5-methylcytosine RNA methylation regulators affect prognosis and tumor microenvironment in lung adenocarcinoma

Taisheng Liu[1,2#^], Xiaoshan Hu[3#], Chunxuan Lin[4#], Xiaoshun Shi[1], Yujing He[1], Jian Zhang[5,6^], Kaican Cai[1^]

[1]Department of Thoracic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China; [2]Department of Thoracic Surgery, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China; [3]Department of Internal Medicine of Oncology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China; [4]Department of Pneumology, Guangdong Provincial Hospital of Integrated Traditional Chinese and Western Medicine, Foshan, China; [5]Department of Radiation Oncology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, State Key Laboratory of Respiratory Diseases, Guangzhou Institute of Respiratory Disease, Guangzhou, China; [6]Guangzhou Medical University, Guangzhou, China

Contributions: (I) Conception and design: T Liu, J Zhang, K Cai; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: X Hu, C Lin, X Shi, Y He; (V) Data analysis and interpretation: T Liu, X Shi, J Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Kaican Cai. Department of Thoracic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China. Email: caican@smu.edu.cn; Jian Zhang. Department of Radiation Oncology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, State Key Laboratory of Respiratory Diseases, Guangzhou Institute of Respiratory Disease, Guangzhou 510095, China; Guangzhou Medical University, Guangzhou 511436, China. Email: zhangjian@gzhmu.edu.cn.

Background: Accumulating evidence has shown that 5-methylcytosinec (m5C) RNA methylation plays an essential role in tumorigenesis. However, the roles of m5C regulators in the prognosis, tumor microenvironment (TME), and immunotherapy responses of lung adenocarcinoma (LUAD) have not been fully analyzed.

Methods: Based on 14 m5C RNA regulators, we evaluated the m5C RNA modification patterns in patients with LUAD (n=594) in The Cancer Genome Atlas (TCGA). Unsupervised clustering analysis was performed to confirm distinct m5C modification patterns. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to investigate the biological functions of differentially expressed genes (DEGs) among different m5C RNA modification patterns. An m5C signature (m5Csig) was constructed using least absolute shrinkage and selection operator (LASSO) algorithms. The GSE72094 cohort (n=442) from the Gene Expression Omnibus (GEO) was used to validate m5Csig. A receiver operating characteristic (ROC) model was constructed to evaluate the sensitivity and specificity of m5Csig. Tumor-infiltrating immune cells (TIICs) between the high- and low-risk groups were estimated using the Cell Type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) algorithm.

Results: We identified 3 m5C RNA modification clusters. Overall survival (OS) differed among the 3 clusters. The m5Csig, including *TRDMT1*, *NSUN1*, *NSUN4*, *NSUN7*, and *ALYREF*, was constructed to classify patients with LUAD into high- and low-risk groups. The high-risk group, with more immune cell infiltration, had a significantly poorer OS than that the low-risk group, which was associated with better response to immune checkpoint blockade therapy.

Conclusions: The present study revealed that m5C RNA regulators play a significant role in TME regulation in LUAD. The m5Csig can predict the prognosis of patients with LUAD and might provide novel strategies for tumor immunotherapy.

^ ORCID: Taisheng Liu, 0000-0002-0132-4707; Jian Zhang, 0000-0002-0557-886X; Kaican Cai, 0000-0002-6805-4107.

Keywords: Lung adenocarcinoma (LUAD); 5-methylcytosine RNA methylation; tumor microenvironment; immunotherapy

Submitted Jan 10, 2022. Accepted for publication Mar 04, 2022.
doi: 10.21037/atm-22-500
View this article at: https://dx.doi.org/10.21037/atm-22-500

Introduction

Lung cancer is one of the leading causes of cancer-related morbidity and mortality worldwide (1). Non-small cell lung carcinoma (NSCLC), accounting for 85% of all lung cancers, mainly comprises lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD), with LUAD being the most common NSCLC subtype (2). With advances in targeted therapy and immunotherapy (3,4), the 5-year survival rate of patients with NSCLC is still unsatisfactory, at 4–17% (5). Patients with the same clinical characteristics can have distinctly different prognoses because of molecular differences. Therefore, there is an urgent need to confirm new molecular targets to improve the clinical treatment outcome in patients with LUAD.

Methylation of RNA, an important epigenetic modification that includes 5-methylcytosine (m5C), N6-methyladenosine (m6A), N1-methyladenosine (m1A), pseudouridine (Ψ), and inosine (I), has been identified to decorate protein-coding messenger RNAs (mRNAs) and noncoding RNAs (ncRNAs) (6-10). Modifications of RNA play crucial roles in RNA translation, transcription, processing, stability, and splicing (11,12), and m5C is one of the most common RNA modifications (13). The m5C RNA methylation can be catalyzed dynamically by a series of significant mediator proteins known as "writers" [tRNA aspartic acid methyltransferase 1 (TRDMT1), NOP2 nucleolar protein (NSUN1), NOP2/Sun RNA methyltransferase 2 (NSUN2), NSUN3-7], "readers" [Aly/REF export factor (ALYREF) and Y-box binding protein 1 (YBX1)], and "erasers" [tetmethylcytosine dioxygenase 1 (TET1), TET2-3]" (13-16). Dysregulation and disorder of m5C are associated with the occurrence of human diseases, including malignancies (17-19).

In recent years, immune checkpoint blockade (ICB) has made great breakthroughs in clinical efficacy for patients with cancer (20). However, only a small number of patients benefit from ICB (21). Numerous studies have identified that the tumor microenvironment (TME), which contains immune cells (such as T and B lymphocytes, natural killer (NK) cells, macrophages, polymorphonuclear cells, dendritic cells, as well as mast cells) and stromal cells, plays a crucial role in tumor progression, immunotherapy response, and immune escape (22,23). However, the relationship among m5C, immunotherapy response, and the TME in LUAD remains unclear. Therefore, a comprehensive understanding of the effect of m5C regulators on the TME might provide new insights into the immune regulation of the TME.

In this study, we analyzed 14 m5C RNA methylation regulators in LUAD from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. We identified that m5C regulators were closely associated with LUAD prognosis, and then constructed an m5C signature (m5Csig) to predict the LUAD survival and evaluate the response to ICB. Overall, the results indicated that m5Csig could act as a biomarker to predict survival and the response of ICB in LUAD. We present the following article in accordance with the TRIPOD reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-500/rc).

Methods

Acquisition of data

The RNA sequencing data (n=594), somatic mutation information (n=561), copy number variation (CNV) information (n=555), and the corresponding clinicopathological features (n=522) of patients with LUAD were downloaded from TCGA (https://portal.gdc.cancer.gov/). To validate the findings in the TCGA database, the validation cohort GSE72094 (n=442) was downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Protein-protein interaction analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/) was used to analyze protein-protein interaction (PPI) information and detect the interactions of 14 m5C regulators. We then

extracted PPI pairs with a combined score of 0.4.

Unsupervised clustering for 14 m5C regulators

Unsupervised clustering analysis was used to identify different m5C RNA modification patterns among patients with LUAD (n=535) from TCGA. The 14 m5C regulators included 8 writers (TRDMT1, NSUN1–7), 2 readers (ALYREF and YBX1), and 4 erasers [TET1–3, AlkB homolog 1, histone H2A dioxygenase (ALKBH1)]. The consensus clustering algorithm was employed to categorize patients with LUAD into different modification patterns (24). The consensus ClusterPlus package (https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html) was used to perform the above steps with a cycle computation of 1,000 iterations to guarantee the stability and reliability of the results (25). The overall survival (OS) rates of patients with the 3 modification patterns were calculated using the Kaplan-Meier method.

Identification of differentially expressed genes between m5C modification patterns

The empirical Bayesian approach in the limma R package (https://bioconductor.org/packages/release/bioc/html/limma.html) was applied to identify differentially expressed genes (DEGs) among the different m5C modification patterns in the standard comparison mode (26). The significance criteria for determining DEGs was set as an adjusted P value <0.001. To identify the potential functions and pathways enriched in the different modification patterns, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed based on the DEGs identified among the different modification patterns (27).

Development of the m5C regulators-related prognostic signature

Univariate Cox regression analysis was performed to identify m5C RNA methylation regulators that were associated with the OS of patients with LUAD. A least absolute shrinkage and selection operator (LASSO) Cox regression algorithm was carried out to construct an optimal m5Csig to predict the prognosis of patients with LUAD. Then, the obtained prognosis-associated genes were used to construct the m5Cscore function to calculate the score for each patient. The m5Cscore formula we used as follows:

$$\text{m5Cscore} = \sum_{i=1}^{n} coef_i \times expr_i \quad [1]$$

where m5Cscore is a prognostic risk score for patients with LUAD patients, $coef_i$ represents the coefficient, and $expr_i$ represents the expression of each prognostic gene. According to the median m5Cscore, the patients with LUAD were classified into high- and low-risk groups. The Kaplan-Meier method with the log-rank test was used to evaluate the OS differences between the high- and low-risk groups, and a receiver operating characteristic curve (ROC) was used to evaluate the prediction accuracy of m5Csig. Univariate and multivariate Cox regression analyses were used to explore prognostic values of m5Csig and clinical characteristics.

Estimation of TME cell infiltration

According to the method used by Newman *et al.* (28), 22 types of tumor-infiltrating immune cells (TIICs) between the high- and low-risk groups were estimated using the Cell Type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) algorithm.

Building and validation of a nomogram

Clinical factors [age, gender, stage and tumor-node-metastasis (TNM) stage] and the m5Cscore were used to develop a prognostic nomogram to evaluate the probability of 1-, 2-, and 3-year OS for patients with LUAD (29). The C-index and a calibration plot were constructed to estimate the accuracy and consistency of the m5Cscore.

Statistical analysis

Spearman and distance correlation analyses were used to compute the correlation coefficients among the expression levels of m5C regulators. The Wilcoxon test was used to analyze the difference between 2 groups, and the Kruskal–Wallis test and one-way analysis of variance (ANOVA) were used among 3 or more groups. LASSO Cox regression and Kaplan-Meier analyses were performed to construct and evaluate the m5Cscore. The area under the receiver operating characteristic curve (AUC) was used to investigate the time-dependent prognostic value of the m5Cscore. Multivariate Cox regression and stratified analysis were used to verify the independence of the m5Cscore from other clinical factors. Statistical significance was set at $P<0.05$, and all statistical P values were 2-sided. All data

were processed using R 4.0.3 software (R Foundation for Statistical Computing, Vienna, Austria).

Results

Landscape of genetic variation among m5C regulators in LUAD

The workflow of our study is shown in Figure S1. A total of 14 m5C regulators including 8 writers (TRDMT1, NSUN1–7), 2 readers (ALYREF and YBX1), and 4 erasers (TET1–3, ALKBH1) were included in our study (Table S1). First, the frequency of CNVs and somatic mutations of the 14 m5C regulators were investigated in LUAD. The CNV alteration frequency indicated that CNV alterations were ubiquitous among the 14 m5C regulators. As shown in *Figure 1A* and Table S2, *NSUN2* (13.69% amplification *vs.* 1.80% deletion), *ALYREF* (10.81% amplification *vs.* 1.44% deletion), *YBX1* (7.39% amplification *vs.* 1.80% deletion), and *NSUN4* (6.13% amplification *vs.* 2.52% deletion) were associated with amplification of the copy number, while *ALYBH1* (4.86% deletion *vs.* 1.80% amplification) and *NSUN1* (5.95% deletion *vs.* 4.32% amplification) were frequently deleted. The distribution of CNV alteration of m5C regulators on chromosomes is shown in *Figure 1B*. The analysis showed that 13.19% of patients with LUAD (n=74) experienced mutations of m5C regulators. The highest mutation frequency was exhibited by *TET1* (4%) followed by *TET2* (2%) and *TET3* (2%), while the genes including the writers (*NSUN1*, *NSUN3–5* and *NSUN7*), readers (*ALYREF* and *YBX1*) had no mutations in the patients with LUAD (*Figure 1C*). To explore the relationship between CNV alteration and the expression of m5C regulators, we analyzed the mRNA expression levels of the regulators. The results indicated that the expression levels of NSUN1, NSUN2, NSUN4–7, ALKBH1, TET1, TET3, and ALYREF were significantly upregulated in LUAD (P<0.001), whereas the expression level of TRDMT1 was significantly downregulated in LUAD (P<0.001). No significant difference was found in the expression levels of TET2, YBX1, and NSUN3 (*Figure 1D,1E*). These analyses showed CNV might play a crucial role in the imbalanced expression of m5C regulators, which could affect the occurrence and progression of LUAD. The clinicopathological features of the patients with LUAD are summarized in Table S3.

Correlation and interaction between m5C regulators

To determine the crosstalk among m5C regulators, correlation analysis was performed, which showed mainly positive correlations among the m5C regulators; however, several regulators exhibited a negative correlation among the m5C regulators. For example, TET2 and NSUN3 had the strongest positive correlation (r=0.7, P<0.001), whereas the correlation between TET2 and NSUN5 was negative (r=–0.30, P<0.001). Weak correlations were observed between TRDMT1 and other regulators (YBX1, NSUN5, ALYREF, NSUN1, NSUN2, NSUN4, and ALKBH1) (*Figure 1F* and Table S4). The PPI network analysis indicated that the 14 m5C regulators interacted with each other and TRDMT1 was one of the hub genes (*Figure 1G,1H*).

Network analyses of m5C regulators

Univariate Cox regression analysis showed the prognostic values of 14 m5C regulators in patients with LUAD, and each regulator had a different prognostic value (*Figure 2A*). Among the m5C regulators, TRDMT1, NSUN1, NSUN4, NSUN7, and ALYREF were related to the prognosis of patients with LUAD (P<0.05). The interactions, connections, and prognostic values of the m5C regulators are depicted in the m5C regulatory network (*Figure 2B*). The strongest positive correlation was observed between TET2 and NSUN3 (r=0.70, P<0.001), while the strongest negative correlation was observed between TET2 and NSUN5 (r=–0.30, P<0.001) (Tables S5,S6).

Consensus clustering of m5C regulators identified 3 clusters with different clinical outcomes

To explore whether the expression levels of m5C regulators were associated with prognosis, consensus clustering analysis was applied to classify patients with LUAD in TCGA cohort into subgroups based on their consensus expression of m5C regulators. It was found that K=3 had optimal clustering stability to classify the patients with LUAD into 3 clusters, namely m5C clusters 1–3 (Figure S2A-S2H). Patients in m5C cluster 3 had a significantly poorer OS than patients in cluster 1 and cluster 2 (*Figure 2C*, P=0.032). Furthermore, to identify enriched functions and pathways among the clusters, GO and KEGG analyses were conducted based on the DEGs identified among the m5C clusters. The results indicated that the DEGs were enriched in various processes, including RNA transport, spliceosome, mitotic nuclear division, and chromosome segregation (*Figure 2D,2E*). All significant (P<0.05) GO terms and KEGG pathways for the DEGs among 3 clusters are shown in Tables S7,S8.

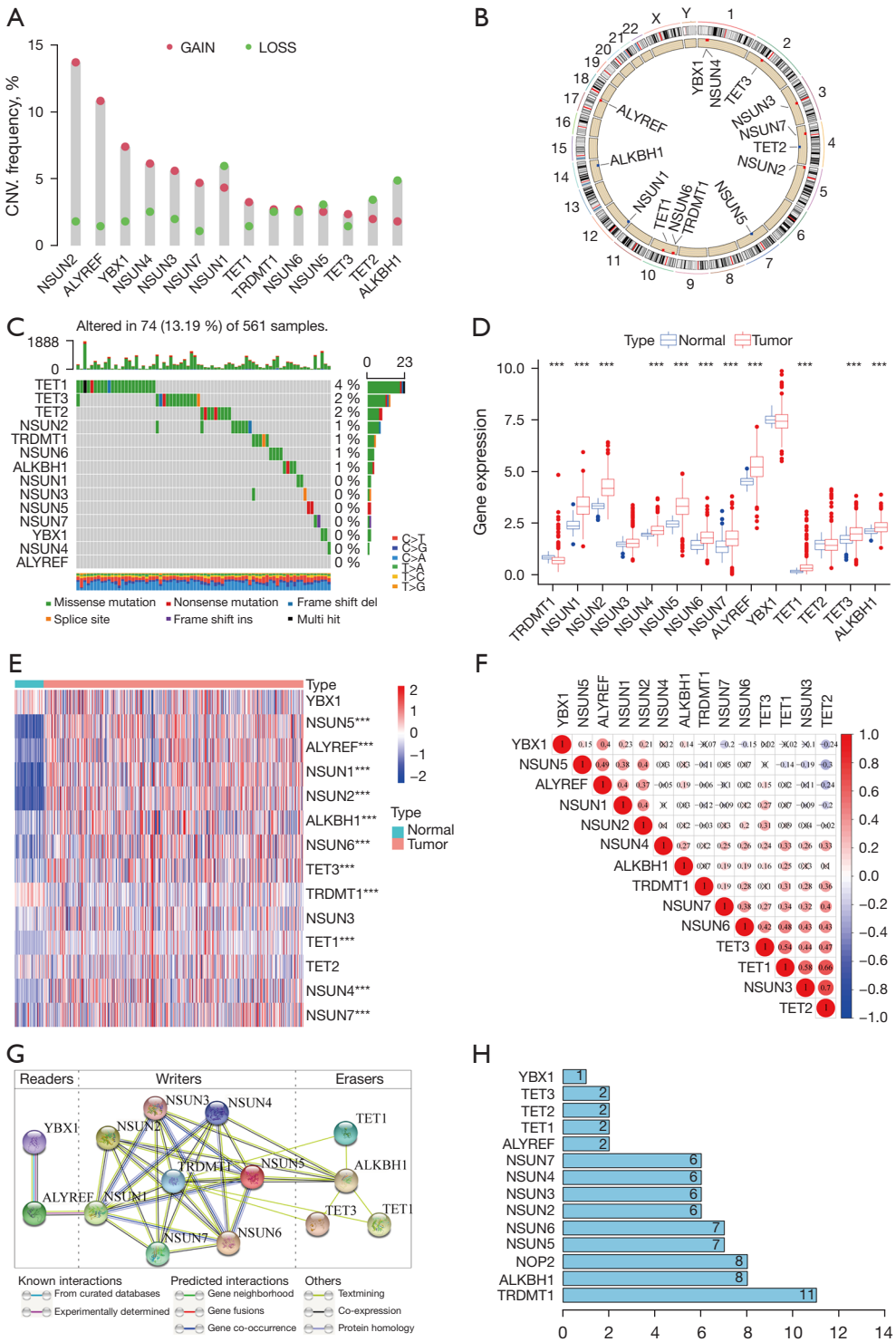

Figure 1 Landscape of the genetics and expression analysis of m5C regulators in TCGA-LUAD. (A) The CNV frequency of 14 m5C regulators; (B) the distribution of CNV alterations of the 14 m5C regulators on 23 chromosomes; (C) the mutation frequency of the 14 m5C regulators in 561 patients with LUAD; (D) the expression levels of the 14 m5C regulators between LUAD tissues and normal tissues; (E) heatmap of the 14 m5C regulators between LUAD tissues and normal tissues; (F) correlations among the 14 m5C regulators; (G) PPI network of the 14 m5C regulators; (H) the interaction numbers of each regulator with the other 13 regulators. ***P<0.001. TCGA-LUAD, The Cancer Genome Atlas lung adenocarcinoma cohort; CNV, copy number variation; PPI, protein-protein interaction.

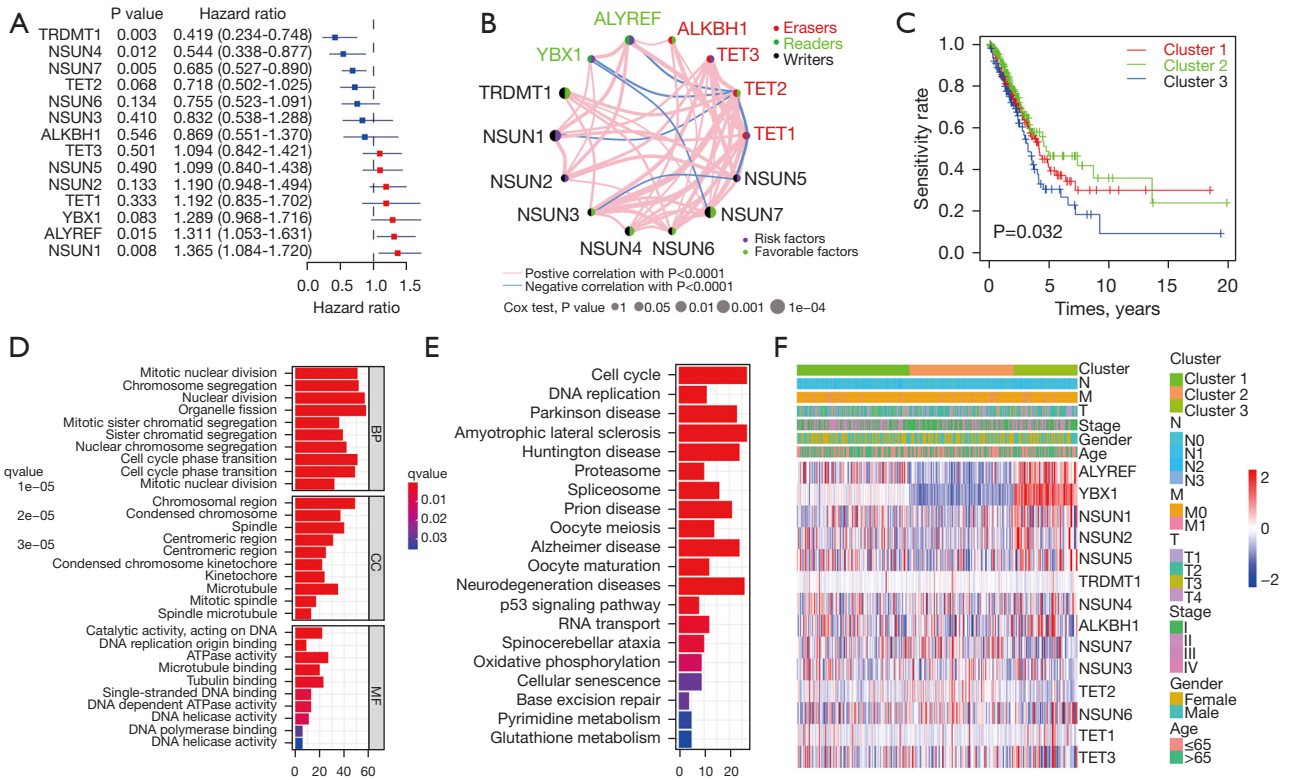

Figure 2 Prognostic analysis of 14 m5C regulators and patterns of m5C methylation modification in TCGA-LUAD. (A) Prognostic analyses for 14 m5C regulators using a univariate Cox regression model. (B) The network of m5C regulators in LUAD. The lines linking regulators represent their interactions, and the thickness of the lines represents the correlation strength between regulators. (C) Kaplan-Meier curve analysis for patients with LUAD in clusters 1–3. (D,E) Functional annotation of DEGs among the 3 clusters using GO (D) and KEGG (E) analysis. (F) Heatmap and clinicopathological features of the three clusters classified by the m5C regulators. TCGA-LUAD, The Cancer Genome Atlas lung adenocarcinoma cohort; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Furthermore, the associations among the 3 clusters, clinicopathological features, and expression levels of the 14 m5C regulators in the TCGA-LUAD cohort were evaluated. As shown in *Figure 2F*, the expression levels of the m5C regulators were higher in cluster 3, especially for ALYREF and YBX1.

Prognostic analysis of m5Csig in the TCGA-LUAD

As mentioned above, 5 m5C regulators, TRDMT1, NSUN1, NSUN4, NSUN7, and ALYREF, were associated with the prognosis of patients with LUAD according to the results of univariate Cox regression analysis (*Figure 3A*). The regulators TRDMT1, NSUN4, and NSUN7 act as protective factors, whereas NSUN1 and ALYREF are associated with risk of LUAD. The 5 m5C regulators were incorporated to build m5Csig according to the LASSO Cox regression algorithm. The regression coefficients of the 5 m5C regulatory factors are as follows: TRDMT1, −0.519056; NOP2, 0.166168; NSUN4, −0.376147; NSUN7, −0.246224; ALYREF, 0.163589. The patients with LUAD with complete clinical information (n=500) were classified into a high-risk group (n=250) and a low-risk group (n=250) according to the median m5Cscore, which was used as the cutoff point. As shown in *Figure 3B*, the expression levels of the risk-associated m5C regulators, NSUN1 and ALYREF, were upregulated in the high-risk group, and those of NSUN4, TRDMT1, and NSUN7 were downregulated in the high-risk group. With the increasing m5Cscore, the number of patients who died increased significantly (*Figure 3C,3D*). The Kaplan-Meier curve revealed that

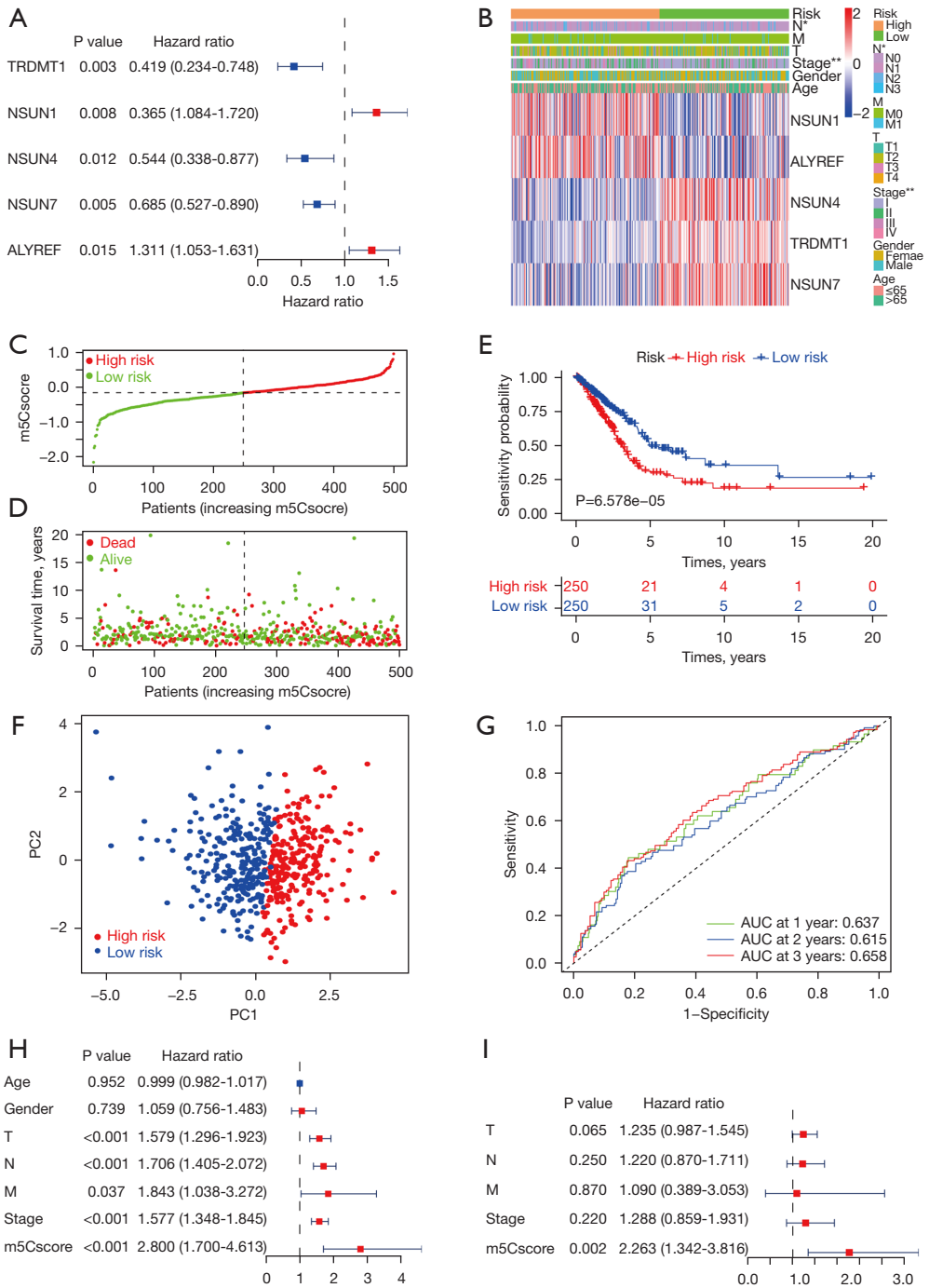

Figure 3 Construction of m5Csig using 5 m5C regulators and the prognostic value of m5Csig in TCGA-LUAD. (A) A forest plot of univariate Cox regression identified 5 m5C regulators associated with overall survival. (B) The relationship between the expression profiles of the m5C regulators stratified by m5Cscore and the clinicopathological features of LUAD. (C,D) Survival status (C) and distribution (D) with increasing m5Cscore of patients with LUAD. (E) Kaplan-Meier survival curve of the high- and low-risk groups. (F) PCA of the TCGA cohort. (G) ROC curve and the AUC value of m5Csig for 1-, 2-, and 3-year overall survival. (H,I) Univariate (H) and multivariate (I) Cox regression analyses of clinicopathological features and m5Cscore. *P<0.05, **P<0.01. TCGA-LUAD, The Cancer Genome Atlas lung adenocarcinoma cohort; PCA, principal component analysis; ROC, receiver operating characteristic; AUC, area under the curve.

the patients in the high-risk group had a significantly poorer OS than those in the low-risk group (P=6.578e-05, *Figure 3E*). Principal component analysis (PCA) analysis showed that the high- and low-risk groups were stratified significantly in 2 different directions, indicating that the patients with LUAD in the high-risk group could be distinguished from those in the low-risk group (*Figure 3F*). The time-dependent ROC analysis indicated that the AUC values of m5Csig for 1-, 2-, and 3-year OS were 0.637, 0.615, and 0.658, respectively (*Figure 3G*). Next, univariate and multivariate Cox regression analyses were used to analyze whether the m5Cscore can be used as an independent prognostic factor. Univariate Cox regression analysis showed that T [hazard ratio (HR) =1.579, 95% confidence interval (CI): 1.296–1.923, P<0.001], N (HR =1.706, 95% CI: 1.405–2.072, P<0.001), M (HR =0.037, 95% CI: 1.038–3.272, P=0.037), stage (HR =1.577, 95% CI: 1.348–1.845, P<0.001) and m5Cscore (HR =2.800, 95% CI: 1.700–4.623, P<0.001) were significantly correlated with OS (*Figure 3H*). However, no significant correlation was found between age and gender and OS. Multivariate Cox regression analysis indicated that only m5Cscore (HR =2.263, 95% CI: 1.342–3.816, P=0.002) can be used as an independent prognostic factor for LUAD (*Figure 3I*). These results indicated that the m5Cscore has the potential to predict prognosis in LUAD patients.

Validation of m5Csig in the GEO database

We validated m5Csig in the GSE72094 cohort. In total, 397 patients with complete clinical information were stratified into the high-risk group (n=198) and the low-risk group (n=199) using the median m5Cscore. As the m5Cscore increased, the number of deaths among the patients increased (*Figure 4A,4B*). Patients in the low-risk group had a better OS than those in the high-risk group (P=1.58e-03, *Figure 4C*). The PCA analysis suggested that patients were appropriately classified into high- and low-risk groups (*Figure 4D*). The ROC curves showed that the AUC values for 1-, 2-, and 3-year OS were 0.651, 0.615, and 0.59 respectively (*Figure 4E*). Univariate and multivariate Cox regression analyses also demonstrated that the m5Cscore can be used as an independent prognostic factor for LUAD patients (*Figure 4F,4G*).

Clinical characteristics between the high- and low-risk groups

A stratification analysis was performed to evaluate whether the m5Cscore could predict survival with the same clinical factor subgroup. Patients were stratified based on clinical parameters, such as age (≤65/>65 years), gender (female/male), T (T1+2/T3+4), N (N0/N1–3), M (M0/M1), and stage (I+II/III+IV). The results showed that the m5Cscore could classify patients of the same stratum of age, gender, and early stage (T1–2, N0, M0 and stage I+II) into high- and low-risk groups (P<0.05). Patients in the high-risk group had a poorer OS than those in the low-risk group in each stratum (Figure S3A-S3L). We further analyzed the differences of clinical characteristics between the high- and low-m5Cscore groups and the difference of m5Cscore among different clinical characteristics. No significant distribution difference was found in terms of age (≤65/>65 years) (P=0.15, Figure S4A,S4B), gender (female/male) (P=0.37, Figure S4C,S4D), stage I and stage II (P=0.19), stage I and stage III (P=0.33), and stage II and stage III (P=0.87) (Figure S4E,S4F). However, significant clinical differences were observed in terms of stage I and stage IV (P=0.043), stage II and stage IV (P=0.018), stage III and stage IV (P=0.023) (Figure S4E,S4F), ever smoking and never smoking (P=0.046, *Figure 5A,5B*), *EGFR* mutation group and *EGFR* wild group (P=4.2e-05, *Figure 5C,5D*), *KRAS* mutation group and *KRAS* wild group (P=0.0032, *Figure 5E,5F*), and *TP53* mutation group and *TP53* wild group (P=0.006, *Figure 5G,5H*).

Tumor mutation burden in the high- and low- risk groups in the TCGA-LUAD database

The tumor mutation burden (TMB) quantification analyses indicated that the high-risk group correlated remarkably with a higher TMB (P<0.001, *Figure 5I*). The m5Cscore and TMB also exhibited a significant positive correlation (R=0.24, P<0.001, *Figure 5J*). There was no difference in OS between the high- and low-TMB groups (P=0.089, *Figure 5K*). As shown in *Figure 5L*, when combined with the m5Cscore, there were significant survival differences among the 4 groups. The high-TMB/low-m5Cscore group had better survival than the high-TMB/high-m5Cscore group, and the low-TMB/high-m5Cscore group had the least favorable OS.

Expression of immune checkpoints and TME cell infiltration characteristics between the high- and low-risk groups in TCGA database

To determine the tumor immune infiltration characteristics,

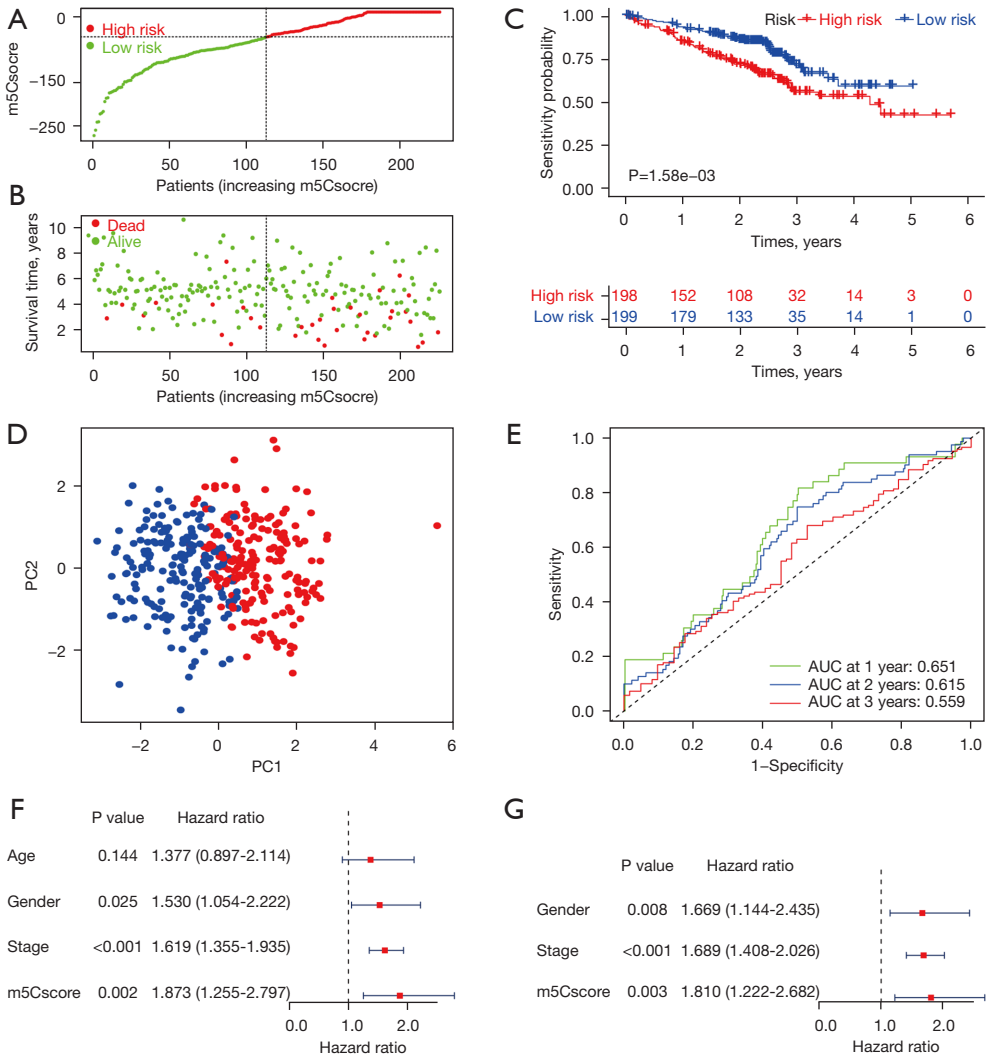

Figure 4 Validation of m5Csig in the GEO database. (A,B) Survival status (A) and distribution (B) of m5Cscore. (C) Kaplan-Meier survival analysis between the two groups. (D) PCA of the GEO cohort. (E) Time-dependent ROC and AUC based on the GEO data for 1-, 2-, and 3-year OS. (F-G) Univariate (F) and multivariate (G) Cox regression analyses of clinicopathological features and m5Cscore. GEO, Gene Expression Omnibus; PCA, principal component analysis; ROC, receiver operating characteristic; AUC, area under the curve; OS, overall survival.

we evaluated the expression of 24 immune checkpoints between the high- and low-risk groups. We found a substantial difference in the expression of 24 immune checkpoints, among which LAG3, PDCD1 (PD-1), TNFRSF4, CD274 (PD-L1), CD276, TNFRSF8, TMIGD2, TNFRSF9, TNFSF4, TNFSF9, KIR3DL1, TNFRSF18, and CD70 were upregulated significantly in the high-risk group (*Figure 6A*). We further analyzed the proportion of immune cells between the high- and low- risk groups in the TCGA-LUAD database. Heterogeneity of LUAD was indicated by the different ratios of each cell type (*Figure 6B*). Furthermore, we compared the infiltration of immune cells between the high- and low- risk groups. As shown in *Figure 6C*, the high-risk group had a higher fraction of CD8 T cells, activated CD4 memory T cells, follicular helper T cells, resting NK cells, and M0 macrophages compared with those in the low-risk group (P<0.05). However, naive B cells, resting CD4 memory T

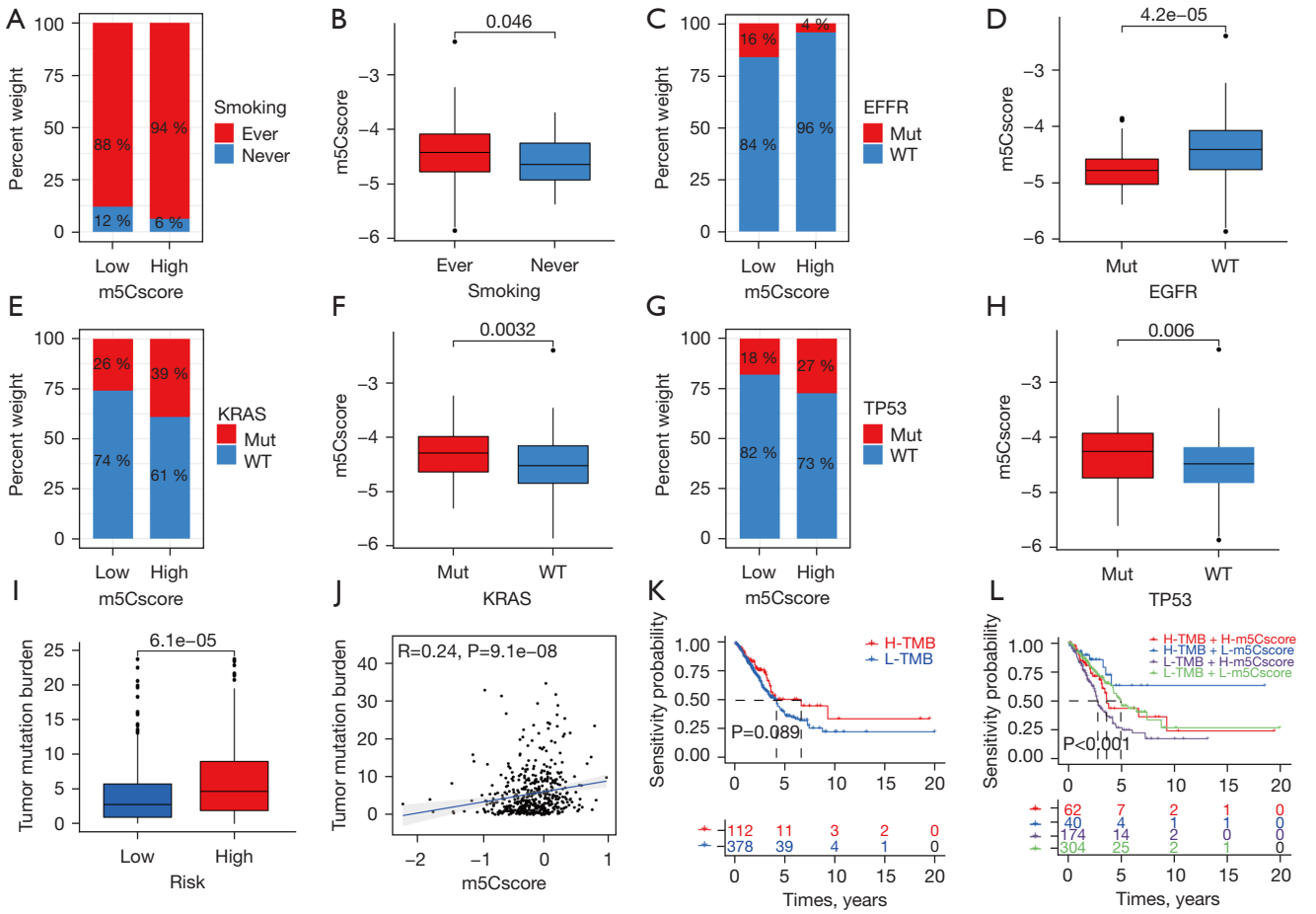

Figure 5 Clinical characteristics and TMB between high- and low-risk groups. (A-D) The proportion and distribution of different clinical characteristics between high- and low-risk groups in GSE72094: ever smoking/never smoking (A,B), EGFR mutation/EGFR wild (C,D), KRAS mutation/KRAS wild (E,F), TP53 mutation/TP53 wild (G,H). (I) Difference in TMB between high- and low-risk groups. (J) Correlation scatter plots between TMB and m5Cscore. (K) Kaplan-Meier curves of OS for high- and low-TMB groups. (L) Kaplan-Meier curves of overall survival stratified by both TMB and m5Cscore. TMB, tumor mutation burden; OS, overall survival.

cells, resting dendritic cells, and resting mast cells were markedly downregulated in the high-risk group (P<0.05).

Construction of a prognostic nomogram for LUAD in the TCGA data

To establish a clinically applicable method to evaluate the prognosis of patients with LUAD, we constructed a prognostic nomogram by integrating clinical factors (age, gender, stage) with the m5Cscore (Figure S5A). Using the bootstrap method, calibration plots showed no significant deviation from the ideal for 1-, 3- and 5-year OS (Figure S5B). These results indicated that the prognostic nomogram could be used to predict the prognosis of patients with LUAD.

Discussion

Abnormalities of m5C modifications have been shown to influence RNA stability, gene expression, and protein synthesis, and thus have an essential role in various cellular, biological, and pathological processes (20-32). The RNA m5C modification and its regulators have been shown to be involved in the progression of various cancers, including hepatocellular carcinoma (33), bladder cancer (34), glioblastoma multiforme (35), breast cancer (36), and head and neck carcinoma (37), indicating that RNA m5C might play an important role in tumorigenesis and progression.

However, the biological functions and mechanism by which m5C modifications affect the TME were previously unknown.

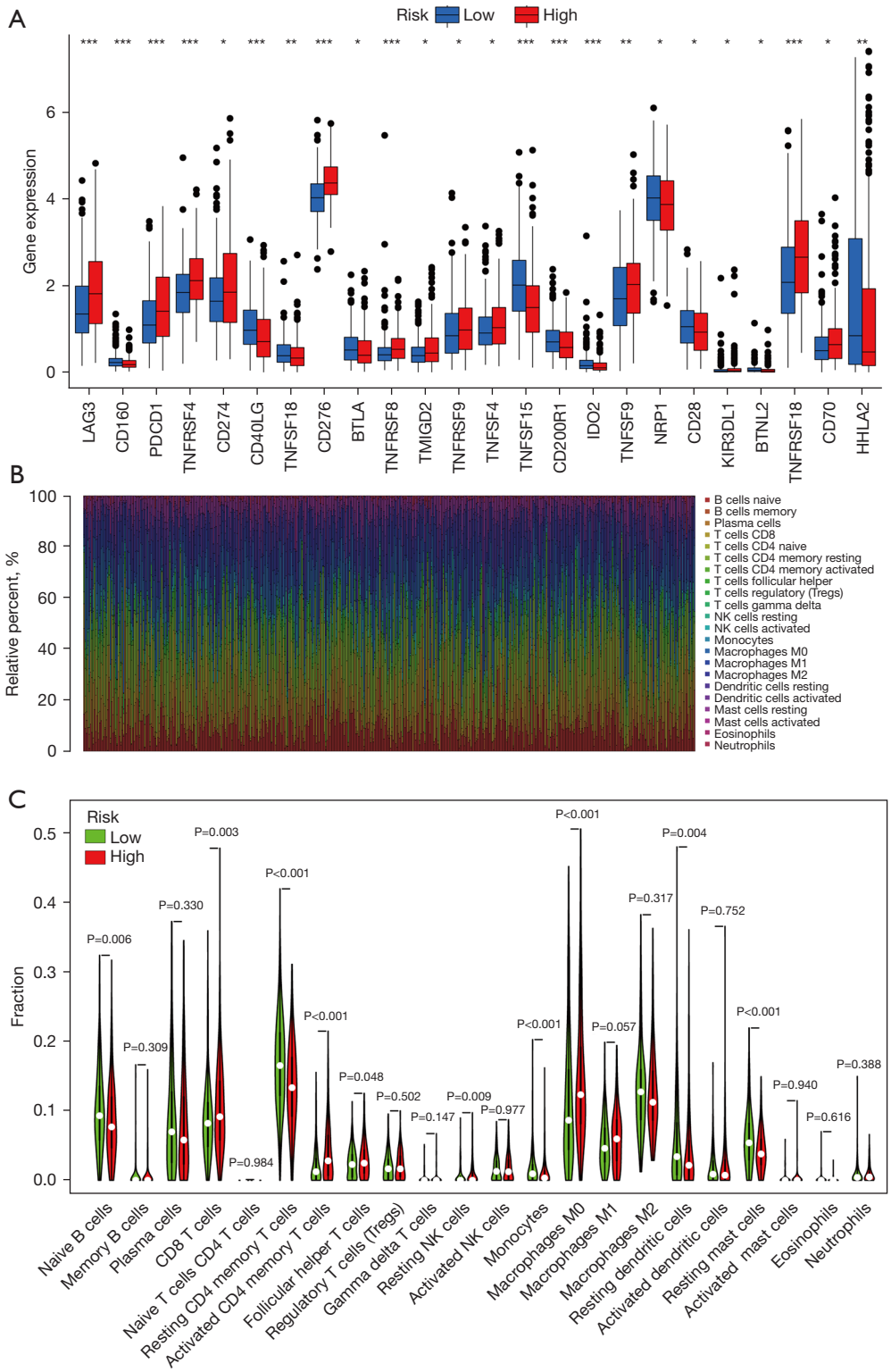

Figure 6 Expression of immune checkpoints and TME cell infiltration characteristics between high- and low-risk groups in the TCGA database. (A) Expression of immune checkpoints between high- and low-risk groups. (B) Barplot of the distribution of 22 immune cells in TCGA-LUAD. (C) TME cell infiltration characteristics between high- and low-risk groups. *P<0.05, **P<0.01, ***P<0.001. TCGA-LUAD, The Cancer Genome Atlas lung adenocarcinoma cohort; TME, tumor microenvironment.

In this study, we found that m5C regulators were significantly differently expressed in LUAD. By analyzing the expression profiles of m5C regulators, we identified 3 m5C clusters associated with different prognoses. Moreover, GO and KEGG function analysis indicated that DEGs among the 3 clusters were closely correlated with biological processes and signaling pathways, such as RNA transport, spliceosome, mitotic nuclear division, and chromosome segregation. We also identified 4 writers (TRDMT1, NSUN1, NSUN4, and NSUN7) and 1 eraser (ALYREF) were correlated with the prognosis of LUAD using LASSO Cox regression. Among the 5 m5C regulators, *TRDMT1* mainly mediates tRNA stability and regulates cell metabolism of the m5C modification (30,38,39). Loss of *TRDMT1* promoted homologous recombination and increased cellular sensitivity to DNA double-strand breaks (40). The NSUN1 protein (also known as NOP2) is a nucleolar-specific protein that plays a crucial role in RNA modification (41), cell cycle progression (41), chromatin organization (42), and HIV-1 latency (43). NSUN4, which forms a complex with MTERF4, is not essential in mitochondrial ribosome biogenesis and mitochondrial translation termination in conditional *Nsun4* mouse knockout mutants (44,45). High expression of NSUN7 has been associated with shorter survival in low-grade gliomas (46), and a deletion mutation of *NSUN7* has been associated with reduced sperm motility in asthenospermic men (47). The mRNA export adaptor, ALYREF, serves as a specific m5C-binding protein and functions in promoting mRNA export (48,49). An ALYREF-MYCN coactivator complex might be involved in neuroblastoma tumorigenesis (50).

An m5Csig was constructed, which divided patients with LUAD into high- and low-risk groups. Patients in the high-risk group had a significantly poorer OS than those in the low-risk group. Univariate and multivariate Cox regression analyses demonstrated that the m5Cscore was an independent prognostic factor for patients with LUAD. Accumulated evidence has demonstrated that patients overexpressing PD-1/PD-L1 and with a high TMB status are associated with an improved and durable ICB response (51-53). The TMB quantification analyses indicated that the high-risk group correlated markedly with a higher TMB. The m5Cscore and TMB also exhibited a significant positive correlation. The high-risk group displayed significantly higher expression levels of PD-1 and PD-L1 than the low-risk group. The above results demonstrated that LUAD with a high m5Cscore might show a better response to ICB therapy.

Accumulating evidence suggested that m5C is closely related to TME. Schoeler *et al.* demonstrated that TET enzymes control antibody production and shape the mutational landscape in germinal centre B cells. TET2 and TET3 guide the transition of germinal centre B cells to antibody-secreting plasma cells (54). Yue *et al.* revealed Tet2/3-deficiency in Treg cells leads to T cell activation and results in an activated phenotype and dysregulated expression of multiple Treg activation and phenotypic molecules in healthy mice (55). In our study, the CIBERSORT results showed that the high-risk group had stronger immune cell invasion compared with that of the low-risk group, for example, the numbers of CD8 T cells and activated CD4 memory T cells were significantly increased. These results suggested that the m5C regulators might be involved in the progression and prognosis of LUAD by modulating TIIC infiltration of the TME. Targeting m5C-related regulators might provide a novel way to improve the efficiency of ICB in LUAD.

However, there were several limitations to our study. First, this was a bioinformatic study based on a public database; therefore, further *in vivo* and *in vitro* experimental studies are needed to explore the potential effect and mechanism of m5C regulators in LUAD. Second, more potential m5C regulators have yet to be discovered. Last, the regulatory mechanism of m5C regulators in the TME was not determined, which requires further investigation to provide a better understanding.

Conclusions

In summary, we comprehensively analyzed the relationship between m5C methylation regulators and TME immune modulation. Based on the characteristics of m5C regulators, m5Csig was constructed to predict the prognosis of patients with LUAD, which might provide novel strategies for ICB therapy.

Acknowledgments

Funding: This study was supported by grants from the Science and Technology Tackling Program of Guangzhou for People's Livelihood (No. 201903010003), the National Natural Science Foundation of China (No. 82003212), the Discipline Construction Project of Guangzhou Medical University During the 14th Five-Year Plan (No. 06-410-2107181), the Guangzhou Key Medical Discipline

Construction Project Fund (No. 02-412-B205002-1004042), and the Medical and Health Technology Projects of Guangzhou, China (No. 2015A011086). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-500/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-500/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Chen Z, Fillmore CM, Hammerman PS, et al. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer 2014;14:535-46.
3. Peters S, Reck M, Smit EF, et al. How to make the best use of immunotherapy as first-line treatment of advanced/metastatic non-small-cell lung cancer. Ann Oncol 2019;30:884-96.
4. Hirsch FR, Suda K, Wiens J, et al. New and emerging targeted treatments in advanced non-small-cell lung cancer. Lancet 2016;388:1012-24.
5. Hirsch FR, Scagliotti GV, Mulshine JL, et al. Lung cancer: current therapies and new targeted treatments. Lancet 2017;389:299-311.
6. Roundtree IA, Evans ME, Pan T, et al. Dynamic RNA Modifications in Gene Expression Regulation. Cell 2017;169:1187-200.
7. Carlile TM, Rojas-Duran MF, Zinshteyn B, et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 2014;515:143-6.
8. Boccaletto P, Machnicka MA, Purta E, et al. MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res 2018;46:D303-7.
9. Li X, Xiong X, Yi C. Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods 2016;14:23-31.
10. Gilbert WV, Bell TA, Schaening C. Messenger RNA modifications: Form, distribution, and function. Science 2016;352:1408-12.
11. Oerum S, Dégut C, Barraud P, et al. m1A Post-Transcriptional Modification in tRNAs. Biomolecules 2017;7:20.
12. Liu J, Jia G. Methylation modifications in eukaryotic messenger RNA. J Genet Genomics 2014;41:21-33.
13. Hussain S, Aleksic J, Blanco S, et al. Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol 2013;14:215.
14. Squires JE, Patel HR, Nousch M, et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res 2012;40:5023-33.
15. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat Biotechnol 2013;31:458-64.
16. Nombela P, Miguel-López B, Blanco S. The role of m6A, m5C and Ψ RNA modifications in cancer: Novel therapeutic opportunities. Mol Cancer 2021;20:18.
17. Jonkhout N, Tran J, Smith MA, et al. The RNA modification landscape in human disease. RNA 2017;23:1754-69.
18. Xue C, Zhao Y, Li L. Advances in RNA cytosine-5 methylation: detection, regulatory mechanisms, biological functions and links to cancer. Biomark Res 2020;8:43.
19. Pan J, Huang Z, Xu Y. m5C RNA Methylation Regulators Predict Prognosis and Regulate the Immune Microenvironment in Lung Squamous Cell Carcinoma.

Front Oncol 2021;11:657466.
20. Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med 2015;372:2018-28.
21. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012;366:2443-54.
22. Fridman WH, Zitvogel L, Sautès-Fridman C, et al. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol 2017;14:717-34.
23. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol 2004;22:329-60.
24. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society 1979;28:100-8.
25. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26:1572-3.
26. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
27. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.
28. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.
29. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26:1364-70.
30. David R, Burgess A, Parker B, et al. Transcriptome-Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs. Plant Cell 2017;29:445-60.
31. King MY, Redman KL. RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. Biochemistry 2002;41:11218-25.
32. Cheng JX, Chen L, Li Y, et al. RNA cytosine methylation and methyltransferases mediate chromatin organization and 5-azacytidine response and resistance in leukaemia. Nat Commun 2018;9:1163.
33. He Y, Yu X, Li J, et al. Role of m5C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma. Am J Transl Res 2020;12:912-22.
34. Chen X, Li A, Sun BF, et al. 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. Nat Cell Biol 2019;21:978-90.
35. Cheray M, Etcheverry A, Jacques C, et al. Cytosine methylation of mature microRNAs inhibits their functions and is associated with poor prognosis in glioblastoma multiforme. Mol Cancer 2020;19:36.
36. Huang Z, Pan J, Wang H, et al. Prognostic Significance and Tumor Immune Microenvironment Heterogenicity of m5C RNA Methylation Regulators in Triple-Negative Breast Cancer. Front Cell Dev Biol 2021;9:657547.
37. Xue M, Shi Q, Zheng L, et al. Gene signatures of m5C regulators may predict prognoses of patients with head and neck squamous cell carcinoma. Am J Transl Res 2020;12:6841-52.
38. Tuorto F, Liebers R, Musch T, et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol 2012;19:900-5.
39. Huang T, Chen W, Liu J, et al. Genome-wide identification of mRNA 5-methylcytosine in mammals. Nat Struct Mol Biol 2019;26:380-8.
40. Chen H, Yang H, Zhu X, et al. m5C modification of mRNA serves a DNA damage code to promote homologous recombination. Nat Commun 2020;11:2834.
41. Kosi N, Alić I, Kolačević M, et al. Nop2 is expressed during proliferation of neural stem cells and in adult mouse and human brain. Brain Res 2015;1597:65-76.
42. Hong J, Lee JH, Chung IK. Telomerase activates transcription of cyclin D1 gene through an interaction with NOL1. J Cell Sci 2016;129:1566-79.
43. Kong W, Biswas A, Zhou D, et al. Nucleolar protein NOP2/NSUN1 suppresses HIV-1 transcription and promotes viral latency by competing with Tat for TAR binding and methylation. PLoS Pathog 2020;16:e1008430.
44. Metodiev MD, Spåhr H, Loguercio Polosa P, et al. NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly. PLoS Genet 2014;10:e1004110.
45. Cámara Y, Asin-Cayuela J, Park CB, et al. MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. Cell Metab 2011;13:527-39.
46. Sato K, Tahata K, Akimoto K. Five Genes Associated With Survival in Patients With Lower-grade Gliomas Were Identified by Information-theoretical Analysis. Anticancer Res 2020;40:2777-85.
47. Khosronezhad N, Hosseinzadeh Colagar A, Mortazavi SM. The Nsun7 (A11337)-deletion mutation, causes reduction of its protein rate and associated with sperm motility defect in infertile men. J Assist Reprod Genet 2015;32:807-15.
48. Zhou Z, Luo MJ, Straesser K, et al. The protein Aly

links pre-messenger-RNA splicing to nuclear export in metazoans. Nature 2000;407:401-5.
49. Yang X, Yang Y, Sun BF, et al. 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m5C reader. Cell Res 2017;27:606-25.
50. Nagy Z, Seneviratne JA, Kanikevich M, et al. An ALYREF-MYCN coactivator complex drives neuroblastoma tumorigenesis through effects on USP3 and MYCN stability. Nat Commun 2021;12:1881.
51. Cohen EEW, Soulières D, Le Tourneau C, et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. Lancet 2019;393:156-67.
52. Fumet JD, Truntzer C, Yarchoan M, et al. Tumour mutational burden as a biomarker for immunotherapy: Current data and emerging concepts. Eur J Cancer 2020;131:40-50.
53. Ahmed ME, Falasiri S, Hajiran A, et al. The Immune Microenvironment in Penile Cancer and Rationale for Immunotherapy. J Clin Med 2020;9:3334.
54. Schoeler K, Aufschnaiter A, Messner S, et al. TET enzymes control antibody production and shape the mutational landscape in germinal centre B cells. The FEBS journal 2019;286:3566-81.
55. Yue X, Lio CJ, Samaniego-Castruita D, Li X, et al. Loss of TET2 and TET3 in regulatory T cells unleashes effector function. Nat Commun 2019;10:2011.

(English Language Editor: J. Jones)

Cite this article as: Liu T, Hu X, Lin C, Shi X, He Y, Zhang J, Cai K. 5-methylcytosine RNA methylation regulators affect prognosis and tumor microenvironment in lung adenocarcinoma. Ann Transl Med 2022;10(5):259. doi: 10.21037/atm-22-500

Supplementary

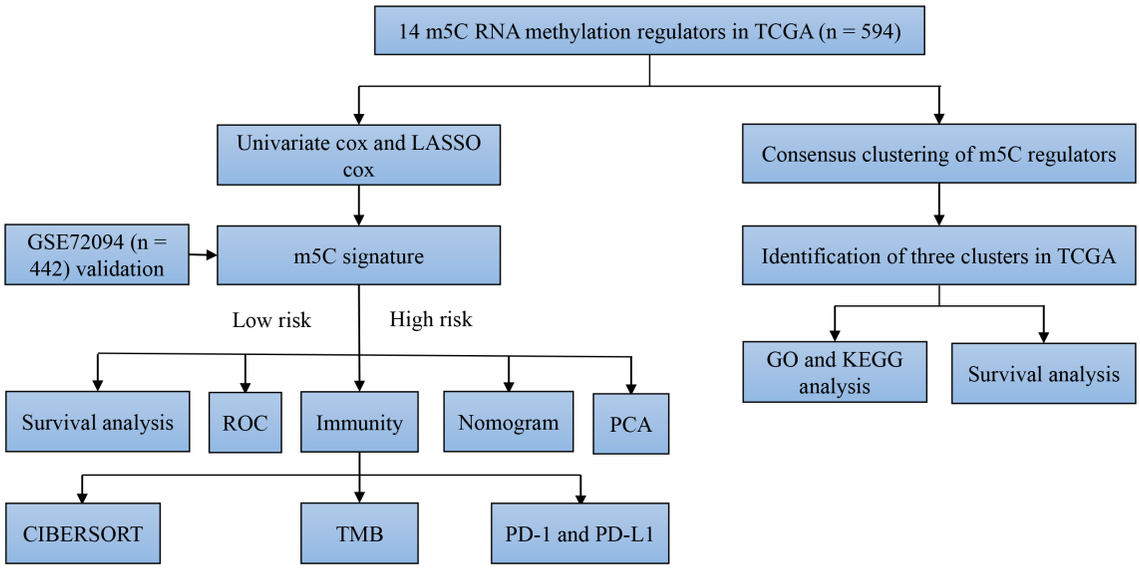

Figure S1 Study flowchart.

Table S1 The 14 RNA m5C methylation regulators enrolled in this study

m5C regulators	Type
TRDMT1	Writers
NSUN1	Writers
NSUN2	Writers
NSUN3	Writers
NSUN4	Writers
NSUN5	Writers
NSUN6	Writers
NSUN7	Writers
ALYREF	Readers
YBX1	Readers
TET1	Erasers
TET2	Erasers
TET3	Erasers
ALKBH1	Erasers

Table S2 CNV frequency of m5C regulators in TCGA-LUAD

Gene	Gain	Loss
NSUN1	4.324324324	5.94594595
ALKBH1	1.801801802	4.86486486
TET2	1.981981982	3.42342342
NSUN5	2.522522523	3.06306306
NSUN4	6.126126126	2.52252252
TRDMT1	2.702702703	2.52252252
NSUN6	2.702702703	2.52252252
NSUN3	5.585585586	1.98198198
NSUN2	13.69369369	1.8018018
YBX1	7.387387387	1.8018018
ALYREF	10.81081081	1.44144144
TET1	3.243243243	1.44144144
TET3	2.342342342	1.44144144
NSUN7	4.684684685	1.08108108

CNV, copy number variation; TCGA-LUAD, The Cancer Genome Atlas lung adenocarcinoma cohort.

https://dx.doi.org/10.21037/atm-22-500

Table S3 Clinical characteristics of the patients in the TCGA-LUAD and the GSE72094 cohort

Characteristics	TCGA-LUAD	GSE72094
n	522	441
Age		
≤65 years	241	127
>65 years	262	294
Unknow	19	21
Gender		
Female	280	240
Male	242	202
Stage		
I	279	265
II	124	69
III	85	63
IV	26	17
Unknow	8	28
T		
T1	172	–
T2	281	–
T3	47	–
T4	19	–
Unknow	3	–
N		
N0	335	–
N1	98	–
N2	75	–
N3	2	–
Unknow	12	–
M		
M0	353	–
M1	25	–
Unknow	144	–
Mutation		
EGFR		
Mutation		47
Wild		395
KRAS		
Mutation		154
Wild		288
TP53		
Mutation		111
Wild		331

TCGA-LUAD, The Cancer Genome Atlas lung adenocarcinoma cohort.

https://dx.doi.org/10.21037/atm-22-500

Table S4 Spearman correlation analysis of the 14 m5C regulators

From	To	Cor	P value	Weight
TET1	*TET2*	0.661743558	4.34E-76	3.970461348
TET1	*TET3*	0.544926884	3.06E-47	3.269561301
TET1	*ALKBH1*	0.250946716	5.55E-10	1.505680297
TET1	*TRDMT1*	0.311450679	7.93E-15	1.868704076
TET1	*NSUN3*	0.576112033	8.05E-54	3.456672195
TET1	*NSUN4*	0.329586303	1.62E-16	1.977517817
TET1	*NSUN6*	0.478985473	2.14E-35	2.873912835
TET1	*NSUN7*	0.335594184	4.21E-17	2.013565101
TET2	*TET3*	0.465789154	2.54E-33	2.794734922
TET2	*ALYREF*	–0.243438883	1.84E-09	1.460633296
TET2	*YBX1*	–0.243353803	1.87E-09	1.46012282
TET2	*TRDMT1*	0.357936995	2.15E-19	2.14762197
TET2	*NOP2*	–0.197848508	1.17E-06	1.187091047
TET2	*NSUN3*	0.699749267	1.58E-88	4.198495604
TET2	*NSUN4*	0.329682267	1.59E-16	1.978093604
TET2	*NSUN5*	–0.299809698	8.39E-14	1.798858185
TET2	*NSUN6*	0.432603482	1.74E-28	2.595620893
TET2	*NSUN7*	0.404451996	8.76E-25	2.426711977
TET3	*NOP2*	0.274301369	1.03E-11	1.645808215
TET3	*NSUN2*	0.313999629	4.66E-15	1.883997774
TET3	*NSUN3*	0.43530397	7.35E-29	2.611823817
TET3	*NSUN4*	0.235522735	6.25E-09	1.413136409
TET3	*NSUN6*	0.421756083	5.11E-27	2.5305365
TET3	*NSUN7*	0.274787916	9.45E-12	1.648727495
ALKBH1	*ALYREF*	0.193981745	1.91E-06	1.163890471
ALKBH1	*NSUN4*	0.266840392	3.84E-11	1.601042351
ALKBH1	*NSUN6*	0.187111572	4.40E-06	1.122669435
ALKBH1	*NSUN7*	0.191851004	2.48E-06	1.151106025
ALYREF	*YBX1*	0.404640807	8.29E-25	2.427844841
ALYREF	*NOP2*	0.399336235	3.79E-24	2.396017408
ALYREF	*NSUN2*	0.366117417	2.80E-20	2.196704503
ALYREF	*NSUN5*	0.488305836	6.44E-37	2.929835015
YBX1	*NOP2*	0.234600394	7.18E-09	1.407602366
YBX1	*NSUN2*	0.206088969	4.05E-07	1.236533812
YBX1	*NSUN7*	–0.195919248	1.50E-06	1.175515486
TRDMT1	*NSUN3*	0.277259403	6.05E-12	1.663556421
TRDMT1	*NSUN6*	0.276129986	7.42E-12	1.656779917
TRDMT1	*NSUN7*	0.186003201	5.03E-06	1.116019207
NOP2	*NSUN2*	0.401768722	1.89E-24	2.410612334
NOP2	*NSUN5*	0.377251562	1.58E-21	2.263509372
NSUN2	*NSUN5*	0.403146217	1.28E-24	2.418877304
NSUN2	*NSUN6*	0.196965583	1.31E-06	1.181793496
NSUN3	*NSUN4*	0.261138589	1.02E-10	1.566831532
NSUN3	*NSUN5*	–0.190912011	2.78E-06	1.145472065
NSUN3	*NSUN6*	0.427830452	7.81E-28	2.566982714
NSUN3	*NSUN7*	0.324232799	5.25E-16	1.945396793
NSUN4	*NSUN6*	0.262270365	8.44E-11	1.573622189
NSUN4	*NSUN7*	0.246962976	1.05E-09	1.481777855
NSUN6	*NSUN7*	0.376461356	1.95E-21	2.258768134

https://dx.doi.org/10.21037/atm-22-500

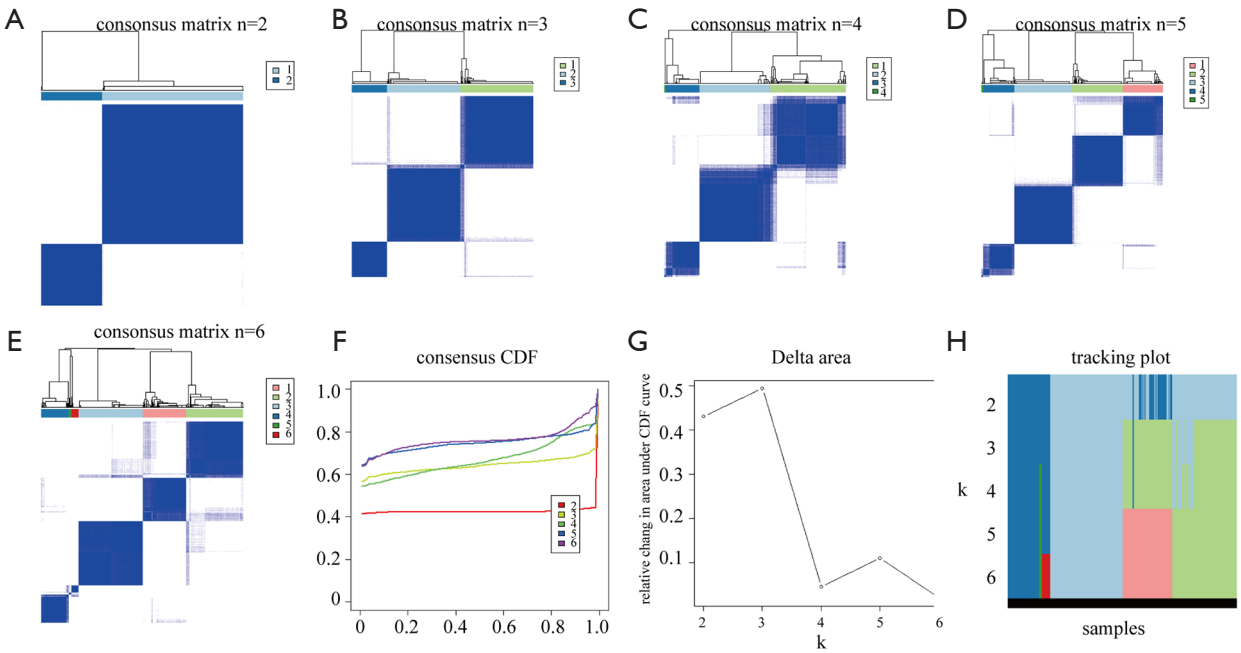

Figure S2 Unsupervised clustering of 14 m5C regulators in TCGA-LUAD. (A-E) Consensus clustering matrix for k=2–6, k=2 (A), k=3 (B), k=4 (C), k=5 (D), and k=6 (E). (F) Consensus clustering CDF for k=2–6. (G) Relative change in the area under the CDF curve for k=2–6. (H) The tracking plot for k=2–6. TCGA-LUAD, The Cancer Genome Atlas lung adenocarcinoma cohort; CDF, cumulative distribution function.

https://dx.doi.org/10.21037/atm-22-500

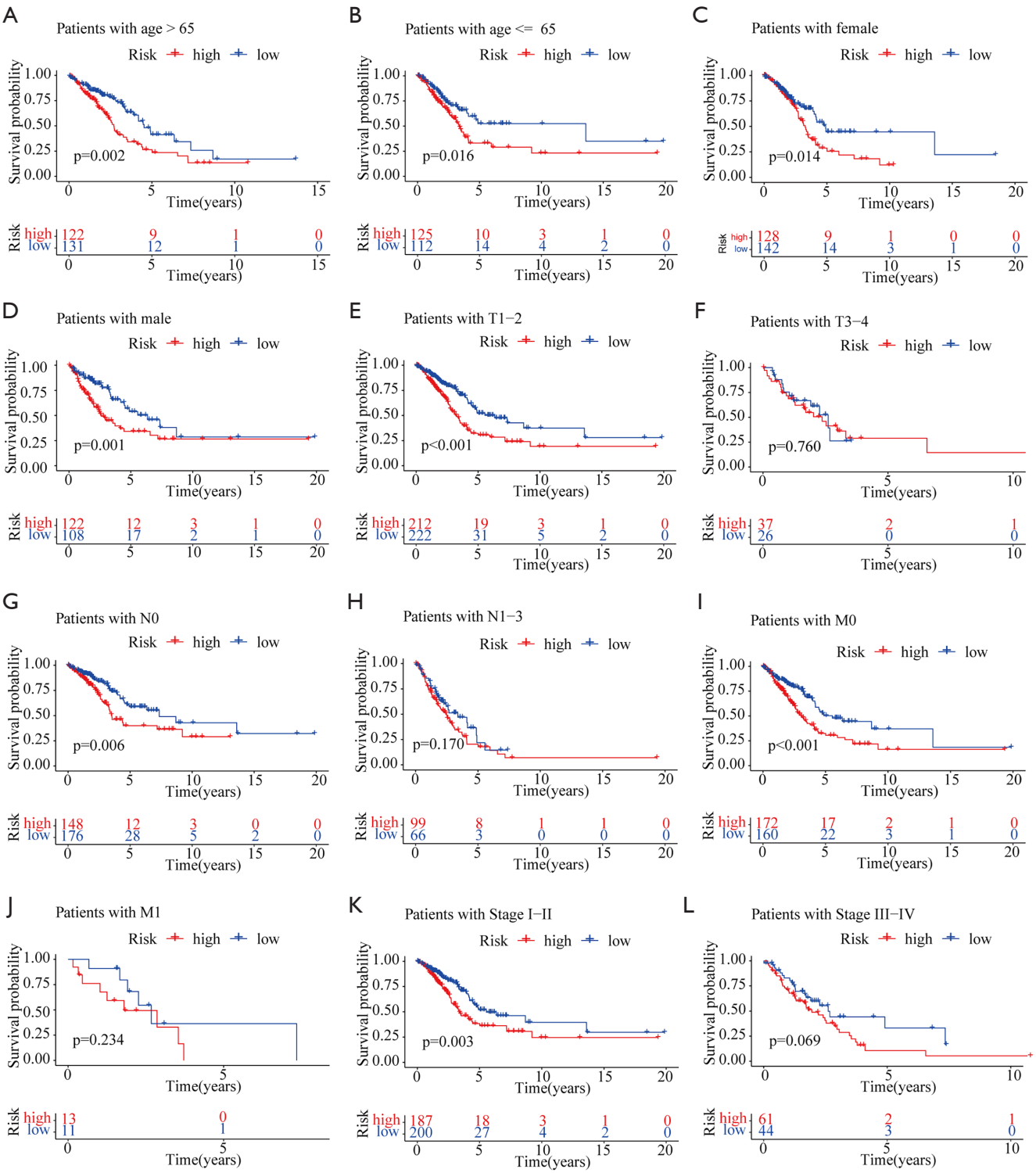

Figure S3 Stratification analysis of m5Csig. (A-L) Kaplan–Meier survival curves for subgroups stratified by different clinical characteristics. Age>65 years (A), age≤65 years (B), female (C), male (D), T1–2 (E), T3–4 (F), N0 (G), N1-3 (H), M0 (I), M1 (J), stage I–II (K) and stage III–IV (L).

https://dx.doi.org/10.21037/atm-22-500

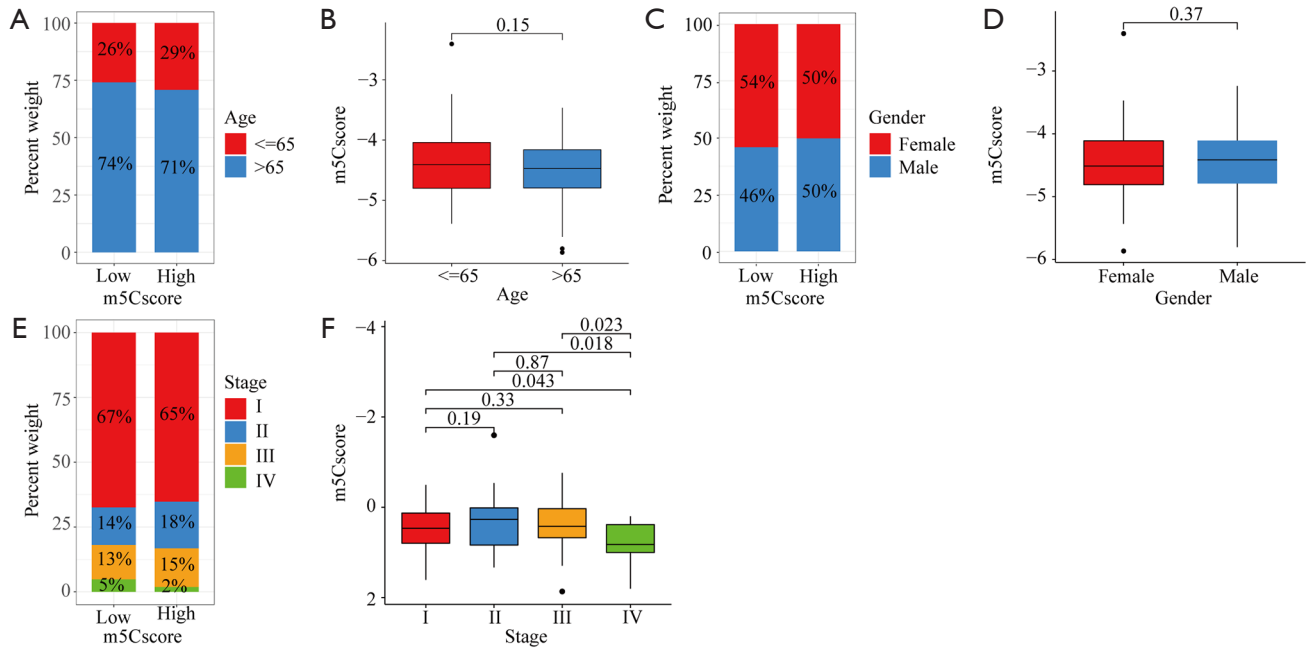

Figure S4 Clinical characteristics between high- and low-risk groups in GSE72094. (A-F) The proportion and distribution of different clinical characteristics between high- and low-risk groups in GSE72094: age 65/>65 years (A,B), female/male (C,D), stage I/II/III/IV (E,F).

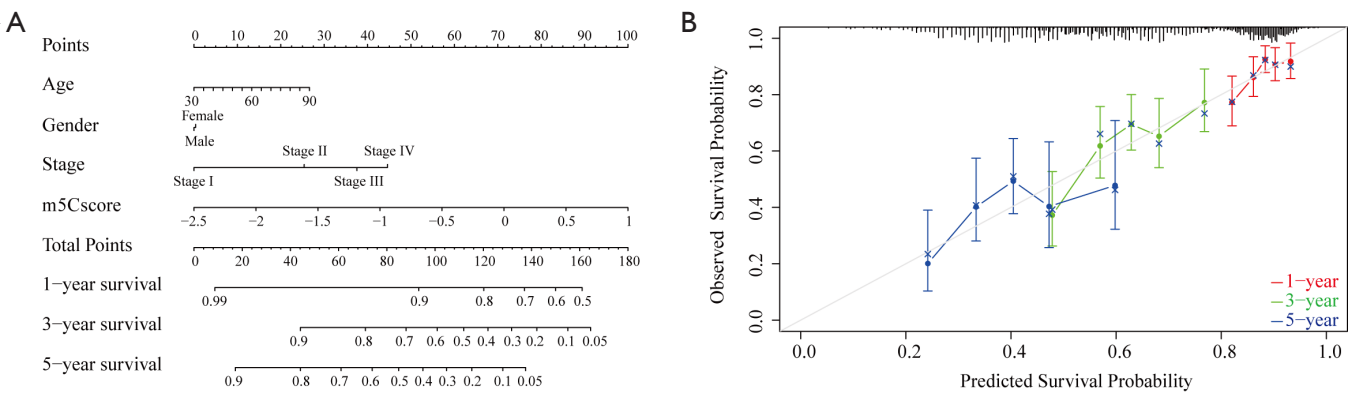

Figure S5 Building and validating the nomogram in the TCGA database. (A) Nomogram based on age, gender, stage, and m5Cscore. (B) Calibration to assess the consistency between the predicted and the actual OS at 1-, 3- and 5 years. TCGA, The Cancer Genome Atlas; OS, overall survival.

https://dx.doi.org/10.21037/atm-22-500

Table S5 Network edges of the 14 m5C regulators

From	To	Cor	P value	Color	Weight
TET2	*NSUN5*	–0.29981	8.39E-14	#6495ED	1.7988582
TET2	*ALYREF*	–0.243439	1.84E-09	#6495ED	1.4606333
TET2	*YBX1*	–0.243354	1.87E-09	#6495ED	1.4601228
TET2	*NOP2*	–0.197849	1.17E-06	#6495ED	1.187091
YBX1	*NSUN7*	–0.195919	1.50E-06	#6495ED	1.1755155
NSUN3	*NSUN5*	–0.190912	2.78E-06	#6495ED	1.1454721
TRDMT1	*NSUN7*	0.1860032	5.03E-06	Pink	1.1160192
ALKBH1	*NSUN6*	0.1871116	4.40E-06	Pink	1.1226694
ALKBH1	*NSUN7*	0.191851	2.48E-06	Pink	1.151106
ALKBH1	*ALYREF*	0.1939817	1.91E-06	Pink	1.1638905
NSUN2	*NSUN6*	0.1969656	1.31E-06	Pink	1.1817935
YBX1	*NSUN2*	0.206089	4.05E-07	Pink	1.2365338
YBX1	*NOP2*	0.2346004	7.18E-09	Pink	1.4076024
TET3	*NSUN4*	0.2355227	6.25E-09	Pink	1.4131364
NSUN4	*NSUN7*	0.246963	1.05E-09	Pink	1.4817779
TET1	*ALKBH1*	0.2509467	5.55E-10	Pink	1.5056803
NSUN3	*NSUN4*	0.2611386	1.02E-10	Pink	1.5668315
NSUN4	*NSUN6*	0.2622704	8.44E-11	Pink	1.5736222
ALKBH1	*NSUN4*	0.2668404	3.84E-11	Pink	1.6010424
TET3	*NOP2*	0.2743014	1.03E-11	Pink	1.6458082
TET3	*NSUN7*	0.2747879	9.45E-12	Pink	1.6487275
TRDMT1	*NSUN6*	0.27613	7.42E-12	Pink	1.6567799
TRDMT1	*NSUN3*	0.2772594	6.05E-12	Pink	1.6635564
TET1	*TRDMT1*	0.3114507	7.93E-15	Pink	1.8687041
TET3	*NSUN2*	0.3139996	4.66E-15	Pink	1.8839978
NSUN3	*NSUN7*	0.3242328	5.25E-16	Pink	1.9453968
TET1	*NSUN4*	0.3295863	1.62E-16	Pink	1.9775178
TET2	*NSUN4*	0.3296823	1.59E-16	Pink	1.9780936
TET1	*NSUN7*	0.3355942	4.21E-17	Pink	2.0135651
TET2	*TRDMT1*	0.357937	2.15E-19	Pink	2.147622
ALYREF	*NSUN2*	0.3661174	2.80E-20	Pink	2.1967045
NSUN6	*NSUN7*	0.3764614	1.95E-21	Pink	2.2587681
NOP2	*NSUN5*	0.3772516	1.58E-21	Pink	2.2635094
ALYREF	*NOP2*	0.3993362	3.79E-24	Pink	2.3960174
NOP2	*NSUN2*	0.4017687	1.89E-24	Pink	2.4106123
NSUN2	*NSUN5*	0.4031462	1.28E-24	Pink	2.4188773
TET2	*NSUN7*	0.404452	8.76E-25	Pink	2.426712
ALYREF	*YBX1*	0.4046408	8.29E-25	Pink	2.4278448
TET3	*NSUN6*	0.4217561	5.11E-27	Pink	2.5305365
NSUN3	*NSUN6*	0.4278305	7.81E-28	Pink	2.5669827
TET2	*NSUN6*	0.4326035	1.74E-28	Pink	2.5956209
TET3	*NSUN3*	0.435304	7.35E-29	Pink	2.6118238
TET2	*TET3*	0.4657892	2.54E-33	Pink	2.7947349
TET1	*NSUN6*	0.4789855	2.14E-35	Pink	2.8739128
ALYREF	*NSUN5*	0.4883058	6.44E-37	Pink	2.929835
TET1	*TET3*	0.5449269	3.06E-47	Pink	3.2695613
TET1	*NSUN3*	0.576112	8.05E-54	Pink	3.4566722
TET1	*TET2*	0.6617436	4.34E-76	Pink	3.9704613
TET2	*NSUN3*	0.6997493	1.58E-88	Pink	4.1984956

https://dx.doi.org/10.21037/atm-22-500

Table S6 Network nodes of the 14 m5C regulators

ID	Group	Color	Shape	Frame	P value	Size
TET1	Erasers	#E41A1C	Circle	Purple	0.3330402	8
TET2	Erasers	#E41A1C	Circle	Green	0.0678145	8
TET3	Erasers	#E41A1C	Circle	Purple	0.5008091	8
ALKBH1	Erasers	#E41A1C	Circle	Green	0.5460156	8
ALYREF	Readers	#FF7F00	Circle	Purple	0.0153939	10
YBX1	Readers	#FF7F00	Circle	Purple	0.0826999	8
TRDMT1	Writers	#999999	Circle	Green	0.0032905	12
NSUN1	Writers	#999999	Circle	Purple	0.0081617	12
NSUN2	Writers	#999999	Circle	Purple	0.1333446	8
NSUN3	Writers	#999999	Circle	Green	0.4098382	8
NSUN4	Writers	#999999	Circle	Green	0.0124132	10
NSUN5	Writers	#999999	Circle	Purple	0.4902566	8
NSUN6	Writers	#999999	Circle	Green	0.1342394	8
NSUN7	Writers	#999999	Circle	Green	0.0045582	12

https://dx.doi.org/10.21037/atm-22-500

Table S7 Significant GO terms for DEGs among the 3 clusters

Ontology	ID	Description	Gene ratio	Bg ratio	P value	P adjust	q value	Gene ID	Count
BP	GO:0002220	Innate immune response activating cell surface receptor signaling pathway	10/309	116/18,670	2.31E-05	0.000253	0.000203	*PSMB2/PSMA5/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	10
BP	GO:0002220	Innate immune response activating cell surface receptor signaling pathway	10/309	116/18,670	2.31E-05	0.000253	0.000203	*PSMB2/PSMA5/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	10
BP	GO:0002218	Activation of innate immune response	10/309	142/18,670	0.000129	0.001176	0.000941	*PSMB2/PSMA5/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	10
BP	GO:0045089	Positive regulation of innate immune response	10/309	214/18,670	0.003088	0.019615	0.015698	*PSMB2/PSMA5/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	10
BP	GO:0071356	Cellular response to tumor necrosis factor	11/309	291/18,670	0.009446	0.053699	0.042975	*PSMB2/PSMA5/TRAIP/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	11
BP	GO:0033209	Tumor necrosis factor-mediated signaling pathway	11/309	167/18,670	0.000109	0.001016	0.000813	*PSMB2/PSMA5/TRAIP/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	11
BP	GO:0050658	RNA transport	10/309	193/18,670	0.001452	0.010146	0.00812	*ALYREF/RAN/NUP37/EIF5A/EIF4A3/CPSF3/SLBP/NUTF2/MAGOH/NUP93*	10
BP	GO:0051236	Establishment of RNA localization	10/309	196/18,670	0.001628	0.011174	0.008943	*ALYREF/RAN/NUP37/EIF5A/EIF4A3/CPSF3/SLBP/NUTF2/MAGOH/NUP93*	10
BP	GO:0008380	RNA splicing	26/309	469/18,670	8.22E-08	1.89E-06	1.52E-06	*YBX1/ALYREF/PPIH/SNRNP40/SNRPC/DNAJC8/EIF4A3/PHF5A/SF3A3/CELF2/SF3B5/CTNNBL1/PPP1R8/POLR2G/CPSF3/ZPR1/CIRBP/LSM10/PPIL1/STRAP/PRPF19/LSM2/MAGOH/SNRPF/SNRPD1/SNRPB*	26
BP	GO:0000398	Mrna splicing, via spliceosome	23/309	379/18,670	9.47E-08	2.09E-06	1.67E-06	*YBX1/ALYREF/PPIH/SNRNP40/SNRPC/DNAJC8/EIF4A3/PHF5A/SF3A3/CELF2/SF3B5/CTNNBL1/POLR2G/CPSF3/CIRBP/PPIL1/STRAP/PRPF19/LSM2/MAGOH/SNRPF/SNRPD1/SNRPB*	23
BP	GO:0006402	Mrna catabolic process	20/309	364/18,670	3.07E-06	4.07E-05	3.26E-05	*YBX1/PSMB2/MRTO4/PSMA5/EIF4A3/PSMD14/PSMB1/PSMC3/PSMD2/POLR2G/PSMB7/CIRBP/PSMA7/PSMA4/PSMD9/LSM2/SERBP1/RPL18A/UBA52/MAGOH*	20
BP	GO:0051028	Mrna transport	9/309	152/18,670	0.000985	0.007171	0.005739	*ALYREF/NUP37/EIF5A/EIF4A3/CPSF3/SLBP/NUTF2/MAGOH/NUP93*	9
CC	GO:0034709	Methylosome	4/313	12/19,717	2.79E-05	0.000242	0.000161	*PRMT1/SNRPF/SNRPD1/SNRPB*	4
CC	GO:0034708	Methyltransferase complex	7/313	113/19,717	0.002186	0.009912	0.006616	*PHF19/RUVBL2/PRMT1/EZH1/SNRPF/SNRPD1/SNRPB*	7
CC	GO:0005689	U12-type spliceosomal complex	6/313	27/19,717	3.41E-06	3.89E-05	2.59E-05	*YBX1/PHF5A/SF3B5/SNRPF/SNRPD1/SNRPB*	6
CC	GO:0097525	Spliceosomal snrnp complex	10/313	99/19,717	3.94E-06	4.29E-05	2.86E-05	*PPIH/SNRNP40/SNRPC/PHF5A/SF3A3/SF3B5/LSM2/SNRPF/SNRPD1/SNRPB*	10
CC	GO:0005732	Small nucleolar ribonucleoprotein complex	5/313	28/19,717	7.11E-05	0.000547	0.000365	*SNRNP40/POP7/RRP9/FBL/SNRPF*	5
CC	GO:0097526	Spliceosomal tri-snrnp complex	5/313	42/19,717	0.000513	0.003022	0.002017	*PPIH/LSM2/SNRPF/SNRPD1/SNRPB*	5
CC	GO:0022624	Proteasome accessory complex	4/313	24/19,717	0.000515	0.003022	0.002017	*PSMD14/PSMC3/PSMD2/PSMD9*	4
CC	GO:1902911	Protein kinase complex	7/313	109/19,717	0.001779	0.008554	0.00571	*CCNB1/CDK1/CCNA2/CCNB2/CDK2/CKS1B/PCNA*	7
CC	GO:0032040	Small-subunit processome	4/313	38/19,717	0.003006	0.012624	0.008426	*UTP18/UTP11/RRP9/FBL*	4
CC	GO:0005759	Mitochondrial matrix	16/313	469/19,717	0.003551	0.014758	0.009851	*MRPL11/MRPL37/CCNB1/RAD51/NUDT1/CDK1/MRPS7/DTYMK/MRPS15/MRPS16/MRPL51/MRPL12/MRPL15/PARK7/MTHFD2/HSD17B10*	16
CC	GO:0032153	Cell division site	4/313	68/19,717	0.022932	0.071185	0.047513	*CEP55/RACGAP1/KIF20A/ANLN*	4
MF	GO:0140097	Catalytic activity, acting on DNA	22/309	213/17,696	2.42E-11	1.13E-08	9.70E-09	*RAD51/RFC2/MCM6/NME1/MCM2/FEN1/MCM4/CDC45/DCLRE1B/POLE3/GINS1/RAD54L/RUVBL2/UNG/EME1/HMGA1/TOP2A/PCNA/MCM7/RFC4/GINS2/PTGES3*	22
MF	GO:0003688	DNA replication origin binding	9/309	24/17,696	1.40E-10	3.25E-08	2.80E-08	*MCM6/MCM2/ORC1/MCM4/CDC6/CDC45/ORC6/MCM7/MCM10*	9
MF	GO:0016887	Atpase activity	27/309	434/17,696	1.24E-08	1.53E-06	1.32E-06	*KIF2C/RAD51/RFC2/MCM6/KIF4A/MCM2/MCM4/ABCC6/CDC45/ATP6V1F/EIF4A3/OLA1/KIF18B/KIF20A/KIF23/GINS1/RAD54L/PSMC3/RUVBL2/KIFC1/TOP2A/KIF11/MCM7/RFC4/GINS2/GET3/DDX49*	27
MF	GO:0008017	Microtubule binding	20/309	246/17,696	1.31E-08	1.53E-06	1.32E-06	*BIRC5/MTUS1/KIF2C/GAPDH/PLK1/KIF4A/FAM83D/DRG1/RACGAP1/KIF18B/KIF20A/KIF23/PRC1/KIFC1/PSRC1/NUSAP1/RCC2/KIF11/DPYSL2/SKA1*	20
MF	GO:0015631	Tubulin binding	23/309	336/17,696	2.73E-08	2.54E-06	2.19E-06	*BIRC5/MTUS1/KIF2C/GAPDH/PLK1/KIF4A/NME1/FAM83D/DRG1/RACGAP1/KIF18B/KIF20A/KIF23/CCT5/PRC1/STMN1/KIFC1/PSRC1/NUSAP1/RCC2/KIF11/DPYSL2/SKA1*	23
MF	GO:0003697	Single-stranded DNA binding	13/309	113/17,696	8.77E-08	5.84E-06	5.02E-06	*YBX1/RAD51/MCM6/NABP2/NME1/MCM2/MCM4/CDC45/RAD51AP1/PRIM1/POLR2G/MCM7/MCM10*	13
MF	GO:0008094	DNA-dependent atpase activity	13/309	113/17,696	8.77E-08	5.84E-06	5.02E-06	*RAD51/RFC2/MCM6/MCM2/MCM4/CDC45/GINS1/RAD54L/RUVBL2/TOP2A/MCM7/RFC4/GINS2*	13
MF	GO:0003678	DNA helicase activity	11/309	81/17,696	1.58E-07	9.18E-06	7.90E-06	*RAD51/RFC2/MCM6/MCM2/MCM4/CDC45/GINS1/RUVBL2/MCM7/RFC4/GINS2*	11
MF	GO:0070182	DNA polymerase binding	6/309	19/17,696	6.05E-07	3.13E-05	2.70E-05	*RAD51/NABP2/CDT1/PCNA/PTGES3/FANCI*	6
MF	GO:0017116	Single-stranded DNA helicase activity	6/309	20/17,696	8.52E-07	3.61E-05	3.11E-05	*RAD51/RFC2/MCM6/MCM2/MCM7/RFC4*	6
MF	GO:0043138	3'-5' DNA helicase activity	6/309	20/17,696	8.52E-07	3.61E-05	3.11E-05	*MCM6/MCM2/CDC45/GINS1/MCM7/GINS2*	6
MF	GO:0004386	Helicase activity	14/309	163/17,696	1.07E-06	3.91E-05	3.37E-05	*RAD51/RFC2/MCM6/MCM2/MCM4/CDC45/EIF4A3/GINS1/RAD54L/RUVBL2/MCM7/RFC4/GINS2/DDX49*	14
MF	GO:0004298	Threonine-type endopeptidase activity	6/309	21/17,696	1.17E-06	3.91E-05	3.37E-05	*PSMB2/PSMA5/PSMB1/PSMB7/PSMA7/PSMA4*	6
MF	GO:0070003	Threonine-type peptidase activity	6/309	21/17,696	1.17E-06	3.91E-05	3.37E-05	*PSMB2/PSMA5/PSMB1/PSMB7/PSMA7/PSMA4*	6
MF	GO:0035173	Histone kinase activity	5/309	17/17,696	8.19E-06	0.000254	0.000219	*CCNB1/CDK1/CDK2/CHEK1/AURKB*	5
MF	GO:0051082	Unfolded protein binding	11/309	131/17,696	1.90E-05	0.000554	0.000477	*PPIH/CCT7/CCT4/TCP1/CCT5/RUVBL2/PPIA/PTGES3/CHAF1A/CCT8/CHAF1B*	11
MF	GO:0043021	Ribonucleoprotein complex binding	10/309	133/17,696	0.000116	0.003173	0.002731	*PPIH/NME1/SNRPC/EIF5A/EIF4A3/OLA1/UNG/PES1/SNRPD1/SNRPB*	10
MF	GO:0003777	Microtubule motor activity	7/309	84/17,696	0.000664	0.017188	0.014792	*KIF2C/KIF4A/KIF18B/KIF20A/KIF23/KIFC1/KIF11*	7
MF	GO:0042393	Histone binding	11/309	197/17,696	0.000711	0.01744	0.015009	*MCM2/PHF19/H2AX/NCAPD2/ASF1B/CKS1B/UHRF1/EZH1/HAT1/NPM3/CHAF1B*	11
MF	GO:0003735	Structural constituent of ribosome	11/309	202/17,696	0.000874	0.020364	0.017526	*MRPL11/MRPL37/MRPS7/MRPS15/MRPS16/MRPL51/MRPL12/MRPL15/RPL39L/RPL18A/UBA52*	11
MF	GO:0003684	Damaged DNA binding	6/309	65/17,696	0.000939	0.020841	0.017936	*FEN1/DCLRE1B/AUNIP/H2AX/UNG/PCNA*	6
MF	GO:0140142	Nucleocytoplasmic carrier activity	4/309	31/17,696	0.001978	0.041893	0.036054	*RAN/KPNA2/CSE1L/NUTF2*	4
MF	GO:0003774	Motor activity	8/309	136/17,696	0.002698	0.054326	0.046754	*KIF2C/KIF4A/KIF18B/KIF20A/KIF23/KIFC1/KIF11/MYL6B*	8
MF	GO:0016891	Endoribonuclease activity, producing 5'-phosphomonoesters	4/309	34/17,696	0.002798	0.054326	0.046754	*FEN1/RNASEH2A/POP7/RPP40*	4
MF	GO:0008409	5'-3' exonuclease activity	3/309	17/17,696	0.00299	0.055732	0.047965	*FEN1/DCLRE1B/CPSF3*	3

GO, Gene Ontology; DEGs, differentially expressed genes.

https://dx.doi.org/10.21037/atm-22-500

Table S8 Significant KEGG terms for DEGs among the three clusters

ID	Description	Gene ratio	Bg ratio	P value	P adjust	q value	Gene ID	Count
hsa04110	Cell cycle	27/162	124/8,108	4.06E-21	6.33E-19	5.30E-19	*CDC20/CCNB1/PLK1/MCM6/CDK1/MCM2/ORC1/CCNA2/CCNB2/MCM4/PKMYT1/CDC6/CDC45/CDK2/MAD2L1/CDC25C/CHEK1/MAD2L2/CDC25A/BUB1/PCNA/ORC6/MCM7/ESPL1/YWHAQ/TTK/BUB1B*	27
hsa03030	DNA replication	11/162	36/8,108	5.63E-11	4.39E-09	3.67E-09	*RFC2/MCM6/MCM2/FEN1/MCM4/POLE3/RNASEH2A/PRIM1/PCNA/MCM7/RFC4*	11
hsa05012	Parkinson disease	23/162	249/8,108	6.51E-10	3.39E-08	2.83E-08	*TUBA1C/TUBA1B/TUBB/PSMB2/PSMA5/SDHB/PSMD14/PSMB1/UQCRH/PSMC3/PSMD2/TUBB6/ATP5MC3/PSMB7/PARK7/PSMA7/PSMA4/CYC1/PSMD9/NDUFS6/UBA52/NDUFA12/COX8A*	23
hsa05014	Amyotrophic lateral sclerosis	27/162	364/8,108	2.54E-09	9.92E-08	8.30E-08	*TUBA1C/TUBA1B/ALYREF/TUBB/PSMB2/PFN1/PSMA5/SDHB/NUP37/BID/PSMD14/PSMB1/UQCRH/PSMC3/PSMD2/TUBB6/ATP5MC3/PSMB7/PSMA7/PSMA4/CYC1/PSMD9/GPX7/NDUFS6/NDUFA12/COX8A/NUP93*	27
hsa05016	Huntington disease	24/162	306/8,108	7.31E-09	2.28E-07	1.91E-07	*TUBA1C/TUBA1B/TUBB/PSMB2/PSMA5/AP2S1/SDHB/PSMD14/PSMB1/UQCRH/PSMC3/PSMD2/TUBB6/ATP5MC3/POLR2G/PSMB7/PSMA7/PSMA4/CYC1/PSMD9/GPX7/NDUFS6/NDUFA12/COX8A*	24
hsa03050	Proteasome	10/162	46/8,108	1.69E-08	4.38E-07	3.67E-07	*PSMB2/PSMA5/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	10
hsa03040	Spliceosome	16/162	147/8,108	3.12E-08	6.96E-07	5.82E-07	*ALYREF/PPIH/SNRNP40/SNRPC/EIF4A3/PHF5A/SF3A3/SF3B5/CTNNBL1/PPIL1/PRPF19/LSM2/MAGOH/SNRPF/SNRPD1/SNRPB*	16
hsa05020	Prion disease	21/162	273/8,108	9.75E-08	1.90E-06	1.59E-06	*TUBA1C/TUBA1B/TUBB/PSMB2/PSMA5/SDHB/PSMD14/PSMB1/UQCRH/PSMC3/PSMD2/TUBB6/ATP5MC3/PSMB7/PSMA7/PSMA4/CYC1/PSMD9/NDUFS6/NDUFA12/COX8A*	21
hsa04114	Oocyte meiosis	14/162	129/8,108	2.46E-07	4.11E-06	3.44E-06	*CDC20/CCNB1/PLK1/CDK1/CCNB2/PKMYT1/CDK2/MAD2L1/CDC25C/MAD2L2/BUB1/ESPL1/YWHAQ/FBXO5*	14
hsa05010	Alzheimer disease	24/162	369/8,108	2.64E-07	4.11E-06	3.44E-06	*TUBA1C/TUBA1B/TUBB/PSMB2/GAPDH/PSMA5/SDHB/BID/PSMD14/PSMB1/UQCRH/PSMC3/PSMD2/TUBB6/ATP5MC3/PSMB7/PSMA7/PSMA4/CYC1/PSMD9/NDUFS6/NDUFA12/COX8A/HSD17B10*	24
hsa04914	Progesterone-mediated oocyte maturation	12/162	100/8,108	6.19E-07	8.77E-06	7.34E-06	*CCNB1/PLK1/CDK1/CCNA2/CCNB2/PKMYT1/CDK2/MAD2L1/CDC25C/MAD2L2/CDC25A/BUB1*	12
hsa05022	Pathways of neurodegeneration - multiple diseases	26/162	475/8,108	2.21E-06	2.88E-05	2.41E-05	*TUBA1C/TUBA1B/TUBB/PSMB2/PSMA5/SDHB/BID/PSMD14/PSMB1/UQCRH/PSMC3/PSMD2/TUBB6/ATP5MC3/PSMB7/PARK7/PSMA7/PSMA4/CYC1/PSMD9/GPX7/NDUFS6/UBA52/NDUFA12/COX8A/HSD17B10*	26
hsa04115	p53 signaling pathway	8/162	73/8,108	9.53E-05	0.001144	0.000957	*CCNB1/RRM2/CDK1/CCNB2/CDK2/GTSE1/BID/CHEK1*	8
hsa03013	RNA transport	12/162	186/8,108	0.000333	0.003716	0.003109	*EIF3I/ALYREF/RAN/TACC3/NUP37/EIF4A3/POP7/RPP40/STRAP/EIF2B3/MAGOH/NUP93*	12
hsa05017	Spinocerebellar ataxia	10/162	143/8,108	0.00056	0.005828	0.004876	*PSMB2/PSMA5/PSMD14/PSMB1/PSMC3/PSMD2/PSMB7/PSMA7/PSMA4/PSMD9*	10
hsa00190	Oxidative phosphorylation	9/162	134/8,108	0.001407	0.013715	0.011475	*SDHB/ATP6V1F/UQCRH/ATP6V0B/ATP5MC3/CYC1/NDUFS6/NDUFA12/COX8A*	9
hsa04218	Cellular senescence	9/162	156/8,108	0.003963	0.034669	0.029008	*CCNB1/CDK1/MYBL2/CCNA2/CCNB2/CDK2/CHEK1/CDC25A/FOXM1*	9
hsa03410	Base excision repair	4/162	33/8,108	0.004	0.034669	0.029008	*FEN1/POLE3/UNG/PCNA*	4
hsa00240	Pyrimidine metabolism	5/162	56/8,108	0.00503	0.041298	0.034554	*TK1/RRM2/NME1/DTYMK/NME2*	5
hsa00480	Glutathione metabolism	5/162	57/8,108	0.005426	0.042326	0.035414	*RRM2/SMS/SRM/GSTO1/GPX7*	5

KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes.

https://dx.doi.org/10.21037/atm-22-500