The Role of N6-Methyladenosine-associated lncRNAs in Immune Microenvironment and Prognosis of Colorectal Cancer

Congfei Yuan  
Lianshui County People's Hospital

Caidong Liu  
Nanjing First Hospital

Shuli Zhao  
Nanjing First Hospital

Xishan Zhang  
Lianshui County People's Hospital

Haifeng Jia  
Lianshui County People's Hospital

Baiyu Chen  
Lianshui County People's Hospital

Maojin Zhang  
Lianshui County People's Hospital

Yuan Zheng  
Lianshui County People's Hospital

Jin Zhou  
Nanjing First Hospital

Yanzhi Bo (✉️ buyz279@hotmail.com)  
Lianshui County People's Hospital

Research Article

Keywords: N6-methyladenosine, colorectal cancer, lncRNA, prognosis, immune microenvironment

Posted Date: April 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1532948/v1

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

Background

The role of N6-methyladenosine long non-coding RNAs (IncRNAs) in colorectal cancer (CRC) is elusive.

Materials and Methods

We identified m6A-associated IncRNAs by using the data gathered from The Cancer Genome Atlas (TCGA) and stratified CRC patients into different subgroups. Cox-regression analysis were performed to construct a m6A-associated IncRNA signature. The role of this signature in immune microenvironment and prognosis was dissected subsequently. Finally, a gene set enrichment analysis (GSEA) was conducted to predict the possible mechanisms based on the signature.

Results

Three m6A-associated clusters were constructed from 866 differentially expressed IncRNAs. Cluster 2 had poor prognosis and low immune cell infiltration. A m6A-associated IncRNA signature consisting of 14 IncRNAs was constructed, and recognized as an independent prognostic indicator of CRC by using survival analysis and receiver operating characteristic (ROC) curves. The clinical features and immune cell infiltration status were significantly different in patients stratified by risk score. Furthermore, GSEA showed that P53 pathway and Natural killer cell mediated cytotoxicity were more enriched in the low-risk group.

Conclusion

Our data revealed that m6A-associated IncRNAs could be potential prognostic indicator of immunogenicity in CRC.

Introduction

Colorectal cancer (CRC) is the third most prevalent gastrointestinal malignancy worldwide\(^1\). A significant number of CRC patients will ultimately relapse after curative treatments\(^2,3\). Hence, there is an urgent need to investigate prognostic markers for CRC.

N6-methyladenosine (M6A) is the most common post-transcriptional modification in RNAs\(^4\). Recent studies have indicated that m6A RNA modification plays important roles in biological processes and cancer pathogenesis\(^5\). Aberrant expressions of m6A regulators (e.g. METTL14, METTL3, KIA, ALKBH5, FTO and YTHDF1/2/3) have been identified in numerous tumors\(^6-10\). A variety of pathological functions, ranging from tumor initiation, invasion, metastasis to tumor stem cell pluripotency, could be mediated by
m6A methylation. Long non-coding RNAs (lncRNAs) are important epigenetic regulators, that play critical roles in diverse physiological and pathological processes. Studies have reported that some lncRNAs participate in tumor initiation and progression. Despite extensive efforts to define the pathogenesis of lncRNAs, the roles of lncRNAs in the m6A modification in CRC remains largely elusive.

Immune microenvironment has been found to be closely associated with the clinical outcome of immunotherapy and tumor development. In the present study, the co-expression network of the m6A-associated lncRNAs was investigated to obtain 68 m6A-associated prognostic lncRNAs. Then we established three m6A-associated clusters in CRC, analyzed the characteristics of immune cell infiltration among tumor cells and investigated whether m6A-associated lncRNAs clusters have prognosis values in CRC patients. Furthermore, we constructed a signature using 14 m6A-associated lncRNAs which could predict the prognosis of CRC patients.

Results

The differential expressions of m6A-associated lncRNAs.

A total of 19604 mRNAs and 14086 lncRNAs were screened from the TCGA database. 1590 m6A-associated lncRNAs were obtained (\(|R|>0.4\) and \(p<0.05\)) according to 23 reported m6A-associated genes, of which, 866 differentially expressed m6A-associated lncRNAs in CRC were detected with a log/fold change (FC) >0.5 and a \(p<0.001\) (Supplementary data 1).

Identification of m6A-associated lncRNAs with prognostic value

As shown in Figure 1A, we annotated m6A-associated lncRNAs and clinical characteristics, then investigated the role of each lncRNA on the prognostic outcome of the patients with CRC. A total of 68 m6A-associated lncRNAs with obvious prognostic value were detected and used for further study.

Establishment of m6A-associated lncRNA clusters

To classify the different m6A clusters based on the lncRNAs, we mapped these 68 m6A-associated lncRNAs to expression profile of CRC samples to perform clustering using the Consensus Cluster Plus (CCP) tool. As shown in Figure 1B, the number of clusters was sequentially set from 1 to 9 and CCP analysis indicated that the results were most stable when these m6A-related lncRNAs were separate into three clusters using Consensus Cluster Plus R package (Figure 1C, D). The OS data of each cluster was calculated using Kaplan-Meier method, and the results displayed that there was significant difference among the survival of CRC patients in these three clusters (Figure 1E).

Clinical characteristics and Immune Score of each cluster in CRC

As compared with cluster 1 and cluster 3, cluster 2 had the highest N stage, M stage and TNM stage (Figure 2A). ESTIMATE-algorithm was employed to evaluate the accurate estimate score (tumor purity), immune score and stromal score in accordance with the gene expression profiles of CRC patients. Our
findings showed that compared with cluster 1 and 3, the cluster 2 had the lowest estimate score, immune score and stromal score (Figure 2B).

**m6A-associated IncRNAs signature construction**

As shown in Figure 3A, a total of 14 m6A-associated IncRNAs that had a co-expression relationship with 8 m6A-associated genes were recognized as effective independent prognostic factors. Among them, AC137932.3, AL391422.4, AC092123.1, AC156455.1, AC132192.2, AC008760.1, RPARP-AS1, LINC02657, AP001619.1, AC003101.2, AL161729.4, TNFRSF10A-AS1, AL121906.2 and AC074117.1 were found to be favorable prognostic factors (Supplementary data 2). The risk score of each CRC patient = AC137932.3*(-1.4041)+AL391422.4*0.9484+AC092123.1*(-1.3865)+AC156455.1*0.1977+AC132192.2*(-0.4822)+AC008760.1*0.5973+RPARP-AS1*0.3572+LINC02657*0.7205+
AP001619.1*0.8025+AC003101.2*1.0959+AL161729.4*0.3047+TNFRSF10A-AS1*(-0.2329)+AL121906.2*1.02629+AC074117.1*0.25582. Based on the median risk score, 426 CRC patients were classified into the low- and high-risk groups. The Kaplan-Meier curves and the distributions of survival status confirmed the poor outcome in the high-risk group (Figure 3 B-D). Our findings showed that the mortality was closely associated with the risk score. Moreover, the area under the curve (AUC) is measured and the value for prognostic risk score was 0.764 which is higher than AUCs of the other clinicopathological factors (Figure 4E). And the AUC values corresponding to 1-, 3-, and 5-year of OS were 0.764, 0.743 and 0.753, respectively (Figure 3F). These data indicated the good prediction accuracy of this model.

**The validation of the signature in CRC**

The prognostic value of the m6A-associated IncRNA signature was investigated in CRC patients from the TCGA dataset. The patients were classified by various clinical parameters, consisting of gender, age, T, N, M and TNM stage. In almost all subgroups, the patients with low-risk score trended to have higher OS rate than that of the high-risk group (Figure 4).

Next, we evaluated the independence and effectiveness of this model in predicting prognosis of CRC patients. Our findings showed that this m6A-associated IncRNA signature could be an effective and independent factor for predicting the outcome of the patients with CRC (Figure 5 A-B). Then, a nomogram was conducted to predict 1-, 3- and 5-year OS of the patients with CRC based on the results of univariate and multivariate Cox-regression analyses, including age, TNM stage and risk score (Figure 5C). The calibration curves demonstrated the well prediction accuracy of this nomogram in CRC patients (Figure 5 D-F).

**Gene set enrichment analysis**

Finally, we evaluated the potential biological mechanisms associated with the risk model by GSEA. As shown in Figure 6, the P53 signaling pathway (NOM p-val=0.0019, FDR q-val=0.155) and Natural Killer cells mediated cytotoxicity (NOM p-val=0.0172, FDR q-val=0.195) were more enriched in the low-risk
group. Our study suggested that this risk-related model could be used for the personalized treatment for CRC patients.

**Discussion**

Previous studies have demonstrated the pivotal roles of m6A modification in various cancers including CRC\(^\text{15-17}\). Investigating the potential prognostic role of m6A-associated lncRNAs will facilitate understanding the molecular mechanisms of CRC. In our work, 68 prognostic m6A-associated lncRNAs were identified, then three m6A-associated lncRNAs cluster groups were constructed using 426 CRC samples from TCGA database. Compared with cluster1 and cluster3, cluster2 had the worst OS time and high pathological stage. In addition, ESTIMATE analyses revealed that the immune score was remarkably reduced in cluster 2. Our data suggested that m6A-associated lncRNAs might be used as a predictive biomarker.

It is generally known that there are currently some CRC prognostic indicators, including TNM stage and tumor grade. However, more accurate prognostic factors are required to predict and analyze the OS rate in CRC patients. Current studies have indicated that lncRNAs play important roles in predicting the outcome and prognosis of various cancers. For instance, Dan Yin, et al\(^\text{21}\) reported that overexpression of LINC01133 was related to the poor prognosis in patients with hepatocellular carcinoma. Shujun Feng, et al\(^\text{22}\) reported that lncRNA-CTS was aberrantly expressed in gastric cancer tissues, and the upregulation of CTS was closely associated with tumor volume, tumor histology, lymph node metastasis and the poor prognosis. Recently, numerous m6A-associated lncRNAs are reported to be potential markers for the prediction of various cancers, Haixu Wang, et al\(^\text{23}\) established a 11 m6A-associated lncRNA signature, and confirmed that it had a good prognostic value and could act as an valid marker for gastric cancer. Feng Xu, et al\(^\text{24}\) established a risk model consisting of 12 m6A-associated lncRNAs, and demonstrated that the model might be a promising prediction of prognosis in lung adenocarcinoma patients. In the present study, a m6A-associated lncRNA signature consists of 14 lncRNAs could predict patients with poor prognosis. Moreover, we assessed the clinical value of the signature in gender, age, T, N, M and TNM stage, and identified that the signature was closely associated with the progression of CRC. Meanwhile, the GSEA analysis preliminary displayed that these lncRNAs were closely involved in P53 pathway and NK cell mediated cytotoxicity. Further studies are needed to demonstrate the mechanisms involved in this lncRNA signature.

**Conclusion**

In summary, our work defined a m6A-associated lncRNA signature which could predict the prognosis of CRC patients. This m6A-associated lncRNA signature will provide guidance for individualized treatment.

**Methods**

**Data acquisition and processing of the CRC dataset**
The public RNA sequencing (RNA-seq) data from 512 patients CRC were downloaded from TCGA(https://portal.gdc.cancer.gov/). Patients without survival information were removed.

Identification of m6A-associated IncRNAs in CRC

The m6A-associated genes were gathered from the TCGA database and selected based on previously published articles\textsuperscript{25,26}. And the m6A-associated IncRNAs were screened by Spearman correlation coefficient formula with $|R|$ value $>$0.6 and $p$ value $<$0.001.

Consensus clustering of m6A-associated IncRNAs

On the basis of the expression levels of m6A-associated IncRNAs, the CRC patients were separately divided into three groups (cluster1, 2 and 3) according to the optimal k-means clustering. Cluster analysis was performed with Consensus Cluster Plus R package. The overall survival (OS) data of each cluster was calculated using Kaplan-Meier method. The correlation between m6A-associated IncRNAs and clinical characteristics was analyzed according to the TCGA database. And ESTIMATE-algorithm was employed to estimate the tumor immune microenvironment.

m6A-associated IncRNA signature construction

The prognostic m6A-associated IncRNAs were identified via univariate cox regression analysis. And the prognostic signature was established via multivariate cox regression analysis. The risk scores of CRC patients were calculated by following formula, Risk score = $\sum$ Expi$^*$bi, here Expi represents the expression and $\beta_i$ represents the coefficient of m6A-associated IncRNAs. And the accuracy of the m6A-associated IncRNAs was assessed via the ROC curve analysis.

Statistical analysis

All data were analyzed via by using R statistical software version 4.0.3. A $p$ value less than 0.05 was statistically significant.

Abbreviations

IncRNAs
long non-coding RNAs
CRC
colorectal cancer
TCGA
The Cancer Genome Atlas
GSEA
gene set enrichment analysis
ROC
receiver operating characteristic
M6A
N6-methyladenosine
FC
fold change
CCP
Consensus Cluster Plus
AUC
area under the curve
RNA-seq
RNA sequencing

References


Declarations

Consent for publication

Not applicable.

Availability of data and materials

The data that support our findings are openly available in TCGA (https://portal.gdc.cancer.gov/) repository.
Competing interests

No author has conflict of interest.

Authors' contributions

Conception and design: Jin Zhou and Yanzhi Bo

Development of methodology: Congfei Yuan and Caidong Liu

Acquisition of data: Congfei Yuan and Caidong Liu, Shuli Zhao

Analysis and interpretation of data: Congfei Yuan, Xishan Zhang, Haifeng Jia, Baiyu Chen, Maojin Zhang and Yuan Zheng

Writing, review, and/or revision of the manuscript: Congfei Yuan, Caidong Liu, Jin Zhou and Yanzhi Bo

Study supervision: Jin Zhou

Figures

Figure 1


Figure 2

Clinical characteristics and immune score of m6A-associated lncRNAs in CRC. (A). Heatmap of the correlation between m6A-associated lncRNAs and clinical characteristics in the TCGA database. (B). Comparison of composition of immune score, stromal score and estimate score in cluster 1, 2 and 3.

Figure 3

The signature based on m6A-associated lncRNAs for CRC patients. (A). The network of 14 m6A-associated lncRNAs. (B). Kaplan-Meier analysis of the low- and high-risk groups. (C-D). The distribution of
risk scores and the survival state of selected m6A-associated lncRNAs. (E). The AUC of risk score and other clinicopathological factors. (F). The AUC for 1-, 3- and 5-year survival rates were 0.764, 0.743, 0.753, respectively.

**Figure 4**

The prognostic value of the m6A-associated lncRNA signature in CRC patients. Kaplan-Meier analysis for the different risk groups classified using clinical factors including age (A), gender (B), M stage (C), N stage(D), T stage(E) and TNM stage (F).

**Figure 5**

The independence and effectiveness of this model in predicting prognosis of CRC patients. Forest plots of univariate (A) and multivariate (B) Cox regression analysis in CRC. Nomogram model (C) to predict 1-, 3- and 5-year survival rates of CRC patients. Calibration graph showed the predicted 1- (D), 3-(E) and 5-year (F) survival rates were close to actual survival rates.

**Figure 6**

The enriched signaling pathways in the low-risk group. The GSEA results of the P53 signaling pathway and nature killer cells mediated cytotoxicity.
Supplementary Files

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

- supplementarydata1.txt
- supplementarydata2.txt