Identification of m6A regulator-mediated RNA methylation modification patterns and key immune-related genes involved in atrial fibrillation

Peng-Fei Zheng[1,2,3,*], Sen-Yu Zhou[3,4,*], Chang-Qing Zhong[1,2,3], Zhao-Fen Zheng[1,2,3], Zheng-Yu Liu[1,2,3], Hong-Wei Pan[1,2,3], Jian-Qiang Peng[1,2,3,&]

[1]Cardiology Department, Hunan Provincial People's Hospital, Furong, Changsha 410000, Hunan, China
[2]Clinical Research Center for Heart Failure in Hunan Province, Furong, Changsha 410000, Hunan, China
[3]Institute of Cardiovascular Epidemiology, Hunan Provincial People's Hospital, Furong, Changsha 410000, Hunan, China
[4]The First Affiliated Hospital of Hunan Normal University (Hunan Provincial People's Hospital), Furong, Changsha 410000, Hunan, China
*Equal contribution

Correspondence to: Hong-Wei Pan, Jian-Qiang Peng; email: panhongwei@hunnu.edu.cn, 2925772400@qq.com

ABSTRACT

The role of m6A in the regulation of the immune microenvironment in atrial fibrillation (AF) remains unclear. This study systematically evaluated the RNA modification patterns mediated by differential m6A regulators in 62 AF samples, identified the pattern of immune cell infiltration in AF and identified several immune-related genes associated with AF. A total of six key differential m6A regulators between healthy subjects and AF patients were identified by the random forest classifier. Three distinct RNA modification patterns (m6A cluster-A, -B and -C) among AF samples were identified based on the expression of 6 key m6A regulators. Differential infiltrating immune cells and HALLMARKS signaling pathways between normal and AF samples as well as among samples with three distinct m6A modification patterns were identified. A total of 16 overlapping key genes were identified by weighted gene coexpression network analysis (WGCNA) combined with two machine learning methods. The expression levels of the *NCF2* and *HCST* genes were different between controls and AF patient samples as well as among samples with the distinct m6A modification patterns. RT−qPCR also proved that the expression of *NCF2* and *HCST* was significantly increased in AF patients compared with control participants. These results suggested that m6A modification plays a key role in the complexity and diversity of the immune microenvironment of AF. Immunotyping of patients with AF will help to develop more accurate immunotherapy strategies for those with a significant immune response. The *NCF2* and *HCST* genes may be novel biomarkers for the accurate diagnosis and immunotherapy of AF.

INTRODUCTION

Atrial fibr1illation (AF) is one of the most common permanent arrhythmia types in the general population, and its pathological changes are characterized by electrical and structural remodeling of the left atrium

[1, 2]. According to relevant reports, AF affects approximately 1%-2% of the population worldwide, its prevalence is proportional to age, and the incidence rate of AF in people over 80 years old reaches 8% [3, 4]. AF can significantly increase the risk of stroke, myocardial infarction and heart failure, which brings a

heavy economic burden to the patient's family and the whole society [5]. Therefore, it is urgent to further clarify its pathogenesis and find more effective treatments. AF is considered to be a multifactorial and complex disease that is usually associated with factors such as age, obesity, hypertension, smoking, sex, diabetes and valvular heart disease [6]. However, the pathophysiology of AF has not been fully elucidated. In a recently published study, we found that the infiltration levels of activated mast cells and regulatory T cells (Tregs) were decreased and the infiltration levels of gamma delta T cells, resting mast cells and M2 macrophages were increased in AF patients compared to those in sinus rhythm (SR) individuals [7]. These results partially elucidate the infiltration of immune cells in AF and suggest that the immune mechanism plays a key role in AF. However, more studies are needed to further fully explore the mechanism of immune infiltration in AF and may help to reveal several new immunotherapies for AF.

Traditional epigenetic modification refers to the reversible modification of proteins (histones) and DNA, which can regulate gene expression without changing the genetic sequence [8]. RNA modification has gradually attracted attention; it is considered the third layer of epigenetics and involves regulation of RNA metabolism and processing [9]. Several RNA modification forms, including N1-methyladenosine (m1A), 5-methylcytosine (m5C), and N6-methyladenosine (m6A), have been found, among which the most common modification is m6A [10]. m6A modification is a homeostatic and reversible process in eukaryotic cells that is mainly regulated by a variety of m6A regulatory factors, including demethylases (erasers), binding proteins (readers), and methyltransferases (writers) [11]. Recent studies have shown that m6A modification may play a crucial role in the regulation of the immune response. Wang et al. noticed that the *HNRNPA2B1* regulator acts as a reader, can promote m6A modification and can trigger the innate immune response by recognizing viral DNA in the context of viral infection [12]. Han et al. found that the *YTHDF1* regulator acts as a reader and is involved in antigen presentation from dendritic cells to CD8+ T cells by enhancing lysosomal cathepsin translation and promoting tumor neoantigen cross-presentation and CD8+ T-cell cross-priming, thereby promoting the immune escape of tumor cells [13]. Moreover, Li et al. found that the homeostatic differentiation of T cells may be severely impaired due to the deletion of a writer, such as *METTL3*, in T cells [14]. However, no research has focused on the role of m6A in the immune microenvironment of AF. Therefore, we investigated the effect of m6A modification on the characteristics of the immune microenvironment in AF in depth and identified key immune-related genes associated

with AF. The implementation of these works will help us to deeply understand the pathogenesis of AF from a completely new perspective.

RESULTS

Data preprocessing

The analysis process of the whole study is shown in Figure 1. The normalized gene expression matrix of the GSE31821, GSE41177, GSE79768 and GSE115574 datasets was obtained by standardizing the data format (Supplementary Figure 1A) and adding missing values by filling in the average value of the variable in all other columns. As shown in Supplementary Figure 1B, principal component analysis (PCA) of the transcriptome profile showed a significant difference between AF samples and normal samples. The integrated expression profile, including 21652 different gene symbols, was obtained from 110 atrial tissue samples after data
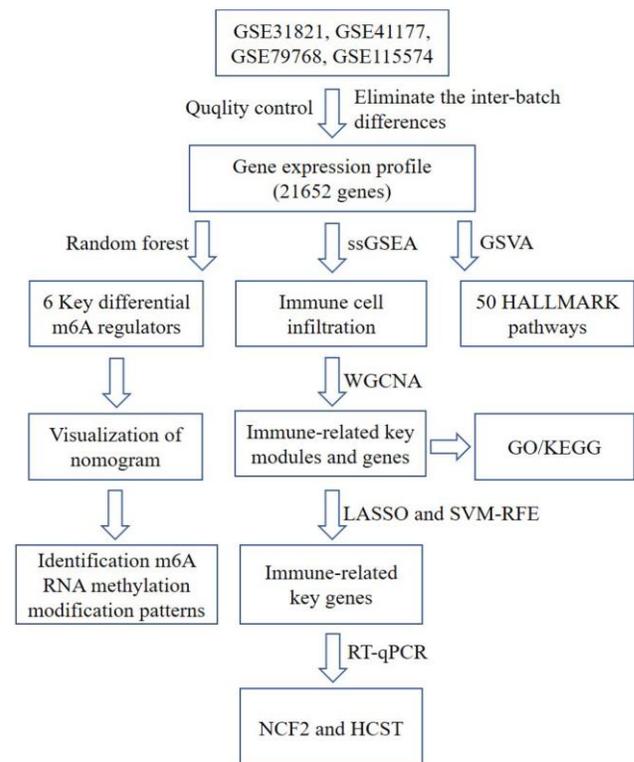

Figure 1. A flow chart of the analysis. ssGSEA, single-sample gene set enrichment analysis; GO, gene ontology annotation; KEGG, kyoto encyclopedia of genes and genomes pathway enrichment analyses; GSVA, gene set variation analysis; WGCNA, weighted gene co-expression network analysis; LASSO, Least Absolute Shrinkage and Selector Operation; SVM-RFE, Support Vector Machine-Recursive Feature Elimination; NCF2, neutrophil cytosolic factor 2; HCST: hematopoietic cell signal transducer.

merging and eliminating the interbatch differences between the GSE31821, GSE41177, GSE79768 and GSE115574 datasets (Supplementary Table 1).

Identification of differential m6A regulators

A total of 23 different m6A regulators, including 7 writers (*METTL3*, *METTL14*, *ZC3H13*, *RBM15*, *RBM15B*, *WTAP* and *CBLL1*), 14 readers (*YTHDC1*, *LRPPRC*, *HNRNPC*, *IGFBP1*, *YTHDC2*, *HNRNPA2B1*, *IGF2BP1*, *YTHDF1*, *FMR1*, *YTHDF3*, *IGFBP2*, *YTHDF2*, *IGFBP3* and *ELAVL1*) and 2 erasers (*ALKBH5* and *FTO*), were analyzed in the current research. As shown in the box plot (Figure 2A) and heatmap plot (Figure 2B), we noticed that the expression levels of *RBM15B*, *IGFBP2*, *IGFBP3* and *ALKBH5* were significantly increased, while the expression levels of *HNRNPC* and *HNRNPA2B1* were significantly decreased in AF samples compared with SR samples.

Random forest screening for key m6A regulators

Cyclic random forest classification was performed for all possible numbers in 1-20 variables, and the average error rate of the pattern mode was calculated. Referring to the relationship plot between the number of decision trees and the model error (Figure 2C), 300 trees were selected as the parameter of the final model, which indicates a stable error in the model. Subsequently, as shown in Figure 2D, six key m6A regulators (*RBM15B*, *IGFBP2*, *IGFBP3*, *ALKBH5*, *HNRNPC* and *HNRNPA2B1*) with importance greater than 2 were identified for subsequent analysis.

Construction and assessment of a nomogram model

As shown in Figure 3A, a predictive nomogram was constructed based on the expression of six key m6A regulators (*RBM15B*, *IGFBP2*, *IGFBP3*, *ALKBH5*,
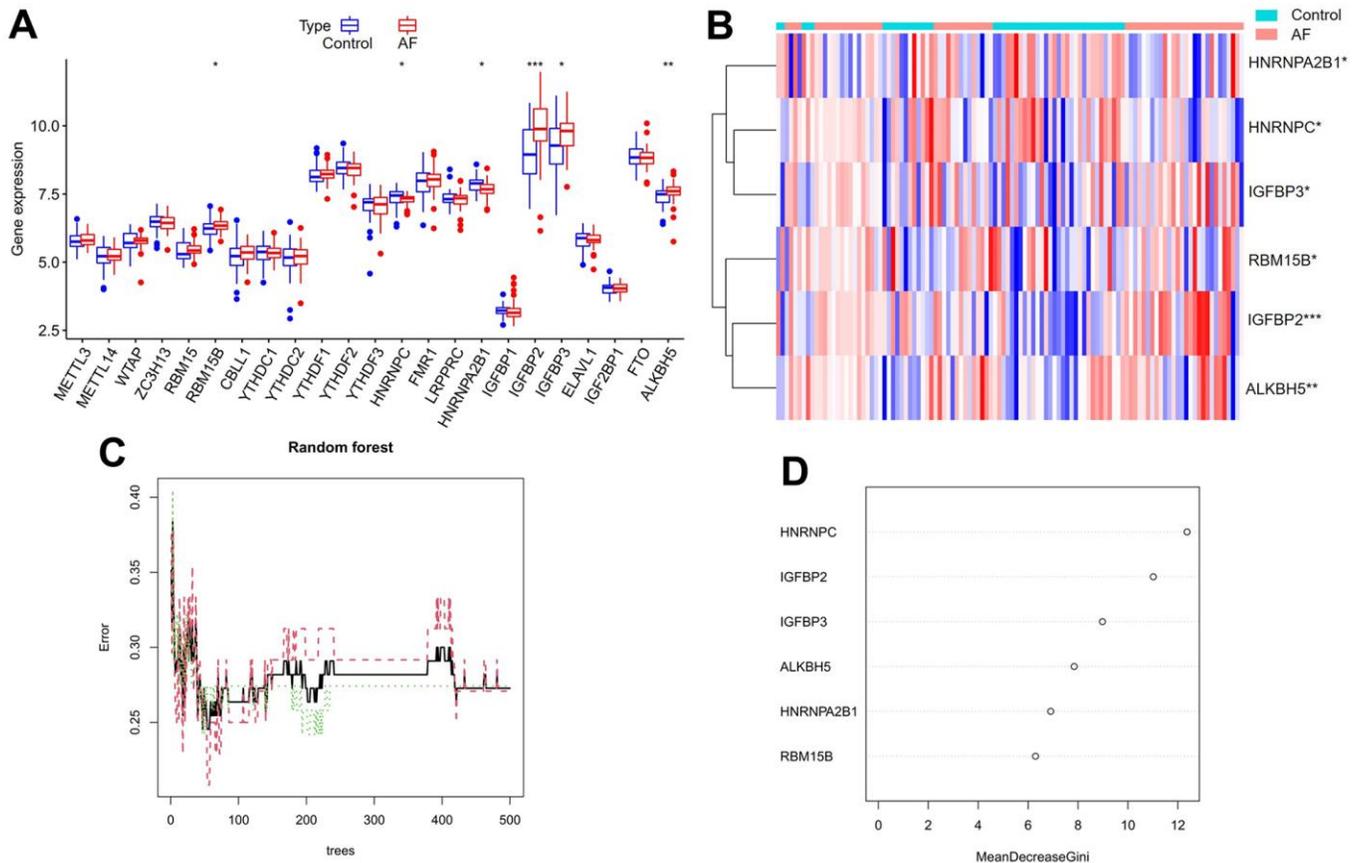

Figure 2. Expression landscape of m6A RNA methylation regulators in AF and random forest model construction to identify key m6A regulators. (**A**) Box plot of differentially expressed m6A regulators. (**B**) Heatmap of differentially expressed m6A regulators. (**C**) Plot of performance in log scale against epoch number. The x-axis represents the number of decision trees, and the y-axis indicates the error rate. When the number of decision trees is approximately 300, the error rate is relatively stable. (**D**) Results of the Gini coefficient method in the random forest classifier. The x-axis indicates the genetic variable, and the y-axis represents the importance index. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

HNRNPC and *HNRNPA2B1*). The calibration curve suggested that the error between the predicted and actual AF risk was very small, suggesting that the nomogram model obtained high accuracy in predicting AF (Figure 3B). Decision curve analysis (DCA) showed that the "nomogram" curve was higher than the gray line, indicating that the nomogram maintains great clinical utility in predicting the morbidity of AF patients (Figure 3C). ROC analysis reconfirmed that the model was effective in distinguishing AF patients from healthy subjects (Figure 3D).

Identification of m6A RNA methylation modification patterns in AF

Unsupervised consistent clustering analysis based on the expression values of six key m6A regulators in AF samples was utilized to study m6A modification patterns in AF (Figure 4A–4C). Three different subtypes of

AF were identified based on qualitatively different expression of six key m6A regulators, including 10 samples in the m6A cluster-A group, 22 samples in the m6A cluster-B group and 30 samples in the m6A cluster-C group (Figure 4D and Supplementary Table 2).

Comparison of immune microenvironment characteristics

The infiltration of many immunocytes was different between the controls and AF patients (Figure 5A), and among the AF samples with three distinct m6A modification patterns (Figure 5B). We found relatively higher infiltration of myeloid-derived suppressor cells (MDSCs), natural killer T cells, plasmacytoid dendritic cells and regulatory T cells in AF subjects than in SR patients. We also found that there was relatively higher infiltration of activated CD4 T cells, activated CD8 T cells, activated dendritic cells, CD56dim natural killer
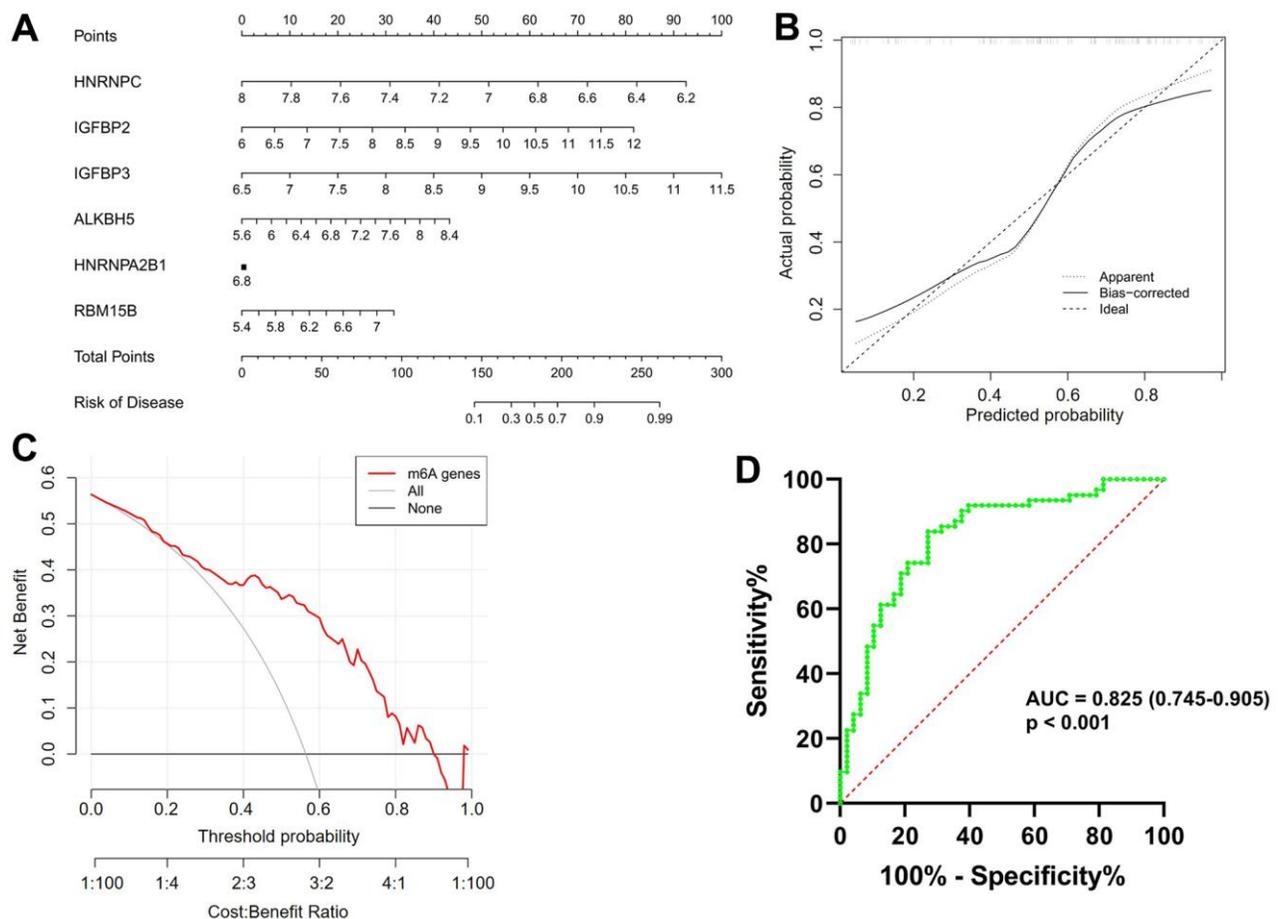

Figure 3. Construction and validation of a predictive nomogram of atrial fibrillation established based on six m6A regulators. (**A**) The nomogram of the model. (**B**) The calibration plot of the nomogram, and the diagonal dotted line represents a perfect prediction by an ideal model. (**C**) Decision curve analysis (DCA) of the nomogram. The solid line represents the performance of the nomogram, of which a closer fit to the diagonal dotted line represents a better prediction. (**D**) Receiver operating characteristic (ROC) analysis of the nomogram confirming that the model was effective in distinguishing atrial fibrillation patients from healthy subjects.

cells, immature dendritic cells, MDSCs, neutrophils, plasmacytoid dendritic cells and type-2 T helper cells in the m6A cluster-C group than in the m6A cluster-A group (Supplementary Table 3).

Comparison of 50 HALLMARKS pathways

Gene set variation analysis (GSVA) was used to compare the HALLMARKS pathways between the healthy subject and AF patient samples (Figure 6A) and among the AF patient samples with three distinct m6A modification patterns (Figure 6B). Compared with the SR subjects, the AF patients had more enriched pathways, such as KRAS signaling up, IL-2/STAT5 signaling, angiogenesis, UV response down, glycolysis,

epithelial mesenchymal transition, mTORC1 signaling, PI3K/Akt/mTOR signaling and DNA repair. Meanwhile, compared with the m6A cluster-A group, the m6A cluster-C group had more enriched pathways, such as allograft rejection, peroxisome, coagulation, UV response up, P53 signaling, fatty acid metabolism, xenobiotic metabolism, epithelial mesenchymal transition, mTORC1 signaling, PI3K/Akt/mTOR signaling, apical junction, myogenesis, adipogenesis, cholesterol homeostasis, and hypoxia (Supplementary Table 4).

Identification of meaningful immune-related modules

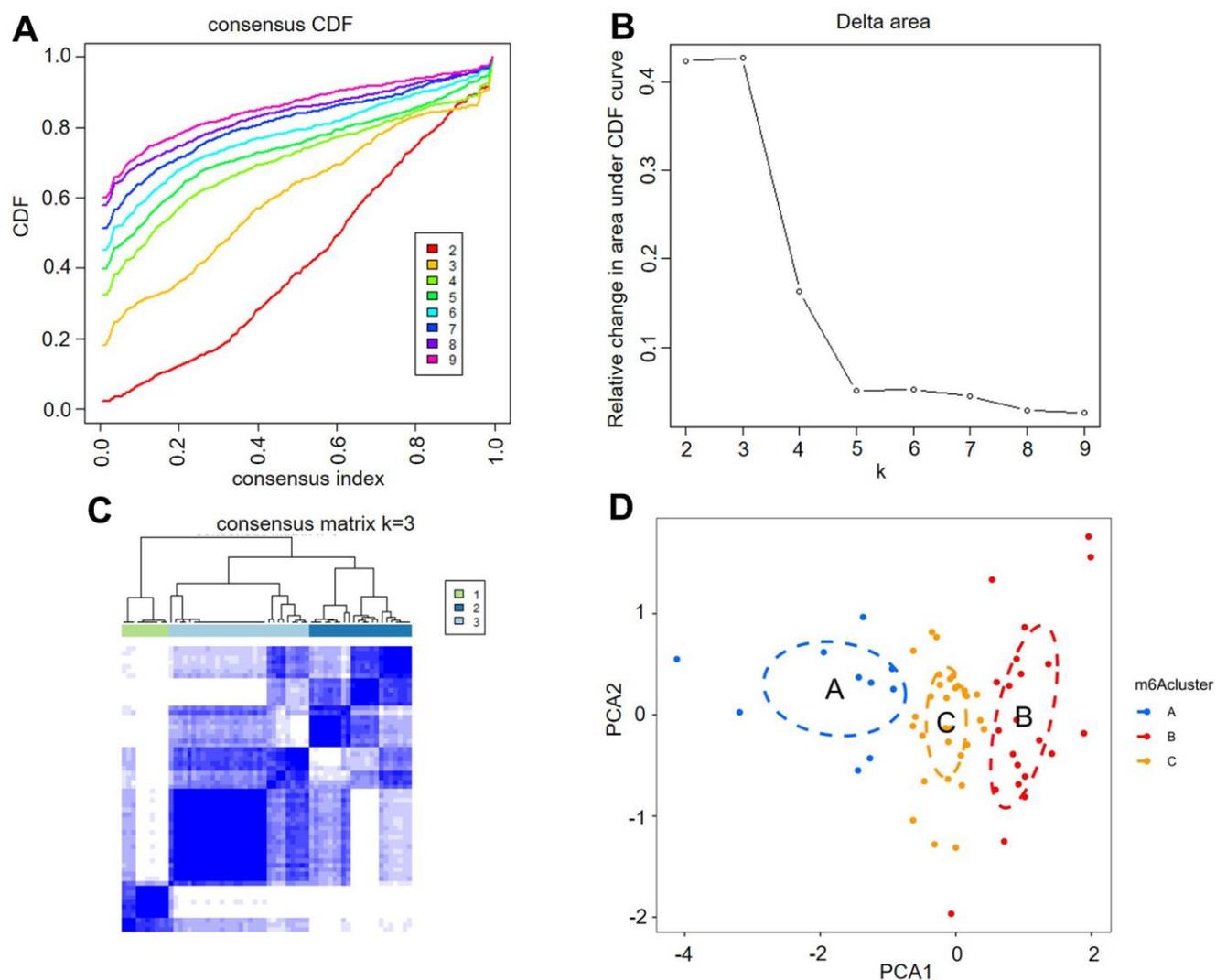Weighted gene coexpression network analysis (WGCNA) was used to identify the meaningful modules that were

Figure 4. Identification of three distinct m6A modification pattern subtypes in atrial fibrillation. (**A**) Consensus clustering cumulative distribution function (CDF) for k = 2-9. (**B**) Relative change in the area under the CDF curve for k = 2-9. (**C**) Heatmap of the matrix of cooccurrence proportions for atrial fibrillation samples. (**D**) Principal component analysis for the transcriptome profiles of three m6A clusters, showing a remarkable difference in the transcriptome between different modification patterns.

significantly associated with infiltrating immune cells. In the WGCNA, when the correlation coefficient was greater than 0.9, the corresponding soft threshold was 14. Therefore, a soft threshold of 14 was selected to construct several gene modules (Figure 7A). Then, the topological overlap matrix combined with the hierarchical mean linkage clustering method was used to identify gene modules in each gene network. As shown
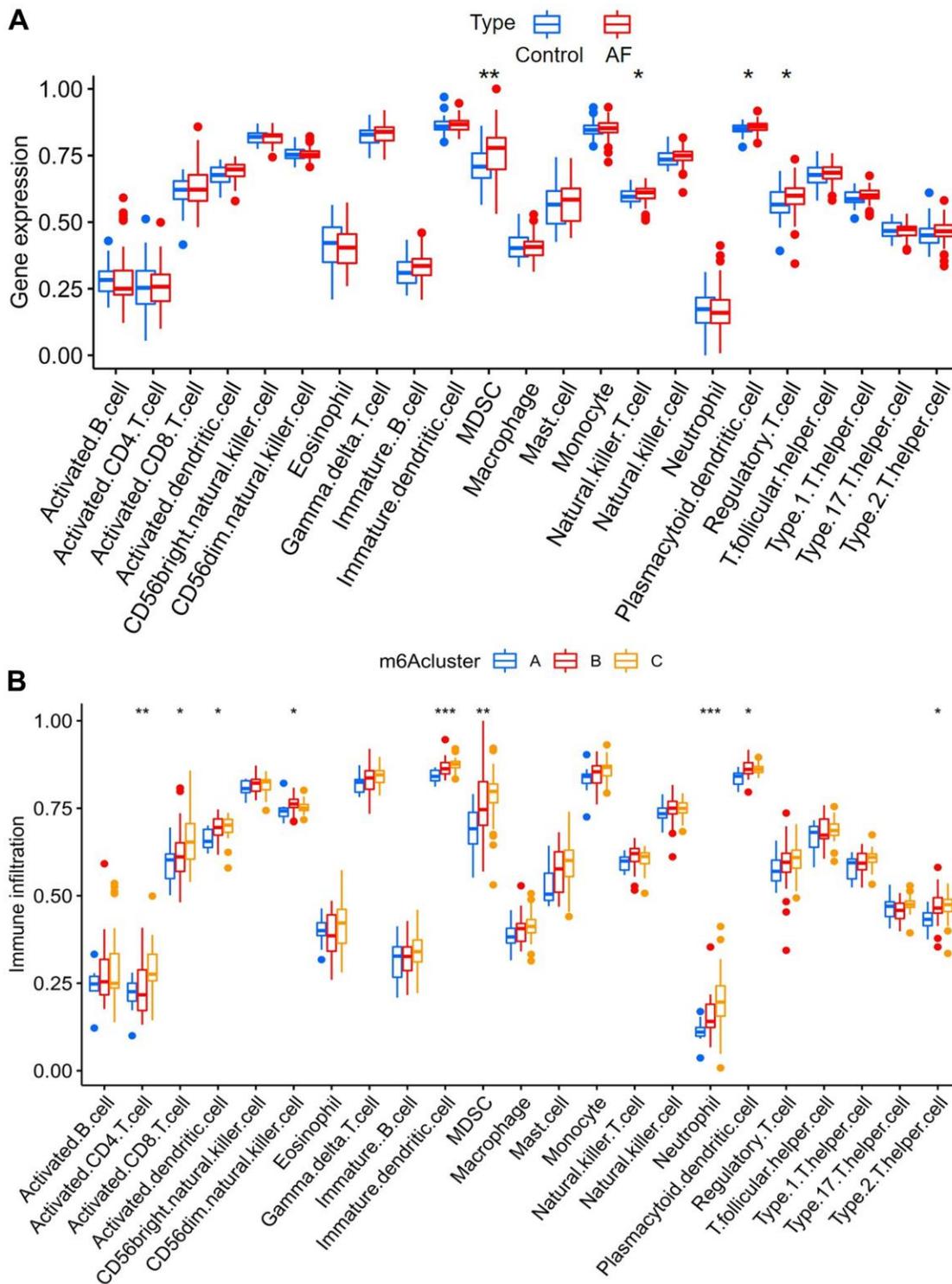

Figure 5. Analysis of immune cell infiltration between different groups. (**A**) Comparison of immunocyte abundance between controls and atrial fibrillation patients. (**B**) Comparison of immunocyte abundance in the 3 clusters.

in Figure 7B, a total of 15 gene modules were identified in the coexpression network. Figure 7C depicts the correlation heatmap between 15 modules. As shown in Figure 7D, the lightcyan module was highly associated with regulatory T cells ($r^2 = 0.90$, $p = 6e-40$), and the lightgreen module was highly associated with activated CD8 T cells ($r^2 = 0.85$, $p = 4e-31$).

Enrichment analysis of the genes in meaningful modules

As shown in Supplementary Table 5, a total of 104 genes in the lightcyan and lightgreen modules were included in the KEGG and GO functional enrichment analyses. The top 10 biological processes, cellular components, and molecular functions are shown in Supplementary Figure 2A. The top 10 KEGG pathways are shown in Supplementary Figure 2B. A total of 104 genes were mainly enriched in the following immune-related biological processes: leukocyte-mediated immunity, positive regulation of leukocyte activation, positive regulation of cell activation, adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains, leukocyte cell–cell adhesion, activation of immune response, leukocyte activation involved in the immune response and cell activation involved in the immune response. In addition, the details of GO and KEGG analyses are presented in Supplementary Tables 6, 7.

Identification of key genes by machine learning

The correlation coefficient between the color module and the gene significance was calculated to determine the significance of the module. As shown in Figure 8, we noticed that the correlation coefficients of gene significance with the lightgreen module (Figure 8A) and the lightcyan module (Figure 8B) were 0.70 ($p = 7.1e-07$) and 0.76 ($p = 2.1e-13$), respectively. Then, LASSO regression and the SVM-RFE algorithm were performed to identify several characteristic genes that were significantly associated with AF based on the expression profile of genes in the lightcyan and lightgreen modules.
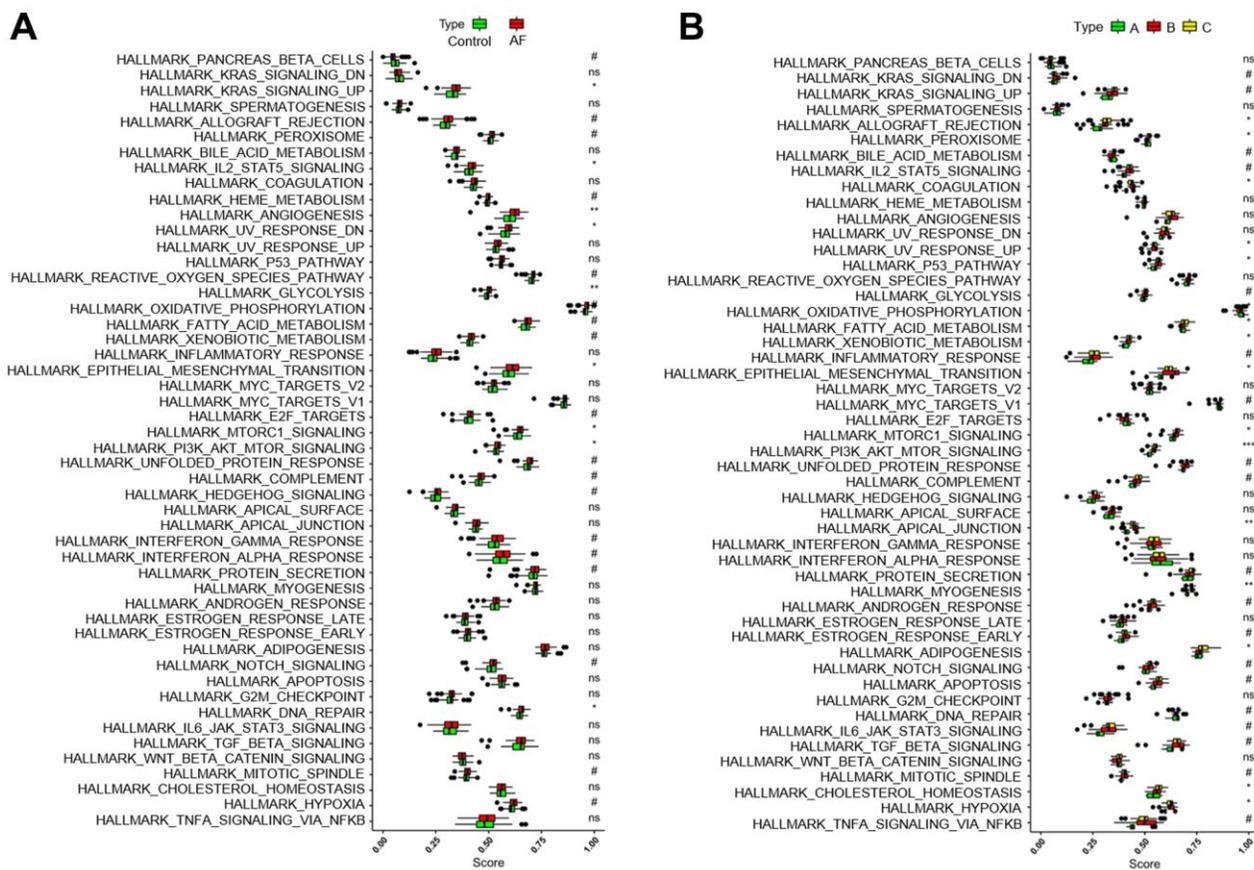

Figure 6. Analysis of HALLMARKS pathway enrichment scores between different groups. (A) Comparison of HALLMARKS pathway enrichment scores between controls and atrial fibrillation patients. (B) Comparison of HALLMARKS pathway enrichment scores in the 3 m6A clusters.

A total of 29 and 28 key genes were identified by LASSO regression (Figure 8C) and the SVM-RFE algorithm (Figure 8D), respectively. Moreover, a total of 16 overlapping genes were identified by both machine learning methods. The detailed gene symbols of several key genes identified by these two machine learning methods are also shown in Supplementary Table 8.

Expression of key genes

As shown in Figure 9A, compared with the SR subjects, the expression levels of neutrophil cytosolic factor 2 (*NCF2*), lysosomal protein transmembrane 5 (*LAPTM5*), hematopoietic cell signal transducer (*HCST*), hematopoietic cell-specific lyn substrate 1 (*HCLS1*), C-X-C motif chemokine ligand 12 (*CXCL12*), coronin 1A (*CORO1A*), complement C1q C chain (*C1QC*) and junction adhesion molecule like (*AMICA1*) were significantly increased in AF patients. Meanwhile,

compared with the m6A cluster-A group, the expression of *NCF2* and *HCST* was significantly increased in the m6A cluster-B and -C groups (Figure 9B). Since the *NCF2* and *HCST* genes differ between SR subjects and AF patients and among different m6A-modified molecular subtypes, *NCF2* and *HCST* were further verified by RT–qPCR. As shown in Figure 10A, we noticed that the expression levels of *NCF2* and *HCST* were significantly increased in AF patients compared with controls. In addition, the AUC values for *NCF2* (Figure 10B) and *HCST* (Figure 10C) were 0.841 (95% CI 0.782–0.900; *p* < 0.001) and 0.862 (95% CI 0.809–0.914; *p* < 0.001), respectively.

Correlation between hub genes and infiltrating immune cells and 50 HALLMARKS pathways

As shown in Supplementary Figure 3A, the *NCF2* and *HCST* genes were significantly positively correlated with activated CD4 T cells, activated CD8 T cells,

Figure 7. Weighted gene coexpression network analysis. (**A**) Analysis of network topology for various soft-thresholding powers. (**B**) Clustering dendrogram of genes. (**C**) Relationship among all the modules. (**D**) Associations between modules and infiltrating immune cells.

activated dendritic cells, CD56 bright natural killer cells, gamma delta T cells, immature B cells, immature dendritic cells, MDSCs, macrophages, mast cells, monocytes, natural killer T cells, natural killer cells, neutrophils, plasmacytoid dendritic cells, regulatory T cells, T follicular helper cells, type 1 T helper cells, and type 2 T helper cells. As shown in Supplementary Figure 3B, the *NCF2* and *HCST* genes were also significantly positively correlated with several HALLMARKS pathways, such as TNF-α signaling via NF-kB, xenobiotic metabolism, beta-catenin signaling, PI3K/Akt/mTOR signaling, the P53 pathway, mTORC1 signaling, KRAS signaling up, the interferon gamma response, the interferon alpha response, the inflammatory response, IL-2/STAT5 signaling, IL-6/JAK/STAT3 signaling, the late estrogen response, the early estrogen response, epithelial mesenchymal transition, complement, coagulation, apoptosis, apical

surface, apical junction, angiogenesis, the androgen response, allograft rejection and adipogenesis.

DISCUSSION

At present, AF is the most common persistent arrhythmia that significantly threatens the health of Chinese residents. Accumulating evidence suggests that immune and inflammatory mechanisms play a key role in the pathogenesis of AF, and these mechanisms appear to be significantly associated with the onset and persistence of AF as well as the prethrombotic state associated with AF [15, 16]. For example, Yamashita et al. demonstrated that macrophage adhesion and recruitment in the cardiac endocardium promoted inflammatory responses in human AF [17]. Hohmann et al. demonstrated that the number of inflammatory CD3+ T cells in the left atrial appendage was significantly increased in patients with
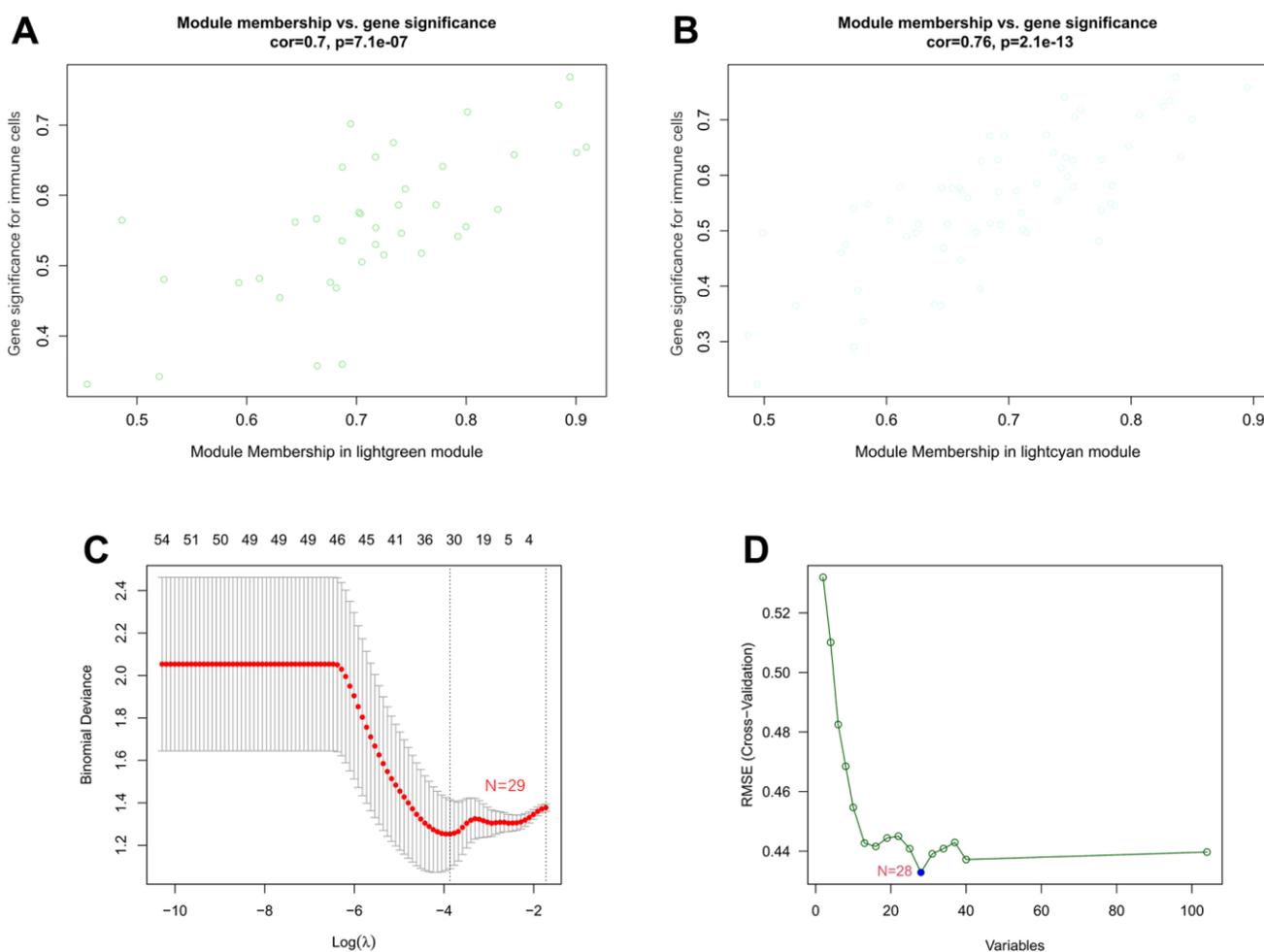

Figure 8. Associations between gene significance and module membership and identification of key genes of AF by machine learning. (**A**) Association between gene significance and module membership in the lightgreen module. (**B**) Association between gene significance and module membership in the lightcyan module. (**C**) The key genes identified by LASSO regression. (**D**) The key genes identified by the SVM-RFE algorithm.
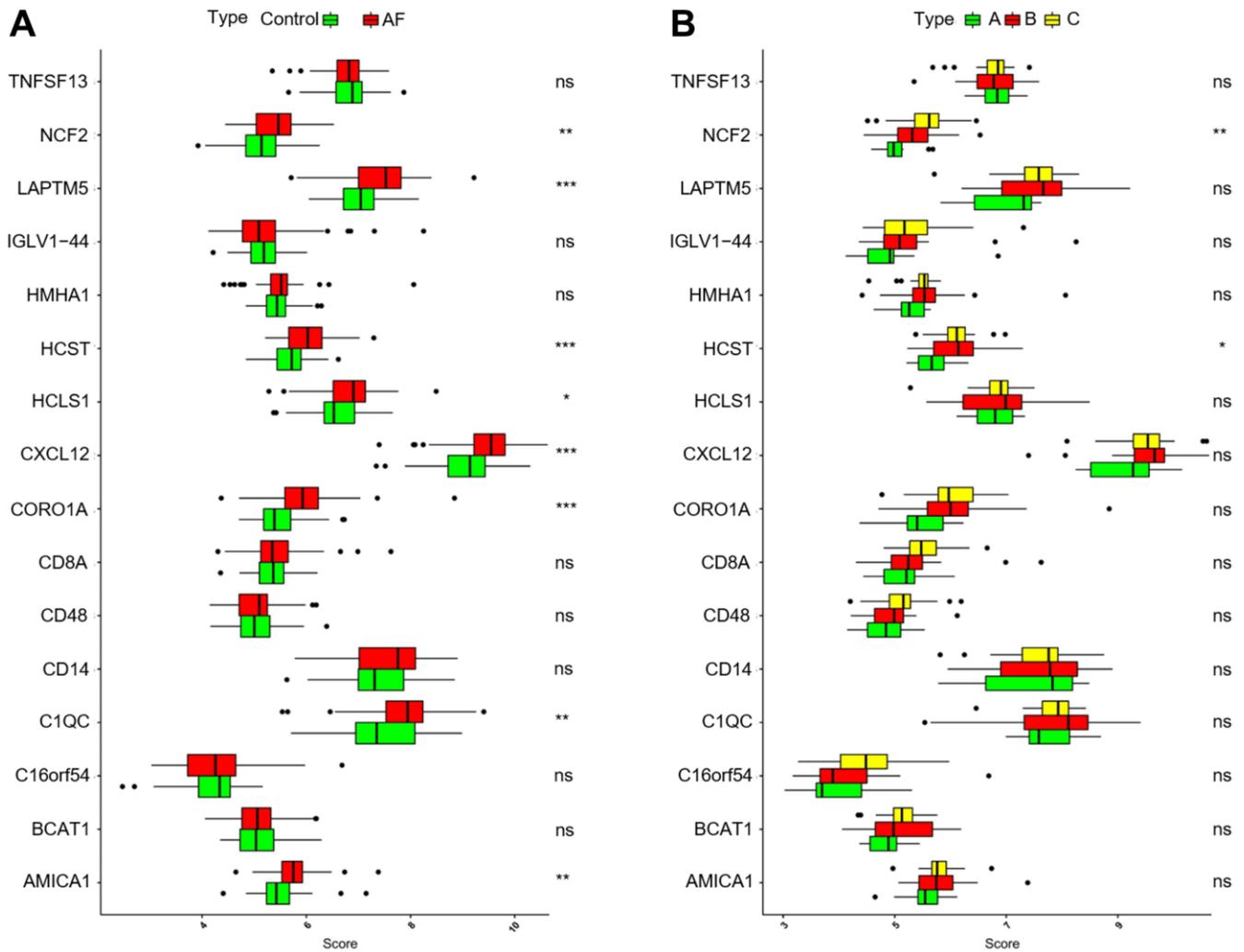
Figure 9. Analysis of the differences in expression of 16 key genes in different groups. (**A**) Comparison of the expression of 16 key genes between controls and atrial fibrillation patients. (**B**) Comparison of the expression of 16 key genes in the 3 m6A clusters. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
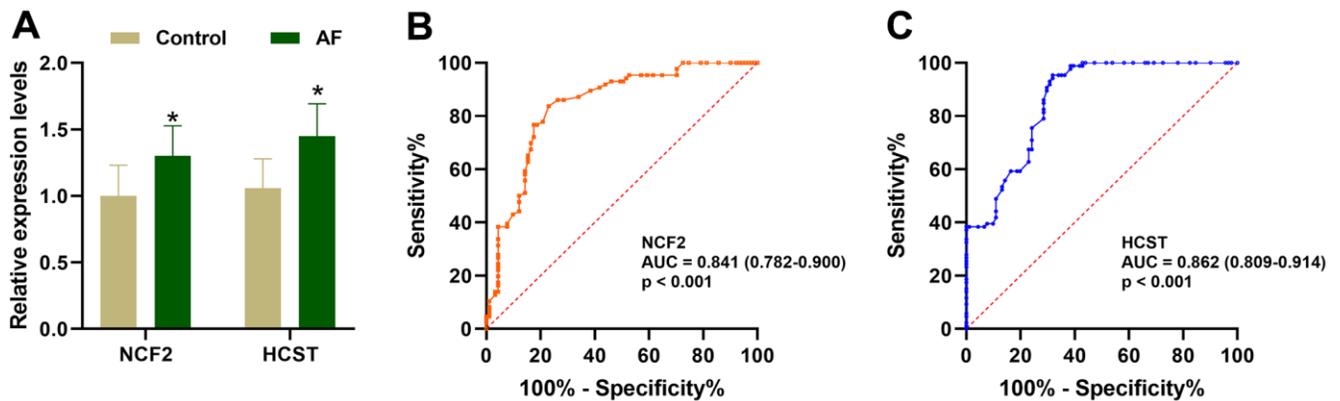

Figure 10. Validation of the *NCF2* and *HCST* genes in clinical samples. (**A**) The relative expression levels of *NCF2* and *HCST* in clinical samples. (**B**) ROC curve analysis of *NCF2*. (**C**) ROC curve analysis of *HCST*. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

SR compared to those with persistent AF [18]. On the other hand, some scholars have found that m6A modification plays an integral role in innate and adaptive immune responses [19]. Several studies have aimed to explore the role of m6A modification in immunity, especially in the infiltration of immune cells in the tumor microenvironment, and the results confirmed that m6A modification plays a fundamental role in tumor immunity [20, 21]. We also believe that the m6A modification pattern plays a similar role in the formation of the immune microenvironment in AF. However, its mechanism has not been clarified. To better clarify these issues, we systematically explored the effects of m6A modification on the infiltration of immune cells and the expression of immune-related genes in AF.

In this study, the effects of m6A modification on immune cell infiltration, inflammation, immune-related pathways, and key immune-related genes in AF were explored, and several meaningful new findings were obtained. First, we found that the expression levels of *RBM15B*, *IGFBP2*, *IGFBP3* and *ALKBH5* were significantly increased, while the expression levels of *HNRNPC* and *HNRNPA2B1* were significantly decreased in AF samples compared with SR samples, suggesting that these six key m6A regulators are involved in the development of AF. Meanwhile, a nomogram model based on the above 6 key m6A regulators suggested a diagnostic value of up to 82.5% to distinguish AF from healthy participants. This meaningful model will help us to screen high-risk individuals susceptible to atrial fibrillation at an early stage in clinical practice, and these m6A regulators may be potential novel molecular targets to help diagnose or treat AF. Second, we identified three distinct m6A modification patterns (m6A cluster-A, -B and -C) using unsupervised clustering of AF samples based on the expression values of six key m6A regulators. We found that there was relatively higher infiltration of activated CD4 T cells, activated CD8 T cells, activated dendritic cells, CD56dim natural killer cells, immature dendritic cells, MDSCs, neutrophils, plasmacytoid dendritic cells and type-2 T helper cells in the m6A cluster-C group than in the m6A cluster-A group, suggesting the essential role of m6A modification in the regulation of the immune microenvironment in AF. Presently, the classification strategy of immune subtypes is widely used in the cancer field, and identifying new immune subtypes can help to formulate better treatment plans for hepatocellular carcinoma [22]. Thus, this classification strategy of immune subtypes can also help us subtype AF samples at the immune level or molecular level and help us to implement individualized immunotherapy for patients with active immune activity.

In recent years, some key molecules related to immune cells have been discovered and are considered new biomarkers to help diagnose disease or predict prognosis [23, 24]. Li et al. found that *PPBP*, *CXCL12* and *CCL4*, as novel molecular markers for the diagnosis of valvular atrial fibrillation, were positively correlated with the infiltration of various immune cells, such as neutrophils, plasma cells and resting dendritic cells [25]. Liu et al. also noticed that *BECN1*, *ATG7*, *BCL2L1* and *MYC*, as novel molecular markers for the diagnosis of valvular atrial fibrillation, were positively correlated with the infiltration of neutrophils and negatively correlated with the infiltration of memory resting CD4 T cells and T follicular helper cells. Meanwhile, they found that the fractions of memory resting CD4 T cells and T follicular helper cells were decreased, and the fractions of resting dendritic cells, plasma cells, neutrophils, and monocytes were significantly increased in AF patients compared with SR subjects [26]. These results partially elucidated the pattern of immune cell infiltration in AF and identified several new molecular markers associated with the pathogenesis of AF. However, the identification of these key genes is based on differential expression analysis of genes. Currently, WGCNA, as a more effective bioinformatics analysis method, is expected to identify key genes ignored by differential gene analysis by constructing a scale-free gene coexpression network and identifying gene modules that are significantly related to phenotypes [27]. Meanwhile, machine learning is often used to improve the accuracy and prediction of several key genes identified based on traditional microarray or next-generation sequencing data [28]. Recently, the SVM-RFE algorithm and LASSO regression analysis have been the most widely used machine learning methods to identify key genes [29]. Thus, in the current research, WGCNA combined with machine learning methods was used to identify key immune-related genes involved in the pathogenesis of AF. We noticed that the lightcyan and lightgreen modules were significantly associated with infiltrating immune cells, such as regulatory T cells, follicular helper T cells, MDSCs and activated CD8 T cells. Then, a total of 16 overlapping genes were identified by both the SVM-RFE algorithm and LASSO regression analysis. Moreover, we also noticed that the expression levels of *NCF2* and *HCST* were significantly different between control and AF patient samples and among AF patient samples with three distinct m6A modification patterns. Meanwhile, RT–qPCR analysis also proved that the expression of *NCF2* and *HCST* was significantly increased in AF patients compared with SR subjects. These results consistently indicate that *NCF2* and *HCST* are significantly related to the pathogenesis of AF, and elucidating the molecular mechanism of their involvement in AF is expected to provide new molecular targets for the prevention and treatment of AF.

Currently, the *NCF2* gene and its genetic variants have been reported to be associated with susceptibility to a variety of inflammation- or immune-related diseases. Cunninghame et al. found that rs10911363, an intronic variant in the *NCF2* gene, was associated with systemic lupus erythematosus (SLE) susceptibility in a European population [30]. Jiao et al. proved that the rs10911362 variant in the *NCF2* gene was correlated with decreased tuberculosis and pulmonary tuberculosis susceptibility in a Chinese population [31]. In addition, Li et al. suggested that the *NCF2* gene is expected to become a molecular marker for the diagnosis of pulmonary tuberculosis, and its expression was significantly increased in patients with pulmonary tuberculosis [32]. Su et al. found that the *NCF2* gene was significantly overexpressed in advanced atherosclerotic plaques and may play a key role in the development of psoriasis complicated with atherosclerosis [33]. Moreover, some scholars have found that *NCF2* may be a key molecule in patients with nonalcoholic fatty liver disease complicated with AF [34], and the expression of *NCF2* is significantly upregulated in patients with AF [35]. On the other hand, Zhou et al. found that overexpression of immune-related genes, such as *HCST,* was significantly correlated with high infiltration of immune cells, especially dendritic cells, in clear cell renal cell carcinoma (ccRCC), and high expression of *HCST* was significantly correlated with poorer prognosis in ccRCC patients [36]. Wang et al. revealed that *HCST* is significantly associated with some immune cells, such as CD8 T cells, B cells, macrophages, CD4 T cells, dendritic cells and neutrophils, and is involved in several inflammatory or immune-related signaling pathways, including the T-cell receptor signaling pathway, cytokine–cytokine receptor signaling pathway, chemokine signaling pathway, pathways involving cell adhesion molecules, FC gamma-mediated phagocytosis and B-cell receptor signaling pathway, and that overexpression of *HCST* can significantly affect the clinical stage, tumor grade and prognosis of kidney renal clear cell carcinoma (KIRC) [37]. However, the correlation between *NCF2*, *HCST* and immune cells and whether the expression levels of *NCF2* and *HCST* are affected by the m6A modification pattern in AF remain unclear. In the current research, we found that *NCF2* and/or *HCST* were positively correlated with multiple immune cells, excluding CD56dim natural killer cells, and their expression levels were significantly upregulated in the m6A cluster-B and -C groups compared with the m6A cluster-A group. Meanwhile, a positive correlation between *NCF2*, *HCST* and several inflammation- or immune-related HALLMARKS pathways was also observed. These results indicate that *NCF2* and *HCST* are key immune regulatory molecules involved in AF.

The current research had several limitations. First, the validation sample size in this study was small, and the

subjects were recruited from a single center, so these findings still need to be validated in other multicenter studies or studies with larger sample sizes. Second, the immune cell analysis in this study adopts the most widely used analysis method to quantify the number of immune cells, but single-cell sequencing is still required to obtain the most accurate number of immune cells. Third, we were unable to obtain more clinical features or serological results of AF samples in the GSE31821, GSE41177, GSE79768 and GSE115574 datasets. Therefore, it is difficult to reveal the key role of m6A modification in immune regulation from multiple perspectives, and the analysis results are relatively singular. Fourth, most of the findings in this study are based on bioinformatics analysis. Although RT–qPCR was used to verify the expression of *NCF2* and *HCST* in clinical samples, the relationship between *NCF2*, *HCST* and immune cells and immune-related pathways still needs to be verified by more *in vitro* and *in vivo* experiments.

In conclusion, m6A modification plays a key role in the complexity and diversity of the immune microenvironment of AF. Immunotyping of patients with AF will help to develop more accurate immunotherapy strategies for patients with a significant immune response. The m6A modification can affect the expression of the immune-related genes *NCF2* and *HCST*, and *NCF2* and *HCST* are expected to become new targets for immunotherapy of AF.

MATERIALS AND METHODS

AF microarray datasets

GSE31821, including 6 atrial tissue samples (2 SR and 4 AF samples), GSE41177, including 19 atrial tissue samples (3 SR and 16 AF samples), GSE79768, including 26 atrial tissue samples (12 SR and 14 AF samples), and GSE115574, including 59 atrial tissue samples (31 SR and 28 AF samples), were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). These datasets were based on the GPL570 Affymetrix Human Genome U133 Plus 2.0 array. An integrated gene expression matrix was obtained after normalization and elimination of interbatch differences between GSE31821, GSE41177, GSE79768 and GSE115574. The gene expression profiles of GSE31821, GSE41177, GSE79768 and GSE115574 were normalized by the "*limma*" package [38]. When a probe corresponded to multiple genes at the same time, it was excluded from the analysis. When multiple different probes corresponded to the same gene, we took the average gene expression value detected by those probes as the true expression value of the gene. The interbatch differences between the GSE31821,

GSE41177, GSE79768 and GSE115574 datasets were eliminated by the ComBat function in the "sva" package in R software.

Identification of key m6A regulators

The Wilcoxon test was used to evaluate the expression status differences of 23 m6A regulators between healthy individuals and AF patients. Then, a random forest model was constructed by the "randomForest" package in R and used to identify key m6A regulatory factors. Specifically, the average model error rate of all m6A regulatory factors was calculated, the optimal number of variables of the binary tree in the node was set as 6, and 300 was selected as the optimal number of trees contained in the random forest. Then, the random forest model was constructed, and the decreasing precision method (Gini coefficient method) was used to obtain the dimension importance value from the random forest model. Factors with importance values greater than 2 [39, 40] were selected as key m6A regulators for subsequent model construction.

Construction and verification of the nomogram

The predictive nomogram was constructed by the "rms" package in R software based on the expression values of six key m6A regulators. Then, the calibration curve was used to assess the predictive power of the nomogram model. Decision curve analysis was used to evaluate the clinical value of the nomogram model. Finally, receiver operating characteristic (ROC) analysis was used to evaluate the diagnostic performance of the nomogram model in distinguishing AF patients from healthy subjects.

Identification of the m6A modification pattern

Based on the expression of six key m6A regulators, unsupervised clustering analysis was used to identify different m6A modification patterns in AF. The robustness and cluster numbers were calculated by the consensus clustering algorithm [41, 42]. The R package "ConsensuClusterPlus" was utilized to perform the above steps for 1000 iterations to guarantee the robustness of the classification [43]. Principal component analysis (PCA) was used to further verify the different m6A modification patterns distinguished by six key m6A regulators.

Single-sample gene set enrichment analysis (ssGSEA) and GSVA enrichment analysis

Single-sample gene-set enrichment analysis (ssGSEA) was used to estimate the number of specific infiltrating immune cells between controls and AF patients as well

as among distinct m6A modification patterns, which defines an enrichment score to represent the degree of absolute enrichment of a gene set in each sample within a given dataset [44]. The list of infiltrating immunocyte gene sets was obtained from a previous study [41]. The Wilcoxon test was utilized to compare enrichment scores representing immunocyte abundance between different m6A modification patterns.

GSVA enrichment analysis was used to evaluate the 50 HALLMARK pathways between control and AF patient samples as well as among AF patient samples with three distinct m6A modification patterns via the 'GSVA' package in R software. The gene sets of 'h.all.v7.0. symbols' were extracted from the MSigDB database [45] (http://software.broadinstitute.org/gsea/msigdb/index.jsp) to run GSVA. Then, the correlations between hub genes and infiltrating immune cells and 50 HALLMARK pathways were determined by Spearman correlation analysis.

WGCNA analysis

A scale-free network based on a gene expression matrix was constructed using WGCNA, which is one of the most commonly used tools in systems biology [46]. Genes with the top 25% of variance were selected for the WGCNA. In the current research, the appropriate soft threshold was defined as 14, and detailed WGCNA steps refer to the method described in our recently published article [47].

Enrichment analysis of genes in meaningful modules

Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) functional enrichment analyses based on the expression profile of genes in the meaningful modules were performed by clusterProfer and the DOSE package [48]. The threshold was defined as FDR < 0.05.

Identification of key genes using machine learning

At present, the least absolute shrinkage and selector operation (LASSO) and support vector machine-recursive feature elimination (SVM-RFE) algorithms are the two most commonly used machine learning methods for identifying key genes with the best prognostic value for disease. Based on the expression profile of genes in the meaningful modules, the "glmnet" package in R was used to perform LASSO logistic regression analysis [49]. In addition, SVM-RFE acts as an effective feature selection technique that finds the best variables by deleting the feature vector generated by SVM [50]. The selected biomarkers in the diagnosis of AF were analyzed and classified by the

SVM classifier based on the SVM function in the e1071 package. The overlapping key genes identified by the above machine learning methods were defined as hub genes. The differential expression of hub genes between controls and AF patients as well as among AF samples with several distinct m6A modification patterns was analyzed by t test and analysis of variance (ANOVA), respectively.

Study population

A total of 91 SR subjects and 86 persistent AF patients were recruited from the Cardiology Department of Hunan Provincial People's Hospital. AF lasting more than 7 days was defined as persistent AF [51]. Patients with a history of type 1 diabetes, hematologic disease, hypertension, coronary heart disease, neoplasia, autoimmune disease, and renal or liver diseases were excluded.

RT–qPCR

According to the manufacturer's instructions, total RNA was extracted from peripheral blood using the UNlQ-10 Column TRIzol Total RNA Isolation Kit (Sangon Biotech, Shanghai, China). The purity and concentration of the extracted RNA was checked by a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), with an A260/A280 between 1.8 and 2.0. The cDNA was reverse-transcribed using the PrimeScriptTM RT Reagent Kit (Takara, Otsu, Japan). Using GAPDH as a reference, we performed quantitative RT–PCR on an ABI 7500 instrument (Applied Biosystems, Waltham, MA, USA) using a Taq PCR Master Mix Kit (Takara, Otsu, Japan). Primer sequences for hub genes and references were designed and validated by Songon Biotech (Songon Biotech, Shanghai, China). The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression level of hub genes. The t test was used to compare the expression of hub genes between controls and AF patients. The ROC curves were constructed based on the expression levels of the hub genes.

Abbreviations

m6A: RNA N6-methyladenosine; AF: atrial fibrillation; WGCNA: weighted gene co-expression network analysis; Tregs: regulatory T cells; SR: sinus rhythm; PCA: principal component analysis; m5C: 5-methylcytosine; m1A: N1-methyladenosine; ssGSEA: single-sample gene-set enrichment analysis; GSVA: gene set variation analysis; DCA: decision curve analysis; ROC: receiver operating characteristic; MDSC: myeloid-derived suppressor cell; GEO: Gene Expression Omnibus; LASSO: Least Absolute Shrinkage and Selector Operation; SVM-RFE: SVM-RFE; NCF2:

neutrophil cytosolic factor 2; LAPTM5: lysosomal protein transmembrane 5; *HCST*: hematopoietic cell signal transducer; HCLS1: hematopoietic cell-specific Lyn substrate 1; CXCL12: C-X-C motif chemokine ligand 12; CORO1A: coronin 1A; C1QC: complement C1q C chain; AMICA1: junction adhesion molecule like; ANOVA: analysis of variance; SLE: systemic lupus erythematosus; HCST: hematopoietic cell signal transducer; ccRCC: clear cell renal cell carcinoma; KIRC: kidney renal clear cell carcinoma; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: gene ontology; MMP: matrix metalloproteinase; Treg: regulatory T; Th17: type-17 T helper; IL: interleukin; MHC: major histocompatibility complex.

AUTHOR CONTRIBUTIONS

Z.P.-F. and Z.S.-Y. conceived the study, participated in the design, performed the statistical analyses, and drafted the manuscript. P.J.-Q. and P.H.-W. conceived the study, participated in the design and helped to draft the manuscript. Z.C.Q., Z.Z.-F. and L.Z.Y. performed the bioinformatics analysis. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

ETHICAL STATEMENT AND CONSENT

Study protocols were developed based on the Ethics Committee of Hunan Provincial People's Hospital (No: KY 2018-141-03) and the 2008 revision of the Declaration of Helsinki of 1975 (http://www.wma.net/en/30publications/10policies/b3/). All subjects provided written and informed consent.

Editorial note

REFERENCES

1. Zhang YF, Meng LB, Hao ML, Yang JF, Zou T. Identification of Co-expressed Genes Between Atrial Fibrillation and Stroke. Front Neurol. 2020; 11:184.
https://doi.org/10.3389/fneur.2020.00184
PMID:32265825

2. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, Castella M, Diener HC, Heidbuchel H, Hendriks J, Hindricks G, Manolis AS, Oldgren J, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Europace. 2016; 18:1609–78.
https://doi.org/10.1093/europace/euw295
PMID:27567465

3. Andrade J, Khairy P, Dobrev D, Nattel S. The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. Circ Res. 2014; 114:1453–68.
https://doi.org/10.1161/CIRCRESAHA.114.303211
PMID:24763464

4. Lozano-Velasco E, Franco D, Aranega A, Daimi H. Genetics and Epigenetics of Atrial Fibrillation. Int J Mol Sci. 2020; 21:5717.
https://doi.org/10.3390/ijms21165717
PMID:32784971

5. Panchal G, Mahmood M, Lip GYH. Revisiting the risks of incident atrial fibrillation: a narrative review. Part 1. Kardiol Pol. 2019; 77:430–6.
https://doi.org/10.33963/KP.14806 PMID:31025653

6. Nguyen TN, Hilmer SN, Cumming RG. Review of epidemiology and management of atrial fibrillation in developing countries. Int J Cardiol. 2013; 167:2412–20.
https://doi.org/10.1016/j.ijcard.2013.01.184
PMID:23453870

7. Zheng PF, Chen LZ, Liu P, Liu ZY, Pan HW. Integrative identification of immune-related key genes in atrial fibrillation using weighted gene coexpression network analysis and machine learning. Front Cardiovasc Med. 2022; 9:922523.
https://doi.org/10.3389/fcvm.2022.922523
PMID:35966550

8. Fu Y, Dominissini D, Rechavi G, He C. Gene expression regulation mediated through reversible m^{6}A RNA methylation. Nat Rev Genet. 2014; 15:293–306.
https://doi.org/10.1038/nrg3724
PMID:24662220

9. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. Cell. 2017; 169:1187–200.
https://doi.org/10.1016/j.cell.2017.05.045
PMID:28622506

10. Patil DP, Chen CK, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature. 2016; 537:369–73.
https://doi.org/10.1038/nature19342 PMID:27602518

11. Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic transcriptomic m^{6}A decoration: writers, erasers, readers and functions in RNA metabolism. Cell Res. 2018; 28:616–24.
https://doi.org/10.1038/s41422-018-0040-8
PMID:29789545

12. Wang L, Wen M, Cao X. Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses. Science. 2019; 365:eaav0758.
https://doi.org/10.1126/science.aav0758
PMID:31320558

13. Han D, Liu J, Chen C, Dong L, Liu Y, Chang R, Huang X, Liu Y, Wang J, Dougherty U, Bissonnette MB, Shen B, Weichselbaum RR, Xu MM, He C. Anti-tumour immunity controlled through mRNA m^{6}A methylation and YTHDF1 in dendritic cells. Nature. 2019; 566:270–4.
https://doi.org/10.1038/s41586-019-0916-x
PMID:30728504
Erratum in: Nature. 2019; 568:E3.

14. Li HB, Tong J, Zhu S, Batista PJ, Duffy EE, Zhao J, Bailis W, Cao G, Kroehling L, Chen Y, Wang G, Broughton JP, Chen YG, et al. m^{6}A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. Nature. 2017; 548:338–42.
https://doi.org/10.1038/nature23450 PMID:28792938

15. Liu Y, Shi Q, Ma Y, Liu Q. The role of immune cells in atrial fibrillation. J Mol Cell Cardiol. 2018; 123:198–208.
https://doi.org/10.1016/j.yjmcc.2018.09.007
PMID:30267749

16. Patel P, Dokainish H, Tsai P, Lakkis N. Update on the association of inflammation and atrial fibrillation. J Cardiovasc Electrophysiol. 2010; 21:1064–70.
https://doi.org/10.1111/j.1540-8167.2010.01774.x
PMID:20455973

17. Yamashita T, Sekiguchi A, Iwasaki YK, Date T, Sagara K, Tanabe H, Suma H, Sawada H, Aizawa T. Recruitment of immune cells across atrial endocardium in human atrial fibrillation. Circ J. 2010; 74:262–70.
https://doi.org/10.1253/circj.cj-09-0644
PMID:20009387

18. Hohmann C, Pfister R, Mollenhauer M, Adler C, Kozlowski J, Wodarz A, Drebber U, Wippermann J,

Michels G. Inflammatory cell infiltration in left atrial appendageal tissues of patients with atrial fibrillation and sinus rhythm. Sci Rep. 2020; 10:1685.
https://doi.org/10.1038/s41598-020-58797-8
PMID:32015492

19. Zheng Q, Hou J, Zhou Y, Li Z, Cao X. The RNA helicase DDX46 inhibits innate immunity by entrapping m⁶A-demethylated antiviral transcripts in the nucleus. Nat Immunol. 2017; 18:1094–103.
https://doi.org/10.1038/ni.3830 PMID:28846086 Erratum in: Nat Immunol. 2017; 18:1361.

20. Yang S, Wei J, Cui YH, Park G, Shah P, Deng Y, Aplin AE, Lu Z, Hwang S, He C, He YY. m⁶A mRNA demethylase FTO regulates melanoma tumorigenicity and response to anti-PD-1 blockade. Nat Commun. 2019; 10:2782.
https://doi.org/10.1038/s41467-019-10669-0
PMID:31239444

21. Zhang C, Zhang M, Ge S, Huang W, Lin X, Gao J, Gong J, Shen L. Reduced m6A modification predicts malignant phenotypes and augmented Wnt/PI3K-Akt signaling in gastric cancer. Cancer Med. 2019; 8:4766–81.
https://doi.org/10.1002/cam4.2360 PMID:31243897

22. Xie J, Chen L, Sun Q, Li H, Wei W, Wu D, Hu Y, Zhu Z, Shi J, Wang M. An immune subtype-related prognostic signature of hepatocellular carcinoma based on single-cell sequencing analysis. Aging (Albany NY). 2022; 14:3276–92.
https://doi.org/10.18632/aging.204012
PMID:35413690

23. Qiu Y, Li H, Xie J, Qiao X, Wu J. Identification of ABCC5 Among ATP-Binding Cassette Transporter Family as a New Biomarker for Hepatocellular Carcinoma Based on Bioinformatics Analysis. Int J Gen Med. 2021; 14:7235–46.
https://doi.org/10.2147/IJGM.S333904
PMID:34737618

24. Xie J, Li H, Chen L, Cao Y, Hu Y, Zhu Z, Wang M, Shi J. A Novel Pyroptosis-Related lncRNA Signature for Predicting the Prognosis of Skin Cutaneous Melanoma. Int J Gen Med. 2021; 14:6517–27.
https://doi.org/10.2147/IJGM.S335396
PMID:34675619

25. Li S, Jiang Z, Chao X, Jiang C, Zhong G. Identification of key immune-related genes and immune infiltration in atrial fibrillation with valvular heart disease based on bioinformatics analysis. J Thorac Dis. 2021; 13:1785–98.
https://doi.org/10.21037/jtd-21-168 PMID:33841968

26. Liu A, Jia K, Liang H, Jin Q. Comprehensive analysis of autophagy-related genes and patterns of immune cell infiltration in valvular atrial fibrillation. BMC Cardiovasc Disord. 2021; 21:132.

https://doi.org/10.1186/s12872-021-01939-1
PMID:33706714

27. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559.
https://doi.org/10.1186/1471-2105-9-559
PMID:19114008

28. Chen R, Liu X, Jin S, Lin J, Liu J. Machine Learning for Drug-Target Interaction Prediction. Molecules. 2018; 23:2208.
https://doi.org/10.3390/molecules23092208
PMID:30200333

29. Wei S, Lu J, Lou J, Shi C, Mo S, Shao Y, Ni J, Zhang W, Cheng X. Gastric Cancer Tumor Microenvironment Characterization Reveals Stromal-Related Gene Signatures Associated With Macrophage Infiltration. Front Genet. 2020; 11:663.
https://doi.org/10.3389/fgene.2020.00663
PMID:32695142

30. Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvänen AC, Rönnblom L, Behrens TW, Graham RR, Vyse TJ. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS Genet. 2011; 7:e1002341.
https://doi.org/10.1371/journal.pgen.1002341
PMID:22046141

31. Jiao L, Song J, Ding L, Liu T, Wu T, Zhang J, Bai H, Chen H, Zhao Z, Ying B. A Novel Genetic Variation in *NCF2*, the Core Component of NADPH Oxidase, Contributes to the Susceptibility of Tuberculosis in Western Chinese Han Population. DNA Cell Biol. 2020; 39:57–62.
https://doi.org/10.1089/dna.2019.5082
PMID:31794672

32. Li HM, Huang Q, Tang F, Zhang TP. Altered NCF2, NOX2 mRNA Expression Levels in Peripheral Blood Mononuclear Cells of Pulmonary Tuberculosis Patients. Int J Gen Med. 2021; 14:9203–9.
https://doi.org/10.2147/IJGM.S339194 PMID:34880659

33. Su W, Zhao Y, Wei Y, Zhang X, Ji J, Yang S. Exploring the Pathogenesis of Psoriasis Complicated With Atherosclerosis via Microarray Data Analysis. Front Immunol. 2021; 12:667690.
https://doi.org/10.3389/fimmu.2021.667690
PMID:34122426

34. Chu Y, Yu F, Wu Y, Yang J, Shi J, Ye T, Han D, Wang X. Identification of genes and key pathways underlying the pathophysiological association between nonalcoholic fatty liver disease and atrial fibrillation. BMC Med Genomics. 2022; 15:150.
https://doi.org/10.1186/s12920-022-01300-1
PMID:35790963

35. Yan T, Zhu S, Zhu M, Wang C, Guo C. Integrative Identification of Hub Genes Associated With Immune Cells in Atrial Fibrillation Using Weighted Gene Correlation Network Analysis. Front Cardiovasc Med. 2021; 7:631775.
https://doi.org/10.3389/fcvm.2020.631775
PMID:33553270

36. Zhou Y, Wang X, Zhang W, Liu H, Liu D, Chen P, Xu D, Liu J, Li Y, Zeng G, Li M, Wu Z, Zhang Y, et al. The Immune-Related Gene HCST as a Novel Biomarker for the Diagnosis and Prognosis of Clear Cell Renal Cell Carcinoma. Front Oncol. 2021; 11:630706.
https://doi.org/10.3389/fonc.2021.630706
PMID:33968730

37. Wang W, Li S, Lin J, Guo X, Xie Y, Li W, Hao Y, Jiang X. The roles and potential mechanisms of HCST in the prognosis and immunity of KIRC via comprehensive analysis. Am J Transl Res. 2022; 14:752–71.
PMID:35273683

38. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47.
https://doi.org/10.1093/nar/gkv007 PMID:25605792

39. Tian Y, Yang J, Lan M, Zou T. Construction and analysis of a joint diagnosis model of random forest and artificial neural network for heart failure. Aging (Albany NY). 2020; 12:26221–35.
https://doi.org/10.18632/aging.202405
PMID:33401250

40. Wu Y, Chen H, Li L, Zhang L, Dai K, Wen T, Peng J, Peng X, Zheng Z, Jiang T, Xiong W. Construction of Novel Gene Signature-Based Predictive Model for the Diagnosis of Acute Myocardial Infarction by Combining Random Forest With Artificial Neural Network. Front Cardiovasc Med. 2022; 9:876543.
https://doi.org/10.3389/fcvm.2022.876543
PMID:35694667

41. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m^{6}A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. Mol Cancer. 2020; 19:53.
https://doi.org/10.1186/s12943-020-01170-0
PMID:32164750

42. Chai RC, Wu F, Wang QX, Zhang S, Zhang KN, Liu YQ, Zhao Z, Jiang T, Wang YZ, Kang CS. m^{6}A RNA methylation regulators contribute to malignant progression and have clinical prognostic impact in gliomas. Aging (Albany NY). 2019; 11:1204–25.
https://doi.org/10.18632/aging.101829
PMID:30810537

43. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26:1572–3.
https://doi.org/10.1093/bioinformatics/btq170
PMID:20427518

44. Shen S, Wang G, Zhang R, Zhao Y, Yu H, Wei Y, Chen F. Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. EBioMedicine. 2019; 40:318–26.
https://doi.org/10.1016/j.ebiom.2018.12.054
PMID:30594555

45. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1:417–25.
https://doi.org/10.1016/j.cels.2015.12.004
PMID:26771021

46. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol. 2008; 4:e1000117.
https://doi.org/10.1371/journal.pcbi.1000117
PMID:18704157

47. Zheng PF, Chen LZ, Liu P, Pan HW. A novel lncRNA-miRNA-mRNA triple network identifies lncRNA XIST as a biomarker for acute myocardial infarction. Aging (Albany NY). 2022; 14:4085–106.
https://doi.org/10.18632/aging.204075
PMID:35537778

48. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–7.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

49. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010; 33:1–22.
PMID:20808728

50. Wang Q, Liu X. Screening of feature genes in distinguishing different types of breast cancer using support vector machine. Onco Targets Ther. 2015; 8:2311–7.
https://doi.org/10.2147/OTT.S85271 PMID:26347014

51. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, Boriani G, Castella M, Dan GA, Dilaveris PE, Fauchier L, Filippatos G, Kalman JM, et al, and ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special

contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J. 2021; 42:373–498.
https://doi.org/10.1093/eurheartj/ehaa612
PMID:32860505
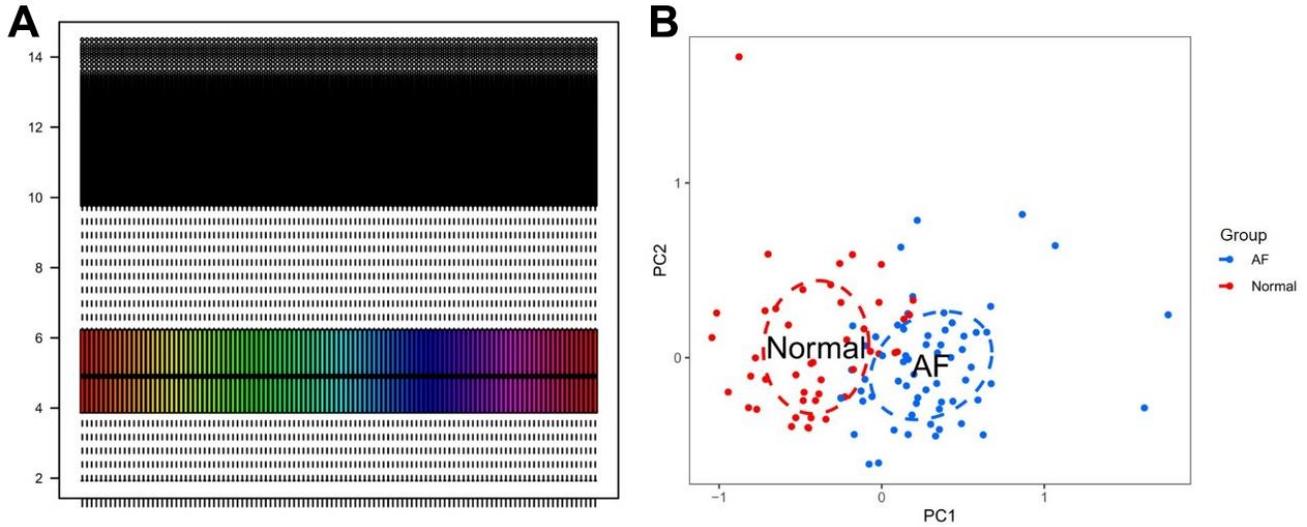
Supplementary Figure 1. Standardization and principal component analysis of all samples. (**A**) Normalized for all samples. (**B**) Principal component analysis for all samples.
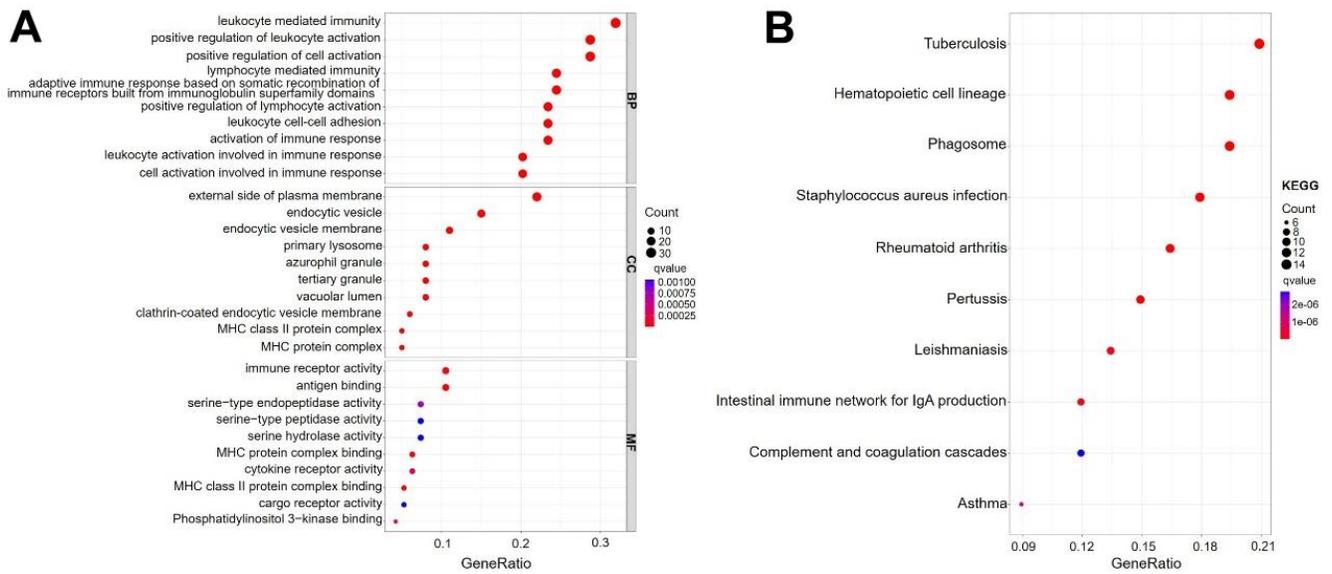
Supplementary Figure 2. Enrichment analysis of the genes in the meaningful modules. (**A**) GO functional enrichment analysis. (**B**) KEGG pathways analysis. The gene number was represented at x-axis, the KEGG pathway and GO terms were presented at y-axis.
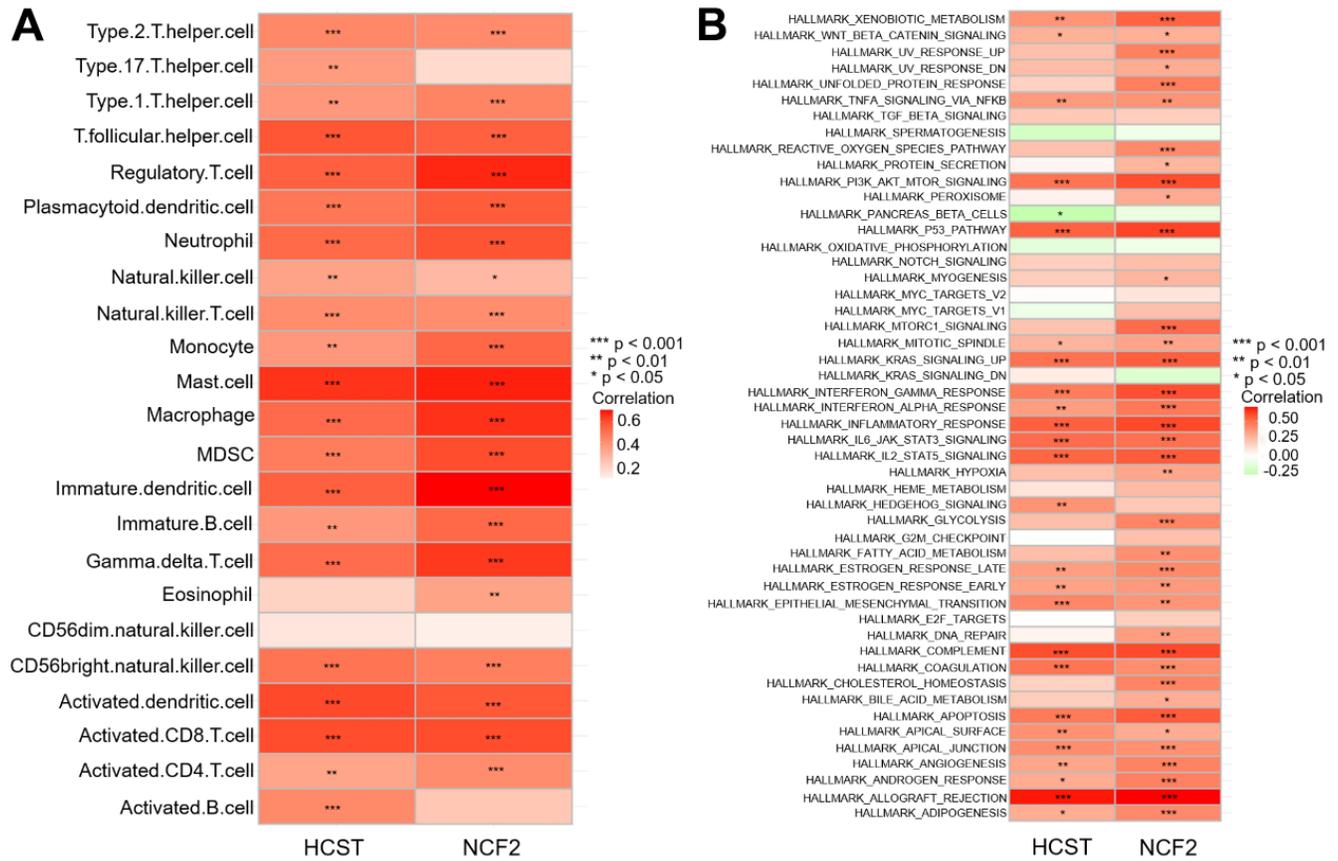
Supplementary Figure 3. Heatmap of correlations between *NCF2* as well as *HCST* genes and infiltrating immune cells and 50 HALLMARKS pathways. (**A**) The relationship between infiltrating immune cells and the *NCF2* and *HCST* genes. (**B**) The relationship between 50 HALLMARKS pathways and the *NCF2* and *HCST* genes. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 3, 4, 5, 6, 7.

Supplementary Table 1. The expression profile of the 21652 genes in the datasets of GSE31821, GSE41177, GSE76796 and GSE115574.

Supplementary Table 2. The m6A RNA methylation modification patterns mediated by 6 key m6A regulators in AF.

ID	RBM15B	HNRNPC	HNRNPA2B1	IGFBP2	IGFBP3	ALKBH5	m6Acluster
GSM789271_treat	5.90138	7.129	8.058032	9.083843	10.027447	7.460989	A
GSM789272_treat	6.585327	7.377221	7.066014	10.813303	10.0168	7.8383	B
GSM789274_treat	6.258503	7.531016	8.161363	9.405052	10.37481	7.624308	C
GSM789275_treat	6.474373	7.389506	7.228629	9.302356	9.066857	7.742908	C
GSM1005424_treat	6.262892	7.408608	7.810999	10.117958	9.868154	7.598898	C
GSM1005426_treat	6.412639	7.339644	7.8383	10.299514	9.623874	7.605746	C
GSM1005428_treat	6.435906	7.372142	7.328194	10.48625	9.491856	7.665631	B
GSM1005430_treat	6.430108	7.380779	7.778062	9.810857	9.88699	7.572645	C
GSM1005432_treat	6.261587	7.384813	7.804135	9.435905	9.538878	7.474808	C
GSM1005434_treat	6.343933	7.352928	7.663704	9.979509	9.990233	7.55616	C
GSM1005436_treat	6.409742	7.402653	7.443494	9.88699	10.092063	7.580118	C
GSM1005438_treat	6.291567	7.39058	7.567494	9.86198	10.066885	7.515343	C
GSM1005440_treat	6.359973	7.374409	7.30202	9.881155	9.963446	7.605746	C
GSM1005442_treat	6.270799	7.399298	7.722304	9.841396	9.57037	7.532237	C
GSM1005444_treat	6.490407	7.477738	7.340784	10.257411	9.510538	7.75741	C
GSM1006245_treat	6.354654	7.442932	7.726283	9.706861	10.114567	7.634785	C
GSM1006247_treat	6.632033	7.283248	7.904238	10.287115	9.906506	7.543983	C
GSM1006249_treat	6.542227	7.186585	7.930539	10.037417	9.349632	7.230754	C
GSM1006251_treat	6.185407	7.41665	7.782726	9.792959	9.588153	7.390041	C
GSM1006253_treat	6.26602	7.377221	7.648592	9.792959	9.445125	7.49398	C
GSM2102184_treat	6.153389	7.511045	7.989245	10.918397	10.166073	7.630532	B
GSM2102185_treat	6.206053	7.49464	7.663039	10.519623	10.941273	8.067351	B
GSM2102186_treat	6.285993	7.525887	7.67074	9.477383	8.425224	7.621636	C
GSM2102187_treat	6.091918	7.607638	8.185945	8.479291	9.242633	7.563582	A
GSM2102188_treat	6.127192	7.277583	7.887248	9.38702	9.69243	6.942886	C
GSM2102189_treat	6.225165	7.273687	7.982342	9.661232	10.536379	7.736113	C
GSM2102190_treat	6.605199	7.331015	7.529836	10.429378	8.932519	7.53881	B
GSM2102191_treat	6.558832	6.843365	7.640226	9.289926	9.755126	7.711509	C
GSM2102192_treat	6.473328	7.131818	7.560531	9.760976	9.11236	7.349055	C
GSM2102193_treat	6.206053	7.380779	7.962586	8.014668	10.450505	7.786423	A
GSM2102194_treat	6.541586	7.331607	7.565579	9.912989	8.996157	7.69189	C
GSM2102195_treat	6.577509	7.306838	6.94679	9.859129	9.332328	8.33254	C
GSM2102196_treat	6.266912	7.335861	7.658575	6.649301	9.947684	7.859217	A
GSM2102197_treat	6.928074	6.971467	7.782726	9.183712	8.754232	7.649273	C
GSM3182680_treat	6.451409	7.344628	7.666274	10.502601	8.380627	7.747689	B
GSM3182681_treat	6.29024	7.246575	7.339644	9.629004	9.993466	7.839076	C
GSM3182682_treat	6.116225	7.191712	7.366893	10.723776	8.908684	7.810999	B
GSM3182683_treat	6.331233	7.402653	7.520917	10.051576	9.925384	7.595759	C
GSM3182684_treat	6.565227	7.342487	7.717524	10.799913	9.144803	7.608857	B
GSM3182685_treat	6.123798	7.309552	7.675719	8.650796	10.084928	7.829262	A

GSM3182686_treat	6.441962	7.041962	7.582069	10.910656	9.046697	7.399848	B
GSM3182687_treat	6.473962	7.263157	7.521482	9.51298	9.906506	7.780009	C
GSM3182688_treat	6.34714	7.189202	7.48677	11.323854	10.102326	7.601435	B
GSM3182689_treat	6.515523	7.527979	8.163929	10.949486	9.631794	7.125972	B
GSM3182690_treat	6.126871	7.380779	7.749103	10.799913	9.943886	7.272573	B
GSM3182691_treat	6.384972	7.450711	7.957127	9.732656	10.491708	7.6409	C
GSM3182692_treat	6.624983	7.20753	8.019662	8.689859	9.43333	7.210897	A
GSM3182693_treat	6.408093	7.499333	7.663039	9.051927	10.261423	7.252255	A
GSM3182694_treat	5.759048	7.147268	7.850979	11.034949	8.870444	6.644807	B
GSM3182695_treat	6.218431	6.932924	7.709471	10.577904	9.996985	7.522144	B
GSM3182696_treat	6.896041	6.749387	6.905003	11.061198	9.426002	7.398668	B
GSM3182697_treat	6.243748	6.926863	7.101934	11.968874	11.093496	7.864051	B
GSM3182698_treat	6.203795	7.425675	7.812532	11.755773	9.300529	8.211354	B
GSM3182699_treat	5.949588	7.403739	7.410937	11.981442	11.248291	8.009548	B
GSM3182700_treat	5.91652	7.239622	8.155872	9.897344	8.386245	7.374409	C
GSM3182701_treat	6.04131	7.221645	7.688407	8.574566	10.730452	7.976127	A
GSM3182702_treat	6.290576	7.355592	7.309552	10.641517	9.20503	7.940356	B
GSM3182703_treat	6.837514	7.388342	7.407493	10.903509	10.460398	8.209532	B
GSM3182704_treat	6.74385	7.222691	7.660359	11.350736	9.260431	7.323158	B
GSM3182705_treat	6.525857	7.165862	8.441135	6.14668	10.778199	5.754425	A
GSM3182706_treat	6.083648	6.797624	7.952319	8.57177	10.218241	7.377736	A
GSM3182707_treat	6.333223	6.878329	7.695903	9.684138	7.761619	7.504573	B

Supplementary Table 3. The immune microenvironment characteristics in control and AF samples.

Supplementary Table 4. The results of gene set variation analysis enrichment analysis.

Supplementary Table 5. The genes in the lightgreen and lightcyan modules.

Supplementary Table 6. Detailed results of the GO enrichment analysis.

Supplementary Table 7. Detailed results of the KEGG enrichment analysis.

Supplementary Table 8. The details genes symbols of several key genes identified by two distinct machine learning methods.

Common elements in LASSO and SVM-RFE :
AMICA1
BCAT1
C16orf54
C1QC
CD14
CD48
CD8A
CORO1A
CXCL12
HCLS1
HCST
HMHA1
IGLV1-44
LAPTM5
NCF2
TNFSF13
Elements only in LASSO :
C1S
CASP1
CD2
CTSS
GZMH
HLA-DMB
HPGDS
ITGAM
LCK
LYZ
MS4A7
SLC7A7
VAMP8
Elements only in SVM-RFE :
HLA-DRA
LYN
HLA-DMA
LYVE1
IGLC1
CCL5
CD163
GIMAP2
RNASE6
MARCKS
CCL19
MRC1