



OPEN RNA m7G methylation regulators and targets significantly contribute to chronic obstructive pulmonary disease

Chenyu Zhu¹, Luyi Tan¹, Xinyu Zhang¹, Wenli Cheng¹, Min Li¹, Yibo Chen²✉ & Wenjuan Zhang¹

Chronic obstructive pulmonary disease (COPD) is one of the most common lung injury diseases, closely associated with aging, air pollution and smoking exposure. The novel epigenetic modification 7-methylguanosine (m7G) RNA methylation affects the pathogenesis and progression of COPD. In this study, the combined roles of m7G methylation regulators were explored in COPD for the first time by integrated bioinformatic methods. The machine algorithms screened 7 disease signature genes relevant to clinical indicators, including CYFIP2, EIF3D, EIF4G3, GEMIN5, METTL1, SNUPN and NCBP2, and METTL1 was related to the progression in COPD. COPD patients could be well divided into two m7G subtypes by consensus clustering, and the two groups had differential immune profiles, visualized by single-cell sequencing and immune infiltration landscapes. More importantly, CAT was found to be a meaningful key target gene in METTL1-CAT axis for m7G methylation in COPD. We also used the cell premature senescence model for the preliminary validation of the above biosignature analysis results. The qRT-PCR and GSEA results revealed the important regulatory roles of the seven disease signature genes in COPD and aging-related diseases. Taken together, METTL1 and its target CAT have played an important role in COPD, as excellent candidates for its prevention and intervention.

Keywords COPD, m7G, METTL1, CAT, Bioinformatics analysis

Abbreviations

m7G	7-Methylguanosine
BP	Biological process
CAT	Catalase
CC	Cellular component
COPD	Chronic obstructive pulmonary disease
CYFIP	Cytoplasmic FMR1 interacting protein
DCPS	Decapping enzyme, scavenger
DEGs	Differentially expressed genes
DO	Disease ontology
EIF3D	Eukaryotic translation initiation factor 3 subunit D
EIF4G3	Eukaryotic translation initiation factor 4 gamma 3
EIF4E	Eukaryotic translation initiation factor 4E
EIF4E3	Eukaryotic translation initiation factor 4E family member 3
GEMIN5	Gem nuclear organelle associated protein 5
GO	Gene ontology
GSEA	Gene set enrichment analysis
GSVA	Gene set variation analysis
HEFs	Human embryonic lung fibroblasts
IPF	Idiopathic pulmonary fibrosis
iRFA	Insufficient radiofrequency ablation

¹Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, Guangdong, China. ²Rice Research Institute, Guangdong Academy of Agricultural Sciences/Guangdong Key Laboratory of New Technology in Rice Breeding, Guangzhou, Guangdong, China. ✉email: cheniyibo@gdaas.cn; zwj2080@126.com

IFIT5	Interferon-induced protein with tetratricopeptide repeats 5
IL-17	Interleukin 17
KEGG	Kyoto encyclopedia of genes and genomes
LARP1	La ribonucleoprotein 1, translational regulator
LASSO	Least absolute shrinkage and selection operator
LSM1	LSM1 homolog, mRNA degradation associated
METTL1	Methyltransferase like 1
MF	Molecular function
NCBP	Nuclear cap binding protein subunit
NUDT10	Nudix hydrolase 10
PS	Premature senescence
PCA	Principal component analysis
PPI	Protein-protein interaction
RF	Random forest
RT-qPCR	Real-time quantitative polymerase chain reaction
STRING	Search tool for the retrieval of interacting genes
ssGSEA	Single sample gene set enrichment analysis
SNUPN	Snurportin 1
SVM-RFE	Support vector machine-recursive feature elimination
tSNE	T-distributed stochastic neighbor embedding
Th1	Type 1T helper
Th2	Type 2T helper
WDR4	WD repeat domain 4

Chronic obstructive pulmonary disease (COPD) was the third leading cause of death in the world¹, affecting approximately 384 million people with low quality of life and lifespan, with a vast social economy and healthcare burden, especially in developing countries. COPD was caused by combined pathophysiological processes, including lung dysplasia, slow lung growth in early life and airway reshaping-related damage, resulted from harmful environmental factors and unhealthy behaviors². The proportion of Chinese COPD patients was approximately 25% of all COPD patients worldwide. From 2012 to 2015, COPD had the fourth prevalence in city and the third in countryside of China, and its prevalence among people aged 40 years and older increased by 67%³.

RNA epigenetic modifications were fundamental in many physiological and pathological processes⁴. More than 100 RNA modifications had been identified, including N6-methyladenosine, 5-methylcytosine, 1-methyladenosine, and 7-methylguanosine (m7G)⁵. With the development of high-throughput sequencing technologies, methods for detecting m7G had been updated, including m7G-MeRIP-Seq, m7G-Seq, m7G-miCLIP-Seq and TRAC-seq⁶, causing increased attention to the association between m7G and diseases. The m7G methylation was a widespread and evolutionarily conserved RNA modification, initially found at the 5' cap of mRNA, and played an essential role in mRNA export, translation, transcriptional elongation, splicing and protein translation⁷, also widely found in tRNA, rRNA and microRNA, for maintaining tRNA stability and regulating intranuclear processing and maturation of 18SrRNA⁸, microRNA biosynthesis⁹, and biological functions¹⁰, with a positively charged modification of mRNA¹¹. The m7G cap-binding protein recruited proteins and mediated cap-related biological functions, including transcription elongation, pre-mRNA splicing, nuclear export and cap-dependent protein synthesis¹², in addition to stabilizing transcripts to prevent extra-nucleotide degradation and regulating the mRNA life cycle¹³.

Methyltransferase like 1 (METTL1) was extensively in mammals as a writer of m7G modification, typically bound to WD repeat domain 4 (WDR4) to form a methyltransferase complex catalyzing m7G modifications in eukaryotic tRNAs¹⁴. METTL1 installed m7G modifications in target mRNAs, and WDR4 promoted their binding of the heterodimeric complex¹⁵. The METTL1/WDR4 methyltransferase complex was associated with many diseases, including dwarfism, Down's syndrome, Galloway-Mowat syndrome and various tumors, etc.¹⁶. The cytoplasmic FMR1-interacting protein families cytoplasmic FMR1 interacting protein 1 (CYFIP1) and cytoplasmic FMR1 interacting protein 2 (CYFIP2) regulated mRNA translation and transport¹⁷. CYFIP2 mutation was associated with Alzheimer's disease, early-onset epileptic encephalopathies, mental retardation and developmental delays¹⁸. The eukaryotic initiation factor family, the NCBP family, and gem nuclear organelle associated protein 5 (GEMIN5) were all m7G cap-binding proteins, playing key roles in mRNA processing, nuclear export, and translation^{19,20}. Eukaryotic translation initiation factor 3 was a eukaryotic translation initiation factor with an important role in translation regulation²¹. Eukaryotic translation initiation factor 3 subunit D (EIF3D) had an oncogenic role in prostate cancer, glioma, melanoma, colon cancer and lung cancer^{22–26}. Knockdown of *EIF3D* inhibited cell proliferation and clone formation^{27,28}. Nuclear cap binding protein subunit 2 (NCBP2) caused cell cycle arrest in colon adenocarcinoma cells, with poor prognosis of colon adenocarcinoma patients²⁹. Snurportin 1 (SNUPN) also took part in nuclear translocation. Nudix hydrolases belonged to RNA decapacitating enzyme regulating vital biological processes such as precursor mRNA processing, cell proliferation, signaling and homeostasis in vivo^{30,31}.

Abnormal expression of m7G-related genes was closely associated with the development and progression of lung diseases. RNA m7G-related genes, including WDR4, la ribonucleoprotein 1, translational regulator (LARP1) and Nuclear cap binding protein subunit 1 (NCBP1), might provide new biomarkers for lung adenocarcinoma prognosis³². The m7G methylation seemed to be a double-edged sword in lung tumors, modifying different RNAs with different effects. The m7G-modified tRNA promoted lung cancer cell growth and invasion via the METTL1/WDR4 complex³³, while the m7G-modified miRNA let-7e suppressed lung cancer cell metastasis

via METTL1¹⁰. The m7G modifications were also closely associated with novel coronavirus pneumonia 2019. Six m7G cap-binding proteins, nudix hydrolase 4B, interferon-induced protein with tetratricopeptide repeats 5 (IFIT5), *LARP1*, eukaryotic translation initiation factor 4E (EIF4E), LSM1 homolog, mRNA degradation associated (LSM1), and nudix hydrolase 4, were used to estimate COVID-19 disease risk³⁴. The m7G methylation had the potential action mechanisms in pulmonary arterial hypertension and the immune microenvironment. Cytoplasmic FMR1 interacting protein 1, EIF4E, and *IFIT5* were the potential novel biomarkers in pulmonary arterial hypertension (PAH)³³.

Currently, studies on the role of RNA m7G methylation regulation in lung diseases are mainly focused on lung cancer. In addition, RNA m7G methylation is also closely associated with other lung diseases, such as PAH³⁵. COPD progression might be closely associated with m7G. The current mechanisms related to m7G methylation mainly focused on cancer in lung diseases, whereas the important roles still remain unclear in COPD. Therefore, we used cluster analysis to identify different categories of m7G patterns and found gene features in different patterns. Then we focused on the disease signature genes screened by the three machine learning methods to explore the association between the signature genes and the clinical indicators of COPD. The single sample gene set enrichment analysis (ssGSEA) and single-cells were also performed to identify the immune microenvironment associated with COPD, and finally we used relevant experimental data for preliminary validation. This study aimed to explore the role and potential targets of m7G methylation and the mediation mechanisms of m7G regulators, providing new evidences and essential clues for epigenetic variations, subtype classification and immunotherapeutic strategies in COPD.

Materials and methods

Data collection

Gene expression data, GSE76925 and GSE196638, were obtained from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). GSE76925 included lung tissue samples from 111 patients with COPD and 40 normal controls. GSE196638 was used to verify the results at the single-cell level. The data were publicly provided by the GEO database with preexisting ethics approval from original studies.

Dataset baseline and identification of differentially expressed genes

The clinical information for the samples was obtained from the GSE76925 dataset. The R package “tableone” (version 0.13.2) was used to derive a table of base characteristics. A baseline investigation of sex, age, BMI and clinical indicators in this dataset is shown in Table 1. Then, m7G-related genes were identified^{36,37}, which included “m7G(5')pppN diphosphatase activity”, “RNA 7-methylguanosine cap binding”, and “RNA cap binding” from the Molecular Signatures Database (<https://www.gsea-msigdb.org/gsea/msigdb/search.jsp>). Twenty-five m7G methylation-related genes were screened and shown in Supplementary Table S1. Then, the “limma” package (version 3.54.1) in R statistical software was used to identify the differential m7G methylation-related genes between the normal and COPD groups in the GSE76925 data. $p < 0.05$ was considered statistically significant.

Characteristics	Control N = 40	COPD N = 111	p value
Sex (%)			
Male	15 (37.50)	52 (46.85)	0.404
Female	25 (62.50)	59 (53.20)	
Age	65.72 (9.01)	63.30 (6.61)	0.074
Age (%)			
< 45	1 (2.63)	1 (0.90)	0.266
45–54	2 (5.26)	8 (7.21)	
55–64	12 (31.58)	56 (50.45)	
≥ 65	23 (60.53)	46 (41.44)	
FEV1/FVC	0.79 (0.05)	0.32 (0.10)	< 0.001
FEV1%	98.84 (12.84)	26.51 (9.38)	< 0.001
BMI	28.08 (5.70)	25.62 (4.62)	0.008
BMI (%)			
< 18.5	1 (2.50)	5 (4.50)	0.100
18.5–24.9	10 (25.00)	48 (43.20)	
25–29.9	14 (35.00)	36 (32.40)	
30–34.9	12 (30.00)	20 (18.00)	
≥ 35	3 (7.50)	2 (1.80)	

Table 1. Baseline information in the GSE76925 dataset. FEV1 forced expiratory volume in one second, FVC forced vital capacity, BMI body mass index.

Identification of molecular subtypes based on significant m7G regulators

Cluster analysis was the technique of partitioning a piece of data into numerous groups based on higher similarity within the same category and lower similarity between categories in order to investigate patterns and motifs in the data between categories. Consensus clustering was used to identify each member and their subgroup number and verify the rationality of clustering, discriminating different m7G patterns based on essential m7G regulators according to the R package “ConsensusClusterPlus” (version 1.62.0). The heatmap and boxplot were used to identify the differential m7G regulatory factors between cluster A and cluster B. The results of the heat map of the consistency matrix show that a matrix color of white means that it is impossible to cluster together and a color of dark blue means that it always clusters together. The colors white and blue-white mixed together are not chosen as much as possible for the clustering values (Figure S1).

Screening of disease signature genes and gene set variation analysis (GSVA)

Three machine learning algorithms, including LASSO, RF and SVM-RFE, were used to screen disease-characteristic genes. LASSO regression was performed using the “glmnet” R package (version 4.1-6), and tenfold cross-validation was used to analyze COPD and normal samples. The LASSO regression reduced the coefficient estimates to zero, with the degree of reduction depending on the additional parameter λ . The optimal value of λ was determined by using tenfold cross-validation. A random forest model was built by using the “randomForest” R package (version 4.7-1.1). The best model was developed using the training set to find the point with the minimum error. Genes with importance scores higher than 2 were selected for inclusion in the subsequent studies. In addition, the SVM-RFE model used the “kernlab” R package (version 0.9-31) to determine the number of best-ranked genes. The SVM-RFE method was advantageous in efficiently selecting relevant features and removing redundant features. SVM-RFE was used for feature selection and ten-fold cross-validation was used to determine which genes should be considered typical. Then, violin and boxplots were used to explore the relationship between the signature genes and the clinical data. GSVA was a non-parametric, unsupervised analysis method to assess whether different pathways were enriched across samples by transforming the expression matrix of a gene across samples into the expression matrix of a gene set across samples using the “GSVA” package (version 1.60.0).

Screening and functional enrichment analysis of differentially expressed genes (DEGs) and protein–protein interaction (PPI) network construction

DEGs of the models were identified between COPD and normal using the “limma” R package by the following statistical criteria: (1) $|\log_2^{FC}| > 2$; (2) $p < 0.05$. The 145 hub COPD genes were obtained from published literature³⁸. The COPD-associated core genes were included by associated analysis of a large number of candidate genes in human COPD and related phenotypes. Most of the genes played potential biological roles in COPD pathophysiology and were strongly associated with COPD, COPD severity and the related phenotypes. Then, the key COPD genes were acquired by taking the intersection of the overlapping DEGs and hub COPD genes (Supplementary Table S3). To analyze the interaction between the significant m7G regulators and essential COPD genes, the PPI network was constructed by search tool for the retrieval of interacting genes (STRING) online.

Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis

The statistically significant DEGs were then subjected to GO and KEGG pathway enrichment analysis in R software using the “clusterProfiler” package (version 4.6.0). The cut-off criterion was $p_{\text{adjust}} < 0.05$.

The ssGSEA immunoassay and evaluation of immune cell infiltration

The ssGSEA was performed by the “GSVA” package in R (version 1.60.0) to assess immune cell abundance in COPD patients. The gene expression levels in the samples were first sorted using ssGSEA. Next, these genes were searched in the input dataset, and the expression levels were summarized. Based on the above evaluation, an abundance of immune cells was obtained in each sample. CIBERSORT was generated by the “e1071” R package (version 1.7-13) to estimate the abundance of immune cells and to visualize differences in immune cell infiltration.

Single-cell RNA sequencing data processing

The single-cell expression matrix came from the GSE196638 dataset, and the gene expression was normalized using the LogNormalize method. Dimensionality was reduced by PCA to reduce variance. Cell clusters were visualized by the t-distributed stochastic neighbor embedding (tSNE) method. Cell type annotations were generated using the “SingleR” package (Version 2.2.0) for automated classification, and then we validated the manual annotations using marker genes, which showed good consistency between the two types of annotations. And the automated annotations were ultimately selected for inclusion in the subsequent analyses due to the highly subjective properties of the manual annotations.

Disease ontology (DO) analysis

DO analyses were performed by the “DOSE” package (version 4.4) to analyze differential gene-disease correlations with enrichment. The top 20 relevant diseases with the highest significance were visualized using boxplot and bubble plots, and the intersection between the diseases was shown using upset plot.

Cell culture and real-time quantitative polymerase chain reaction (qRT-PCR)

Based on our cell premature senescence (PS) model of the human embryonic lung fibroblasts (HEFs) for the preliminary validation^{39,40}. HEFs were obtained from the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, and cultured with 10% foetal bovine serum, 100 U/mL penicillin, and 0.1 mg/mL streptomycin (Gibco, Grand Island, NE, USA) in low-sugar Dulbecco's modified Eagle medium (L-DMEM, Gibco, Grand Island, NE, USA). Constant temperature incubator (Thermo Fisher, Waltham, MA, USA) conditions were strictly controlled at 37 °C, 95% relative humidity and 5% CO₂. We exposed 22PDL (control) HEFs to 400 μmol/L H₂O₂ (BDH Chemicals Ltd, Poole, UK) for 2 h per day for 4 consecutive days to obtain the PS group. Total RNA was extracted by Trizol reagent (Invitrogen, Carlsbad, CA, USA). The cDNA was synthesised from total RNA using PrimeScript premix (TaKaRa, Kyodo, Japan), and synthesised from total RNA using the SYBR Premix Ex Taq™ II kit (Tli RNaseH Plus. TaKaRa, Kyodo, Japan) and a fluorescent quantitative PCR instrument (CFX connect, BIO-RAD, Hercules, CA, USA) to quantify the mRNA levels. The primers used were shown in Table 2.

RIP-Q-PCR

The cell precipitates were lysed using Polysome lysis buffer, Protease inhibitor and RNase inhibitor; DNase salt stock and DNase were added and incubated for 10 min, and the DNA was removed by EDTA, DTT and 0.5 M EGTA; the supernatants collected above were divided into The supernatant collected above was divided into two samples of IP group and Input group respectively; the IP group was incubated with IP-grade METTL1 primary antibody for more than 16 h; after equilibrating the magnetic beads, the samples of each group were washed; RNA samples of the IP group and Input group were extracted and amplified on the machine. The %Input content can be calculated based on the following RIP-Q-PCR formula:

$$\Delta Ct \text{ [normalized IP]} = (Ct \text{ [IP]} - (Ct \text{ [Input]} - \text{Log2 (Input Dilution Factor)}))$$
$$\%Input = 2^{(-\Delta Ct \text{ [normalized IP]})} \times 100\%$$

Gel electrophoresis

Prepare a 2% agarose gel, add 6 μL of nucleic acid dye, solidify at room temperature; one side of the spiking well is located in the negative pole, pour 1×TBE buffer so that its liquid height is higher than that of the gel; according to the ratio of 1:5, add 6×DNA up-sampling buffer to the RT-qPCR amplified product; electrophoresis conditions are 120 V, 20 min. place the gel in the gel imager, turn on the switch, adjust the computer coefficients, and photographed to save the developed image.

Statistical analysis

All data processing in the study was dependent on R version 4.2.3. The correlations within the continuous variables were analyzed by Pearson's test. The *p*<0.05 was considered statistically significant.

Results

Differential analysis of 25 m7G RNA methylation regulators in COPD

As shown in Fig. 1A,B, differential expression analysis of 25 m7G RNA methylation regulators was shown between COPD patients and control samples, and 15 kinds of m7G regulators had significant differences. Among the set of 15 genes, CYFIP1, CYFIP2, decapping enzyme, scavenger (DCPS), EIF3D, eukaryotic translation initiation factor 4E family member 3 (EIF4E3), eukaryotic translation initiation factor 4 gamma 3 (EIF4G3),

Primer name	Sequence
GAPDH-F	GTCTCCTCTGACTTCAACAGCG
GAPDH-R	ACCACCCTGTTGCTGTAGCCAA
CYFIP2-F	CCACCACCAACTGAAGGACATCATC
CYFIP2-R	TCTATGAGGAGGCAGAACAGGATGG
EIF3D-F	CAACAAGCAGGTCATCCGAGTCTAC
EIF3D-R	CCTCCTCTTCCTCCTCATCCTCTTC
GEMIN5-F	CATTGGCTACAAGGATGGCA
GEMIN5-R	GTTGGTAATTTAGCTTCTTCTGA
METTL1-F	CCGACCCACATTTCAGCG
METTL1-R	TCCAGCACATCGGTTATGGTA
NCBP2-F	AGCGTGTGGGTTCTGTTTCGTG
NCBP2-R	CATACTGCCTGCCCTCCTTAAAGC
SNUPN-F	TATCGTGGCCTCCAGGGGTT
SNUPN-R	GTTGAGTTTCGCCTGTTGCC
CAT-F	CGCCTGTGTGAGAACATTGC
CAT-R	TGTGAATGGCGTTCTTTGGT

Table 2. The primer sequences used in this study for qPCR.

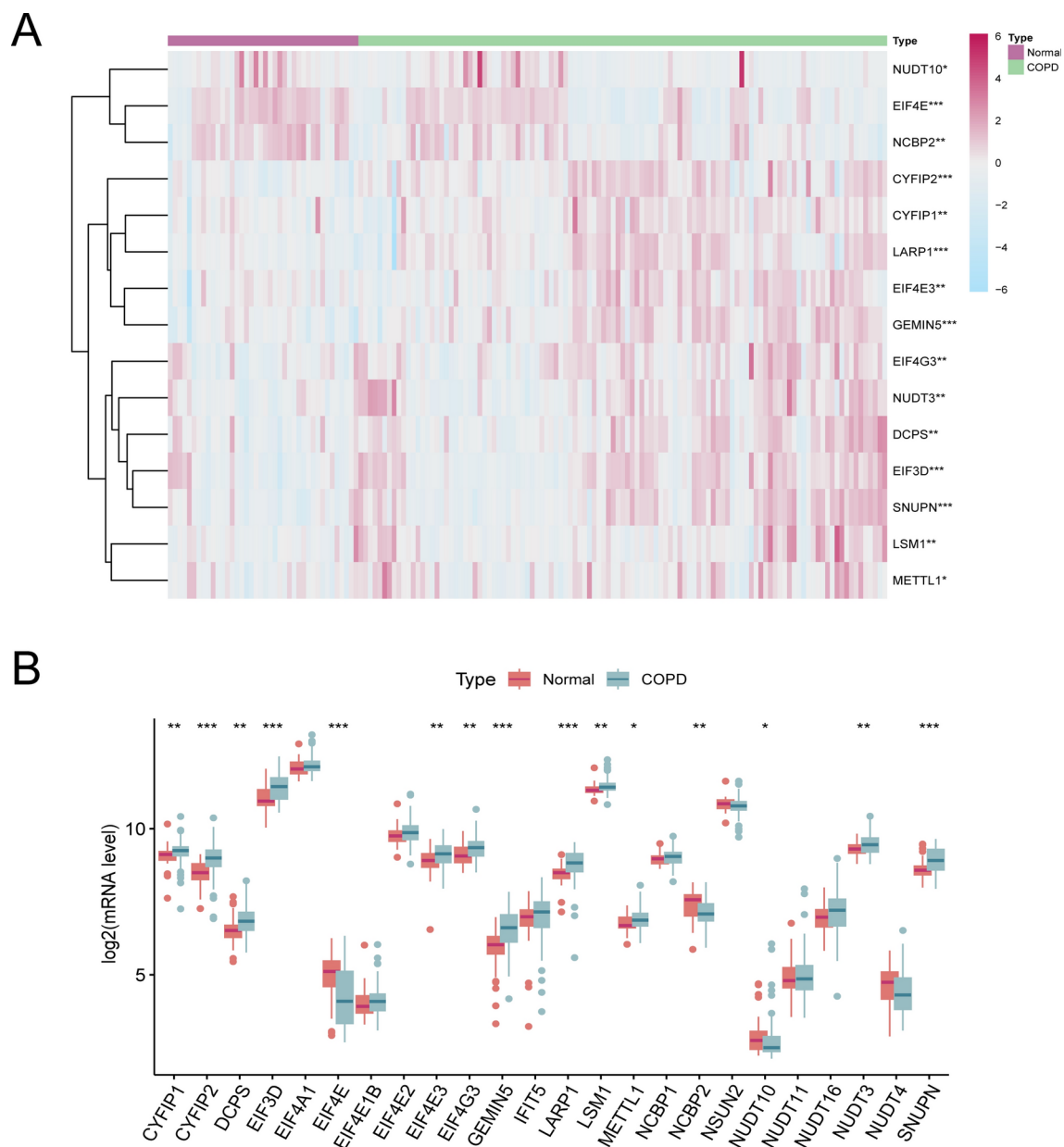


Fig. 1. Gene differential expression analysis. (A) Differential expression of m7G-regulated genes in COPD patients and normal controls. (B) Differential expression boxplot of the 25 identified m7G regulators. The t test was used to analyze the difference in the expression levels of m6A regulators. Compared with the controls, * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

GEMIN5, LARP1, LSM1, METTL1, nudix hydrolase 3 (NUDT3), and SNUPN were overexpressed in COPD patients compared with those in normal controls, while EIF4E, NCBP2, and nudix hydrolase 10 (NUDT10) were downregulated. Therefore, the expression levels of significant m7G regulatory factors were generally upregulated, and m7G regulatory factors played an important role in the development of COPD.

Two distinct m7G patterns identified from significant regulators

Correlation analysis was performed of 15 kinds of m7G regulators with significant differences as shown in Fig. 2A,B. There were different degrees of correlation among these 15 kinds of m7G regulators. The EIF3D and SNUPN genes exhibited strong synergistic effects ($r = 0.76$), whereas the EIF3D and EIF4E genes showed an antagonistic effect ($r = -0.81$). In addition, the CYFIP2 gene was positively correlated with EIF3D ($r = 0.39$) and GEMIN5 (correlation coefficients = 0.33), whereas the NCBP2 gene was negatively correlated with both SNUPN ($r = -0.60$) and EIF4G3 ($r = -0.41$). From the network diagram, significant m7G regulators affected COPD disease progression via synergistic interactions with each other.

To classify m7G methylation patterns, we performed a consensus unsupervised cluster analysis of these 15 kinds of m7G regulators. According to Fig. 2C and Fig. S1, the difference between the two clusters was greatest

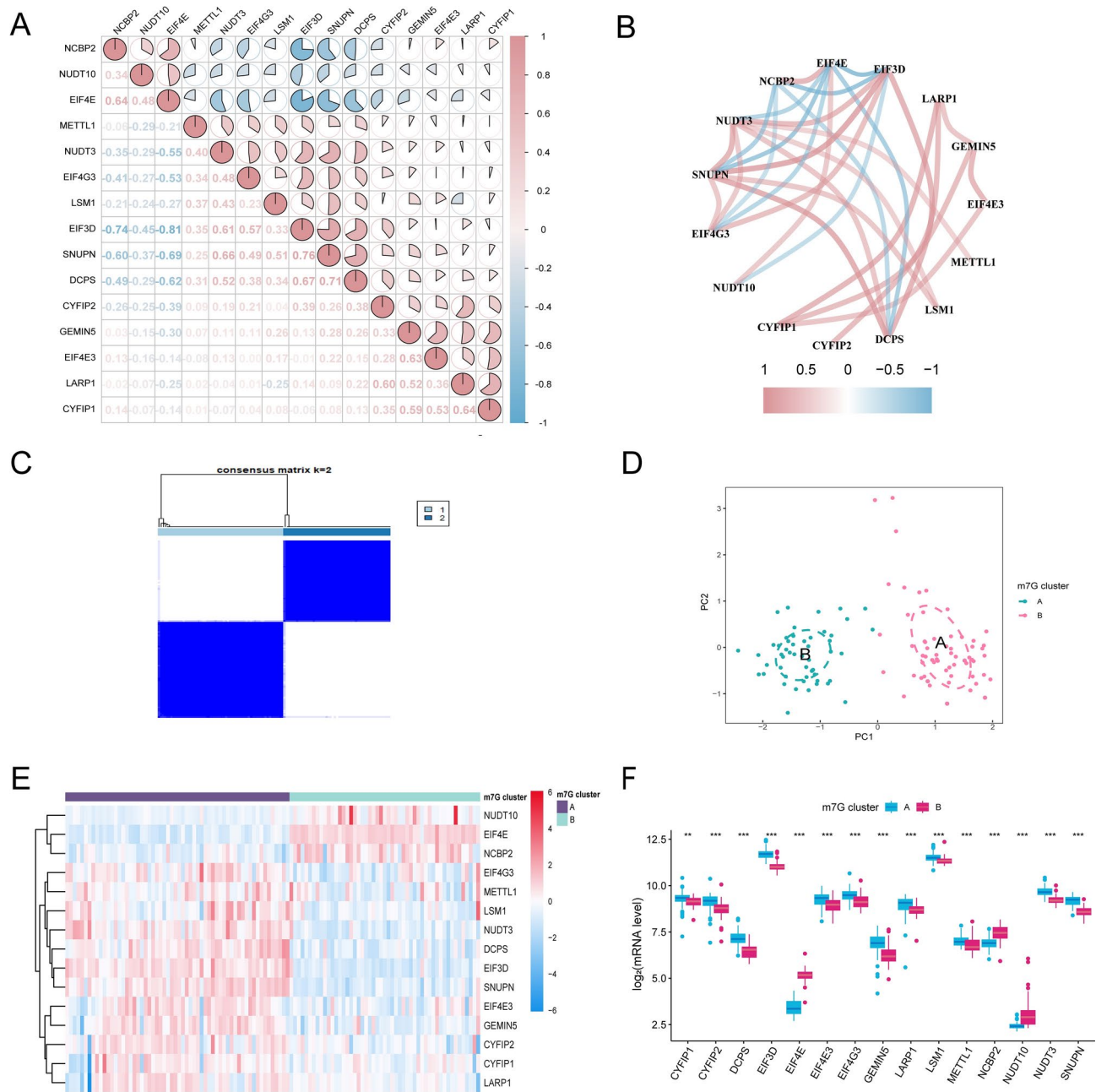


Fig. 2. Consistent clustering of m7G genes. **(A)** Correlations and correlation coefficients among 15 m7G methylation-related genes. **(B)** Correlation network of 15 m7G methylation-related genes. **(C)** Consensus matrices of the 15 significant m7G regulators for $k=2$. **(D)** PCA. Different colors represent different groups. As shown in the diagram, groups A and B are separated without any intersection. **(E)** Heatmap of the distribution of 15 m7G methylation-related genes between cluster A and cluster B. **(F)** Expression of 15 kinds of m7G regulators in clusters. The abscissa represents m7G differential genes, and the ordinate represents the expression distribution of these related genes. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

when $k=2$ and the difference between each clustered sample was smallest and uniform and the clusters obtained more meaningful for subsequent analysis. Based on the consensus clustering, we got two homogeneous clusters, including cluster A and B. The principal component analysis (PCA) found that the 15 significant m7G regulators could discriminate the two m7G patterns better (Fig. 2D). Furthermore, the differential expression levels of 15 crucial m7G regulators were depicted using a heatmap and a boxplot (Fig. 2E,F). The gene expression of CYFIP1, CYFIP2, DCPS, EIF3D, EIF4E3, EIF4G3, GEMIN5, LARP1, LSM1, METTL1, NUDT3 and SNUPN displayed lower levels in cluster B than those in cluster A, while the gene expression of EIF4E, NCBP2, and NUDT10 showed the opposite pattern, with higher levels. Combining the results of Fig. 2E,F and Figure S2, most of the RNA m7G methylation regulators showed a higher level of expression in cluster B compared with that of cluster A. Cluster A samples had smaller FEV1/FVC values for the clinical indicators of COPD, suggesting that the

ventilation impairment was more severe relative to cluster B. The high expression of RNA m7G methylation regulators may be closely related to the severity of pulmonary function ventilation impairment with important implications.

Screening of disease signature genes and the correlation with clinical features and GSVA

We then analyzed 15 kinds of m7G regulators with significant differences to screen for disease signature genes by three machine learning algorithms. As shown in Fig. 3A, the 15 kinds of m7G regulators were filtered through the least absolute shrinkage and selection operator (LASSO) regression analysis, with CYFIP2, EIF3D, GEMIN5, METTL1, NCBP2 and SNUPN as signature genes of COPD. Next, these 15 COPD genes were screened by the random forest (RF) method, with selected genes with scores greater than 2 for subsequent analysis (Fig. 3B), including LARP1, EIF4E, GEMIN5, NUDT3, CYFIP2, EIF4E3, EIF4G3, NCBP2, METTL1, EIF3D, DCPS, LSM1, SNUPN and NUDT10. In Fig. 3C, the characteristic genes of COPD were determined by the support vector machine-recursive feature elimination (SVM-RFE) algorithm, with the change curve of the cross-validation error of each gene. Therefore, the fifteen genes with the slightest cross-validation error were selected, including LARP1, GEMIN5, CYFIP2, NCBP2, EIF4E, EIF3D, EIF4G3, SNUPN, DCPS, CYFIP1, EIF4E3, LSM1, METTL1, NUDT3 and NUDT10. Finally, we took the intersection of the three machine learning algorithms and obtained 7 overlapping genes, consisting of CYFIP2, EIF3D, EIF4G3, GEMIN5, METTL1, NCBP2 and SNUPN (Fig. 3D).

Next, the correlation was visualized between the expression of the seven disease signature genes and the risk of FEV1/FVC and the COPD degree separately by a violin diagram plus box diagram. From Fig. 3E,F, compared with FEV1/FVC < 0.7, the gene expression levels of CYFIP2, EIF3D, EIF4G3, GEMIN5, METTL1 and SNUPN were downregulated in FEV1/FVC ≥ 0.7, while the NCBP2 gene was significantly upregulated in FEV1/FVC ≥ 0.7. The gene expression levels of CYFIP2, EIF3D, EIF4G3, GEMIN5, NCBP2 and SNUPN were upregulated in COPD severity grades III and IV compared with those in the controls. In particular, the METTL1 gene was upregulated in COPD severity grade IV compared with the controls, and in grade IV compared with grade III. The levels of disease signature genes were correlated with FEV1/FVC and COPD severity grades, and all the genes were upregulated in COPD except for NCBP2. Multiple m7G disease signature genes were commonly overexpressed in the pulmonary function assessment index FEV1/FVC and COPD classification, strongly associated with COPD progression.

Association of DEGs with hub COPD genes and the key targets of m7G regulators

A total of 1574 DEGs were screened, including 1500 up-regulated and 74 down-regulated genes as shown in Fig. 4A. From the intersection of 1574 differentially expressed genes with 145 hub COPD genes, 4 genes were obtained as critical genes, including alpha-2-macroglobulin, catalase (CAT), glutamate-cysteine ligase catalytic subunit and matrix metalloproteinase 1 (Fig. 4B). The 15 important m7G methylation regulators and 4 key genes were analyzed for co-expression relationships in COPD by PPI network. These important m7G regulators could directly or indirectly interact with some COPD hub genes, as shown in Fig. 4C. In addition, the COPD hub gene CAT was closely related to m7G regulators as an important target for m7G methylation.

To further assess whether the CAT gene was associated with COPD, we evaluated the CAT expression levels in healthy controls and COPD patients. As shown in Fig. 4D, the gene expression level of CAT was lower in COPD patients than that in controls ($p < 0.0001$). Additionally, the gene expression level of CAT was negatively correlated with GEMIN5 ($r = -0.26$, $p = 0.001$), CYFIP2 ($r = -0.48$, $p < 0.001$), METTL1 ($r = -0.51$, $p < 0.001$), SNUPN ($r = -0.66$, $p < 0.001$), EIF4G3 ($r = -0.66$, $p < 0.001$), and EIF3D ($r = -0.85$, $p < 0.001$). It was also positively correlated with NCBP2 ($r = 0.67$, $p < 0.001$), as shown in Fig. 4E. The altered CAT gene expression in COPD lung tissues was closely related to m7G regulatory factors as a key target.

GO and KEGG analyses

To a deeper understanding of the biological functions of the DEGs, we performed GO annotation and KEGG pathway enrichment analysis. The first 10 enriched GO terms were shown in Fig. 5A and divided into 3 sections: biological process (BP), cellular component (CC) and molecular function (MF). BP showed that DEGs participated in the response to ribonucleoprotein complex biogenesis, protein localization to the cell periphery, and Golgi vesicle transport et al. CC analysis showed that DEGs were significantly enriched in nuclear speck, cell-substrate junction, focal adhesion, and COPII-coated ER to Golgi transport vesicle et al. The top three enriched terms for MF were DNA-binding transcription factor binding, ubiquitin protein ligase binding, and catalytic activity acting on DNA. Considering all these results, the majority of DEGs were enriched in categories of the inflammatory and immune response. The enriched KEGG pathways included pancreatic secretion, interleukin 17 (IL-17) signaling pathway, and basal transcription factors et al. (Fig. 5B). COPD was a multifactorial disease, and all of the above pathways were directly or indirectly closely related to COPD. Pancreatic secretion was associated with COPD-related metabolic abnormalities, and the IL-17 signaling pathway played an important role in COPD inflammatory and immune processes. And, multiple cell-signaling pathways would act in COPD through oxidative stress.

The m7G regulators associated with immune cell infiltration

Based on the results of GO and KEGG analyses, DEGs were centrally enriched in pathways such as immunity and inflammation, so we further assessed the immune differences between COPD and control groups. The expression levels of 22 immune cells were evaluated in the GSE76925 samples, as shown in Fig. 6A. There were 10 significantly different classes of immune cells between healthy controls and COPD patients (Fig. 6B). Compared with those in control group, the number of plasma cells, activated CD8⁺ T cells and T follicular helper cells were upregulated in COPD group, while there were nine types of downregulated cells, including

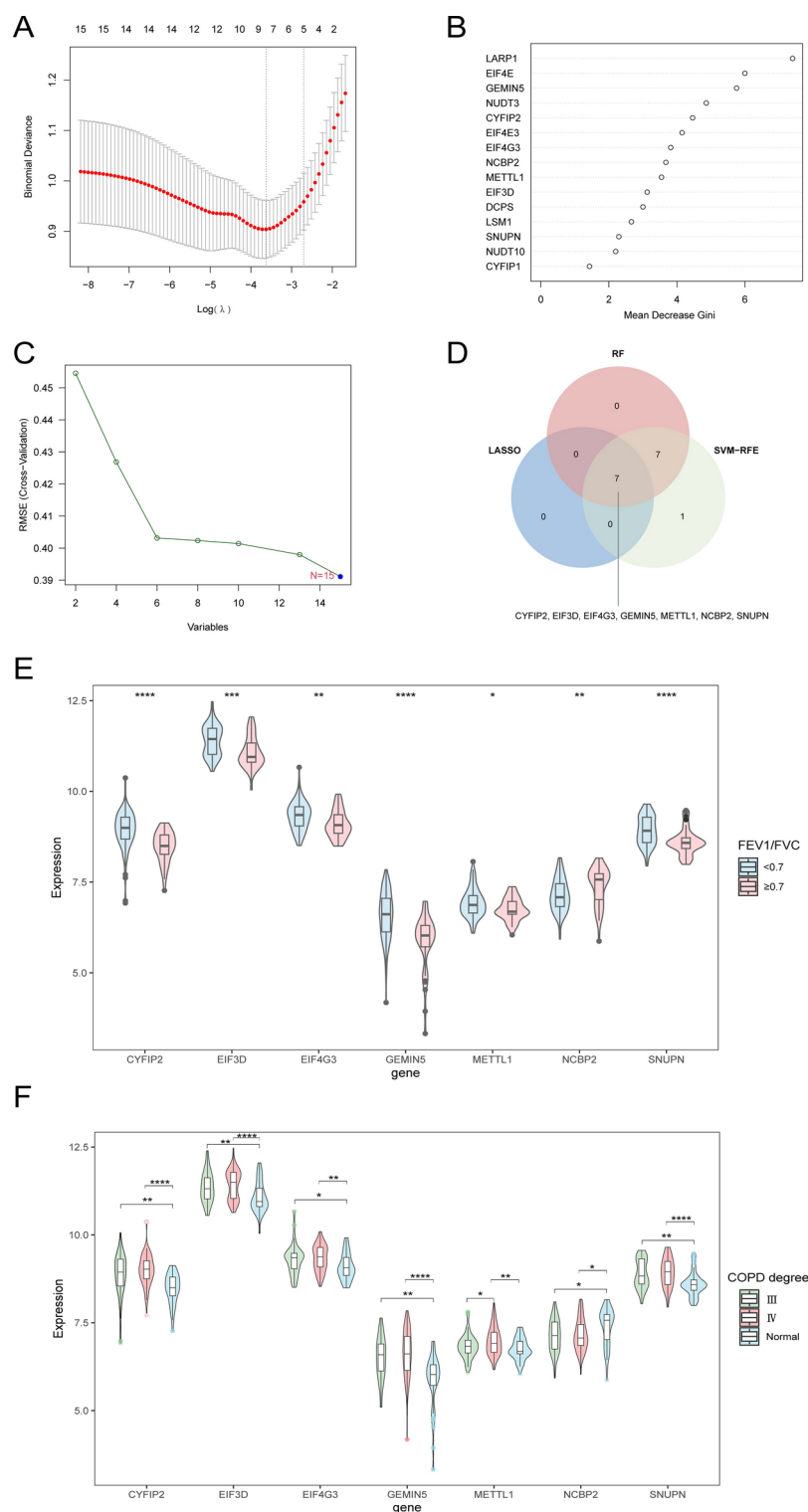


Fig. 3. Screening of disease signature genes. **(A)** Selection of the best $\text{Log}(\lambda)$ value for LASSO regression. The x-axis represents the $\text{Log}(\lambda)$ value, and the y-axis represents the error rate of cross-validation errors. **(B)** Results of the Ginico efficient method in the random forest classifier. The x-axis indicates the genetic variable, and the y-axis represents the importance index. **(C)** Variation curve of gene cross-validation error in the SVM-RFE algorithm. **(D)** Venn diagram showing 3 disease signature genes shared by LASSO, RF and SVM-RFE arithmetic methods. **(E)** Expression of 7 disease signature genes at different levels of the clinical indicator FEV1/FVC. **(F)** Expression of 7 disease signature genes in different clinical grades of COPD. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

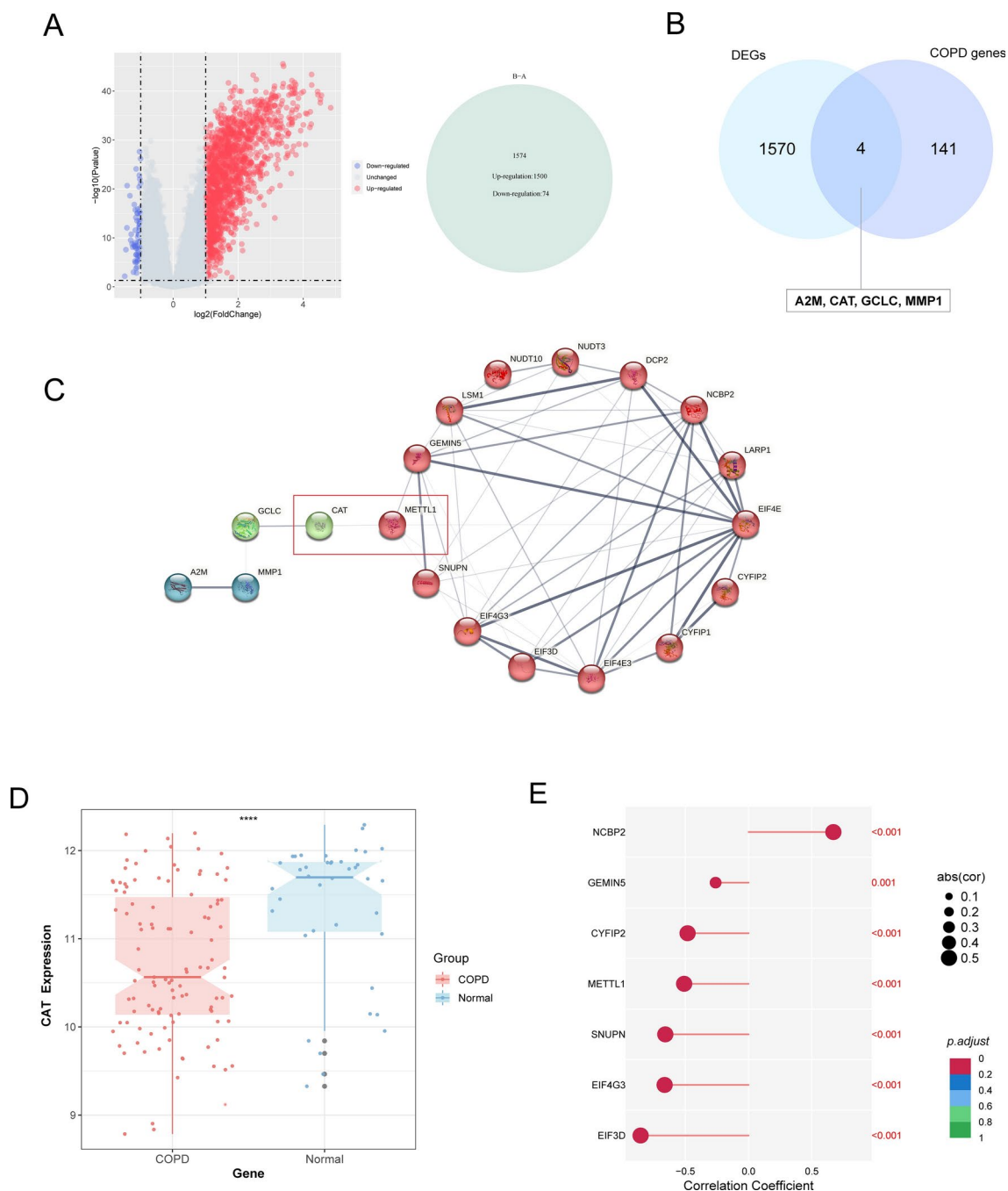


Fig. 4. Screening of DEGs and correlation of CAT with disease signature genes. **(A)** Volcano plot and Venn diagram of DEGs ($|\log_2 \text{FC}| > 2$; $p < 0.05$). Upregulated genes are colored in red, and downregulated genes are colored in blue. **(B)** Venn diagram showing 4 overlapping DEGs and hub COPD genes. **(C)** The PPI network was constructed by 15 m7G regulators as well as 4 key AD genes, with the disconnected proteins being hidden. Wider lines indicate stronger evidence of protein interactions. **(D)** The t test was conducted to analyze the difference in the expression of CAT between normal controls and COPD patients. **(E)** Pearson's test was performed to analyze the correlation between CAT and m7G disease signature genes. The results with significant differences are marked as **** $p < 0.0001$.

activated dendritic cells, activated CD4⁺ T memory cells and resting CD4⁺ T memory cells et al. On account of the two typing patterns of m7G clusters A and B, the differences in immune cells were analyzed then. The results showed that 16 immune cells differed in these two subtypes (Fig. 6C). Activated CD8⁺ T cells, CD56 dim natural killer cells, myeloid-derived suppressor cells, macrophages, monocytes, natural killer T cells, regulatory T cells, T follicular helper cells and type 1 T helper (Th1) cells were upregulated in cluster A, while activated dendritic cells, eosinophils, gamma delta T cells, natural killer cells, neutrophils, and type 2 T helper (Th2) cells were upregulated in cluster B. Cluster A was associated with Th1 cell dominant immunity,

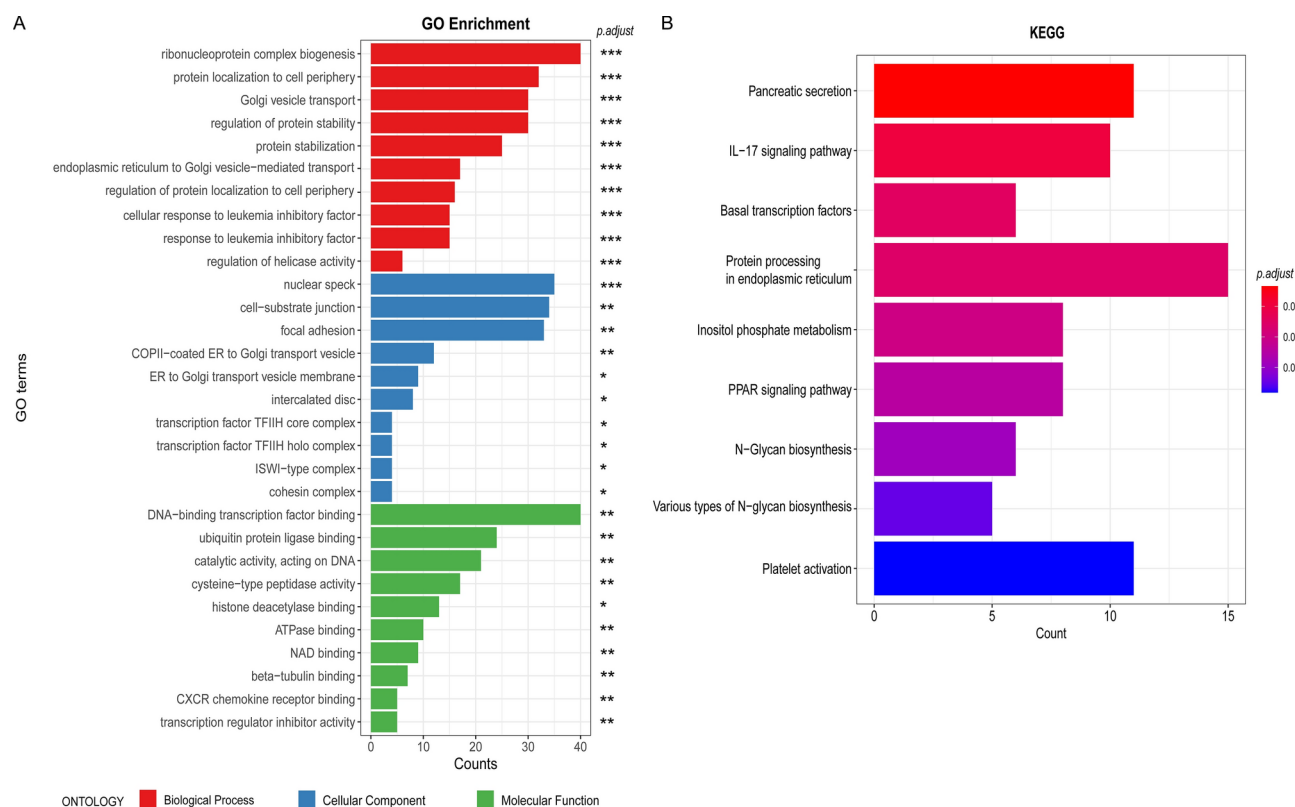


Fig. 5. The relationships of m7G regulators and key COPD genes. **(A)** GO enrichment analysis of DEGs ($p < 0.05$ and $q < 0.05$). **(B)** KEGG pathway enrichment results for DEGs^{41–43} ($p < 0.05$ and $q < 1$). * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

whereas cluster B was associated with Th2 cell prevalent immunity. Moreover, six genes of the m7G regulators CYFIP1, CYFIP2, EIF4E3, GEMIN5, LARP1 and METTL1 were positively associated with many immune cells, while NUDT10 was negatively associated with many immune cells (Fig. 6D). The results of immune infiltration showed a close association between the expression level of m7G regulators and the immune infiltrating cells in COPD, especially T cells and B cells.

Single-cell sequencing and further sub-clustering of T cells and B cells

In support of the verification of the immune infiltration results described above, we additionally found a dataset GSE196638 of emphysema tissue to assist of the immune microenvironment in COPD patients. The control and emphysema cell clusters and cell type annotations were visualized separately, with separation of seven cell types, including natural killer cells, CD4⁺ memory cells, monocyte-derived dendritic cells, B cells, CD16⁺ and CD 14⁺ monocytes, naive T cells and plasmacytoid dendritic cells (Fig. 7A,B). Next, the levels of seven disease characteristic genes were visualized in normal lung tissue cells and emphysema tissue cells in each cell cluster (Fig. 7C,D). Genes with low expression were not shown in the graph. Compared with normal lung tissue cells, the CYFIP2 gene expression was slightly reduced in natural killer cells and significantly increased in plasmacytoid dendritic cells, while the EIF3D gene expression was significantly increased in plasmacytoid dendritic cells, with little change in the remaining six cell clusters. Additionally, the EIF4G3 gene expression was significantly increased in plasmacytoid dendritic cells. Additionally, the NCBP2 gene was reduced considerably in B cells and dramatically increased in both naive T cells and plasmacytoid dendritic cells.

According to the results of immune infiltration, T and B cells were more actively expressed in COPD, and we then concentrated on two immune subclusters, T and B cells as followed. Compared with those in control, T cells subcluster map had more intermediate monocytes, natural killer cells and Vδ2^(neg) γδ T cells in the emphysema group and fewer MAIT cells and T regulatory cells in the emphysema group as shown in Fig. 8A. According to Fig. 8B, the two disease signature genes, CYFIP2 and EIF3D, were closely associated with T cell subcluster, with CYFIP2 being expressed more highly in MAIT cells, while EIF3D was expressed more highly in intermediate monocytes cells, MAIT cells and Vδ2^(neg) γδ T cells. As shown in Fig. 8C,D, compared with those in control, the five types of B cells increased in the emphysema group. The four disease signature genes, CYFIP2, EIF3D, EIF4G3 and NCBP2, were strongly associated with B cells subcluster.

DO analysis

We explored the enrichment of differential genes in disease pathways by conducting DO analysis of 1574 DEGs. According to the bar and bubble plots in Fig. 9A,B, DEGs were associated with a variety of lung diseases,

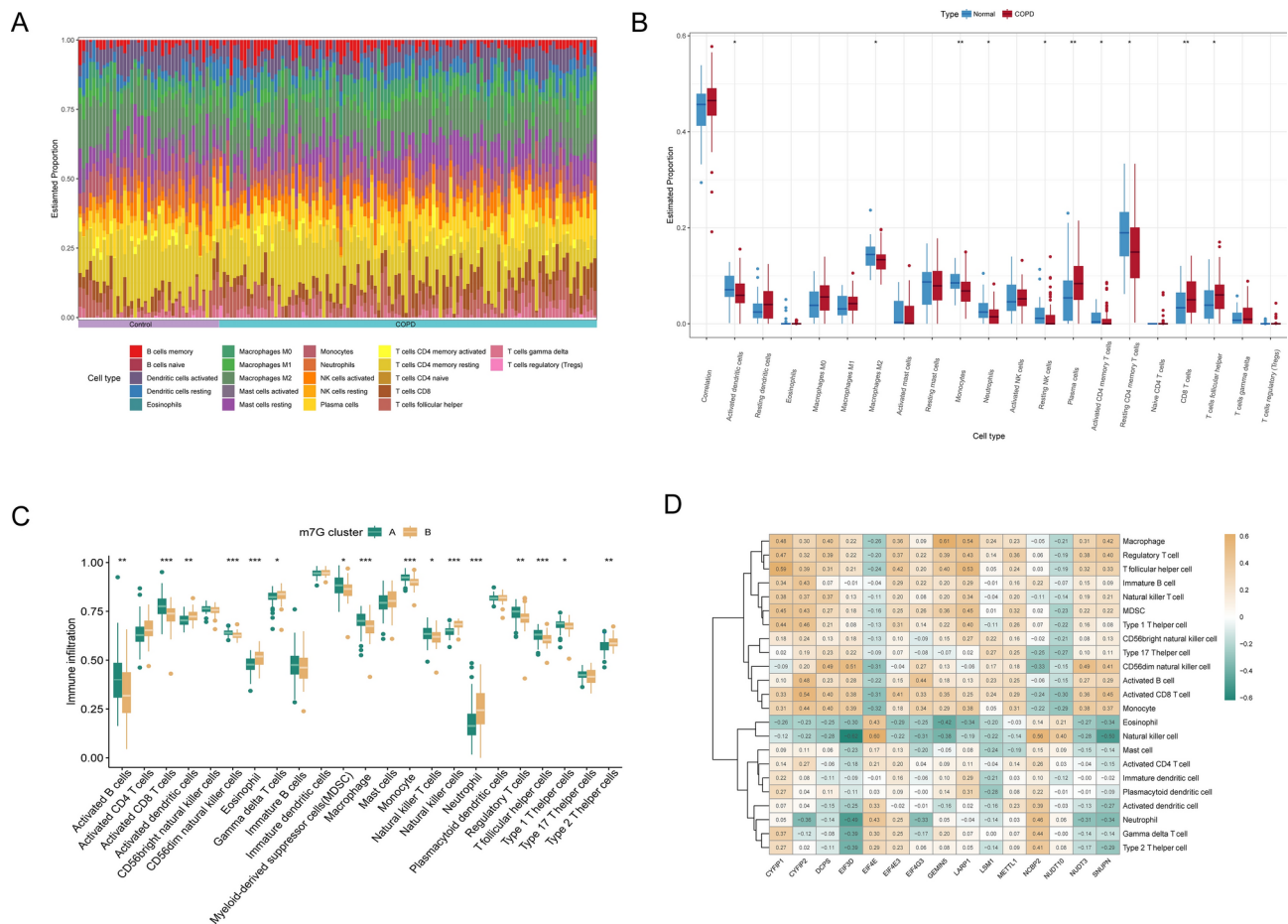


Fig. 6. Immune cell infiltration. (A,B) Expression of 22 immune cells between COPD patients and healthy controls. (C) Difference analysis of immune cells in the m7G cluster. (D) Correlation analysis of immune cells and m7G regulators. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

including COPD, obstructive lung disease, bronchial disease, asthma, interstitial lung disease, and pneumonia et al. Furthermore, DEGs were also mostly enriched in diseases, closely related to immunity, such as dermatitis and bacterial infections disease. There was cross-over of multiple DEGs in COPD, obstructive lung disease, bronchial disease, asthma and interstitial lung disease, suggesting that multiple genes could collaboratively play vital roles in these five lung-associated diseases, as shown in Fig. 9C.

The mRNA levels of signature genes in cell senescence and Gene set enrichment analysis (GSEA)

To further validate the above results, we carried out a preliminary validation of the mRNA levels of seven disease signature genes and the target gene CAT by qRT-PCR using the HEFs cell premature senescence model. The expression differences of these eight genes in the 22PDL and PS groups (Fig. 10A) were consistent with those in the healthy control and COPD groups in the GSE76925 dataset (Figs. 4D and 10B). The differences in the gene expression of CYFIP2, EIF4G3, METTL1, NCBP2, SNUPN and CAT were statistically significant. In this study, we focused on the association between RNA m7G methylation and aging and aging-related diseases. We finally screened four pathways from multiple statistically significant GSEA pathways that were closely associated with promoting premature senescence or acting on senescence. We performed GSEA of the differentially expressed genes in the validation dataset (Fig. 10C) and GSE76925 dataset (Fig. 10D) and found that both datasets showed upregulation in the enrichment of pathways such as oxidative stress and inflammation, including the IL-17 signaling pathway, the NF- κ B signaling pathway, the NOD-like receptor signaling pathway and the TNF signaling pathway. These results were consistent with the above results, suggesting that m7G-related disease signature genes played important roles in COPD disease and were closely associated with inflammatory immune pathways.

By RNA immunoprecipitation technique, RNA m7G methyltransferase METTL1 and its binding RNAs were successfully isolated, and the primers of CAT gene were amplified by RT-qPCR to deeply explore the regulatory mechanism of METTL1-CAT axis. Our results have proved the combination and abundance of METTL1 for CAT. The IP group showed distinct destination bands and the RNA m7G methyltransferase METTL1 could effectively bind to CAT (Fig. 11A). As shown in Fig. 11B, the binding level of METTL1 to CAT was significantly

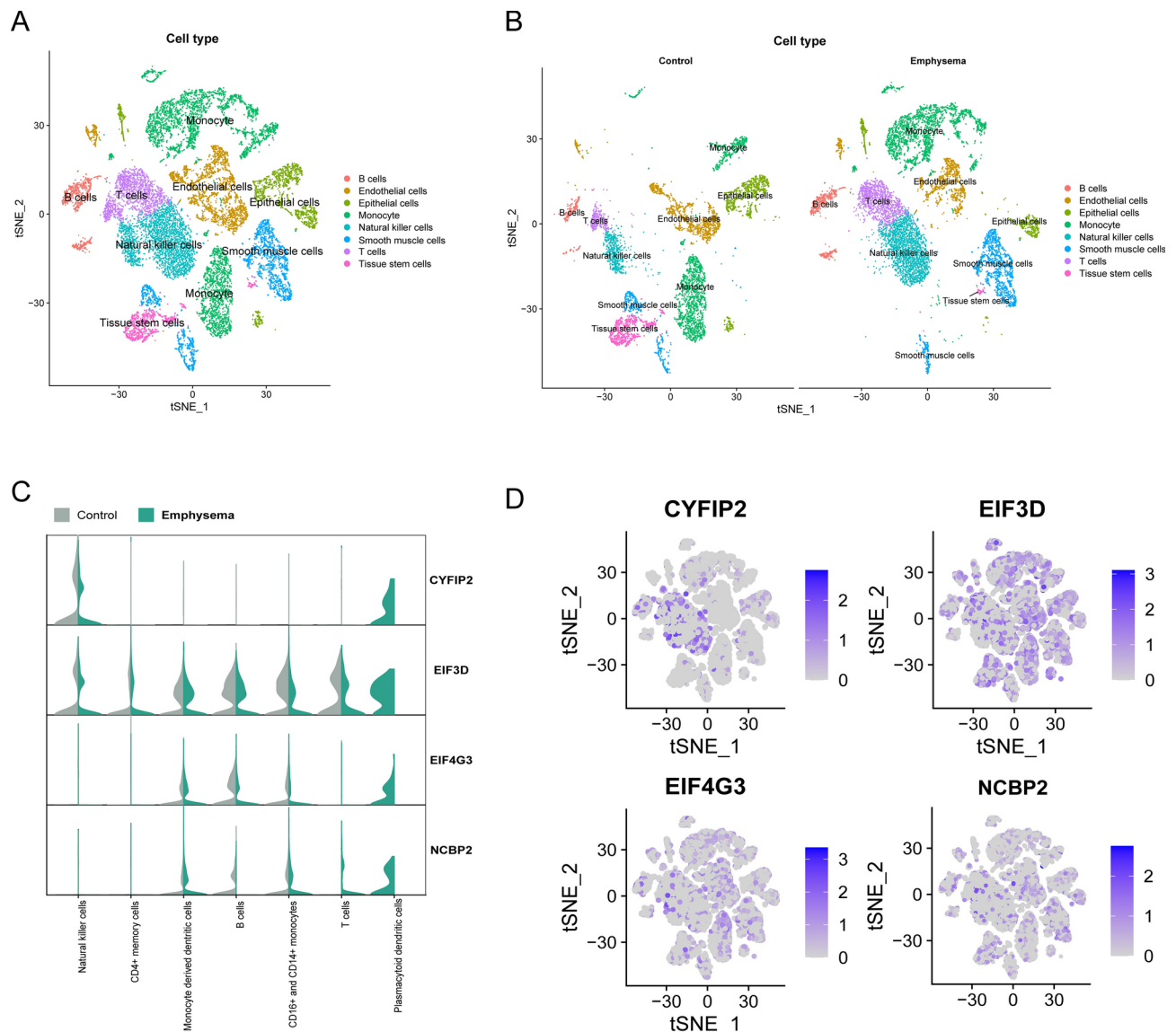


Fig. 7. Single-cell sequencing. (A,B) The tSNE clustering was used to display the distribution of various types of cells between normal controls and emphysema cells. (C) The violin plot shows the expression of 7 disease-characteristic genes in normal control cells and emphysema cells. Undistributed genes have been hidden. (D) Distribution of 7 disease signature genes in tSNE clustering. Undistributed genes have been hidden.

increased in the PS group, which was 6.2 times higher than that in the control group with statistical significance ($p < 0.05$). The results suggested that high expression of METTL1 in lung cell premature senescence promoted RNA m7G methylation and could stimulate elevated levels of METTL1-CAT binding, contributing to the development of COPD.

Discussion

COPD is a chronic lung condition associated with an elevated risk of lung cancer⁴⁴. The role of m7G methylation had garnered increasing attention due to its involvement in various diseases. RNA m7G methylation is closely linked to the development of lung cancer, interstitial fibrosis and pulmonary hypertension³³. While m7G regulators play a critical role in lung diseases, the connection between these regulators and COPD pathogenesis remains poorly understood. In this study, we employed integrative bioinformatics to identify signature genes of m7G regulators in COPD, systematically explored their differences in immune infiltration, and identified potential targets of m7G methylation for the first time.

Fifteen crucial m7G regulators were identified based on differential expression analysis between normal and COPD lung tissues. Most of the m7G regulators were elevated and correlated with COPD, suggesting that m7G modifications were generally increased in COPD lung tissues, affecting the expression of target genes during the development of COPD. Two m7G subtype patterns were identified for cell clusters A and B based on 15 crucial m7G regulators, which were useful for the prevention, intervention and precise treatment of COPD.

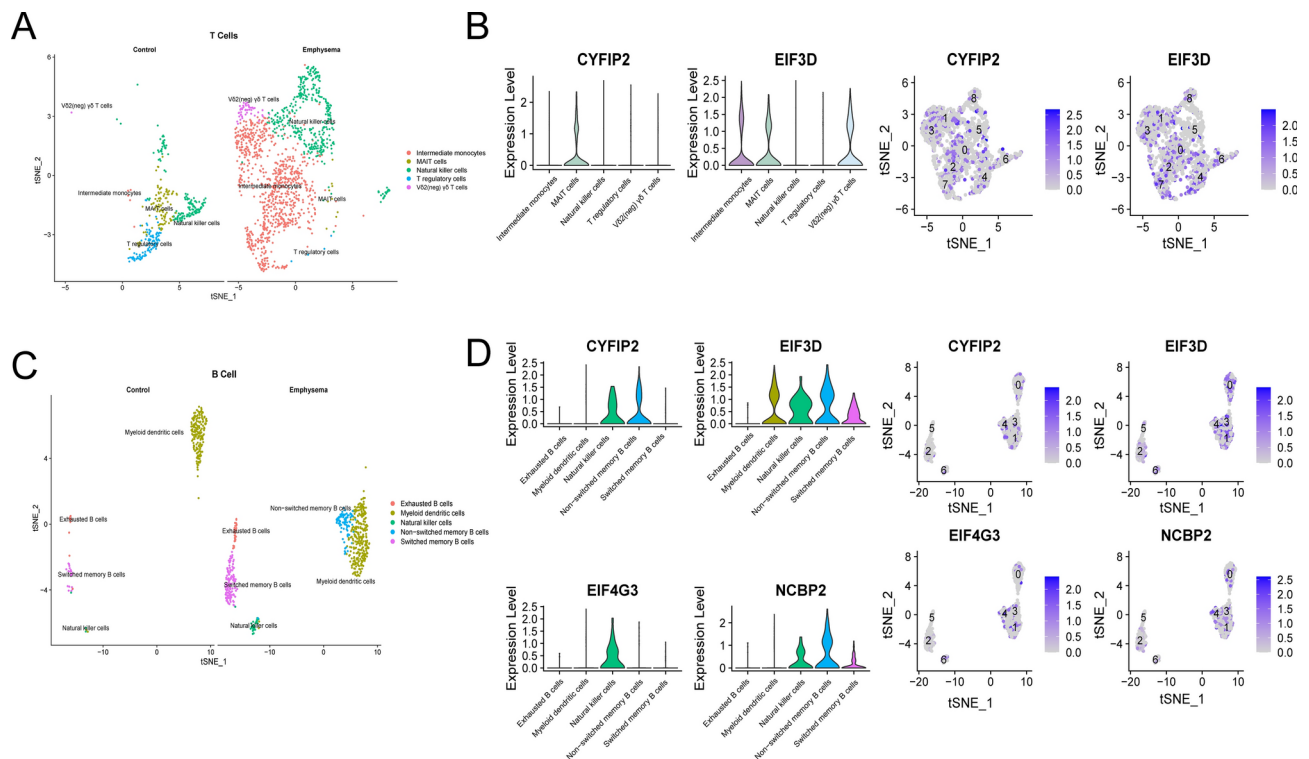


Fig. 8. Sub-clustering analysis of T-cells and B-cells. **(A,B)** Differences in enrichment of T-cell subclusters between control and emphysema groups, expression levels of the two disease signature genes in T-cell clusters. **(C,D)** Differences in enrichment of B-cell subclusters between control and emphysema groups, expression levels of the four disease signature genes in B-cell clusters.

Three machine learning methods screened seven important disease signature genes, including METTL1, EIF3D, CYFIP2, GEMIN5, NCBP2 and SNUPN. The METTL1 gene expression level was upregulated in lung cancer³⁷, idiopathic pulmonary fibrosis (IPF)⁴⁵, and bronchoalveolar lavage fluid cells from IPF patients and was associated with poor prognosis. Overexpression of EIF3D promoted cell proliferation in ovarian, renal cell, gastric and gallbladder cancers⁴⁶. Moreover, a high level of EIF3D was a valuable independent prognostic indicator of shorter overall survival in lung adenocarcinoma patients and could coregulate the translation of SLC7A11 with RBMS1 in lung cancer, closely related to lung injury diseases^{47,48}. Recent studies reveal that the m7G-related regulators NCBP2 and EIF3E3 may modulate the tumor immune microenvironment in head and neck squamous cell carcinoma through the downstream molecules CCL4 and CCL5⁴⁹. In our results, the METTL1 and EIF3D genes were increased in lung tissues of COPD patients and were closely associated with FEV1/FVC < 0.7 and COPD severe grades III and IV, suggesting that METTL1 and EIF3D affect the development and progression of COPD. In addition, the other five disease signature genes, CYFIP2, EIF4G3, GEMIN5, SNUPN were upregulated in FEV1/FVC < 0.7 and COPD severe grades III and IV, while the NCBP2 gene was downregulated. Although the roles of these genes in lung- or aging-related diseases are not fully understood, this study highlights their significant involvement in regulating FEV1/FVC and COPD progression through RNA m7G methylation. These findings provide valuable insights for COPD classification and intervention strategies.

We screened the target gene CAT by PPI protein interactions. CAT was expressed predominantly in bronchial epithelial cells⁵⁰ and alveolar macrophages⁵¹ and was the main endogenous antioxidant involved in lung, catalyzing H₂O₂ and protecting body tissues from damages. Levels of MDA (a byproduct of peroxidation of polyunsaturated fatty acids) in sputum from patients with acute exacerbation of chronic obstructive pulmonary disease were correlated positively with CAT activity, suggesting a direct relationship between the degree of inflammation peroxidation and antioxidant activity in the respiratory tract⁵². Smoking exposure⁵³, harmful gases^{54,55} and heavy metals⁵⁶ could contribute to disease development by increasing MDA, causing oxidative stress and decreasing superoxide dismutase and CAT. Our results were consistent with these previous findings that CAT expression levels were upregulated in COPD patients and closely related to COPD disease progression. The CAT allele was statistically associated with COPD risk⁵⁷. Therefore, some potential regulatory mechanisms had existed between m7G regulators and the above genes contribute to COPD pathogenesis and progression, but the specific mechanisms and roles of these potential target genes needed to be explored further. We also identified METTL1 and CAT as possible critical links between m7G methylation and COPD. Current clinical studies had shown that insufficient radiofrequency ablation (iRFA)-induced hepatocellular carcinoma could be attenuated by downregulation of METTL1, and that high expression of METTL1/WDR4 mediated methylation modification of m7G tRNAs, which in turn upregulated the resistance of nasopharyngeal carcinoma cells to cisplatin and docetaxel chemotherapy⁵⁸. Additionally, studies had shown that peroxisomal viruses (containing

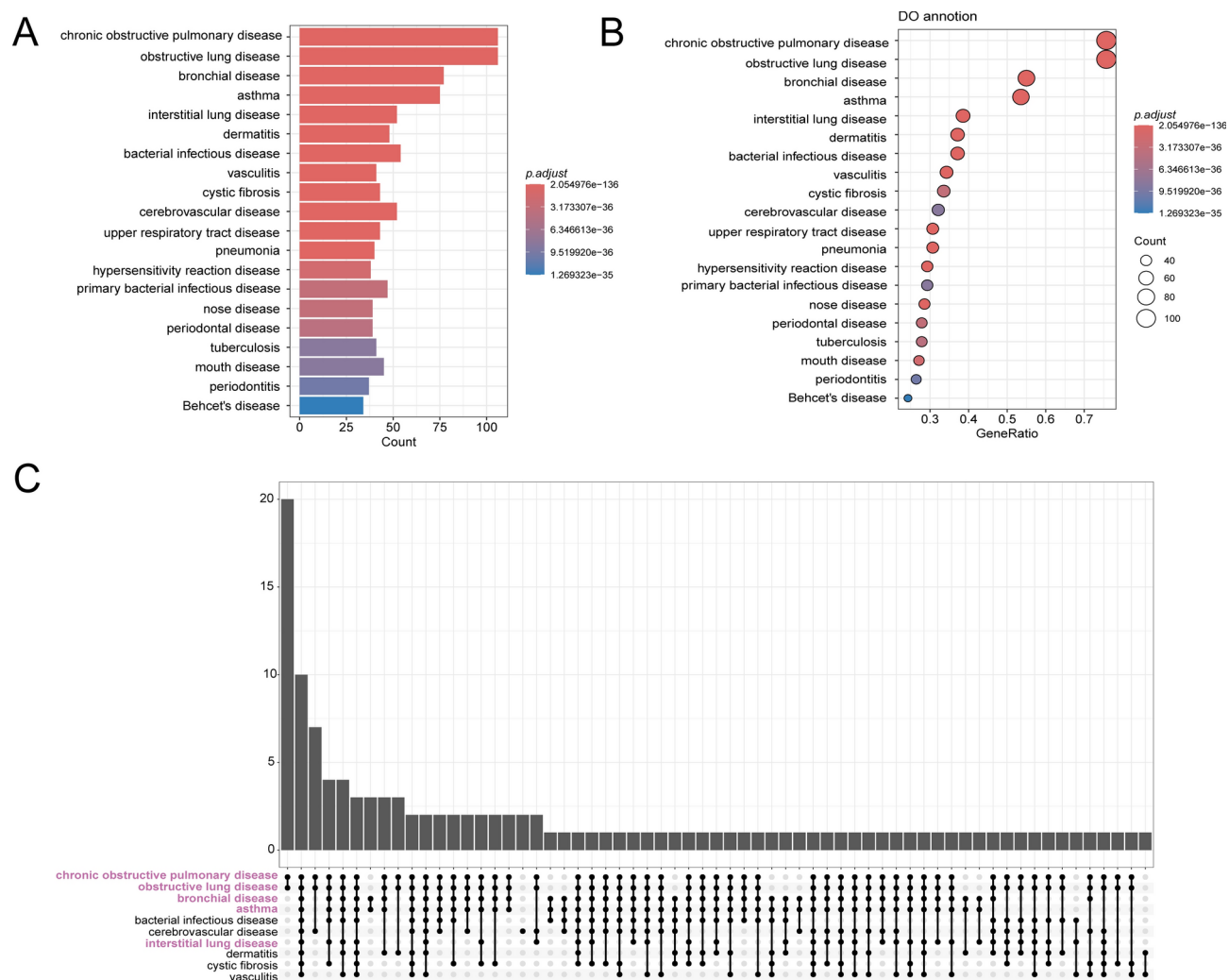


Fig. 9. DO analysis. (A,B) The bar and bubble plots of disease pathways enriched for 1574 DEGs. (C) The upset maps of overlapping genes between different diseases.

the CAT gene) ameliorate inflammatory diseases such as ocular optic neuritis, and a novel AAV-9 viral vector for therapeutic anti-aging contained CAT gene. According to gene therapy strategy, the aging immunosuppressive microenvironment was formed through the METTL1-CAT axis, thus contributing to the prevention and intervention of COPD.

A total of 1573 DEGs were obtained, comprising 1500 upregulated and 74 downregulated genes. These DEGs were primarily enriched in pathways regulating helicase activity and the cellular response to leukemia inhibitory factors. Helicase regulation is crucial for alveolar destruction and airway remodeling in COPD, while inhibition of the MAVS/RIG-I-like pathway is directly associated with COPD severity and inversely correlated with lung function and quality of life⁵⁹. Viral infections, which exacerbate COPD, are linked to increased expression of leukemia inhibitory factor in lung tissues, worsening symptoms and lung injury⁶⁰. The DEGs were also involved in protein processing in the endoplasmic reticulum, pancreatic secretion and the IL-17 signaling pathway. Smoking exposure induced endoplasmic reticulum stress, leading to the inflammatory response in COPD⁶¹. The porcine pancreatic elastase PPE induced the expression and secretion of placental growth factors in lung epithelial cells in vitro and in vivo, promoting apoptosis and emphysema⁶². IL-17 cytokines mediated the recruitment of inflammatory cells, with essential treatment potential for acute exacerbations of COPD⁶³.

Several studies have shown that m7G regulatory factors are inextricably linked to immune infiltration^{58,64}. Our results showed that the m7G regulators were also associated with a wide range of immune cells in COPD. Differences in immune cell infiltration between the two m7G patterns showed higher levels of activated B cells, CD8⁺ T cells, monocytes, natural killer T cells, regulatory T cells, follicular helper T cells, and Th1 cells in cluster A compared to cluster B, while $\gamma\delta$ T cells, natural killer cells, and Th2 cells were lower. Single-cell sequencing of emphysema tissues further supported the relationship between COPD and immune cells, consistent with findings in PAH³³. This study is the first to suggest that COPD phenotypes are closely linked to immune cells and that m7G signature genes may interact with immune cells. The four disease signature genes were significantly upregulated in emphysema tissues in plasmacytoid dendritic cells. Dendritic cells were one of

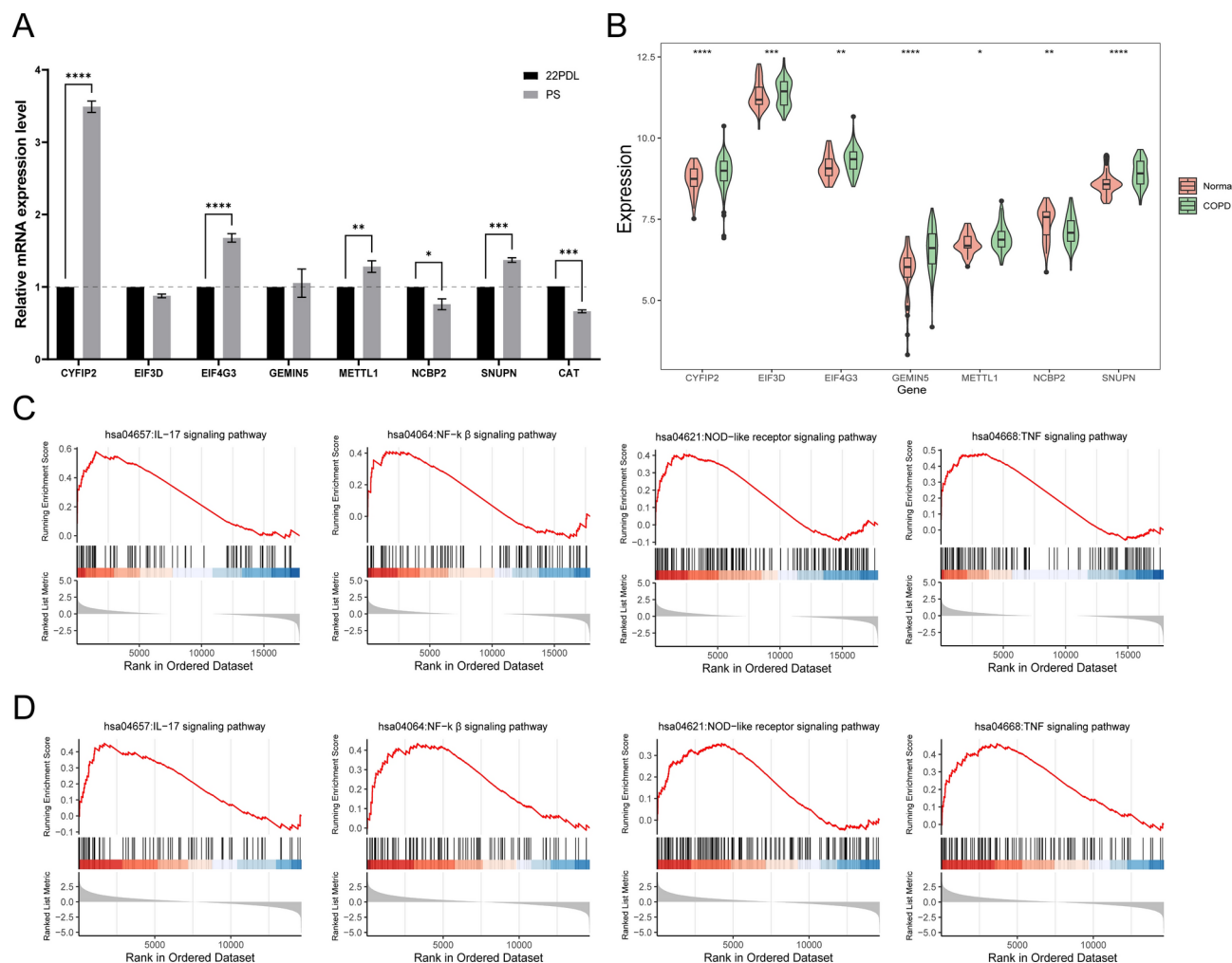


Fig. 10. Validation of the HEFs dataset. (A) The mRNA expression of 7 disease signature genes and target gene CAT between the 22PDL and PS groups. (B) Expression of 7 disease signature genes between the control and COPD groups. (C-D) GSEA of the validation dataset and GSE76925 dataset. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

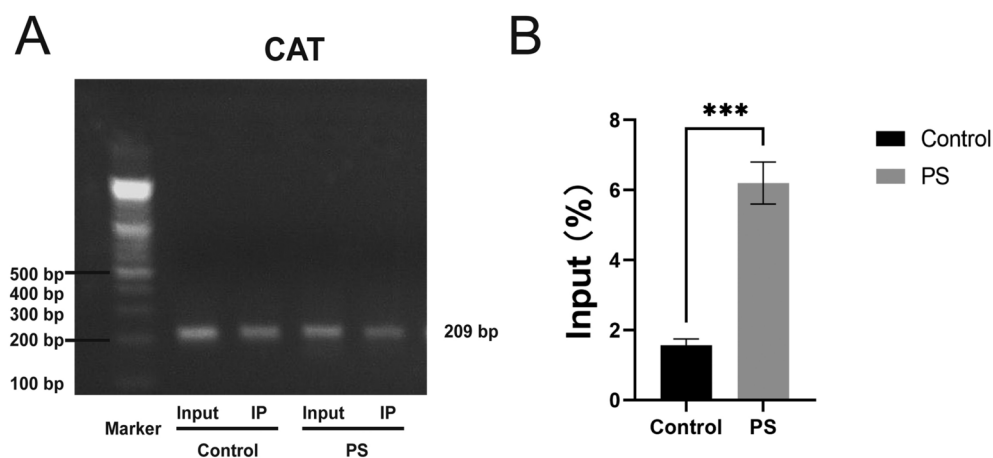


Fig. 11. RIP-Q-PCR. (A) CAT gene PCR product gel electrophoresis plots in different groups. (B) Levels of CAT gene binding to METTL1. $n = 3$, *** $p < 0.001$.

the critical regulators of the immune response and inflammation in asthma and COPD⁶⁵. Both the four disease signature genes were significantly increased in plasmacytoid dendritic cells in emphysematous tissue, suggesting that m7G could regulate the COPD immune response and inflammation through dendritic cells. It has been shown that dysregulation of m7G-associated regulatory factors is associated with ocular immune responses, particularly by stimulating T cells and leukocytes to promote local immunity in the ophthalmoscope⁶⁶. Both the CYFIP2 and EIF3D genes were abundantly expressed in the subclusters of T and B cells, revealing that these two disease signature genes played key immunomodulatory roles in emphysema tissues, which was consistent with the previous results. In addition, EIF4G3 and NCBP2 were also abundantly expressed in B-cell clusters. All these results suggested that RNA m7G regulators in COPD and emphysema had key roles in the immune microenvironment and could influence the onset and progression of COPD via immune pathways.

DO analysis validated the association of DEGs with lung diseases. To further validate our bioinformatics findings, we used a cell premature senescence model. We observed that the mRNA levels of all the genes were consistent with those in GSE76925 between the premature senescence and control groups, and the pathway results of the four senescence-related pathways of GSEA were also consistent. These findings suggest that the seven disease signature genes act as RNA m7G regulators in COPD and aging-related diseases, with the METTL1-CAT axis potentially playing a key role in senescence and disease progression. And that the METTL1-CAT axis might play a key regulatory role in senescence and related diseases. The GSEA pathway we focused on was based on the inflammatory pathway in cell premature senescence and aging diseases. It had been shown that increased IL-17 positively regulated eosinophils and neutrophils in vivo, which exerted a pro-inflammatory effect and exacerbated diseases such as asthma⁶⁷. The IL-17A could induce endothelial cell senescence by the NF- κ B-related oxidative stress pathway⁶⁸. Meanwhile, excessive ROS could activate the NF- κ B pathway to trigger inflammation and induce the expression of various inflammatory mediators, including TNF- α ⁶⁹. In addition, NOD1 and NOD2 could also activate the NF- κ B signaling pathway in the form of inflammatory vesicles to promote inflammation under certain stimulation⁷⁰. These pathways were consistently upregulated in both the biological dataset and the cell premature senescence model, validating the role of inflammatory immune responses in emphysema and COPD progression. Importantly, these results suggest that COPD, as an aging-related disease, is driven by the activation of inflammatory pathways that accelerate lung cell senescence, exacerbating disease progression.

This study is the first to comprehensively analyze the role of m7G regulators in COPD pathogenesis, subtype classification, immune infiltration, and clinical indicators. The seven disease signature genes screened by three machine learning, may affect COPD progression by regulating FEV1/FVC and immune infiltration, providing meaningful references to COPD progression and deterioration. The CAT gene may be a potential target for m7G methylation through PPI network interactions analysis, which provides new issues for COPD prevention and intervention.

In conclusion, this study tapped the potential METTL1-CAT axis mechanism from the perspective of RNA m7G methylation and COPD disease, and the mechanism of RNA m7G methylation in COPD is closely related to the immune microenvironment. However, our study has some limitations. The sample size of the dataset selected for this study was limited, and there were individual differences. In the future, we need further experimental validation in vitro studies and clinical samples. For example, the regulation of the METTL1-CAT axis and the alteration of the immune microenvironment.

Conclusion

In summary, m7G regulators play essential roles in the development, subtype classification, clinical indicators, and the immune infiltration landscape of COPD. The 15 important m7G regulators are closely associated with the development of COPD and have a suitable categorization of COPD patients into two subtypes. The immune infiltration landscape has visualized the differences in immune cells in COPD and nonpatients. Furthermore, disease signature genes are closely related to key COPD genes, and CAT, as a potential modification target, provides new clues for the new biomarker of COPD. The specific mechanisms of m7G methylation are still need to be clarified further with great potential in COPD in the future.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 25 November 2024; Accepted: 10 March 2025

Published online: 20 March 2025

References

- Kahnert, K. A., Jörres, R., Behr, J. & Welte, T. The diagnosis and treatment of COPD and its comorbidities. *Dtsch Arztebl Int.* **120**, 434–444 (2023).
- Zhu, B., Wang, Y., Ming, J., Chen, W. & Zhang, L. Disease burden of COPD in China: A systematic review. *Int. J. Chron. Obstruct. Pulmon. Dis.* **13**, 1353–1364 (2018).
- Wang, C. et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): A national cross-sectional study. *The Lancet.* **391**, 1706–1717 (2018).
- Wang, X., Guo, Z. & Yan, F. RNA epigenetics in chronic lung diseases. *Genes.* **13**, 2381 (2022).
- Wiener, D. & Schwartz, S. The epitranscriptome beyond m6A. *Nat. Rev. Genet.* **22**, 119–131 (2021).
- Lin, S., Liu, Q., Jiang, Y.-Z. & Gregory, R. I. Nucleotide resolution profiling of m7G tRNA modification by TRAC-Seq. *Nat. Protoc.* **14**, 3220–3242 (2019).
- Malbec, L. et al. Dynamic methylome of internal mRNA N7-methylguanosine and its regulatory role in translation. *Cell Res.* **29**, 927–941 (2019).

8. 7-Methylguanosine Modifications in Transfer RNA (tRNA). *Int. J. Mol. Sci.* **19**, 4080 (2018).
9. Haag, S., Kretschmer, J. & Bohnsack, M. T. WBSCR22/Merm1 is required for late nuclear pre-ribosomal RNA processing and mediates N7-methylation of G1639 in human 18S rRNA. *RNA*. **21**, 180–187 (2015).
10. Pandolfini, L. et al. METTL1 promotes let-7 MicroRNA processing via m7G methylation. *Mol. Cell*. **74**, 1278–1290.e9 (2019).
11. Furuichi, Y. Discovery of m7G-cap in eukaryotic mRNAs. *Proc. Jpn Acad. Ser. B Phys. Biol. Sci.* **91**, 394–409 (2015).
12. Lewis, J. D. & Izaurflde, E. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247**, 461–469 (1997).
13. Muthukrishnan, S., Both, G. W., Furuichi, Y. & Shatkin, A. J. 5'-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation. *Nature*. **255**, 33–37 (1975).
14. Lin, S. et al. Mettl1/Wdr4-mediated m7G tRNA methylome is required for normal mRNA translation and embryonic stem cell self-renewal and differentiation. *Mol. Cell*. **71**, 244–255 (2018).
15. Zhang, L.-S. et al. Transcriptome-wide mapping of internal N7-methylguanosine methylome in mammalian mRNA. *Mol. Cell*. **74**, 1304–1316 (2019).
16. Luo, Y. et al. The potential role of N7-methylguanosine (m7G) in cancer. *J. Hematol. Oncol.* **15**, 63 (2022).
17. Cioni, J. M. et al. Axon-axon interactions regulate topographic optic tract sorting via CYFIP2-dependent WAVE complex function. *Neuron*. **97**, 1078 (2018).
18. Tiwari, S. S. et al. Alzheimer-related decrease in CYFIP2 links amyloid production to tau hyperphosphorylation and memory loss. *Brain*. **139**, 2751–2765 (2016).
19. Nakashima, M. et al. De novo hotspot variants in CYFIP2 cause early-onset epileptic encephalopathy. *Ann. Neurol.* **83**, 794–806 (2018).
20. Zweier, M. et al. Spatially clustering de novo variants in CYFIP2, encoding the cytoplasmic FMRP interacting protein 2, cause intellectual disability and seizures. *Eur. J. Hum. Genet. EJHG*. **27**, 747–759 (2019).
21. Park, Y. et al. Translation mediated by the nuclear cap-binding complex is confined to the perinuclear region via a CTIF-DDX19B interaction. *Nucleic Acids Res.* **49**, 8261–8276 (2021).
22. Ryu, I. & Kim, Y. K. Translation initiation mediated by nuclear cap-binding protein complex. *BMB Rep.* **50**, 186–193 (2017).
23. Yin, Y. et al. The function and clinical significance of eIF3 in cancer. *Gene*. **673**, 130–133 (2018).
24. Lin, Z., Xiong, L. & Lin, Q. Knockdown of eIF3d inhibits cell proliferation through G2/M phase arrest in non-small cell lung cancer. *Med. Oncol. Northw. Lond. Engl.* **32**, 183 (2015).
25. Yu, X., Zheng, B. & Chai, R. Lentivirus-mediated knockdown of eukaryotic translation initiation factor 3 subunit D inhibits proliferation of HCT116 colon cancer cells. *Biosci. Rep.* **34**, e00161 (2014).
26. Ren, M., Zhou, C., Liang, H., Wang, X. & Xu, L. RNAi-mediated silencing of EIF3D alleviates proliferation and migration of glioma U251 and U87MG cells. *Chem. Biol. Drug Des.* **86**, 715–722 (2015).
27. Gao, Y. et al. The oncogenic role of EIF3D is associated with increased cell cycle progression and motility in prostate cancer. *Med. Oncol. Northwood Lond. Engl.* **32**, 518 (2015).
28. Li, H. et al. Knockdown of EIF3D suppresses proliferation of human melanoma cells through G2/M phase arrest. *Biotechnol. Appl. Biochem.* **62**, 615–620 (2015).
29. Bu, H. et al. Diagnostic and prognostic potential of the novel biomarker nuclear cap binding protein subunit 2 (NCBP2) in colon adenocarcinoma. *J. Gastrointest. Oncol.* **13**, 1782–1792 (2022).
30. Wright, R. H. G. & Beato, M. Role of the NUDT enzymes in breast cancer. *Int. J. Mol. Sci.* **22**, 2267 (2021).
31. Page, B. D. G. et al. Targeted NUDT5 inhibitors block hormone signaling in breast cancer cells. *Nat. Commun.* **9**, 250 (2018).
32. Tao, X., Huang, R., Xu, R., Zheng, S. & Yue, J. A novel m7G methylation-related signature associated with chromosome homeostasis in patients with lung adenocarcinoma. *Front. Genet.* **13**, 998258 (2022).
33. Wang, D., Mo, Y., Zhang, D. & Bai, Y. Analysis of m7G methylation modification patterns and pulmonary vascular immune microenvironment in pulmonary arterial hypertension. *Front. Immunol.* **13**, 1014509 (2022).
34. Lu, L. et al. The m7G modification level and immune infiltration characteristics in patients with COVID-19. *J. Multidiscip. Healthc.* **15**, 2461–2472 (2022).
35. Teng, P.-C. et al. RNA modifications and epigenetics in modulation of lung cancer and pulmonary diseases. *Int. J. Mol. Sci.* **22**, 10592 (2021).
36. Tomikawa, C. 7-Methylguanosine modifications in transfer RNA (tRNA). *Int. J. Mol. Sci.* **19**, 4080 (2018).
37. Lu, F. et al. Construction of a novel prognostic model in lung adenocarcinoma based on 7-methylguanosine-related gene signatures. *Front. Oncol.* **12**, 876360 (2022).
38. Bossé, Y. Updates on the COPD gene list. *Int. J. Chron. Obstruct. Pulmon. Dis.* **7**, 607–631 (2012).
39. Wu, F. et al. Dynamic alteration profile and new role of RNA m6A methylation in replicative and H2O2-induced premature senescence of human embryonic lung fibroblasts. *Int. J. Mol. Sci.* **23**, 9271 (2022).
40. Wang, Y. et al. Biological and epigenetic alterations of mitochondria involved in cellular replicative and hydrogen peroxide-induced premature senescence of human embryonic lung fibroblasts. *Ecotoxicol. Environ. Saf.* **216**, 112204 (2021).
41. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
42. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
43. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci. Publ. Protein Soc.* **28**, 1947–1951 (2019).
44. Silvestri, G. A. & Young, R. P. Strange bedfellows: The interaction between COPD and lung cancer in the context of lung cancer screening. *Ann. Am. Thorac. Soc.* **17**, 810–812 (2020).
45. Huang, T. & He, W.-Y. Construction and validation of a novel prognostic signature of idiopathic pulmonary fibrosis by identifying subtypes based on genes related to 7-methylguanosine modification. *Front. Genet.* **13**, 890530 (2022).
46. Zhang, F. et al. EIF3D promotes gallbladder cancer development by stabilizing GRK2 kinase and activating PI3K-AKT signaling pathway. *Cell Death Dis.* **8**, e2868 (2017).
47. Wang, D., Jia, Y., Zheng, W., Li, C. & Cui, W. Overexpression of eIF3D in lung adenocarcinoma is a new independent prognostic marker of poor survival. *Dis. Markers*. **2019**, 6019637 (2019).
48. Zhang, W. et al. RBMS1 regulates lung cancer ferroptosis through translational control of SLC7A11. *J. Clin. Invest.* **131**, e152067(2021).
49. Xu, X. et al. m7G-related genes—NCBP2 and EIF4E3 determine immune contexture in head and neck squamous cell carcinoma by regulating CCL4/CCL5 expression. *Mol. Carcinog.* **62**, 1091–1106 (2023).
50. Betsuyaku, T. et al. Bronchiolar epithelial catalase is diminished in smokers with mild COPD. *Eur. Respir. J. Eur. Respir. Soc.* **42**, 42–53 (2013).
51. Kaarteenaho-Wiik, R. & Kinnula, V. L. Distribution of antioxidant enzymes in developing human lung, respiratory distress syndrome, and bronchopulmonary dysplasia. *J. Histochem. Cytochem. Off J. Histochem. Soc.* **52**, 1231–1240 (2004).
52. Antus, B., Paska, C., Simon, B. & Barta, I. Monitoring Antioxidant Enzyme Activity during Exacerbations of Chronic Obstructive Pulmonary Disease. *COPD J Chronic Obstr Pulm Dis.* **15**, 496–502 (2018).
53. Chunhua, M. et al. Betulin inhibited cigarette smoke-induced COPD in mice. *Biomed. Pharmacother.* **85**, 679–686 (2017).
54. Shah, S. W. A. et al. The effect of ammonia exposure on energy metabolism and mitochondrial dynamic proteins in chicken thymus: Through oxidative stress, apoptosis, and autophagy. *Ecotoxicol. Environ. Saf.* **206**, 111413 (2020).

55. Guo, J., Xing, H., Cai, J., Zhang, H. & Xu, S. H₂S exposure-induced oxidative stress promotes LPS-mediated hepatocyte autophagy through the PI3K/AKT/TOR pathway. *Ecotoxicol. Environ. Saf.* **209**, 111801 (2021).
56. Chen, J. et al. Immunosuppression, oxidative stress, and glycometabolism disorder caused by cadmium in common carp (*Cyprinus carpio* L.): Application of transcriptome analysis in risk assessment of environmental contaminant cadmium. *J. Hazard Mater.* **366**, 386–394 (2019).
57. Zhou, T. et al. Association between the oxidative stress gene polymorphism and chronic obstructive pulmonary disease risk: a meta-analysis. *BMC Pulm. Med.* **23**, 384 (2023).
58. Zeng, X. et al. Eliminating METTL1-mediated accumulation of PMN-MDSCs prevents hepatocellular carcinoma recurrence after radiofrequency ablation. *Hepatology*. **77**, 1122 (2023).
59. Kang, M.-J. et al. Suppression of NLRX1 in chronic obstructive pulmonary disease. *J. Clin. Invest.* **125**, 2458–2462 (2015).
60. Poon, J. et al. Cigarette smoke exposure reduces leukemia inhibitory factor levels during respiratory syncytial viral infection. *Int. J. Chron. Obstruct. Pulmon. Dis.* **14**, 1305–1315 (2019).
61. Mahalanobish, S., Dutta, S., Saha, S. & Sil, P. C. Melatonin induced suppression of ER stress and mitochondrial dysfunction inhibited NLRP3 inflammasome activation in COPD mice. *Food Chem. Toxicol.* **144**, 111588 (2020).
62. Hou, H. H. et al. Elastase induces lung epithelial cell autophagy through placental growth factor: A new insight of emphysema pathogenesis. *Autophagy*. **10**, 1509–1521 (2014).
63. Ritzmann, F. & Beisswenger, C. Preclinical studies and the function of IL-17 cytokines in COPD. *Ann. Anat. Anat. Anz.* **237**, 151729 (2021).
64. Chen, Z. & Hua, Y. Identification of m7G-related hub biomarkers and m7G regulator expression pattern in immune landscape during the progression of osteoarthritis. *Cytokine*. **170**, 156313 (2023).
65. Kool, M. & Lambrecht, B. N. Dendritic cells in asthma and COPD: Opportunities for drug development. *Curr. Opin. Immunol.* **19**, 701–710 (2007).
66. Wu, P. et al. Development of a prognostic risk model of uveal melanoma based on N7-methylguanosine-related regulators. *Hereditas*. **161**, 22 (2024).
67. Huang, S., Zhou, R., Yuan, Y. & Shen, Y. Stigmasterol alleviates airway inflammation in OVA-induced asthmatic mice via inhibiting the TGF- β 1/Smad2 and IL-17A signaling pathways. *Aging*. **16**, 6478–6487 (2024).
68. Zhang, L. et al. Th17/IL-17 induces endothelial cell senescence via activation of NF- κ B/p53/Rb signaling pathway. *Lab. Invest.* **101**, 1418–1426 (2021).
69. Zhong, J. et al. Shikonin ameliorates D-galactose-induced oxidative stress and cognitive impairment in mice via the MAPK and nuclear factor- κ B signaling pathway. *Int. Immunopharmacol.* **83**, 106491 (2020).
70. Hasegawa, M. et al. A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF- κ B activation. *EMBO J.* **27**, 373–383 (2008).

Author contributions

The conception and design of the study were collaboratively developed by C.Z., L.T., and X.Z. C.Z. and L.T. undertook the cell culture and molecular biology experiments. Data collection and analysis were carried out by C.Z. The preparation of figures and tables for the manuscript was managed by X.Z. and C.Z. The initial manuscript draft was edited by W.C. and M.L. And the revised manuscript and funding were edited and acquired by Y.C. and W.Z. All authors have read and approved the final manuscript for publication, thereby ensuring the integrity and accuracy of the work.

Funding

This work was supported by the Natural Science Foundation of Guangdong Province of China (2025A15150123 72, 2023A1515010353) and the National Natural Science Foundation of China (81473014).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-93901-w>.

Correspondence and requests for materials should be addressed to Y.C. or W.Z.

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

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

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

© The Author(s) 2025