Gene Signature And Prognostic Values of m^5C-Related Regulators In Colon Adenocarcinoma

Yuancheng Huang
Guangzhou University of Chinese Medicine
Chaoyuan Huang
Guangzhou University of Chinese Medicine
Xiaotao Jiang
Guangzhou University of Chinese Medicine
Yanhua Yan
Guangzhou University of Chinese Medicine
Kunhai Zhuang
Guangzhou University of Chinese Medicine
Fengbin Liu
Guangzhou University of Chinese Medicine
Peiwu Li
Guangzhou University of Chinese Medicine
Yi Wen (doctorwenyigzucm@163.com)
Guangzhou University of Traditional Chinese Medicine First Affiliated Hospital

Research

Keywords: colon adenocarcinoma, epigenetics, RNA modification, 5-methylcytosine, gene expression profile, prognostic signature, TCGA.

Posted Date: October 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-952509/v1

Abstract

Objectives: The purpose of this study was to investigate the role of 13 m^5C-related regulators in colon adenocarcinoma (COAD) and determine their prognostic value.

Main Methods: Gene expression and clinicopathological data were obtained from The Cancer Genome Atlas (TCGA) datasets. The expression of m^5C-related regulators were analyzed with clinicopathological characteristics and alterations within m^5C-related regulators. Subsequently, different subtypes of patients with COAD were identified. Then, the prognostic value of m^5C-related regulators in COAD were confirmed via univariate Cox regression and least absolute shrinkage and selection operator (LASSO) Cox regression analyses. The prognostic value of risk scores was evaluated using the Kaplan-Meier method, receiver operating characteristic (ROC) curves, and univariate and multivariate regression analyses. Additionally, Gene Set Enrichment Analysisc (GSEA), Kyoto Encyclopedia of Genes and Genomes c (KEGG) pathways, and Gene Ontologyc (GO) analysis were performed for biological functional analysis.

Results: m^5C-related regulators were found to be differentially expressed in COAD with different clinicopathological features. We observed a high alteration frequency in these genes, which were significantly correlated with their mRNA expression levels. Two clusters with different prognostic features were identified. Based on two independent prognostic m^5C-related regulators (NSUN6 and ALYREF), a risk signature with good predictive significance was constructed. Univariate and multivariate Cox regression analyses suggested that the risk score was an independent prognostic factor. Biological processes and pathways associated with cancer, immune response, and RNA processing were identified.

Conclusion: We revealed the genetic signatures and prognostic values of m^5C-related regulators in COAD. Together, this has improved our understanding of m^5C RNA modification and provided novel insights to identify predictive biomarkers and develop molecular targeted therapy for COAD.

Introduction

Changes in gene expression are closely associated with the development of disease, and epigenetic processes are heritable changes in gene expression that do not alter the nucleotide sequence [1]. Traditional epigenetic modifications, including chromatin remodeling, DNA methylation, and histone modifications, are involved in various biological processes related to the occurrence and progression of tumors, including gastrointestinal cancers [2–4]. With considerable progress in zymology and high-throughput sequencing technology, epitranscriptomics has attracted significant attention recently [5–10]. Research investigating the physiological and pathological functions of RNA modifications have identified multiple dynamic modifications of RNA, including N6-methyladenosine (m^6A), 2-O-dimethyladenosine (m^6Am), 5-methylcytosine (m^5C), 7-methylguanosine (m^7G), N1-methyladenosine (m^1A), and pseudouridylation (Ψ) [11, 12]. Increasing evidence suggests that RNA modifications play critical roles in tumorigenesis and the progression of different cancers [13, 14]. m^5C RNA modification is found in a variety of RNAs, including messenger RNAs (mRNAs), transfer RNAs (tRNAs), ribosomal RNAs(rRNAs). This modification introduces a methyl group in the 5th carbon atom of cytosine [15]. Based on published data, m^5C RNA modification plays a critical role in the translation, transport, and stability of mRNAs, and is also closely associated with the biogenesis and function of other RNA species [16, 17]. As a dynamic and reversible process, m^5C RNA modification is primarily regulated by “writers” (adenosine

methyltransferases) and “erasers” (demethylases), and achieves different functions by interacting with “readers” (m^5C-binding proteins). The “writers” include the NOL1/NOP2/sun domain RNA methyltransferase family NSUN1-NSUN7 and DNMT2. m^5c “erasers” include enzymes in the TET family (TET1,TET2,TET3) and ALKBH1. The “readers”, such as ALYREF and YBX1, recognize and bind to methylated RNAs to realize different functions [18, 19].

Globally, colorectal cancer (CRC) is the third most common cancer and the second most deadly neoplasm [20]. Colon adenocarcinoma (COAD) is the most common pathological type of CRC, and despite considerable progress in diagnosis and therapeutic strategies for COAD, the prognosis of patients with COAD remains poor due to advanced stage and postsurgical recurrence [21, 22]. Therefore, identification of novel biomarkers for early detection and effective therapeutic targets for treating patients with COAD is critical and urgent.

In this study, we analyzed a TCGA dataset for m^5C-related regulators involved in COAD, the correlation between the expression levels of 13 m^5C-related regulators and clinicopathological features, as well as potential independent prognostic m^5C-related regulators and a risk signature to predict the prognosis of patients with COAD.

Material And Methods

Acquisition of datasets

The RNA-seq transcriptome data (fragments per kilobase million, FPKM) from 437 samples [23], copy number variant (CNV) data from 825 samples, single nucleotide variant (SNV) data from 399 samples, and clinical information from 385 patients with COAD in TCGA database (http://cancergenome.nih.gov/) were downloaded for our study. Patients with complete clinicopathological and survival information were included for further assessment (**Table 1**).

Selection of m^5C-related regulators

Based on published data,14 m^5C-related regulators, including NOP2(NSUN1), NSUN2, NSUN3, NSUN4, NSUN5, NSUN6, NSUN7, DNMT2, TET1, TET2, TET3, ALKBH1, ALYREF, and YBX1 were used in our study. DNMT2 was not found to be expressed in COAD from TCGA datasets. Therefore, the remaining 13 m^5C-related regulators were used for further analysis.

Tumor classification and principal component analysis

To explore the function of m^5C-related regulators in COAD, a consistent clustering algorithm was used to determine the clustering of samples and estimate the stability of the clustering. Using the “Consensus ClusterPlus” R package [24], two different subgroups (cluster □ and cluster □) were identified based on the following classification parameters: 1) slow growth rate of the cumulative distribution function value; 2) high correlation in the subgroup; and 3) no small clusters in the clustering data. Furthermore, principal component analysis (PCA) was used to assess gene expression patterns in different subgroups using the “Limma” R package [25].

Analysis of clinicopathological features and prognosis

The correlation between m^5C-related regulators and clinicopathological features was analyzed. Then, to filter the m^5C-related regulators that were highly correlated with overall survival (OS), univariate Cox regression analysis was performed. Next, the Lasso Cox regression algorithm was used to identify m^5C-related regulators with powerful prognostic significance. According to the best penalty parameter λ, the selected regulators' coefficients were calculated. The risk score (RS) was estimated using the following formula:

$$RS = \sum_{i=1}^{n} Coef(i)X(i)$$

where Coef(i) is the coefficient and X(i) represents the expression levels of the selected m^5C-related regulators. Using the obtained median risk score as the demarcation value, patients with COAD were classified in two groups: high-risk group and low-risk group. The OS and clinicopathological features were compared between these subgroups. Kaplan-Meier analysis and the receiver operating characteristic (ROC) curves were used to validate the predictive efficiency [26]. Additionally, the prognostic value of the RS was verified using univariate and multivariate Cox regression analyses. The hazard ratio (HR) with 95% confidence intervals and log-rank *P*-value were calculated using the "glmnet" and "survival" R packages [27].

Biological function analysis

To explore the biological functions associated with m^5C RNA modification, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, Gene Ontology (GO) analysis and Gene Set Enrichment Analysis (GSEA) were performed.The genes that were differentially expressed between high-risk group and low-risk group were functionally annotated using GO analysis and KEGG pathway analysis. Next, GSEA was conducted to determine the signaling pathways related to different clusters. Later, to explore the latent biological function of the m^5C-related genes in COAD, GSEA for the m^5C-related regulatory genes with powerful prognostic value was performed. The flow chart of bioinformatic analysis was shown in **Figure 1**.

Statistical analysis

The expression data of m^5C-related regulators in tumor tissues and adjacent mucosa of COAD obtained from TCGA were compared using one-way analysis of variance (ANOVA); the clinical characteristics and m^5C-related regulators of different groups were compared using the chi-square test; the Kaplan-Meier method was used to perform a bilateral logarithmic rank test in overall survival analysis; *P*-values < 0.05 were regarded as statistically significant. All statistical analyses were implemented using Rv4.0.3 (https://www.r-project.org/).

Results

RNA-seq transcriptome data of m^5C-related regulators in COAD

Based on RNA-seq transcriptome data of COAD from TCGA database, the expression of 13 m^5C-related regulators between tumor tissues and adjacent mucosa was compared (**Figure 2**). With the exceptions of TET1 and TET3, the expression levels of the other 11 factors were significantly different in the tumor tissues and the adjacent mucosal tissues. Compared with the adjacent mucosa, the expression of NSUN3 ($P < 0.001$) and TET2 ($P < 0.001$) in the tumor group was significantly downregulated. The expression of ALKBH1 ($P = 0.036$), ALYREF ($P < 0.001$), NOP2 ($P < 0.001$), NSUN2 ($P < 0.001$), NSUN4 ($P < 0.001$), NSUN5 ($P < 0.001$), NSUN6 ($P < 0.001$), NSUN7 ($P = 0.006$), and YBX1($P < 0.001$) were significantly upregulated in tumor tissues compared with the adjacent mucosa.

Correlation and interaction of m^5C-related regulators in COAD

The correlations between the m^5C-related regulators were analyzed using the “corrplot” package in R and their interrelationships were retrieved from the STRING database (https://string-db.org/). The expression levels of the seven “writers” were correlated with each other, except for NSUN2 and NSUN7, NSUN5 and NSUN7, NSUN2 and NSUN3, and NSUN5 and NSUN6. There were also close and complicated relationships between each regulator in the protein-protein interaction (PPI) network. We also found that the expression of TET family genes (TET1, TET2, TET3) were highly related to each other and had little correlation with ALKBH1. However, the TET family was associated with ALKBH1 in the PPI network and had interrelationships with the “writer” genes via ALKBH1. In addition, there was evidence supporting the interaction between the “reader” genes ALYREF and YBX1 in the PPI network. The expression of these genes was also positively associated with each other (**Figure 3**).

CNVs and SNPs of m^5C-related regulators in COAD

Regarding CNVs, we found that 10 of the 13 m^5C-related regulators were significantly different between the tumor tissue and the adjacent mucosa from 825 samples with CNV data. Furthermore, it was found that CNVs affect the expression of m^5C-related regulators. The highest frequency of CNVs occurred in the “writer” gene NSUN5 (24.47%), followed by the “eraser” gene ALKBH1 (19.53%). The “eraser” gene TET3 had the lowest CNV frequency (2.35%) (**Table 2)**. The “writer” genes NOP2, NSUN2, NSUN5, and NSUN7, the “eraser” genes TET2 and ALKBH, and the “reader” gene ALYREF displayed a significant difference in expression due to CNVs (**Figure 4**).

Regarding SNPs, we found that all of the m^5C-related regulators had missense mutations, and missense mutations were the highest frequency mutation in 399 COAD cases with available sequencing data. Among them, the m^5C “eraser” gene TET2 had the highest frequency of mutation events (96/399), followed by TET3 and TET1 (both 39/399). In addition, the “writer” genes NSUN2 and NSUN7, the “eraser” gene TET2, and the “reader” gene ALYREF displayed significant differences in expression levels due to SNPs. Next, we evaluated the effect of SNPs on patient prognosis, but no difference was observed due to the relatively few numbers of mutations (**Figure 5**).

Consensus clustering of patients with COAD and the survival rate of subgroups

Based on the expression levels of 13 m^5C-related regulators, a consistent clustering analysis of patients with COAD was performed, and they were clustered into two subgroups because there was minimal interference between the two subgroups (**Figure 6A, B, C, D**).

PCA showed that the RNA expression levels in patients with COAD in clusters I and II were specific (**Figure 6E**). Nevertheless, there were many overlapping areas between each cluster on the whole, indicating that the clusters had something in common; however, there were no significant differences between the two subgroups. Notably, the overall survival rate of cluster I and cluster II when analyzed using the Kaplan-Meier method was significantly different (**Figure 6F**). Specifically, the 5-year survival rate was 57.4% in cluster I and 66.7% in cluster II.

Prognostic value of m^5C-related regulators in COAD prognosis

To evaluate the prognostic value of these 13 m^5C-related regulators in COAD, univariate Cox regression analysis was used to identify m^5C-related regulators that were highly correlated with the OS in patients with COAD, and two regulators with prognostic significance ($P < 0.05$) were found: NSUN6 and ALYREF. Specifically, ALYREF was considered a protective factor with HR < 1 in patients with COAD and NSUN6 was considered as a risk factor with HR > 1 (**Figure 7A**). To further evaluate the prognostic significance of these two m^5C-related regulators, LASSO Cox regression analysis was performed and it was revealed that NSUN6 (Coef=0.300256795278519) and ALYREF (Coef=0.00796895949684636) could serve as powerful prognostic factors in COAD (**Figure 7B, C, D**).

Based on NSUN6 and ALYREF, a risk signature was constructed and the risk score was calculated.Using the median risk score as the demarcation value, patients with COAD (n = 525) were classified into two groups, namely the high-risk and low-risk groups. To test the prognostic role of the two gene risk signatures. Survival and ROC curve analyses were conducted; the low-risk group had significantly longer survival time than the high-risk group (**Figure 7E**). In particular, compared with the 46.4% 5-year survival rate in the high-risk group, that of the low-risk group was 78.7%. The area under the curve (AUC) value in the time-dependent ROC curve was 0.754, suggesting good prediction performance of the survival model (**Figure 7F**).

Relationship between the risk score, the expression of the two selected m^5C-related regulators, and clinicopathological characteristics in COAD

The expression of NSUN6 and ALYREF and the distribution of clinicopathological characteristics in the high-risk and low-risk groups are displayed as a heatmap (**Figure 8A**). Evident differences between the two groups according to stage T ($P < 0.05$) and fustat ($P < 0.01$) were observed.

To examine whether the risk score was an independent prognostic factor, univariate and multivariate Cox regression analyses were conducted. This revealed that the risk score was significantly associated with OS in univariate analysis, in addition to age at diagnosis, pathological stage, and TNM stage ($P < 0.05$). However, only the age at diagnosis and risk score were correlated with OS ($P < 0.05$) in the multivariate Cox regression analysis (**Figure 8B,C**).

Biological functional analysis of differentially expressed genes between different subgroups and m5C-related regulatory gene

As we clustered the patients with COAD into cluster and cluster , genes that were significantly upregulated (fold change > 1 and $p < 0.05$) or downregulated (fold change < 1 and $p < 0.05$) between the high-risk group and low-risk group were identified using the "edgeR" package in R. GO and KEGG pathway analysis were used for biological functional analysis.

With regard to GO analysis, the differentially expressed genes were associated with immune-related biological processes, such as "antigen binding" and "immunoglobulin receptor binding," and pre-mRNA-related biological processes, such as "pre-mRNA 5'-splice site binding" and "pre-mRNA binding."(**Figure 9A**). KEGG pathway analysis results were correlated with immune-related pathways, including "complement and coagulation cascades" and "NOD-like receptor signaling pathway," and RNA-related pathways, including "RNA transport" and "spliceosome." Moreover, cancer-related pathways were enriched, such as "transcriptional misregulation in cancer" and "MAPK signaling pathway" (**Figure 9B**).

Next, we used GSEA to predict the functional difference between clusters I and II. The results showed that cluster I had a worse OS and lower 5-year survival rate associated with malignancy-associated pathways, including the ATP-binding cassette transporter (NES = 1.79, normalized $P = 0.006$) and phosphatidylinositol signaling system (NES = 1.63, normalized $P = 0.03$) (**Figure 9C,D**).

Furthermore, as NSUN6 and ALYREF were shown to be important regulators of m^5C in our study, GSEA was performed to investigate the potential biological processes associated with NSUN6 and ALYREF in COAD pathogenesis. GSEA suggested that increased expression of NSUN6 and ALYREF is involved in various biological functions in RNA processing, such as spliceosome, RNA polymerase, and RNA degradation. Upregulation of these genes was associated with malignancy-associated pathways, such as the cell cycle (**Figure 9E,F**).

Discussion

RNA modifications have been increasingly demonstrated in tumorigenesis and tumor progression, suggesting that RNA epigenetic regulators may play an important role in COAD. Previous studies have shown that m^6A RNA modification not only plays a critical role in the tumorigenesis and progression of CRC, but also has powerful significance in the diagnosis and prognosis of CRC patients [28]. However, the literature on CRC and m^5C has largely focused on DNA methylation [29]. Little is known about the relationship between m^5C-related RNA modifications and CRC, which calls our attention to investigate the aberrant expression of m^5C-related regulators in COAD and explore whether m^5C-related regulators could serve as ideal biomarkers for COAD prognosis and participate in COAD initiation and progression.

In our study, we showed that the expressions of m^5C-related regulators were significantly altered between tumor tissues and adjacent mucosa and had a strong correlation with the tumor progression and prognosis. This indicated that m^5C-related regulators play crucial role in COAD. First, the "writer" genes NSUN1-NSUN7, the "eraser" genes TET2 and ALKBH1, and the "reader" genes ALYREF and YBX1 were significantly upregulated or downregulated in tumor tissues, suggesting these genes may be critical in m^5C-related occurrence and progression of COAD. To investigate the relationship between CNVs or SNPs of m^5C-related regulators and their mRNA expression levels, COAD samples with CNV or SNP data from TCGA were analyzed. Regarding CNVs, the copy number of seven m^5C-related regulators increased or was lost, and their mRNA expression was upregulated or downregulated accordingly and were significantly correlated. SNPs in TET2 and ALYREF were highly correlated

with their high mRNA expression, while SNPs of NSUN2 and NSUN7 were significantly correlated with their low mRNA expression levels.

Thereafter, based on the expression of the m^5C-related regulators, patients with COAD were clustered into two subgroups (cluster and cluster), which had significant differences in the OS rate. To further study the effect of m^5C-related regulators on the prognosis and clinicopathological characteristics of COAD, we constructed a prognostic risk signature using two identified m^5C-related regulators (NSUN6 and ALYREF) and were able to assign patients with COAD into high- and low-risk groups. The correlation between the groups and clinicopathological characteristics were assessed, which revealed that the high-risk group was linked with stage T and fustat. Based on the risk value, the established ROC curve showed a satisfactory prediction performance. Moreover, the risk score can be used as an independent prognostic factor for COAD, suggesting that NSUN6 and ALYREF may be vital m^5C-related regulators and significant prognostic factors for patients with COAD.

Recently, many studies have indicated that m^5C RNA modification is involved in all types of human cancer. NSUN2 is the most studied m^5C methyltransferase and participates in various cancers, such as bladder cancer, gallbladder carcinoma, and hepatocellular carcinoma [30–32]. It was reported that NSUN2 is highly expressed in colon cancers [33], which was corroborated in our results. NSUN2 mainly exerts an oncogenic role by maintaining the stability of oncogenic RNA [34], but whether NSUN2 plays the same role in COAD requires further research. With respect to the two m^5C-related regulators (NSUN6 and ALYREF) identified in our results, there have been some studies on cancer and related mechanisms. The role of NSUN6 in regulating cell proliferation and pancreatic cancer tumor growth was recently confirmed, and NSUN6 performs well in evaluating tumor recurrence and survival among pancreatic cancer patients [35]. Next, ALYREF was found to be upregulated in hepatocellular carcinoma and oral squamous cell carcinoma, and it may have an effect on tumorigenesis via cell cycle regulation and mitosis [36, 37].

To provide a comprehensive analysis, GO, KEEG pathway and GSEA analyses of m^5C-related regulators were also conducted. Several biological processes and pathways associated with the occurrence and progression of COAD were enriched, including “MAPK signaling pathway” and “cell cycle” [38, 39]. Moreover, previous studies have reported that m^5C-related RNA modifications are closely associated with mRNA translation, transport, and stability. Here, we found that the m^5C-related regulators were associated with “pre-mRNA 5'-splice site binding” and “spliceosome,” suggesting they play important roles in RNA processing. In addition, it should be noted that a number of biological processes and pathways associated with immune response were identified. While extensive literature reports have demonstrated that N6-methyladenosine plays important role in immune evasion and immune response [40, 41], there have been few reports about the relationship between m^5C-related RNA modifications and immune response, suggesting that further research is required.

Conclusion

In this study, we first found that there was a significant correlation between the expression of m^5C-related regulators and clinicopathological features and OS of patients with COAD. This revealed that a prognostic signature obtained using m^5C-related regulators (NSUN6 and ALYREF) had significant value in COAD and could effectively predict the survival of patients with COAD. Additionally, biological processes and pathways associated with m^5C-related RNA modifications were identified, which may facilitate the malignant development of COAD,

thus improving our understanding of the role of m^5C-related RNA modifications in the occurrence and progression of COAD. This work also provides important evidence towards the development of predictive biomarkers and molecular targeted therapy for COAD.

Declarations

Ethics Statement

This study met the publication guidelines stated by TCGA (https://cancergenome.nih.Gov/publications/publicationguidelines). All data used in the study were obtained from TCGA and ethics approval and informed consent were not required.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Consent for Publication

Not applicable.

Data Availability Statement

Publicly available datasets were analyzed in this study, which can be found in the Cancer Genome Atlas (TCGA) database.

Author Contributions

YH and YW conceived and designed the study. CH, XJ, and YY organized the database and performed statistical analyses. YH and YW wrote the first draft of the manuscript. CH, XJ and YY prepared the figures and tables and were involved in manuscript writing. KZ, PL and FL revised and proofread the manuscript. All authors contributed to manuscript revision and approved the submitted version.

Funding

This study was supported by The National Natural Science Foundation (No. 81904139, No. 81973819 and No. 81904145), Collaborative Innovation Team Project of First-Rate Universities and Disciplines and High-level University Discipline of Guangzhou University of Chinese Medicine (No. 2021xk47), The Natural Science Foundation of Guangdong Province (No. 2019A1515011145), Guangdong Medical Science and Technology Research Fund (No. B2021089, No. A2020186), Clinical Research Project of Innovation Hospital in the First Affiliated Hospital of Guangzhou University of Chinese Medicine (No. 2019□T19) and High-level Hospital Construction project of the First Affiliated Hospital of Guangzhou University of Chinese Medicine (No. 2019QN01).

Acknowledgments

This study benefited from the Cancer Genome Atlas (TCGA) databases. We appreciate the data platform and the authors up-loaded their data.We would like to thank Editage (www.editage.cn) for English language editing.

References

1. Wu C, Morris JR, "Genes, genetics, and epigenetics: a correspondence," *SCIENCE*, vol. 293, no. 5532, pp. 1103-1105, 2001.
2. Grady WM, Yu M, Markowitz SD, "Epigenetic Alterations in the Gastrointestinal Tract: Current and Emerging Use for Biomarkers of Cancer," *GASTROENTEROLOGY*, vol. 160, no. 3, pp. 690-709, 2021.
3. Dawson MA, Kouzarides T, "Cancer epigenetics: from mechanism to therapy," *CELL*, vol. 150, no. 1, pp. 12-27, 2012.
4. Darwiche N. "Epigenetic mechanisms and the hallmarks of cancer: an intimate affair". American Journal of Cancer Research. 2020;10(7):1954–78.
5. Schaefer MR. "The Regulation of RNA Modification Systems: The Next Frontier in Epitranscriptomics?" *Genes (Basel)*, vol. 12, no. 3, pp.2021.
6. Angelova MT, Dimitrova DG, Dinges N, et al. "The Emerging Field of Epitranscriptomics in Neurodevelopmental and Neuronal Disorders". Front Bioeng Biotechnol. 2018;6:46.
7. Zhao Z, Meng J, Su R, et al. "Epitranscriptomics in liver disease: Basic concepts and therapeutic potential". JOURNAL OF HEPATOLOGY. 2020;73(3):664–79.
8. Minervini CF, Parciante E, Impera L, et al, "Epitranscriptomics in Normal and Malignant Hematopoiesis" *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, vol. 21, no. 18, pp.2020.
9. Song H, Liu D, Dong S, et al. "Epitranscriptomics and epiproteomics in cancer drug resistance: therapeutic implications". Signal Transduct Target Ther. 2020;5(1):193.
10. Porcellini E, Laprovitera N, Riefolo M, Ravaioli M, Garajova I, Ferracin M. "Epigenetic and epitranscriptomic changes in colorectal cancer: Diagnostic, prognostic, and treatment implications". CANCER LETTERS. 2018;419:84–95.
11. Shi H, Chai P, Jia R, Fan X. "Novel insight into the regulatory roles of diverse RNA modifications: Re-defining the bridge between transcription and translation". Molecular Cancer. 2020;19(1):78.
12. Roundtree IA, Evans ME, Pan T, He C. "Dynamic RNA Modifications in Gene Expression Regulation" *CELL*. 2017;169(7):1187–200.
13. Begik O, Lucas MC, Liu H, Ramirez JM, Mattick JS, Novoa EM. "Integrative analyses of the RNA modification machinery reveal tissue- and cancer-specific signatures. GENOME BIOLOGY. 2020;21(1):97.
14. Barbieri I, Kouzarides T. "Role of RNA modifications in cancer". NATURE REVIEWS CANCER. 2020;20(6):303–22.
15. Chen YS, Yang WL, Zhao YL, Yang YG. "Dynamic transcriptomic m(5) C and its regulatory role in RNA processing" *Wiley Interdiscip Rev RNA*, vol., pp. e1639, 2021.
16. Hussain S, "The Emerging Roles of Cytosine-5 Methylation in mRNAs," *TRENDS IN GENETICS*, vol., pp.2021.
17. 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.
18. Nombela P, Miguel-López B, Blanco S. "The role of m(6)A, m(5)C and Ψ RNA modifications in cancer: Novel therapeutic opportunities". Molecular Cancer. 2021;20(1):18.
19. Xie S, Chen W, Chen K, et al. "Emerging roles of RNA methylation in gastrointestinal cancers". Cancer Cell Int. 2020;20(1):585.

20. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". CA Cancer J Clin. 2018;68(6):394–424.
21. White A, Joseph D, Rim SH, Johnson CJ, Coleman MP, Allemani C, "Colon cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study," *CANCER*, vol. 123 Suppl 24, no. Suppl 24, pp. 5014-5036, 2017.
22. Miller KD, Nogueira L, Mariotto AB, et al. "Cancer treatment and survivorship statistics, 2019". CA Cancer J Clin. 2019;69(5):363–85.
23. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. "Mapping and quantifying mammalian transcriptomes by RNA-Seq". NATURE METHODS. 2008;5(7):621–8.
24. Wilkerson MD, Hayes DN, "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking," *BIOINFORMATICS*, vol. 26, no. 12, pp. 1572-1573, 2010.
25. Ritchie ME, Phipson B, Wu D, et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies". NUCLEIC ACIDS RESEARCH. 2015;43(7):e47.
26. Hanley JA, McNeil BJ, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *RADIOLOGY*, vol. 143, no. 1, pp. 29-36, 1982.
27. Simon N, Friedman J, Hastie T, Tibshirani R. "Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent". J Stat Softw. 2011;39(5):1–13.
28. Li J, Liang L, Yang Y, Li X, Ma Y, "N(6)-methyladenosine as a biological and clinical determinant in colorectal cancer: progression and future direction," *Theranostics*, vol. 11, no. 6, pp. 2581-2593, 2021.
29. Zhu Y, Lu H, Zhang D, et al. "Integrated analyses of multi-omics reveal global patterns of methylation and hydroxymethylation and screen the tumor suppressive roles of HADHB in colorectal cancer". Clinical Epigenetics. 2018;10:30.
30. 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*, vol. 39, no. 45, pp. 6906-6919, 2020.
31. Gao Y, Wang Z, Zhu Y, et al. "NOP2/Sun RNA methyltransferase 2 promotes tumor progression via its interacting partner RPL6 in gallbladder carcinoma". CANCER SCIENCE. 2019;110(11):3510–9.
32. Chen X, Li A, Sun BF, et al. "5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs". NATURE CELL BIOLOGY. 2019;21(8):978–90.
33. Okamoto M, Hirata S, Sato S, et al. "Frequent increased gene copy number and high protein expression of tRNA (cytosine-5-)-methyltransferase (NSUN2) in human cancers". DNA AND CELL BIOLOGY. 2012;31(5):660–71.
34. Chellamuthu A, Gray SG. "The RNA Methyltransferase NSUN2 and Its Potential Roles in Cancer" *Cells*, vol. 9, no. 8, pp.2020.
35. Yang R, Liang X, Wang H, et al, "The RNA methyltransferase NSUN6 suppresses pancreatic cancer development by regulating cell proliferation," *EBioMedicine*, vol. 63, pp. 103195, 2021.
36. He Y, Yu X, Li J, Zhang Q, Zheng Q, Guo W. "Role of m(5)C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma". American Journal of Translational Research. 2020;12(3):912–22.
37. Saito Y, Kasamatsu A, Yamamoto A, et al. "ALY as a potential contributor to metastasis in human oral squamous cell carcinoma". J Cancer Res Clin Oncol. 2013;139(4):585–94.

38. Malki A, ElRuz RA, Gupta I, Allouch A, Vranic S, Al MA. "Molecular Mechanisms of Colon Cancer Progression and Metastasis: Recent Insights and Advancements" *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, vol. 22, no. 1, pp.2020.
39. Koveitypour Z, Panahi F, Vakilian M, et al. "Signaling pathways involved in colorectal cancer progression". Cell Bioscience. 2019;9:97.
40. Lou X, Wang JJ, Wei YQ, Sun JJ. "Emerging role of RNA modification N6-methyladenosine in immune evasion". Cell Death Dis. 2021;12(4):300.
41. Shulman Z, Stern-Ginossar N. The RNA modification N(6)-methyladenosine as a novel regulator of the immune system. NATURE IMMUNOLOGY. 2020;21(5):501–12.

Tables

TABLE 1|Clinicopathological features of patients included in this study

	Total patients(337)		High-risk group(163)		Low-risk group(168)		*p*-value
	Number	Percentage(%)	Number	Percentage(%)	Number	Percentage(%)	
Fustat							0.009
Alive	279	82.8	125	76.7	148	88.1	
Dead	58	17.2	38	23.3	20	11.9	
Age							0.178
≤65	135	40.1	72	44.2	61	36.3	
65	202	59.9	91	55.8	107	63.7	
gender							0.714
female	156	46.3	78	47.9	76	45.2	
male	181	53.7	85	52.1	92	54.8	
Stage							
I	59	17.5	28	17.2	30	17.9	
II	137	40.7	61	37.4	73	43.5	
III	87	25.8	42	25.8	44	26.2	
IV	54	16.0	32	19.6	21	12.5	
Stage T							0.016
T1	7	2.1	5	3.1	2	1.2	
T2	59	17.5	26	16.0	32	19.0	
T3	235	69.7	106	65.0	124	73.8	
T4	36	10.7	26	16.0	10	6.0	
Stage M							0.105
M0	283	84.0	131	80.4	147	87.5	
M1	54	16.0	32	19.6	21	12.5	
Stage N							0.202
N0	203	60.2	92	56.4	107	63.7	
N1	76	22.6	37	22.7	38	22.6	
N2	58	17.2	34	20.9	23	13.7	

TABLE 2|Copy number variants (CNV) of m5C related regulators in colon adenocarcinoma

Function	Genes	Diploid	Deletion	Amplification	CNV sum	Deletion %	Amplification %	Percentage
Writers	NOP2	379	6	40	46	13.04%	86.96%	10.82%
	NSUN2	385	7	33	40	17.50%	82.50%	9.41%
	NSUN3	406	5	14	19	26.32%	73.68%	4.47%
	NSUN4	406	17	2	19	89.47%	10.53%	4.47%
	NSUN5	321	1	103	104	0.96%	99.04%	24.47%
	NSUN6	405	9	11	20	45.00%	55.00%	4.71%
	NSUN7	391	32	2	34	94.12%	5.88%	8.00%
Erasers	TET1	405	13	7	20	65.00%	35.00%	4.71%
	TET2	396	26	3	29	89.66%	10.34%	6.82%
	TET3	415	2	8	10	20.00%	80.00%	2.35%
	ALKBH1	342	78	5	83	93.98%	6.02%	19.53%
Readers	ALYREF	377	13	35	48	27.08%	72.92%	11.29%
	YBX1	401	19	5	24	79.17%	20.83%	5.65%

Figures

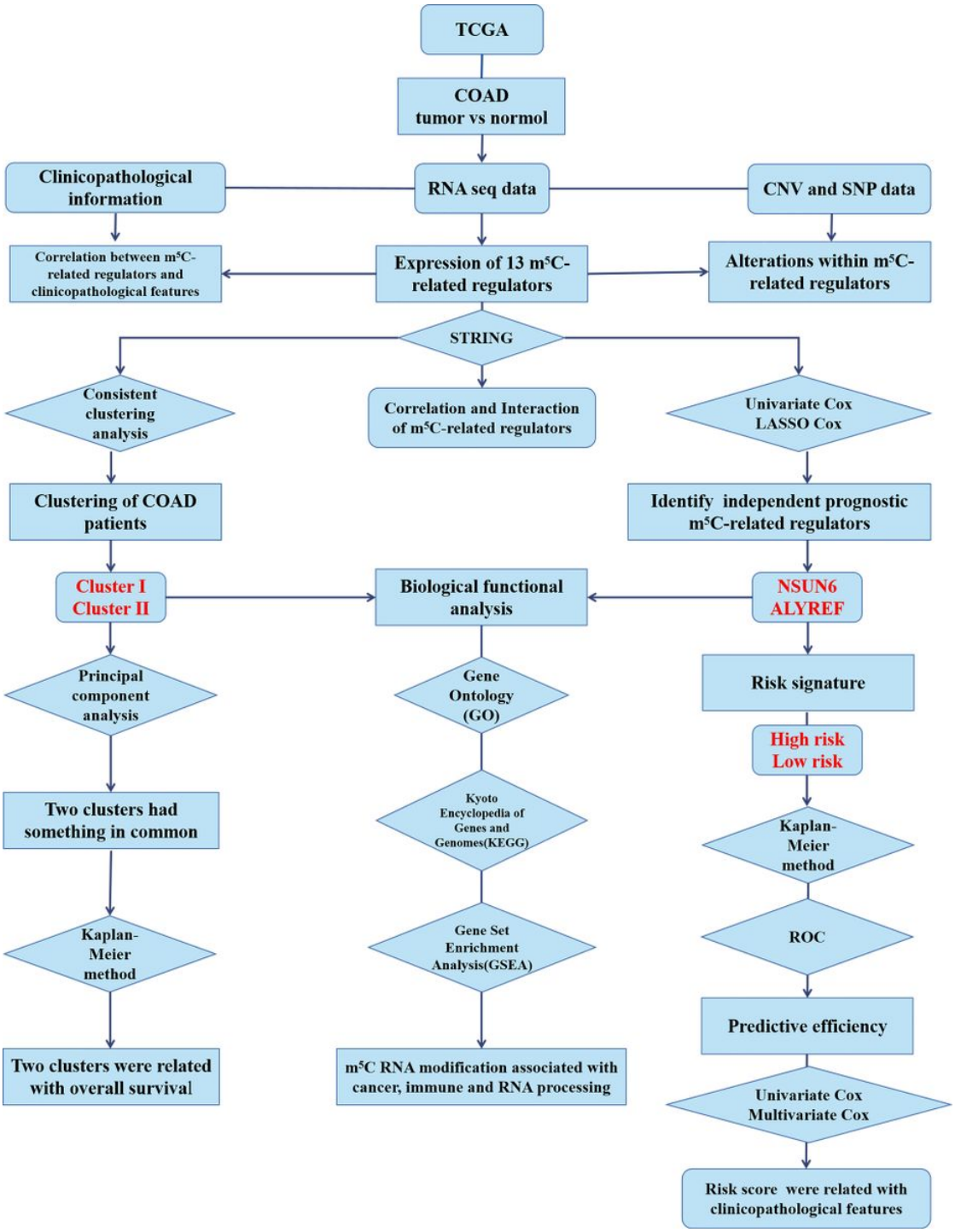

Figure 1

The flow chart of the study design and analysis.

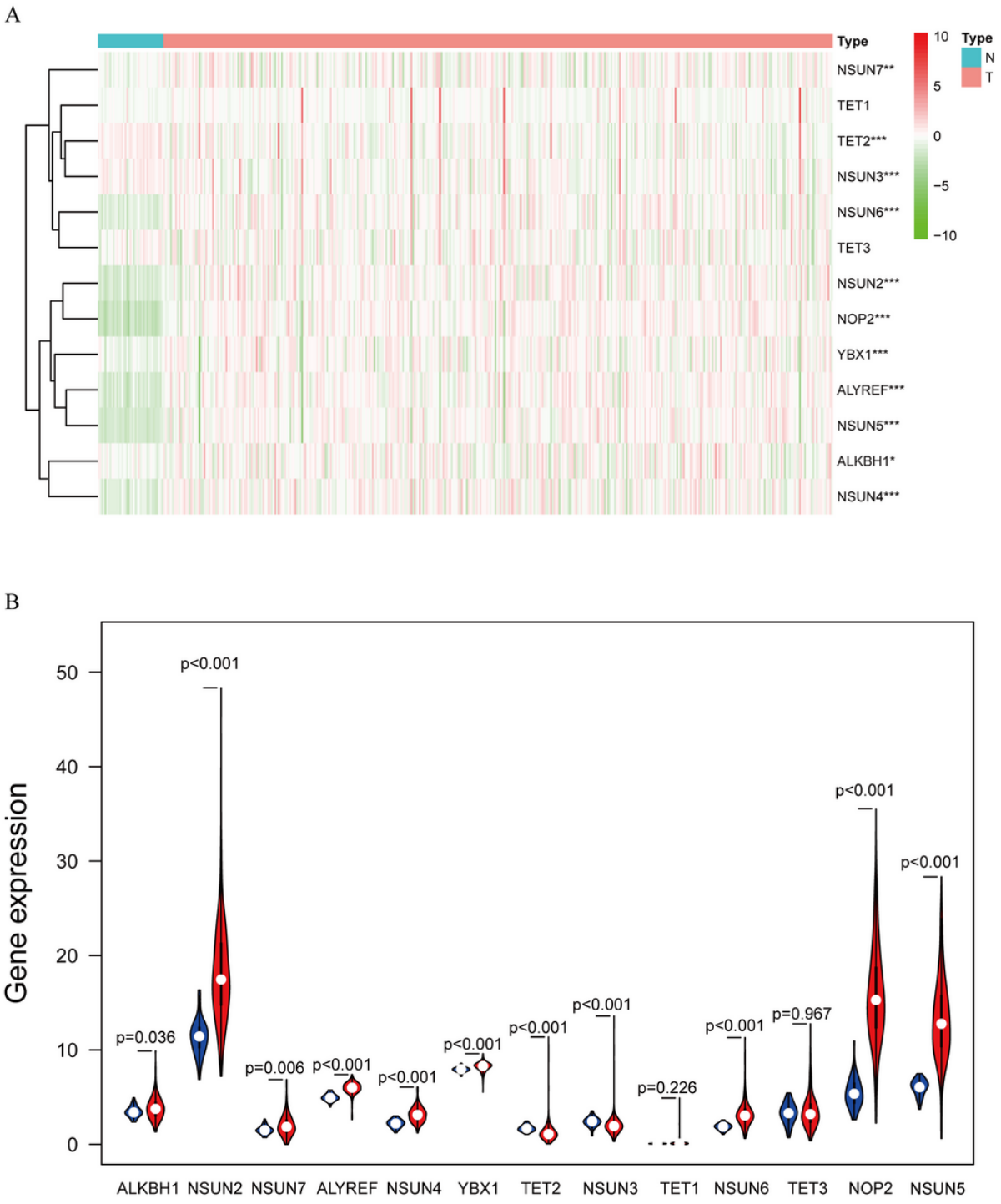

Figure 2

The expression of 13 m5C-related regulators in TCGA database between the tumor group and the normal group. (A) Heatmap of the expression of 13 m5C-related regulators. The depth of red represents the level of high expression, and the depth of green represents the level of low expression * p<0.05, ** <0.01, ***<0.001. (B) The violin diagram showed the median expression of 13 m5C-related regulators in COAD, and position of white spots on the way represented the median value of expression.

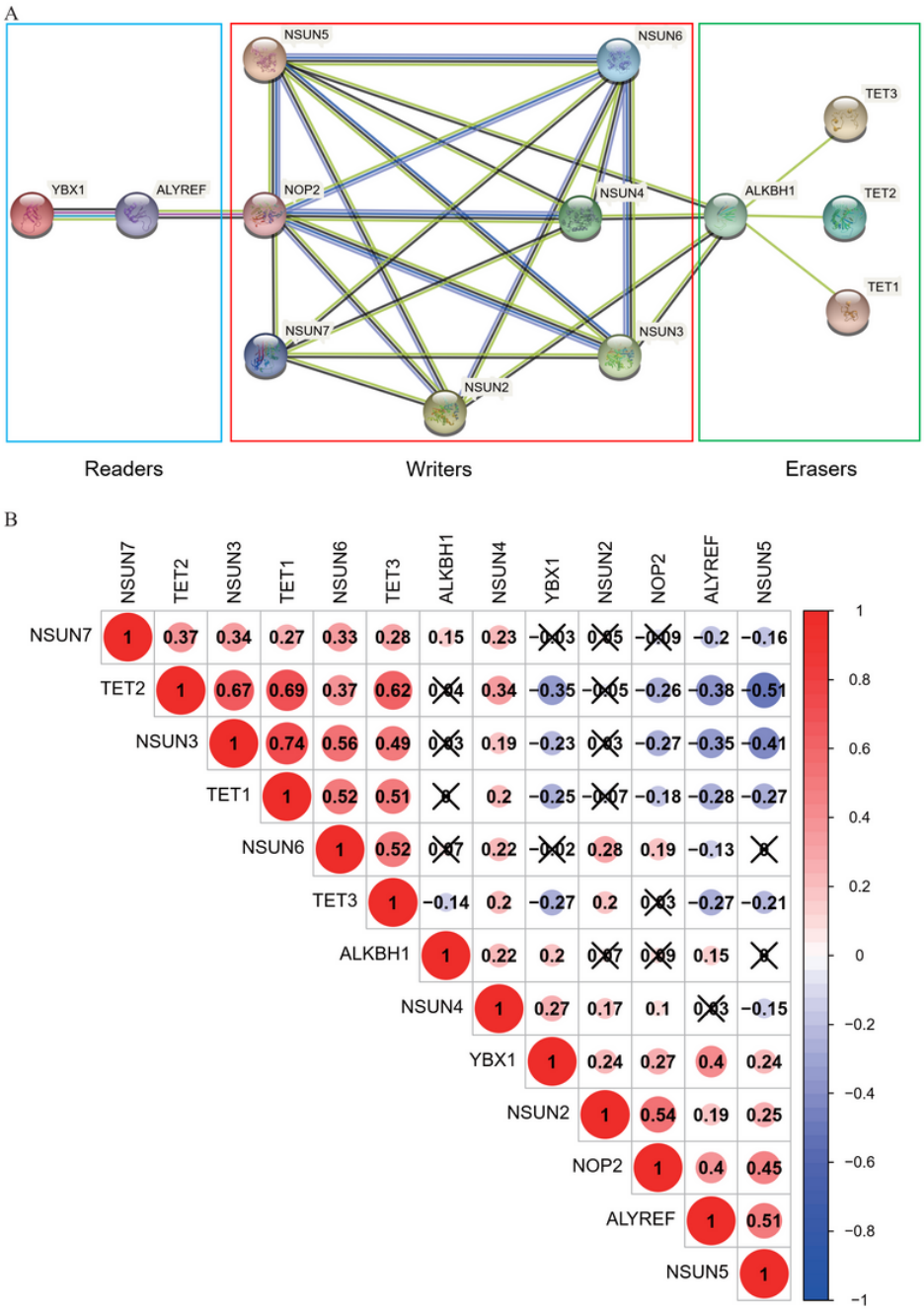

Figure 3

Correlation and interaction of m5C-related regulators in COAD. (A)The PPI network of the 13 m5C-related regulators constructed using STRING. (B) Spearman correlation analysis of the 13 m5C-related regulators.

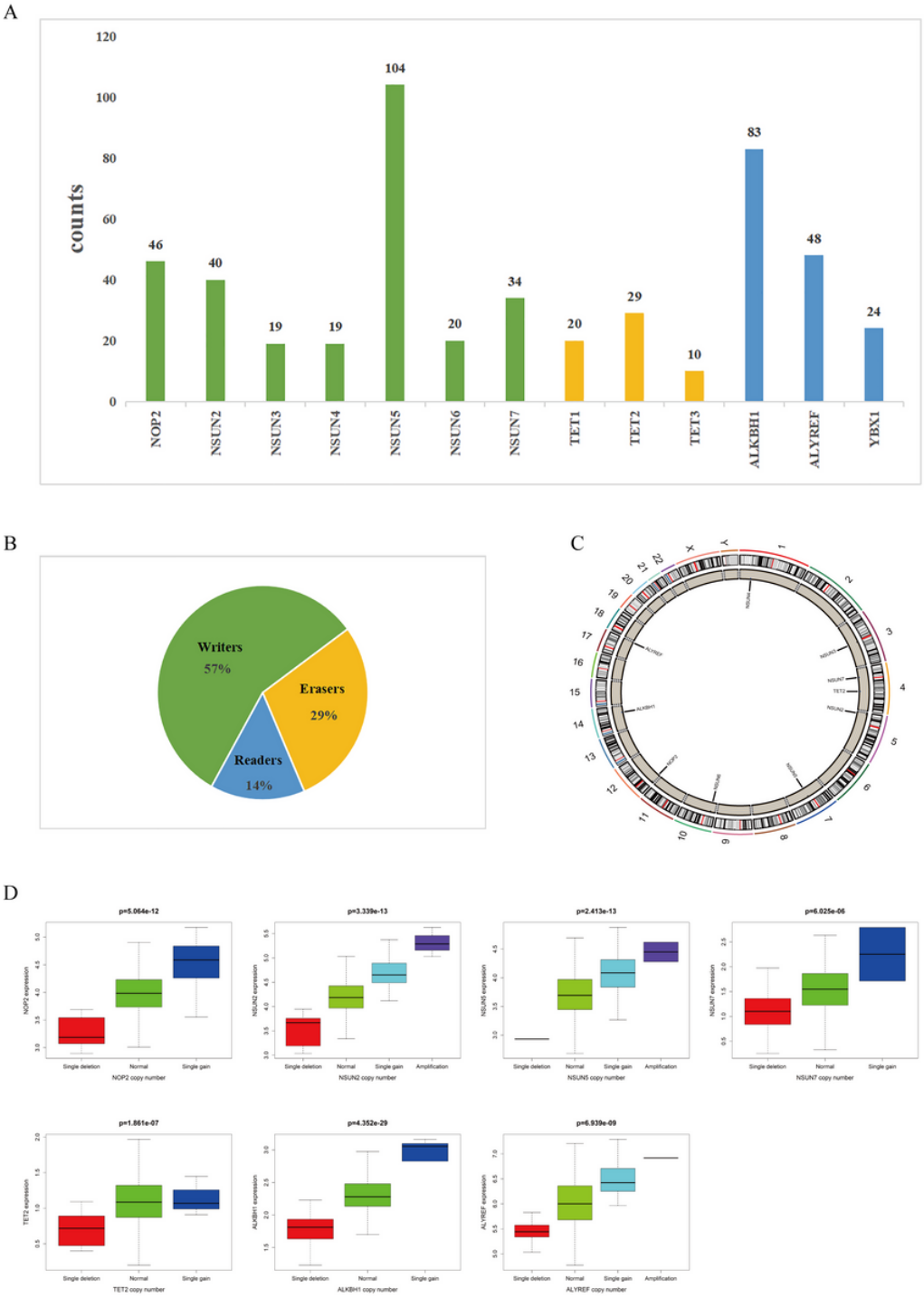

Figure 4

The landscape of CNV of m5C-related regulators in COAD.(A,B) Frequency of CNV of 13 m5C-related regulators in COAD. (B) Percentage of CNV of 13 m5C-related regulators in COAD.(C) Location of CNV alteration of 13 m5C-related regulators on chromosomes.(D) NOP2, NSUN2, NSUN5, NSUN7, TET2, ALKBH, and ALYREF displayed a significant difference in expression due to CNVs.

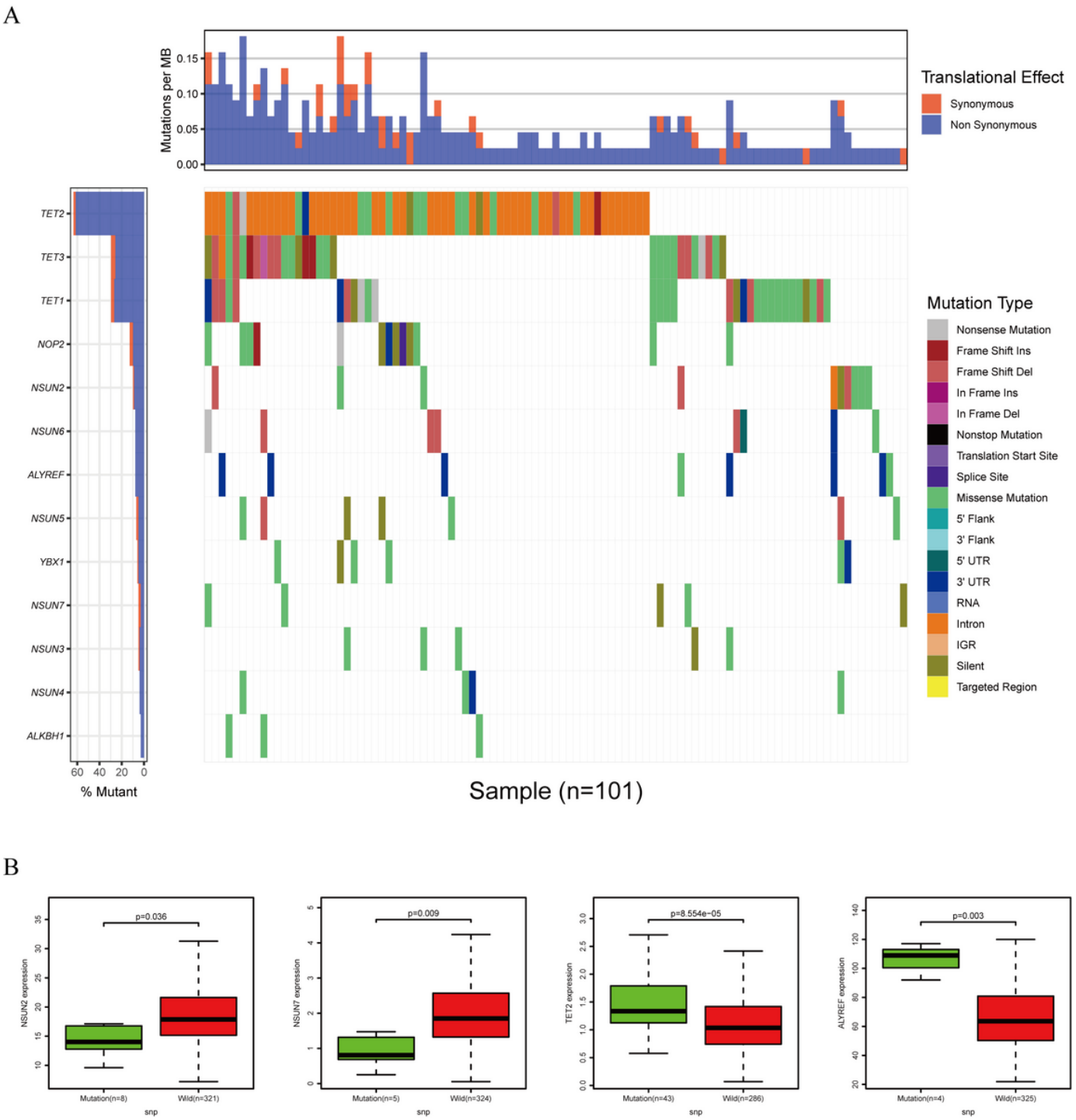

Figure 5

The landscape of SNP of m5C-related regulators in COAD. (A) Waterfall plot of SNP of 13 m5C-related regulators in COAD. (B) NSUN2, NSUN7, TET2, and ALYREF displayed significant differences in expression levels due to SNPs..

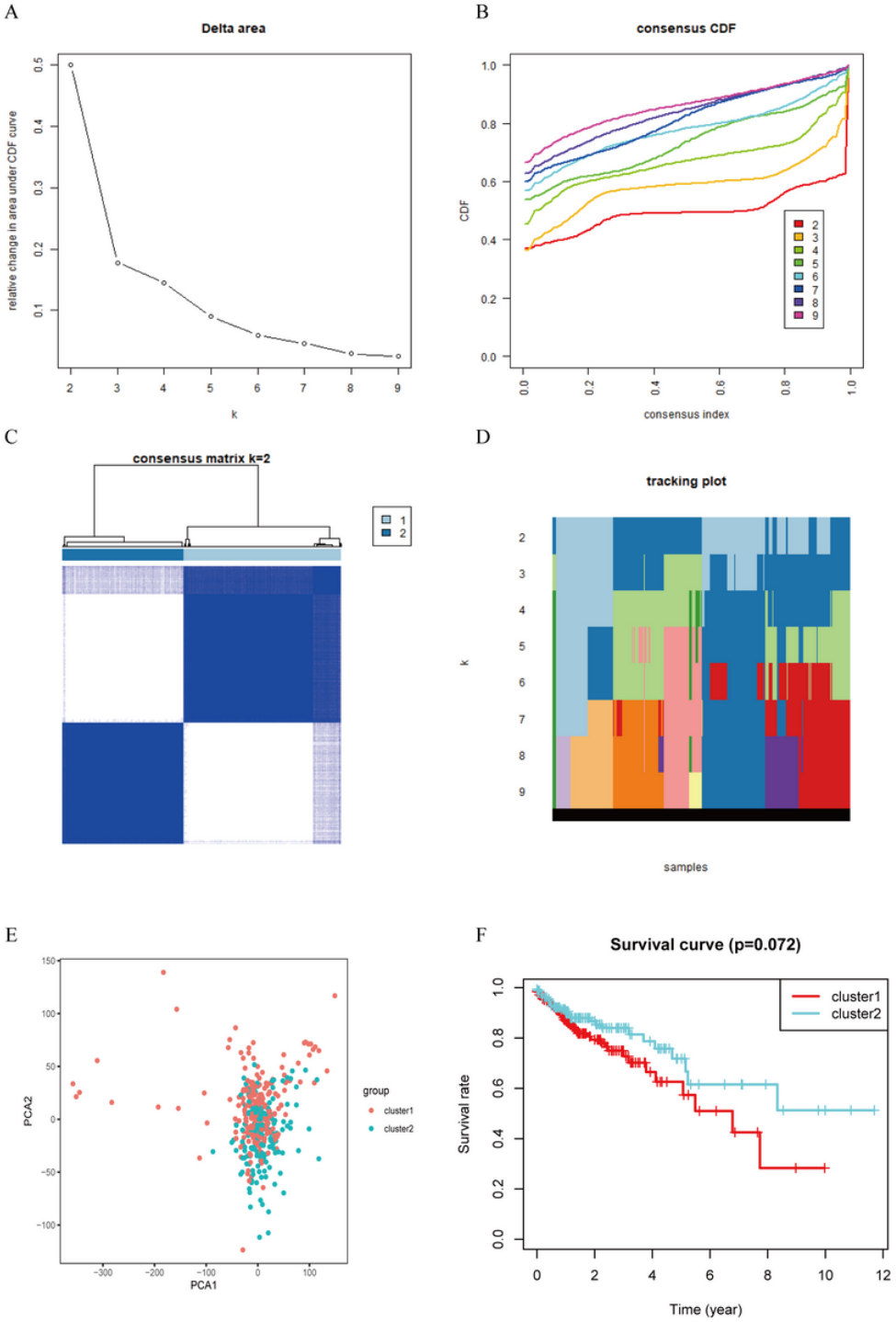

Figure 6

Consistent cluster analysis and principal component analysis of COAD. (A) The consistency clustering cumulative distribution function (CDF) when k is between 2 and 10. (B) The relative change of the area under the CDF curve from 2 to 10 of k. (C) At k=2, the correlation between groups. (D) The distribution of the sample when k is between 2 and 10. (E) Principal component analysis of 2 clusters of total RNA expression profile after consistency analysis. (F) Comparison of Kaplan-Meier overall survival (OS) curve for COAD patients in cluster and .

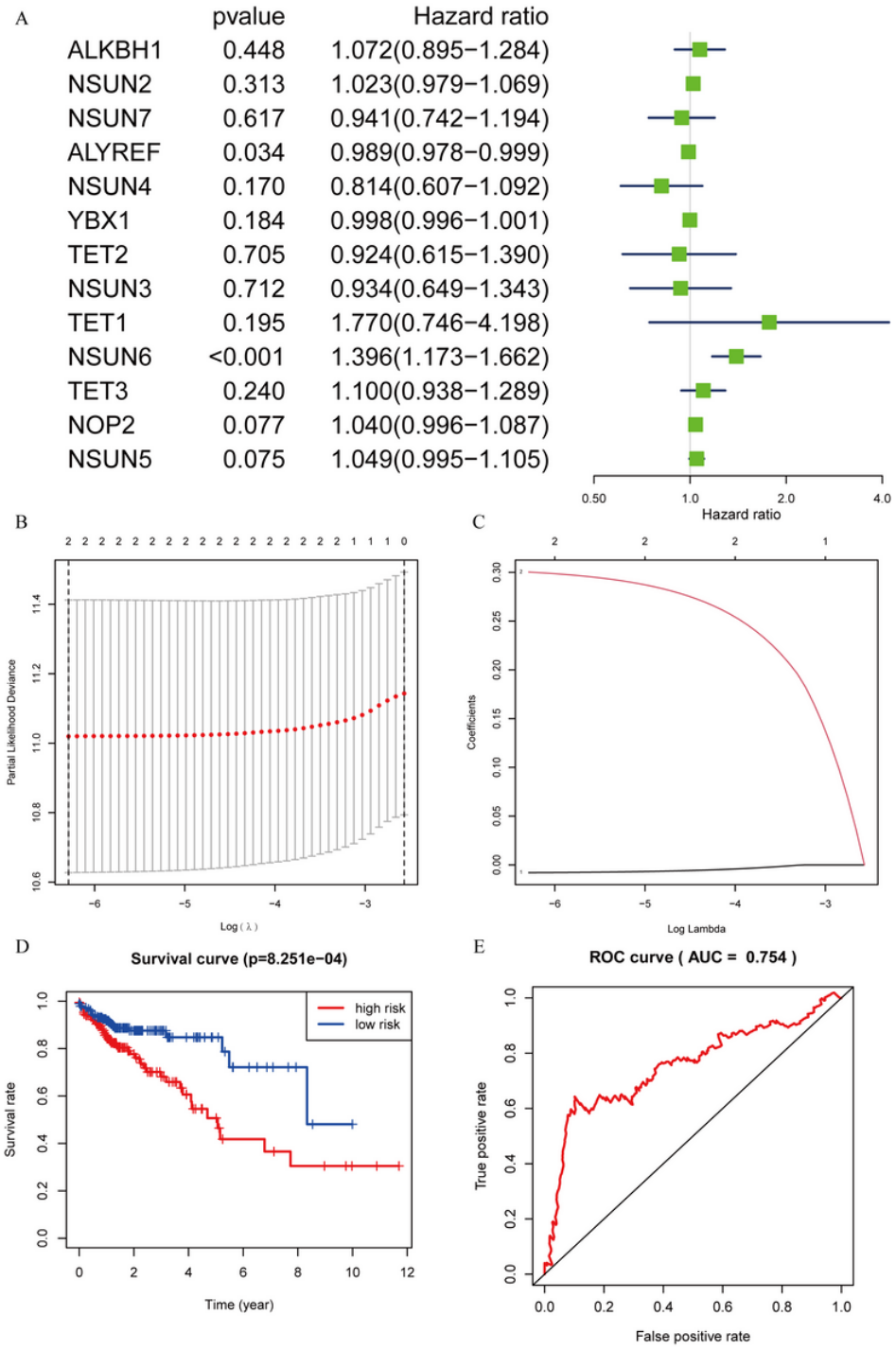

Figure 7

The process of constructing the signature based on NSUN6 and ALYREF and evaluating its prognostic value. (A) The Hazard ratio (HR), 95% confidence interval (CI) of 13 m5C-related regulators estimated by univariate Cox regression. (B) The point with the smallest cross verification error corresponds to the number of factors included in the Lasso regression model. (C) The lines of different colors represent the trajectory of the correlation coefficient of different factors in the model with the increase of Log Lamda. (D) Kaplan-Meier overall survival curves for patients in high-risk group- and low-risk group divided according to the risk score. (E) ROC analysis and AUC value of the ROC curve suggesting the sensitivity and specificity for risk signature.

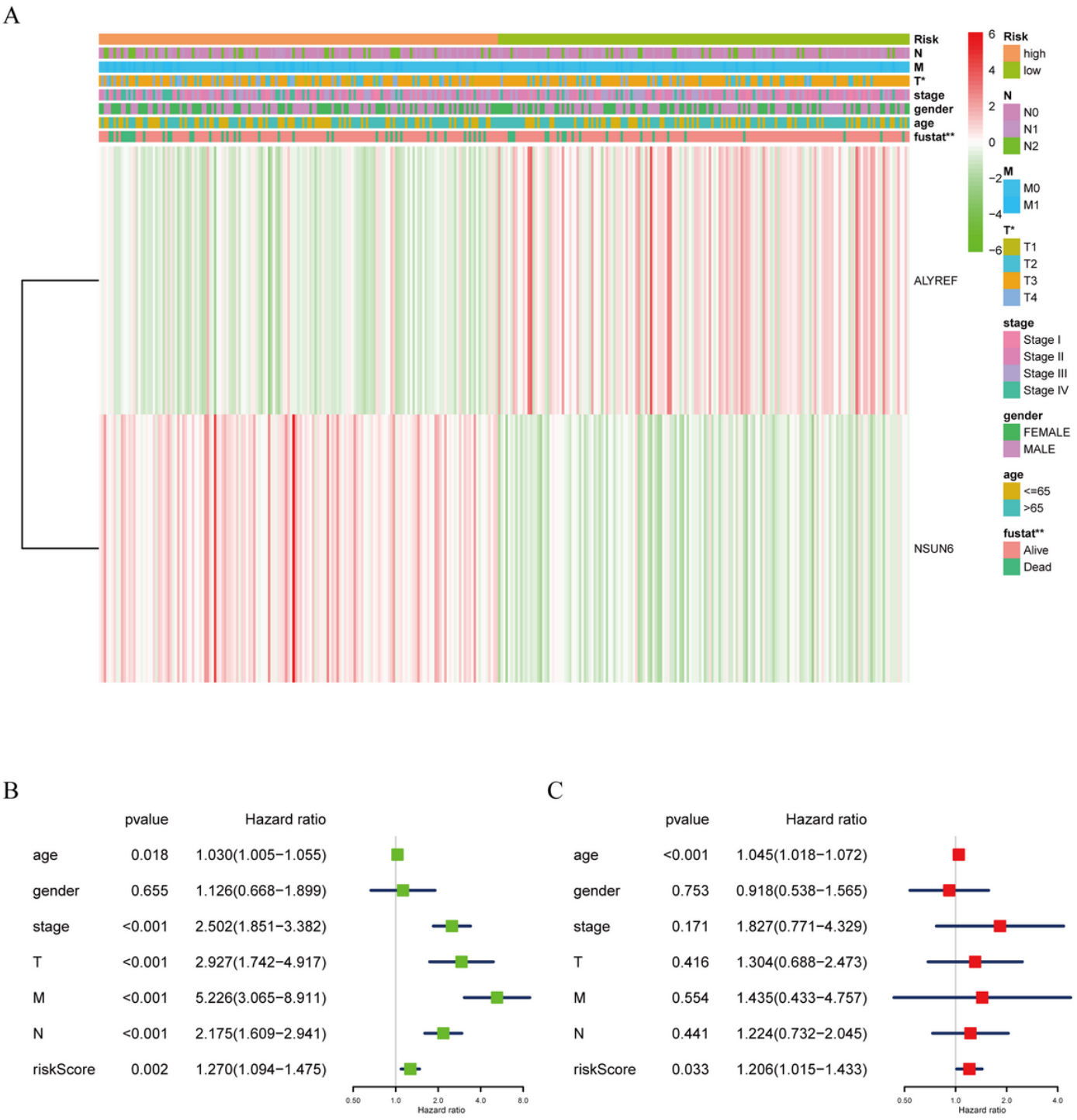

Figure 8

Relationship between the risk score,the expression of NSUN6 and ALYREF and clinicopathological characteristics. (A) The heatmap shows the expression of NSUN6 and ALYRE in high-risk and low-risk.The distribution of clinicopathological characteristics was compared between the high-risk and low-risk groups. * p <0.05, ** p <0.01. (B) Univariate Cox regression analysis of the associated between clinicopathological factors (including risk score) and overall survival of patients. (C) Multivariate Cox regression analysis of the associated between clinicopathological factors (including risk score) and overall survival of patients.

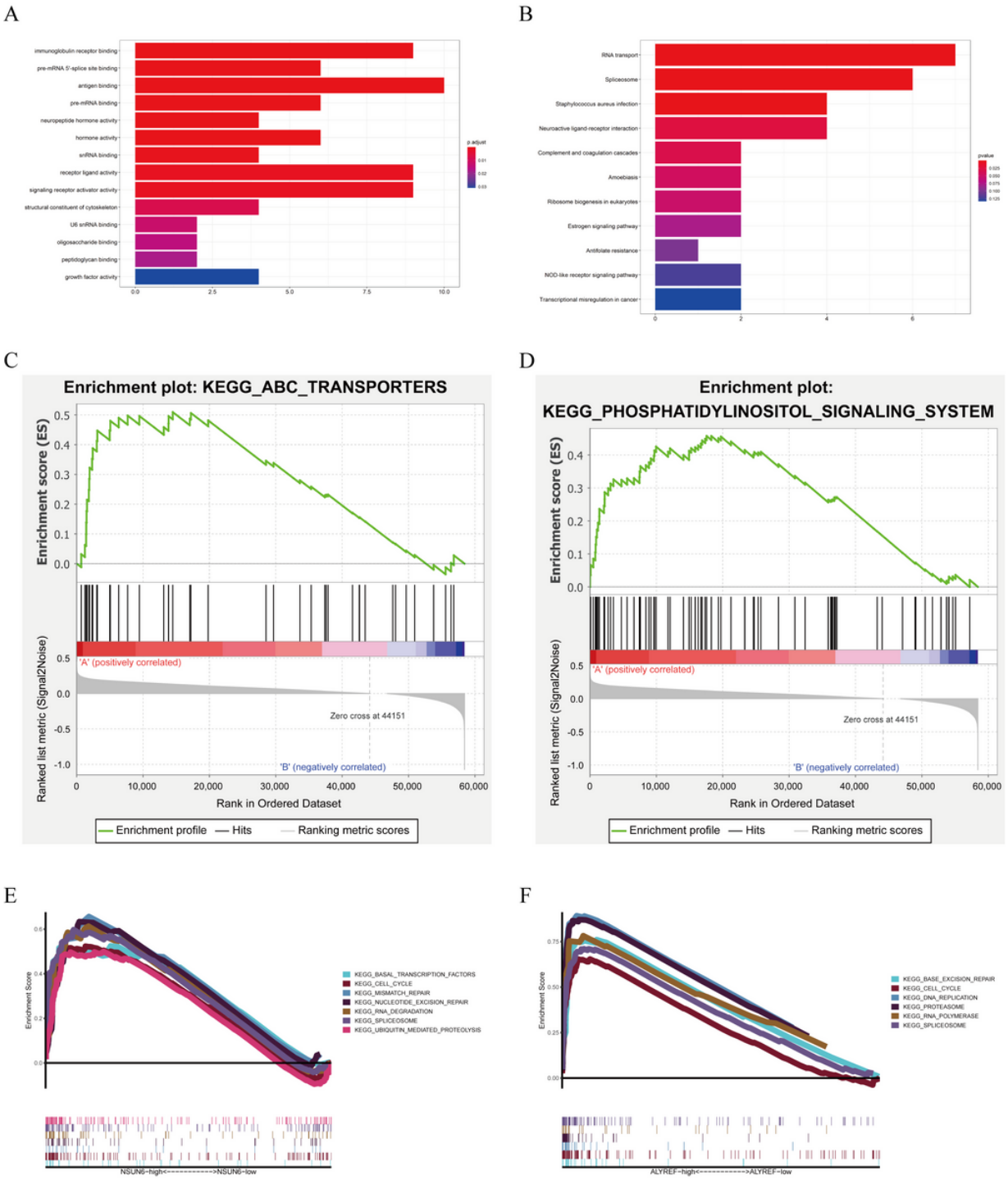

Figure 9

Biological functional analysis.(A-B) GO analysis and KEEG pathways analysisof the genes significantly upregulated or downregulated between cluster and cluster . (C-D) Cluster I had a worse overall survival (OS) and lower 5-year survival rate associated with malignancy-associated pathways, including the ATP-binding cassette transporter and phosphatidylinositol signaling system. (E) GSEA results for NSUN6. (F) GSEA results for ALYREF.