Original Article

M^5C regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in lung adenocarcinoma

Hui Chen[1#], Xiao-Lin Ge[1#], Zhao-Yue Zhang[1#], Ming Liu[2,3,4], Rui-Yan Wu[2,3,4], Xiao-Fei Zhang[2,3,4], Li-Ping Xu[1], Hong-Yan Cheng[5], Xin-Chen Sun[1], Hong-Cheng Zhu[1,2,3,4]

[1]Department of Radiation Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; [2]Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Shanghai, China; [3]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China; [4]Shanghai Key Laboratory of Radiation Oncology, Shanghai, China; [5]Department of Synthetic Internal Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Contributions: (I) Conception and design: H Chen, HC Zhu; (II) Administrative support: HC Zhu, HY Cheng, XC Sun; (III) Provision of study materials or patients: XL G, ZY Zhang; (IV) Collection and assembly of data: H Chen, ZY Zhang, M Liu, RY Wu, XF Zhang; (V) Data analysis and interpretation: XL Ge, HC Zhu, LP Xu, M Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Xin-Chen Sun. Department of Radiation Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 300 Guangzhou Road, Nanjing 210029, China. Email: sunxinchen@njmu.edu.cn; Hong-Cheng Zhu. Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Shanghai 200032, China. Email: zhuhc90@163.com.

Background: In recent years, immunotherapy has made great progress, and the regulatory role of epigenetics has been verified. However, the role of 5-methylcytosine (m^5C) in the tumor microenvironment (TME) and immunotherapy response remains unclear.

Methods: Based on 11 m^5C regulators, we evaluated the m^5C modification patterns of 572 lung adenocarcinoma (LUAD) patients. The m^5C score was constructed by principal component analysis (PCA) algorithms in order to quantify the m^5C modification pattern of individual LUAD patients.

Results: Two m^5C methylation modification patterns were identified according to 11 m^5C regulators. The two patterns had a remarkably distinct TME immune cell infiltration characterization. Next, 226 differentially expressed genes (DEGs) related to the m^5C phenotype were screened. Patients were divided into three different gene cluster subtypes based on these genes, which had different TME immune cell infiltration and prognosis characteristics. The m^5C score was constructed to quantify the m^5C modification pattern of individual LUAD patients. We found that the high m^5C score group had a better prognosis. The role of the m^5C score in predicting prognosis was also verified in the dataset GSE31210.

Conclusions: Our study revealed that m^5C modification played a significant role in TME regulation of LUAD. Investigation of the m^5C regulation mode may have some implications for tumor immunotherapy in the future.

Keywords: M^5C; tumor microenvironment (TME); immunotherapy; lung adenocarcinoma (LUAD)

Submitted Mar 16, 2021. Accepted for publication May 21, 2021.
doi: 10.21037/tlcr-21-351
View this article at: http://dx.doi.org/10.21037/tlcr-21-351

Introduction

According to the ribonucleic acid (RNA) modification database (MODOMICS), over 150 RNA modifications have been detected, including 5-methylcytosine (m^5C), N6-methyladenosine (m^6A), and N1-methyladenosine (m^1A) (1,2). 5-methylcytosine (m^5C) is one common methylation

© Translational Lung Cancer Research. All rights reserved.

modification, and plays significant roles in various biological process. M^5C modification is a kind of post-transcriptional modification regulated by "writers", "erasers", and "readers", which are methyltransferases, demethylases, and binding proteins, respectively.

Methylation of the cytosine at the fifth carbon position (m^5C) is mediated by methyltransferases consisting of NOL1/NOP2/Sun domain family, member 1-7 (NSUN1-7), DNA methyltransferase1 (DNMT1), DNMT2, DNMT3A, and DNMT3B, while the removal process is catalyzed by demethylases such as ten-eleven translocation 2 (TET2) . In addition, a group of specific RNA-binding proteins can read the m^5C motif, thereby mediating its function. It has been found that m^5C modification in messenger RNA (mRNA) is primarily enriched in the non-translated region (3'UTR and 5'UTR), guanine-cytosine (GC)-rich regions, and near the argonaute (AGO) protein binding site, which has a conserved sequence of AU(m^5C)GANGU (3-6).

Immunotherapy has been an effective treatment against cancer, and is represented by immunological checkpoint blockades (ICBs). However, the overall response rates are still unsatisfying, especially for cancers with low mutational burdens (7). In recent years, with the development of immunotherapy, the therapeutic options for cancer treatment have undergone significant changes (8-11). Among the various immunotherapies, ICBs work by blocking the interaction between immunosuppressive receptors (immune checkpoints) expressed on the surface of immunocytes and their ligands. ICBs include a series of monoclonal antibody-based therapies. Cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), programmed death 1 (PD-1) and programmed death-ligand 1 (PD-L1) are the main targets of immunotherapy (12,13). ICB has attracted widespread attention due to its persistence in reaction and its impact on the overall survival of patients. However, the challenge for clinicians is to determine who will respond to immunotherapy. The number of patients who actually benefit from immunotherapy remains small (14-16).

The complexity and strong interrelationship of the tumor microenvironment (TME), which comprises immune cells (such as macrophages, mast cells, polymorphonuclear cells, dendritic cells, natural killer (NK) cells, as well as T and B lymphocytes) and non-immune cells (including endothelial cells and stromal cells), play a key role in its development and progression (17,18). The immune cell components of the tumor are the basis for determining the fate of the tumor, as well as its invasion and metastasis. Interacting with other TME components either directly or indirectly can lead to a variety of biological behavioral changes in tumor cells, including proliferation and angiogenesis, apoptosis, hypoxia, and immune tolerance (19). An increasing number of studies have shown that the TME has a crucial impact on tumor progression, immunotherapy response, and immune escape (20,21). Recently, one research has also showed that under a scenario of balanced autophagy in the tumor microenvironment, the infiltrating immune cells control cytokine production and secretion (22). Sacco *et al.* indicated that tumor-infiltrating immune cells could affect the tumor immunosurveillance by regulating the iron metabolism (23). Therefore, the characteristics of tumor immune infiltration can provide new strategies for the prediction of immunotherapeutic effect, the improvement of immunotherapeutic response rate, and the development of novel immunotherapeutic targets.

Recently, several studies have shown a strong association between m^5C modification and TME infiltrating immune cells. Schoeler *et al.* reported that TET enzymes control antibody production and shape the mutational landscape in germinal center B cells. They found that TET2 and TET3 guide the transition of germinal center B cells to antibody-secreting plasma cells (24). Also, Li *et al.* revealed that the TET family modulates the activation of dendritic cells. TET1-inhibited monocyte-derived dendritic cells were found to significantly decrease the percentage of $CD45RA^-FoxP3^{hi}$-activated regulatory T (Treg) cells in the allergic rhinitis group, which might be linked to immune activation (25). Yue *et al.* found that TET2/3-deficiency in Treg cells leads to T cell activation, TET2/3 double-knockout (DKO) Treg cells exhibited a dysregulated cell surface phenotype, and TET2/3 DKO $CD4^+$ T cells induced disease in healthy mice (26). Moreover, some researches have focused on the intrinsic pathways of cancer cells, such as genomic variation and the disordered expression of m^5C regulators. Chen *et al.* indicated that numerous oncogene RNAs with hypermethylated m^5C sites were causally linked to their upregulation in urothelial carcinoma of the bladder, and demonstrated that Y-box binding protein 1 (Ybx1) is an m^5C "reader" (27). Lastly, USUN5 expression has been verified to be related to shorter survival in glioblastoma and the high expression of USUN7 is correlated to the poor survival in low-grade gliomas (28,29).

However, the above studies only mentioned the role and mechanism of one or two regulatory factors of m^5C in antitumor and immune processes, while the potential cross-talk between regulators remains uncharacterized in human cancers. Therefore, establishing a comprehensive

© Translational Lung Cancer Research. All rights reserved.

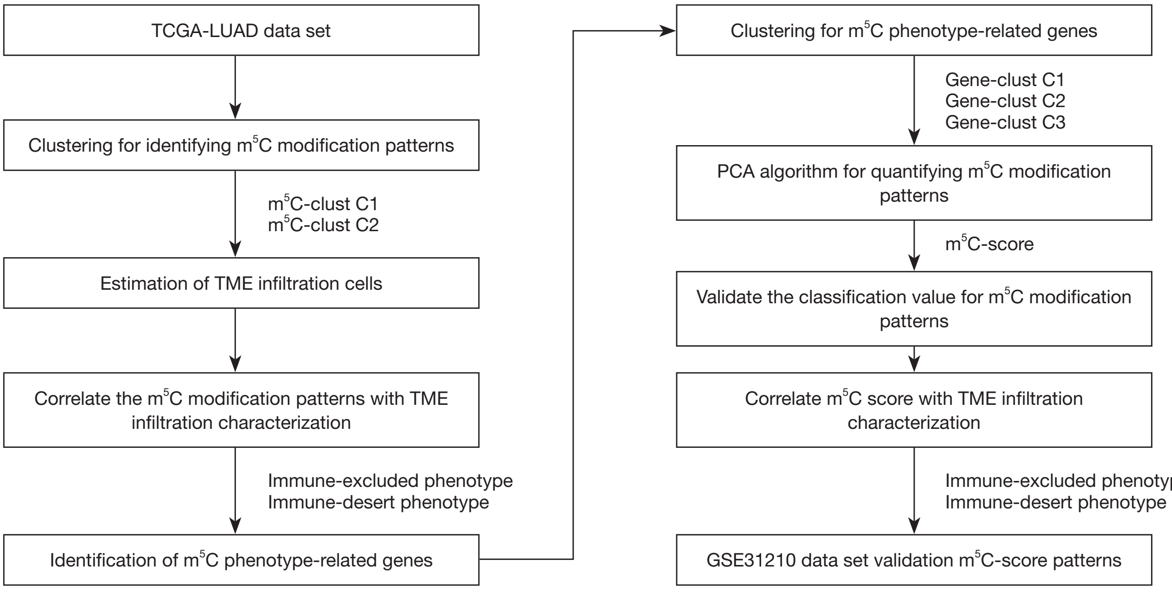


Figure 1 Flowchart of bioinformatics analysis in our study. TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; TME, tumor microenvironment.

understanding of the TME cell infiltration characterization mediated by m^5C regulators will offer insight into TME immune regulation. In this study, we analyzed the gene mutation of The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD), the mRNA expression data, and the clinical information of patients. We also investigated the mechanisms through which m^5C affected the prognosis of patients during the occurrence of LUAD, and further verified the results in an external dataset (GSE31210). We present the following article in accordance with the MDAR reporting checklist (available at http://dx.doi.org/10.21037/tlcr-21-351).

Methods

Dataset source and preprocessing

We conducted a systematic search of TCGA and the Gene-Expression Omnibus (GEO) databases for LUAD. Standardized matrix files of each cohort were downloaded for further analysis. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The procedure of data preprocessing lists was as follows: (I) we downloaded data of TCGA LUAD single nucleotide variation (SNV) (MuTect2 Annotation), which included 570 samples; (II) we downloaded data of TCGA LUAD copy number variation (CNV), which included 544 samples. We re-annotated the CNV of 13 genes using Bedtools software using hg38 as a reference (30); (III) we downloaded data of TCGA LUAD FPKM (Fragments Per Kilobase Million), which included 572 samples (513 tumor samples and 59 normal samples) and 513 follow-up data; and (IV) we downloaded 246 samples of the expression profile and follow-up data of GSE31210 from National Coalition Building Institute (NCBI) GEO, which included 226 tumor samples and 20 normal samples. The study roadmap is shown in *Figure 1*.

Unsupervised clustering for 11 m^5C regulators

We extracted 11 regulators related to m^5C that had expression in TCGA datasets for LUAD analysis using the prCOMP function in R language (13 genes related to m^5C modification were detected, but only 11 had expression). These 13 regulators comprised 11 writers (*NSUN2*, *NSUN3*, *NSUN4*, *NSUN5*, *NSUN6*, *NSUN7*, *DNMT1*, *DNMT2*, *DNMT3B*, *NSUN1*, *DNMT3A*), one eraser (TET2), and one reader (Aly/REF export factor, ALYREF). In order to identify different m^5C modification patterns and classify patients for further study, unsupervised clustering analysis was applied. The 11 m^5C regulators were clustered with LUAD tumor samples by non-negative matrix factorization (NMF). The NMF method selected the standard "Brunet" and carried out 100 iterations. The number of clusters was set from 2 to 10, and we determined

© Translational Lung Cancer Research. All rights reserved.

the average contour width of the common member matrix using the R package "NMF", setting the minimum members of each subclass to 10. We selected the optimal clustering number as 2 based on the cophenetic, dispersion, and silhouette.

Gene set variation analysis (GSVA) and functional annotation

In order to explore the biological behavior between these different m^5C modification patterns, GSVA enrichment analysis was carried out using the R language GSVA package (31), and the "c2.cp.kegg.v7.0.symbols.gmt" gene set was used as the background. GSVA, in a non-parametric and unsupervised method, is commonly employed for estimating the variations in pathway and biological process activity in the samples of an expression dataset. Differential pathways were screened by $|t| > 6$ using the R package limma.

Estimation of TME cell infiltration

The cell type identification by estimating relative subsets of RNA transcripts (CIBERSORT) method was used to analyze the composition and relative abundance of m^5C-modified immune cells of the two patterns. Since T cells. CD4.memory.activated was 0 in all samples, we removed the cells and calculated the correlation and significance of 11 m^5C-related genes and TME infiltration types through the rcorr function of the R language Hmisc package. We also used the ESTIMATE algorithm to quantify the immune, matrix, and ESTIMATE scores between groups of high and low expression regulators.

Identification of differentially expressed genes (DEGs) between m^5C distinct phenotypes

Previously, two m^5C modification patterns were classified by clustering m^5C-related genes. In the next step, we carried out principal component analysis (PCA) of these two subtypes, and the two patterns were separated from each other. Using the R package limma package for difference analysis, 226 differential genes were screened by $|\log2\text{fold change}| > 1$, false discovery rate (FDR) <0.05. The patients were divided into different gene clusters by unsupervised clustering of 226 m^5C phenotype-related genes (the distance between samples was calculated by complete and Euclidean).

Generation of the m^5C gene signature

Due to the heterogeneity and complexity of m^5C modification, we constructed a scoring system to quantify the m^5C modification pattern of individual LUAD patients based on these phenotypic genes, which was called the m^5C score. We then performed a prognostic analysis on each gene in the signature using a univariate Cox regression model. We screened 124 genes related to prognosis with $P<0.05$ from 226 DEGs, and subsequently analyzed the 124 genes by PCA, scored PC1 and PC2, and calculated the m^5C score of each sample. The formula was as follows:

$$m^5C \text{ score} = \Sigma\,(PC1i + PC2i)$$

where i is the expression of 125 m^5C phenotype-related genes.

Statistical analysis

Spearman and distance correlation analyses were utilized to compute correlation coefficients between the expression of m^5C regulators and TME infiltrating immune cells. To analyze difference between two groups, the Wilcoxon test was used, and in cases of three or more groups, difference comparisons were conducted using Kruskal-Wallis tests and one-way ANOVA (analysis of variance). For verification of the external dataset GSE31210, m^5C score model samples were divided into high and low score subgroups according to the median. Using the survminer R package, survival curves were generated using log-rank tests and the Kaplan-Meier (KM) method. Statistical significance was set at $P<0.05$, and all statistical P values were two-sided. All data was processed using R 3.6.1 software.

Results

Genetic variation of m^5C regulators in LUAD

Thirteen m^5C regulators were identified in this study, including 11 writers, one eraser, and one reader. We first summarized the incidence of SNV and CNV in the 13 m^5C regulators in LUAD. Figure S1 shows the dynamic and reversible regulation of m^5C RNA methylation.

SNV analysis of m^5C related genes

Of the 570 LUAD patients, gene mutations of the 13 m^5C regulators appeared in 99 independent samples, with a frequency of 15.75%. The writer, DNMT3A, exhibited the highest incidence of mutation, followed by NSUN2,

© Translational Lung Cancer Research. All rights reserved.

TET2, and DNMT3B, while “reader” genes had fewer mutations than “writer” and “eraser” genes. *Figure 2A* displays the mutations in the top 10 genes associated with m^5C, including variant classification, type, and variants per sample.

CNV analysis of m^5C related genes

In addition to SNV, CNVs are also present as genetic variations, including amplification (Segment_Mean >0.2), diploid (−0.2< Segment_Mean <0.2), and deletion (Segment_Mean<−0.2). *Table 1* shows the proportion of amplification and deletion of the 11 genes. We examined the incidence of CNV and the mRNA expression of these regulators to explore the relationship between gene variations and the expression levels of m^5C regulators (*Figure 2B*), and found that CNV could be the key factor leading to the disordered expression of m^5C regulators. The expression of m^5C regulators in LUAD tissue was significantly higher than that in normal lung tissue (except NSUN3 and TET2) (Figure S2).

In total, nine CNV gene mutations had quantitative values in the gene expression profile. We observed that genes that experienced amplification showed higher expression, while those that experienced deletion exhibited lower expression. NSUN2, DNMT3B, ALYREF, and NSUN5 had a high frequency of CNV amplification, while DNMT1 and TET2 exhibited a high frequency of CNV deletion. These gene mutations may affect the transmission of the m^5C signal in cells and result in cellular functional disorder. Among them, NSUN2, DNMT3b, NSUN5 and DNMT1 are writers, ALYREF is a reader, and TET2 is an eraser. Mutations of NSUN2, DNMT3b, ALYREF, NSUN5, DNMT1, and TET2 suggested that the function of m^5C in tumor cells may be abnormal. The above analyses demonstrated the high heterogeneity of the genetic and expressional alteration landscape in m^5C regulators between LUAD samples, indicating that the expression imbalance between m^5C regulators plays a crucial role in the occurrence and progression of LUAD.

M^5C methylation modification patterns mediated by 11 regulators

PCA analyses of m^5C-related genes

We extracted 11 m^5C-related genes from TCGA and performed PCA analyses using prCOMP (there were 13 genes related to m^5C modification, but only 11 genes with a quantitative expression level). The first three principal components were shown by pca3d in *Figure 3A*. The 11 m^5C-related samples could be completely distinguished between tumor samples and normal samples.

Network analyses of m^5C-related genes

LUAD tumor samples from TCGA with available overall survival (OS) data and clinical information were enrolled into one meta-cohort. The prognostic values of the 11 m^5C regulators were revealed via a univariate Cox regression model (*Figure 3B*). The 11 regulators were not related to the prognosis of LUAD patients, except for NSUN7, which also indicated that these 11 genes may indirectly interfere with the prognosis of LUAD patients. The m^5C regulatory network described the integrated view of the mutual effect of m^5C regulators, regulator connection, and their prognostic value for LUAD patients (*Figure 3C* and Table S1). The 11 genes were divided into four clusters. We found a correlation between expression and functional category of similar m^5C regulators. ALYREF may be a key gene of m^5C regulators, which affects the prognosis of LUAD through forward and reverse regulation of the other 10 genes.

TME cell infiltration characteristics in distinct m^5C modification patterns

Identification of m^5C modified subtypes (m^5C clusters)

We used the NMF R package to classify patients into two distinct modification patterns via unsupervised clustering, according to the expression quantity of 11 m^5C regulators (*Figure 4A,B*). A total of 504 samples were included, including 152 samples for cluster C1 and 352 samples for cluster C2. We termed these patterns: m^5C cluster C1 and C2, respectively. Furthermore, prognostic analysis for the two main m^5C modification subtypes was also performed, and the results showed significant differences in OS between cluster C1 and C2 (*Figure 4C*). The m^5C cluster C2 modification pattern exhibited a significant survival advantage. Then, we analyzed the expression of 11 m^5C regulators in the two main m^5C modification subtypes. The expression of the 7 genes among 11 regulators were significantly different between cluster C1 and C2, and all 7 genes' expression is higher in the cluster C1 (*Figure 4D*).

Functional enrichment of m^5C modified subtypes

In order to explore the biological behavior of these different m^5C modification patterns, enrichment analysis of GSVA was carried out using R language GSVA package, with the c2 cp.kegg.v7.0.symbols.gmt gene set as a background. A

© Translational Lung Cancer Research. All rights reserved.

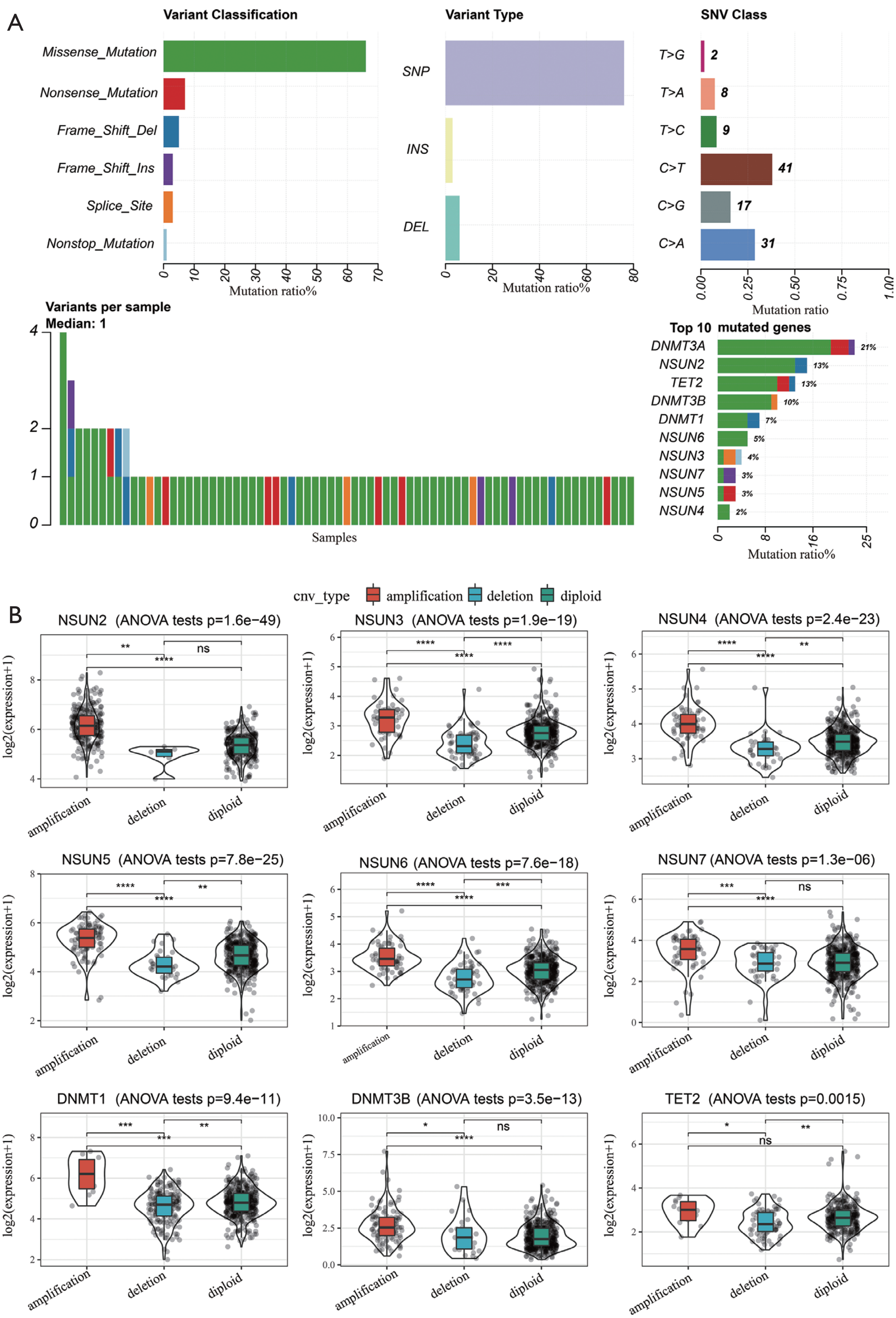


Figure 2 Landscape of m^{5}C regulators in LUAD. (A) Mutations of the first 10 genes related to m^{5}C; (B) the relationship between CNV and expression of nine genes related to m^{5}C modification. ns, no significant difference; *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001. LUAD, lung adenocarcinoma.

© Translational Lung Cancer Research. All rights reserved.

Table 1 The proportion of amplification and deletion of 11 genes related to m^{5}C modification

Role	Gene symbol	Amplification	Diploid	Deletion	CNV_sum	Amplification %	Deletion%
Writers	*NSUN2*	279	826	6	1111	25.1	0.540054
	NSUN3	53	985	64	1102	4.8	5.807623
	NSUN4	64	998	42	1104	5.8	3.804348
	NSUN5	102	964	32	1098	9.289617	2.91439
	NSUN6	67	978	57	1102	6.079855	5.172414
	NSUN7	62	990	48	1100	5.636364	4.363636
	DNMT1	12	919	174	1105	1.085973	15.74661
	DNMT2	72	1017	10	1099	6.55141	0.909918
	DNMT3B	112	967	28	1107	10.11743	2.529359
	NSUN1	–	–	–	–	–	–
	DNMT3A	–	–	–	–	–	–
Erasers	*TET2*	13	433	80	526	2.471483	15.20913
ALYREF	*ALYREF*	146	937	15	1098	13.2969	1.36612

total of 187 pathways were enriched, and 39 differential pathways were screened by |t|>6. The m^{5}C C1 subgroup was enriched in 14 pathways, mainly related to matrix pathways such as cell cycle and DNA (deoxyribonucleic acid) repair, while C2 was enriched in 25 pathways, mainly related to signal transduction and immune pathways (such as Fc epsilon RI signaling pathway and the mitogen-activated protein kinase (MAPK) signaling pathway) (*Figure 5*).

TME analyses of m^{5}C modified subtypes

The CIBERSORT method was used to analyze the composition of immune cells of two m^{5}C modification patterns (32). C1 was primarily composed of naïve B cells, activated CD4 memory T cells, follicular T helper cells, resting NK cells, M0 macrophages, and M1 macrophages, while C2 was mainly composed of memory B cells, resting CD4 memory T cells, monocytes, M2 macrophages, resting dendritic cells, resting mast cells, neutrophils, and eosinophils (*Figure 6A*).

The correlation between m^{5}C-related genes and TME infiltration type was calculated using the rcorr function of Hmisc package in R language. As shown in *Figure 6B*, the DNMT3B gene was significantly associated with 10 TME infiltrating immune cell groups, of which, six were composed of m^{5}C modified C1 immune cells (naïve B cells, activated CD4 memory T cells, follicular T helper cells, resting NK cells, M0 macrophages and M1 macrophages). The remaining four were composed of immune cells of the C2 subgroup (memory B cells, resting CD4 memory T cells, resting dendritic cells and resting mast cells). We used the ESTIMATE algorithm to quantify DNMT3B (*Figure 6C*). DNMT3B expression was inversely correlated with the immune, matrix, and ESTIMATE scores. Furthermore, we analyzed the expression of DNMT3B in 21 immune cells, and found that the low expression of DNMT3B was significantly increased in the 21 immune cells (*Figure 7A*).

Next, we analyzed the relationship between the expression of DNMT3B and ICB inhibitors. Abnormal expression of DNMT3B was associated with immune function disorder (*Figure 7B*). Subsequent analyses of pathway enrichment revealed that tumors with high DNMT3B expression exhibited enrichment in the Nod-like receptor (NLR) signaling pathway, cytosolic DNA-sensing pathway, and RIG-I-like receptor (RLR) signaling pathway (*Figure 7C*). Furthermore, we analyzed the OS of high and low expression groups of DNMT3B. The results showed that low DNMT3b gene expression group was associated with immunity and had a better prognosis (*Figure 7D*).

Generation of m^{5}C gene signatures and functional annotation

Using the limma package from R language, 226 DEGs were screened by |log2fold change| >1 and FDR <0.05,

© Translational Lung Cancer Research. All rights reserved.

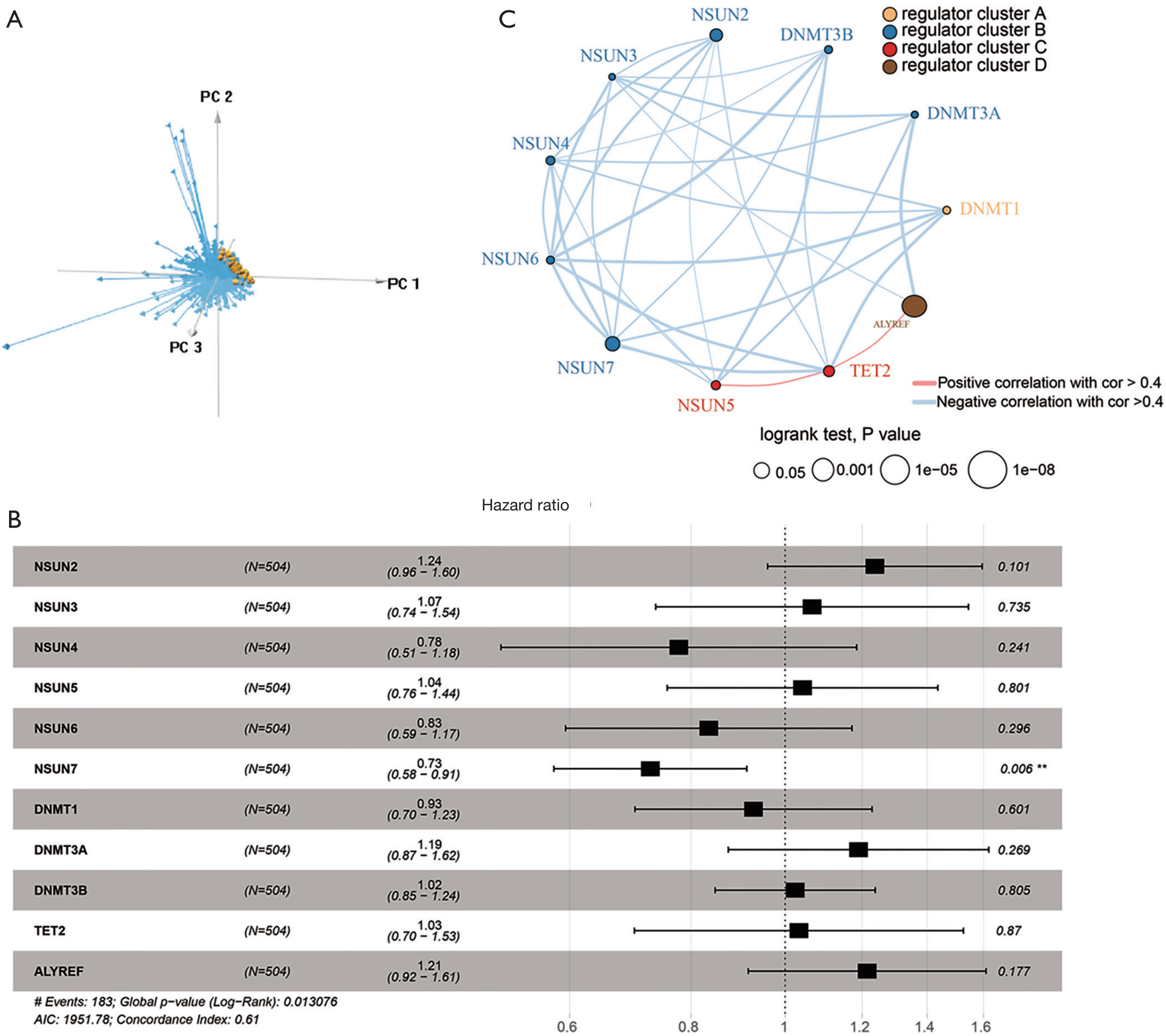


Figure 3 M^{5}C methylation modification patterns mediated by 11 regulators. (A) PCA for the expression of 11 m^{5}C regulators to distinguish tumors from normal samples. Tumors were marked with blue, and normal samples were marked with yellow; (B) the prognostic analyses for 11 m^{5}C regulators using a univariate Cox regression model; (C) the interaction between m^{5}C regulators in LUAD. **, P<0.01. PCA, principal component analysis; LUAD, lung adenocarcinoma.

all of which were related to the m^{5}C phenotype. The patients were divided into three different gene cluster subtypes through unsupervised clustering of 226 m^{5}C phenotype-related genes (the cluster method was complete and Euclidean was used to calculate the distance between samples). PCA analysis demonstrated that they were separated from each other (*Figure 8A*). These three clusters were named m^{5}C gene cluster C1–C3. We also observed the distribution of the 11 genes in the three m^{5}C gene clusters (*Figure 8B*), and found that most samples of gene cluster C2 and C3 were included in m^{5}C cluster C2, and most samples of gene cluster C1 coincided with m^{5}C cluster C1. In order to further determine which biological processes these 226 genes were primarily involved in, R language WebGestaltR

© Translational Lung Cancer Research. All rights reserved.

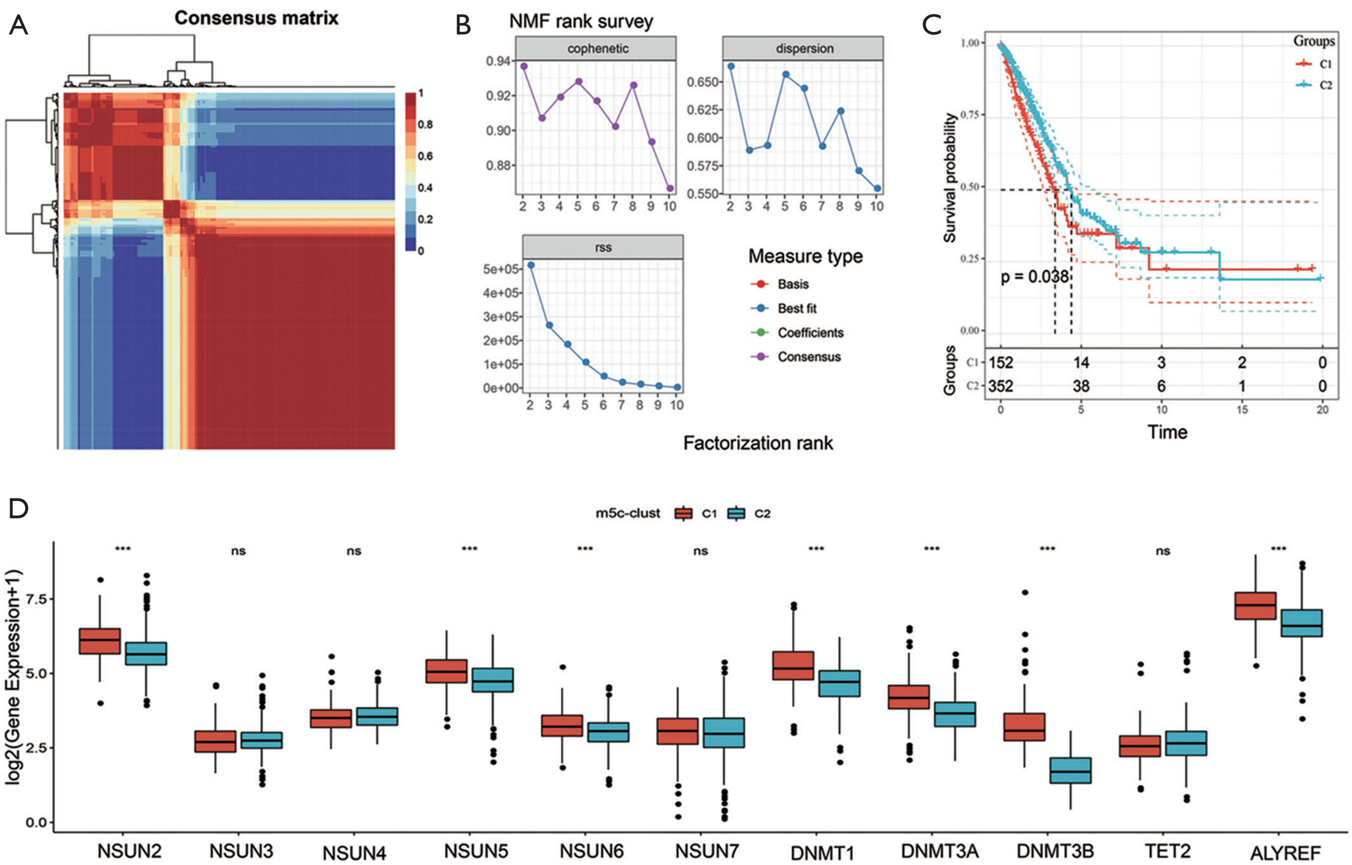


Figure 4 Identification of m^{5}C modified subtypes. (A) Consensus map of NMF clustering; (B) Cophenetic, RSS, and dispersion distributions with rank =2–10; (C) OS survival curves of m^{5}C clusters C1 and C2; (D) expression of 11 genes in two m^{5}C modification clusters. ns, no significant difference; ***, P<0.001. NMF, non-negative matrix factorization; RSS, residual sum of squares; OS, overall survival.

package was used for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (31441146) (33). We screened a total of five pathways (by P<0.05): cell cycle, oocyte meiosis, progesterone mediated oocyte maturation, cellular sense, and the p53 signaling pathway. The 226 genes were associated with m^{5}C modification and were significantly related to tumorigenesis (*Figure 8C*).

Subsequently, the distribution of 21 immune cells in the three subtypes of the m^{5}C gene cluster was analyzed. As shown in *Figure 8D*, the three subtypes were statistically significant in 14 cells. Thus, it was clear that m^{5}C modification had a critical role in TME, and the 226 genes modified by m^{5}C also played an important role in the TME. We further analyzed the KM curve of gene clusters C1–C3, and found that these three subtypes were associated with prognosis (P<0.05, *Figure 8E*). Although the samples were divided into three subtypes, there were only nine cases of C3 samples. These results were consistent with the classification of m^{5}C modification patterns. The prognosis of C2 was superior to that of C1.

Establishment of the m^{5}C score model

Due to the individual heterogeneity and complexity of m^{5}C modification, a scoring system was constructed to quantify the m^{5}C modification pattern of individual LUAD patients, which was called the m^{5}C score. Firstly, we screened 124 genes related to prognosis (P<0.05) from 226 isoform differential genes. Table S2 shows the results of the univariate COX analysis of 124 genes. PCA analysis was then performed on the 124 genes, PC1 and PC2 scores were taken, and the m^{5}C score of each sample was calculated as follows: m^{5}C-score=ΣPC1i+PC2i. The m^{5}C score results of the 513 samples are displayed in Table S3.

We divided the high and low score groups according to the median of the m^{5}C score and used the alluvial diagram

© Translational Lung Cancer Research. All rights reserved.

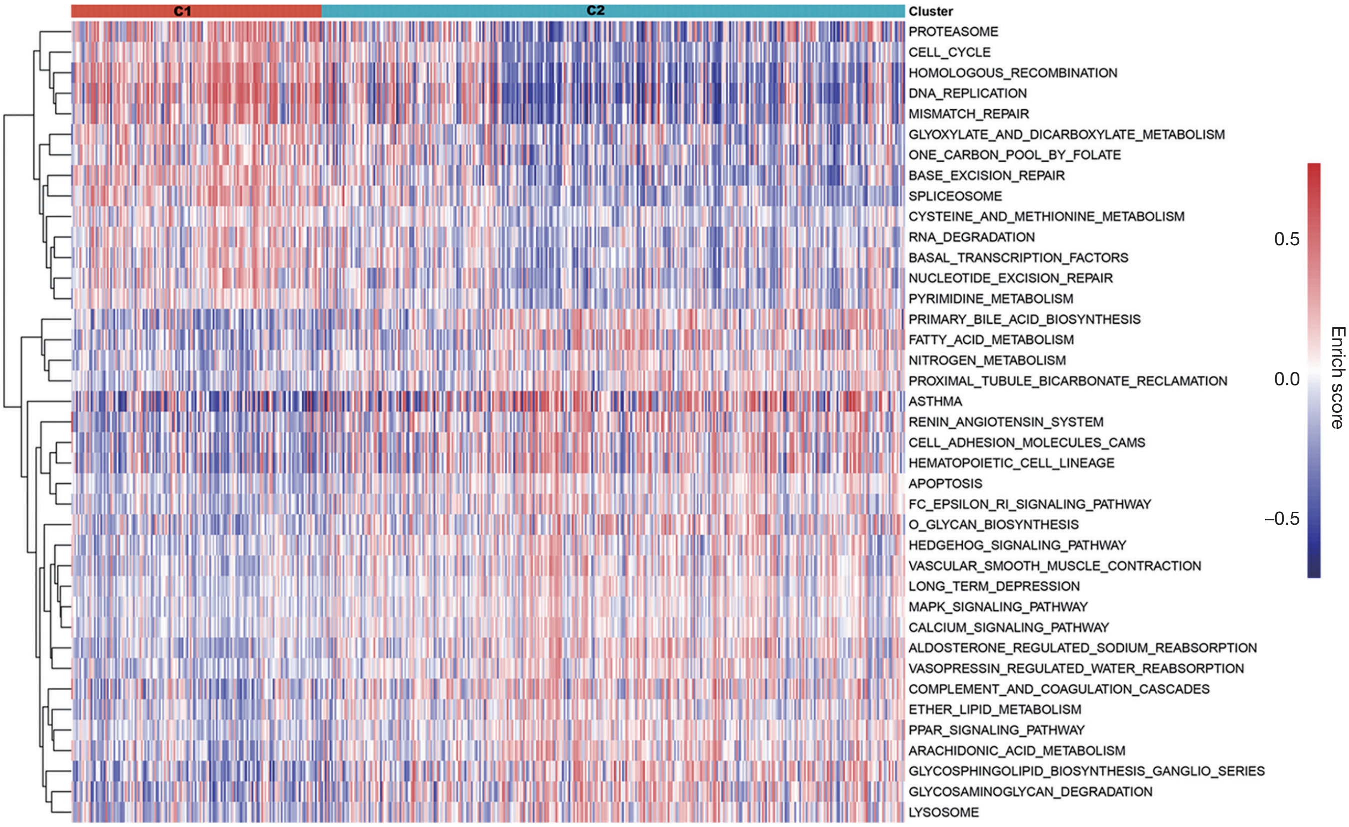


Figure 5 Expression of 39 pathways in the GSVA analysis of two m^{5}C modification clusters. GSVA, gene set variation analysis.

to demonstrate the changes between m^{5}C clusters, gene clusters, and m^{5}C scores (*Figure 9A*). We found that most of the samples of the m^{5}C cluster C2 subtype with good prognosis were identical with those of the gene cluster C2 subtype, and patients with good prognosis primarily exhibited a high m^{5}C score.

To further verify the relationship between our m^{5}C score model and the prognosis of LUAD, we divided the high and low score groups according to the median m^{5}C score. Survival analysis was then performed between these two groups. We observed that the high m^{5}C score group had a better prognosis, which was consistent with the results of the previous analysis (*Figure 9B*).

As shown in *Figure 9C*, there was a significant difference in the m^{5}C scores among the three gene cluster subtypes, with cluster C2 scoring the highest, and cluster C1 the lowest, which also verified that a high m^{5}C score had a good prognosis. Additionally, m^{5}C score difference was also statistically significant between the two m^{5}C cluster subtypes (*Figure 9D*). The score of the C2 subtype was markedly higher than that of the C1 subtype, and the prognosis of C2 was better than that of C1, which further verified that a high m^{5}C score had a better prognosis. Therefore, a high m^{5}C score may predict a good prognosis for LUAD patients, while a low m^{5}C score may predict a poor prognosis.

We also performed GSVA analysis to further explore the biological process involved in the m^{5}C score difference. We found that the low m^{5}C score group was mainly related to pathways of DNA repair, cell cycle, and stroma, while the high m^{5}C score group was primarily associated with immune-related pathways and MAPK signaling pathways (*Figure 9E*). Furthermore, through multivariate Cox regression model analysis, we found that m^{5}C score was an independent prognostic factor (sample with missing clinical information removed) (*Figure 9F*).

Moreover, we analyzed the expression of 11 m^{5}C regulators in the high and low m^{5}C score groups. The expression of seven regulators exhibited significant correlation with m^{5}C score. As shown in *Figure 10*, in addition to TET2, a high m^{5}C score also corresponded to low gene expression (*NSUN2*, *NSUN5*, *DNMT1*, *DNMT3A*, *DNMT3B*, and *ALYREF*).

© Translational Lung Cancer Research. All rights reserved.

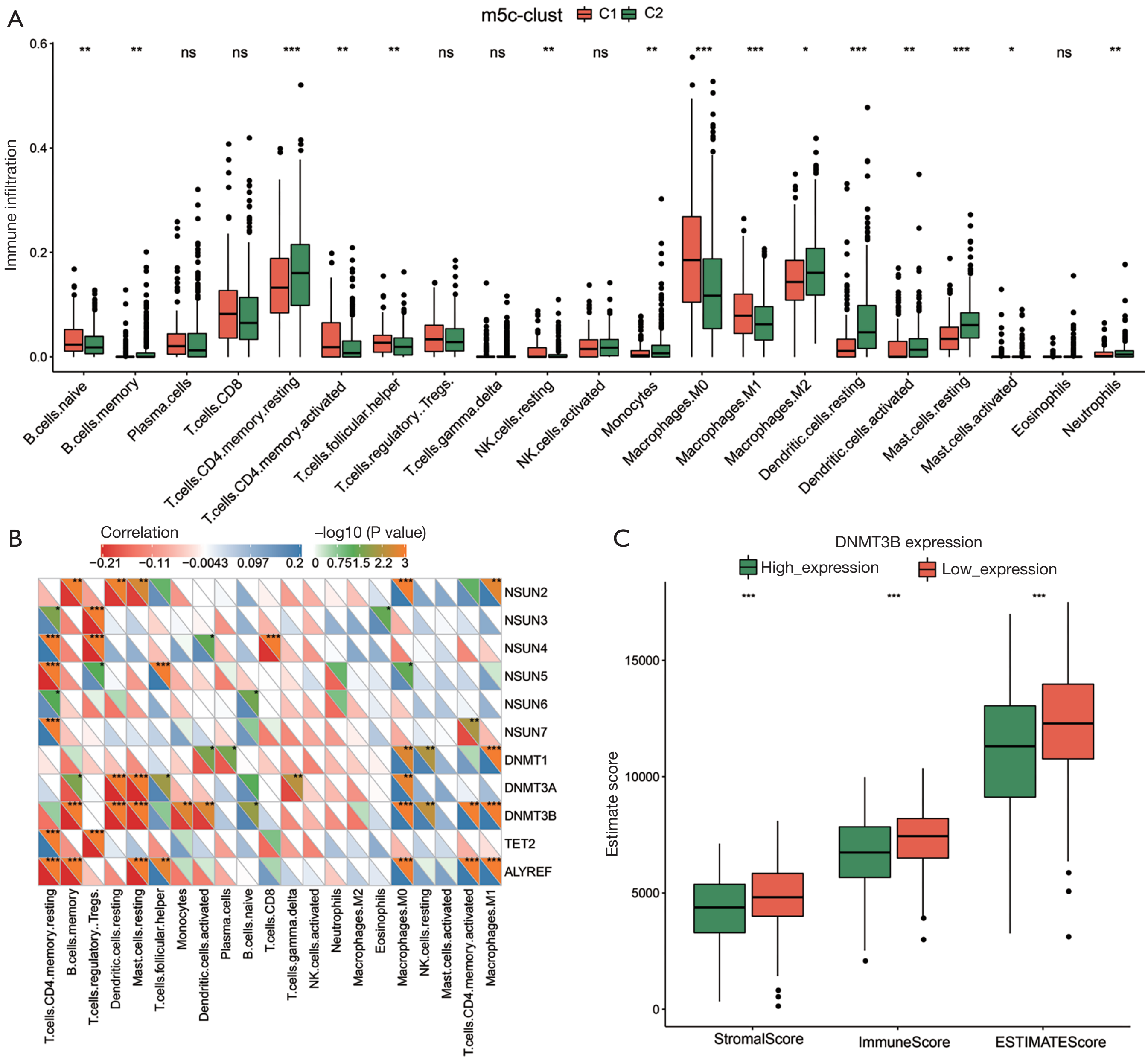


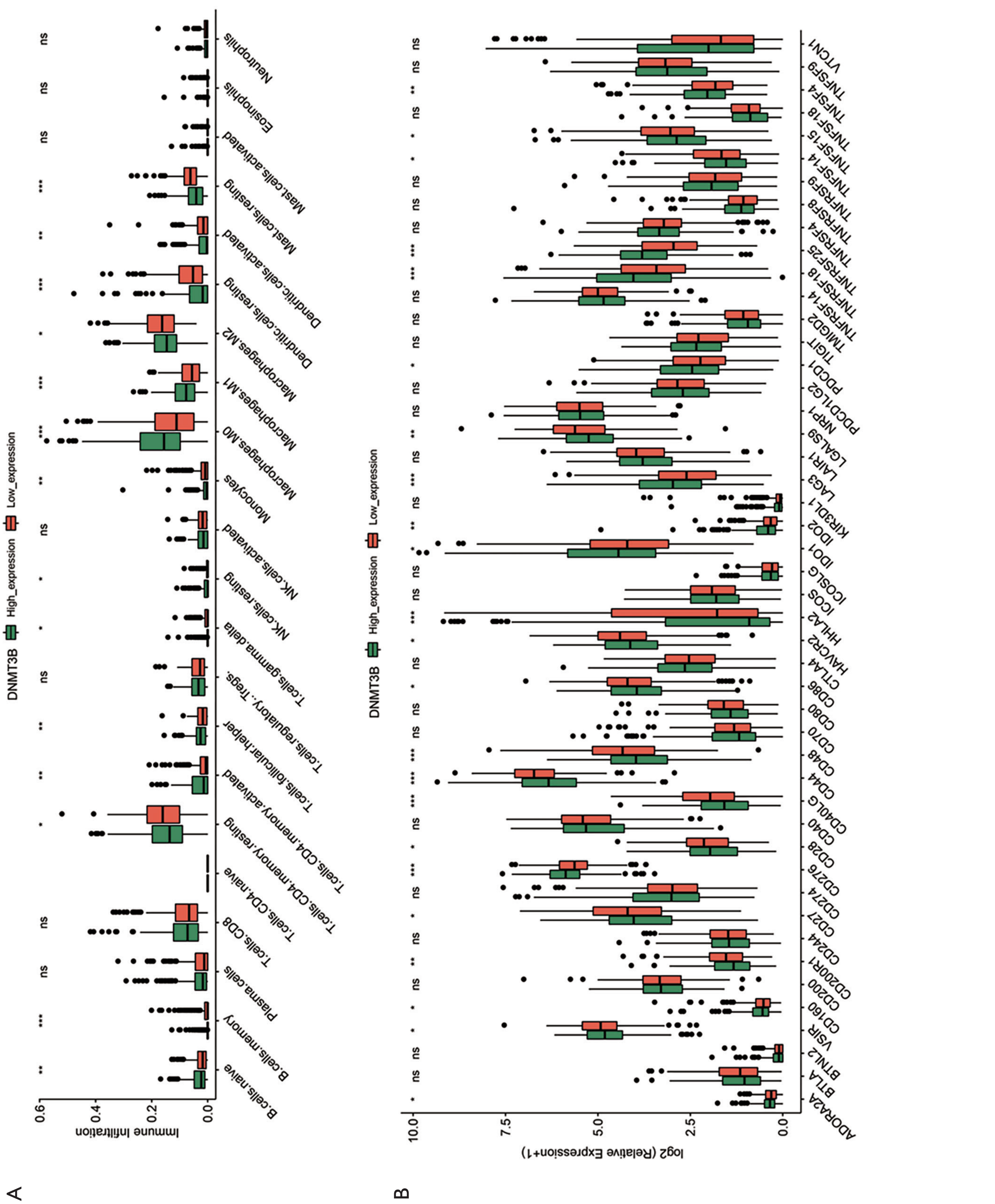
Figure 6 TME cell infiltration characteristics and transcriptome traits in distinct m^{5}C modification patterns. (A) The abundance of each TME infiltration cell in two m^{5}C modification patterns; (B) the correlation between each TME infiltration cell type and each m^{5}C regulator using Spearman analyses; (C) difference in stromal, immune, and ESTIMATE scores between high and low DNMT3B expression groups. ns, no significant difference; *, P<0.05; **, P<0.01; ***, P<0.001. TME, tumor microenvironment.

Validation of external datasets

Establishment of GSE31210 dataset m^{5}C score

The PCA analysis results of 125 genes obtained from the previous analysis were used to establish a new m^{5}C score model based on the GSE31210 dataset. In total, 116 genes were identified in the GSE31210 dataset, which were used to establish the m^{5}C score model for 226 tumor samples in GSE31210. First, through PCA analysis, PC1 and PC2 of the 116 genes were calculated, and the m^{5}C score was calculated for each sample. *Figure 11A* shows the distribution of the 11 genes in the high and low m^{5}C

© Translational Lung Cancer Research. All rights reserved.



© Translational Lung Cancer Research. All rights reserved.

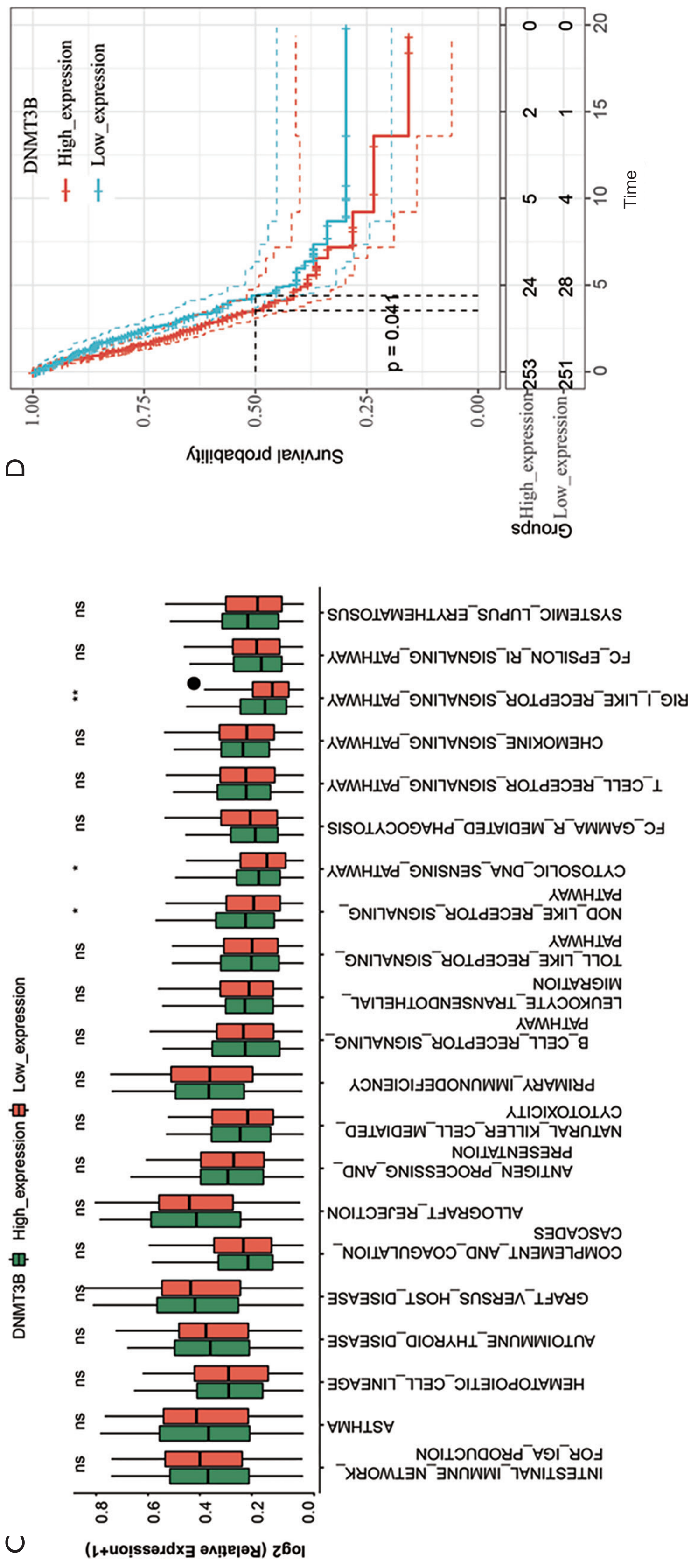


Figure 7 The expression of DNMT3B is associated with TME. (A) Differences in the abundance of each TME infiltration cell between high and low DNMT3B expression groups; (B) differences in the expression of immune checkpoint between high and low DNMT3B expression groups; (C) differences in the immune related pathways between high and low DNMT3B expression groups; (D) survival analyses for patients with low or high DNMT3B expression. ns, no significant difference; *, P<0.05; **, P<0.01; ***, P<0.001. TME, tumor microenvironment.

© Translational Lung Cancer Research. All rights reserved.

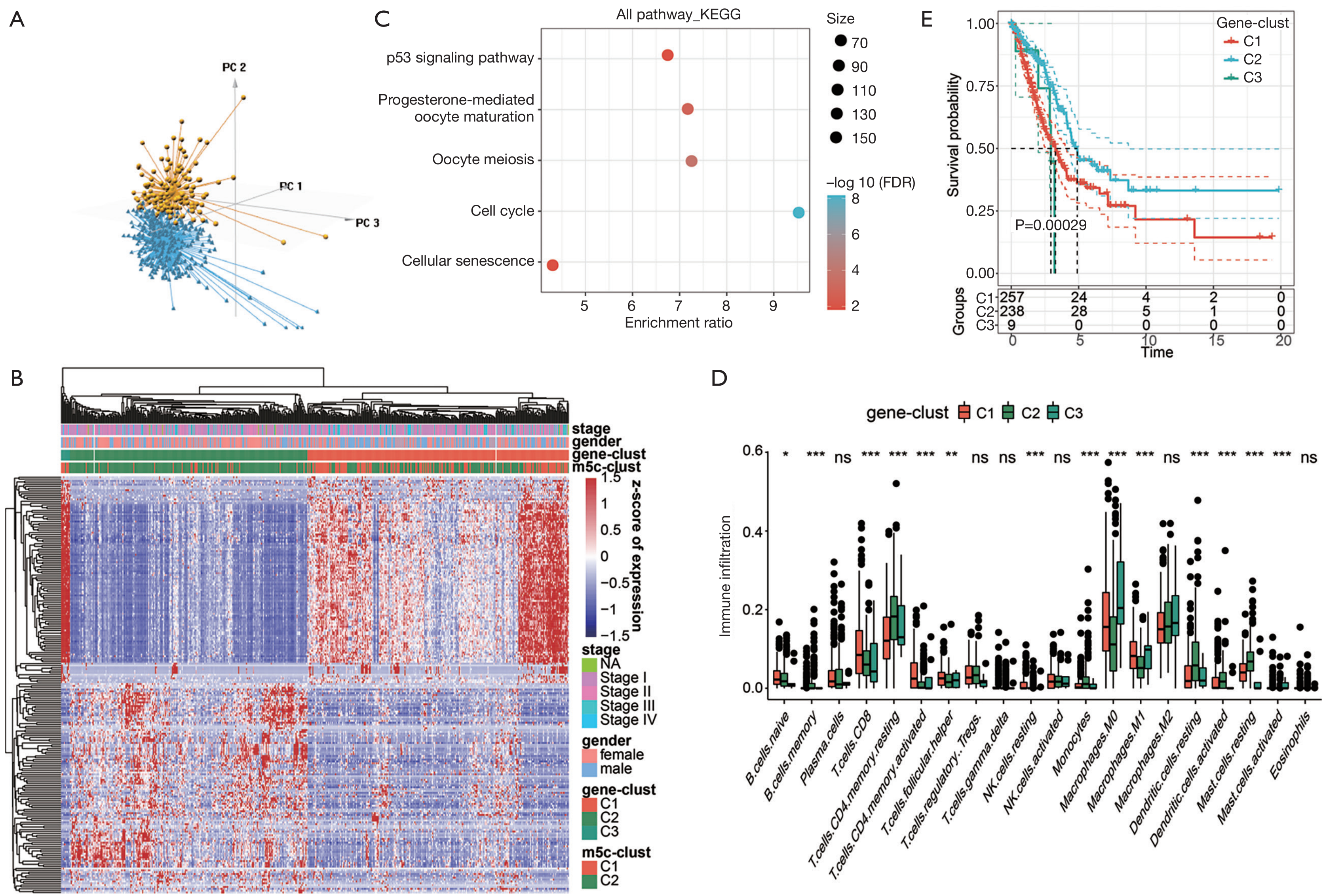


Figure 8 Generation of m^{5}C gene signatures and functional annotation. (A) Principal component analysis for two m^{5}C modification patterns to distinguish m^{5}C clusters C1 and C2; (B) unsupervised clustering of overlapping m^{5}C phenotype-related genes was performed to classify patients into different genomic subtypes, termed m^{5}C gene clusters C1–C3, respectively. The gene clusters, m^{5}C clusters, gender, and stage were used as patient annotations; (C) KEGG enrichment analysis of 226 m^{5}C phenotype-related genes; (D) the abundance of each TME infiltration cell in three m^{5}C gene clusters; (E) survival analyses for patients of m^{5}C gene cluster C1, C2, and C3. ns, no significant difference; *, P<0.05; **, P<0.01; ***, P<0.001. TME, tumor microenvironment.

© Translational Lung Cancer Research. All rights reserved.

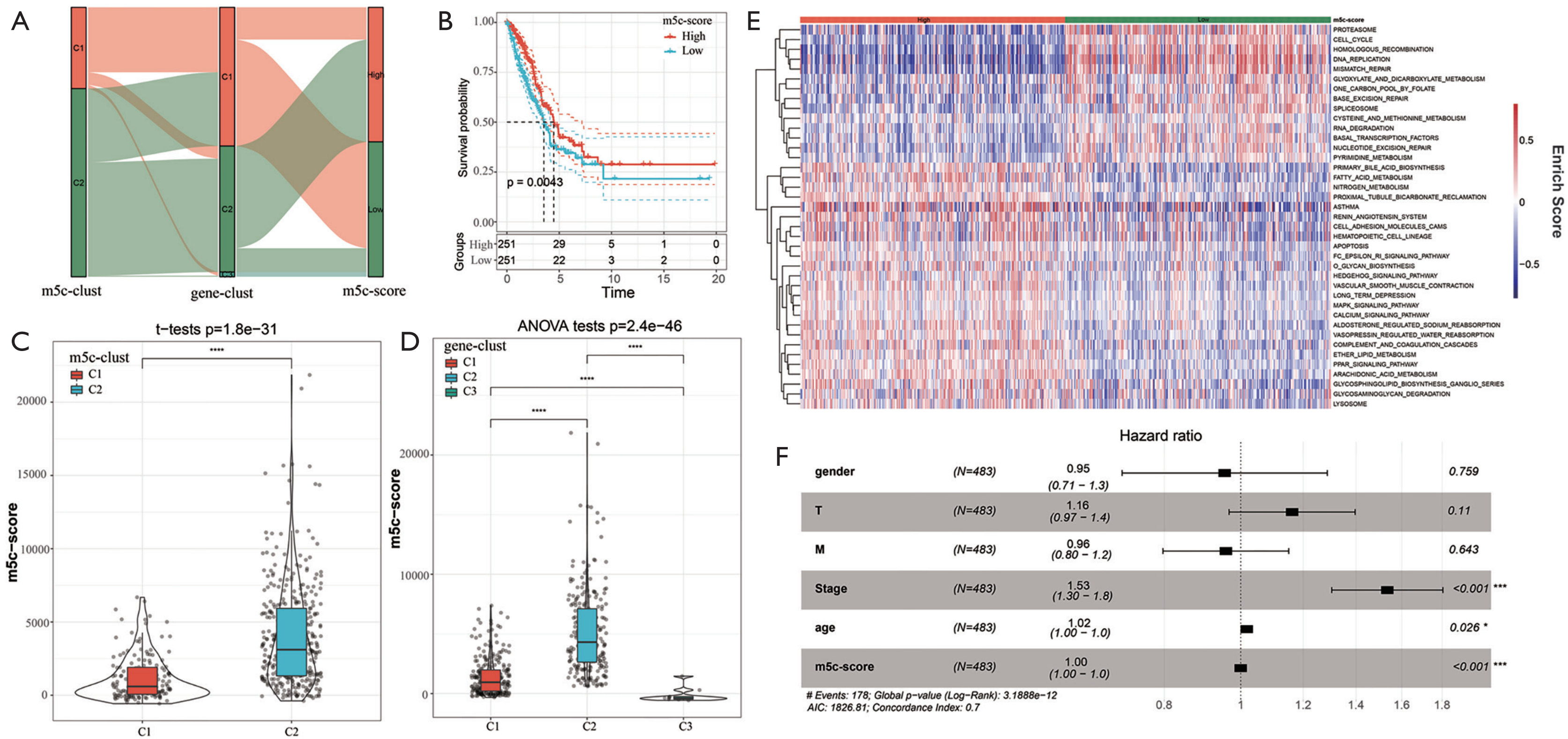


Figure 9 Establishment of m^{5}C-score model. (A) Alluvial diagram showing the changes of m^{5}C clusters, gene clusters, and m^{5}C scores; (B) survival analyses for both low and high m^{5}C score patient groups; (C) distribution of patients with different m^{5}C scores in two m^{5}C modification clusters; (D) distribution of patients with different m^{5}C scores in three m^{5}C gene clusters; (E) GSVA enrichment analysis showing the activation states of biological pathways in both high and low m^{5}C score groups; (F) multivariate Cox regression analysis for m^{5}C score in LUAD patients shown by forest plot. *, P<0.05; ***, P<0.001; ****, P<0.0001. GSVA, gene set variation analysis; LUAD, lung adenocarcinoma.

© Translational Lung Cancer Research. All rights reserved.

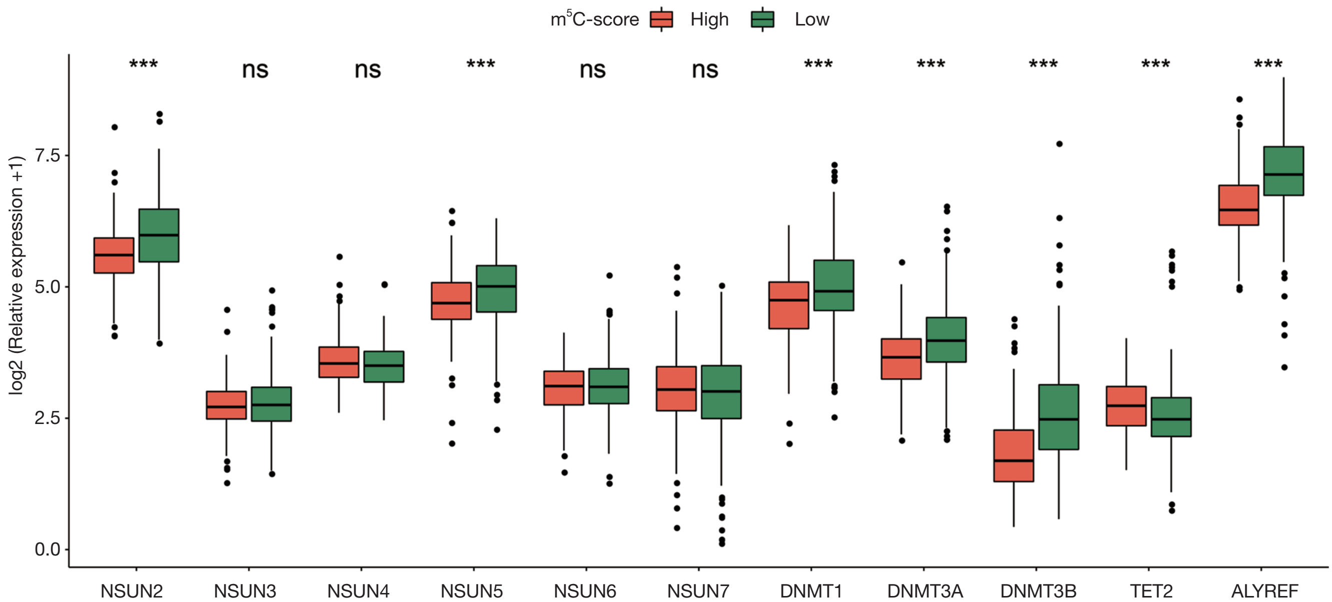


Figure 10 The expression of 11 m5C regulators in both high and low m5C score groups. ns, no significant difference; ***, P<0.001.

scores. *Figure 11B* displays the prognosis of the high and low m5C score groups; a high m5C score may lead to a good prognosis, which was consistent with the results of TCGA.

GSVA analysis of the high and low m5C score groups

In order to further investigate mechanisms through which the m5C score affected biological processes, we performed GSVA analysis using the R language. The results showed that the high m5C score group was associated with immune pathways, such as the complement and coagulation cascades, leukocyte transendothelial migration, and the intestinal immune network for immunoglobulin A (IgA) production. Meanwhile, the low score group was associated with the pathways related to the stroma, such as basal resection and repair, cell cycle, etc. (*Figure 11C*).

These results verified that high m5C scores were related to an immune desert type, which predicted a good prognosis, while low m5C scores indicated an immune exclusion phenotype, which suggested a poor prognosis. The table online (https://cdn.amegroups.cn/static/public/tlcr-21-351-1.xlsx) exhibits the enrichment scores of all samples in 186 pathways, and Table S4 shows the enrichment results of the high and low m5C score groups.

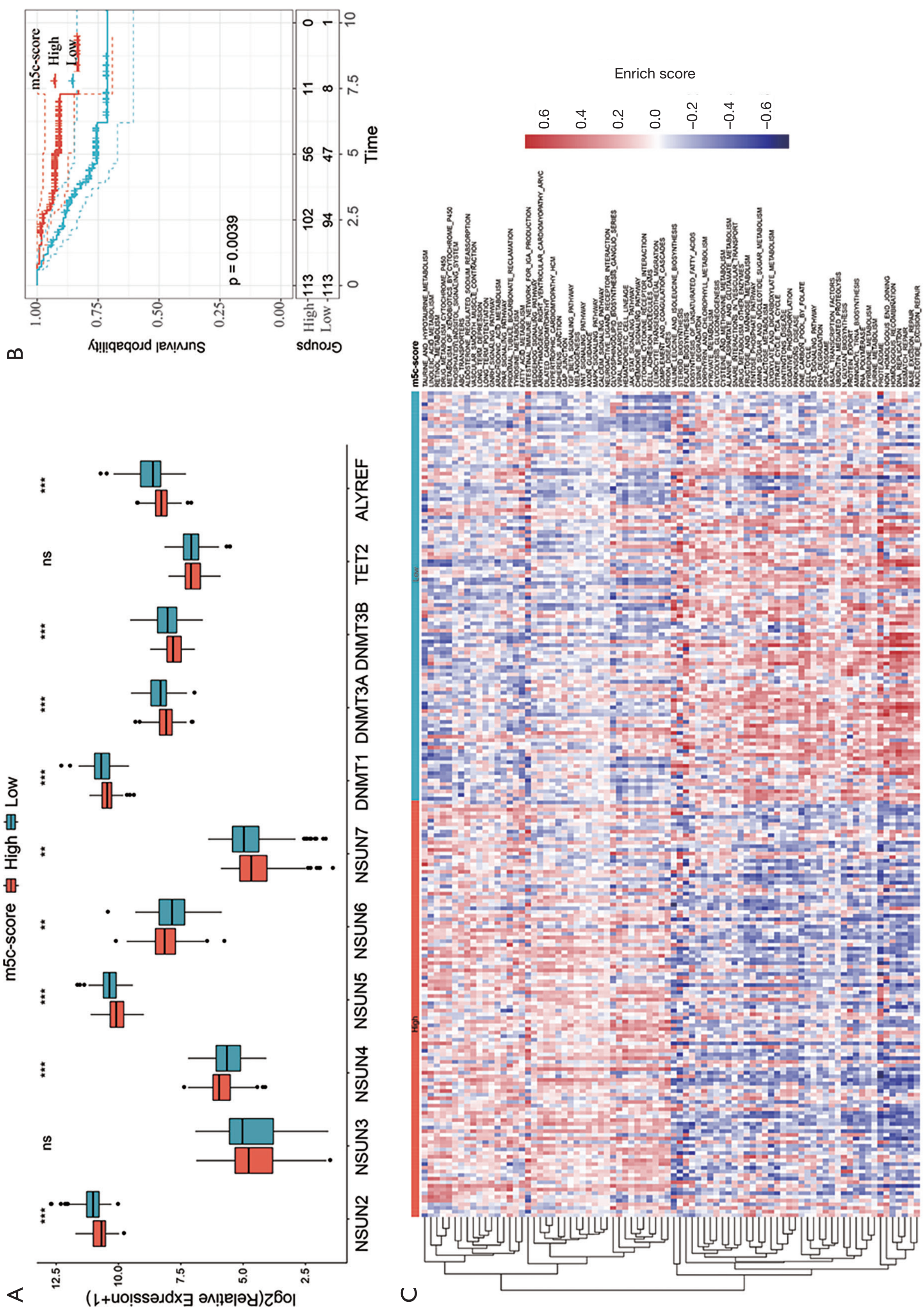
Composition of immune cells in the high and low levels of the m5C score model

To further verify the immunophenotype of the high and low m5C score groups of the dataset, we used CIBERSORT to analyze the composition of immune cells in the high and low m5C score groups (*Figure 11D*). The high m5C score exhibited more infiltration of resting CD4 memory T cells and resting mast cells, as well as less infiltration of M0 and M1 macrophages, which was similar to the immunocyte infiltration of gene cluster C2.

Discussion

With the development of deep sequencing and mass spectrometry (30), accumulating evidence has suggested that m5C modification is very important for maintaining the normal physiological function of cells and organisms (31-36), while its abnormal distribution and expression are closely related to tumor development. Studies have confirmed that m5C is involved in the progression of hepatocellular carcinoma (37,38). Also, there is increasing evidence that methylation regulatory factors can be used as prognostic and diagnostic markers of cancer (39-43). For example, the high expression of NSUN1 has been identified as a prognostic marker for non-small cell lung cancer (44-46). Recent studies have also confirmed that m5C may affect the behavior of immune cells, such as CD^+ T cells (47). Since most studies have focused on the effect of single TME cell types or regulators on tumor development, there remains a lack of comprehensive recognition of TME

© Translational Lung Cancer Research. All rights reserved.


A
m5c-score
High
Low
log2(Relative Expression+1)
12.5
10.0
7.5
5.0
2.5
***
ns
**
NSUN2
NSUN3
NSUN4
NSUN5
NSUN6
NSUN7
DNMT1
DNMT3A
DNMT3B
TET2
ALYREF
B
m5c-score
High
Low
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.0039
Groups
High
Low
113
102
56
11
0
94
47
8
1
0
2.5
5
7.5
10
Time
C
Enrich score
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
m5c-score

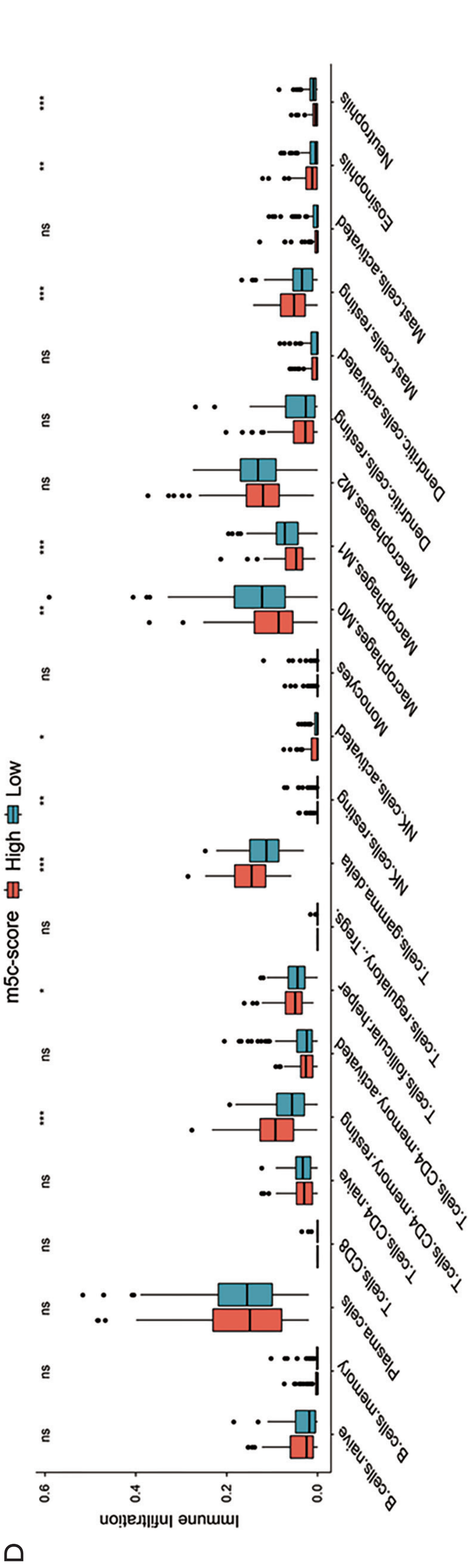


Figure 11 GSVA analysis of the high and low m^5C score groups. (A) The expression of 11 m^5C regulators in both high and low m^5C score groups in the GSE31210 dataset; (B) survival analysis of both high and low m^5C score patient groups in the GSE31210 dataset; (C) GSVA enrichment analysis showing the activation states of biological pathways of both high and low m^5C score groups in the GSE31210 dataset; (D) the abundance of each TME infiltration cell in both high and low m^5C score groups in the GSE31210 dataset. ns, no significant difference; *, P<0.05; **, P<0.01; ***, P<0.001. GSVA, gene set variation analysis.

infiltration mediated by multiple m^5C regulators. Further understanding of the role of different m^5C modification patterns in the infiltration of TME cell will help to improve our understanding of the TME antitumor immune response and provide novel immunotherapy strategies.

In this study, two m^5C methylation modification patterns were revealed according to 11 m^5C regulators, which had remarkably distinct TME immune cell infiltration characterization. Also, three genomic subtypes of the m^5C gene were identified based on 226 m^5C phenotype-related DEGs, which were also significantly related to tumor occurrence. This further revealed the important role of m^5C modification in influencing the TME landscape. Identification of the m^5C modification patterns of individual tumors was crucial due to the individual heterogeneity of m^5C modification. Thus, a scoring system was constructed to assess the m^5C modification pattern of LUAD patients. The m^5C cluster C2 exhibited a higher m^5C score, and patients in the m^5C cluster C2 showed better prognosis. The high m^5C score group had a better prognosis, while the low m^5C score group had a poor prognosis. These results were further verified in the GSE31210 dataset, which indicated that the m^5C score was a reliable method for the integrated evaluation of distinct tumor m^5C modification patterns. Comprehensive analyses also proved that the m^5C score was an independent prognostic marker in LUAD. Functional enrichment analyses in the groups with better prognosis tended to be associated with immunity; m^5C cluster C2 exhibited enrichment pathways related to immunity, such as the Fc epsilon RI signaling pathway, and the high m^5C score group in the GSE31210 dataset was correlated with immune pathways, such as the complement and coagulation cascades, leukocyte trans-endothelial migration, and the intestinal immune network for IgA production. *NSUN2*, *NSUN5*, *DNMT1*, *DNMT3A*, *DNMT3B*, and *ALYREF* were highly expressed in m^5C cluster C2, as well as in TCGA and GSE31210 low m^5C score groups, which had a poor prognosis. Above, we analyzed immune cell infiltration, immune checkpoint characteristics, and functional enrichment analysis among different expression levels of DNMT3B in LUAD.

Our study provides some insight for clinical application. Our m^5C score system could serve as a reliable and independent biomarker for predicting the prognosis of patients with LUAD. Our findings may help to screen suitable patients who can benefit from immune checkpoint inhibitor therapy. Further research based on these m^5C regulators, which regulate TME immune

© Translational Lung Cancer Research. All rights reserved.

cell infiltration, may contribute to the discovery of novel immune drug combination treatment strategies or new immunotherapeutic agents, and promote the development of individual tumor immunotherapy.

The methylation modification patterns of gastric cancer, LUAD, and other cancers, which are mediated by the m^6A modulator, and the invasion characteristics of the TME have been studied, and the m^6A modulator is closely related to the tumor immunophenotype (48-53). Studies have also revealed that cross-talk between m^6A and m^5C regulators is associated with tumor immunogenicity and prognosis in 33 cancer types (54). In future studies, we will also aim to explore whether m^5C and m^6A have a synergistic effect on LUAD tumor microenvironmental characteristics and the patients' response to immunotherapy. We will also further investigate how genes (*NSUN2*, *NSUN5*, *DNMT1*, *DNMT3A*, *DNMT3B* and *ALYREF*) that are highly expressed in groups with poor prognosis work. In addition, we cannot rule out the possibility that m^5C regulatory factors affect the behavior of the matrix in the TME. Some researchers have found that m^5C is related to PM2.5-induced pulmonary fibrosis in mice (55), thus the regulatory behavior of m^5C on the TME may be complex.

Our study had limitations that should be noted. Firstly, we did not consider the correlation between immune infiltration location and TME heterogeneity. Secondly, due to the limited clinical annotation in public datasets, the clinicopathological parameters detected in this study are not comprehensive, which may contribute to potential bias in the predictive performance when the m^5C score signature served as a prognosis biomarker. Thirdly, due to the time constraints and lack of enough budget, we haven't carried out relevant experiments now. In future work, we will conduct further experiments to validate the results. Finally, due to the lack of overall clinical information in the datasets involved, we could not directly analyze the correlation between m^5C score and the response of LUAD patients to immunotherapy.

Conclusions

In this study, we found that m^5C modification played a significant role in formation of TME diversity and complexity. Based on the characteristics of m^5C modification, a score model was constructed to predict the prognosis of LUAD patients, which was also verified in the external dataset. We believe that m^5C modification will have some implications for tumor immunotherapy in the future.

Acknowledgments

This research was partly presented as an e-Poster at European Lung Cancer Virtual Congress during 25–27 March 2021.
Funding: This work was supported by the Chinese Society of Clinical Oncology/Beijing Xisike Clinical Oncology Research Foundation (Y-Young2020-0003) and the Shanghai Sailing Program (20YF1408300, 19YF1409200).

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at http://dx.doi.org/10.21037/tlcr-21-351

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tlcr-21-351). 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).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Boccaletto P, Machnicka MA, Purta E, et al. MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res 2018;46:D303-D307.
2. Roundtree IA, Evans ME, Pan T, et al. Dynamic RNA Modifications in Gene Expression Regulation. Cell 2017;169:1187-200.
3. Chen K, Zhao BS, He C. Nucleic Acid Modifications in Regulation of Gene Expression. Cell Chem Biol

© Translational Lung Cancer Research. All rights reserved.

2016;23:74-85.
4. García-Vílchez R, Sevilla A, Blanco S. Post-transcriptional regulation by cytosine-5 methylation of RNA. Biochim Biophys Acta Gene Regul Mech 2019;1862:240-52.
5. 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.
6. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat Biotechnol 2013;31:458-64.
7. He X, Xu C. Immune checkpoint signaling and cancer immunotherapy. Cell Res 2020;30:660-9.
8. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. N Engl J Med 2015;373:123-35.
9. Wu YL, Lu S, Cheng Y, et al. Nivolumab Versus Docetaxel in a Predominantly Chinese Patient Population With Previously Treated Advanced NSCLC: CheckMate 078 Randomized Phase III Clinical Trial. J Thorac Oncol 2019;14:867-75.
10. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016;387:1540-50.
11. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. N Engl J Med 2015;373:1627-39.
12. Qureshi OS, Zheng Y, Nakamura K, et al. Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. Science 2011;332:600-3.
13. Sharpe AH, Pauken KE. The diverse functions of the PD1 inhibitory pathway. Nat Rev Immunol 2018;18:153-67.
14. Champiat S, Dercle L, Ammari S, et al. Hyperprogressive Disease Is a New Pattern of Progression in Cancer Patients Treated by Anti-PD-1/PD-L1. Clin Cancer Res 2017;23:1920-8.
15. van Kesteren MTR, Meeter M. How to optimize knowledge construction in the brain. NPJ Sci Learn 2020;5:5.
16. Martini C, Marrucci W, Lucacchini A, et al. Specific inhibition of benzodiazepine receptor binding by some 1,2,3-triazole derivatives. J Pharm Sci 1988;77:977-80.
17. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
18. Fridman WH, Pages F, Sautes-Fridman C, et al. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer 2012;12:298-306.
19. Petitprez F, Meylan M, de Reynies A, et al. The Tumor Microenvironment in the Response to Immune Checkpoint Blockade Therapies. Front Immunol 2020;11:784.
20. Fridman WH, Zitvogel L, Sautes-Fridman C, et al. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol 2017;14:717-34.
21. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol 2004;22:329-60.
22. Gonçalves RC, Freire PP, Coletti D, et al. Tumor Microenvironment Autophagic Processes and Cachexia: The Missing Link? Front Oncol 2021;10:617109.
23. Sacco A, Battaglia AM, Botta C, et al. Iron Metabolism in the Tumor Microenvironment-Implications for Anti-Cancer Immune Response. Cells 2021;10:303.
24. Schoeler K, Aufschnaiter A, Messner S, et al. TET enzymes control antibody production and shape the mutational landscape in germinal centre B cells. FEBS J 2019;286:3566-81.
25. Li H, Lu T, Sun W, et al. Ten-Eleven Translocation (TET) Enzymes Modulate the Activation of Dendritic Cells in Allergic Rhinitis. Front Immunol 2019;10:2271.
26. Yue X, Lio CJ, Samaniego-Castruita D, et al. Loss of TET2 and TET3 in regulatory T cells unleashes effector function. Nat Commun 2019;10:2011.
27. 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.
28. Janin M, Ortiz-Barahona V, de Moura MC, et al. Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program. Acta Neuropathol 2019;138:1053-74.
29. 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.
30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-2.
31. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
32. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.
33. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci 2019;28:1947-51.
34. Glasner H, Riml C, Micura R, et al. Label-free, direct localization and relative quantitation of the RNA

© Translational Lung Cancer Research. All rights reserved.

nucleobase methylations m6A, m5C, m3U, and m5U by top-down mass spectrometry. Nucleic Acids Res 2017;45:8014-25.
35. Chen H, Yang H, Zhu X, et al. m(5)C modification of mRNA serves a DNA damage code to promote homologous recombination. Nat Commun 2020;11:2834.
36. Trixl L, Lusser A. The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. Wiley Interdiscip Rev RNA 2019;10:e1510.
37. Sun Z, Xue S, Zhang M, et al. Aberrant NSUN2-mediated m(5)C modification of H19 lncRNA is associated with poor differentiation of hepatocellular carcinoma. Oncogene 2020;39:6906-19.
38. Fish GD, Stanley JH, Miller KS, et al. Postbiopsy pneumothorax: estimating the risk by chest radiography and pulmonary function tests. AJR Am J Roentgenol 1988;150:71-4.
39. Sun L, Liu WK, Du XW, et al. Large-scale transcriptome analysis identified RNA methylation regulators as novel prognostic signatures for lung adenocarcinoma. Ann Transl Med 2020;8:751.
40. He Y, Yu X, Li J, et al. Role of m(5)C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma. Am J Transl Res 2020;12:912-22.
41. Wang P, Wu M, Tu Z, et al. Identification of RNA: 5-Methylcytosine Methyltransferases-Related Signature for Predicting Prognosis in Glioma. Front Oncol 2020;10:1119.
42. McHam ML, Fulton A. Albinism. Int Ophthalmol Clin. Winter 1992;32:185-200.
43. 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.
44. Saijo Y, Sato G, Usui K, et al. Expression of nucleolar protein p120 predicts poor prognosis in patients with stage I lung adenocarcinoma. Ann Oncol 2001;12:1121-5.
45. Sato G, Saijo Y, Uchiyama B, Kumano N, Sugawara S, Fujimura S, et al. Prognostic value of nucleolar protein p120 in patients with resected lung adenocarcinoma. J Clin Oncol 1999;17:2721-7.
46. Uchiyama B, Saijo Y, Kumano N, et al. Expression of nucleolar protein p120 in human lung cancer: difference in histological types as a marker for proliferation. Clin Cancer Res 1997;3:1873-7.
47. Guo G, Wang H, Shi X, Yet al. Disease Activity-Associated Alteration of mRNA m(5) C Methylation in CD4(+) T Cells of Systemic Lupus Erythematosus. Front Cell Dev Biol 2020;8:430.
48. Zhang B, Wu Q, Li B, et al. m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. Mol Cancer 2020;19:53.
49. Li Y, Gu J, Xu F, et al. Molecular characterization, biological function, tumor microenvironment association and clinical significance of m6A regulators in lung adenocarcinoma. Brief Bioinform 2020. [Epub ahead of print]. doi: 10.1093/bib/bbaa225.
50. Xu S, Tang L, Dai G, et al. Expression of m6A Regulators Correlated With Immune Microenvironment Predicts Therapeutic Efficacy and Prognosis in Gliomas. Front Cell Dev Biol 2020;8:594112.
51. Xu F, Zhang Z, Yuan M, et al. M6A Regulatory Genes Play an Important Role in the Prognosis, Progression and Immune Microenvironment of Pancreatic Adenocarcinoma. Cancer Invest 2021;39:39-54.
52. Lin S, Xu H, Zhang A, et al. Prognosis Analysis and Validation of m(6)A Signature and Tumor Immune Microenvironment in Glioma. Front Oncol 2020;10:541401.
53. Fang J, Hu M, Sun Y, et al. Expression Profile Analysis of m6A RNA Methylation Regulators Indicates They Are Immune Signature Associated and Can Predict Survival in Kidney Renal Cell Carcinoma. DNA Cell Biol 2020. [Epub ahead of print]. doi: 10.1089/dna.2020.5767.
54. Chen YT, Shen JY, Chen DP, et al. Identification of cross-talk between m(6)A and 5mC regulators associated with onco-immunogenic features and prognosis across 33 cancer types. J Hematol Oncol 2020;13:22.
55. Han X, Liu H, Zhang Z, et al. Epitranscriptomic 5-Methylcytosine Profile in PM2.5-induced Mouse Pulmonary Fibrosis. Genomics Proteomics Bioinformatics 2020;18:41-51.

Cite this article as: Chen H, Ge XL, Zhang ZY, Liu M, Wu RY, Zhang XF, Xu LP, Cheng HY, Sun XC, Zhu HC. M^5C regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in lung adenocarcinoma. Transl Lung Cancer Res 2021;10(5):2172-2192. doi: 10.21037/tlcr-21-351

© Translational Lung Cancer Research. All rights reserved.

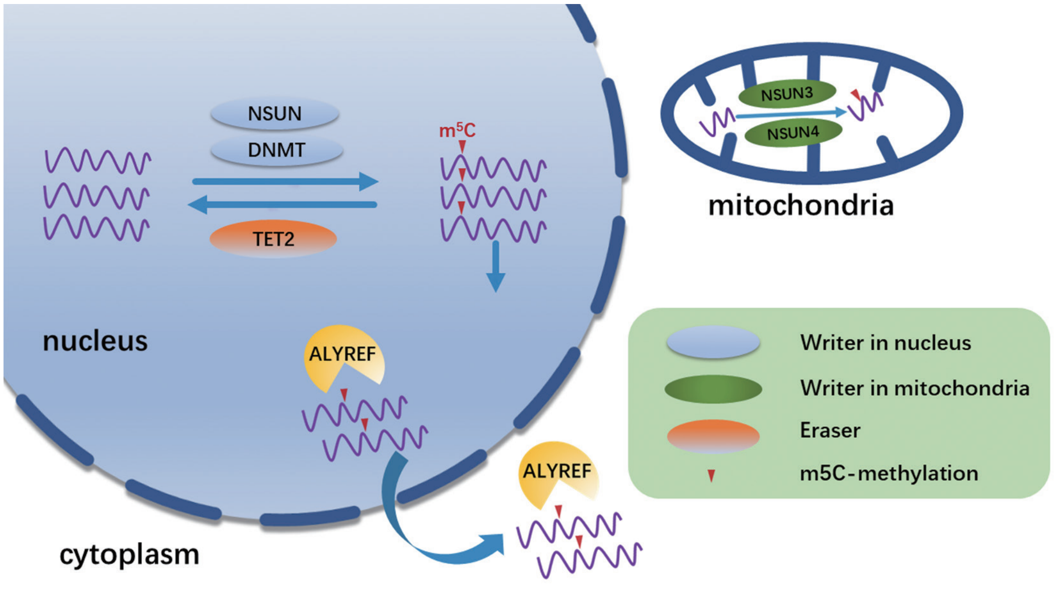


Figure S1 The process of m^{5}C RNA methylation mediated by regulators.

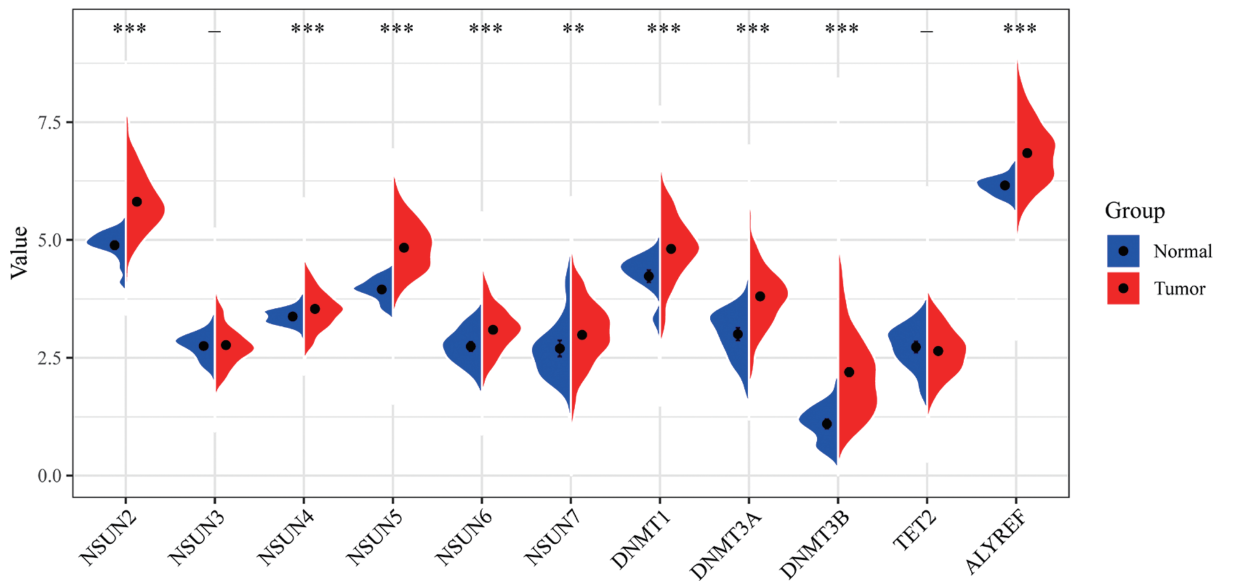


Figure S2 The expression of 11 m^{5}C regulators between normal tissues and LAUD tissues. Tumor, red; Normal, blue. **, P<0.01; ***, P<0.001. LUAD, lung adenocarcinoma.

© Translational Lung Cancer Research. All rights reserved.
http://dx.doi.org/10.21037/tlcr-21-351

Table S1 The interaction between m^5C regulators in LUAD

from	to	pvalue	cor	weight	color
NSUN3	NSUN2	9.28E-05	0.17195286	4.0324767	#C6DBEF
NSUN4	NSUN2	1.28E-09	0.2648461	8.89121472	#C6DBEF
NSUN5	NSUN2	0	0.41406303	Inf	#C6DBEF
NSUN6	NSUN2	5.62E-11	0.28530344	10.2503662	#C6DBEF
NSUN7	NSUN2	2.44E-07	0.22613398	6.61262551	#C6DBEF
NSUN6	NSUN3	2.49E-12	0.30416041	11.6043272	#C6DBEF
NSUN7	NSUN3	2.83E-07	0.22494816	6.54869035	#C6DBEF
DNMT1	NSUN3	1.45E-08	0.24775946	7.83919998	#C6DBEF
DNMT3A	NSUN3	3.74E-10	0.273108	9.42682302	#C6DBEF
NSUN6	NSUN4	8.42E-09	0.25169354	8.07484959	#C6DBEF
NSUN7	NSUN4	9.51E-14	0.32246131	13.021756	#C6DBEF
DNMT1	NSUN4	5.23E-07	0.2199184	6.28129578	#C6DBEF
DNMT3A	NSUN4	7.47E-08	0.23544234	7.12643018	#C6DBEF
NSUN7	NSUN5	9.69E-05	0.17149368	4.0135231	#C6DBEF
DNMT3A	NSUN5	8.91E-10	0.26732271	9.04990545	#C6DBEF
DNMT3B	NSUN5	2.59E-12	0.3039158	11.586107	#C6DBEF
NSUN7	NSUN6	1.88E-13	0.31876927	12.7266334	#C6DBEF
DNMT1	NSUN6	1.09E-13	0.32172987	12.9628856	#C6DBEF
DNMT3B	NSUN6	2.44E-15	0.3411941	14.6128325	#C6DBEF
TET2	NSUN6	3.09E-13	0.31602253	12.5102525	#C6DBEF
DNMT1	NSUN7	7.32E-10	0.26865004	9.13560993	#C6DBEF
DNMT3B	NSUN7	6.84E-06	0.19753793	5.16496991	#C6DBEF
TET2	NSUN7	8.11E-14	0.32331914	13.0910608	#C6DBEF
TET2	DNMT1	2.34E-13	0.31756247	12.6312402	#C6DBEF
ALYREF	DNMT3A	3.90E-15	0.33892201	14.4094214	#C6DBEF

© Translational Lung Cancer Research. All rights reserved. http://dx.doi.org/10.21037/tlcr-21-351

Table S3 The m^5C score of the 513 samples were calculated

sample	m5c_score
TCGA.05.4244.01	3941.97655
TCGA.05.4249.01	3864.35489
TCGA.05.4250.01	780.180689
TCGA.05.4382.01	408.241152
TCGA.05.4384.01	6194.83536
TCGA.05.4389.01	906.416537
TCGA.05.4390.01	1882.09467
TCGA.05.4395.01	-78.918747
TCGA.05.4396.01	2889.33874
TCGA.05.4397.01	-384.2075
TCGA.05.4398.01	458.908982
TCGA.05.4402.01	1388.67565
TCGA.05.4403.01	4677.00271
TCGA.05.4405.01	2253.90708
TCGA.05.4410.01	1925.60939
TCGA.05.4415.01	3.47116387
TCGA.05.4417.01	7887.49791
TCGA.05.4418.01	48.8789004
TCGA.05.4420.01	1233.90258
TCGA.05.4422.01	1680.37005
TCGA.05.4424.01	2435.33192
TCGA.05.4425.01	7096.72494
TCGA.05.4426.01	4604.16893
TCGA.05.4427.01	2058.8011
TCGA.05.4430.01	3944.22255
TCGA.05.4432.01	1629.18656
TCGA.05.4433.01	4034.92269
TCGA.05.4434.01	948.42533
TCGA.05.5420.01	1273.14467
TCGA.05.5423.01	10110.8607
TCGA.05.5425.01	2025.68067
TCGA.05.5428.01	968.056514
TCGA.05.5429.01	91.8937812
TCGA.05.5715.01	6023.68652
TCGA.35.3615.01	5772.89192
TCGA.35.4122.01	1394.88773
TCGA.35.4123.01	164.581805
TCGA.35.5375.01	-25.937518
TCGA.38.4625.01	1860.7779
TCGA.38.4626.01	15673.2899
TCGA.38.4627.01	1333.04341
TCGA.38.4628.01	1142.54305
TCGA.38.4629.01	277.750047
TCGA.38.4630.01	-248.92446
TCGA.38.4631.01	-185.6258
TCGA.38.4632.01	1210.92849
TCGA.38.6178.01	1393.75607
TCGA.38.7271.01	4114.25511
TCGA.38.A44F.01	2685.91889
TCGA.44.2655.01	6039.39027
TCGA.44.2656.01	4979.58626
TCGA.44.2657.01	7498.6512
TCGA.44.2659.01	3695.85733
TCGA.44.2661.01	6397.53399
TCGA.44.2662.01	218.989521
TCGA.44.2665.01	833.320462
TCGA.44.2666.01	2396.49823
TCGA.44.2668.01	62.1589986
TCGA.44.3396.01	680.046906
TCGA.44.3398.01	5017.70898
TCGA.44.3917.01	125.105017
TCGA.44.3918.01	785.803954
TCGA.44.3919.01	5345.53402
TCGA.44.4112.01	1709.63054
TCGA.44.5643.01	-24.15859
TCGA.44.5644.01	-36.283756
TCGA.44.5645.01	1522.84879
TCGA.44.6145.01	811.825087
TCGA.44.6146.01	1113.52062
TCGA.44.6147.01	3539.02999
TCGA.44.6148.01	8308.15145
TCGA.44.6774.01	633.436196
TCGA.44.6775.01	642.528978
TCGA.44.6776.01	11178.1986
TCGA.44.6777.01	11040.1782
TCGA.44.6778.01	4850.93417
TCGA.44.6779.01	11.8548335
TCGA.44.7659.01	15620.1999
TCGA.44.7660.01	-74.309889
TCGA.44.7661.01	3080.60099
TCGA.44.7662.01	-28.823859
TCGA.44.7667.01	-386.02993
TCGA.44.7669.01	1269.15387
TCGA.44.7670.01	471.317021
TCGA.44.7671.01	7869.23606
TCGA.44.7672.01	1496.78774
TCGA.44.8117.01	2196.37979
TCGA.44.8119.01	808.515243
TCGA.44.8120.01	2055.60871
TCGA.44.A479.01	527.745095
TCGA.44.A47A.01	6409.6207
TCGA.44.A47B.01	5676.50853
TCGA.44.A47G.01	2010.45743
TCGA.44.A4SS.01	99.8906326
TCGA.44.A4SU.01	2147.13062
TCGA.49.4486.01	6686.72643
TCGA.49.4487.01	803.956963
TCGA.49.4488.01	2258.78361
TCGA.49.4490.01	1086.97051
TCGA.49.4494.01	5244.53529
TCGA.49.4501.01	3584.2459
TCGA.49.4505.01	2069.54329
TCGA.49.4506.01	-65.873768
TCGA.49.4507.01	-168.40967
TCGA.49.4510.01	7550.35307
TCGA.49.4512.01	2199.71929
TCGA.49.4514.01	772.347067
TCGA.49.6742.01	2737.14383
TCGA.49.6743.01	646.904233
TCGA.49.6744.01	2969.3549
TCGA.49.6745.01	4094.74643
TCGA.49.6761.01	1499.87838
TCGA.49.6767.01	-172.76777
TCGA.49.AAQV.01	874.288009
TCGA.49.AAR0.01	5929.19762
TCGA.49.AAR2.01	734.280895
TCGA.49.AAR3.01	-63.723148
TCGA.49.AAR4.01	510.325626
TCGA.49.AAR9.01	-325.17214
TCGA.49.AARE.01	119.897344
TCGA.49.AARN.01	2976.47429
TCGA.49.AARO.01	779.585433
TCGA.49.AARQ.01	240.188431
TCGA.49.AARR.01	8811.38223
TCGA.4B.A93V.01	365.952922
TCGA.50.5044.01	2306.575
TCGA.50.5045.01	2641.72037
TCGA.50.5049.01	803.818844
TCGA.50.5051.01	3784.26278
TCGA.50.5055.01	2075.16394
TCGA.50.5066.01	510.999804
TCGA.50.5068.01	2067.55375
TCGA.50.5072.01	1616.24525
TCGA.50.5930.01	1007.27883
TCGA.50.5931.01	-176.91232
TCGA.50.5932.01	5674.32865
TCGA.50.5933.01	333.38433
TCGA.50.5935.01	8573.13062
TCGA.50.5936.01	1466.84922
TCGA.50.5939.01	1327.23609
TCGA.50.5941.01	2797.25124
TCGA.50.5942.01	14411.9263
TCGA.50.5944.01	4840.1935
TCGA.50.5946.01	248.379049
TCGA.50.6590.01	-177.00065
TCGA.50.6591.01	-311.30397
TCGA.50.6592.01	-85.88019
TCGA.50.6593.01	1882.88459
TCGA.50.6594.01	327.85932
TCGA.50.6595.01	-58.483418
TCGA.50.6597.01	1878.49545
TCGA.50.6673.01	1710.46671
TCGA.50.7109.01	1287.46908
TCGA.50.8457.01	5306.97234
TCGA.50.8459.01	610.059706
TCGA.50.8460.01	8530.18193
TCGA.53.7624.01	-497.95908
TCGA.53.7626.01	4107.73733
TCGA.53.7813.01	1014.68358
TCGA.53.A4EZ.01	213.182959
TCGA.55.1592.01	2871.48627
TCGA.55.1594.01	948.352445
TCGA.55.1596.01	124.606326
TCGA.55.5899.01	-129.24002
TCGA.55.6543.01	8647.51353
TCGA.55.6642.01	1096.91236
TCGA.55.6712.01	687.16129
TCGA.55.6968.01	-284.00356
TCGA.55.6970.01	2908.14124
TCGA.55.6971.01	3994.23867
TCGA.55.6972.01	2670.26515
TCGA.55.6975.01	-122.5366
TCGA.55.6978.01	67.4422753
TCGA.55.6979.01	1304.91729
TCGA.55.6980.01	1273.91531
TCGA.55.6981.01	957.909388
TCGA.55.6982.01	2369.70059
TCGA.55.6983.01	4033.34272
TCGA.55.6984.01	436.831688
TCGA.55.6985.01	2041.85773
TCGA.55.6986.01	5284.73482
TCGA.55.6987.01	1337.57188
TCGA.55.7227.01	3414.21204
TCGA.55.7281.01	3451.58231
TCGA.55.7283.01	10332.1132
TCGA.55.7284.01	3460.00451
TCGA.55.7570.01	0.468144
TCGA.55.7573.01	7140.65284
TCGA.55.7574.01	5007.26717
TCGA.55.7576.01	1897.21928
TCGA.55.7724.01	3553.33491
TCGA.55.7725.01	11011.6991
TCGA.55.7726.01	-30.594751
TCGA.55.7727.01	5350.54249
TCGA.55.7728.01	1292.97041
TCGA.55.7815.01	6774.97058
TCGA.55.7816.01	1333.09921
TCGA.55.7903.01	3164.83114
TCGA.55.7907.01	790.515828
TCGA.55.7910.01	2016.05007
TCGA.55.7911.01	2781.60741
TCGA.55.7913.01	336.617338
TCGA.55.7914.01	2736.29994
TCGA.55.7994.01	4370.80727
TCGA.55.7995.01	240.807393
TCGA.55.8085.01	1103.13144
TCGA.55.8087.01	6342.09816
TCGA.55.8089.01	1434.18662
TCGA.55.8090.01	5575.85793
TCGA.55.8091.01	2777.15685
TCGA.55.8092.01	56.1820804
TCGA.55.8094.01	382.848887
TCGA.55.8096.01	636.233868
TCGA.55.8097.01	5570.33088
TCGA.55.8203.01	3455.65341
TCGA.55.8204.01	1145.36317
TCGA.55.8205.01	973.891583
TCGA.55.8206.01	5932.61793
TCGA.55.8207.01	9616.98676
TCGA.55.8208.01	1348.43274
TCGA.55.8299.01	3151.84234
TCGA.55.8301.01	1924.14818
TCGA.55.8302.01	1708.37974
TCGA.55.8505.01	34.5646559
TCGA.55.8506.01	7378.60222
TCGA.55.8507.01	5409.28701
TCGA.55.8508.01	544.46915
TCGA.55.8510.01	2620.25165
TCGA.55.8511.01	629.908608
TCGA.55.8512.01	6390.1738
TCGA.55.8513.01	7373.21603
TCGA.55.8514.01	2909.76716
TCGA.55.8614.01	2231.00147
TCGA.55.8615.01	3861.8192
TCGA.55.8616.01	2182.1482
TCGA.55.8619.01	6328.89602
TCGA.55.8620.01	78.5053468
TCGA.55.8621.01	10690.3714
TCGA.55.A48X.01	3844.10261
TCGA.55.A48Y.01	1369.29757
TCGA.55.A48Z.01	511.648877
TCGA.55.A490.01	557.5491
TCGA.55.A491.01	414.612837
TCGA.55.A492.01	9409.22737
TCGA.55.A493.01	-120.58349
TCGA.55.A494.01	2042.75923
TCGA.55.A4DF.01	262.731581
TCGA.55.A4DG.01	3730.78097
TCGA.55.A57B.01	2623.26357
TCGA.62.8394.01	1084.14876
TCGA.62.8395.01	4302.64847
TCGA.62.8397.01	2883.87329
TCGA.62.8398.01	1343.49955
TCGA.62.8399.01	1518.85015
TCGA.62.8402.01	83.1897684
TCGA.62.A46O.01	-240.76651
TCGA.62.A46P.01	3699.56308
TCGA.62.A46R.01	1911.43357
TCGA.62.A46S.01	3885.23984
TCGA.62.A46V.01	6685.38794
TCGA.62.A46Y.01	9689.70574
TCGA.62.A470.01	3462.75257
TCGA.62.A471.01	-34.262579
TCGA.62.A472.01	3482.64014
TCGA.64.1676.01	815.625505
TCGA.64.1677.01	696.630749
TCGA.64.1678.01	264.627684
TCGA.64.1679.01	751.605918
TCGA.64.1680.01	9742.51932
TCGA.64.1681.01	4681.75416
TCGA.64.5774.01	1572.69703
TCGA.64.5775.01	-149.83565
TCGA.64.5778.01	2273.97207
TCGA.64.5779.01	321.745939
TCGA.64.5781.01	-22.462696
TCGA.64.5815.01	2022.8242
TCGA.67.3770.01	8060.66359
TCGA.67.3771.01	218.369674
TCGA.67.3772.01	3367.31601
TCGA.67.3773.01	15764.6667
TCGA.67.3774.01	6867.79493
TCGA.67.4679.01	6896.33944
TCGA.67.6215.01	9543.86548
TCGA.67.6216.01	8898.23214
TCGA.67.6217.01	8255.95106
TCGA.69.7760.01	1130.56835
TCGA.69.7761.01	67.6740245
TCGA.69.7763.01	5582.39187
TCGA.69.7764.01	7415.81541
TCGA.69.7765.01	4312.50551
TCGA.69.7973.01	3037.42801
TCGA.69.7974.01	5856.92515
TCGA.69.7978.01	1376.99064
TCGA.69.7979.01	288.881739
TCGA.69.7980.01	1059.66228
TCGA.69.8253.01	8747.79233
TCGA.69.8254.01	7178.12397
TCGA.69.8255.01	491.527978
TCGA.69.8453.01	8217.42203
TCGA.69.A59K.01	2598.28375
TCGA.71.6725.01	4899.81336
TCGA.71.8520.01	2989.50554
TCGA.73.4658.01	2830.13357
TCGA.73.4659.01	6116.0832
TCGA.73.4662.01	7508.71538
TCGA.73.4666.01	378.937035
TCGA.73.4668.01	1404.72074
TCGA.73.4670.01	-47.325392
TCGA.73.4675.01	5610.07846
TCGA.73.4676.01	1495.81014
TCGA.73.4677.01	8315.46306
TCGA.73.7498.01	7643.79589
TCGA.73.7499.01	1336.87595
TCGA.73.A9RS.01	-72.683826
TCGA.75.5122.01	984.910677
TCGA.75.5125.01	361.266645
TCGA.75.5126.01	3542.19964
TCGA.75.5146.01	6933.144
TCGA.75.5147.01	736.05185
TCGA.75.6203.01	14352.2091
TCGA.75.6205.01	777.337611
TCGA.75.6206.01	5763.73223
TCGA.75.6207.01	1944.79193
TCGA.75.6211.01	150.644921
TCGA.75.6212.01	6374.17849
TCGA.75.6214.01	39.3784336
TCGA.75.7025.01	3994.07828
TCGA.75.7027.01	-51.705756
TCGA.75.7030.01	3448.25891
TCGA.75.7031.01	2259.90469
TCGA.78.7143.01	3355.74677
TCGA.78.7145.01	1580.63016
TCGA.78.7146.01	-169.78491
TCGA.78.7147.01	1748.27939
TCGA.78.7148.01	4289.46764
TCGA.78.7149.01	3784.49351
TCGA.78.7150.01	-199.04474
TCGA.78.7152.01	7295.47829
TCGA.78.7153.01	2925.94999
TCGA.78.7154.01	558.557041
TCGA.78.7155.01	-571.94825
TCGA.78.7156.01	3824.31329
TCGA.78.7158.01	5062.13942
TCGA.78.7159.01	1863.59035
TCGA.78.7160.01	2423.9348
TCGA.78.7161.01	4347.55764
TCGA.78.7162.01	15147.9379
TCGA.78.7163.01	3302.40239
TCGA.78.7166.01	54.2330105
TCGA.78.7167.01	8131.52952
TCGA.78.7220.01	281.983709
TCGA.78.7535.01	2344.21618
TCGA.78.7536.01	-12.650543
TCGA.78.7537.01	7275.7536
TCGA.78.7539.01	2509.16458
TCGA.78.7540.01	2094.71261
TCGA.78.7542.01	768.102809
TCGA.78.7633.01	7505.80554
TCGA.78.8640.01	1374.89241
TCGA.78.8648.01	3097.54838
TCGA.78.8655.01	4683.62004
TCGA.78.8660.01	3097.08477
TCGA.78.8662.01	265.527203
TCGA.80.5607.01	4805.5455
TCGA.80.5608.01	4874.05928
TCGA.80.5611.01	139.826883
TCGA.83.5908.01	-152.3545
TCGA.86.6562.01	2405.21772
TCGA.86.6851.01	608.559902
TCGA.86.7701.01	1468.64937
TCGA.86.7711.01	-83.308441
TCGA.86.7713.01	2746.02499
TCGA.86.7714.01	8417.53842
TCGA.86.7953.01	312.495943
TCGA.86.7954.01	1431.8526
TCGA.86.7955.01	-24.53177
TCGA.86.8054.01	1459.79006
TCGA.86.8055.01	703.795171
TCGA.86.8056.01	2707.20992
TCGA.86.8073.01	1820.6947
TCGA.86.8074.01	721.745019
TCGA.86.8075.01	997.084962
TCGA.86.8076.01	3016.27455
TCGA.86.8278.01	3112.79512
TCGA.86.8279.01	2549.79462
TCGA.86.8280.01	6147.09958
TCGA.86.8281.01	5137.09477
TCGA.86.8358.01	-480.10508
TCGA.86.8359.01	1775.33285
TCGA.86.8585.01	732.878214
TCGA.86.8668.01	6231.01438
TCGA.86.8669.01	5108.58979
TCGA.86.8671.01	6599.39691
TCGA.86.8672.01	152.40442
TCGA.86.8673.01	1397.97339
TCGA.86.8674.01	4824.02292
TCGA.86.A456.01	6216.4754
TCGA.86.A4D0.01	133.140751
TCGA.86.A4JF.01	6203.16743
TCGA.86.A4P7.01	5606.57966
TCGA.86.A4P8.01	10724.8689
TCGA.91.6828.01	5598.10903
TCGA.91.6829.01	1045.63051
TCGA.91.6830.01	2361.92228
TCGA.91.6831.01	107.061719
TCGA.91.6835.01	3206.56102
TCGA.91.6836.01	296.00954
TCGA.91.6840.01	1015.37428
TCGA.91.6847.01	-30.902848
TCGA.91.6848.01	-196.34632
TCGA.91.6849.01	5662.67586
TCGA.91.7771.01	3686.44108
TCGA.91.8496.01	20932.5705
TCGA.91.8497.01	6366.14438
TCGA.91.8499.01	-150.78657
TCGA.91.A4BC.01	67.4911041
TCGA.91.A4BD.01	21855.3423
TCGA.93.7347.01	7969.17601
TCGA.93.7348.01	2258.05625
TCGA.93.8067.01	1231.45311
TCGA.93.A4JN.01	1636.47001
TCGA.93.A4JO.01	1764.46487
TCGA.93.A4JP.01	8187.09823
TCGA.93.A4JQ.01	1293.88732
TCGA.95.7039.01	1859.36071
TCGA.95.7043.01	1017.84492
TCGA.95.7562.01	445.27342
TCGA.95.7567.01	3318.84507
TCGA.95.7944.01	-69.836201
TCGA.95.7947.01	1118.72579
TCGA.95.7948.01	1247.65073
TCGA.95.8039.01	5323.50266
TCGA.95.8494.01	3112.07375
TCGA.95.A4VK.01	1102.2972
TCGA.95.A4VN.01	2399.3251
TCGA.95.A4VP.01	5162.33166
TCGA.97.7546.01	6268.25383
TCGA.97.7547.01	9610.70165
TCGA.97.7552.01	4091.63524
TCGA.97.7553.01	11230.7849
TCGA.97.7554.01	3182.86711
TCGA.97.7937.01	1296.19984
TCGA.97.7938.01	5700.71881
TCGA.97.7941.01	3795.66732
TCGA.97.8171.01	5180.62486
TCGA.97.8172.01	5540.18914
TCGA.97.8174.01	7629.5792
TCGA.97.8175.01	4018.94818
TCGA.97.8176.01	919.307488
TCGA.97.8177.01	4849.9011
TCGA.97.8179.01	4085.2693
TCGA.97.8547.01	3656.00198
TCGA.97.8552.01	13127.9162
TCGA.97.A4LX.01	3083.63997
TCGA.97.A4M0.01	8524.74047
TCGA.97.A4M1.01	9236.46086
TCGA.97.A4M2.01	7078.13039
TCGA.97.A4M3.01	6230.47237
TCGA.97.A4M5.01	4209.779
TCGA.97.A4M6.01	1981.90008
TCGA.97.A4M7.01	3402.69355
TCGA.99.7458.01	6206.84548
TCGA.99.8025.01	1310.8076
TCGA.99.8028.01	2264.62083
TCGA.99.8032.01	2298.71376
TCGA.99.8033.01	581.140177
TCGA.99.AA5R.01	7885.08108
TCGA.J2.8192.01	1968.0161
TCGA.J2.8194.01	4132.25907
TCGA.J2.A4AD.01	423.615682
TCGA.J2.A4AE.01	6563.84342
TCGA.J2.A4AG.01	3070.13575
TCGA.L4.A4E5.01	556.884176
TCGA.L4.A4E6.01	11111.0219
TCGA.L9.A443.01	4273.07397
TCGA.L9.A444.01	2975.93399
TCGA.L9.A50W.01	3136.12251
TCGA.L9.A5IP.01	-69.975676
TCGA.L9.A743.01	4003.57802
TCGA.L9.A7SV.01	431.657386
TCGA.L9.A8F4.01	248.401772
TCGA.MN.A4N1.01	412.498166
TCGA.MN.A4N4.01	259.464807
TCGA.MN.A4N5.01	3571.3851
TCGA.MP.A4SV.01	1488.13415
TCGA.MP.A4SW.01	6382.63229
TCGA.MP.A4SY.01	1176.90218
TCGA.MP.A4T4.01	2969.98648
TCGA.MP.A4T6.01	2721.88549
TCGA.MP.A4T7.01	4774.05022
TCGA.MP.A4T8.01	962.91559
TCGA.MP.A4T9.01	4901.60136
TCGA.MP.A4TA.01	-106.36706
TCGA.MP.A4TC.01	1555.11372
TCGA.MP.A4TD.01	2745.62194
TCGA.MP.A4TE.01	2812.01692
TCGA.MP.A4TF.01	1458.70439
TCGA.MP.A4TH.01	5492.45622
TCGA.MP.A4TI.01	1367.11092
TCGA.MP.A4TJ.01	3840.87619
TCGA.MP.A4TK.01	742.070263
TCGA.MP.A5C7.01	3309.56017
TCGA.NJ.A4YF.01	314.500561
TCGA.NJ.A4YG.01	8163.23167
TCGA.NJ.A4YI.01	4201.83162
TCGA.NJ.A4YP.01	2390.92585
TCGA.NJ.A4YQ.01	576.724724
TCGA.NJ.A55A.01	8504.42135
TCGA.NJ.A55O.01	7401.38172
TCGA.NJ.A55R.01	1466.17695
TCGA.NJ.A7XG.01	14648.501
TCGA.O1.A52J.01	8902.8433
TCGA.S2.AA1A.01	1106.87712

Table S3 (*continued*)

© Translational Lung Cancer Research. All rights reserved.

Table S2 The univariate COX analysis identified 124 genes related to prognosis

Gene symbol	p.value	HR	Low 95% CI	High 95% CI
ESPL1	0.0125163	1.0237293	1.00505846	1.04274699
CDCA3	0.00150631	1.03454726	1.01307215	1.05647759
SPAG5	0.01602571	1.00681244	1.00126479	1.01239083
EME1	0.0318387	1.03913299	1.00334032	1.0762025
CDC45	0.02448267	1.00925326	1.00118612	1.01738541
CENPA	0.02959417	1.01104594	1.00108887	1.02110204
SPC24	0.02790471	1.0115681	1.00124928	1.02199327
TPX2	0.00047155	1.00307183	1.00134871	1.00479791
TEDC2	0.03274375	1.02411384	1.00195757	1.04676006
NCAPH	0.0059505	1.01045558	1.00299396	1.01797271
RAD54L	0.04190695	1.02024865	1.00073415	1.04014369
KIF2C	0.00300378	1.00715765	1.00242423	1.01191342
GTSE1	0.00093638	1.02644023	1.01069597	1.04242974
KIFC1	0.03417985	1.00428373	1.0003189	1.00826427
MCM10	0.0228088	1.02028659	1.00279735	1.03808085
HJURP	0.00021403	1.01547163	1.00725053	1.02375983
BUB1B	0.00395048	1.01394948	1.00444213	1.02354681
FOXM1	0.00017214	1.00794722	1.00379315	1.01211848
EXO1	1.24E-05	1.02360496	1.01294926	1.03437274
KIF4A	5.73E-05	1.01800259	1.00919284	1.02688924
CDCA5	0.00011692	1.0119799	1.00586729	1.01812966
PKMYT1	0.00518002	1.02764177	1.00818374	1.04747533
KIF23	0.00367715	1.01544613	1.00499747	1.02600342
CDT1	0.0376879	1.00682024	1.00038689	1.01329496
ZWINT	0.00898083	1.00407562	1.00101676	1.00714382
CDC20	0.00118778	1.00330745	1.00130649	1.0053124
CDCA8	0.00790741	1.0058386	1.00152679	1.01016897
PLK1	1.54E-05	1.01490283	1.00811969	1.02173161
CENPF	0.0023956	1.01103059	1.00389622	1.01821566
SKA1	0.00256932	1.02190639	1.00761201	1.03640356
KIF14	0.0013409	1.03399369	1.01308404	1.05533492
CDC6	0.01283523	1.00826499	1.00174923	1.01482314
BUB1	0.00314237	1.01183703	1.00396624	1.01976953
ASPM	0.0129994	1.01586051	1.00332419	1.02855347
CDCA2	0.00031497	1.04116396	1.01856364	1.06426575
AURKA	0.00024554	1.00829872	1.00385447	1.01276265
NDC80	0.00041249	1.01416952	1.00628192	1.02211895
ORC1	0.00096886	1.02422875	1.00976617	1.03889848
SKA3	0.00010485	1.03137047	1.01539895	1.04759322
KIF20A	4.17E-05	1.01507392	1.00783579	1.02236403
NEK2	4.29E-05	1.0148936	1.00773163	1.02210647
DLGAP5	2.97E-05	1.01251719	1.00662298	1.01844591
NCAPG	0.00202128	1.01504701	1.00546777	1.02471751
TTK	0.00692508	1.01952686	1.00531678	1.0339378
PRC1	2.94E-05	1.0134835	1.00713588	1.01987113
BIRC5	0.00111196	1.00637487	1.00253799	1.01022643
KPNA2	2.03E-05	1.00291817	1.00157494	1.00426319
UBE2C	0.00864351	1.00151884	1.00038494	1.00265402
MELK	0.00173031	1.01093098	1.00407879	1.01782994
CDKN3	0.00030399	1.0132551	1.00604063	1.02052129
CCNB2	0.01429593	1.00504572	1.00100676	1.00910098
PIMREG	0.01756341	1.01424803	1.00247386	1.02616049
RAD51AP1	0.02602186	1.01091354	1.00129759	1.02062184
KIF11	0.00200755	1.0109095	1.00397398	1.01789294
UBE2S	0.00019974	1.00959197	1.00452593	1.01468356
NUSAP1	0.00426708	1.00562704	1.00176433	1.00950465
PRR11	0.0003022	1.01722318	1.0078446	1.02668904
FAM83D	0.00987526	1.00879844	1.00210788	1.01553366
DEPDC1	0.0013399	1.025639	1.00989406	1.04162942
CKAP2L	0.00022282	1.03144537	1.01462919	1.04854026
MKI67	0.00055306	1.01028382	1.00443479	1.01616691
CCNA2	0.00012613	1.00860177	1.00419516	1.01302772
UHRF1	0.0047405	1.01761467	1.00535748	1.0300213
TRIP13	0.03187744	1.00662893	1.00057277	1.01272175
SAPCD2	0.01463769	1.01212271	1.0023783	1.02196185
UBE2T	0.00520567	1.00366397	1.00109235	1.00624219
CENPW	0.04776129	1.00403828	1.00003975	1.0080528
PBK	0.00179879	1.01028727	1.00381586	1.01680039
CCNB1	1.19E-05	1.00523804	1.00288997	1.00759161
ADH1B	0.01413662	0.99269281	0.98689452	0.99852516
CEP55	0.00215438	1.00839835	1.00302518	1.0138003
MGP	0.01459993	0.99908722	0.99835527	0.99981971
INMT	0.04421502	0.994163	0.98850985	0.99984848
CDK1	0.00410337	1.00628174	1.00198775	1.01059413
SCN7A	0.00515894	0.96309514	0.93804856	0.98881048
RRM2	2.67E-05	1.00710052	1.00378063	1.01043138
C2orf48	2.67E-05	1.00710052	1.00378063	1.01043138
C16orf89	0.04432789	0.99968632	0.9993807	0.99999204
MAMDC2	0.01224184	0.97658138	0.95864191	0.99485655
CTSV	0.01386895	1.00555839	1.00112849	1.01000789
CD1C	0.04223957	0.98937815	0.97923554	0.9996258
ANLN	3.03E-09	1.01044691	1.00698197	1.01392376
IGF2BP3	0.00343338	1.01570465	1.00515799	1.02636198
TK1	4.09E-05	1.00292052	1.0015242	1.0043188
NAPSA	0.01815889	0.99986083	0.99974539	0.99997629
C7	0.01391943	0.99654169	0.99379415	0.99929681
ADGRF5	0.01790826	0.99838104	0.99704275	0.99972112
CFAP221	0.03493106	0.97014044	0.9431928	0.99785798
FMO2	0.00474838	0.9831934	0.9716938	0.9948291
IL33	0.01847859	0.99198264	0.98536166	0.99864812
SLC22A3	0.0415354	0.99532187	0.99084383	0.99982015
MFAP4	0.00590622	0.99816589	0.99686213	0.99947134
CD1E	0.03216115	0.97742045	0.95721054	0.99805707
CPA3	0.00940746	0.99499014	0.99122563	0.99876894
SLC7A5	1.01E-05	1.00244697	1.0013597	1.00353543
SLC34A2	0.00100296	0.99977288	0.99963758	0.99990819
RNASE1	0.02591289	0.99985683	0.99973087	0.9999828
SFTPB	0.03298292	0.99997984	0.99996131	0.99999837
IGF2BP1	9.18E-05	1.0143412	1.0071291	1.02160494
AQP4	0.03423508	0.99726693	0.9947436	0.99979666
SCGB3A1	0.01486137	0.99989419	0.99980905	0.99997933
NDNF	0.0118608	0.99712582	0.99489278	0.99936386
CYP4B1	0.00411154	0.99828911	0.99712219	0.9994574
SFTPA1	0.01709152	0.99995813	0.99992372	0.99999254
GGTLC1	0.02685584	0.99684876	0.99406689	0.99963842
SUSD2	0.00979165	0.99853924	0.99743229	0.99964742
SFTPA2	0.01275597	0.99996403	0.99993573	0.99999234
PHGDH	0.03164422	1.00532661	1.00046714	1.01020969
CHRDL1	0.01332615	0.98839001	0.97929148	0.99757309
FOXA2	0.0372601	0.99570486	0.99168008	0.99974598
SCNN1B	0.04453229	0.99670789	0.99350666	0.99991944
VEGFD	0.0011589	0.96626929	0.9464727	0.98647994
ATP13A4	0.01808718	0.99199442	0.98540523	0.99862767
CRABP1	0.00300559	1.00089306	1.00030306	1.00148342
HAGLR	0.01819089	0.99391015	0.98888439	0.99896144
TPSB2	0.02526535	0.99442618	0.98956887	0.99930733
SFTA3	0.00352533	0.99470854	0.99116983	0.99825988
SFTPC	0.00930057	0.9998637	0.999761	0.99996642
CACNA2D2	0.01739149	0.99554816	0.99189432	0.99921546
SCGB1A1	0.01291537	0.99978894	0.99962258	0.99995533
CRTAC1	0.04567201	0.99445741	0.98905129	0.99989307
AGER	0.03024055	0.9988547	0.99781991	0.99989056
IRX2	0.0089347	0.99375213	0.98909409	0.9984321

© Translational Lung Cancer Research. All rights reserved. http://dx.doi.org/10.21037/tlcr-21-351

Table S4 Enrichment results of the high and low m^5C score groups were analysed

KEGG	subtype
KEGG_HOMOLOGOUS_RECOMBINATION	C1
KEGG_BASE_EXCISION_REPAIR	C1
KEGG_SPLICEOSOME	C1
KEGG_DNA_REPLICATION	C1
KEGG_CELL_CYCLE	C1
KEGG_MISMATCH_REPAIR	C1
KEGG_PROTEASOME	C1
KEGG_NUCLEOTIDE_EXCISION_REPAIR	C1
KEGG_RNA_DEGRADATION	C1
KEGG_PYRIMIDINE_METABOLISM	C1
KEGG_GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM	C1
KEGG_BASAL_TRANSCRIPTION_FACTORS	C1
KEGG_CYSTEINE_AND_METHIONINE_METABOLISM	C1
KEGG_ONE_CARBON_POOL_BY_FOLATE	C1
KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_GANGLIO_SERIES	C2
KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	C2
KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS	C2
KEGG_VASOPRESSIN_REGULATED_WATER_REABSORPTION	C2
KEGG_PPAR_SIGNALING_PATHWAY	C2
KEGG_RENIN_ANGIOTENSIN_SYSTEM	C2
KEGG_HEDGEHOG_SIGNALING_PATHWAY	C2
KEGG_LONG_TERM_DEPRESSION	C2
KEGG_ARACHIDONIC_ACID_METABOLISM	C2
KEGG_GLYCOSAMINOGLYCAN_DEGRADATION	C2
KEGG_APOPTOSIS	C2
KEGG_ASTHMA	C2
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	C2
KEGG_LYSOSOME	C2
KEGG_O_GLYCAN_BIOSYNTHESIS	C2
KEGG_NITROGEN_METABOLISM	C2
KEGG_PROXIMAL_TUBULE_BICARBONATE_RECLAMATION	C2
KEGG_FATTY_ACID_METABOLISM	C2
KEGG_ETHER_LIPID_METABOLISM	C2
KEGG_CELL_ADHESION_MOLECULES_CAMS	C2
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	C2
KEGG_MAPK_SIGNALING_PATHWAY	C2
KEGG_HEMATOPOIETIC_CELL_LINEAGE	C2
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	C2
KEGG_CALCIUM_SIGNALING_PATHWAY	C2

© Translational Lung Cancer Research. All rights reserved.
http://dx.doi.org/10.21037/tlcr-21-351