Observational Study

Medicine

OPEN

The Cancer Genome Atlas (TCGA) based m6A methylation-related genes predict prognosis in rectosigmoid cancer

Wei Zhou, MD[a], Junchao Lin, MD[a], Zeng Li[a], Min Li[b], Daiming Fan, MD, PhD[a], Liu Hong, MD, PhD[a,*]

Abstract

N6-methyladenosine (m6A) methylation plays an important role in the occurrence and development of tumors. This study aimed to explore the effects of m6A methylation regulatory genes on rectosigmoid cancer (RSC). RNA-seq data and related clinical information in The Cancer Genome Atlas database were analyzed. The Wilcoxon test was used to analyze the different expression levels of m6A methylation regulatory genes between the tumor and normal samples. Least absolute shrinkage and selection operator Cox regression analysis was used to construct a risk prognosis model between the m6A methylation regulatory genes and RSC. The median risk score was used to classify RSC patients into high and low-risk groups. Kaplan–Meier survival analysis and receiver operating characteristic curves were used to evaluate the sensitivity and specificity of the prediction model. The expression of m6A methylation regulation genes was different between the tumor and normal samples, 6 genes were overexpressed in tumor and 2 genes were down-regulated. Four m6A methylation regulatory genes, YTHDF3, KIAA1429, ALKBH5 and METTL3, were screened by least absolute shrinkage and selection operator Cox regression analysis. The overall survival of high-risk group was significantly lower than that of low-risk group ($P = 4.681 \times 10^{-4}$). The area under the curve value in the receiver operating characteristic curve was 0.935, indicating that the prediction model was effective. Univariate and multivariate Cox regression were used to test the effectiveness of the model. m6A methylation regulators YTHDF3, KIAA1429, ALKBH5, and METTL3 can be used to construct predictive models to predict overall survival in different clinical subgroups of RSC patients.

Abbreviations: CRC = colorectal cancer, ICD = International Classification of Diseases of Cancer, HCC = hepatocellular carcinoma, m6A = N6-methyladenosine, OS = overall survival, RNA = ribonucleic acid, ROC = receiver operating characteristic, RSC = rectosigmoid cancer, TCGA = the Cancer Genome Atlas database.

Key words: bioinformatics, clinical prognosis, m6A methylation, rectosigmoid cancer

1. Introduction

Colorectal cancer (CRC) is the third most common cancer in the world.[1] As the dividing line between colon and rectum, rectosigmoid junction is the common position of colon cancer. It is reported that sigmoid colon cancer and rectosigmoid cancer (RSC) account for 27% of colon cancer.[2] Although the rectosigmoid junction is anatomically considered to be the distal part of the sigmoid colon, it is considered to be part of the rectum because it shares an important vascular system with the rectum above the peritoneal reflex.[3] In fact, the Third Edition of the International Classification of Diseases of Cancer (ICD-O-3) stated that the rectosigmoid junction should be classified as a separate segment of the large intestine (ICD-O; C-19), rather than categorizing it as colon (ICD-O; C-18) or rectum (ICD-O; C-20).[4] The pathogenesis, treatment and prognosis of RSC may be different from rectal cancer and sigmoid colon cancer.[5] Therefore, the search for biological targets and molecular mechanisms that affect the prognosis of RSC patients is attracting more and more attention.

N6 methyladenine (m6A) refers to the methylation modification at the N6 position of adenine base. It is an RNA modification with the highest endogenous abundance and participates in almost all RNA metabolism processes including RNA transport, splicing, translation or degradation.[6–8] By regulating the expression of tumor related genes, m6A plays an important role in tumor development such as proliferation, invasion and metastasis.[9] Hua et al[10] found that as an m6A reader, YTHDF2 expression was elevated in patients with multiple myeloma and was associated with poor prognosis. Further studies have confirmed that YTHDF2 inhibits MAP2K2/P-ERK expression by mediating non-phosphorylated STAT5A, thereby promoting tumor proliferation, and targeting YTHDF2 may be a promising therapeutic strategy.

WZ and JL contributed equally to this work.

This work was supported by grants from the National Natural Science Foundation of China (no. 82073210).

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

[a] Xijing Hospital of Digestive Diseases, The Fourth Military Medical University, Xi'an, China, [b] Shaanxi University of Traditional Chinese Medicine, Xianyang, China.

**Correspondence: Liu Hong, Xijing Hospital of Digestive Diseases, The Fourth Military Medical University, Xi'an 710032, China (e-mail: hongliu1@fmmu.edu.cn).*

How to cite this article: Zhou W, Lin J, Li Z, Li M, Fan D, Hong L. The Cancer Genome Atlas (TCGA) based m6A methylation-related genes predict prognosis in rectosigmoid cancer. Medicine 2022;XX:XX(e32328).

Received: 5 April 2022 / Received in final form: 24 November 2022 / Accepted: 29 November 2022

http://dx.doi.org/10.1097/MD.0000000000032328

Although more and more studies have proved the role of m6A methylation in the occurrence and development of many tumors, the relationship between m6A methylation and RSC is still unclear, and there is no prognostic model based on m6A. In this study, The Cancer Genome Atlas (TCGA) data were used to analyze the expression differences of 23 m6A methylation regulatory genes in RSC and normal tissues, and 4 genes were selected by least absolute shrinkage and selection operator regression analysis to construct a risk model. Then verify its predictive role in RSC. This study found that m6A regulatory genes plays an important role in the occurrence and development of RSC, and can be used as an independent predictive gene to predict the prognosis of RSC patients, so as to provide predictive conditions for further individualized treatment.

2. Materials and Methods

2.1. The data collection

The transcriptome data and corresponding clinical data of RSC were obtained from TCGA database (https://portal.gdc.cancer). The mRNA expression data were collected from 66 tumors and 6 normal tissues, as well as clinical data from 66 patients with RSC.

2.2. m6A methylation regulates gene selection

Twenty-three recognized m6A regulatory genes were selected: There are eight 23 putative m6A regulatory genes were selected, including 8 writers (METTL3, METTL14, METTL16, WTAP, KIAA1429/VIRMA, ZC3H13, RBM15, and RBM15B), 13

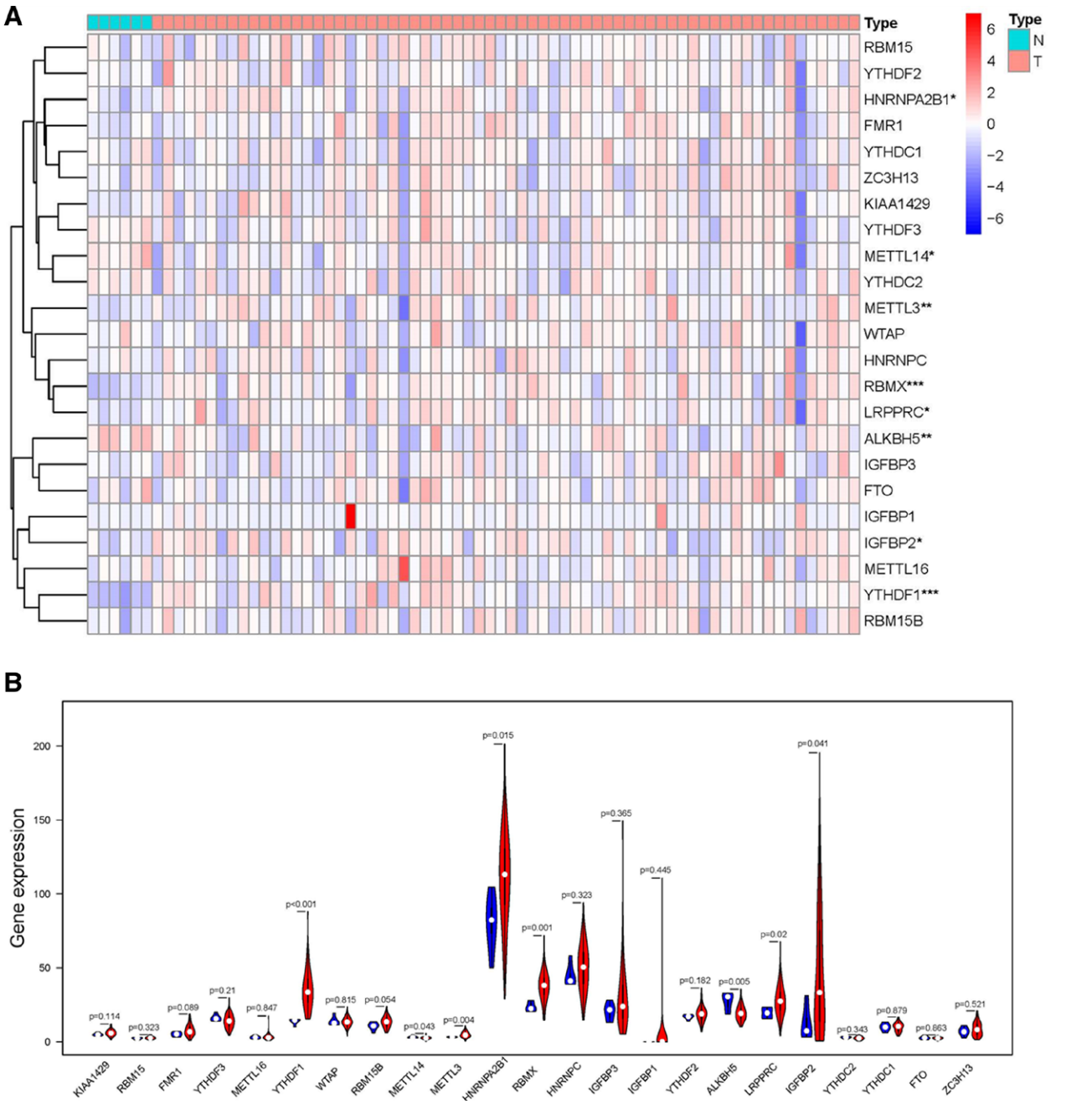

Figure 1. Expression of m6A methylation regulator genes in RSC. The expression of m6A methylation regulators in RSC heat map (A) and violin plots (B) of 23 m6A RNA methylation regulators between RSC tumor samples and normal samples. *$P < .05$, **$P < .01$, ***$P < .001$. RSC = rectosigmoid cancer.

readers (YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, HNRNPA2B1, HNRNPC, IGFBP1, IGFBP2, IGFBP3, FMR1, LRPP, RC, and RBMX), and 2 Erasers (ALKBH5 and FTO). The expression data were obtained from the TCGA database, and the differential gene screening was performed in normal and tumor tissues using R's limma package and Wilcox test, and clinicopathological data were compared.

2.3. Bioinformatics analysis

Biological information was analyzed using R language software. The interactions between m6A RNA methylation regulators were analyzed using STRING database (http://www.string-db.org/). The correlation analysis of 23 m6A regulation genes was performed using pheatMap package and Pearson correlation analysis using corrplot package. ConsensusClusterPlu package was used to identify different subgroups of clinical samples based on the comprehensive expression of 23 genes, and principal component analysis was used to verify the grouping results. The survival curves of the 2 subgroups were plotted by KM method. Least absolute shrinkage and selection operator Cox risk regression analysis was performed for 23 m6A regulatory genes using glmnet package, and the risk scores were calculated as follows:

$$\text{Risk} \quad \text{score} = \sum_{i=1}^{n} Coefi \times xi$$

Where Coefi is the coefficient, Xi is the expression value of each selected gene, and this formula is used to calculate the risk score of each patient. Four genes related to prognosis were selected, and the patients were divided into high and low risk groups according to the risk value. Survival package was used for survival analysis of high and low risk groups and survival ROC package was used for receiver operating characteristic (ROC) analysis to verify the validity of the model. Cox regression analysis was used to analyze the relationship between risk value and age, gender, stage, T stage, N stage and M stage. The forest map was drawn with forest plot package.

2.4. Statistical analysis

All statistical analyses were performed using R software (version 4.0.3) and Cytoscape software (version 3.7.2). Wilcoxon test was used to analyze the differential expression of m6A methylation regulation genes in tumor samples and normal samples. The cutoff value that distinguishes patients as high and low risk is the risk value obtained above. Kaplan–Meier method was used to analyze the overall survival (OS), and Chi-square test was used to analyze risk value and clinical characteristic variables. Finally, the constructed risk model was validated by survival analysis in clinical subgroups. $P < .05$ was considered statistically significant.

2.5. Ethical statements

As all data were obtained from public database, this study did not require ethical approval.

3. Results

3.1. Expression of m6a RNA methylation regulator in RSC

Heat maps were used to visualize the expression of 23 m6A methylation regulators in RSC samples and normal samples. Asterisks indicate differentially expressed m6A methylation regulators ($*P < .05$, $**P < .01$, $***P < .001$). In the heat map, red indicates high expression and blue indicates low expression (Fig. 1A). Violin plots showed that YTHDF1, RBMX, METTL3, HNRNPA2B1, LRPPRC and IGF2BP1 were significantly increased in tumor samples, while ALKBH5 and METTL14 were significantly lower in tumor samples than in normal samples, while other molecules showed no significant changes in RSC data (Fig. 1B).

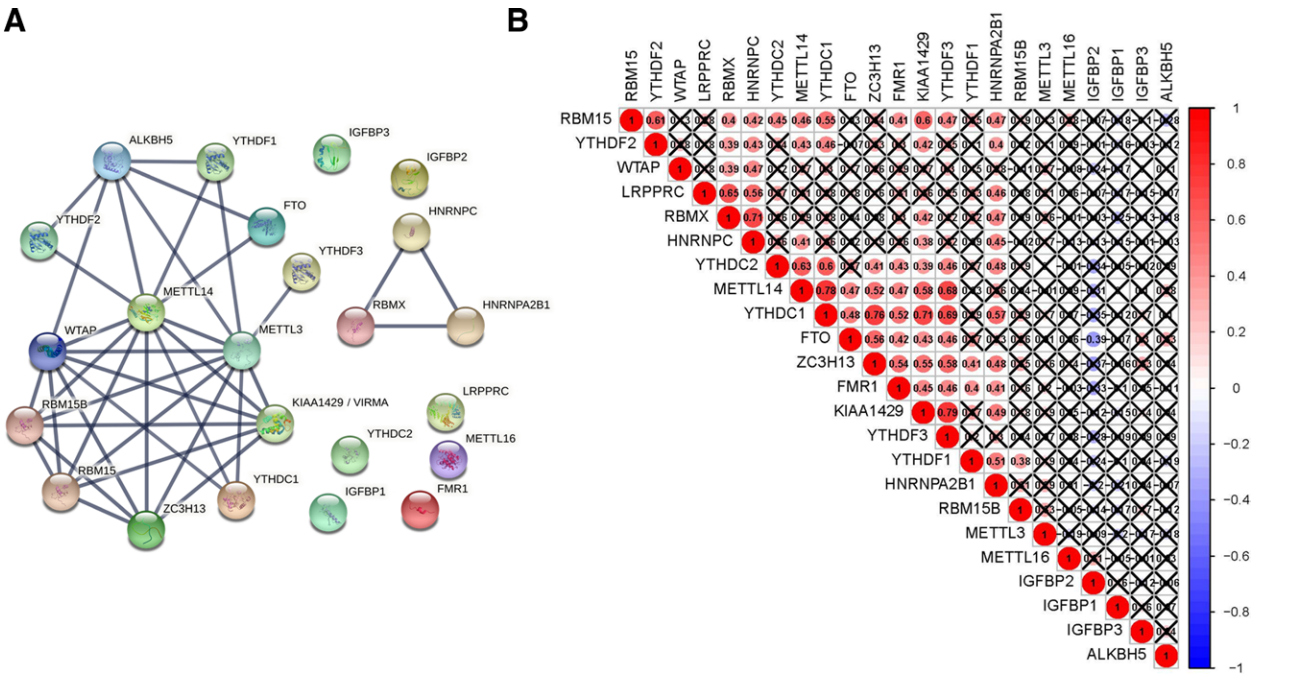

Figure 2. Interaction and correlation between 23 m6A methylation regulators in RSC. (A) Interaction network between m6A methylation regulators. (B) Pearson correlation between m6A RNA methylation regulators. RSC = rectosigmoid cancer.

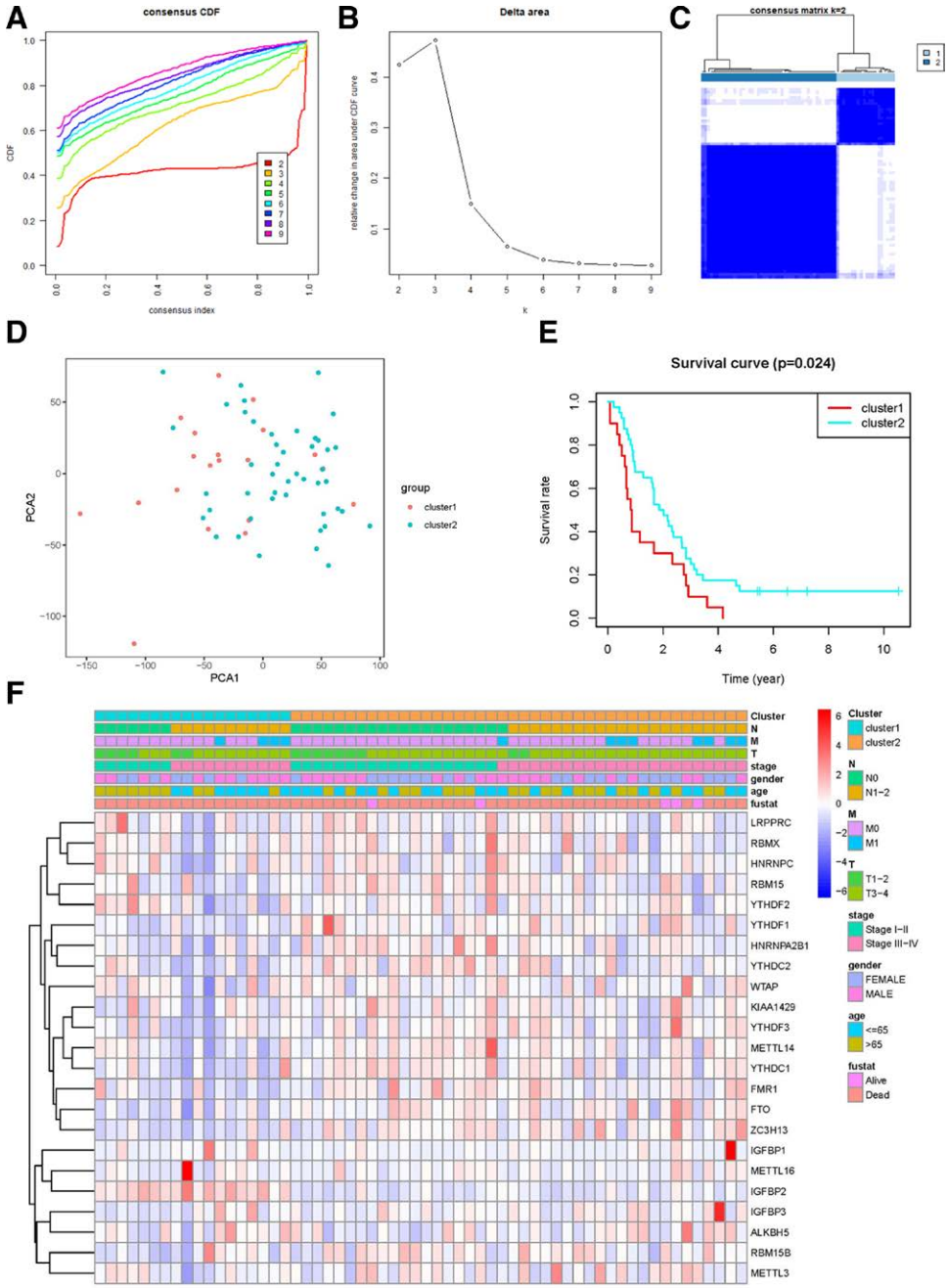

Figure 3. Difference in OS rate and grade of cancer genome map between 2 groups of patients with different RSC. (A) Consensus clustering CDF with $k = 2 – 9$. (B) Relative change of area under CDF curve when $k = 2 – 9$. (C) When $k = 2$, the TCGA queue is divided into 2 different clusters. (D) PCA of the 2 subgroups. (E) Kaplan Meier survival curves of the 2 subgroups showed that the OS rate of group 1 was significantly lower than that of group 2 ($P < .05$). (F) Heat map showed the clinicopathological features of these 2 subgroups. CDF = cumulative distribution function, OS = overall survival, PCA = principal component analysis, RSC = rectosigmoid cancer, TCGA = The Cancer Genome Atlas.

3.2. Interaction between expression of m6a RNA methylation regulators

We submitted the information of 23 m6A regulators to the STRING interaction network analysis, set the truncated confidence interval to 0.900, and the results showed the correlation network of 23 m6A regulators (Fig. 2A), in which METTL14 appeared to be the hub gene in the expression network. Pearson correlation analysis showed the correlation between m6A regulators. As can be seen from the figure, KIAA1429 had the highest correlation with YTHDF3, with a correlation coefficient of 0.79. YTHDC1 was also highly correlated with METTL14, ZC3H13, KIAA1429 and YTHDF3, with correlation coefficients of 0.78, 0.76, 0.71 and 0.69, respectively (Fig. 2B).

3.3. Consensus - clustering analysis

In consensus cluster analysis, RSC samples were divided into 2 clusters according to the similarity of the expression of 23 m6A regulators ($k = 2$; Fig. 3A–C). Principal component analysis shows that this classification method can effectively distinguish samples (Fig. 3D). According to Kaplan–Meier survival analysis, there was a significant difference in OS rate between Cluster1 and Cluster2, $P = .024$ (Fig. 3E). This suggests that these m6A

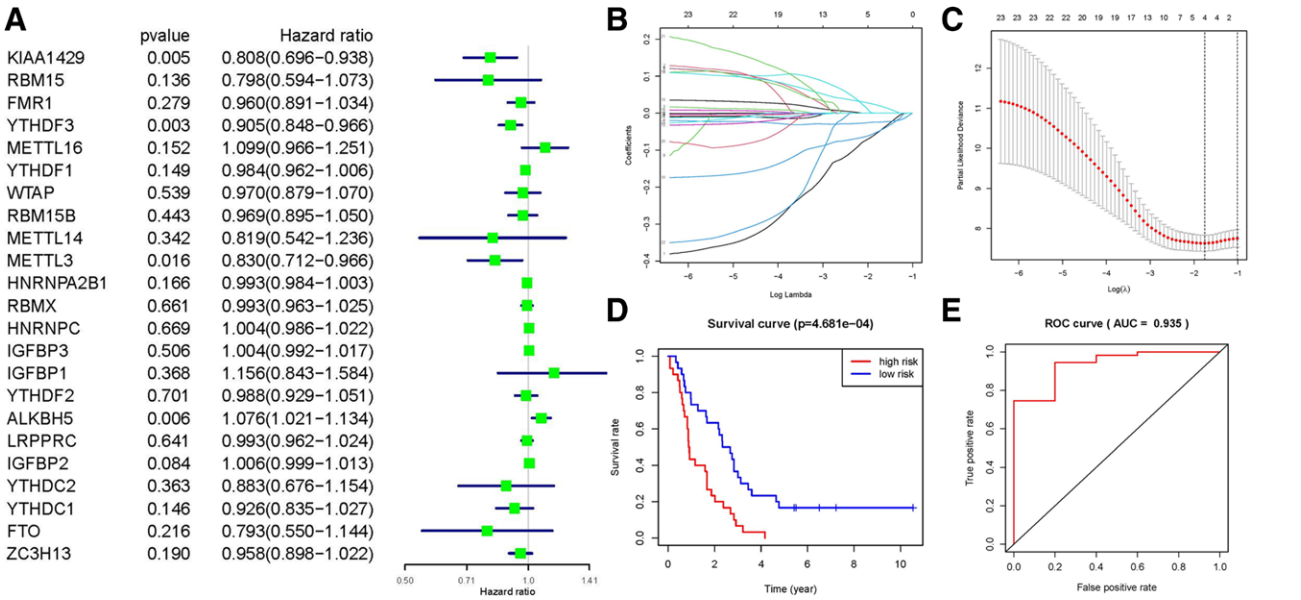

Figure 4. Gene selection and survival analysis in prognosis prediction of RSC. (A) Hazard ratio forest map of m6A methylation regulators associated with RSC survival. (B) Coefficients of selected features are shown by lambda parameter. (C) Partial likelihood deviation versus log (λ) was plotted using LASSO Cox regression model. (D) Relationship between risk score and OS in patients with RSC. (E) ROC curves of RSC survival model (AUC = 0.935). AUC = area under the curve, LASSO = least absolute shrinkage and selection operator, OS = overall survival, ROC = receiver operating characteristic, RSC = rectosigmoid cancer.

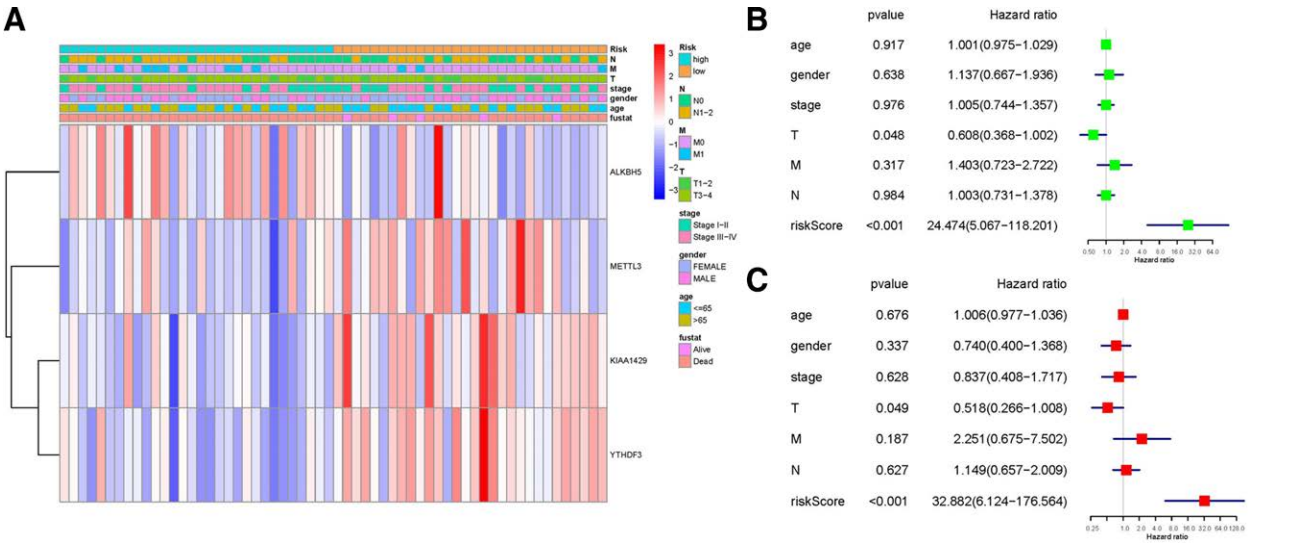

Figure 5. Effects of risk score and clinicopathological variables on the prognosis of patients with RSC. (A) Heat map shows the expression of 3 m6A methylation regulators and the distribution of clinicopathological variables between high-risk and low-risk groups. (B) Cox univariate analysis of clinicopathological variables (including risk score) and OS. (C) Cox multivariate analysis of clinicopathological variables (including risk score) and OS. OS = overall survival, RSC = rectosigmoid cancer.

molecules can be used to classify RSCS at a prognostic level. Figure 3F shows the clinicopathological features of these 2 clusters by heat map.

3.4. Prognostic role of m6a RNA methylation regulator

To establish the risk profile associated with m6A methylation regulators, univariate Cox regression analysis was used to assess the relationship between 23 m6A regulators and OS. We found that YTHDF3, KIAA1429, ALKBH5 and METTL3 were significantly correlated with OS ($P < .05$; Fig. 4A). To better explain this internal relationship, we performed LASSO logistic regression analysis on these 4 molecules (Fig. 4B and C). We constructed the risk profile of m6A methylation regulators composed of YTHDF3, KIAA1429, ALKBH5 and METTL3. Kaplan–Meier survival analysis was used to analyze RSC patients in the high-risk and low-risk groups. We found significant improvement in OS in the low-risk group compared to the high-risk group (Fig. 4D; $P = 4.681 \times 10^{-4}$). The area under the curve of 5-year ROC for this risk feature was 0.935, indicating that this risk feature had good specificity and sensitivity (Fig. 4E).

3.5. Value at risk in RSC patients

We further demonstrated the relationship between different risk subgroups and clinicopathological features (Fig. 5A). Finally, we included different clinicopathological features and risk scores based on m6A methylation regulators in univariate and

multivariate cox regression analyses. Univariate and multivariate prognostic analyses showed that tumor T stage and riskScore were significantly correlated with OS (Fig. 5B and C; $P < .05$).

4. Discussion

CRC is the second leading cause of cancer-related death, with about 1.84 million new cases and 900,000 deaths each year, and RSC accounts for about a quarter of these patients.[1] The incidence rate of RSC is increasing in young adults, whether male or female,[11] which undoubtedly bring difficulties to clinical research. Tumorigenesis is a process of multi-gene and multi-stage gradual evolution, which is a pathological process of oncogene activation and tumor suppressor gene inactivation.[12] More and more studies have shown that m6A methylation regulators play an important role in inflammation, tumor immunity and antitumor therapy.[13–15] Therefore, it is necessary to investigate the effect of m6A methylation regulators on RSC.

In this study, we first analyzed the expression of m6A methylation regulators in RSC and normal tissues, and the relationship between their expression and different clinicopathological variables. The expression of m6A methylation regulator was different from that of different clinicopathological variables in patients with RSC. YTHDF3 overexpression is clinically associated with brain metastasis in breast cancer patients, which can promote the interaction between cancer cells and brain endothelial cells and astrocytes, as well as blood-brain barrier extravasation and angiogenesis. Researchers further demonstrated the potential of YTHDF3-targeted therapy for life-threatening brain metastasis through the data of stable ablation of YTHDF3 to inhibit brain metastases.[16] In CRC, YTHDF3 participates in the YAP signal transduction process, and reversibly and selectively binds TO m6A methylation of GAS5 to trigger its decay and form a negative feedback loop, thus providing a promising method for the treatment of CRC.[17] As an important component of m6A methyltransferase complex, KIAA1429 is significantly upregulated in hepatocellular carcinoma (HCC), and its high expression is associated with poor prognosis of HCC patients, while silencing KIAA1429 can inhibit cell proliferation and metastasis in vitro and in vivo.[18] In the study of CRC, it was found that KIAA1429 plays a carcinogenic role in CRC cells by inhibiting the expression of WEE1 in an m6A independent manner, and is associated with the poor prognosis of CRC patients. The above results suggest that KIAA1429 may be a potential prognostic marker of CRC.[19] ALKBH5, another m6A methyltransferase, is down-regulated in gastric cancer and is associated with distant metastasis and lymph node metastasis. Further studies showed that ALKBH5 regulated the expression of PKMYT1 in an m6A dependent manner, thereby affecting the invasion and migration of gastric cancer cells. Therefore, ALKBH5 represents a new therapeutic target for gastric cancer metastasis.[20] Another study found that ALKBH5 was down regulated in HCC, and the decreased expression of ALKBH5 was an independent prognostic factor for the reduced survival rate of HCC patients. ALKBH5, as a tumor suppressor, reduces the expression of LYPD1 in HCC cells in an m6A dependent manner, thereby inhibiting the proliferation and invasion of HCC cells in vitro and in vivo, providing new insights into potential biomarkers and therapeutic targets for HCC treatment.[21] METTL3 is an effective therapeutic target for a variety of cancers, such as CRC, myeloid leukemia and breast cancer.[22–24] METTL3 is overexpressed in prostate cancer, which is associated with poor prognosis of patients, and the migration and invasion ability of tumor cells are significantly inhibited after METTL3 knockdown.[25]

In recent years, the prognosis prediction model based on m6A related genes has been used to evaluate and predict the prognosis of patients with various tumors and guide the formulation of subsequent individualized treatment plans.[26–29] In this study, a prognostic prediction model of m6A-related genes in RSC was established by LASSO regression analysis of m6A-related genes significantly differentially expressed in RSC tissues. Kaplan–Meier survival curve showed that the 5-year survival rate of low-risk group was significantly better than that of high-risk group, and the time-dependent ROC curve showed that the model had a good effect in predicting the prognosis and survival of RSC patients (area under the curve = 0.935). The follow-up study proved that the model can be used as an independent prognostic factor through univariate Cox regression analysis. The limitation of this study is that our study is based on the patient data in TCGA and GEO databases and has not been verified by an independent cohort. Therefore, more clinical data need to be included in the future to constantly revise and improve this study.

5. Conclusion

In summary, this study constructed a risk scoring model for predicting the prognosis of RSC patients based on 4 m6A-related genes. This model has good sensitivity and specificity, and can be used as an independent prognostic factor to evaluate the prognosis of RSC patients, which has important reference value for the development of reasonable and effective individualized treatment plan. However, there were some limitations to the present study. This study only uses the data of TCGA database to verify the Cox regression model, and we should also practice in other public databases. Moreover, the application of this model still needs the support of a large number of clinical research data and the verification of large-scale, multi-center evidence-based medical evidence.

Author contributions

Conceptualization: Liu Hong, Daiming Fan.
Data curation: Min Li, Liu Hong.
Methodology: Wei Zhou, Daiming Fan.
Software: Junchao Lin, Min Li.
Supervision: Liu Hong.
Validation: Junchao Lin, Zeng Li.
Writing – original draft: Wei Zhou, Zeng Li, Liu Hong.

References

[1] Grady WM, Yu M, Markowitz SD. Epigenetic alterations in the gastrointestinal tract: current and emerging use for biomarkers of cancer. Gastroenterology. 2020;160:690–709.
[2] Ng OH, Lunevicius R, Arthur JD. Rising incidence of colorectal cancer in individuals younger than 50 years and increasing mortality from rectosigmoid cancer in England. Colorectal Dis. 2021;23:2637–46.
[3] Mukai M, Kishima K, Yamazaki M, et al. Stage II/III cancer of the rectosigmoid junction: an independent tumor type? Oncol Rep. 2011;26:737–41.
[4] Guan X, Jiang Z, Ma TY, et al. Radiotherapy dose led to a substantial prolongation of survival in patients with locally advanced rectosigmoid junction cancer: a large population based study. Oncotarget. 2016;7:28408–19.
[5] Wang L, Lo CH, He XS, et al. Risk factor profiles differ for cancers of different regions of the colorectum. Gastroenterology. 2020;159:241–56.e13.
[6] He LE, Li HY, Wu AQ, et al. Functions of N6-methyladenosine and its role in cancer. Mol Cancer. 2019;18:176.
[7] Wang TY, Kong S, Tao M, et al. The potential role of RNA N6-methyladenosine in Cancer progression. Mol Cancer. 2020;19:88.
[8] Dixit D, Prager BC, Gimple RC, et al. The RNA m6A reader YTHDF2 maintains oncogene expression and is a targetable dependency in glioblastoma stem cells. Cancer Discov. 2021;11:480–99.
[9] Liu ZX, Li LM, Sun HL, et al. Link between m6A modification and cancers. Front Bioeng Biotechnol. 2018;6:89.
[10] Hua Z, Wei RF, Guo MJ, et al. YTHDF2 promotes multiple myeloma cell proliferation via STAT5A/MAP2K2/p-ERK axis. Oncogene. 2022;41:1482–91.
[11] Yang L, Liu SS, Xiong ZC, et al. Changes in colorectal cancer incidence by site and age from 1973 to 2015: a SEER database analysis. Aging Clin Exp Res. 2021;33:1937–46.

[12] Liu ZY, Lin Y, Gao X, et al. Construction of a comprehensive multiomics map of hepatocellular carcinoma and screening of possible driver genes. Front Genet. 2020;11:634.
[13] Hou JJ, Zhang H, Liu J, et al. YTHDF2 reduction fuels inflammation and vascular abnormalization in hepatocellular carcinoma. Mol Cancer. 2019;18:163.
[14] Zhang HJ, Hu J, Liu AN, et al. An N6-methyladenosine-related gene set variation score as a prognostic tool for lung adenocarcinoma. Front Cell Dev Biol. 2021;9:651575.
[15] Liu YD, Zhang QM, Xing BC, et al. Immune phenotypic linkage between colorectal cancer and liver metastasis. Cancer Cell. 2022.
[16] Chang GQ, Shi L, Ye YQ, et al. YTHDF3 induces the translation of m 6 A-enriched gene transcripts to promote breast cancer brain metastasis. Cancer Cell. 2020;38:857–71.e7.
[17] Ni W, Yao S, Zhou YX, et al. Long noncoding RNA GAS5 inhibits progression of colorectal cancer by interacting with and triggering YAP phosphorylation and degradation and is negatively regulated by the m 6 A reader YTHDF3. Mol Cancer. 2019;18:143.
[18] Lan T, Li H, Zhang DL, et al. 1429 contributes to liver cancer progression through N6-methyladenosine-dependent post-transcriptional modification of GATA3. Mol Cancer. 2019;18:186.
[19] Ma L, Lin Y, Sun SW, et al. KIAA1429 is a potential prognostic marker in colorectal cancer by promoting the proliferation via downregulating WEE1 expression in an m6A-independent manner. Oncogene. 2022;41:692–03.
[20] Hu YY, Gong CL, Li ZB, et al. Demethylase ALKBH5 suppresses invasion of gastric cancer via PKMYT1 m6A modification. Mol Cancer. 2022;21:34.
[21] Chen YH, Zhao YC, Chen JR, et al. ALKBH5 suppresses malignancy of hepatocellular carcinoma via m 6 A-guided epigenetic inhibition of LYPD1. Mol Cancer. 2020;19:123.
[22] Pan JS, Liu F, Xiao XL, et al. METTL3 promotes colorectal carcinoma progression by regulating the m6A-CRB3-Hippo axis. J Exp Clin Cancer Res. 2022;41:19.
[23] Yankova E, Blackaby W, Albertella M, et al. Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. Nature. 2021;593:597–601.
[24] Wan WJ, Ao X, Chen Q, et al. METTL3/IGF2BP3 axis inhibits tumor immune surveillance by upregulating N 6-methyladenosine modification of PD-L1 mRNA in breast cancer. Mol Cancer. 2022;21:60.
[25] Chen YB, Pan C, Wang XT, et al. Silencing of METTL3 effectively hinders invasion and metastasis of prostate cancer cells. Theranostics. 2021;11:7640–57.
[26] Li W, Gao YC, Jin XJ, et al. Comprehensive analysis of N6-methylandenosine regulators and m6A-related RNAs as prognosis factors in colorectal cancer. Mol Ther Nucleic Acids. 2021;27:598–10.
[27] Song W, Ren J, Yuan WZ, et al. N6-methyladenosine-related lncRNA signature predicts the overall survival of colorectal cancer patients. Genes (Basel). 2021;12:1375.
[28] Zhang YY, Zhu BY, He MH, et al. N6-methylandenosine-related lncRNAs predict prognosis and immunotherapy response in bladder cancer. Front Oncol. 2021;11:710767.
[29] Zhao JH, Lin X, Zhuang JM, et al. Relationships of N6-methyladenosine-related long non-coding RNAs with tumor immune microenvironment and clinical prognosis in lung adenocarcinoma. Front Genet. 2021;12:714697.