than B. Besides, MUC5B promoter variant rs35705950 is a high-risk factor for IPF¹¹, the cluster A exhibited higher MUC5B expression than B. These results suggested that cluster A was related to IPF. We also found higher immune and stromal scores in m5C cluster A than B, implying that immune element plays an underlying impact on cluster A. PD-1/PD-L1 was significantly upregulated in cluster A than B, indicating that patients in the former cluster were more likely to promote fibrosis and have better immune efficacy than their cluster B counterparts.

GO analysis revealed that the identified DEGs were mainly enriched in NOD-like receptor and FoxO signaling pathways, which have been shown to play important roles in inflammatory response^{51,52}, indicating that their immune components involved in the pathological processes and immunoregulatory mechanisms of CHP and IPF. We found two distinct DEGs-related patterns, with further analysis of clinical characteristics and immune microenvironment revealing patterns that were consistent with those observed in m5C. In general, the two m5C patterns have potential to distinguish patients with CHP from IPF.

The present study had some limitations. Firstly, the GSE150910 used herein lacked data on pathological classification and imaging features for patients with CHP or IPF. Although the probability of misdiagnosis in these samples was low, it cannot be ruled out. Secondly, although we searched and downloaded all interstitial pneumonia datasets from the GEO database, no other dataset contained both IPF and CHP, thus we had no

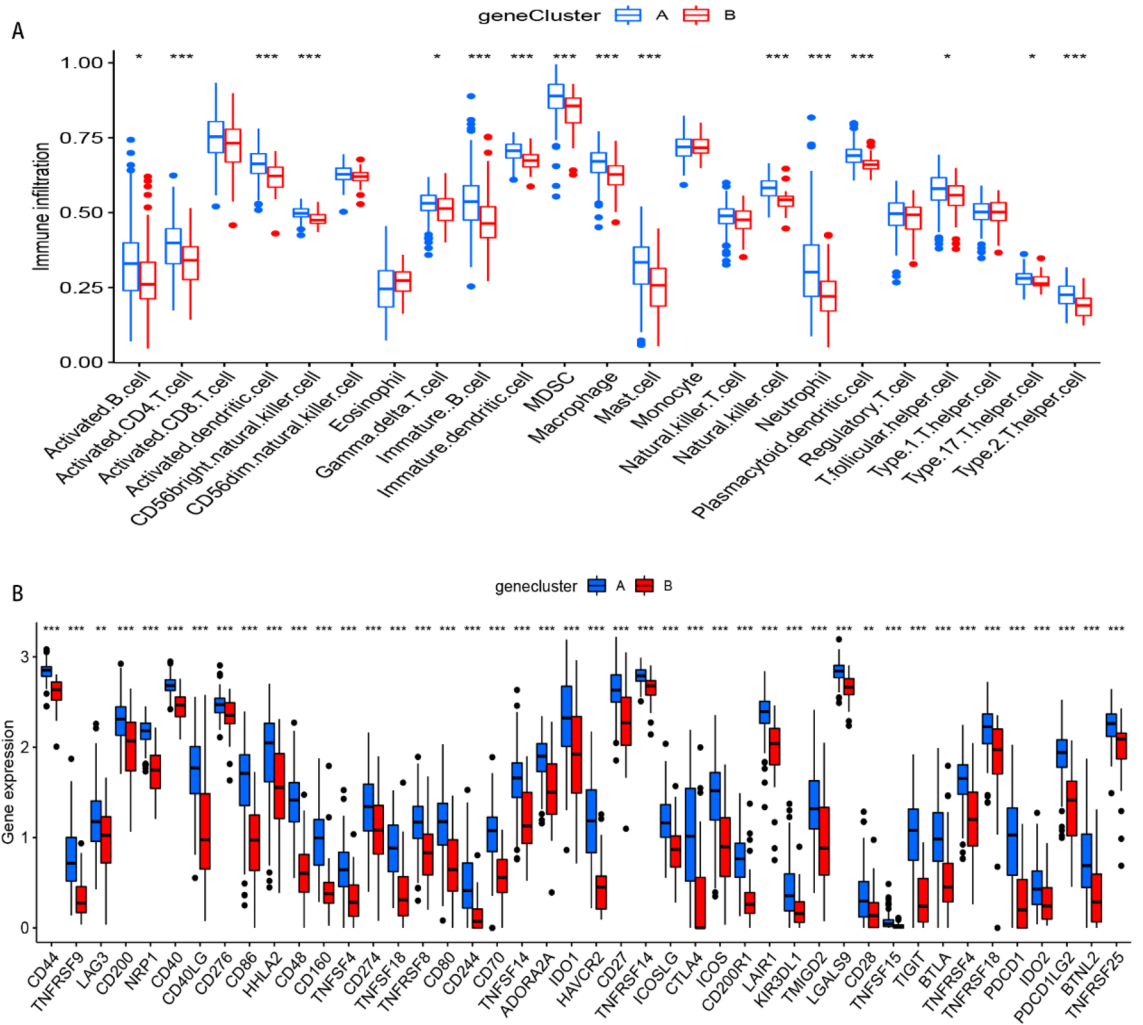


Figure 10. Analyses of immune cells infiltration and immune checkpoints in two m5C clusters. **(A)** The relationship between immune cells infiltration with two m5C patterns. **(B)** The boxplot of immune checkpoints in the two m5C patterns. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

validation dataset. In future, we expect to use NGS of clinical specimens coupled with experimental studies based on pulmonary fibrosis mouse models to validate these findings.

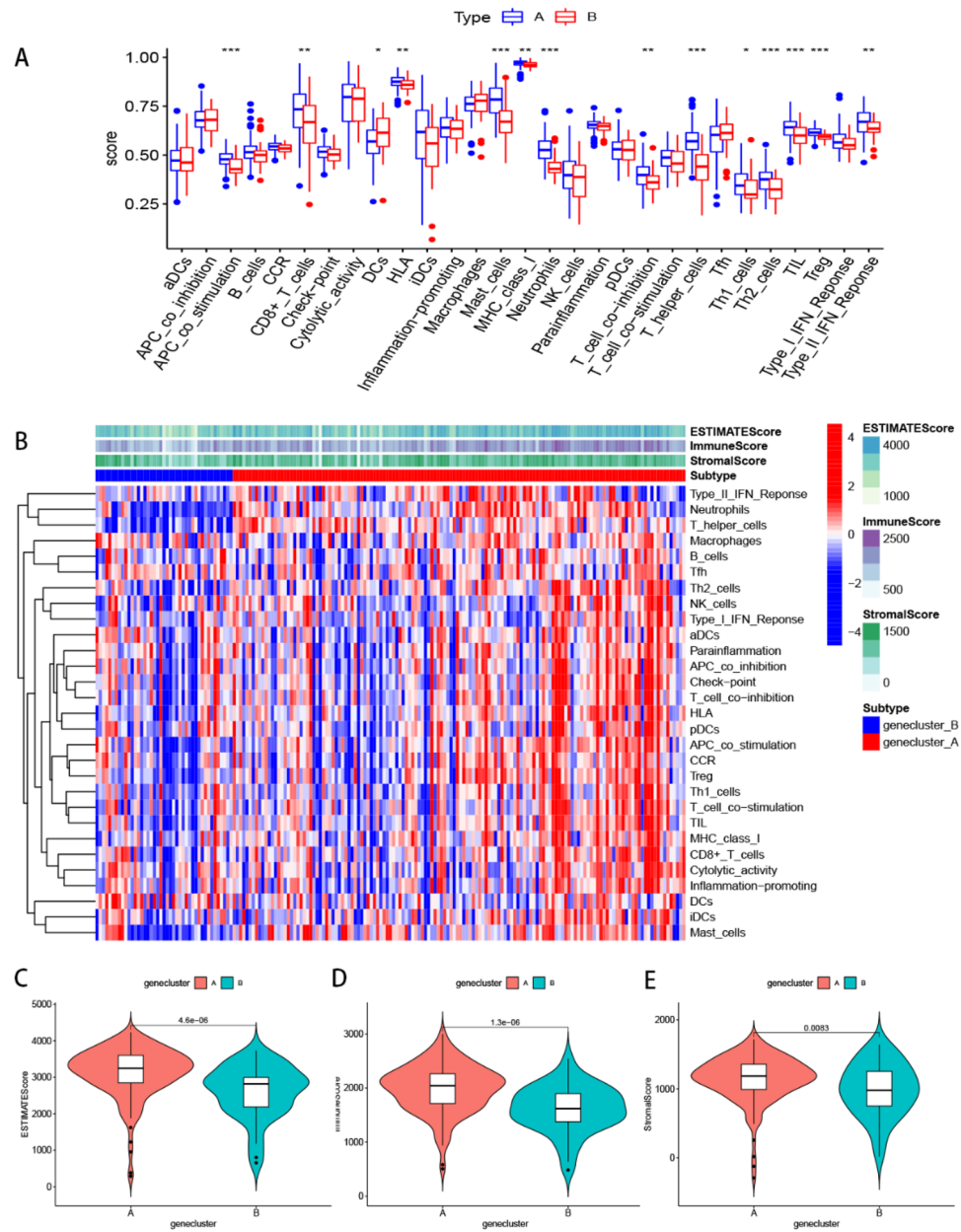


Figure 11. Single sample gene set enrichment analysis of the two gene clusters. **(A)** The boxplot of immune function in the two gene patterns. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. **(B)** The heatmap of immune function in the two gene patterns. Other characteristics including estimate score, immune score, and stromal score. Differential estimate score **(C)**, immune score **(D)**, and stromal score **(E)** between m5Ccluster A and B.

Conclusion

In summary, our findings revealed the role of m5C methylation regulators in development of pulmonary fibrosis. These genetic factors have potential in distinguishing CHP and IPF patients, thus are expected to be used in developing new immunotherapy strategies for these patients.

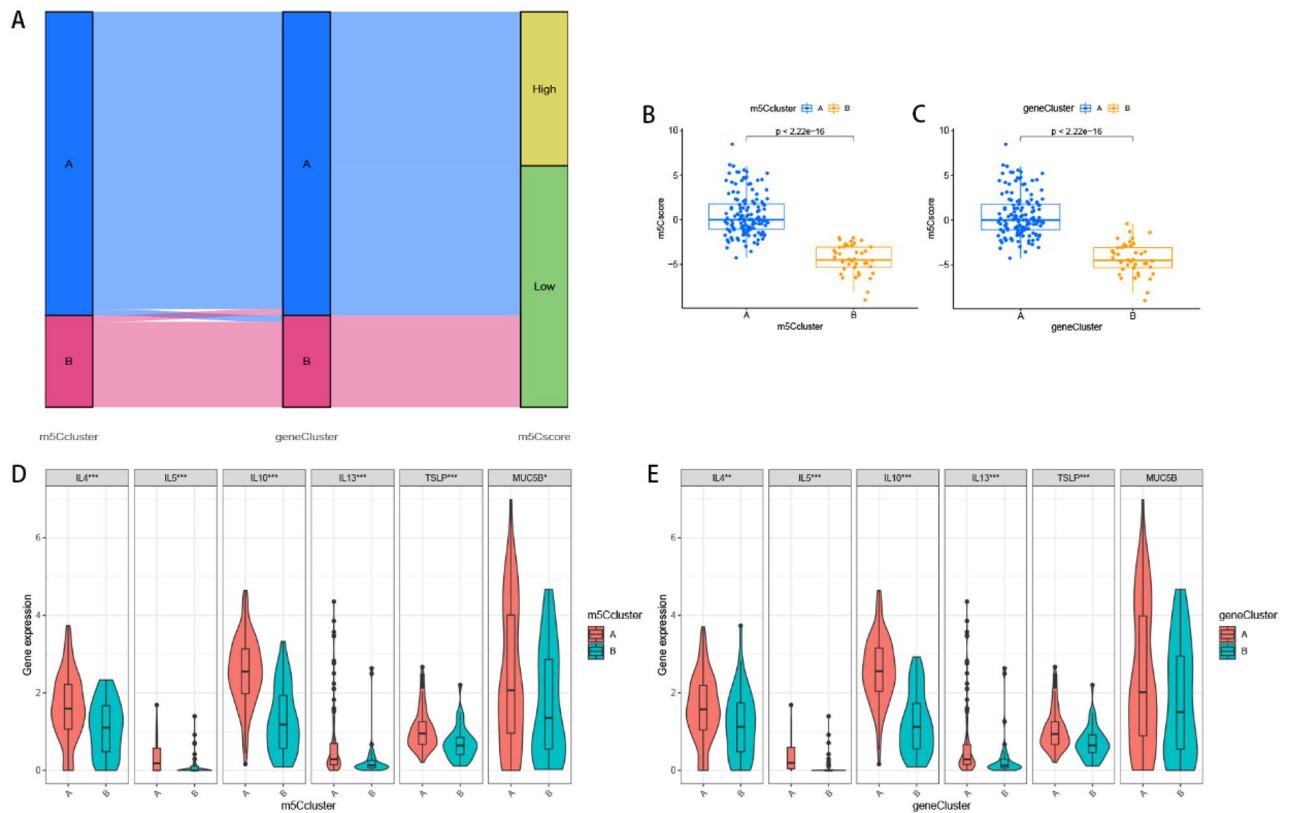


Figure 12. Comparison of m5C patterns and gene patterns. (A) Sankey diagram of the relationship between two m5C patterns, two gene patterns, and m5C scores. Differences in m5C score based on PCA algorithm between the two m5C patterns (B) or the two gene patterns (C). Differential expression levels of IL-4, IL-5, IL-10, IL-13, TSLP and MUC5B between m5C cluster A and cluster B (D) or gene cluster A and cluster B (E). * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Data availability

The dataset is available in online repositories (NCBI-GEO, <https://www.ncbi.nlm.nih.gov/geo>), under accession number GSE150910. All data to support our conclusions are included in this article.

Received: 22 October 2022; Accepted: 28 March 2023

Published online: 12 April 2023

References

- Koster, M. A. *et al.* Diagnosis of hypersensitivity pneumonitis in adults, 2020 clinical practice guideline: Summary for clinicians. *Ann. Am. Thorac. Soc.* **18**(4), 559–566 (2021).
- Morell, F. *et al.* Chronic hypersensitivity pneumonitis in patients diagnosed with idiopathic pulmonary fibrosis: A prospective case-cohort study. *Lancet Respir. Med.* **1**(9), 685–694 (2013).
- Invernizzi, R. *et al.* The respiratory microbiome in chronic hypersensitivity pneumonitis is distinct from that of idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **203**(3), 339–347 (2021).
- Morell, F. *et al.* Bird fancier's lung: A series of 86 patients. *Medicine (Baltimore)* **87**(2), 110–130 (2008).
- Lederer, D. J. & Martinez, F. J. Idiopathic pulmonary fibrosis. *N. Engl. J. Med.* **378**(19), 1811–1823 (2018).
- Hanak, V., Golbin, J. M., Hartman, T. E. & Ryu, J. H. High-resolution CT findings of parenchymal fibrosis correlate with prognosis in hypersensitivity pneumonitis. *Chest* **134**(1), 133–138 (2008).
- Vourlekis, J. S. *et al.* The effect of pulmonary fibrosis on survival in patients with hypersensitivity pneumonitis. *Am. J. Med.* **116**(10), 662–668 (2004).
- Johansson, K. A. *et al.* A diagnostic model for chronic hypersensitivity pneumonitis. *Thorax* **71**(10), 951–954 (2016).
- Lacasse, Y. *et al.* Clinical diagnosis of hypersensitivity pneumonitis. *Am. J. Respir. Crit. Care Med.* **168**(8), 952–958 (2003).
- Sharma, A., Ferraro, V., Renzoni, E. A. & Morisset, J. Chronic hypersensitivity pneumonitis: Real world diagnostic criteria. *Curr. Opin. Pulm. Med.* **27**(5), 414–421 (2021).
- Furusawa, H. *et al.* Chronic hypersensitivity pneumonitis, an interstitial lung disease with distinct molecular signatures. *Am. J. Respir. Crit. Care Med.* **202**(10), 1430–1444 (2020).
- Churg, A., Bilawich, A. & Wright, J. L. Pathology of chronic hypersensitivity pneumonitis what is it? What are the diagnostic criteria? Why do we care?. *Arch. Pathol. Lab Med.* **142**(1), 109–119 (2018).
- Cohn, W. E. Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: Isolation, structure, and chemical characteristics. *J. Biol. Chem.* **235**, 1488–1498 (1960).
- Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**(7), 1187–1200 (2017).
- Lee, M., Kim, B. & Kim, V. N. Emerging roles of RNA modification: m(6)A and U-tail. *Cell* **158**(5), 980–987 (2014).

16. Xu, Y. *et al.* Role of Main RNA methylation in hepatocellular carcinoma: N6-methyladenosine, 5-methylcytosine, and n1-methyladenosine. *Front. Cell Dev. Biol.* **9**, 767668 (2021).
17. Gilbert, W. V., Bell, T. A. & Schaening, C. Messenger RNA modifications: Form, distribution, and function. *Science* **352**(6292), 1408–1412 (2016).
18. Oerum, S., Degut, C., Barraud, P. & Tisne, C. m1A Post-Transcriptional Modification in tRNAs. *Biomolecules* **7**(1), 20 (2017).
19. Liu, J. & Jia, G. Methylation modifications in eukaryotic messenger RNA. *J. Genet. Genomics* **41**(1), 21–33 (2014).
20. Chen, Y. S., Yang, W. L., Zhao, Y. L. & Yang, Y. G. Dynamic transcriptomic m(5) C and its regulatory role in RNA processing. *Wiley Interdiscip. Rev. RNA* **12**(4), e1639 (2021).
21. Garcia-Vilchez, R., Sevilla, A. & Blanco, S. Post-transcriptional regulation by cytosine-5 methylation of RNA. *Biochim. Biophys. Acta Gene Regul. Mech.* **1862**(3), 240–252 (2019).
22. He, Z., Xu, J., Shi, H. & Wu, S. m5CRegpred: Epitranscriptome Target Prediction of 5-Methylcytosine (m5C) Regulators Based on Sequencing Features. *Genes (Basel)*. **13**(4), 677 (2022).
23. Sun, G. *et al.* Multi-omics analysis of expression and prognostic value of NSUN members in prostate cancer. *Front Oncol.* **12**, 965571 (2022).
24. Sun, Z. *et al.* Aberrant NSUN2-mediated m(5)C modification of H19 lncRNA is associated with poor differentiation of hepatocellular carcinoma. *Oncogene* **39**(45), 6906–6919 (2020).
25. Wong, K. K. DNMT1: A key drug target in triple-negative breast cancer. *Semin. Cancer Biol.* **72**, 198–213 (2021).
26. Chen, W. T. *et al.* a simple algorithm using ventilator parameters to predict successfully rapid weaning program in cardiac intensive care unit patients. *J. Pers. Med.* **12**(3), 501 (2022).
27. Wang, J. *et al.* Causative classification of ischemic stroke by the machine learning algorithm random forests. *Front. Aging Neurosci.* **14**, 788637 (2022).
28. Hou, Y. *et al.* Characterization of immune-related genes and immune infiltration features in epilepsy by multi-transcriptome data. *J. Inflamm. Res.* **15**, 2855–2876 (2022).
29. Lamb, J. *et al.* The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**(5795), 1929–1935 (2006).
30. Subramanian, A. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**(6), 1437–1452 (2017).
31. Li, Y. *et al.* Molecular characterization and clinical relevance of m(6)A regulators across 33 cancer types. *Mol. Cancer* **18**(1), 137 (2019).
32. Duitman, J., van den Ende, T. & Spek, C. A. Immune checkpoints as promising targets for the treatment of idiopathic pulmonary fibrosis?. *J. Clin. Med.* **8**(10), 1547 (2019).
33. Li, N. *et al.* Bioinformatic analysis of differentially expressed genes and pathways in idiopathic pulmonary fibrosis. *Ann. Transl. Med.* **9**(18), 1459 (2021).
34. Celada, L. J. *et al.* PD-1 up-regulation on CD4(+) T cells promotes pulmonary fibrosis through STAT3-mediated IL-17A and TGF-beta1 production. *Sci. Transl. Med.* **10**(460), eaar8356 (2018).
35. Li, Y. *et al.* CD247, a potential T cell-derived disease severity and prognostic biomarker in patients with idiopathic pulmonary fibrosis. *Front. Immunol.* **12**, 762594 (2021).
36. Yang, I. V. *et al.* Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* **68**(12), 1114–1121 (2013).
37. Seibold, M. A. *et al.* A common MUC5B promoter polymorphism and pulmonary fibrosis. *N. Engl. J. Med.* **364**(16), 1503–1512 (2011).
38. Zhou, Y., Fang, C., Sun, Q. & Dong, Y. Relevance of RNA N6-methyladenosine regulators for pulmonary fibrosis: Implications for chronic hypersensitivity pneumonitis and idiopathic pulmonary fibrosis. *Front. Genet.* **13**, 939175 (2022).
39. Matsuno, Y. *et al.* Overexpression of GATA-3 protects against the development of hypersensitivity pneumonitis. *Am. J. Respir. Crit. Care Med.* **176**(10), 1015–1025 (2007).
40. Andrews, K. *et al.* Chronic hypersensitivity pneumonitis caused by *Saccharopolyspora rectivirgula* is not associated with a switch to a Th2 response. *Am. J. Physiol. Lung Cell Mol. Physiol.* **310**(5), L393–402 (2016).
41. Jimenez-Alvarez, L. *et al.* Inflammatory response and dynamics of lung T cell subsets in Th1, Th2 biased and Th2 deficient mice during the development of hypersensitivity pneumonitis. *Exp. Mol. Pathol.* **88**(3), 407–415 (2010).
42. Barrera, L. *et al.* Functional diversity of T-cell subpopulations in subacute and chronic hypersensitivity pneumonitis. *Am. J. Respir. Crit. Care Med.* **177**(1), 44–55 (2008).
43. Wallace, W. A., Ramage, E. A., Lamb, D. & Howie, S. E. A type 2 (Th2-like) pattern of immune response predominates in the pulmonary interstitium of patients with cryptogenic fibrosing alveolitis (CFA). *Clin. Exp. Immunol.* **101**(3), 436–441 (1995).
44. Wygrecka, M. *et al.* Mast cells and fibroblasts work in concert to aggravate pulmonary fibrosis: Role of transmembrane SCF and the PAR-2/PKC-alpha/Raf-1/p44/42 signaling pathway. *Am. J. Pathol.* **182**(6), 2094–2108 (2013).
45. Delgado, L. *et al.* Suberosis: involvement of bronchoalveolar +mastocytes in the genesis of interstitial involvement. *Arch Bronconeumol.* **35**(2), 71–78 (1999).
46. Cha, S. I. *et al.* Lung mast cell density defines a subpopulation of patients with idiopathic pulmonary fibrosis. *Histopathology* **61**(1), 98–106 (2012).
47. Ohrn, M. B. *et al.* Sarcoidosis patients have bronchial hyperreactivity and signs of mast cell activation in their bronchoalveolar lavage. *Respiration* **62**(3), 136–142 (1995).
48. Hamada, H. *et al.* Mast cell basic fibroblast growth factor in silicosis. *Am. J. Respir. Crit. Care Med.* **161**(6), 2026–2034 (2000).
49. Wang, S. *et al.* Ameliorative effects of eosinophil deficiency on immune response, endoplasmic reticulum stress, apoptosis, and autophagy in fungus-induced allergic lung inflammation. *Respir. Res.* **22**(1), 173 (2021).
50. Soler, P. *et al.* Role of mast cells in the pathogenesis of hypersensitivity pneumonitis. *Thorax* **42**(8), 565–572 (1987).
51. Liu, P. *et al.* NOD-like receptor signaling in inflammation-associated cancers: From functions to targeted therapies. *Phytomedicine* **64**, 152925 (2019).
52. Hammami, A., Abidin, B. M., Heinonen, K. M. & Stager, S. HIF-1alpha hampers dendritic cell function and Th1 generation during chronic visceral leishmaniasis. *Sci. Rep.* **8**(1), 3500 (2018).

Acknowledgements

We thank the authors who provided the GEO public datasets.

Author contributions

Y.Z. put forward the conception and contributed to the writing of this study. Z.H. and Q.S. improved the framework and provided the materials. Y.Z. and Z.H. performed and verified the data analysis of this experiment. Y.D. and Q.S. contributed to significant modification of this manuscript. All authors participating in the article approved the submitted version.

Funding

The study was supported by funds from the National Natural Science Foundation of China: TRIM37 inhibits the activation and inflammatory cytokine production of dendritic cell and the development of asthma (Grants 81800018).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-32452-4>.

Correspondence and requests for materials should be addressed to Y.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023