Journal of Inflammation Research

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Identification of Five NK Cell-Related Hub Genes in COPD Using Single-Cell RNA Sequencing Analysis

Xiaojie Deng[1,*], Xiahui Yang[1,*], Zhihua Gan[2], Huaxing Huang[1], Jun Yang[1]

[1]Department of Pulmonary and Critical Care Medicine, The Second Affiliated Hospital of Guangzhou Medical University Guangzhou, Guangdong, People's Republic of China; [2]Department of Pulmonary and Critical Care Medicine, The Affiliated Brain Hospital, Guangzhou Medical University Guangzhou, Guangdong, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jun Yang, Email yangjun2001@gzhmu.edu.cn

Background: COPD is a healthcare problem. However, the underlying mechanism remains unclear. Our study aimed to explore the key genes involved in immune infiltration in COPD using bioinformatic tools.

Methods: In this study, scRNA-seq analysis was utilized to explore specific marker genes of each immune cell subtype in COPD. TSNE analysis was used to evaluate the relationship between each immune cell cluster. Lasso regression identified 21 genes as characteristics of COPD modulated by the single-cell NK cell subpopulation. The "limma" package was used for differentially expressed analysis. The pseudotime analysis reveals the continuous changes of NK cells along their developmental trajectory. Further, we constructed a hub gene network to examine the correlation between hub genes and immune factors, transcriptional regulation factors, and potential therapeutic drugs. GO and KEGG enrichment analysis revealed the biological functions of the hub genes. RT-qPCR was used for validation of the five hub in COPD patients.

Results: NK cell subtypes are closely related to other immune cell subtypes and considered as the most important immune cells in the immune microenvironment of COPD patients. LASSO regression identified 21 genes as NK cells-characteristic genes for COPD. The GSE57148 as the training set has a AUC of 0.9489 and GSE8581 as the validation set has a AUC of 0.7303. The GO semantic similarity further confirmed five NK cell-related hub genes, C1orf56, S100A6, IGFBP7, ANXA1, and PTPN7. RT-qPCR experiment revealed that the mRNA expression of five hub genes in the normal group was lower than that in the disease group. We also found that five hub genes correlated with immune cell infiltration. The potential therapeutic agents for COPD may be zalcitabine, PP-2, PD-98059, and TGX-221 based on the CMap database prediction.

Conclusion: We proposed that peripheral NK cells may play a role in the pathogenesis of COPD through bioinformatic analysis. These hub genes may provide insights into mechanistic research and new targets for new therapies in patients with COPD.

Keywords: COPD, hub genes, single-cell RNA-sequencing, NK cell, immune infiltration

Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease that causes a healthcare burden and is the third leading cause of death worldwide. The prevalence of COPD ranges from 13.1% in the adult population in Europe, aged over 40 years, to 13.7% in Chinese.[1–5] Despite the progress in diagnostic criteria and treatment strategies for decades, the definition and therapeutic strategies are still based on clinical characteristics.[6,7] Thus, the exploration of complex mechanisms and development of better prognostic biomarkers are urgently needed.

Increasing evidence suggests that COPD pathogenesis is related to various biological functions including cell proliferation, apoptosis, autophagy, and inflammation.[8] The hub genes with the most key module were identified to be the biomarkers in the biological mechanism. With the development of bioinformatic analysis and the application of gene expression profiles, hub genes have been identified with respect to the etiology of COPD.[9,10] However, there is still a lack of sufficiently specific and sensitive biomarkers owing to disease heterogeneity and confounding factors.[11,12]

Received: 14 September 2024
Accepted: 31 December 2024
Published: 12 February 2025

In recent years, increasing studies have proved that the immune cell infiltration plays a role in the occurrence and development of COPD.[13] There was various evidence showed that natural killer cell (NK cell) populations may be associated with high CMV seropositivity, current therapy and the risk of COPD patients, especially smokers.[14,15] However, the molecular variations in systemic immune system of patients in COPD are poorly understood.

Applications of a single-cell RNA-sequencing (scRNA-seq) technologies were used to provide a more accurate classification and more information of the immune microenvironment of inflammatory disease.[16] ScRNA-seq studies on COPD may help find novel insights into the molecular mechanisms and identify new treatment targets.

In this study, we investigated microarray datasets from the GEO database (GSE167295, GSE57148, GSE8581) to determine the types of immune cells that are specifically involved in COPD. We used GSE57148 as the training set, and GSE8581 as the validation set. LASSO regression identified 21 genes as single-cell NK_cell-featured genes for COPD, and the GO semantic similarity analysis confirmed that the hub genes C1orf56, S100A6, IGFBP7, ANXA1, and PTPN7 were confirmed in our study. Single-cell analysis revealed the expression model of hub genes plays a role in COPD immune cell clusters. Furthermore, we investigated the transcriptional regulation of hub genes, identified related miRNAs, and predicted targeted medications. Our study aimed to explore the potential modulatory mechanisms in COPD via bioinformatics analysis of single-cell RNA data.

Materials and Methods

Data Download

The GENE EXPRESSION OMNIBUS (GEO) database is a gene expression profile constructed by the National Center for Biotechnology Information (NCBI). The GEO database contains microarray, next-generation sequencing, and high-throughput sequencing data. The single-cell file, GSE167295, was downloaded from the GEO database for single-cell analysis. The Series Matrix File data file GSE57148, annotated with file GPL11154, including the expression profile data of 189 patients, was divided into 91 cases in the control group and 98 cases in the disease group. The Series Matrix File data file of GSE8581, annotated with GPL570, including the expression profile data of 35 patients, was divided into 19 control and 16 disease patients.

According to item 1 and 2 of Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects dated February 18, 2023, China, if the use of human information data or biological samples for life science and medical research involving humans does not cause harm to the human body, does not involve sensitive personal information or commercial interests, ethical review can be exempted to reduce unnecessary burden on researchers and promote the development of life science and medical research involving humans: (1) legally obtained public data or data generated through observation without interfering with public behavior for research; (2) anonymized information data.

The GEO database used in our study is legally obtained public and anonymized information data, in which the data were obtained without interfering with public behavior. Therefore, this database is in accord with the scope of exemption approval. Thus, our study is exempt from ethical approval based on national legislation guidelines.

Single Cell Analysis

First, the Seurat package is used for preprocessing and quality control of single-cell RNA sequencing data. We calculate the expression ratio of mitochondrial genes in each cell using the Percentage Eigen Set function and store it as percent.mt. To evaluate the data quality, we visualized nFeature-RNA, nCunt-RNA, and percent.mt using violin plots. Second, a scatter plot was created using EigenScatter to analyze the relationship between nCunt-RNA, percent.mt, and nFeature-RNA. Based on quality control standards, we screened cells with nFeature-RNA >50 and percent.mt <5 to remove low-quality cells. Next, the data was normalized using Normalize Data, using the Log Normalize method and setting the normalized scale factor to 10000. Then, FindVariable Features was used to screen out 5000 highly variable genes, and the variation of these genes was visualized through Variable Feature Plot. In the principal component analysis (PCA) stage, the data were first standardized, and then PCA was run to select the top 20 principal components. Finally, t-SNE dimensionality reduction was performed on the data using RunTSNE to further analyze the clustering of cells, and FindClusters was used for cell clustering analysis with a resolution of 0.5. To further annotate the cell types of single cells, we used the SingleR method, which predicts cell types by comparing sample data with a reference database. In this study, we selected Human Primary Cell Atlas Data from the celldex package as the reference dataset for cell-type annotation.

Analysis of Ligand–Receptor Interactions

We conducted a significant analysis of ligand receptor relationships in the features of single-cell expression profiles using the CellphoneDB software package. We randomly arranged the cluster labels of all cells 1000 times and determined the average expression levels of receptors in the clusters and ligands in the interacting clusters. For each receptor ligand pair compared between two cell types, a null distribution (also known as Bernoulli distribution or two-point distribution) is generated. Finally, we selected some ligand receptor pairs to showcase their relationship.

Differential Expression Analysis

R-Pack "limma" was to identify differentially expressed genes between the control and disease groups. The screening conditions for differentially expressed genes were P.value <0.05 and | logFc |>0.585. Volcano and heat maps of the differentially expressed genes were drawn.

Collection of COPD Patient Samples

COPD patients were collected at the Second Affiliated Hospital of Guangzhou Medical University, China. The patients were fully informed of the study's purpose prior to providing their consent and our study complied with the Declaration of Helsinki. Before using these clinical materials for research purposes, consent of the patient and the approval of the Ethics Committee of the Second Affiliated Hospital of Guangzhou Medical University were obtained. The reference number is Y2021-23-02. It was obtained from patients as part of a routine hospital procedure.

RNA Extraction and Real-Time PCR

Total RNA was isolated by RNAiso Plus (Simgen). Total RNA (1 μg) was transcribed into cDNA by use of PrimeScriptTM RT reagent kit with gDNA eraser (Simgen). The RT-PCR assay was performed with TB Green™ Premix Ex Taq™ II (RR820A, Takara, Japan) and a LightCycler 96 instrument. The relative expression levels were estimated using GAPDH as an internal reference.

Fold changes were calculated by a comparative threshold cycle (Ct) method using the formula 2-ΔΔCT. The experiments were conducted three times with biological duplicates. The primer sequences are as follows (Table 1).

Predictive Model Construction

LASSO regression was used to construct a prediction-relation model for the candidate gene set. We then constructed the risk score of the patients and weighted the estimated regression coefficient using LASSO regression analysis. The ROC curve was

Table 1 Primer Sequences Used

Gene	Sequence (5to 3)
H-C1orf56-F	GTGGGTCCTGCTGCTGAATC
H-C1orf56-R	TGAAGACCCATCCTCCTCGT
H-S100A6-F	CTCCCTACCGCTCCAAGC
H-S100A6-R	CACCTCCTGGTCCTTGTTCC
H-IGFBP7-F	AAGAGGCGGAAGGGTAAAGC
H-IGFBP7-R	TGCCCTTATGGGTTGCTAACT
H-ANXA1-F	TTGCAAGAAGGTAGAGATAAAGACA
H-ANXA1-R	AAGGCCCTGGCATCTGAATC
H-PTPN7-F	GTCCACTGTCTACCTCATGGC
H-PTPN7-R	CCAGGCCGAAAAGAGGATGT

Abbreviations: H, homo sapiens; F, forward; R, reverse.

used to show the accuracy of prediction model. We constructed a prediction model using the following equation: RiskScore = ID2 × (- 0.267066685376779) + S100A6 × (−0.247077185441186) + PRDX5 × (−0.168845200184146) + CTSD × (−0.162467714926961) + KLRB1 × (−0.0846706164685527) + CD7 × (−0.0633802002394572) + HCST × (−0.0439467444664161) + PTPN7 × (−0.0344859224551446) + CST7 × (−0.0172768735682816) + PYHIN1 × (−0.0128728545833332) + IGFBP7 x (−0.0109431282642951) + GZMM × (−0.0104487517062616) + DUSP2 x0.00615757376172074 + SYTL3 x0.0161742719120735 + HSP90AA1 x0.0324601018354379 + ANXA1 × 0.0389798369030959 + DUSP1 x0.0509495275171626 + GLUL × 0.0742002610535615 + RUNX3 x0.148344169248718 + FTH1 × 0.194942531632122 + C1orf56 x0.230552622146433.

GO Semantic Similarity

Based on GO semantic similarity, we ranked proteins based on the functional similarity between proteins and interacting proteins. GO semantic similarity has been verified by the correlation between gene expression profiles,[17] which provides a basis for functional comparison of gene, such as protein–protein interaction analysis,[18] pathway analysis[19] and gene function prediction.[20] Here, we measured the functional similarity between proteins in terms of cellular components (CC) and biological pathways (BP) to explore the relationship between each protein and its interacting proteins. The semantic similarity between interacting histones in CC and BP is determined using the GoSemsim software package,[21] which is implemented more accurately by considering the GO topology.[22]

Immune Cell Infiltration Analysis

The ssGSEA method distinguished 29 human immune cell phenotypes, including T, B, and NK cells. In this study, the ssGSEA algorithm was used to quantify the immune cells in the expression profile.

GSEA Analysis

The differential signal pathways between the high- and low-expression groups were analyzed using GSEA. The background gene set was the version 7.0 annotation gene set downloaded from the MSIGDB database. Differential expression analysis of pathways between subtypes was performed using the annotation gene set of subtype pathways. Enriched gene sets (adjusted p value less than 0.05) were sorted according to their consistency scores.

GSVA (Gene Set Variance Analysis)

In this study, we downloaded gene sets from a molecular signature database and used the GSVA algorithm to measure each gene set and potential changes in biological functions.

Analysis of Transcriptional Regulation of Hub Genes

In this study, transcription factors were predicted by the R package "RcisTarget". In addition to the motifs annotated by the source data, we inferred further annotation files based on the motif similarity and gene sequences. The first step in estimating the overexpression of each motif in a gene set was to calculate the area under the curve (AUC) for each motif–motif pair. The NES of each motif was calculated based on the AUC distribution of all the motifs in the gene set.

CMap Drug Prediction

The Connectivity Map (CMap) is a gene expression profiling database based on interventional gene expression developed by the Broad Institute. It is primarily used to reveal associations between small molecules, gene expression, and disease that contains data on 1309 small-molecule drugs from five human cell lines.

Statistical Analysis

All statistical analyses were performed using the R software (version 4.2.2). The predictive power of immune cells and the efficacy of the diagnostic model were assessed using receiver operating characteristic (ROC) curves. The Wilcoxon analysis is used to analyze the differences between the two groups, including the comparison of 29 immune cells and candidate signatures between COPD patients. All statistical tests were two-sided, and statistical significance was set at $P < 0.05$.

Result

Single Cell Level Analysis in scRNA-Seq Data

The data samples were preliminarily screened based on nCount_RNA (percentage mt <5) and nFeature_RNA (nfeature_RNA>50) (Figure 1A-B). The 10 ranked genes with the highest standards are displayed (Figure 1C). PCA dimensionality reduction analysis revealed that the batch effect of each sample was not significant (Figure 1D). The optimal number of PCs observed in the elbow plot is 15 (Figure 1E). Finally, 20 subgroups were examined using TSNE analysis (Figure 1F). The ten marker genes with high expression levels are shown in the figure (Figure S1A).

Single Cell Data Cell Subpopulation Annotation

In this study, each isoform was annotated using an R package. The 20 clusters were annotated into nine cell categories: monocytes, T cells, macrophages, NK cells, epithelial cells, B cells, tissue stem cells, neutrophils, and endothelial cells (Figure 1G). Furthermore, we screened the specific marker genes of each cell subtype from the single-cell data using the FindAllMarkers function (CellMarkers.txt) (Figure S1B).

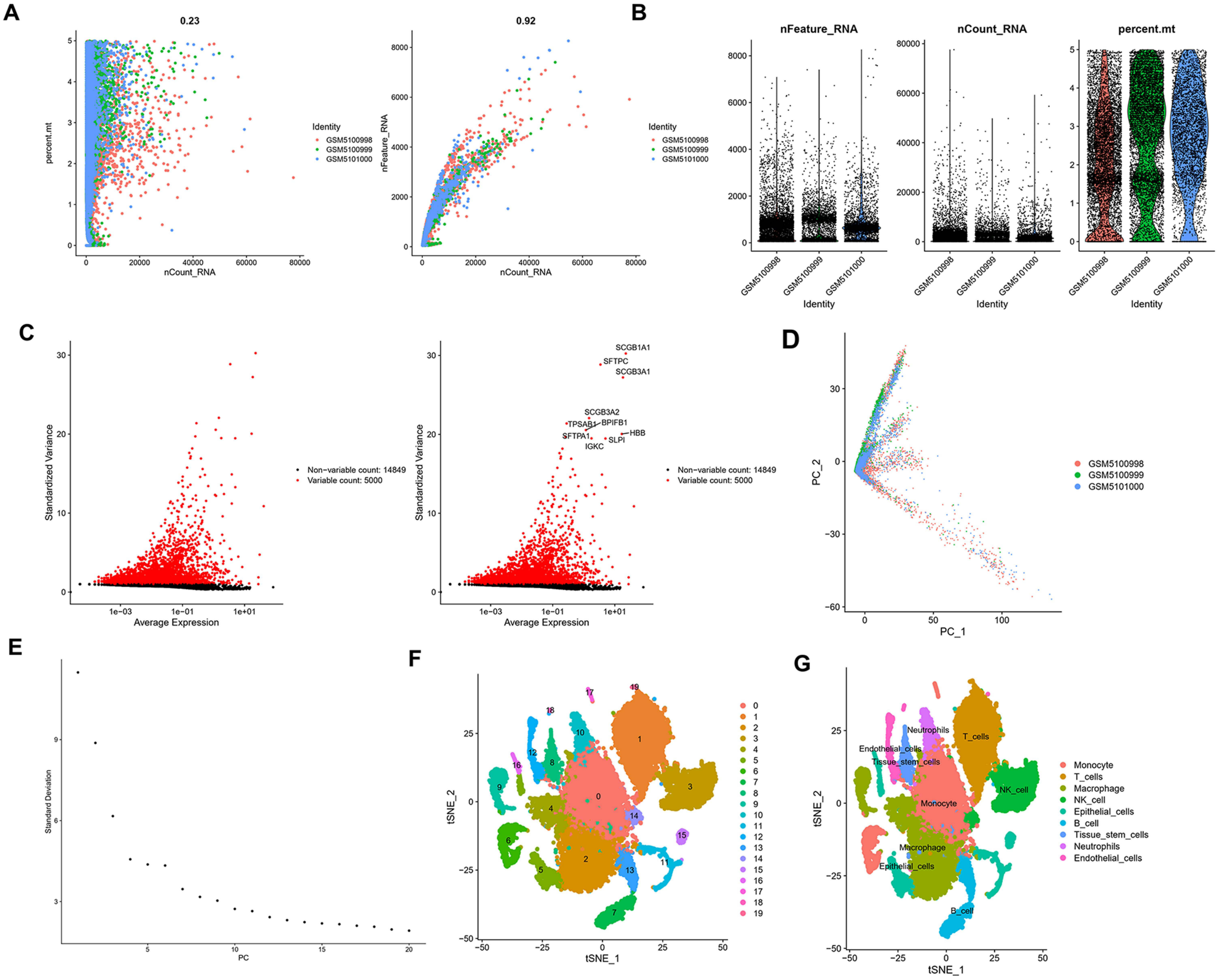

Figure 1 Identification of NK-cell marker genes by scRNA-seq analysis. (**A**) PCA was used for dimensionality reduction. (**B**) Quality control of scRNA-seq data from COPD samples. (**C**) The variance plot showed the 10 genes with the highest standard deviation. (**D** and **E**) 15 PCs were identified based on *P*-value< 0.05. (**F**) 19 clusters were visualized based on the t-SNE algorithm. (**G**) Cell subpopulations identified by marker genes.

https://doi.org/10.2147/JIR.S491298

Analysis of Receptor–Ligand Relationship Pairs

We used the Cellphonedb package to analyze the ligand receptor correlation of the characteristics in the single-cell expression profile (Figure 2A). We found that the interactions of NK_ Cell|NK_ Cell, T_ Cells|NK_ Cell, HLA-B_ KIR3DL2, and HLA-C_ KIR2DL3 had high interaction scores. Moreover, there were a number of potential ligand receptor pairs among NK _ Cell, Endothelial_ Cells, T_ cells, and other cells (Figure 2B). We counted the ligand receptor gene pairs corresponding to each cell group and found that the NK_ Cell subtypes had the most interactions (Figure 2C).

Differential Expression Analysis

We downloaded the dataset GSE57148 from the GEO database. The expression profile included 189 patients in the control group (n = 91) and disease group (n = 98). The limma package was used to calculate the differentially expressed genes between the control and disease groups. The differential conditions were P-value<0.05 and logFC>0.585. A total of 161 genes, including 106 upregulated genes and 55 downregulated genes, were screened (Figure 3A and Figure S1C).

Prediction Model Construction of Hub Genes

We used GSE57148 as the training set, and GSE8581 as the validation set. Modular genes of the single-cell NK_cell subpopulation were selected for feature screening using LASSO regression. The results showed that LASSO regression identified 21 genes as NK cell-characteristic genes for COPD (Figure 3B-D). Furthermore, we used these 21 genes as genes to construct a prediction model. The results showed that the prediction model constructed for these 21 genes had good diagnostic efficacy with an AUC of 0.9489 (Figure 3E). We further used the dataset GSE8581 as the validation set to further validate the diagnostic model, the results showed that the model had strong stability, and the GSE8581-AUC curve was 0.7303 (Figure 3F). Subsequently, we screened the hub genes using GO semantic similarity, and the results showed that C1orf56, S100A6, IGFBP7, ANXA1, and PTPN7 have most relationship with the NK cells in COPD (Figure S1D).

The Validation of Hub Genes in COPD Patients

We performed to confirm the expression of hub genes in the serum of 84 COPD patients and 20 normal patients. We collected the serum of patients with COPD and measured the mRNA expression levels of the hub genes via RT-qPCR. As shown in Figure 3, we found that the mRNA expression of C1orf56 (Figure 3G), S100A6 (Figure 3H), IGFBP7 (Figure 3J), ANXA1 (Figure 3K), and PTPN7 (Figure 3L) in the normal group were lower than the disease group.

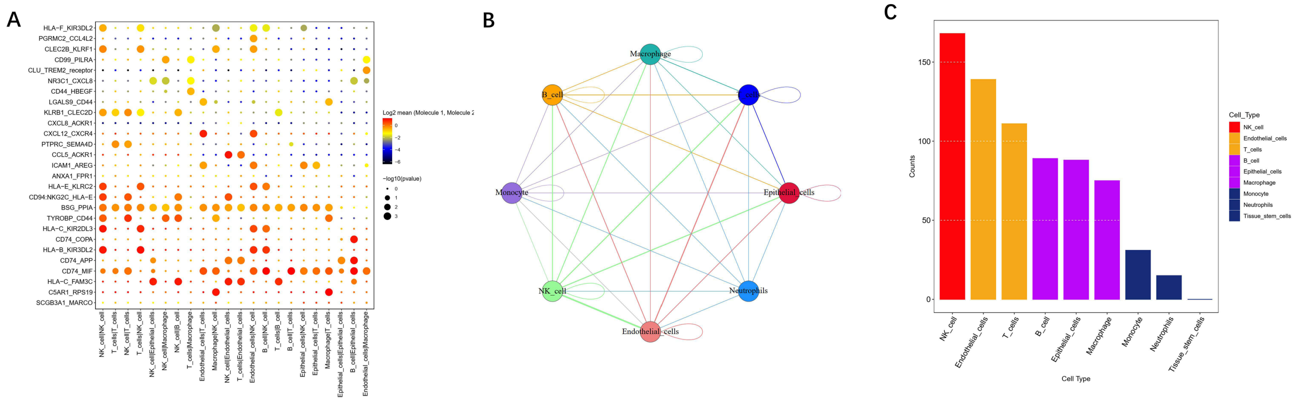

Figure 2 Analysis of receptor–ligand relationship pairs. (**A**) The heatmap showed the ligand receptor relationship in the single cell expression profile. (**B**) The interaction between NK_ Cells, endothelial_ Cells, Monocyte cells, Neutrophils Cells and T cells. (**C**) The interactions between ligand receptor gene and each cell group.

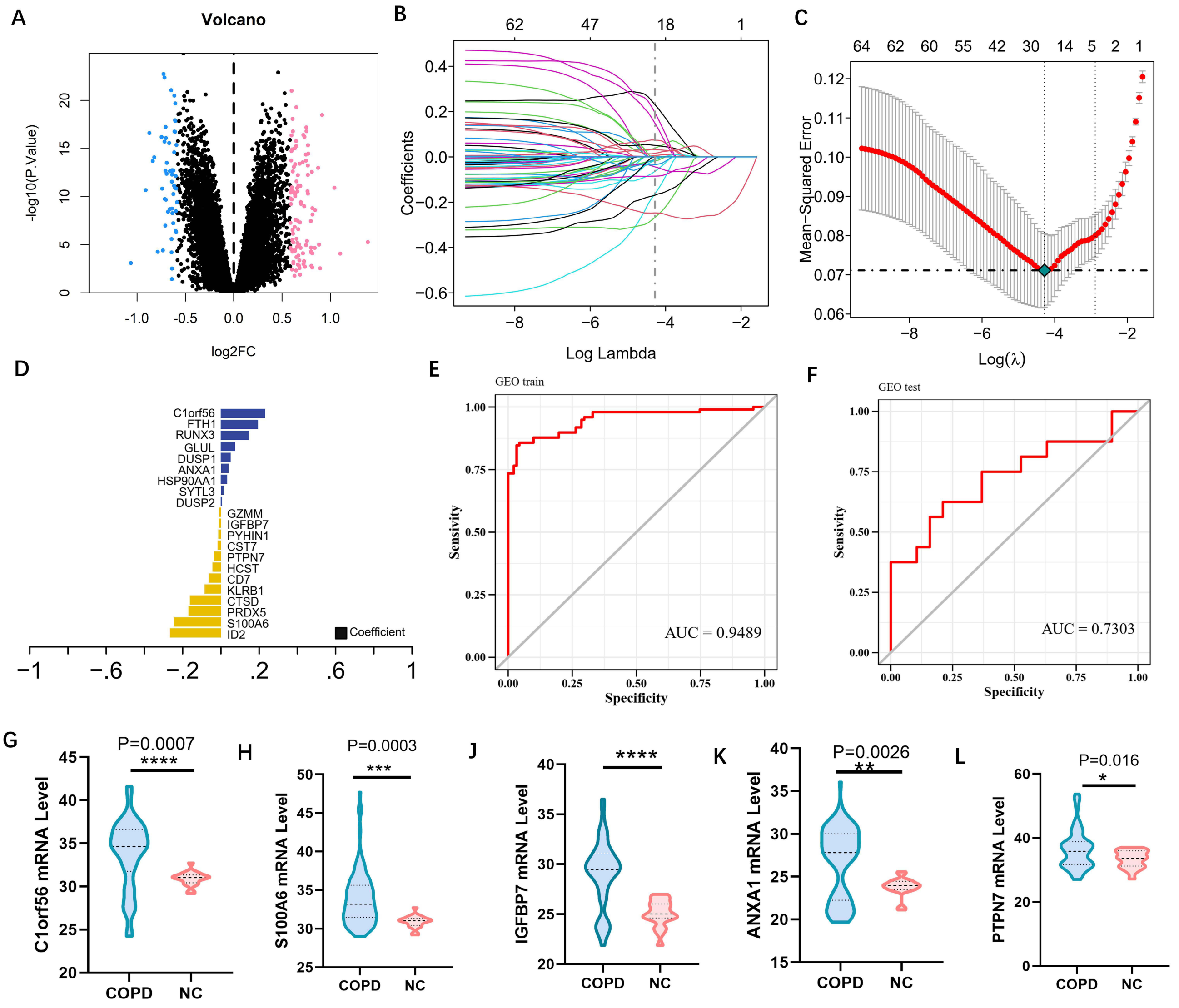

Figure 3 Construction and validation of prediction models. (**A**) The 161differential genes, including 106 up-regulated genes and 55 down-regulated genes, were plotted by Volcano. (**B** and **C**) LASSO regression analysis. Pink represents upregulation in COPD, blue represents downregulation in COPD. (**D**) Forest plot of multivariate Cox regression result. (**E** and **F**) The Kaplan–Meier curves in the training cohort, testing cohort and GEO cohort, respectively. (**G–H**) The mRNA levels of hub genes in serum were determined. NC, patients without COPD; COPD, COPD patients. Data are mean ± SD (n = 3). *P < 0.05, **P<0.01, ***P < 0.001, ****P<0.0001.

The Cell Differentiation Trajectory Image of NK Cells and Hub Genes in Single Cell Data

Subsequently, we extracted NK cells to calculate the similarity between cells and construct cell differentiation trajectories. Then, the cell differentiation trajectory image constructed by pseudo time can be generated to display the developmental process of cells. The results showed the developmental trajectory of NK cells from early precursor to mature state (Figure S2A and B). We then analyzed the relationship between hub genes and the developmental trajectory of NK cells based on the expression levels of key genes. The results showed that ANXA1 and PTPN7 increased with the increase of NK cell development time, while S100A6 decreased with the increase of NK cell development time (Figure 4A-E).

Expression Profile of Hub Genes in Single Cell Data and Immune Infiltration Analysis

We analyzed the expression of hub genes in single cells. The expression of hub genes in monocytes, T_ Cells, macrophages, NK_ Cell, Epithelial_ Cells, B_ Cell, Tissue_ Stem_ Cells, Neutrophils, and Endothelial_ Cells is shown in the Figures 5A and S3A. The proportion of immune cells in COPD patients and their correlation with each other are shown in the Figure S3B and C. In

https://doi.org/10.2147/JIR.S491298

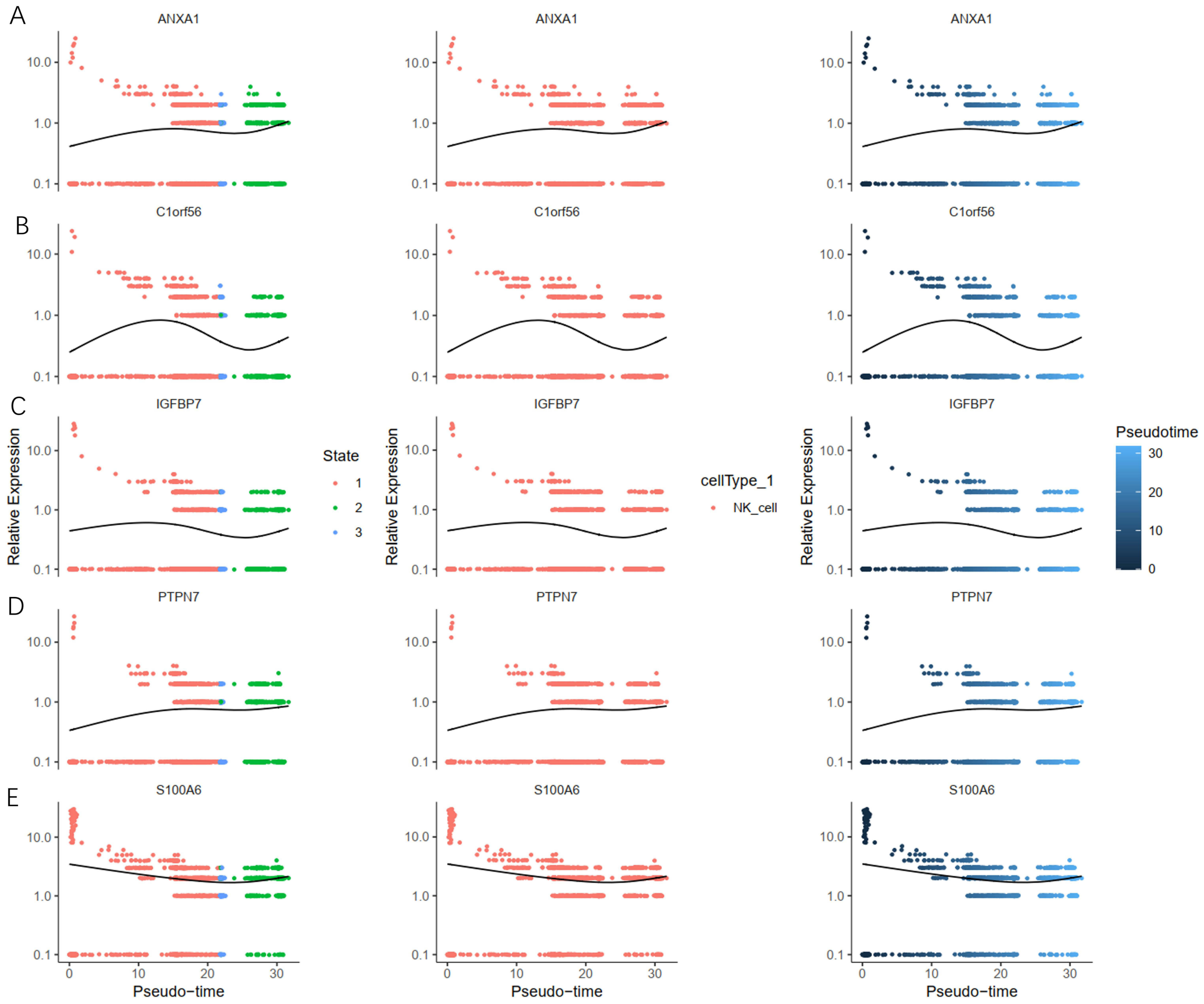

Figure 4 Pseudotime analysis of genes. (**A–E**) The correlation between the NK cells and hub genes in the cell differentiation trajectory. The X-axis represents the cell of trajectory analysis, and the Y-axis represents the relative expression of the gene.

addition, compared to the control patients, the aDCs, APC_ Co_ Stimulation, CCR, neutrophils, Treg and type II IFN responses were significantly higher in COPD patients (Figure S3D). Furthermore, the hub genes were highly correlated with immune cells (Figure 5B). We measured the correlation between these hub genes and immune factors, including immunosuppressive factors, immunostimulating factors, chemokines and receptors (Figure S4). These results suggested that hub genes are related to immune cell infiltration and play a role in the COPD immune microenvironment.

Signaling Pathways Involved in Hub Genes

We analyzed the specific signaling pathways enriched by the five hub genes. GSEA results showed that the pathways enriched were the IL − 17 signaling pathway, p53 signaling pathway, B-cell receptor signaling pathway and T-cell receptor signaling pathway (Figure 6A-E).

In addition, GSVA results showed that the high expression of hub genes was enriched in TNFa_ Signaling, p53_ Pathway, IL6_ JAK_ STAT3_ Signaling, IL2_ STAT5_ Signaling, and other signaling pathways (Figure 6F-J). This suggests that hub genes may affect the progress of chronic obstructive pulmonary disease through these immune factor-related pathways.

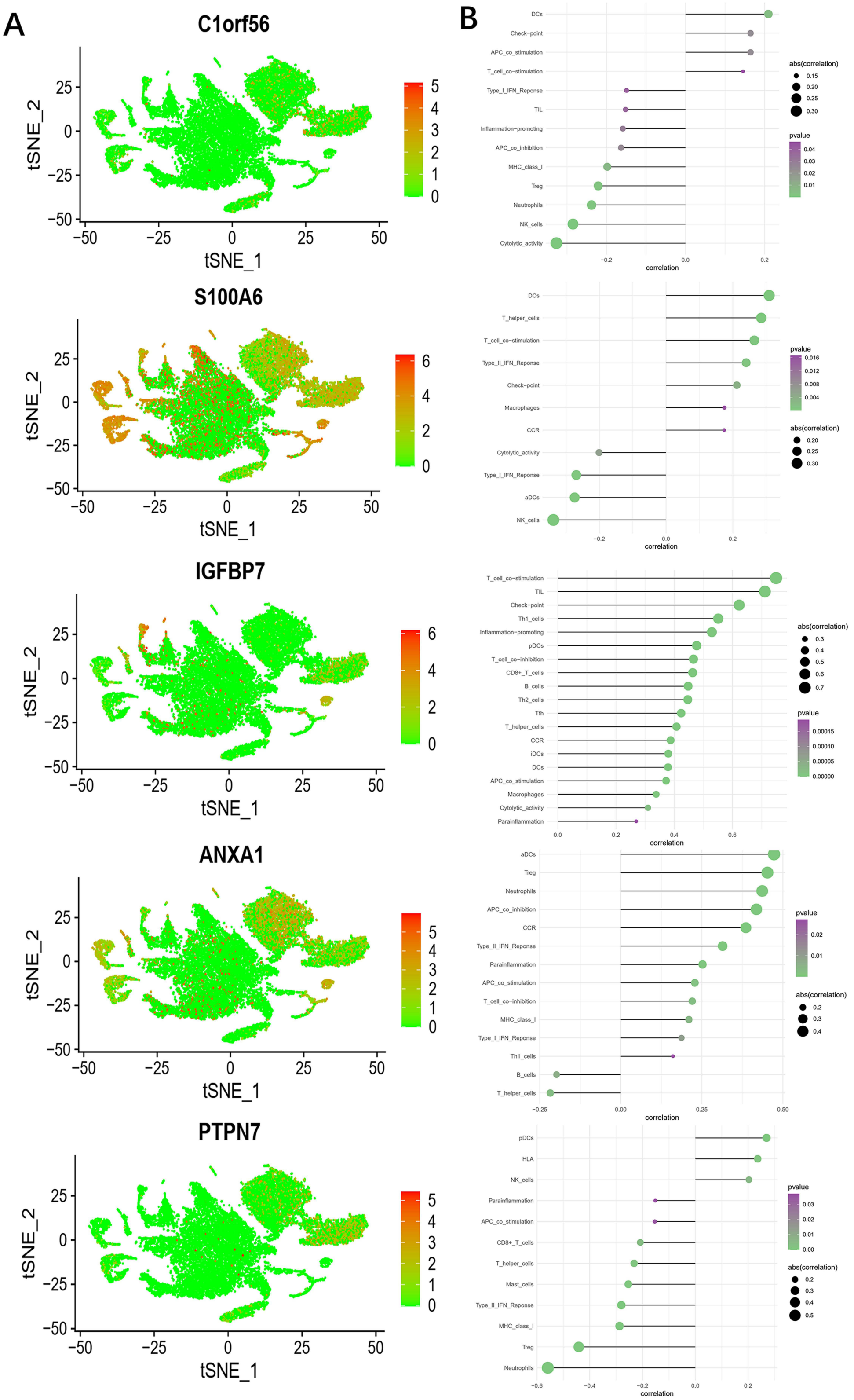

Figure 5 Expression profile of hub genes in single cell data and immune infiltration analysis. (**A**) The expression of hub genes in monocyte and t_ Cells, macrophage, NK_ Cell, epithelial_ Cells, B_ Cell, tissue_ Stem_ Cells, neutrophils, endothelial cells is shown. (**B**) The relationship between hub genes and immune cells.

https://doi.org/10.2147/JIR.S491298

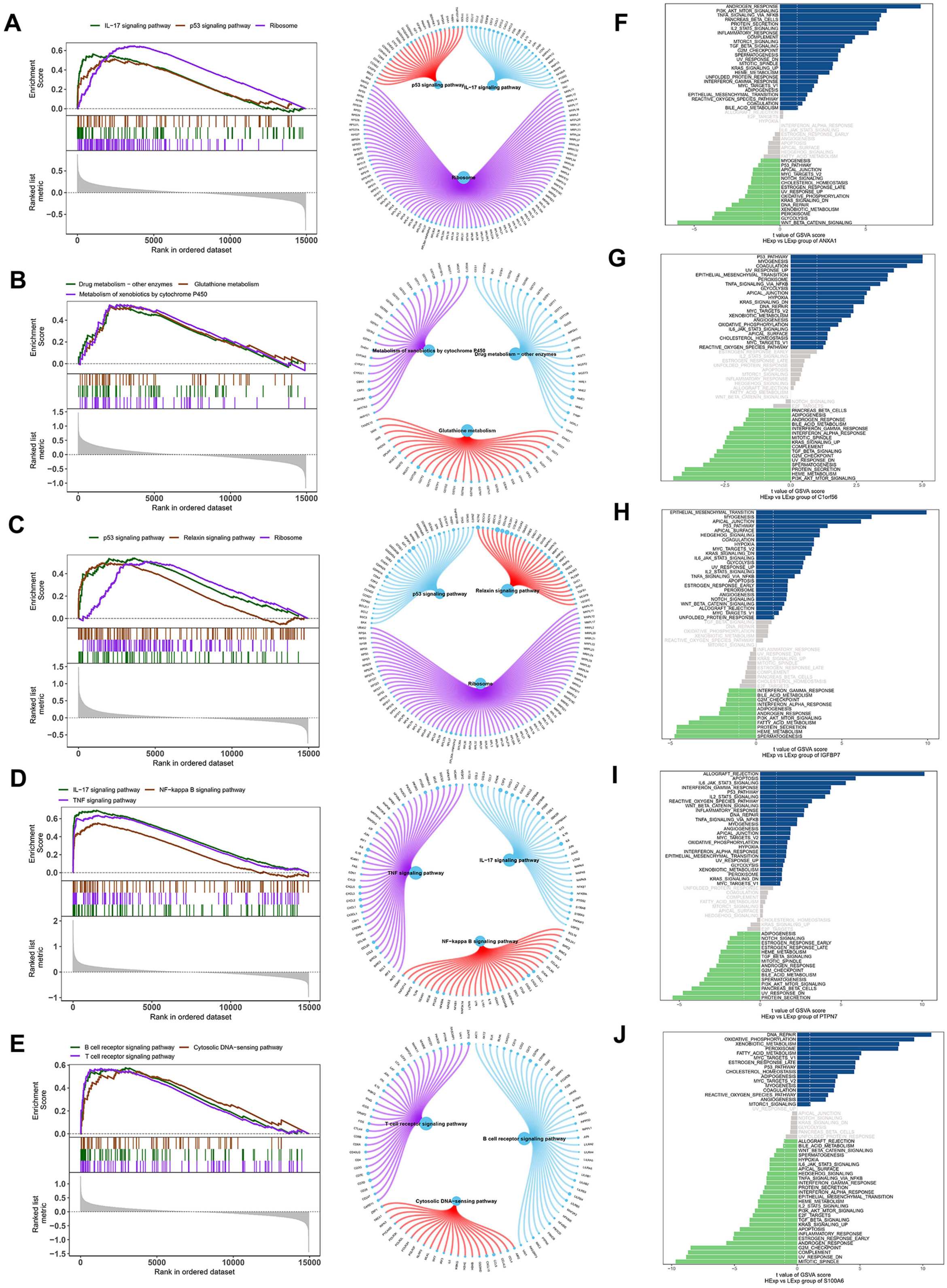

Figure 6 Signaling pathways involved in hub genes. (**A–E**) Gene set enrichment analysis. (**F–J**) GSVA analysis.

miRNA Network Construction of Hub Genes and Related Transcriptional Regulation Analysis

We inversely predicted five hub genes using the mircode database and obtained 66 miRNAs, with a total of 116 pairs of mRNA miRNA relationship pairs, which were visualized using Cytoscape (Figure 7A). In addition, we found that these five hub genes are regulated by multiple transcription factors. These transcription factors were enriched using the

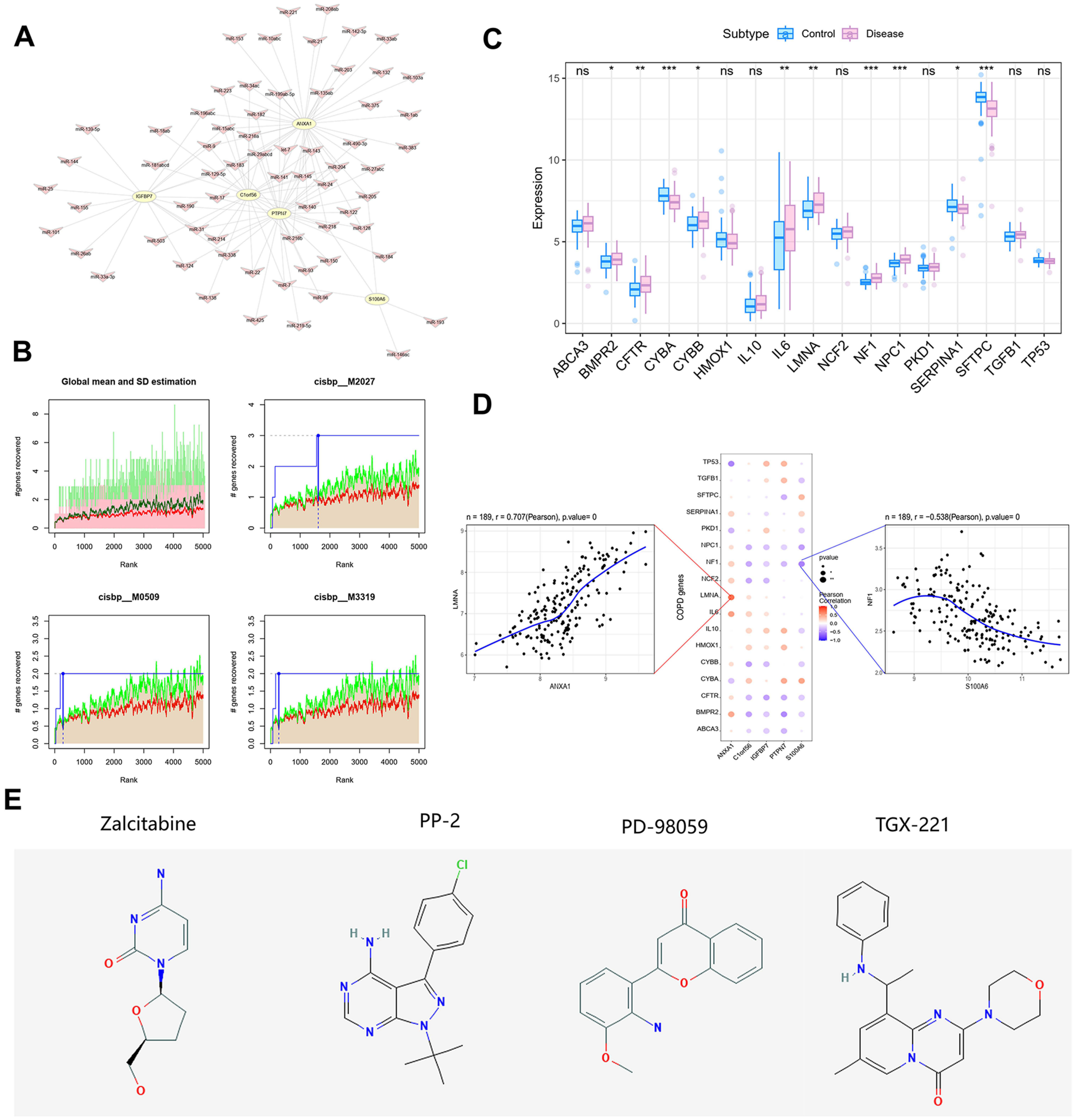

Figure 7 Motiftranscriptional regulation analysis. (**A**) miRNA networks of hub genes, pink for mRNA and purple for miRNA. (**B**) The three motifs with the highest AUC values. The red line is the average of the recovery curve of each motif, the green line is the mean + standard deviation, and the blue line is the recovery curve of the current motif. The maximum distance point (mean + sd) between the current motif and the green curve is the maximum enrichment level selected. (**C**) The expression levels of the top 20 genes between the two groups. Blue represents healthy controls; red represents COPD patients. ns, no significance; *P < 0.05, **P < 0.01,***P < 0.001. (**D**) The expression levels of five hub genes were significantly correlated with the expression levels of multiple disease-related genes. (**E**) Molecular structure diagram of drug.

https://doi.org/10.2147/JIR.S491298

cumulative recovery curves. Motif TF annotation and selection analysis of the important genes showed that the motif with the highest standardized enrichment score (n = 6.45) was cisbp _ M2027 (Figure 7B).

Correlation of Hub Genes With Disease-Modifying Genes

Genes related to COPD were obtained from the GenBank database. We analyzed the expression levels of the top 20 disease-related genes between COPD and control groups. We found that the expressions of BMPR2, CFTR, CYBA, CYBB, IL6, LMNA, NF1, NPC1, serpina1, and SFTPC have a significant correlation between the two groups (Figure 7C). In addition, the expression levels of five hub genes were significantly correlated with these TOP10 disease-related genes, in which ANXA1 was significantly positively correlated with LMNA (r = 0.707) and S100A6 was significantly negatively correlated with NF1 (r = −0.538) (Figure 7D).

CMap Drug Prediction

We divided the TOP50 genes into two groups and performed drug prediction using the Connectivity Map database. The results showed that the expression profiles of drug perturbations, such as zalcitabine, PP-2, PD-98059, and TGX-221, were more significantly negatively correlated with the expression profiles of disease perturbations, suggesting that these drugs could alleviate or even reverse the disease state (Figure 7E).

Discussion

COPD is a complex and persistent respiratory disease that causes a major healthcare burden worldwide every year. In 2015, COPD led to a global death rate of approximately 3.2 million patients, making it the third most common disease among the elderly disease.[23] COPD is usually characterized by consistent respiratory symptoms and airflow limitation due to airway and alveolar abnormalities caused by significant exposure to noxious particles or gases; COPD is usually characterized by consistently respiratory symptoms and airflow limitation.[24] Currently, there is a lack of effective therapies for preventing COPD progression. Therefore, a better understanding of the genetic and biological mechanisms underlying this disease is required for innovative drug development.[25]

NK cells which is a type of innate immune cell in pulmonary immunity had been shown to exhibit hyper-responsiveness in COPD. Several clinical trials have indicated that COPD treatment is related to respiratory infections occurrence have a relationship.[26] Thus, we hypothesized that NK cells may contribute to enhanced respiratory inflammation during COPD exacerbations.

In our study, we chose a single-cell dataset that included three COPD samples for single-cell annotation analysis. We found that NK cell subtype has the highest potential interactions with other immune cell subtypes. This result indicated the importance of NK cell in the pathology of COPD. Previous study had suggested that NK cells play a role in enhancing lung inflammation during influenza-induced COPD exacerbations.[27] Pei et al used single-cell profiling of PBMCs in COPD to identify the connection between impaired immune function and COPD development.[28] These studies were similar to our results.

Subsequently, the GEO database (GSE57148 and GSE8581) from COPD was subjected to various bioinformatic methods. LASSO regression identified 21 genes as single-cell NK_cell-featured genes for COPD, and these 21 genes had good diagnostic efficacy in COPD. Additionally, we used the dataset GSE8581 as a validation set to confirm the same results. Using GO semantic similarity analysis, we found that the hub genes C1orf56, S100A6, IGFBP7, ANXA1, and PTPN7 were the most critical in the whole network.

In our study, we collected the serum of patients with COPD and measured the mRNA expression levels of the hub genes using RT-qPCR. The result found that the mRNA expressions of C1orf56, S100A6, IGFBP7, ANXA1, and PTPN in the normal group were lower than that in the disease group, which indicating the five hub genes may be a favor factor in the development of COPD.

In many single-cell studies, individual cells perform gene expression processes in asynchronous ways, in which each cell was a moment of the transcription process. Monocle was to sort individual cells within pseudotime, utilizing their asynchronous processes to place them on trajectories corresponding to biological processes such as cell differentiation.[29] In our study, the images of cell coloring by pseudotime value (pseudotime is the probability

https://doi.org/10.2147/JIR.S491298

calculated by monocle2 based on cell gene expression information, representing the order of time) and cell type showed that NK cells develop from right to left. The findings further illustrated the importance of NK cells to the immune microenvironment of COPD.

We also analyzed the relationship between hub genes and the developmental trajectory of NK cells. The expression of ANXA1, PTPN7and C1orf56 is lower in the early stages of NK cells than in the mature stage of NK cells, indicating that they may be involved in the maturation of NK cells. While the expression of S100A6 shows a downward trend in time fitting analysis, gradually decreasing from early high expression, suggesting that it may play a role in the early stages of NK cell development. The above results suggest that the functions of hub genes were associated with most NK cell-associated activities.

Immune cell infiltration plays a prominent role in chronic inflammation in COPD.[30] We analyzed the correlation between hub genes and immune infiltration in the COPD dataset to clarify the underlying molecular mechanisms by which hub genes influence COPD progression. After evaluating the expression profile of hub genes in single cells, we found that the hub genes were not only significantly higher in the immune cells of the COPD group than in the normal group but were also correlated with different immune factors.

A previous study has shown that the acrosomal region of spermatozoa possesses C1orf56 immunoreactive signals.[31] S100A6 and ANXA1 are potential biomarkers that are upstream molecules involved in innate immunity in lung fibrosis.[32,33] Li et al found that IGFBP7, an immune-related hub gene, is an independent variable diagnostic biomarker for idiopathic pulmonary fibrosis.[34] PTPN7 expression is a promising predictive biomarker for immunotherapy of multiple cancers.[35] These findings are similar to our experimental results, suggesting that these hub genes also play critical roles in the immune response in COPD.

MicroRNAs (miRNAs) are endogenous small noncoding RNA, consisting of 21–25 nucleotides that combine with the 3′ UTR of target mRNAs to cause mRNA destruction.[36] Numerous studies have shown that dysregulated miRNAs are associated with COPD onset and development of COPD.[37,38] In this study, we used *miRcode* to identify regulatory transcription factors and corresponding binding motifs for hub genes. We constructed an mRNA–miRNA regulatory network, in which miR-9, miR-7, miR-31, and miR-17 had the highest average connectivity among hub genes. Previous study found that miR-7 stimulated macrophage differentiation of COPD exacerbation.[39] According to Chen et al, miR-17 is transported from primary human bronchial epithelial cells, mesenchymal stem cells, and dendritic cells.[40] Through a recovery curves, we predicted the potential transcriptional-binding sites to further explore the potential molecular mechanisms of COPD.[41]

Additionally, we confirmed top 10 genes related to COPD using the GeneCards database. The expression levels of the five hub genes were significantly correlated with the expression levels of multiple disease-related genes, such as ANXA1, LMNA, S100A6 and NF1. Furthermore, our study also predicted potential therapeutic drugs for the five hub genes, Zalcitabine, PP-2, PD-98059, TGX-221, which will support future studies on the treatment of the disease.

However, our study has some limitations. First, our research was based on bioinformatics analysis and prospective clinical studies, which should be performed to verify the prognostic characteristics. Second, we only studied the mRNA expression levels of the hub genes in peripheral blood leukocytes. The protein levels of the hub genes and their molecular mechanisms underlying the impact of NK cell in COPD remain unclear. Third, animal experiments are required to identify potential therapeutic agents. Finally, our single-cell analysis is based on three COPD samples and does not include a control group or grouping based on different disease severity levels. Due to the limitations of this single-cell dataset, we were unable to directly compare immune cells between control and COPD patients and differ the immune cells under different severity levels of COPD.

Conclusion

To sum up, the above analysis shows that immunity plays an important role in the progression of COPD and that immune cell infiltration is a favorable prognostic factor for COPD development. Peripheral NK cells may contribute to the pathogenesis and genetic mechanisms of COPD. These five hub genes may provide insights into mechanistic research and new targets for new therapies for patients with COPD. This provides a reliable, in-depth study of the mechanisms of COPD and the development of novel COPD treatments in the future.

https://doi.org/10.2147/JIR.S491298

Author Contributions

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

Funding

This study was supported by the Natural Science Foundation of Guangdong Province (Grant No. 2022A1515110107).

Disclosure

The authors declare that they have no conflicts of interest in this work.

References

1. Celli B, Fabbri L, Criner G, et al. Definition and Nomenclature of Chronic Obstructive Pulmonary Disease: time for Its Revision. *Am J Respir Crit Care Med*. 2022;206:1317–1325. PMID: 35914087 Free PMC article. No abstract available. doi:10.1164/rccm.202204-0671PP
2. Confalonieri M, Braga L, Salton F, Ruaro B, Confalonieri P. Chronic Obstructive Pulmonary Disease Definition: is It Time to Incorporate the Concept of Failure of Lung Regeneration? *Am J Respir Crit Care Med*. 2023;207(3):366–367.PMID: 36174210. doi:10.1164/rccm.202208-1508LE
3. Atsou K, Chouaid C, Hejblum G. Variability of the chronic obstructive pulmonary disease key epidemiological data in Europe: systematic review. *Bmc Med*. 2011;9:7. doi:10.1186/1741-7015-9-7
4. Christenson SA, Smith BM, Bafadhel M, Putcha N. Chronic obstructive pulmonary disease. *Lancet*. 2022;399(10342):2227–2242. doi:10.1016/S0140-6736(22)00470-6
5. Celli BR, Wedzicha JA. Update on Clinical Aspects of Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2019;381(13):1257–1266. doi:10.1056/NEJMra1900500
6. NICE. *Inhaled Triple Therapy: Chronic Obstructive Pulmonary Disease in Over 16s: Diagnosis and Management: Evidence Review I*. London: National Institute for Health and Care Excellence (NICE); 2019.
7. Wedzicha JA, Wilkinson T. Impact of chronic obstructive pulmonary disease exacerbations on patients and payers. *Proc Am Thorac Soc*. 2006;3(3):218–221. doi:10.1513/pats.200510-114SF
8. Singh D, Agusti A, Anzueto A, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. *Eur Respir J*. 2019;53(5):1900164. doi:10.1183/13993003.00164-2019
9. Lin YZ, Zhong XN, Chen X, Liang Y, Zhang H, Zhu DL. Roundabout signaling pathway involved in the pathogenesis of COPD by integrative bioinformatics analysis. *Int J Chron Obstruct Pulmon Dis*. 2019;14:2145–2162. doi:10.2147/COPD.S216050
10. Huang X, Li Y, Guo X, et al. Identification of differentially expressed genes and signaling pathways in chronic obstructive pulmonary disease via bioinformatic analysis. *Febs Open Bio*. 2019;9(11):1880–1899. doi:10.1002/2211-5463.12719
11. Yu H, Guo W, Liu Y, Wang Y. Immune Characteristics Analysis and Transcriptional Regulation Prediction Based on Gene Signatures of Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis*. 2021;16:3027–3039. doi:10.2147/COPD.S325328
12. Deng M, Yin Y, Zhang Q, Zhou X, Hou G. Identification of Inflammation-Related Biomarker Lp-PLA2 for Patients With COPD by Comprehensive Analysis. *Front Immunol*. 2021;12:670971. doi:10.3389/fimmu.2021.670971
13. Duvall MG, Barnig C, Cernadas M, et al. Natural killer cell-mediated inflammation resolution is disabled in severe asthma.National Heart. *Lung Blood Institutes Severe Asthma Res Program-3 Investigators Sci Immunol*. 2017;2(9):1.
14. Lin SJ, Yan DC, Lee WI, Kuo ML, Hsiao HS, Lee PY. Effect of azithromycin on natural killer cell function. *Int Immunopharmacol*. 2012;13(1):8–14. doi:10.1016/j.intimp.2012.02.013
15. Nabekura T, Girard JP, Lanier LLJI. IL-33 receptor ST2 amplifies the expansion of NK cells and enhances host defense during mouse cytomegalovirus infection.. *Journal of Immunology (Baltimore, Md.: 1950)*. 2015;194(12):5948–5952. doi:10.4049/jimmunol.1500424
16. Zheng Y, Yang X. Yang X.Spatial RNA sequencing methods show high resolution of single cell in cancer metastasis and the formation of tumor microenvironment. *Biosci Rep*. 2023;43(2):BSR20221680. doi:10.1042/BSR20221680
17. Sevilla JL, Segura V, Podhorski A, et al. Correlation between gene expression and GO semantic similarity. *IEEE/ACM Trans Comput Biol Bioinform*. 2005;2(4):330–338. doi:10.1109/TCBB.2005.50
18. Jain S, Bader GD. An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology. *BMC Bioinf*. 2010;11:562. doi:10.1186/1471-2105-11-562
19. Guo X, Shriver CD, Hu H, Liebman MN. Analysis of metabolic and regulatory pathways through Gene Ontology-derived semantic similarity measures. *AMIA Annu Symp Proc*. 2005;2005:972.
20. Tedder PM, Bradford JR, Needham CJ, McConkey GA, Bulpitt AJ, Westhead DR. Gene function prediction using semantic similarity clustering and enrichment analysis in the malaria parasite Plasmodium falciparum. *Bioinformatics*. 2010;26(19):2431–2437. doi:10.1093/bioinformatics/btq450
21. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. 2010;26(7):976–978. doi:10.1093/bioinformatics/btq064
22. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. *Bioinformatics*. 2007;23(10):1274–1281. doi:10.1093/bioinformatics/btm087
23. Global. regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1459–1544.

24. Uwagboe I, Adcock IM, Lo BF, Caramori G, Mumby S. New drugs under development for COPD. *Minerva Med*. 2022;113(3):471–496. doi:10.23736/S0026-4806.22.08024-7
25. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med*. 2017;195(5):557–582. doi:10.1164/rccm.201701-0218PP
26. Folli C, Chiappori A, Pellegrini M, et al. COPD treatment: real life and experimental effects on peripheral NK cells, their receptors expression and their IFN-γ secretion. *Pulm Pharmacol Ther*. 2012;25(5):371–376. doi:10.1016/j.pupt.2012.06.009
27. Wortham BW, Eppert BL, Motz GT, et al. NKG2D mediates NK cell hyperresponsiveness and influenza-induced pathologies in a mouse model of chronic obstructive pulmonary disease. *J Immunol*. 2012;188(9):4468–4475. doi:10.4049/jimmunol.1102643
28. Pei Y, Wei Y, Peng B, et al. Combining single-cell RNA sequencing of peripheral blood mononuclear cells and exosomal transcriptome to reveal the cellular and genetic profiles in COPD. *Respir Res*. 2022;23(1):260. doi:10.1186/s12931-022-02182-8
29. Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med*. 2020;52(9):1428–1442. doi:10.1038/s12276-020-0420-2
30. Agustí A, Hogg JC. Update on the Pathogenesis of Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2019;381(13):1248–1256. doi:10.1056/NEJMra1900475
31. Wang Y, Zhao W, Mei S, et al. Identification of Sialyl-Lewis(x)-Interacting Protein on Human Spermatozoa. *Front Cell Dev Biol*. 2021;9:700396. doi:10.3389/fcell.2021.700396
32. Landi C, Bargagli E, Carleo A, et al. Bronchoalveolar lavage proteomic analysis in pulmonary fibrosis associated with systemic sclerosis: S100A6 and 14-3-3ε as potential biomarkers. *Rheumatology*. 2019;58(1):165–178. doi:10.1093/rheumatology/key223
33. Jia Y, Morand EF, Song W, Cheng Q, Stewart A, Yang YH. Regulation of lung fibroblast activation by annexin A1. *J Cell Physiol*. 2013;228(2):476–484. doi:10.1002/jcp.24156
34. Li K, Liu P, Zhang W, et al. Bioinformatic identification and analysis of immune-related chromatin regulatory genes as potential biomarkers in idiopathic pulmonary fibrosis. *Ann Transl Med*. 2022;10(16):896. doi:10.21037/atm-22-3700
35. Wang F, Wang X, Liu L, et al. Comprehensive analysis of PTPN gene family revealing PTPN7 as a novel biomarker for immuno-hot tumors in breast cancer. *Front Genet*. 2022;13:981603. doi:10.3389/fgene.2022.981603
36. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*. 2004;5(7):522–531. doi:10.1038/nrg1379
37. Conickx G, Mestdagh P, Avila CF, et al. MicroRNA Profiling Reveals a Role for MicroRNA-218-5p in the Pathogenesis of Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2017;195(1):43–56. doi:10.1164/rccm.201506-1182OC
38. Mohamed A, Kunda NK, Ross K, Hutcheon GA, Saleem IY. Polymeric nanoparticles for the delivery of miRNA to treat Chronic Obstructive Pulmonary Disease (COPD). *Eur J Pharm Biopharm*. 2019;136:1–8. doi:10.1016/j.ejpb.2019.01.002
39. Jiang Y, Wang J, Zhang H, Min Y, Gu T. Serum Exosome-Derived MiR-7 Exacerbates Chronic Obstructive Pulmonary Disease by Regulating Macrophage Differentiation. *Iran J Public Health*. 2023;52(3):563–574. doi:10.18502/ijph.v52i3.12139
40. Chen J, Hu C, Pan P. Extracellular Vesicle MicroRNA Transfer in Lung Diseases. *Front Physiol*. 2017;8:1028. doi:10.3389/fphys.2017.01028
41. Kuret K, Amalietti AG, Jones DM, Capitanchik C, Ule J. Positional motif analysis reveals the extent of specificity of protein-RNA interactions observed by CLIP. *Genome Biol*. 2022;23(1):191. doi:10.1186/s13059-022-02755-2