The Role of RNA Methylation Modification Related Genes in Prognosis and Immunotherapy of Colorectal Cancer

Ruoyang Lin  
The First Affiliated Hospital of Wenzhou Medical University

Renpin Chen  
The First Affiliated Hospital of Wenzhou Medical University

Lechi Ye  
The First Affiliated Hospital of Wenzhou Medical University

Zhiming Huang  
The First Affiliated Hospital of Wenzhou Medical University

Xianfan Lin  
The First Affiliated Hospital of Wenzhou Medical University

Tanzhou Chen (✉ cctzz@163.com)  
The First Affiliated Hospital of Wenzhou Medical University

Research Article

Keywords:

Posted Date: December 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2314779/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

This study aimed to analyze the effects of RNA methylation regulatory genes in prognosis and treatment of CRC. Prognostic signature associated with CRCs were constructed by differential expression analysis, Cox and LASSO analyses. ROC and Kaplan-Meier survival analyses were used to validate the reliability of the developed model. Finally, normal and cancerous tissue were collected to validate gene by qRT-PCR. A prognostic risk model based on LRPPRC and UHRF2 was constructed and relevant to the OS of CRC. Ultimately, PCR validation showed that a significant upregulation in the expression of LRPPRC and UHRF2 in cancerous tissue. A prognostic risk model based on LRPPRC and UHRF2 was successfully built, and the model could predict the immunotherapy efficacy of CRC.

1. Introduction

CRC is one of the leading causes of death, which has high lethality rate worldwide. Despite the rapid development of diagnosis and treatment techniques for CRC in recent years, each therapeutic method, such as chemotherapy drugs, targeted drugs, immunotherapy, still has its own limitations, and the prognosis of CRC cases is poor, with only 60% a 5-year survival rate[1]. In addition, there are still few prognostic markers for CRC. Therefore, investigating the diagnosis, prognostic evaluation and molecular markers related to clinical features of CRC is very meaningful[2].

Epigenetic modification includes DNA methylation, histone modification and RNA methylation[3]. These modifications greatly enrich the diversity of RNA functions and genetic information[4]. At present, RNA methylation is a research hotspot, which is a post-transcriptional level regulation mode. RNA methylation modification includes a variety of modification modes, including N\textsuperscript{6}-methyladenosine (m6A), N\textsuperscript{5}-methylcytosine (m5C), N\textsuperscript{1}-methyladenosine (m1A) and N\textsuperscript{7}-Methylguanosine (m7G), which is important in RNA metabolism, cell differentiation and protein production. Studies have shown that RNA methylation participated in the regulation of the development, invasion, and metastasis of tumors[5, 6]. The dysregulation and mutation of RNA modification genes are relevant to the occurrence of many cancers and can impact the prognosis of tumors[7, 8]. For example, methyltransferase-like 3 (METTL3) is an important m6A regulatory enzyme in the development of CRC. METTL3 contributed to the initiation and advancement of CRC by inducing the M6A- glucose transporter 1 (GLUT1) - mechanistic target of rapamycin complex 1 (mTORC1) axis[9]. In addition, the highly expressed NOP2/Sun RNA methyltransferase 5 (NSUN5) promoted tumor cells proliferation by controlling the cell cycle in CRC[10]. However, a comprehensive analysis of the effect of methylation regulatory enzymes on the prognosis of CRC has not been reported.

Tumor immunotherapy has received increasing attention in recent years. Immune checkpoint inhibitor is a treatment for many advanced malignancies, offering a new paradigm of cancer treatment through the immune system of the patient. RNA methylation regulates RNA splicing, translocation and translation efficiency and has a critical effect in immune cell proliferation, differentiation and immune function. Therefore, studying RNA modification regulatory genes can better understand the process of tumor
development and better guide tumor treatment [11]. In this paper, we downloaded CNV, single nucleotide variation (SNV) and mRNA data of CRC, and then analyzed the mutations of m1A, m5C, m6A and m7G regulatory genes and CNV in CRC cases. Then, a survival risk model was constructed based on m1A, m5C, m6A and m7G, which was verified by clinical specimens. This paper was the first comprehensive analysis of the effects of m1A, m5C, m6A, and m7G-related methylated genes in the prognosis of CNV, SNV, and transcriptome data, and to explore the correlation with immunotherapy and tumor microenvironment.

2. Materials And Methods

2.1 Data sources

The TCGA CRC RNA sequencing, CNV, and somatic mutation data were acquired from TCGA database (https://portal.gdc.cancer.gov/). 365 CRC cases were collected from TCGA database construction after removing samples with no survival status, survival time and survival time of 0 (or negative). Moreover, transcriptome data of 365 cases, CNV data of 362 cases, and somatic mutation data of 343 cases were available. Similarly, the RNA-seq data and clinical information of 556 CRC cases in the GSE39582 dataset which was acquired from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) as an external validation cohort.

2.2 RNA methylation regulatory genes

By searching the published literature, 54 RNA methylation regulatory genes were identified. In accordance with previous reports, genes TRMT10C, TRMT61B, TRMT6, TRMT61A, ALKBH3, ALKBH1, YTHDC1, YTHDF1, YTHDF2 and YTHDF3 were acquired as regulators of m1A[7, 12]. Some m5C regulatory genes were collected, including DNMT1, DNMT3A, DNMT3B, MBD1, MBD2, MBD3, MBD4, MECP2, NEIL1, NTHL1, SMUG1, TDG, UHRF1, UHRF2, UNG, ZBTB33, ZBTB38, ZBTB4, TET1, TET2 and TET3[13, 14]. 21 m6A regulators were acquired, comprising METTL3, METTL14, RBM15, RBM15B, WTAP, KIAA1429, CBLL1, ZC3H13, ALKBH5, FTO, YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, IGF2BP1, HNRNPA2B1, HNRNPC, FMR1, LRPPRC and ELAVL1[15]. METTL1, WDR4, WBSCR22, TRMT112, NCBP2 and NCBP1 were used as m7G-related regulators.

2.3 Building and proving of the prognostic gene signature

To expound the prognostic correlation between RNA methylation genes and CRC cases, the samples from TCGA-CRC database were classified a training cohort (n = 256) and an internal validation cohort (n = 109) with7:3 at random. The GEO dataset GSE39582 (CRC samples n = 556) was used as external validation cohort to verify the prognostic signature. In TCGA-CRC training cohort, the prognostic of methylation regulatory genes was gotten by a univariate Cox regression analysis (R package “survival”, version 3.2-3). The LASSO was used to get gene signature. The risk score of each case was computed in TCGA-CRC training, internal and GEO validation cohorts by following formula. Cox regression analysis were used to determine independent prognostic factors, such as age, gender, stage, T, N, M stages and risk score.
Riskscore = \sum_{i=1}^{n} \text{Expr}_{\text{gene}(i)} \times \text{Coeff}_{\text{gene}(i)}

2.4 Nomogram

To study the prognostic value of the model, the CRC patients survival prediction nomogram was constructed by R package “rms” (version 6.2-0), and the risk model factors were scored separately. The scores were added to obtain an overall score, which used to predict survival probability. And the higher scores associated with lower survival probability. Then constructed the correction curve of the nomogram and observed its prediction accuracy.

2.5 Functional enrichment analysis

To explore the function of genes, functional enrichment analysis was implemented. DEGs between high and low risk cohort were gotten by differential expression analysis. The “clusterProfiler” (Version 3.18.0) R package was used for enrichment analysis based on KEGG and GO databases to find the common functions and related pathways of differentially expressed genes (DEGs). We used GSVA package (Version 1.38.2) to obtain the enrichment score of the samples in each pathway. Then, the differences in pathways between the high and low risk groups were gotten by t-test.

2.6 Estimation of immune cell infiltration and immune therapy response

The immune infiltration levels were analyzed via ESTIMATE and CIBERSORT. Rank sum test used to compare the expression of 9 kinds of immune checkpoint in high-low risk groups. The immunotherapy response between high and low risk groups were contrasted by Submap (https://cloud.genepattern.org/gp).

2.7 Validation of the gene expression

Then we collected 10 pericarcinomatous tissues and 10 cancer tissues from CRC patients in the First Affiliated Hospital of Wenzhou Medical University for qRT-PCR validation, and this study was consented to by all participants. This study was approved by the Ethics committee at the XX hospital. Total RNA extracted using TRIzol Reagent (Ambion, USA). Total RNA reverse transcribed to cDNA using sweScript RT I First strand cDNA Synthesis All-in-OneTM First-Strand cDNA Synthesis Kit (Servicebio, China), and qRT-PCR was conducted using 2xUniversal Blue SYBR Green qPCR Master Mix (Servicebio, China). We used GAPDH as an internal control. The forward primer of LRPPRC was “ATCCACCAACATAAGTTTCCC”. The reverse primer of LRPPRC was “CAGTCTCGCCTTTCTCTACCA”. The forward primer of UHRF2 was “CTGGCTCCTTGGACCTCTG”. The reverse primer of UHRF2 was “TGGGCTGCTTCTTTGACTG”. The forward primer of GAPDH was “CCCATCACCATCTTCCAGG”. The reverse primer of GAPDH was “CATCACGCCACAGTTCCTCC”.

3. Results
3.1 The landscape of genetic variations of RNA methylation genes in CRC

We analyzed and visualized the mutation frequency of 53 RNA methylation regulatory genes (TRMT10C without data). For the SNV, we found that the most frequent mutation of 53 genes was missense mutation. Mutation frequencies of the top 5 genes were ZC3H13, TET1, TET3, KIAA1429 and YTHDC2 (Fig. 1A, 1B). Moreover, as shown in Fig. 1C, the amplification frequencies of YTHDF1, DNMT3B, ZC3H13, HNRNPA2B1 and KIAA1429 were the highest. The genes MBD2, MBD1, ZBTB4, ALKBH5 and NEIL1 showed higher frequency copy number deletion.

3.2 Correlation analysis of gene expression and somatic mutation & CNV patterns

In order to obtain genes significantly associated with somatic mutations, we analyzed the relationship between somatic mutations and the expression levels of RNA methylation regulatory genes. As can be seen from the Fig. 2A, the expressions of DNMT1, METL1, METL14, NCBP1, RBM15, TET12, ZBTB38, ZC3H13 were significantly correlated with their mutations. Results of genes which not significantly related to mutation type were shown in the Figure S1. The differences of gene expression in the different CNV groups, as shown in Fig. 2B. The abundance of some genes (such as ALBKH1, CBLL1, DNMT1, ALBKH5 and ELAVL1, Etc.) were significantly correlated with the occurrence of CNV. And results of genes which not significantly related to CNVs were shown in the Figure S2.

3.3 Relationship between RNA methylation genes and clinical stages

The gene expression at different T, N, M stages were compared by rank-sum test. It can be seen that expression of genes CBLL1, HNRNPC, ZBTB38 and YTHDC2 were significantly correlated with T stages in Fig. 3A. The differences in gene expression between N0 and N+ (N+ vs. N0, N+ is the staging except N0) were analyzed and compared, we found that the expressions of DNMT3A, FTO, IGF2BP1, KIAA1429, LRPPRC, MBD2, MECP2 and METTL14 were significantly correlated with N stages (Fig. 3B). And the expressions of TET3, UHRF1 and YTHDF2 were significantly correlated with M stages (Fig. 3C).

3.4 Prognostic value of clinical stages

The above analysis showed that RNA methylation regulatory genes were related to TNM stages, then we analyzed the relationship between stages and prognosis via K-M curve analysis. Figure 4A showed that T, N, M stages had significant prognostic significance. In this study, we revealed that 23 genes were associated with at least one T, N, M stage, and the list of genes was in the Table S1. The expression of this genes was shown in the Fig. 4B.
3.5 Identification and confirmation of RNA methylation regulators signature

Firstly, two RNA methylation regulatory genes were gotten by univariate Cox analysis, and the univariate Cox analysis results were shown in the Table S2. As shown in Fig. 4C, LRPPRC was a protective factor for survival of CRC patients, and UHRF2 was a risk factor. Then we set parameter family as Cox to achieve LASSO logistic regression to reduce feature dimensions. There were two corresponding model genes when lambdamin was 0.002 (Fig. 4D). Therefore, LASSO analysis was used to build a prognostic signature with LRPPRC, UHRF2, and the risk coefficients of each gene were shown in Table 1.

<table>
<thead>
<tr>
<th>Gene</th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRPPRC</td>
<td>-0.7862</td>
<td>0.4556</td>
<td>0.2876</td>
<td>-2.734</td>
<td>0.00626</td>
</tr>
<tr>
<td>UHRF2</td>
<td>0.684</td>
<td>1.9818</td>
<td>0.2859</td>
<td>2.392</td>
<td>0.01675</td>
</tr>
</tbody>
</table>

To assess the prognostic of risk model, cases in TCGA training were classified into high and low-risk group by. The risk curve was composed of upper, middle and lower parts, and the dotted line was Median Riskscore (0.963) (Fig. 5A). By survival analysis, the OS of cases in the low-risk group were longer than that in the high-risk group in the TCGA-CRC training cohort (Fig. 5B). ROC curves were shown in Fig. 5C. The AUC of ROC curves in the training cohort were all greater than 0.6, indicating that the risk model had good prognostic value.

Then we also separated the TCGA internal validation cases into the high and low risk group. After analyzing the risk curve, scatter plot (Fig. 5D), survival curves (Fig. 5E) and ROC curve (Fig. 5F), it was found that the same high-risk group had lower survival time ($p = 0.036$). And in the ROC curve, AUC of nodes at 1, 3 and 5 years were all greater than 0.6. Finally, the risk model was externally validated using GSE48075. Corresponding to the results of the training cohort and the internal validation cohort, the same low-risk group had a better survival prognosis ($p = 0.031$) (Fig. 5G-I).

3.6 Prognostic risk scores is related with pathological characteristics in CRC

Correlation analysis was conducted between age, Stage, T, N, M stages and risk score of CRC samples from TCGA database ($n = 365$). And the differences of the risk score under different clinical information were gotten by the X2 test. Results showed that risk score was correlated with T Stages (Fig. 6).

3.7 The Risk Score was independent
Uni-Cox independent prognostic analysis was performed on 365 TCGA CRC samples with 6 clinicopathological factors including age, gender, Stage, TNM stages and risk score (Fig. 7A). The P values of riskScore, age, Stage and TNM stages were < 0.05. Ultimately, the risk score was an independent prognostic indicator for OS in CRC patients after multi-Cox independent prognostic analysis (Fig. 7B). In this study, relevant clinical information and risk score of CRC samples from TCGA-CRC database (n = 365) were used to construct a nomogram as shown in Fig. 7C, and the C index of the nomogram was 0.762, indicating that the nomogram model had a nice prediction value. The correction curve of the line graph (Fig. 7D) was built, and it was found that the prediction accuracy of the model for the survival probability of patients was relatively high, indicating that the nomogram could be used as an effective model.

3.8 Functional enrichment analysis

In this study, 2018 genes as the DEGs, which were annotated for GO analysis, and the biological significance of each gene was explored. We can see that, the most biological process (BP) terms were collagen fibrous tissue, chondrogenesis, ossification and other processes, including ion channel complex, complex of collagen trimers, cation channel complex, external side of plasma membrane, postsynaptic membrane, endoplasmic reticulum lumen, synaptic membrane, and collagen trimer etc. The DEGs were mainly involved in BP, including ion channel complex, lateral plasma membrane, and postsynaptic membrane. In molecular function (MF) terms, the most enriched terms were growth factor activity, receptor-ligand activity, integrin and heparin binding, including receptor ligand activity, growth factor binding, sulfur compound binding, collagen binding, extracellular matrix structural constituent conferring tensile strength, heparin binding, and glycosaminoglycan binding (Fig. 8A). Functional enrichment analysis showed that 67 related pathways were obtained, and these DEGs were significantly correlated with phagocytosis, cell adhesion molecules, PI3K-Akt signaling pathway, calcium signaling and other pathways (Fig. 8B).

The enrichment scores and differences in each pathway were analyzed for samples from the high-risk groups by GSVA. A total of 108 pathways were different between high and low risk cohort. The enrichment score of top 20 pathways was visualized by heat map (Fig. 8C). We found that Glucose metabolism, fat metabolism, amino acid metabolism and other pathways were different between high and low risk cohorts.

3.9 Analysis of immune infiltration and immunotherapy responses in high and low risk groups

First, we analyzed differences in immune microenvironment of CRC. The ImmuneScore, StromalScore and ESTIMATEScore of CRC samples (n = 365) were obtained according to the ESTIMATE algorithm (Fig. 9A-C). It can be seen that there were significant differences in these scores of high-low risk groups ($p < 0.05$). Then we further analyzed the scores for each immune cell by ssGSEA. The 19 immune cells were different, including NK cells, CD8 T cells, Tem, etc. The aDC, NK CD56dim cells, NK CD56bright cells, T
helper cells and TReg had no different between high-low groups (Fig. 9D). Lastly, the abundance of each immune cell in CRC sample was counted, and the corresponding statistical value was calculated. We remained 112 samples ($p < 0.05$), and the abundance of immune cells was shown in Fig. 9E. We observed only Macrophages M2 and T cells CD4 memory resting were different (Fig. 9F). To further validate the model, explain differences in immunotherapy between the high and low risk groups. Primarily, we compared the abundance of immune checkpoint genes. In addition to IDO1, the CD27, CD274, HAVCR2, ICOS, LAG3, PDCD1, PDCD1LG2, and TIGIT were significantly different between the high and low risk groups, and showed up-regulated expression in the high-risk group (Fig. 10A). Then the abundance profiles of CRC and melanoma were compared. According to the comparison results, immunotherapy with CTAL4 showed differences between high and low risk patients (Fig. 10B).

### 3.10 Validation of the expression of LRPPRC and UHRF2

we used qRT-PCR to compare gene expression levels of LRPPRC and UHRF2 in normal and cancerous tissue to verify the efficiency of the signature. The qRT-PCR showed an up-regulation in the abundance of LRPPRC and UHRF2 in cancerous tissue (Fig. 11A, B).

### 4. Discussion

Early CRC has no obvious symptoms, and most cases are diagnosed in the middle to late stages, which miss the best treatment time and result in poorer outcomes. Although immunotherapy has made great progress, however, current studies have suggested that only patients with MSI-H/dMMR and high tumor mutational Burden (TMB) can have a better effect on immunotherapy (IT)[16], most CRC patients still can not benefit from it. Improving the effectiveness of IT treatment and the prognosis of CRC patients is the focus of current research. So far, more than 170 post-transcriptional modifications of RNA have been discovered[17], M6A was discovered first and studied most among them[18]. Our study showed that LRPPRC and UHRF2 were overexpressed in CRC tissues. The prognostic risk model based on these two genes suggested that there was a difference in patient survival between the high and low-risk group, and the higher the risk score, the worse the patient survival. The proteins produced by LRPPRC are mitochondria-associated proteins that belonged to the family of pentatricopeptide repeat (PPR)-containing proteins that bound to RNA and regulate transcription, splicing, stability, editing, and translation-related[19, 20]. LRPPRC located in mitochondria and is identified as a mitochondria related protein. It was first discovered in HepG2 cells in 1994, and the gene encoding this protein is 4.8 kb in length[21]. Current studies have shown that LRPPRC is increased in cancer tissues and cell lines, including prostate cancer[22], gastric cancer[23], esophageal squamous cell carcinoma[24], colon cancer[25], etc. High LRPPRC combined with low microtubule-associated protein 1 family (MAP1S) is significantly associated with poor prognosis. MAP1S connects mitochondria and microtubules for transport, affecting the biogenesis and degradation of autophagy, thus increasing autophagy and inhibiting the occurrence of tumors[26]. RPPRC also interacts with Beclin 1 and Bcl2 to form ternary complexes to maintain stability of Bcl2 and inhibit autophagy[27]. Liu et al. found that LncRNA SnoRNA Host Gene 17 (SNHG17) may inhibit c-Myc ubiquitination through interaction with LRPPRC, thus
improving c-Myc and promoting the proliferation of tumor cells[28]. UHRF2 was another RNA methylation regulatory gene associated with CRC prognosis identified in this paper. UHRF2 is a multi-domain E3 ubiquitin ligase, which is involved in the development of many cancers, but in cancer it is controversial. For example, in the study of human esophageal squamous cell carcinoma, low abundance of UHRF2 was found to be associated with vascular invasion of squamous cell carcinoma and was an independent prognostic factor for poor prognosis[29]. The abundance of UHRF2 in lung cancer was lower, and UHRF2 knockout in A549 and 95-D cell lines enhanced the proliferation, invasion and migration of NSCLC cells, while overexpression of UHRF2 inhibited the proliferation, invasion and migration of NSCLC cells[30]. Our results show that this gene acts as an oncogene in colon cancer, which is consistent with other studies[31, 32]. UHRF2 is upregulated at both transcription and translation levels in CRC tumor tissue. Increased UHRF2 cytoplasmic expression was significantly correlated with clinical stage, invasion depth, lymph node metastasis, tumor histological grade, and metastasis[31]. The Wnt signaling pathway is important in the initiation and development of CRC. It is reported that UHRF2 can enhance Wnt signaling pathway transduction and promote the occurrence of CRC[32].

Prediction models for CRC patients have been widely reported[33–35]. But previous literature tended to be based on a single aspect of CRC development, such as N6 methylation alone, and some were relatively limited with either no internal tests or no external validation. In this paper, the effects of methylation regulatory enzymes m1A, m5C, m6A and m7G on the prognosis of CRC were comprehensively analyzed for internal verification and external verification, and also combined with TNM staging analysis. T, N, M staging is currently the most classic and commonly used method for CRC prognosis analysis, which can indicate the prognosis of patients[36]. However, it cannot predict the immunotherapy effect of patients. Our prognostic model showed that there were statistical differences in the abundance of immune microenvironment, immune cells, immune checkpoint molecules and response to immunotherapy in the high-low risk groups.

Tumor immune microenvironment (TME) affects tumor genesis, development and prognosis, and also plays a crucial role in immunotherapy. The tumor immune microenvironment is a complex and dynamic system [37, 38] involving the action of multiple cells and cytokines. NK cells are effector cells of the immune system. However, the level of NK cells in CRC tissues was lower, suggesting that less infiltration of NK cells may be one of the mechanisms of TME immune escape. Phenotypic changes of NK cells in peripheral blood of CRC cases promote tumor progression[39, 40]. TME is rich in macrophages (TAMs), which are considered to be the most abundant population of immune cells in solid tumor tissue[41]. In CRC, TAMs participates in angiogenesis and promoted the growth and invasion of CRC cells through epithelial-mesenchymal transformation (EMT) remodeling[42]. Studies have showed that in up to 80% of cases, elevated TAM levels predict a worse outcome. Current literature suggested that the M2-like phenotype of macrophages was more predictive of poor prognosis of CRC[38, 43]. In this study, ssGSEA and CIBERSORT algorithms were used to get the differences of immune cells. There were some differences between the two conclusions, which may be related to the race, gender, age and sample quality of the samples[44]. But overall, the tumor immune microenvironment was different. Immune
checkpoint blocking (ICB) therapies such as anti-PD-1/PD-L1 and anti-CTLA-4 have been clinically proven to be effective in CRC patients with MSI-H tumors\cite{45,46}. Our study showed that CTAL4 immunotherapy was more effective in the high-risk group.

In conclusion, this study constructed a gene related prognosis model for RNA methylation regulation, which have potential immunotherapy guidance significance. However, there were still some defects, the risk model needs to be further confirmed by increasing the sample size and conducting prospective multi-center studies.

**Abbreviations**

**CRC**- colorectal cancer

**CNV**- copy number variation

**TCGA**- The Cancer Genome Atlas

**OS**- overall survival

**K-M**- Kaplan–Meier

**LASSO** - Least Absolute Shrinkage and Selection Operator

**ROC**- Receiver operating characteristic

**GO**- Gene Ontology

**GSVA**- Gene set variation analysis

**KEGG**- Kyoto Encyclopedia of Genes and Genomes

**qRT-PCR** - quantative real-time PCR

**LRPPRC**- leucine rich pentatricopeptide repeat containing

**UHRF2**- ubiquitin-like with PHD and ring finger domains 2

**m6A**- N$\textsuperscript{6}$-methyladenosine

**m5C** - N$\textsuperscript{5}$-methylcytosine

**m1A** - N$\textsuperscript{1}$-methyladenosine

**m7G**- N$\textsuperscript{7}$-methylguanosine

**METTL3**- methyltransferase-like 3
GLUT1 - glucose transporter 1

mTORC1 - mechanistic target of rapamycin complex 1

BP - biological process

NSUN5 - NOP2/Sun RNA methyltransferase 5

SNV - single nucleotide variation

GEO - Gene Expression Omnibus

DEGs - differentially expressed genes

BP - biological process

MF - molecular function

TMB - tumor mutational Burden

PPR - pentatricopeptide repeat

MAP1S - microtubule-associated protein 1 family

SNHG17 - SnoRNA Host Gene 17

TME - Tumor immune microenvironment

TAMs - macrophages

EMT - epithelial-mesenchymal transformation

ICB - Immune checkpoint blocking

**Declarations**

**Acknowledgements**

None.

**Ethical Approval**

The study was approved by the Ethics Committee of the First Affiliated Hospital of Wenzhou Medical University. Signed written informed consents were obtained from the patients and/or guardians.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**

Ruoyang Lin and Tanzhou Chen conceived and designed the project, Ruoyang Lin, Lechi Ye and Xianfan Lin acquired the data, Ruoyang Lin and Renpin Chen analysed and interpreted the data, Ruoyang and Zhiming Huang wrote the paper, Tanzhou Chen revised the paper.

**Funding**

None.

**Availability of data and materials**

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

**References**


39. Halama N, Braun M, Kahlert C, Spille A, Quack C, Rahbari N, et al. Natural killer cells are scarce in colorectal carcinoma tissue despite high levels of chemokines and cytokines[J]. Clinical cancer
research 2011;17(4):678–689.


Figures
Figure 1

The landscape of genetic variations of RNA methylation regulatory genes in colorectal cancer. **A.** Mutation analysis of 53 RNA methylation regulatory genes. **B.** The frequency of mutations in genes showed by waterfall plot, the five genes with the highest mutation frequency were *ZC3H13, TET1, TET3, KIAA1429* and *YTHDC2*. **C.** Bar chart shows amplification and deletion frequency of RNA methylation regulating genes.
Figure 2

The correlation analysis of gene expression and somatic mutation & CNV patterns. 

**A.** Boxplot showing the effect of somatic mutations on gene expression. Pink is the mutant group, green is the wild group. (*P < 0.05)

**B.** Boxplot showing the effect of CNV on gene expression. Blue represents copy number amplification, orange represents copy number deletion, and green represents normal diploid. (*P < 0.05, **P < 0.01)
Figure 3

The relationship between RNA methylation regulatory genes expression and clinical stages.  

A. Relationship between gene expression and T staging. Blue represents T3 and T4s stage, and orange represents T1 and T2 stage.  

B. Relationship between gene expression and N staging. Blue represents N0 stage, and orange represents N+ stage.  

C. Relationship between gene expression and M staging. Blue represents M0 stage, and orange represents M+ stage. (*P<0.05, **P<0.01, ***P<0.001)
Figure 4

Correlation analysis of RNA methylation regulatory genes and prognosis. **A.** Kaplen-Meier survival curve analysis showed the correlation between T, N, M stage and prognosis. **B.** Heat map showing the correlation between RNA methylation genes and T, N, M staging. **C.** Forest map of RNA methylation regulation genes by univariate Cox analysis. Red dots represent risk factors and blue dots represent protective factors. **D.** The risk model was built by LASSO regression analysis.
Figure 5

Evaluation and validation of prognostic models. **A.** Risk curve, scatter plot and model gene expression heat map of high and low risk groups in training cohort. **B.** Kaplen-Meier survival curves for high and low risk groups in training cohort. Red represents the high risk group and blue represents the low risk group. **C.** ROC curves for high and low risk groups at 1 year, 3 years, 5 years in training cohort. **D.** Risk curve, scatter plot and model gene expression heat map of high and low risk groups in internal validation cohort. **E.**
Kaplen-Meier survival curves for high and low risk groups in internal validation cohort. Red represents the high risk group and blue represents the low risk group. F. ROC curves for high and low risk groups at 1 year, 3 years, 5 years in internal validation cohort. G. Risk curve, scatter plot and model gene expression heat map of high and low risk groups in external validation cohort. H. Kaplen-Meier survival curves for high and low risk groups in external validation cohort. Red represents the high risk group and blue represents the low risk group. I. ROC curves for high and low risk groups at 1 year, 3 years, 5 years in external validation cohort.

<table>
<thead>
<tr>
<th></th>
<th>Total (N=365)</th>
<th>high (N=182)</th>
<th>low (N=183)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>56 (15.3%)</td>
<td>19 (10.4%)</td>
<td>37 (20.2%)</td>
<td>0.07</td>
</tr>
<tr>
<td>Stage II</td>
<td>131 (35.9%)</td>
<td>68 (37.4%)</td>
<td>63 (34.4%)</td>
<td></td>
</tr>
<tr>
<td>Stage III</td>
<td>111 (30.4%)</td>
<td>61 (33.5%)</td>
<td>50 (27.3%)</td>
<td></td>
</tr>
<tr>
<td>Stage IV</td>
<td>51 (14.0%)</td>
<td>25 (13.7%)</td>
<td>26 (14.2%)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>16 (4.4%)</td>
<td>9 (4.9%)</td>
<td>7 (3.8%)</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>248 (67.9%)</td>
<td>124 (68.1%)</td>
<td>124 (67.8%)</td>
<td>0.787</td>
</tr>
<tr>
<td>M1</td>
<td>48 (13.2%)</td>
<td>22 (12.1%)</td>
<td>26 (14.2%)</td>
<td></td>
</tr>
<tr>
<td>MX</td>
<td>61 (16.7%)</td>
<td>32 (17.6%)</td>
<td>29 (15.8%)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>8 (2.2%)</td>
<td>4 (2.2%)</td>
<td>4 (2.2%)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>199 (54.5%)</td>
<td>92 (50.5%)</td>
<td>107 (58.5%)</td>
<td>0.162</td>
</tr>
<tr>
<td>N1</td>
<td>97 (26.6%)</td>
<td>48 (26.4%)</td>
<td>49 (26.8%)</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>64 (17.5%)</td>
<td>40 (22.0%)</td>
<td>24 (13.1%)</td>
<td></td>
</tr>
<tr>
<td>NX</td>
<td>2 (0.5%)</td>
<td>1 (0.5%)</td>
<td>1 (0.5%)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3 (0.8%)</td>
<td>1 (0.5%)</td>
<td>2 (1.1%)</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>10 (2.7%)</td>
<td>2 (1.1%)</td>
<td>8 (4.4%)</td>
<td>0.024</td>
</tr>
<tr>
<td>T2</td>
<td>55 (15.1%)</td>
<td>20 (11.0%)</td>
<td>35 (19.1%)</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>250 (68.5%)</td>
<td>130 (71.4%)</td>
<td>120 (65.6%)</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>46 (12.6%)</td>
<td>28 (15.4%)</td>
<td>18 (9.8%)</td>
<td></td>
</tr>
<tr>
<td>Tis</td>
<td>1 (0.3%)</td>
<td>1 (0.5%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3 (0.8%)</td>
<td>1 (0.5%)</td>
<td>2 (1.1%)</td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>162 (44.4%)</td>
<td>82 (45.1%)</td>
<td>80 (43.7%)</td>
<td>0.879</td>
</tr>
<tr>
<td>male</td>
<td>203 (55.6%)</td>
<td>100 (54.9%)</td>
<td>103 (56.3%)</td>
<td></td>
</tr>
<tr>
<td>age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;=65</td>
<td>193 (52.9%)</td>
<td>96 (52.7%)</td>
<td>97 (53.0%)</td>
<td>1</td>
</tr>
<tr>
<td>&lt;65</td>
<td>171 (46.8%)</td>
<td>86 (47.3%)</td>
<td>85 (46.4%)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.3%)</td>
<td>0 (0%)</td>
<td>1 (0.5%)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6
Comparison of risk scores among different clinical subgroups.

A

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
<td>&lt;0.001</td>
<td>1.892 (1.482-2.464)</td>
</tr>
<tr>
<td>T</td>
<td>&lt;0.001</td>
<td>2.783 (1.775-4.363)</td>
</tr>
<tr>
<td>N</td>
<td>&lt;0.001</td>
<td>1.791 (1.379-2.328)</td>
</tr>
<tr>
<td>M</td>
<td>&lt;0.001</td>
<td>2.181 (1.399-3.421)</td>
</tr>
<tr>
<td>age</td>
<td>0.001</td>
<td>1.030 (1.012-1.048)</td>
</tr>
<tr>
<td>riskScore</td>
<td>0.002</td>
<td>1.675 (1.211-2.318)</td>
</tr>
<tr>
<td>gender</td>
<td>0.308</td>
<td>1.256 (0.812-1.944)</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>&lt;0.001</td>
<td>1.043 (1.023-1.054)</td>
</tr>
<tr>
<td>Stage</td>
<td>0.090</td>
<td>1.625 (1.001-2.638)</td>
</tr>
<tr>
<td>M</td>
<td>0.410</td>
<td>1.276 (0.718-2.246)</td>
</tr>
<tr>
<td>N</td>
<td>0.390</td>
<td>1.204 (0.806-1.753)</td>
</tr>
<tr>
<td>T</td>
<td>0.033</td>
<td>1.811 (1.056-3.122)</td>
</tr>
<tr>
<td>riskScore</td>
<td>0.024</td>
<td>1.638 (1.006-2.516)</td>
</tr>
</tbody>
</table>

Figure 7
The nomogram model construction and validation. A. Uni-Cox independent prognostic analysis. B. Multi-Cox independent prognostic analysis. C. A nomogram as shown, the C index of the nomogram was 0.761719. D. The correction curve of the line graph.
Results of functional enrichment analysis. **A.** The GO function enrichment bar chart of differentially expressed genes. The ordinate represents the enriched GO Term, the bar length represents the number of differentially expressed genes enriched in the GO Term. **B.** Bars of KEGG pathway enrichment of differentially expressed genes. The ordinate indicates the enrichment of KEGG Pathway, the bar length indicates the number of Pathway genes. **C.** Heat map of enrichment of TOP20 differential pathways in high and low risk groups. Each small square represents the ssGSEA score of each sample, and the color represents the size of ssGSEA score. The larger the square, the darker the color. Red indicates a high ssGSEA score and blue indicates a low ssGSEA score.
Figure 9

Results of immune infiltration analysis in high and low risk groups. A. Violins of ImmuneScore in the high and low risk groups. B. Violins of StromalScore in the high and low risk groups. C. Violins of ESTIMATEScore in the high and low risk groups. The abscissa represents groups, the ordinate represents scores, the blue represents low risk group, and the pink represents high risk group. D. Analyze the differences of immune cells in high and low risk groups by ssGSEA. E. The bar chart of proportion of
immune cells in high and low risk groups by CIBERSORT. F. Violin diagram of immune cells in the high and low risk groups by CIBERSORT.

Figure 10

Results of immunotherapy differences between high and low risk groups. A. Boxplot of immune checkpoint gene expression differences in high and low risk groups. Red is the high risk group, blue is the
low risk group. B. Heat maps of immunotherapy differences between high and low risk groups.

Figure 11

Results of Gene expression levels in normal and cancerous tissue using qRT-PCR. A. LRPPRC. B. UHRF2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfile.docx