


## RESEARCH ARTICLE

# Comprehensive analysis of N<sup>6</sup>-methyladenosine-related long non-coding RNAs for prognosis prediction in liver hepatocellular carcinoma

Hong-Xu Zhu<sup>1</sup> | Wen-Jie Lu<sup>2</sup> | Wei-Ping Zhu<sup>1</sup> | Song Yu<sup>3</sup> 

<sup>1</sup>Department of Hepatic Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, Shanghai, China

<sup>2</sup>Department of General Surgery, School of Medicine, Second Affiliated Hospital of Zhejiang University, Hangzhou, China

<sup>3</sup>Department of General Surgery, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China

## Correspondence

Song Yu and Wei-Ping Zhu, Department of General Surgery, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, No. 600 Yi Shan Road, Shanghai 200233, China.

Emails: yusong1668@163.com; wpzhush@hotmail.com

## Abstract

**Background:** Liver hepatocellular carcinoma (LIHC) is a lethal cancer. This study aimed to identify the N<sup>6</sup>-methyladenosine (m<sup>6</sup>A)-targeted long non-coding RNA (lncRNA) related to LIHC prognosis and to develop an m<sup>6</sup>A-targeted lncRNA model for prognosis prediction in LIHC.

**Methods:** The expression matrix of mRNA and lncRNA was obtained, and differentially expressed (DE) mRNAs and lncRNAs between tumor and normal samples were identified. Univariate Cox and pathway enrichment analyses were performed on the m<sup>6</sup>A-targeted lncRNAs and the LIHC prognosis-related m<sup>6</sup>A-targeted lncRNAs. Prognostic analysis, immune infiltration, and gene DE analyses were performed on LIHC subgroups, which were obtained from unsupervised clustering analysis. Additionally, a multi-factor Cox analysis was used to construct a prognostic risk model based on the lncRNAs from the LASSO Cox model. Univariate and multivariate Cox analyses were used to assess prognostic independence.

**Results:** A total of 5031 significant DE mRNAs and 292 significant DE lncRNAs were screened, and 72 LIHC-specific m<sup>6</sup>A-targeted binding lncRNAs were screened. Moreover, a total of 29 LIHC prognosis-related m<sup>6</sup>A-targeted lncRNAs were obtained and enriched in cytoskeletal, spliceosome, and cell cycle pathways. An 11-m<sup>6</sup>A-lncRNA prognostic model was constructed and verified; the top 10 lncRNAs included LINC00152, RP6-65G23.3, RP11-620J15.3, RP11-290F5.1, RP11-147L13.13, RP11-923I11.6, AC092171.4, KB-1460A1.5, LINC00339, and RP11-119D9.1. Additionally, the two LIHC subgroups, Cluster 1 and Cluster 2, showed significant differences in the immune microenvironment, m<sup>6</sup>A enzyme genes, and prognosis of LIHC.

**Conclusion:** The m<sup>6</sup>A-lncRNA prognostic model accurately and effectively predicted the prognostic survival of LIHC. Immune cells, immune checkpoints (ICs), and m<sup>6</sup>A enzyme genes could act as novel therapeutic targets for LIHC.

## KEYWORDS

liver hepatocellular carcinoma, long non-coding RNA, N<sup>6</sup>-methyladenosine, prognosis prediction

Hong-Xu Zhu and Wenjie Lu should be regarded as co-first authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Clinical Laboratory Analysis* published by Wiley Periodicals LLC.

## 1 | INTRODUCTION

Liver hepatocellular carcinoma (LIHC) is the primary malignant tumor of the liver and a key cause of cancer-related deaths worldwide.<sup>1</sup> LIHC is mostly found in the late stages of the disease, and there are few effective treatments that can improve survival.<sup>2</sup> Currently, in addition to traditional methods including liver transplantation, sorafenib, and other available treatments for LIHC,<sup>3</sup> many combination methods including combinations of different immune checkpoint inhibitors (ICIs) are under investigation.<sup>4,5</sup> For prognostic prediction of patients with LIHC, the patient's physical status, degree of liver dysfunction, tumor expansion, and alpha-fetoprotein (AFP) concentration are the main factors.<sup>6,7</sup> Therefore, the identification and comprehensive analysis of novel molecular biomarkers will benefit the ability to predict patient prognosis in LIHC.

Long non-coding RNA (lncRNA) is a type of non-coding RNA with a length is greater than 200 nucleotides.<sup>8</sup> Importantly, specific lncRNAs can be used as prognostic biomarkers for cancer.<sup>9,10</sup> Many studies have confirmed the role of lncRNAs in LIHC; for example, YAP and Rnux2 inhibit lncRNA MT1DP in LIHC cells.<sup>11</sup> However, the prognostic value of lncRNA regulation requires further exploration in LIHC.

N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) is one of the most common RNA modifications. Changes in m<sup>6</sup>A levels may affect the characteristics of cancer, including maintaining proliferation signals, resisting cell death, and activating invasion and metastasis and may have carcinogenic or suppressive effects in malignant tumors.<sup>12,13</sup> In LIHC, m<sup>6</sup>A genes such as N<sup>6</sup>-adenosine-methyltransferase complex catalytic subunit (METTL3),<sup>14</sup> vir like m<sup>6</sup>A methyltransferase-associated VIRMA (KIAA1429),<sup>15</sup> WT1 associated protein (WTAP),<sup>16</sup> YTH domain family 2 (YTHDF2),<sup>17</sup> and insulin-like growth factor 2 mRNA binding protein 1 (IGF2BP1)<sup>18</sup> promote the occurrence of LIHC by affecting the proliferation and migration of liver cancer cells. In addition, it has been proven that lncRNAs modified by m<sup>6</sup>A methylation could affect the function of target genes through RNA-protein binding, ceRNA mechanism, or RNA-RNA binding to regulate the biological functions of cancer cells.<sup>14,19</sup> For example, m<sup>6</sup>A-mediated up-regulation of LINC00958 increased the expression of hepatocellular carcinoma-derived growth factors, thereby promoting LIHC progression and adipogenesis.<sup>20,21</sup> However, most of the current research on lncRNAs modified by m<sup>6</sup>A methylation is based on the correlation between the expression level of the m<sup>6</sup>A enzyme genes and the expression level of lncRNAs; therefore, the lncRNA obtained may not necessarily be targeted or functionally related to the m<sup>6</sup>A enzyme genes. In addition, the potential prognostic mechanism of m<sup>6</sup>A-lncRNAs in LIHC remains unclear, and further research is needed.

In this study, the specific lncRNAs binding to m<sup>6</sup>A were obtained from the TCGA-LIHC and m<sup>6</sup>A2Target databases. Multiple sets of data mining analysis methods were used for m<sup>6</sup>A-related LIHC subgroups, and a series of in-depth analyses of immune infiltration and prognosis of different subgroups were performed. This research may contribute to improving our understanding of the molecular mechanisms of LIHC from the perspective of m<sup>6</sup>A-related lncRNAs.

The data analysis flowchart of the m<sup>6</sup>A-targeted lncRNAs in LIHC in this study is shown in Figure 1.

## 2 | MATERIALS AND METHODS

### 2.1 | Data acquisition and preprocessing

After downloading transcriptome sequencing data (FPKM) and clinical data (phenotype) from the TCGA-LIHC database (website: <https://xenabrowser.net/>),<sup>22</sup> a total of 418 samples were obtained, of which 368 were tumor samples. The Gencode database (<https://www.gencodegenes.org/>) was used for gene annotation of the transcriptome (hg38, V22),<sup>23</sup> which converted Ensembl\_ID to Symbol\_ID (if multiple Ensembl\_IDs corresponded to the same symbol ID, the average value of Ensembl\_ID was defined as the final expression value of Symbol\_ID). The annotated "protein\_coding" was mRNA, and annotated "lincRNA", "non\_coding", "3prime\_overlapping\_ncrna", "sense\_intronic", "processed\_transcript", "sense\_overlapping", "antisense". Finally, mRNA and lncRNA expression values were obtained.

### 2.2 | Differential expression analysis

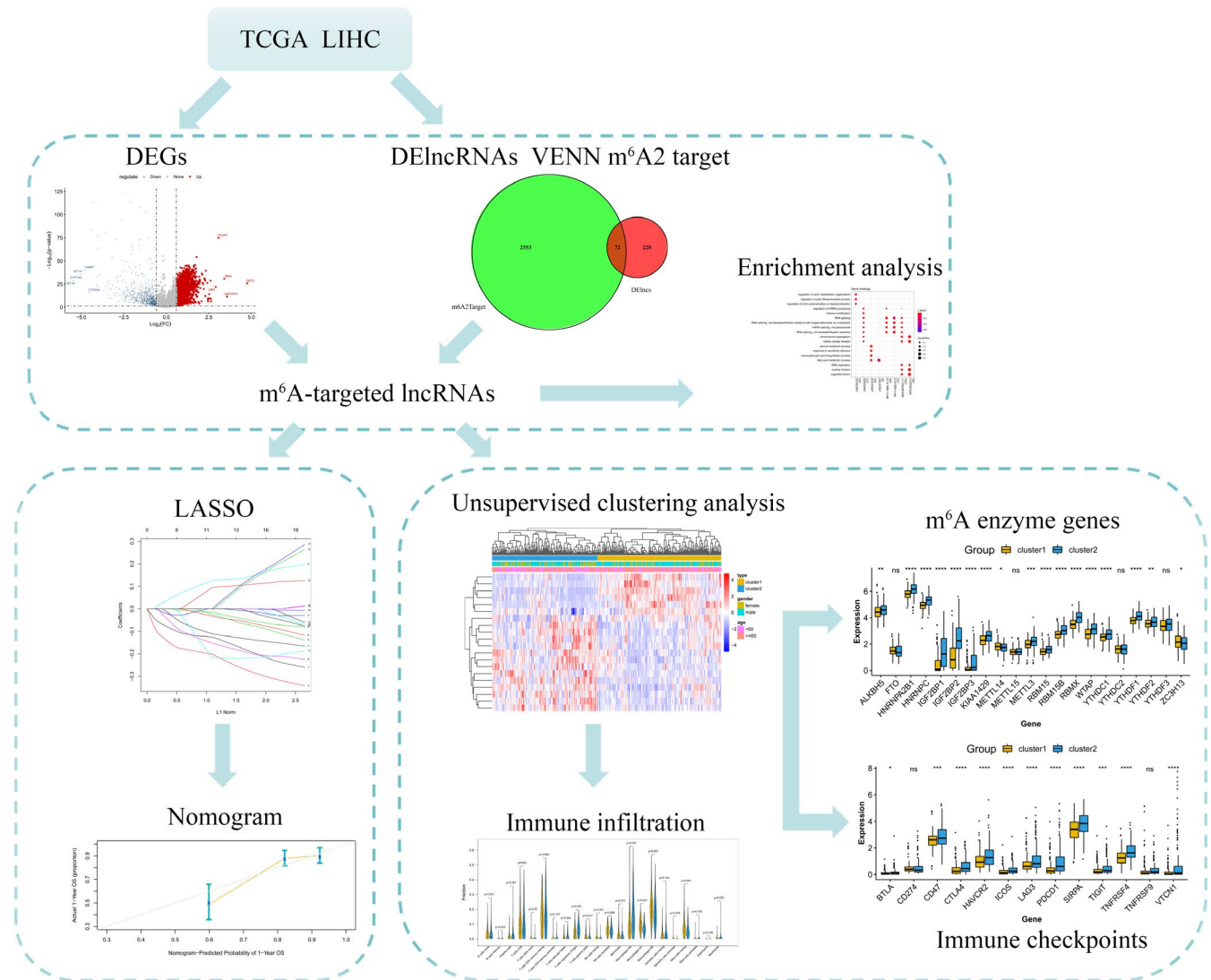
The *t* test method in the R language limma package (Version 3.10.3, website: <http://www.bioconductor.org/packages/release/bioc/html/limma.html>) was used to test the difference in the mean value of gene expression between normal and tumor samples.<sup>24</sup> The difference was represented by the *P* value. In addition, to make the results more reliable, multiple testing adjustments (Benjamini & Hochberg method) were performed for the *P* value correction. Differential expression analysis was performed, and RNAs with *P* < .05, and |logFC| > 0.585 were finally regarded as differentially expressed (DE) mRNAs and lncRNAs.

### 2.3 | Recognition of M<sup>6</sup>A-targeted lncRNAs

The experimentally verified lncRNAs that had a target binding relationship with m<sup>6</sup>A enzyme were obtained from the m<sup>6</sup>A WER Target Gene Database (m6A2Target, <http://m6a2target.canceromics.org/#/>).<sup>25</sup> Moreover, by taking the intersection with the above-mentioned DElncRNAs, unique m<sup>6</sup>A-targeted lncRNAs in LIHC were obtained.

### 2.4 | Screening of lncRNAs related to LIHC prognosis

The clinical information related to prognosis in TCGA LIHC, including overall survival (OS) and OS status, was collected. In addition, survival (Version 3.2-7, website: <http://bioconductor.org/packages/survival/>) was used to sequentially perform univariate Cox regression analysis on the unique m<sup>6</sup>A-targeted binding lncRNAs in LIHC.<sup>26</sup>



**FIGURE 1** Flow chart of data analysis of N<sup>6</sup>-methyladenosine (m<sup>6</sup>A)-targeted long non-coding RNAs (lncRNAs) in liver hepatocellular carcinoma (LIHC)

In addition, a  $P < .05$  was considered as the threshold for identifying lncRNAs that were obviously related to the prognosis of LIHC.

## 2.5 | Functional analysis of lncRNAs

To screen m<sup>6</sup>A-targeted lncRNAs related to LIHC prognosis that were significantly related to DE mRNAs at the gene expression level, correlation analysis was performed on the DE mRNAs and the above-mentioned LIHC prognosis-related lncRNAs via the `cor` function in the R language, and the Pearson correlation coefficient (PCC) was calculated. The threshold was set to  $PCC > 0.6$  and  $P < .05$ . In addition, the enrichment analysis tool `clusterProfiler` was used for mRNA analysis of co-expression relationship pairs (Version 3.16.0, website: <http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>).<sup>27</sup> The number of enriched genes count  $\geq 2$  and the significance threshold  $P < .05$  were regarded as significant enrichment.

## 2.6 | Unsupervised clustering analysis

Based on the lncRNAs in the co-expression relationship obtained above, the `heatmap`<sup>28</sup> (Version 1.0.8, website: [ps://cran.r-project.org/web/packages/heatmap/index.html](https://cran.r-project.org/web/packages/heatmap/index.html)) in R was used to perform unsupervised clustering analysis for the acquisition and identification of different LIHC subgroups.

## 2.7 | Prognostic Kaplan-Meier (K-M) survival analysis and clinical information association analysis between subgroups

To explore the differences in the prognosis of LIHC between the different subgroups, the K-M survival prognostic curve was used. In addition, the clinical information in each subgroup, including sex ratio, age distribution, TNM classification (TNM), and tumor stage (stage i, ii, iii, iv), was statistically collected to explore the clinical information

association between the subgroups. The chi-square test was used to calculate the *P* value, which evaluated differences in the different clinical characteristics of each subgroup. Clinical factors, with a *P* < .05, were considered to be related to the subgroup.

## 2.8 | Analysis for the abundance of immune infiltration of subgroups

To explore differences in the abundance of immune infiltration between subgroups, the tumor tissue expression profile data were used to analyze the infiltration of 22 immune cells in each subgroup based on the Cibersort algorithm.<sup>29,30</sup> The LM22 data set was used as the template gene expression profile, and the parameters were set to perm = 100 and QN = F.

## 2.9 | Analysis of immune checkpoints and M<sup>6</sup>A enzyme gene expression levels between subgroups

To detect the immune checkpoint (ICS) and M<sup>6</sup>A gene expression levels in different subgroups, information about the expression levels of 13 ICs and 21 M<sup>6</sup>A enzyme genes was extracted from the tumor samples in the TCGA data set.

## 2.10 | Construction of prognostic risk prediction model

The samples were randomly divided into two sets at a ratio of 1:1, namely the training set (125 cases) and the validation set (125 cases). First, based on the lncRNAs in the co-expression network, 10-fold cross-validation was performed on the training set using the LASSO Cox regression model of the R package glmnet (Version 4.0-2, website: <https://cran.r-project.org/web/packages/glmnet/index.html>) to obtain the best lambda value. In addition, lncRNAs with a regression coefficient of non-zero (ie, the optimal combination of characteristic lncRNA markers) were obtained. In addition, the prognostic coefficients of each element in the lncRNA combination and their expression levels in samples from the training set obtained by the multi-factor Cox regression analysis of the survival package were applied to construct the following prognostic risk model. The calculation formula is as follows:

$$\text{Risk score} = \sum \text{Coef gene} \times \text{Exp gene} \left( \begin{array}{l} \text{(Coef gene represented the prognostic coefficient of multivariate Cox regression analysis,} \\ \text{and Exp gene represented the gene expression level in the TCGA data set)} \end{array} \right)$$

## 2.11 | Prognostic risk model validation

To verify that the risk score model constructed by the lncRNAs could separate the samples into high- and low-risk groups with significant

differences in survival rates, the training set was used to calculate the  $\beta$  coefficients of the above model, and the validation set was used to reconstruct the model using the  $\beta$  coefficients calculated by the training set. Finally, the corresponding K-M survival curve was drawn to observe whether the prognoses of the high-risk and low-risk groups were significantly different.

### 2.11.1 | Independence analysis of prognosis model and Nomogram chart establishment

Univariate and multivariate Cox analyses were used to evaluate whether the prognostic model was an independent prognostic factor. A log-rank *P* < .05 was chosen as the threshold for screening significant correlation. In addition, a normality chart was drawn, and the index of concordance (C-index) was obtained to evaluate the predictive ability of the nomogram chart according to independent prognostic factors.

## 3 | RESULTS

### 3.1 | Analysis of differential expression of mRNAs and lncRNAs

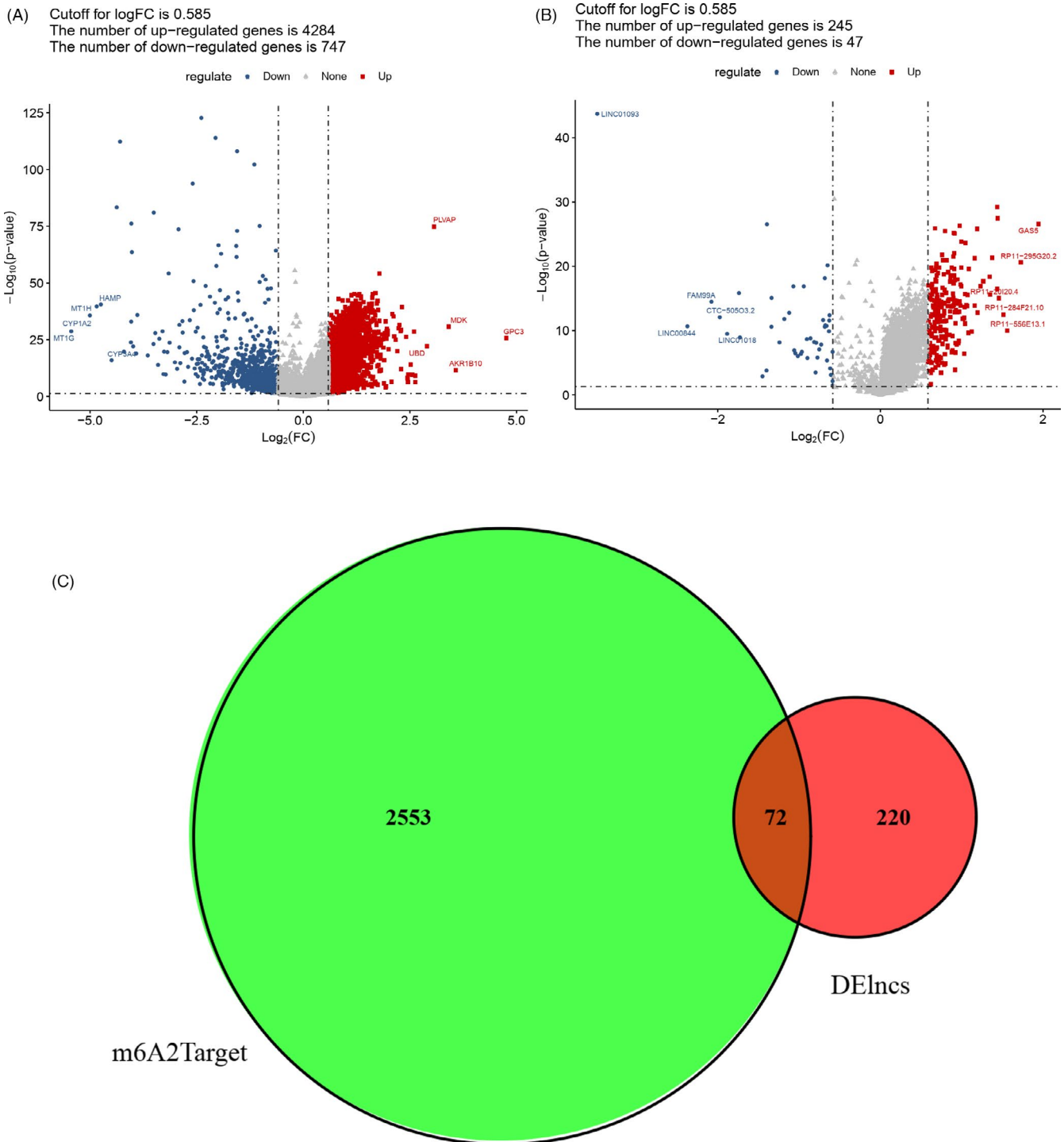
After re-annotation, a total of 19712 mRNAs and 3873 lncRNAs were obtained. A total of 5031 significantly DE mRNAs and 292 significantly DE lncRNAs were screened after taking  $|\log FC| > 0.585$  and *P* < .05 as the significance threshold. The volcano plot of DE mRNAs and lncRNAs are shown in Figure 2A,B.

### 3.2 | M<sup>6</sup>A-targeted lncRNAs related to the prognosis of LIHC

The 2625 m<sup>6</sup>A enzyme-targeted lncRNAs were intersected with the DE lncRNAs, and a total of 72 LIHC-specific m<sup>6</sup>A-targeted binding lncRNAs were subjected to single-factor Cox regression analysis (Figure 2C). Furthermore, 29 lncRNAs that were significantly associated with LIHC prognosis were identified (Table 1).

### 3.3 | Enrichment analysis

By calculating the correlation between the expression levels of lncRNAs and mRNAs that were significantly related to the prognosis of



**FIGURE 2** Analysis of liver hepatocellular carcinoma (LIHC) prognostic-related m<sup>6</sup>A-targeted lncRNAs. A, Volcano plot of differentially expressed mRNAs. B, Volcano plot of differentially expressed lncRNAs. The red squares represented up-regulated genes, and the blue circles represented down-regulated genes. The dots showing the names of genes were the top 5 genes with up-regulated or down-regulated expression. C, Venn diagram of m<sup>6</sup>A enzyme-targeted lncRNAs and differentially expressed lncRNAs

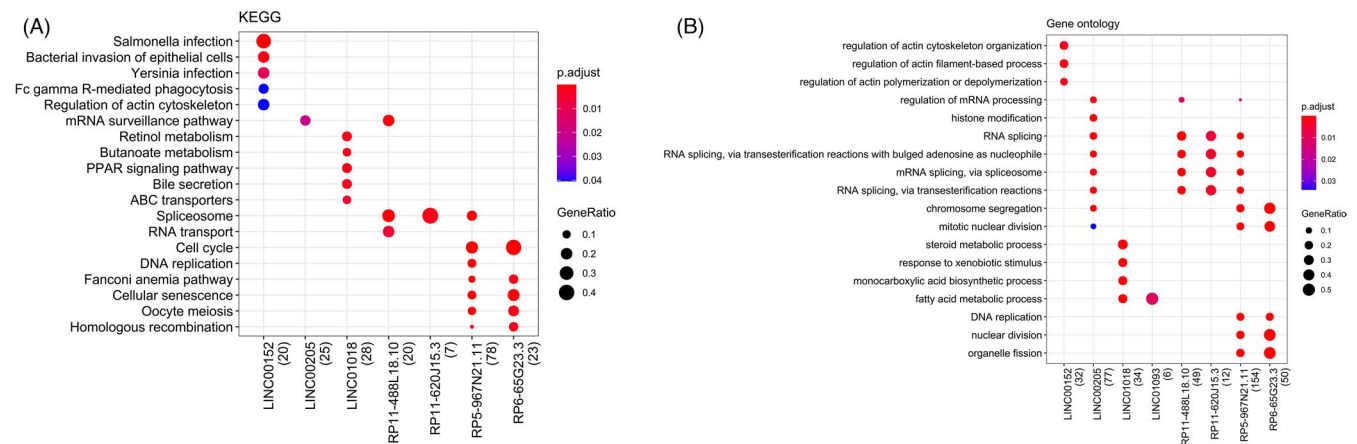
LIHC, a total of 522 pairs of co-expression relationships were obtained, involving 408 DEmRNAs and 20 DElncRNAs. In addition, the possible pathways and functions of lncRNAs can be predicted based on the co-expression relationship. Finally, 19 KEGG pathways and 18 GO terms were enriched (Figure 3). According to the *P* value and

gene ratio, the top three enriched KEGG pathways were as follows: (1) spliceosome; (2) cell cycle; and (3) Salmonella infection. The top three enriched GO terms were as follows: (1) RNA splicing; (2) RNA splicing via transesterification reactions with bulged adenosine as a nucleophile; and (3) mRNA splicing via the spliceosome.



TABLE 1 Total 29 lncRNAs significantly related to the prognosis of LIHC

symbol	Hazard. Ratio	P value	symbol	Hazard. Ratio	P value
LINC00152	1.360 (1.156–1.600)	0	LINC00339	1.447 (1.075–1.948)	.015
LINC01138	1.967 (1.427–2.712)	0	LINC01018	0.895 (0.819–0.979)	.015
RP6-65G23.3	1.629 (1.250–2.123)	0	RP11-1094M14.11	1.327 (1.049–1.680)	.018
CMB9-22P13.1	1.357 (1.154–1.595)	0	RP11-119D9.1	0.775 (0.628–0.958)	.018
RP11-620J15.3	1.463 (1.162–1.842)	.001	RP11-160O5.1	1.228 (1.036–1.457)	.018
LINC00205	1.550 (1.185–2.026)	.001	NUP50-AS1	1.312 (1.044–1.649)	.02
RP11-290F5.1	0.732 (0.607–0.884)	.001	PVT1	1.354 (1.044–1.755)	.022
RP11-147L13.13	1.555 (1.172–2.065)	.002	RP11-488L18.10	1.243 (1.029–1.501)	.024
LINC01554	0.883 (0.815–0.956)	.002	RP5-967N21.11	1.277 (1.033–1.577)	.024
AC007405.6	1.453 (1.131–1.869)	.004	RAB30-AS1	1.479 (1.045–2.095)	.027
RP11-923I11.6	1.336 (1.089–1.639)	.005	RP11-390P24.1	0.759 (0.591–0.974)	.03
AC092171.4	1.357 (1.093–1.684)	.006	RP11-513G11.3	0.786 (0.633–0.977)	.03
KB-1460A1.5	1.410 (1.104–1.800)	.006	LINC00261	0.852 (0.733–0.990)	.036
LINC00665	1.295 (1.068–1.572)	.009	LINC01093	0.875 (0.770–0.994)	.04
AC079466.1	1.137 (1.027–1.257)	.013			



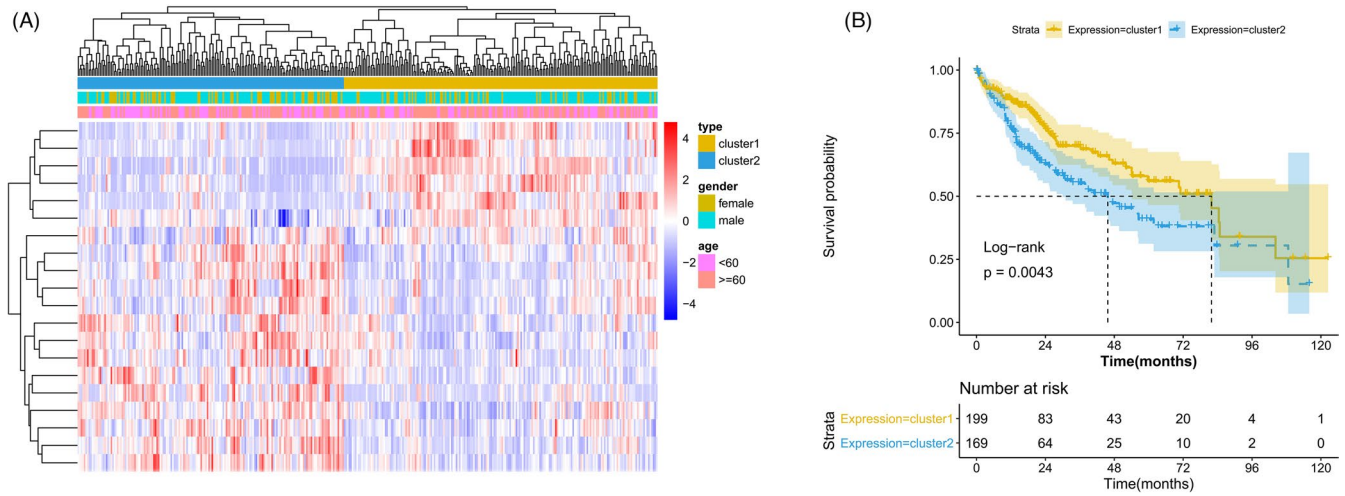
**FIGURE 3** Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene ontology (GO) enrichment analysis of liver hepatocellular carcinoma (LIHC) prognosis-related genes. A, KEGG pathways enrichment bubble chart. B, GO terms enrichment bubble chart. The horizontal axis represented the number of enriched genes, and the vertical axis represented the name of the pathway. The size of the dot was the ratio of the number of enriched genes to the total number of genes. The larger the ratio, the bigger the dot. The redder the dot color, the more significant the difference (indicated by *P* value)

### 3.4 | Prognostic K-M survival analysis and clinical information association analysis

The expression matrix of 20 lncRNAs related to LIHC prognosis in co-expression relationship pairs was classified through unsupervised clustering analysis, and two LIHC subgroups, namely, Cluster 1 and Cluster 2, were obtained (Figure 4A). For the two subgroups, the survival data of TCGA LIHC were used for the K-M survival analysis. The results suggested that there were significant differences in survival rates between the two subgroups (Figure 4B, *P* = .0043). Moreover, the results of clinical information association analysis showed significant differences in gender and pathology\_T between Clusters 1 and 2 (Table 2. The *P* values were .008 and .021, respectively).

### 3.5 | Abundance analysis for immune infiltration between subgroups

The 22 immune infiltration abundances between Clusters 1 and 2 were calculated using the CIBERSORT algorithm. After the 22 cells were subjected to the Mann-Whitney *U* test, the results showed that there was a significant difference in the proportion of eight immune cells between the two subgroups (Figure 5A). For example, compared with Cluster 1, the proportion of immune cells in Cluster 2, such as follicular helper T cells, regulatory T cells (Tregs), and delta gamma T cells, significantly increased. Conversely, the proportions of resting NK cells, M1 macrophages, M2 macrophages, and resting mast cells were significantly reduced.



**FIGURE 4** Prognostic analysis and clinical phenotype of subgroups. A, Kaplan-Meier survival analysis. B, The distribution of WHO classifications in different subgroups

### 3.6 | Immune checkpoints and M<sup>6</sup>A gene expression levels in different subgroups

For the ICs and M<sup>6</sup>A enzyme genes, the expression levels were detected and analyzed between Clusters 1 and 2. The results indicated that the expression levels of 13 ICs and 21 m<sup>6</sup>A genes were different between the two subgroups (Figure 5B,C). Compared with Cluster 2, except that the expression levels of CD274 molecule (*CD274*) and TNF receptor superfamily member 9 (*TNFRSF9*) were not significantly different between the two subgroups, the expression levels of the remaining 11 genes were significantly down-regulated in Cluster 1 (all *P* < .05). In addition, compared with Cluster 2, the expression levels of key LIHC m<sup>6</sup>A oncogenes *KIAA1429*, *WTAP*, *YTHDF2*, and *IGF2BP1* were significantly decreased in Cluster 1 (all *P* < .05).

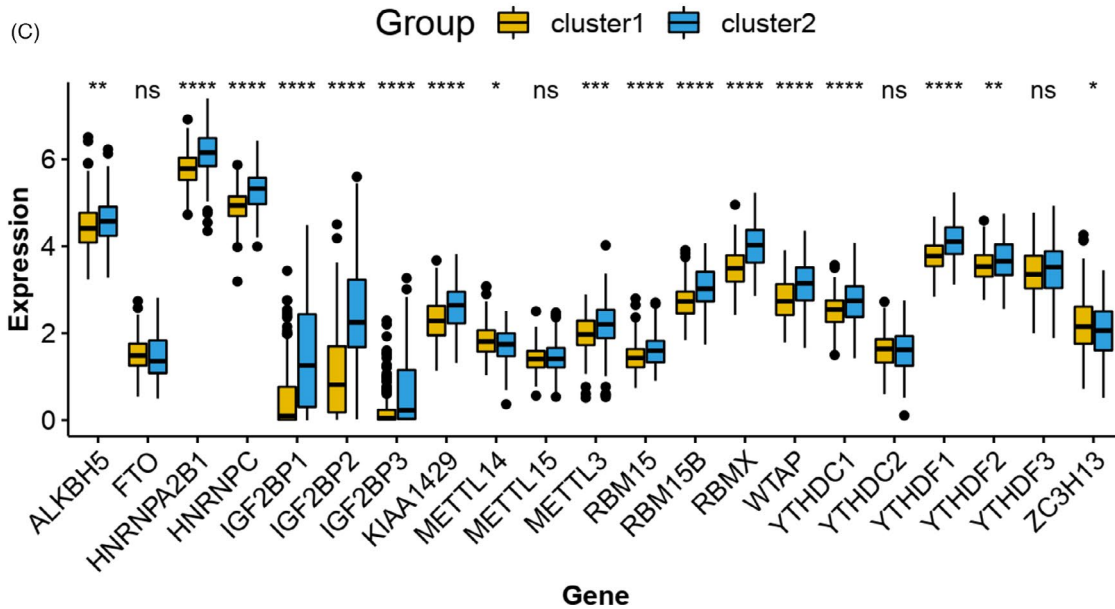
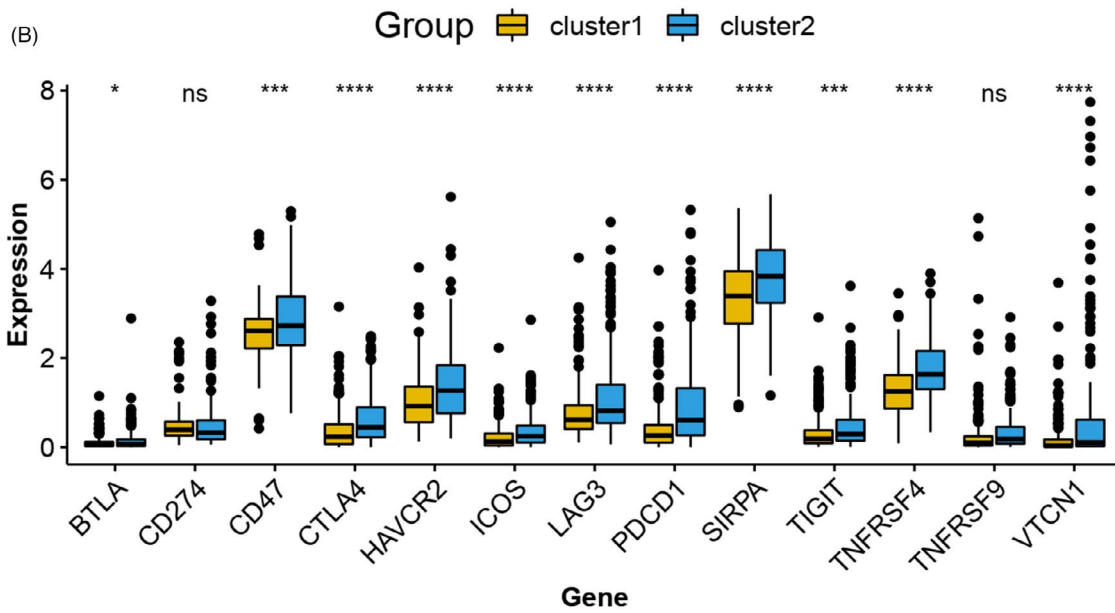
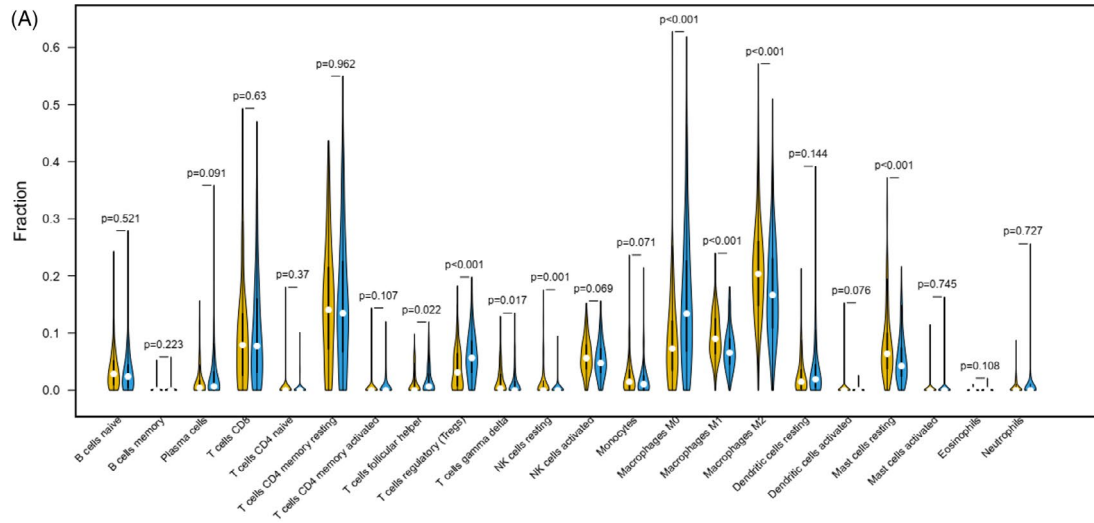
### 3.7 | LIHC prognostic risk model construction and verification

For prognostic lncRNAs targeted by m<sup>6</sup>A in the co-expression relationship, a total of 11 lncRNAs were further screened using the LASSO Cox regression model. The 11 lncRNAs, namely LINC00152, RP6-65G23.3, RP11-620J15.3, LINC00205, RP11-290F5.1, RP11-147L13.13, RP11-923I11.6, AC092171.4, KB-1460A1.5, LINC00339, and RP11-119D9.1, were used to construct an 11-lncRNA signature. Because the LIHC prognostic risk model was based on multiple m<sup>6</sup>A-targeted lncRNAs, it was also named the multi-m<sup>6</sup>A-targeted lncRNA model.

The β coefficients in the Riskscore calculation formula after constructing the model using multi-factor Cox regression analysis are shown in Table 3. Furthermore, according to whether the risk score was greater than the median value, the 11-lncRNA signature divided patients into high-risk and low-risk groups in the training and validation sets, respectively. Combining the overall survival time and survival status of the sample, the log-rank test was applied to

**TABLE 2** Comparison of clinical characteristics distribution differences between different LIHC subgroups

	Cluster 1 (N = 199)	Cluster 2 (N = 69)	P value
Gender			
Female	52.0 (26.1%)	67.0 (39.6%)	.008
Male	147 (73.9%)	102 (60.4%)	
Age (years)			
<60	81.0 (40.7%)	84.0 (49.7%)	.104
≥60	118 (59.3%)	85.0 (50.3%)	
Stage			
NA	14.0 (7.0%)	10.0 (5.9%)	.055
Stage i	105 (52.8%)	67.0 (39.6%)	
Stage ii	41.0 (20.6%)	44.0 (26.0%)	
Stage iii	36.0 (18.1%)	47.0 (27.8%)	
Stage iv	3.00 (1.5%)	1.00 (0.6%)	
T			
T1	112 (56.3%)	70.0 (41.4%)	.021
T2	43.0 (21.6%)	49.0 (29.0%)	
T3	33.0 (16.6%)	45.0 (26.6%)	
T4	8.00 (4.0%)	5.00 (3.0%)	
TX	1.00 (0.5%)	0 (0%)	
Missing	2.00 (1.0%)	0 (0%)	
N			
N0	130 (65.3%)	120 (71.0%)	.175
N1	1.00 (0.5%)	3.00 (1.8%)	
NX	68.0 (34.2%)	45.0 (26.6%)	
Missing	0 (0%)	1.00 (0.6%)	
M			
M0	136 (68.3%)	129 (76.3%)	.231
M1	2.00 (1.0%)	1.00 (0.6%)	
MX	61.0 (30.7%)	39.0 (23.1%)	





**FIGURE 5** Results of abundance analysis for immune infiltration, immune checkpoints genes, and m<sup>6</sup>A genes between subgroups. A, Violin plot of the abundance distribution of 22 immune infiltrating cells. B, The expression levels of immune checkpoints genes in Cluster 1 and Cluster 2. C, The expression levels of m<sup>6</sup>A genes in Cluster 1 and Cluster 2. Yellow and blue graphics represent Cluster 1 and Cluster 2, respectively

perform the K-M survival analysis. The survival curves are shown in Figure 6. The results showed that the prognosis of the high-risk group was obviously worse than that of the low-risk group (*P* values in the training and validation sets were both less than .05). In addition, the clinical information, subgroups, and significant differences in the two different risk level groups were separately counted. The results are presented in Table 4. Except for M and T, the *P* values of other clinical characteristics between the high-risk and low-risk groups were greater than 0.05. Therefore, the prognostic value of

the multi-m<sup>6</sup>A-targeted lncRNA model was significantly related to the clinical features of M and T.

### 3.8 | Screening of independent prognostic clinical factors

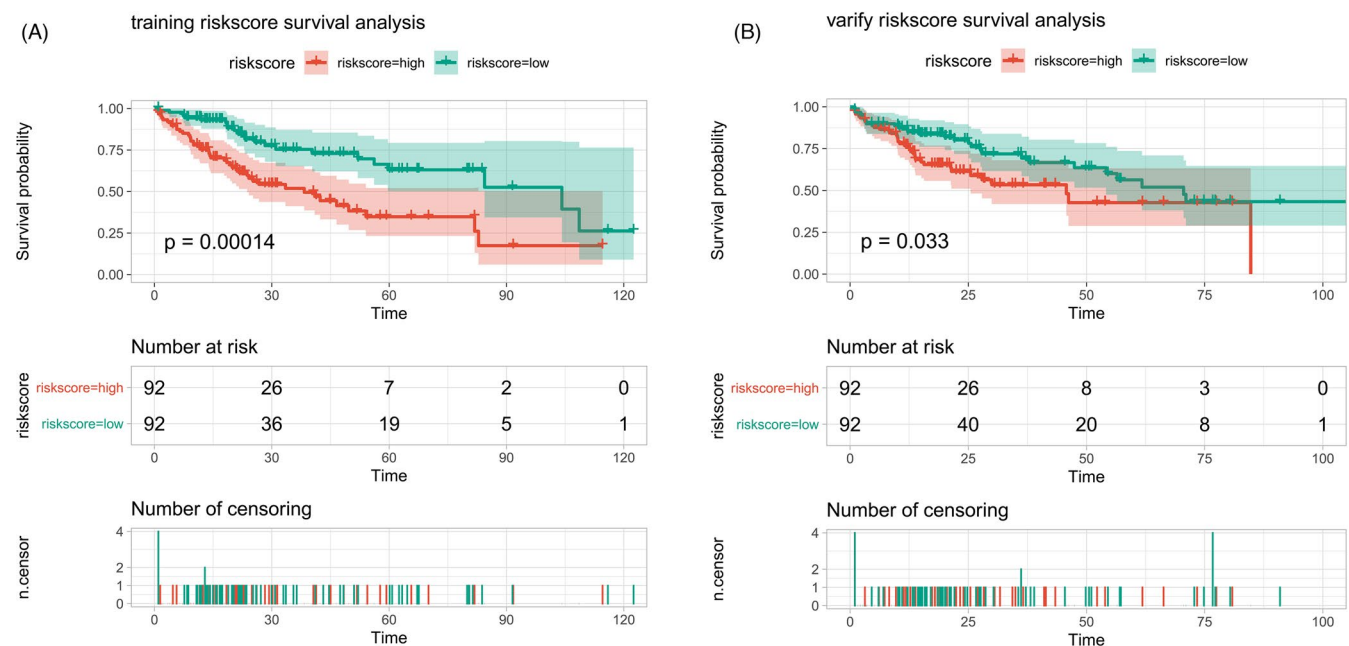
After selecting a log-rank *P* < .05, as the threshold for screening significant correlation, the Cox analysis results are shown in Figure 7A. Univariate and multivariate Cox regression analyses showed that the independent prognostic factors might be pathological T and risk groups. In addition, a normality chart was drawn and verified as an independent prognostic factor (Figure 7B,C). Pathology\_T and risk groups constituted a nomogram survival prediction model, which had good prediction ability for the survival rate of patients with LIHC. Therefore, the nomogram model, which consisted of pathological T and risk score evaluation models, could be used as an independent factor to predict LIHC prognosis.

**TABLE 3**  $\beta$  coefficients in the Riskscore calculation formula corresponding to lncRNAs in the prognostic risk model

LncRNA	$\beta$
LINC00152	0.121444
RP6-65G23.3	0.262191
RP11-620J15.3	-0.05283
LINC00205	-0.17789
RP11-290F5.1	-0.10845
RP11-147L13.13	0.339291
RP11-923I11.6	0.28707
AC092171.4	-0.09625
KB-1460A1.5	0.197882
LINC00339	0.582752
RP11-119D9.1	-0.06094

## 4 | DISCUSSION

The high morbidity and mortality associated with LIHC seriously endangers human health.<sup>31</sup> An increasing number of reports have confirmed the pathological significance of m<sup>6</sup>A in cancers, including LIHC.<sup>32</sup> However, early research has mainly focused on the



**FIGURE 6** Kaplan-Meier (K-M) survival curve in the high- and low-risk groups of training-set and valid-set models. A, Risk score survival analysis in training set. B, Risk score survival analysis in valid set

**TABLE 4** Statistics of clinical characteristics of high-risk and low-risk groups

	High risk (N = 184)	Low risk (N = 184)	P value
<b>Gender</b>			
Female	58.0 (31.5%)	61.0 (33.2%)	.824
Male	126 (68.5%)	123 (66.8%)	
<b>Age (years)</b>			
<60	84.0 (45.7%)	81.0 (44.0%)	.834
≥60	100 (54.3%)	103 (56.0%)	
<b>T</b>			
T1	74.0 (40.2%)	108 (58.7%)	<.001
T2	57.0 (31.0%)	35.0 (19.0%)	
T3	48.0 (26.1%)	30.0 (16.3%)	
T4	4.00 (2.2%)	9.00 (4.9%)	
TX	0 (0%)	1.00 (0.5%)	
Missing	1.00 (0.5%)	1.00 (0.5%)	
<b>N</b>			
N0	130 (70.7%)	120 (65.2%)	.48
N1	2.00 (1.1%)	2.00 (1.1%)	
NX	51.0 (27.7%)	62.0 (33.7%)	
Missing	1.00 (0.5%)	0 (0%)	
<b>M</b>			
M0	141 (76.6%)	124 (67.4%)	.049
M1	0 (0%)	3.00 (1.6%)	
MX	43.0 (23.4%)	57.0 (31.0%)	
<b>Subset</b>			
Cluster 1	71.0 (38.6%)	128 (69.6%)	<.001
Cluster 2	113 (61.4%)	56.0 (30.4%)	

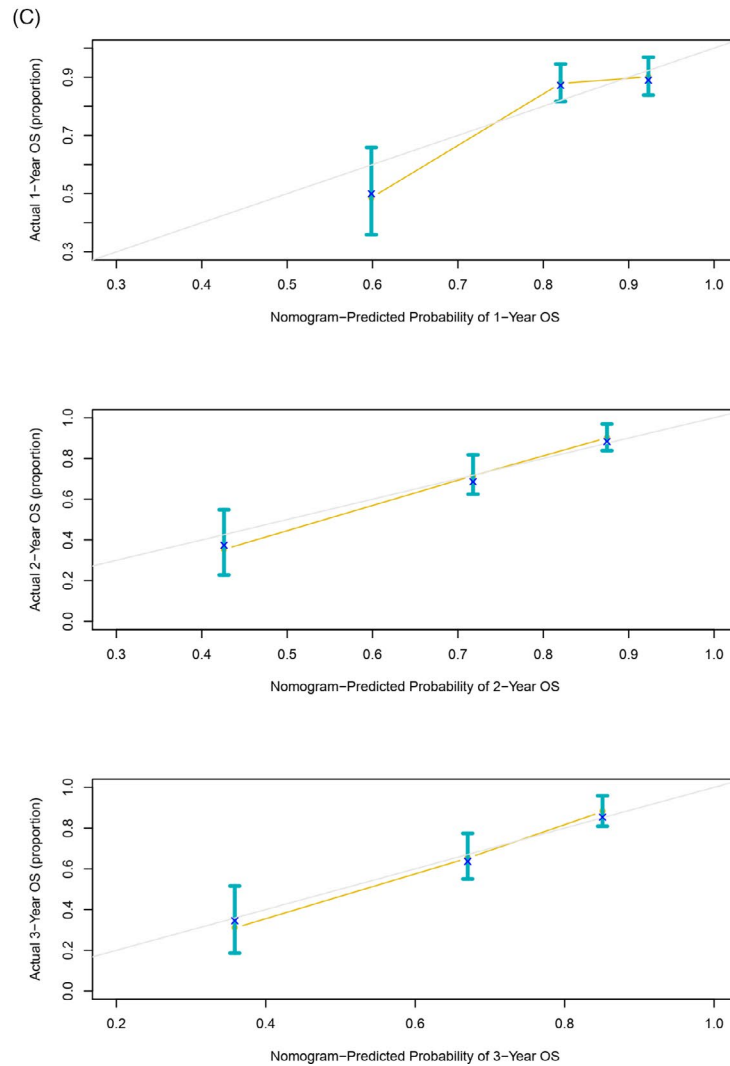
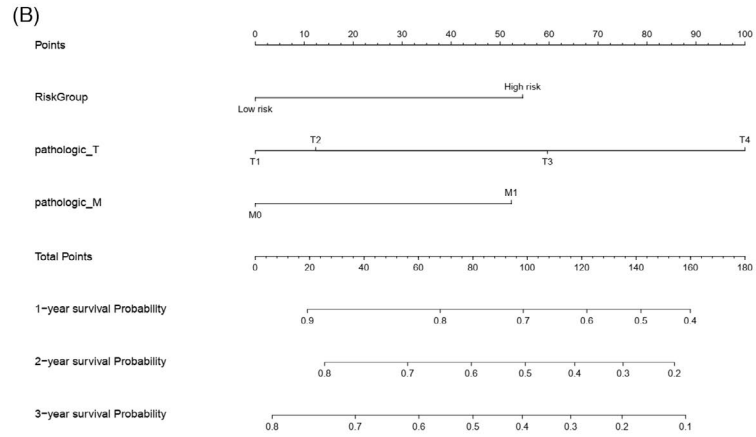
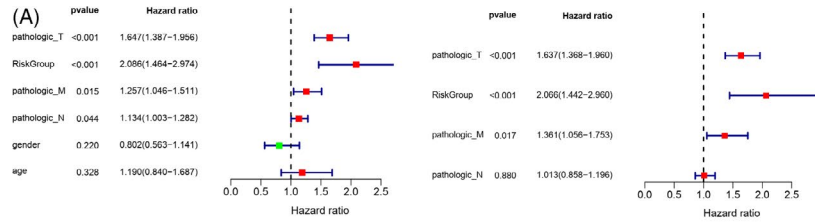
relevance of specific m<sup>6</sup>A-related genes or their corresponding pathways in tumor diagnosis and treatment. Therefore, the lack of evidence for a comprehensive analysis of m<sup>6</sup>A-related lncRNAs in LIHC should be resolved immediately. The identification and analysis of m<sup>6</sup>A-related lncRNAs in large LIHC cohorts is important for the research direction and objectives of LIHC. In this study, 29 prognostic m<sup>6</sup>A-lncRNAs were found to have different expression levels in normal tissues and tumor tissues, and the role of m<sup>6</sup>A-lncRNAs in LIHC was further explored.

It can be seen from the functional analysis of LIHC prognosis-related m<sup>6</sup>A-targeted lncRNAs that the number of lncRNAs involved in cytoskeleton regulation, spliceosome, and cell cycle pathways was the largest. Correspondingly, the lncRNAs related to prognosis involving the above pathways were LINC00152, RP11-620J15.3, and RP6-65G23.3, respectively, which were also identified and used to construct the LIHC prognostic risk model. LINC00152 is a cytoskeleton regulator RNA and has been identified as an effective oncogene in various cancer types, participating in cancer cell proliferation and invasion by combining tumor suppressor microRNAs as competitive endogenous RNA with gene promoters.<sup>33,34</sup> Although LINC00152 can be used for prognosis and potential biomarker therapy,<sup>35</sup> its

molecular regulation mechanism for the prognosis of LIHC has not yet been reported. The GO terms enrichment results showed that LINC00152 was mainly enriched in actin cytoskeleton organization, actin filament-based process, and actin polymerization or depolymerization regulatory pathways. Moreover, LINC00152 expression was modified by m<sup>6</sup>A methylation and participated in the prognosis of LIHC patients through the cytoskeleton regulation pathway. Therefore, LINC00152 can be used as a novel potential LIHC prognostic marker. As for RP11-620J15.3, it was the former transcript of hepatocarcinogenesis (GIHCG), and its overexpression was related to poor prognosis and immune infiltration of hepatocellular carcinoma (HCC).<sup>36,37</sup> However, the molecular regulatory mechanism of RP11-620J15.3 on the prognosis of LIHC remains unclear. Through the enrichment analysis of KEGG pathways and GO terms, RP11-620J15.3 was confirmed to be enriched in the spliceosome and RNA splicing pathways, which were linked to the occurrence and prognosis of many cancers, including LIHC.<sup>38-41</sup> For example, a meta-analysis of gene expression profiles conducted by Xu et al. showed that genes in the spliceosome pathway (including *HSPA1A*, *SNRPE*, *SF3B2*, *SF3B4*, and *TRA2A* genes) were up-regulated in LIHC.<sup>42</sup> Therefore, it was speculated that the m<sup>6</sup>A methylation target gene RP11-620J15.3 directly or indirectly participated in the splicing, RNA splicing, and mRNA splicing pathways and affected the prognosis of patients with LIHC. As for the cell cycle pathway, a variety of lncRNAs, including *CDKN2B-AS1* and *HOTAIR*, are considered to be involved in promoting LIHC cell proliferation and growth.<sup>43,44</sup> However, the role of these lncRNAs in the prognosis of LIHC has not been studied clearly. In this study, RP6-65G23.3, a promising biomarker for the prognosis of LIHC, was identified as the target gene of the m<sup>6</sup>A enzyme gene. In summary, the LINC00152, RP11-620J15.3, and RP6-65G23.3 identified in this study were not only the target genes of the m<sup>6</sup>A enzyme gene, but also novel prognostic predictors of LIHC.

In addition to obtaining promising novel predictors of LIHC prognosis, through a comprehensive analysis of m<sup>6</sup>A-targeted lncRNAs, the correlation between these lncRNAs and the prognosis as well as immune microenvironment of TCGA patients was also explored and analyzed. After obtaining two LIHC subgroups that were highly correlated with the clinical characteristics and performing the analysis of immune infiltration, it was speculated that the good prognosis of Cluster 1 was due to the highest proportion of immune cells such as follicular helper T cells, regulatory T cells (Tregs), and delta gamma T cells. These T cells not only participate in fibrotic diseases by regulating fibrosis,<sup>45</sup> but are also related to the prognosis of LIHC. Previous studies have shown that the low number of delta gamma T cells in the liver tissue around the tumor was related to the higher recurrence rate of LIHC, which could predict postoperative recurrence, especially in patients with early LIHC.<sup>46,47</sup> In addition, the analysis results of the expression levels of ICs and M<sup>6</sup>A enzyme genes in Cluster 1 and Cluster 2 showed that the good prognosis of Cluster 1 was also related to the down-regulation of most ICs and M<sup>6</sup>A enzyme genes. It can be seen that the LIHC subgroups obtained based on the clustering analysis of m<sup>6</sup>A-targeted lncRNAs were effective.

**FIGURE 7** Analysis of the liver hepatocellular carcinoma (LIHC) prognostic risk model. A, Univariate and multivariate Cox analysis of prognostic model. B, Nomogram chart of independent prognostic clinical factors. C, Nomogram verification chart for independent prognostic clinical factors. The abscissa represents the survival rate of patients predicted by the model, and the ordinate represents the survival rate of patients actually observed



In previous studies, m<sup>6</sup>A-lncRNA might affect the migration and proliferation of LIHC cells through the regulation of the Notch signaling pathway.<sup>48</sup> Furthermore, the m<sup>6</sup>A-lncRNA-related prognosis model developed by LASSO regression was used to determine the survival rate of patients with LIHC.<sup>49</sup> In our current research, the m<sup>6</sup>A-lncRNA-related prognosis model was constructed according to the 11-lncRNA. In addition, both univariate Cox analysis and multivariate Cox analysis indicated that pathological T and risk groups could be independent prognostic factors. The nomogram model, which consisted of pathological T and risk score evaluation model, could be used as an independent factor to predict the prognosis of LIHC. Therefore, the LIHC prognostic prediction model established based on m<sup>6</sup>A-targeted lncRNAs has been well verified, which once again showed the important role of these lncRNAs in the prognosis of LIHC.

The types of LIHC samples in this study were incomplete, which had certain limitations on the results of the study. In addition, the research on lncRNAs that had predictive significance for the prognosis of LIHC was not very in-depth. Therefore, the next research plan is to expand the number of samples and enrich the types of samples (including LIHC samples of different nationalities, etc.) to further confirm the role of data mining analysis and clustering analysis of m<sup>6</sup>A-targeted lncRNAs in predicting the prognosis of LIHC. Moreover, the target genes of lncRNAs related to the 11-lncRNA signature should be further screened, which is conducive to a better understanding of the molecular mechanism of LIHC prognosis regulation. In addition, relevant experiments are needed to validate the lncRNAs identified from the bioinformatics analysis.

## 5 | CONCLUSION

In conclusion, the 11-lncRNA prognostic prediction signature constructed in this study could help to evaluate the prognosis of patients with LIHC. Besides, the key immune cells (macrophages, resting NK cells, follicular helper T cells, etc.) play an important role in the progression of LIHC. This research has strengthened the understanding of m<sup>6</sup>A-related lncRNAs and immune infiltration in LIHC, which may provide new targets for LIHC therapy and prognostic-related biomarkers for further research and evaluation.

### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

### AUTHOR CONTRIBUTIONS

Song Yu conceptualized and designed the research. Wenjie Lu involved in acquisition of data, and analysis and interpretation of data. Hong-Xu Zhu involved in statistical analysis and drafting the manuscript. Wei-Ping Zhu revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

### DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this published article.

### ORCID

Song Yu  <https://orcid.org/0000-0002-4296-0830>

### REFERENCES

- Balogh J, Victor D III, Asham EH, et al. Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma*. 2016;3:41.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
- Tejeda-Maldonado J, García-Juárez I, Aguirre-Valadez J, et al. Diagnosis and treatment of hepatocellular carcinoma: an update. *World J Hepatol*. 2015;7(3):362.
- Longo V, Brunetti O, Gnoni A, et al. Emerging role of immune checkpoint inhibitors in hepatocellular carcinoma. *Medicina*. 2019;55(10):698.
- Xu F, Jin T, Zhu Y, Dai C. Immune checkpoint therapy in liver cancer. *J Exp Clin Cancer Res*. 2018;37(1):1-12.
- Attwa MH, El-Etreby SA. Guide for diagnosis and treatment of hepatocellular carcinoma. *World J Hepatol*. 2015;7(12):1632.
- Ouyang X, Fan Q, Ling G, Shi Y, Hu F. Identification of diagnostic biomarkers and subtypes of liver hepatocellular carcinoma by multi-omics data analysis. *Genes*. 2020;11(9):1051.
- Uszczyńska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet*. 2018;19(9):535-548.
- Bolha L, Ravnik-Glavač M, Glavač D. Long noncoding RNAs as biomarkers in cancer. *Dis Markers*. 2017;2017:1-14.
- Quan J, Pan X, Zhao L, et al. LncRNA as a diagnostic and prognostic biomarker in bladder cancer: a systematic review and meta-analysis. *Oncol Targets Ther*. 2018;11:6415.
- Yu W, Qiao Y, Tang X, et al. Tumor suppressor long non-coding RNA, MT1DP is negatively regulated by YAP and Runx2 to inhibit FoxA1 in liver cancer cells. *Cell Signal*. 2014;26(12):2961-2968.
- Vu LP, Cheng Y, Kharas MG. The biology of m<sup>6</sup>A RNA methylation in normal and malignant hematopoiesis. *Cancer Discov*. 2019;9(1):25-33.
- Zhao W, Qi X, Liu L, Ma S, Liu J, Wu J. Epigenetic regulation of m<sup>6</sup>A modifications in human cancer. *Mol Ther Nucleic Acids*. 2020;19:405-412.
- Chen M, Wei L, Law CT, et al. RNA N<sup>6</sup>-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. *Hepatology*. 2018;67(6):2254-2270.
- Cheng X, Li M, Rao X, et al. KIAA1429 regulates the migration and invasion of hepatocellular carcinoma by altering m<sup>6</sup>A modification of ID2 mRNA. *Oncol Targets Ther*. 2019;12:3421.
- Chen Y, Peng C, Chen J, et al. WTAP facilitates progression of hepatocellular carcinoma via m<sup>6</sup>A-HuR-dependent epigenetic silencing of ETS1. *Mol Cancer*. 2019;18(1):1-19.
- Yang Z, Li J, Feng G, et al. MicroRNA-145 modulates N<sup>6</sup>-methyladenosine levels by targeting the 3'-untranslated mRNA region of the N<sup>6</sup>-methyladenosine binding YTH domain family 2 protein. *J Biol Chem*. 2017;292(9):3614-3623.
- Müller S, Glaß M, Singh AK, et al. IGF2BP1 promotes SRF-dependent transcription in cancer in a m<sup>6</sup>A- and miRNA-dependent manner. *Nucleic Acids Res*. 2019;47(1):375-390.
- Kwok C-T, Marshall AD, Rasko JE, Wong JJ. Genetic alterations of m<sup>6</sup>A regulators predict poorer survival in acute myeloid leukemia. *J Hematol Oncol*. 2017;10(1):1-6.
- Rong D, Dong Q, Qu H, et al. m<sup>6</sup>A-induced LINC00958 promotes breast cancer tumorigenesis via the miR-378a-3p/YY1 axis. *Cell Death Discovery*. 2021;7(1):1-9.
- Unfried JP, Sangro P, Prats-Mari L, Sangro B, Fortes P. The landscape of lncRNAs in hepatocellular carcinoma: a translational perspective. *Cancers*. 2021;13(11):2651.

22. Goldman M, Craft B, Hastie M, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv*. 2019. <https://doi.org/10.1101/326470>
23. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9):1760-1774.
24. Smyth GK, Ritchie M, Thorne N, Wettenhall J. LIMMA: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Statistics for Biology and Health. 2005. [https://doi.org/10.1007/0-387-29362-0\\_23](https://doi.org/10.1007/0-387-29362-0_23)
25. Deng S, Zhang H, Zhu K, et al. M6A2Target: a comprehensive database for targets of m<sup>6</sup>A writers, erasers and readers. *Brief Bioinform*. 2021;22(3):bbaa055.
26. Therneau TM. *Extending the cox model*; Proceedings of the First Seattle Symposium in Biostatistics. 1997;123:51-84.
27. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.
28. Kolde R, Kolde MR. Package 'pheatmap'. *R Package*. 2015;1(7):790.
29. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Cancer systems biology*. Springer; 2018:243-259.
30. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457.
31. Mazzaferro V, Citterio D, Bhoori S, et al. Liver transplantation in hepatocellular carcinoma after tumour downstaging (XXL): a randomised, controlled, phase 2b/3 trial. *Lancet Oncol*. 2020;21(7):947-956.
32. Li Y, Qi D, Zhu B, Ye X. Analysis of m<sup>6</sup>A RNA methylation-related genes in liver hepatocellular carcinoma and their correlation with survival. *Int J Mol Sci*. 2021;22(3):1474.
33. Moradi M-T, Hatami R, Rahimi Z. Circulating CYTOR as a potential biomarker in breast cancer. *Int J Mol Cell Med*. 2020;9(1):83.
34. Seo D, Kim D, Kim W. Long non-coding RNA linc00152 acting as a promising oncogene in cancer progression. *Genomics Inform*. 2019;17(4):e36.
35. Zhang J, Li W. Long noncoding RNA CYTOR sponges miR-195 to modulate proliferation, migration, invasion and radiosensitivity in nonsmall cell lung cancer cells. *Biosci Rep*. 2018;38(6):BSR20181599.
36. Xiao S, Huang S, Yang J. Overexpression of GIHCG is associated with a poor prognosis and immune infiltration in hepatocellular carcinoma. *Onco Targets Ther*. 2020;13:11607-11619.
37. Yi T, Luo H, Qin F, et al. LncRNA LL22NC03-N14H11. 1 promoted hepatocellular carcinoma progression through activating MAPK pathway to induce mitochondrial fission. *Cell Death Dis*. 2020;11(10):1-17.
38. Lei Y, Shi Y, Duan J, et al. Identification of alternative splicing and lncRNA genes in pathogenesis of small cell lung cancer based on their RNA sequencing. *Adv Clin Exp Med*. 2019;28(8):1043-1050.
39. Tang Y, Cheung BB, Atmadibrata B, et al. The regulatory role of long noncoding RNAs in cancer. *Cancer Lett*. 2017;391:12-19.
40. Yan J, Zhou C, Guo K, Li Q, Wang Z. A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma. *J Cell Biochem*. 2019;120(1):213-223.
41. Yang S, Zhou Y, Zhang X, et al. The prognostic value of an autophagy-related lncRNA signature in hepatocellular carcinoma. *BMC Bioinform*. 2021;22(1):1-16.
42. Xu W, Huang H, Yu L, Cao L. Meta-analysis of gene expression profiles indicates genes in spliceosome pathway are up-regulated in hepatocellular carcinoma (HCC). *Med Oncol*. 2015;32(4):96.
43. Li H, An J, Wu M, et al. LncRNA HOTAIR promotes human liver cancer stem cell malignant growth through downregulation of SETD2. *Oncotarget*. 2015;6(29):27847.
44. Zhuang H, Cao G, Kou C, Li D. Overexpressed lncRNA CDKN2B-AS1 is an independent prognostic factor for liver cancer and promotes its proliferation. *J BUON*. 2019;24(4):1441-1448.
45. Zhang M, Zhang S. T cells in fibrosis and fibrotic diseases. *Front Immunol*. 2020;11:1142.
46. Cai X-Y, Wang J-X, Yi Y, et al. Low counts of  $\gamma\delta$  T cells in peritumoral liver tissue are related to more frequent recurrence in patients with hepatocellular carcinoma after curative resection. *Asian Pac J Cancer Prev*. 2014;15(2):775-780.
47. Zhao Y, Niu C, Cui J. Gamma-delta ( $\gamma\delta$ ) T cells: friend or foe in cancer development? *J Transl Med*. 2018;16(1):1-13.
48. Meurette O, Mehlen P. Notch signaling in the tumor microenvironment. *Cancer Cell*. 2018;34(4):536-548.
49. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T. N<sup>6</sup>-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res*. 2017;45(10):6051-6063.

**How to cite this article:** Zhu H-X, Lu W-J, Zhu W-P, Yu S. Comprehensive analysis of N<sup>6</sup>-methyladenosine-related long non-coding RNAs for prognosis prediction in liver hepatocellular carcinoma. *J Clin Lab Anal*. 2021;35:e24071. <https://doi.org/10.1002/jcla.24071>