Construction of a Prognostic m6A-related lncRNA Signature to Predict Immunotherapy and Chemotherapy Response in Hepatocellular Carcinoma

Yiping Zou
Guangdong Provincial People's Hospital

Zhihong Chen
Guangdong Provincial People's Hospital

Hongwei Han
Guangdong Provincial People's Hospital

Qi Lou
Guangdong Provincial People's Hospital

Yuanpeng Zhang
Guangdong Provincial People's Hospital

Zhenrong Chen
Guangdong Provincial People's Hospital

Zuyi Ma
Guangdong Provincial People

Haosheng Jin
Guangdong Provincial People's Hospital

Ning Shi (✉ shining_doc@163.com)
Guangdong Provincial People's Hospital

Research Article

Keywords: LncRNA, hepatocellular carcinoma, tumor-infiltrating immune cells, chemotherapy, methylation

Posted Date: November 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1040436/v1

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

Hepatocellular carcinoma (HCC) is one of the main causes of cancer-related deaths worldwide. N6-methyladenosine (m6A) and long noncoding RNA (lncRNA) are two common modifications that affect tumor development and play vital roles in the prognosis of HCC. Therefore, we comprehensively analyzed transcriptome data from the Cancer Genome Atlas (TCGA) and identified lncRNAs related to m6A regulators. Univariate, LASSO, and multivariate Cox regression analyses were used to build a prognostic model. Patients were then divided into low- and high-risk groups according to the optimal cut-off point. The prognosis value of the novel model was evaluated using Kaplan-Meier analysis and the receiver operating characteristic curve. Besides, mutation and immune profiles together with immune checkpoint expressions were further explored to identify the difference in somatic alteration and tumor immune landscape between the two groups. Additionally, a better response to conventional chemotherapy and immunotherapy was found in patients with the high-risk group, but they were more resistant to sorafenib. The m6A-related lncRNAs model might be used to predict the prognosis and drug responses in patients with HCC.

1. Introduction

Hepatocellular carcinoma (HCC), the most common histological subtype of primary liver cancer, is one of the leading causes of cancer deaths worldwide.[1] Cirrhosis from any etiology, such as chronic alcohol consumption and infection by hepatitis C virus (HCV) or hepatitis B virus (HBV) is the primary risk factor for HCC.[2] Radical resection is the only choice for HCC patients to receive a long-term survival. However, most HCC patients were diagnosed at an advanced stage and systemic therapies were indicated. [3, 4]

Doxorubicin, cisplatin, fluorouracil, and other chemotherapy drugs were administered as the systemic chemotherapy regimens for palliative treatment in advanced HCC, but the response rate was very low. [5–7] Sorafenib and lenvatinib are multikinase Inhibitors (MKIs) approved as first-line systemic therapies for unresectable HCC. [8, 9] Recently, immune checkpoint inhibitors such as nivolumab, pembrolizumab, and camrelizumab monotherapy have been approved for the treatment of HCC patients who have progressed on or after standard sorafenib therapy, which launched the era of immunotherapy in HCC. [10–12] However, both current first-line and second-line systemic treatments have the disadvantage of low objective response rates (ORR). Immunotherapy combined with targeted therapy or double immunotherapy could increase the ORR to 30% or more. [13–15] Despite such advancements, there is still a large proportion of patients with HCC who do not respond to systematic therapy. Therefore, there is an urgent need to identify the potential predictors for improving the response rate to systematic therapy.

N6-methyladenosine (m6A), the most common RNA modification in eukaryotic cells, is considered to affect all aspects of RNA metabolism, including RNA translocation, splicing, stability, and translation into protein.[16] There are three types of homologous factors in m6A modification: methyltransferases (writers), signal transducers (readers), and demethylases (erasers).[17] Previous studies have indicated that abnormal m6A methylation modification is involved in oncogenesis and tumor development. [18]
Long non-coding RNAs (lncRNAs) are a type of RNA with a transcript length of more than 200 nucleotides and play an important role in almost all vital bioprocesses. [19, 20] Dysregulation of lncRNAs has been observed in several malignant tumors and implicated as a novel biomarker in the diagnosis, prognosis, and therapeutic response of cancer. [21–24] Therefore, exploring the underlying mechanism of m6A-related lncRNAs may guide the prognostic and treatment of HCC.

Our study aimed to construct a novel model based on m6A-related lncRNAs to predict the prognosis of patients with HCC. The predictive efficiency of overall survival (OS), tumor immune infiltration, chemotherapeutic efficacy, and immunotherapy response were further estimated.

2. Materials And Methods

2.1. Data collection

The transcriptomic profiles (FPKM format), mutation data (simple nucleotide variation), and clinical data of LIHC were downloaded from The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/). We downloaded the GTF files from Ensembl (http://asia.ensembl.org/) to distinguish the mRNAs and lncRNAs from the mRNA expression data. The LIHC dataset included 374 tumor samples and 50 normal samples. All samples were used for gene screening and differential expression analyses. Patients with missing survival information or an overall survival (OS) time of less than one month were excluded from the survival and subgroup analyses. A total of 365 samples were used in this study.

2.2 Identification of m6A-related lncRNA

23 m6A regulators including 8 writers (METTL14, METTL3, METTL16, VIRMA, WTAP, RBM15B, RBM15, and ZC3H13), 2 erasers (FTO and ALKBH5), and 13 readers (YYTHDC1, HNRNPC, YTHDF1, YTHDC2, YTHDF3, YTHDF2, FMR1, LRPPRC, IGFBP1, IGFBP2, IGFBP3, RBMX, and HNRNPA2B1) were extracted to identify m6-A related lncRNAs. We conducted a Pearson correlation analysis between all lncRNAs and m6A regulators. The m6A-related lncRNAs were selected using the following criteria, R (correlation coefficient) > 0.5 and p < 0.001. Moreover, univariate cox regression analysis was applied to identify lncRNAs related to prognosis. LncRNAs with p values < 0.01 were screened for further analysis.

2.3 Construction and evaluation of a novel m6A-related signature

To conduct dimensionality reduction and construct a signature, LASSO addition with stepwise multivariate Cox regression analyses was further performed. [25] The risk score was calculated using the following formula:

\[ \text{risk score} = \sum (\beta \text{ln} x \times \text{Exp} \text{ln} x) \]

Where “\( \beta \)” was the coefficient value for each lncRNA, while “\( x \)” was the number of prognostic lncRNAs, and “\( \text{Ex} \)” was lncRNAs expression value.
We then divided the patients into high- and low-risk groups according to the cutoff point determined by the R package “survminer”. Additionally, we used the principal component analysis (PCA) to visualize the sample distinguishing ability of the signature. In addition, the Kaplan-Meier survival method was applied to compare the difference of OS between high- and low-risk groups. Subsequently, Kaplan-Meier analyses between the high- and low-risk groups were conducted, stratified by patients’ age, gender, tumor grade, or tumor stage. The receiver operating characteristic (ROC) was further applied to evaluate the prognostic accuracy of the predictive signature. Univariate and multivariate Cox regression analyses were performed between the risk model and other clinicopathological variables to evaluate the independent prognostic value of the risk signature. Finally, the association of prognosis m6A-related lncRNAs and the clinicopathological characteristics were visualized by heatmap.

2.4 Mutation data analysis

The R package “maftools” was utilized to assess and visualize the variance of somatic mutations between high- and low-risk groups. [26] Mutation information of each gene was exhibited in the waterfall plot. The tumor mutational burden (TMB) was calculated according to following formula: (total mutation/total covered bases) ×10^6. The difference of TMB between high- and low-risk groups was evaluated. The prognostic value of TMB was also analyzed.

2.5 Evaluation of Immune microenvironment and immunosuppressive molecules expressing

The CIBERSORT package was applied to evaluate the infiltration levels of 22 immune cells. [27] Then the association between immune cell infiltration and risk signature was explored. The samples with p < 0.05 were included in further analysis. Wilcoxon test was utilized to analyze differences of immune cell infiltration in low- and high-risk groups. Additionally, spearman correction analysis was also adopted. We evaluated the relationship between immune cell infiltration and risk score. The expression levels of immune checkpoint molecules (PD1, PDL1, TIM3, LAG3, TIGIT, and CTLA4) were further analyzed.

2.6 Therapeutic response prediction

The R package pRRophetic was utilized to predict the half inhibitory centration (IC50) of antitumor drugs such as sorafenib, cisplatin, mitomycin, and doxorubicin obtained from the GDSC website. [28] Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was performed to assess the sensitivity of the immunotherapeutic response.[29] Patients with low TIDE were more likely to show the immunotherapeutic response. Furthermore, the Wilcoxon test was conducted to compare the difference of IC50 and TIDE between the low- and high-risk groups.

2.7 Gene set enrichment analysis (GSEA)

We conduct the Kyoto Encyclopedia of Genes and Genomes (KEGG) terms of GSEA according to the weighted enrichment statistic method. [30] The false discovery rate (FDR) was calculated. When FDR was
less than 0.25 and the nominal p value was less than 0.05, the enriched gene set was considered statistically significant. Enrichment results were visualized using the R package “ggplot2”.

2.8 Statistical Analyses

All statistical analysis of this study was performed using R software, version 4.0.5. Without special description, the p < 0.05 was considered statistically significant.

3. Results

3.1. Acquisition and analysis of m6A-related lncRNA data of HCC patients

Figure 1 was the work chart of this study. Firstly, we identified 203 m6A-related lncRNAs based on 23 m6A regulators (Figure 2A). Univariate Cox regression analysis was conducted to identify m6A-related lncRNA with prognostic roles. And 43 m6A-related lncRNAs which were significantly related to the OS of HCC patients were identified (Figure 2B). Figure 2C and 2D were the boxplot and heatmap to show the expression level of 43 OS-related lncRNAs. The result indicated that all the expression of 43 lncRNAs were significantly upregulated in tumor tissue than normal tissue (p< 0.05).

3.2 Construction of a m6A-related lncRNAs signature

The LOSSO regression was used in the expression profiles of 43 m6A-related lncRNAs associated with OS and 8 m6A-related lncRNAs were further selected (Figure 3A). A stepwise multivariate Cox regression analysis was performed and finally, a 5 m6A-related lncRNAs risk signature was constructed (Figure 3B). The risk score was computed with the formula: risk score = (0.302121 × expression of AC145207) + (0.34025 × expression of AL031985) + (0.198432 × expression of SNHG4) + (0.051602× expression of AC099850) + (1.312568×expression of AC026412) (Table 1). Subsequently, we calculated the risk score for each patient and classified them into a high-risk group (n = 44) and a low-risk group (n = 321) according to the cut-off point of 1.91, which was obtained by the R package “survminer” (Figure 3C).

The result of PCA indicated that the clusters of the low-risk group were independent of the high-risk group without obvious intersection in m6A-related lncRNAs expression level (Figure 3D). ROC analysis indicated that the value of m6A-related signature in predicting OS (AUC= 0.754) was higher than those of tumor stage (AUC= 0.664), age (AUC= 0.535), tumor grade (AUC= 0.505) and gender (AUC= 0.513) (Figure 3E). Further analysis showed poorer prognosis and more deaths for the patients with higher risk scores (Figure 3F). A significantly shorter OS was shown in the high-risk group by Kaplan-Meier survival analysis (p < 0.001) (Figure 3G).

3.3 Relationship between m6A-related lncRNAs signature and clinicopathological variables

Figure 4A summarized the clinical characteristics of patients in different risk groups. Patients in low-risk group were older than patients in the low-risk group, with higher tumor grades and more advanced tumor stage (p< 0.05). Besides, all of 5 m6A-related lncRNAs in the risk signature were augmented in the high-
risk group. The univariate Cox regression analysis revealed that the risk score was associated with a poor OS indicator (HR=1.402, 95% CI= 1.285- 1.535, p< 0.001, Figure 4B). Besides, multivariate Cox regression analysis indicated that the risk score was an independent risk predictor (HR= 1.400, 95% CI=1.270–1.542, p < 0.001, Figure 4C). In the subgroups classified by age, gender, grade, or tumor stage, patients in the low-risk group continued to have longer OS compared with those in the low-risk group (all p< 0.001, Figure 5A-5H).

3.4 M6A-related IncRNAs signature and cancer somatic variants

The distributions of HCC somatic variants between the high- and low-risk groups were represented in Figure 6A-B. The top 20 driver genes with the highest alteration frequency were shown by “maftools”. Compared with the low-risk group, the alteration frequency of TP53 (p< 0.001) was significantly higher in the high-risk group, while the alteration frequency of CTNNB1 (p= 0.022) was significantly lower. However, the TMB between high and low-risk groups did not meet the significant difference (p= 0.5, Figure 6C). Patients with low TMB had better OS outcomes than those with high TMB (p< 0.001, Figure 6D). Stratified survival analysis in subgroups showed significant differences in OS according to the risk score among the TMB subgroups (p< 0.001, Figure 6E). The result showed that the m6A-related IncRNAs signature could be a survival predictor independent of the TMB.

3.5 Estimation of the immune cell infiltration and immune checkpoints expression

We applied the CIBERSORT method to analyze the enrichment situation of 22 immune cells. The samples with p < 0.05 were used for further analysis (Figure 7A). High-risk group had lower infiltration level of naive B cells (p= 0.010) and regulatory T cells (p= 0.004, Figure 7B). Furthermore, the result of Spearman correlation analysis revealed that activated NK cell was negatively correlated with risk score (r= -0.28, p= 0.038, Figure 7C).

Next, we compared the expression of immune checkpoints expression between these two subgroups. The results indicated that immune checkpoints such as CTLA4, PDCD1, LAG3, TIGIT, and HAVCR2 were upregulated in the high-risk group (all p< 0.05, Figure 7D-7H). Although the expression of PDL1 (Figure 7I) was not significantly different between the two subgroups, Spearman correlation analysis indicated that all these immune checkpoints expression (CTLA4, PDCD1, PDL1, LAG3, TIGIT, and HAVCR2) were positively correlated with risk score (Figure 8A-8F). These results revealed a more suppressed immune status in the high-risk group.

3.6 Analysis of the therapeutic responses and underlying regulatory mechanism

Firstly, we used the TIDE score to predict the efficiency of immunotherapy. As Figure 9A showed, the TIDE score in the high-risk group was significantly lower than the low-risk group (p< 0.001). This result suggested that patients in the high-risk group had a superior response to immunotherapy.

Besides, we analyzed the association between risk subgroups and the efficacy of common chemotherapeutics and targeted therapy in treating HCC (Figure 9B-9E). These results verified that
patients in the high-risk group had a lower IC50 of chemotherapeutics such as cisplatin (p = 0.028), doxorubicin (p < 0.001), mitomycin (p < 0.001), while with a higher IC50 for sorafenib (p < 0.001). Thus, the results indicated that the m6A-related IncRNA risk signature could be a potential predictor for drug sensitivity.

The result of GSEA demonstrated that the patients with the high-risk score were enriched in some cancer and proliferation pathways: cell cycle, mismatch repair, oocyte meiosis, pathway in cancer, and RNA degradation. Interestingly, patients with low-risk scores were mainly enriched in drug metabolism, such as peroxisome, drug metabolism cytochrome, and metabolism of xenobiotics by cytochrome P450 (Figure 9F). These results could explain why the low-risk group was resistant to conventional chemotherapy.

4. Discussion

Patients’ treatment strategies need to identify the prognosis and treatment response of patients with HCC. M6A modifications (in both coding and non-coding RNAs) play an important role in cancer pathogenesis and drug response/resistance.[31] Our study constructed an m6-A related IncRNAs signature based on HCC transcriptome data from TCGA. We evaluated the m6A-related signature under various clinical and genetic settings including OS, clinicopathological variables, immune cells infiltration, immune checkpoints expression, chemotherapy response, and immunotherapy response. Our results showed a relatively higher prediction accuracy of the novel model than clinicopathological variables. This m6A-related IncRNAs signature was significantly correlated with some clinicopathological variables, such as tumor stage and grade, and was an independent predictor for OS in HCC patients. A higher risk score was negatively related to the infiltration of active NK cells. Additionally, TP53 mutations were more frequent in the high-risk group, while CTNNB1 mutations were more frequent in the low-risk group. The expressions of immune checkpoints (CTLA4, PDCD1, PDL1, LAG3, TIGIT, and HAVCR2) were positively related to the risk score. Except for PDL1, all these genes were significantly enhanced in the high-risk group. Interestingly, our study also indicated that patients in the high-risk group were more resistant to sorafenib, but they were more sensitive to conventional chemotherapy (cisplatin, doxorubicin, and mitomycin) and immunotherapy.

Our study demonstrated that patients with higher risk scores were in a more suppressive immune microenvironment, with lower infiltration of active NK cells and higher expression of various immune checkpoints. Previous studies also demonstrated that a higher number of NK cells predicted a better outcome in HCC patients. [32, 33] The ability of NK cells in anti-tumor immunity is to kill cancer cells without prior sensitization which plays an important role in tumor immunosurveillance. However, NK cells lost their functions through exposure to inhibitory molecules produced by the tumor in the immunosuppressive microenvironment [34]. Immune checkpoint TIGIT involved in the dysfunction of NK cells. [35] Nowadays, anti-TIGIT antibody drugs, such as tiragolumab, are developing in clinical trials. [36] Our study also found that the expression of TIGIT was positively related with risk score. TIDE analysis showed that patients in high-risk group were more sensitive to immunotherapy. Thus, we could look
forward to the application effect of TIGIT inhibitors alone or in combination with other immune checkpoint inhibitors in patients with HCC, especially in high-risk patients.

Previous deep-sequencing studies had found that TP53 and CTNNB1 were frequently mutated in HCC. Generally, mutations of TP53 and CTNNB1 in HCC were frequently mutually exclusive. [37, 38] Our study indicated that TP53 mutated more frequently in high-risk group, while mutation frequency of CTNNB1 was significantly lower. TP53 mutation was demonstrated to be an indicator of poor treatment response for patients treated with TKI but showed remarkable clinical benefit to PD-1 inhibitors in patients with advanced non-small cell cancer. [39, 40] A recent study also indicated TP53 mutation could be a predictor of better efficacy of immunotherapy in HCC patients. [41] Intriguingly, Wnt/CTNNB1 mutations were also characterized as the immune-excluded class (cold tumors) recently, which might represent the biomarkers predicting resistance to immunotherapy. [42] These studies support the results in our study that the higher response of immunotherapy and the lower response of sorafenib in HCC patients with high-risk group in somatic alterations level. However, further researches need to be conducted with a larger external validation cohort or experiment to provide a better understanding of the above results.

5. conclusions

In summary, we constructed a novel risk signature based on 5 m6A-related lncRNAs to predict the OS of patients with HCC and distinguished the patients who might respond better to different drug therapies. Our study might provide novel insight into the screening strategies of HCC patients in systemic therapy. Specifically, compared with patients in the high-risk group, patients in the low-risk group tended to have better OS, lower response rate to chemotherapy and immunotherapy, and higher response rate to sorafenib.

Declarations

CRediT authorship contribution statement


Data Sharing Statement

All the data generated or analyzed in this study could be download from TCGA database (https://cancergenome.nih.gov/).

Ethical Statement
As the work is a bioinformatics analysis article, ethical approval was not necessary and all the data were retrieved from the free online databases.

**Funding**

This grant of the study was from the Matching start-up fund of the Natural Science Foundation of China (8200110843) and the Natural Science Foundation of Guangdong Province, Grant/Award Number: No.2021A1515011066.

**Conflict of Interest**

The authors declare that there is no conflict of interest.

**References**


**Figures**
Figure 1

Flow chart.
Figure 2

(A) Interaction network of m6A-related lncRNAs and m6-A regulators. (B) Univariate Cox regression analysis of the selected lncRNAs significantly correlated with OS. (C-D) Boxplot and heatmap of the expressions of prognosis m6A-related lncRNAs between normal tissues and tumor tissues. OS, overall survival; *P < 0.05; **P < 0.01; ***P < 0.001.
Figure 3

(A) LASSO COX regression for OS of m6A-related IncRNAs. (B) Multivariate COX for OS of 5 m6A-related IncRNAs in HCC. (C) Picking the best cut-off value of risk score. (D) PCA between the high- and low-risk groups. (E) ROC curves of the risk score and clinical characteristics. (F) Risk score distribution and survival overview assigned to low- and high-risk groups based on the risk score. (G) Kaplan-Meier survival curves of the OS in patients with the low- and high-risk groups. LASSO, least absolute shrinkage, and selection operator. PCA, principal component analysis. OS, overall survival. ROC, receiver operational feature curve.
Figure 4

(A) The heat map showing the expression levels of 5 m6A-related IncRNAs in the low- and high-risk groups. The distribution of clinicopathological features was compared between the low- and high-risk groups. (B) Univariate COX analyses for OS of risk score and clinicopathological variables. (C) Multivariate COX analyses for OS of risk score clinicopathological variables and. *P < 0.05; **P < 0.01; ***P < 0.001.
Figure 5

Kaplan-Meier survival curve of OS for patients with (A) age> 65, (B) age 65, (C) female, (D) grade 1-2, (E) grade 3-4, (F) stage I-II, (G) stage III-IV.
Figure 6

(A) The waterfall plot of genetic alteration established by patients with the high-risk group. (B) The waterfall plot of genetic alteration established by patients with low-risk group. (C) TMB difference in the low- and high-risk groups. (D) The correlation between TMB and risk score. (E) Kaplan-Meier survival curves of the OS in patients with the high- and low TMB. (F) Kaplan-Meier curve analysis of OS is shown.
for patients classified according to the TMB and risk score. TMB, tumor mutational burden. OS, overall survival.

Figure 7

(A) Composition of infiltrated immune cells between low- and high-risk groups with CIBERSORT $p < 0.05$ for all eligible samples. (B) Comparisons of immune cells between high- and low-risk groups. (C) The negative correlation between active NK cell and risk score. (D-I) Comparisons of expression levels of CTLA4, LAG3, PDCD1, TIGIT, HAVCR2, and CD274 between low- and high-risk groups.
Figure 8

(A-F) The correlation of CTLA4, PD1, PDL1, LAG3, TIGIT, and HAVCR2 expressions with the risk score.
Figure 9

(A) Comparison of TIDE score between low- and high-risk groups. (B-E) Boxplot demonstrated the differences in the IC50 values for cisplatin, doxorubicin, mitomycin, and sorafenib estimated by the pRRophetic algorithm between low- and high-risk groups. (F) GSEA of high- and low-risk groups. TIDE, Tumor Immune Dysfunction and Exclusion. IC50, half-maximal inhibitory concentration. GSEA, Gene Set Enrichment Analysis.