Am J Cancer Res 2022;12(7):3259-3279
www.ajcr.us /ISSN:2156-6976/ajcr0143625

Original Article

Development and validation of prognostic m6A-related lncRNA and mRNA model in thyroid cancer

Yu Zhu[1*], Tian Yu[2,3*], Ju Huang[4], Xitao Ma[5], Tao Shen[5], Annuo Li[1], Rensong Yue[1]

[1]Department of Endocrinology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610075, Sichuan, P. R. China; [2]Department of General Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 1 Shuaifuyuan, Wangfujin, Dongcheng District, Beijing 100730, P. R. China; [3]Graduate School, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100005, P. R. China; [4]Department of Oncology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610075, Sichuan, P. R. China; [5]Internal Medicine, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610075, Sichuan, P. R. China. [*]Equal contributors.

Received April 19, 2022; Accepted June 25, 2022; Epub July 15, 2022; Published July 30, 2022

Abstract: Although N6-methyladenosine (m6A) regulators and lncRNAs influence the carcinogenesis of thyroid cancer (THCA), the association between m6A-related lncRNAs and THCA remains unexplored. Therefore, we have developed and validated a prognostic model based on m6A-related lncRNAs and mRNAs in THCA. Data from the Cancer Genome Atlas were used to analyze the expression and prognostic characteristics of m6A-related lncRNAs and mRNAs in THCA. Univariate Cox regression analysis was used to screen out independent prognostic factors, while Lasso Cox regression was performed to construct m6A-related lncRNA and mRNA models. The correlation between the prognostic models and gene mutation, immune cell infiltration, tumor microenvironment score, tumor mutational burden, and microsatellite instability were assessed. The prognostic models showed excellent accuracy in predicting the prognosis of patients with THCA. Our study established an m6A-related nomogram capable of predicting the prognosis of patients with THCA. In addition, the hub lncRNAs and mRNAs provide insight into improving the prognosis of THCA. These findings can improve our understanding of m6A modifications in THCA and the prognosis and treatment strategies of THCA.

Keywords: N6-methyladenosine, endocrine, cancer, long non-coding RNA, tumor, nomogram

Introduction

The incidence of thyroid cancer, the most common type of endocrine cancer, has significantly increased over the past 10 years, thereby posing a significant public health threat [1-3]. Miller [4] has studied the incidence of tumors in adolescents and young adults in the United States and found that overall cancer incidence in that population has increased during 2007-2016, largely attributed to thyroid cancer, which rose by approximately 3% annually among those aged 20-39 years and 4% among those aged 15-19 years.

Thyroid cancer can be divided into different pathological types [5, 6], the most common which is papillary thyroid carcinoma derived from thyroid follicular epithelial cells and accounts for 80% of all reported cases. Meanwhile, the other pathological types include follicular thyroid cancer, poorly differentiated thyroid cancer, and undifferentiated thyroid cancer [6]. Although thyroid cancer generally exhibits indolent behavior and has favorable prognosis, approximately 20-30% of all patients with THCA experience recurrence or distant metastasis following the initial treatment [5], some patients do not respond well to conventional treatment, leading to poor prognosis [5].

N6-methyladenosine (m6A) modification is the most common post-transcriptional modification of messenger RNA (mRNA) and non-coding RNA that plays a vital role in RNA maturation, stability, translation, export, and decay [7, 8]. To date, m6A modifications have been identified in more than 7,600 genes and 300 non-coding RNAs in mammals [7]. As a reversible dynamic RNA epigenetic process, the molecular components of m6A RNA methylation involve intracellular methyltransferase, demethylase, and signal trans-

ducers [9], which regulate gene expression and interact with various biological functions, such as RNA splicing, export, stability, translation, and ncRNA biogenesis [10]. Evidence has shown that m6A affects several important life processes including cell division and proliferation, focal death, and apoptosis. Abnormal m6A modification plays a key role in the occurrence and development of a variety of tumors including THCA [8]. Song et al. [11] demonstrated that β-catenin represses miR455-3p to stimulate m6A modification of HSF1 mRNA and promote its translation in colorectal cancer. However, data on the prognostic value of m6A-related genes in THCA remain limited.

Long non-coding RNA (lncRNA) is a non-protein-coding RNA transcript of more than 200 nucleotides [12, 13]; with the development of high-throughput sequencing technology, more than 20,000 lncRNAs have been studied [12]. Long non-coding RNAs are involved in various processes such as transcription, splicing, and translation and have a significant influence in various human diseases, including cancer [14-17]. They are abnormally expressed in various types of tumors and play a key regulatory role [17], and play a significant role in various biological events such as cell cycle, differentiation, proliferation, apoptosis, metastasis, and invasion of various tumor cells [15]. Previous studies have explored the lncRNA expression profile of THCA and demonstrated their potential use as biomarkers for the prognosis and diagnosis of THCA [18-20]. Notably, lncRNA modified by m6A can affect the stability of the transcript and gene expression, resulting in abnormal regulation, which in turn affects the occurrence and development of tumors [21, 22]. Yang et al. [23] have reported that METTL14 suppresses proliferation and metastasis of colorectal cancer by downregulating oncogenic lncRNA X-inactive specific transcript. However, the effect of m6A-modified lncRNA on the prognosis of THCA remains poorly understood.

In this study, we explore the correlation between m6A-related lncRNA and mRNA and the clinicopathological characteristics of THCA. A prognostic model of m6A-related lncRNA and m6A-related mRNA was constructed. The independent prognostic factors obtained via multivariate Cox analysis were integrated into a nomogram to determine the probability of predicting the 1-, 3-, and 5-year progression-free survival in patients with thyroid cancer. The calibration curve was used to evaluate the predictive effect of the nomogram. Furthermore, we assessed the correlation between the prognostic models of m6A-lncRNA and m6A-mRNA with gene mutations, tumor microenvironment, and clinicopathological characteristics. The results of this study provides new insights into the potential mechanism underlying THCA development and facilitate clinical prognosis prediction.

Methods

Data sources and preprocessing

In this study, 23 m6A methylation-related regulators were obtained from previously published studies [24-27]. First, the Cancer Genome Atlas (TCGA) somatic mutation data were downloaded using the TCGAbiolinks R package [28] and visualized using the maftoolR package. The count sequencing data of tumor tissues from 502 patients with THCA and 58 matched control tissues were downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). The corresponding clinical data were obtained from the University of California Santa Cruz (UCSC) Xena platform (http://genome.ucsc.edu). By integrating clinical data with sequencing data and excluding samples with progression-free-survival of shorter than one month, we included a total of 497 patients with THCA. The caret (classification and regression training) package in R was used to divide the datasets of 497 patients with THCA into training set and validation set at a 6:4 ratio for subsequent model construction. To ensure the comparability of gene expression between different samples, RNA sequencing data were standardized to TPM (transcripts per kilobase million) format data. Based on the annotation data of the human genome, the genes of the sequencing data were divided into protein-coding and lncRNAs genes. The Hmisc package in R was used to analyze the correlation between the m6A gene and lncRNA and other mRNA molecules. The absolute value of the correlation coefficient > 0.8 (0.6) and P value < 0.05 were used as the screening criteria for m6A-related mRNAs (m6A-related lncRNAs). The TPM datasets of TCGA and GTEx were downloaded from the UCSC Xena platform for subsequent pan-cancer analysis of gene expression. Finally, the merged dataset con-

tained a total of 33 types of tumors and 31 types of normal tissue sequencing expression profiles.

Construction and validation of prognostic model based on m6A-related lncRNAs and m6A regulators

In this study, data from 298 patients with THCA were used as the internal training set for constructing the prognostic risk model. Univariate Cox regression analysis was used to identify m6A-related lncRNA and mRNA closely related to the prognosis of progression-free survival (*P* value < 0.05 was used as the screening index). Then, the glmnet package in R was used to perform Lasso Cox regression analysis to identify the genes associated with progression-free survival in patients with THCA [29]. To determine the appropriate lambda value and lasso (least absolute shrinkage and selection operator) correlation coefficient, 1,000 repetitions of 10-fold cross-validation lasso regression analysis were performed on the training set to determine the minimum lambda value as the criterion for screening the most suitable gene set. The risk score of each THCA patient is based on the following algorithm:

$$Risk\ Score = \sum_i^n (Exp_i Coef_i) \quad (1)$$

where Exp_i is the expression value of each lncRNA or mRNA, and $Coef_i$ represents the regression coefficient of the corresponding lncRNA or mRNAs. Based on the correlation between the risk score in the training set and the patient's progression-free survival, we used the Sur_cutpoint function in the Survminer package to calculate the best cut-off that divides the population into high and low risk groups [30]. The minimum number of samples not below 40% of the total number of samples in the dataset was set. The Kaplan-Meier survival curve combined with the log-rank test in the survival package in R was used to distinguish the prognostic difference between high and low risk groups. In the training set, validation set, and the entire dataset, the prediction accuracy of the lncRNA or mRNA prognostic risk model was evaluated using the receiver operating characteristic (ROC) curve.

Interaction network construction

The interaction between m6A-related mRNAs and m6A molecules was visualized using the Cytoscape software (version 3.7.2). The CytoHubba plug-in was used to calculate the maximum correlation standard value of the nodes in the interaction network [31]. The top 25 hub genes in the network were obtained, and the Pearson's correlation coefficient between m6A-mRNA and m6A-lncRNA molecules was calculated. The absolute value of the correlation coefficient > 0.85 and *P* value < 0.05 were used as the screening criteria for the correlation between mRNA and lncRNA. Cytoscape software was then used to visualize the mRNA-lncRNA interaction.

Gene enrichment analysis, gene set enrichment analysis, and gene set variation analysis

To evaluate the biological processes associated with m6A-related mRNA molecules and mRNAs in the lncRNA-mRNA network, we used the clusterProfiler package to perform gene ontology (GO) or Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis [32]. The entry screening criteria were Adj*P*.value < 0.05 and *q*.value < 0.05, with the Benjamini-Hochberg used as a *P* value correction method. The clusterProfiler package was also used to perform gene set enrichment analysis on the high- and low-risk subgroups of the mRNA prognostic model in the entire TCGA dataset to identify the enrichment pathways closely associated with the survival of the two risk groups. The parameters used in the gene set enrichment analysis were as follows: the seed was 2,020, the number of calculations was 1,000, the number of genes contained in each gene set was at least 10, and the number of genes contained at most was 500; the Benjamini-Hochberg used as a *P* value correction method. The reference gene set was derived from the c2.cp.v7.2.symbols.gmt file, while the screening criteria for significant enrichment were set to P < 0.05 with a false discovery rate of less than 0.20. The gene set variation analysis (GSVA) was performed on the molecular expression profile of m6A-related mRNAs through the GSVA package in R [33]. First, the enriched pathways were quantitatively scored, and the difference between the tumor and normal tissues was analyzed. Furthermore, pathways significantly enriched between the two groups were obtained. The selection conditions of the enrichment pathway were *P* value < 0.05 and |logFC| > 1.5. The reference enriched

gene set was obtained from the c2.cp.v7.4.symbols.gmt file.

Identification of independent prognostic factors

To identify the prognostic factors independently related to progression-free survival, we included the mRNA risk score, age, gender, and tumor grade to perform univariate and multivariate Cox regression analyses. $P < 0.1$ was used as the screening condition for independent prognostic factors.

Construction and validation of the novel prognostic nomogram

The independent prognostic factors obtained using multivariate Cox analysis were integrated into a nomogram to determine the probability of predicting the 1-, 3-, and 5-year progression-free survival of patients with THCA. The calibration curve was used to evaluate the prediction effect of the nomogram. To confirm reliability, decision curve analysis was used to compare nomograms with tumor grades and risk scores.

Assessment of tumor microenvironment

In this study, the ESTIMATE algorithm in the estimate package in R was used to calculate the immune and interstitial scores of patients with thyroid cancer [34]. The CIBERSORT algorithm was used to estimate the percentage of 22 immune cell subtypes in each tumor sample [35]. The number of replacements in the algorithm was 100, while no interquartile standardization was performed. The difference in the immune infiltration between the high-risk and low-risk groups was estimated using this result.

Statistical analysis

All analyses were performed using R statistical software (Version 4.0.5). Continuous variables were expressed as mean ± standard deviation. The Wilcoxon rank sum test was used to compare the two groups. The Kruskal-Wallis test method was used to compare three or more groups. The comparison of progression-free survival between the two groups was performed via Kaplan-Meier analysis combined with the log-rank test. $P < 0.05$ indicated significant difference.

Results

Expression profile of m6A RNA methylation regulatory gene in THCA

m6A regulatory genes play an important role in the progression of malignant tumors; hence, we explored the differences in the expression of 23 m6A methylation-related genes between THCA tumor tissues and normal tissues. As shown in **Figure 1A**, **1B**, the expression of m6A methylation genes was significantly different between tumor and normal tissues, with 19 m6A regulators differentially expressed between tumor and normal tissues. Except for ELAVL1 and IGF2BP2 genes, which are highly expressed in tumor tissues, the remaining 17 m6A genes were significantly low in tumor tissues, including all Erasers (ALKBH5 and FTO), 7 Writers (CBLL1, KIAA1429, METL14, METL3, RBM15, WTAP and ZC3H13), and 8 Readers (HNRNPA2B1, IGF2BP1, IGF2BP3, LRPPRC, YTHDC1, YTHDC2, YTHDF1, YTHDF3). These findings indicate that m6A RNA methylation regulators play an important role in THCA.

Identification of m6A-related lncRNAs and mRNAs in THCA

To identify potential m6A-related lncRNA and mRNA molecules, we calculated Pearson's correlation coefficient between m6A methylation regulatory genes and the expression of lncRNA and mRNA molecules. In this study, based on the screening criteria of the absolute value of the correlation coefficient > 0.6 and P value < 0.05, 2,253 pairs of m6A regulatory gene-lncRNA correlations were obtained, including 771 lncRNAs. In addition, based on the screening criteria of the absolute value of the correlation coefficient > 0.8 and the P value < 0.05, a total of 5,531 m6A regulator-mRNA correlation pairs, including 1,898 mRNAs, was obtained. **Figure 1C** illustrates the interaction network of m6A regulators and related mRNAs. The red nodes are m6A RNA methylation regulatory genes, while the blue nodes are m6A-related mRNAs. The top 25 hub genes of the maximum correlation standard ranking were obtained from the interaction network diagram through Cytohubba plug-in of Cytoscape, including 8 non-m6A genes (**Figure 1D**). Figure S1A shows a heat map of the differential expression of eight genes in tumor and normal tissues. We

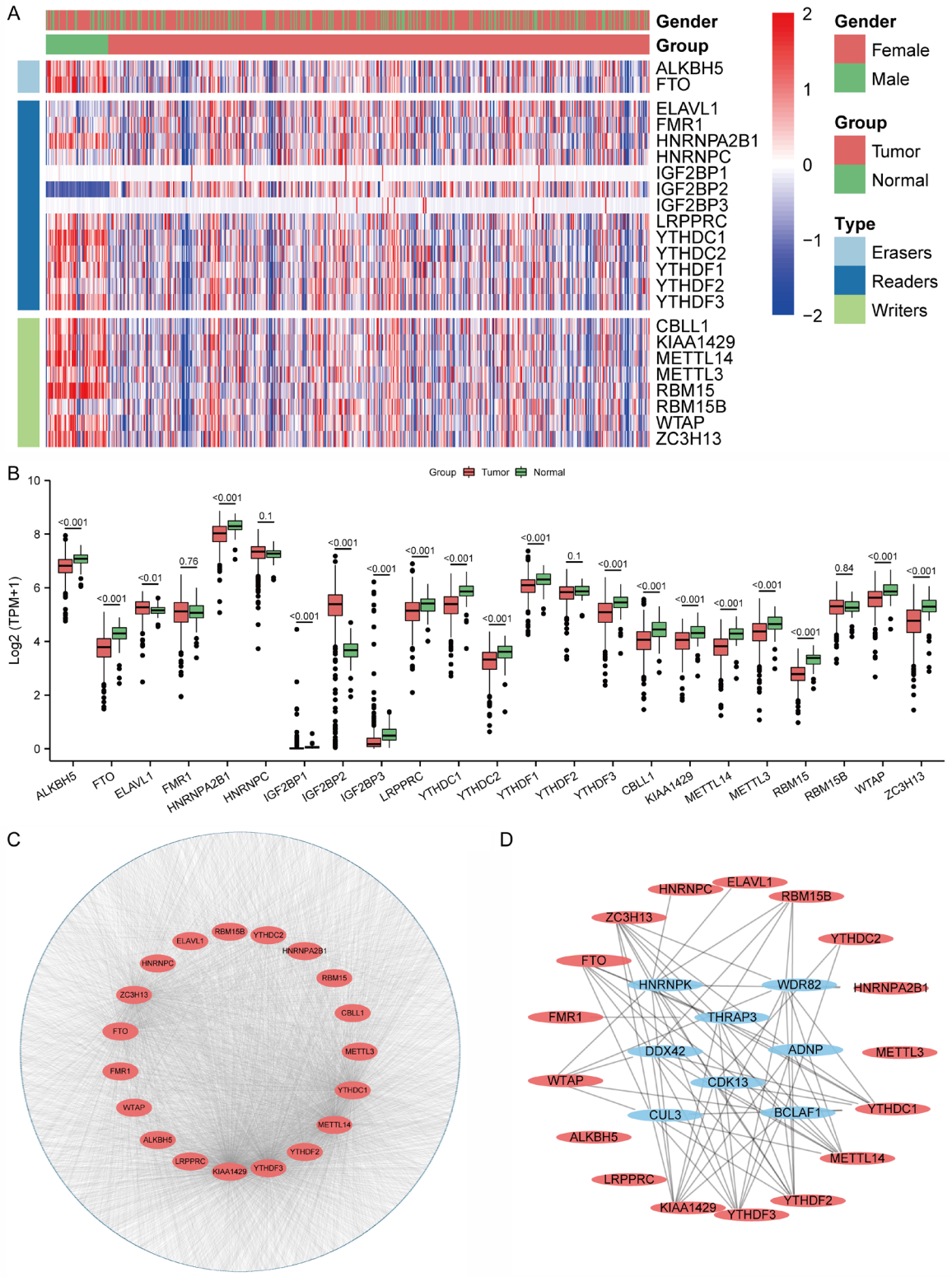


Figure 1. Expression of m6A regulatory genes between the thyroid cancer group and adjacent tissues, and the construction of the molecular interaction network between m6A gene and its related mRNAs. A. Heat map of the expression of 23 m6A regulatory genes in thyroid cancer and adjacent tissues; B. Expression of 23 m6A regulatory genes between thyroid cancer and adjacent tissues; C. Interaction network between m6A regulatory genes and their related mRNAs; D. 25 hub genes in the interaction network between m6A regulatory genes and their related mRNAs; red nodes represent m6A regulatory genes, and blue nodes represent m6A-mRNAs molecules.

Table 1. Patient characteristics of LncRNA prognostic model in the TCGA training and validation datasets

Characteristics	Train (n = 298)	Test (n = 199)	*P*-value
Age (mean (sd))	45.84±15.46	49.54±16.16	0.12
Gender (%)			0.77
Female	219 (73.49)	143 (71.86)	
Male	79 (26.51)	56 (28.14)	
T_category (%)			0.80
T1	101 (33.89)	69 (34.67)	
T2	96 (32.21)	67 (33.67)	
T3	84 (28.19)	56 (28.14)	
T4	16 (5.37)	6 (3.02)	
TX	1 (0.34)	1 (0.5)	
N_category (%)			0.66
N0	127 (42.62)	93 (46.73)	
N1	141 (47.32)	87 (43.72)	
NX	30 (10.07)	19 (9.55)	
M_category (%)			0.92
M0	5 (1.68)	4 (2.01)	
M1	126 (42.28)	81 (40.7)	
MX	167 (56.04)	114 (57.29)	
UICC_stage (%)			0.35
I	177 (59.4)	103 (51.76)	
II	31 (10.4)	21 (10.55)	
III	61 (20.47)	50 (25.13)	
IV	29 (9.73)	25 (12.56)	
Papillary Carcinoma Type (%)			0.12
Classical/usual	5 (1.68)	4 (2.01)	
Follicular (≥ 99 follicular patterned)	219 (73.49)	133 (66.83)	
Tall Cell (≥ 50 tall cell features)	59 (19.8)	41 (20.6)	
Other, specify	15 (5.03)	21 (10.55)	

further compared the expression levels of these 8 genes in 33 tumor tissues and their corresponding normal tissues and showed that the expression of eight genes was different between the tumor and normal tissues (Figure S1B-I).

Construction and validation of prognostic model based on m6A-related lncRNAs

The complete prognostic data from 497 patients with THCA were divided into training (n = 298) and validation sets (n = 199) (**Table 1**) at a 6:4 ratio. Using single-factor Cox regression and lasso Cox regression analysis methods, a prognostic model of 20 lncRNA molecules was constructed on the training set (**Figure 2A**, **2B**). For each patient in the training set, validation set, and the entire dataset, the risk score was calculated using equation 2: AC000123.4* 0.48053 + LINC00271* - 0.39758 + LINC01012*0.23698 + ARHGAP5-AS1* - 0.44110 + CTD-2541M15.1*0.14856 + CTD-3157E16.2* - 0.63425 + PRKAR2A - AS1* 0.60911 + PSMG3 - AS1*0.28977 + RP1 - 12G14.7* - 0.07105 + RP1 - 153G14.4* - 0.48361 + RP11 - 196G18.22*0.41636 + RP11 - 216B9.6* - 0.88477 + RP11 - 240G22.5*2.63020 + RP11 - 313D6.3* 0.14101 + RP11 - 337C18.9*0.10700 + RP11 - 359B12.2*1.57736 + RP11-680F8.3* - 0.51279 + RP4 - 568C11.4* - 0.57081 + SSSCA1 - AS1*0.81404 + XXbac - B444P24.13* - 1.26468 [2].

Using the survival data of the training set and the surv_cutpoint function of the survminer package, -5.935 was set as the cut-off value to distinguish between the high-risk group (n = 132) and the low-risk group (n = 166). This cut-

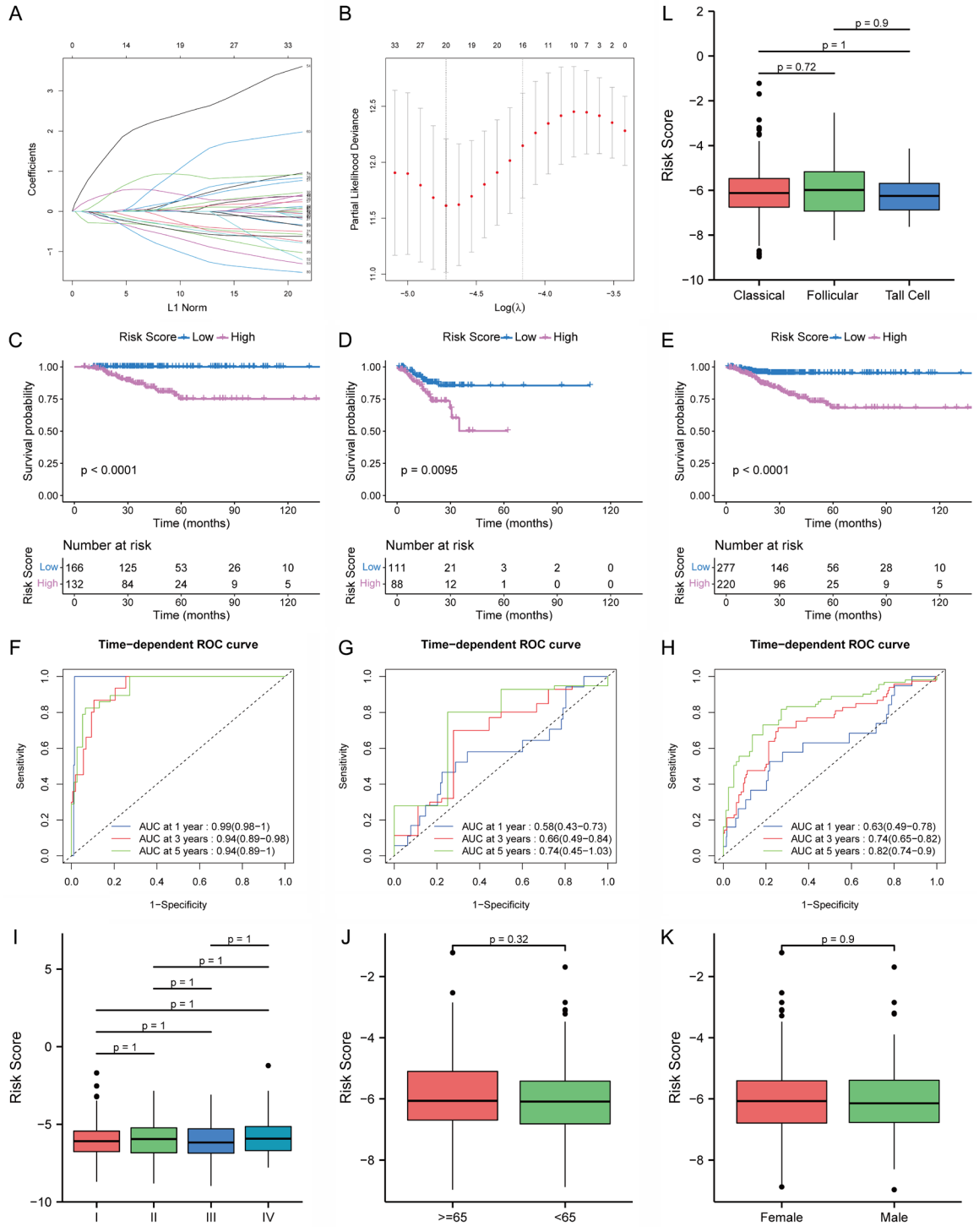


Figure 2. Construction and verification of m6A-lncRNA prognostic model. A. Lasso regression coefficients of 81 genes in the THCA training set; B. In lasso regression, based on the minimum lambda value, the most suitable gene set for constructing the model is screened; C-E. Kaplan-Meier survival analysis of the m6A-lncRNA prognostic model in the training set, validation set, and the entire dataset; F-H. ROC curve of m6A-lncRNA prognostic model in training set, validation set, and entire dataset; I-L. Correlation analysis between m6A-lncRNA prognostic model and tumor grade, age, sex and pathological categories.

off value was also used as the distinguishing value between the high and low risk groups of the validation set and the entire dataset. According to the Kaplan-Meier analysis (**Figure**

2C-E), the prognosis of the low-risk group was significantly better than that of the high-risk group ($P < 0.05$). To further evaluate the prognosis accuracy of the lncRNA model, the area under the ROC curve (AUC value) in the training, validation, and the entire dataset was calculated for 1-, 3- and 5-year progression-free survival predictions, which resulted in 0.99, 0.58, 0.63; 0.94, 0.66, 0.74; 0.94, 0.74, 0.82, respectively (**Figure 2F-H**). Therefore, the lncRNA model accurately predicted the prognosis of THCA patients. The correlations between risk score and tumor grade, age, gender, and papillary tumor subtypes of thyroid cancer were displayed as box plots to evaluate the relationship between lncRNA models and clinical indicators. The results showed no significant correlation between the risk score and these indicators ($P > 0.05$) (**Figure 2I-L**).

Correlation analysis of molecules in lncRNA model and enrichment analysis of m6A-related mRNAs

A significant positive correlation among 20 lncRNAs was observed in this study (**Figure 3A**). Potential signal pathways or functions related to m6A-mRNAs were enriched with GO and KEGG through the clusterProfiler package. The m6A-related mRNA molecules were mainly enriched in RNA splicing and covalent chromatin modification in biological processes. In terms of cell components, they were mainly enriched in the nuclear speck and chromatin; in terms of molecular functions, they were enriched in ubiquitin-like protein transferase activity and ubiquitin-protein transferase activity; in terms of molecular pathways, they were enriched in spliceosome and ubiquitin mediated proteolysis (**Figure 3B**, **3C**; **Table 3**). The GSVA algorithm was used to perform functional annotations on m6A-related mRNAs, which showed significantly different enrichment pathways between the groups. Tumor samples were enriched in the olfactory signaling pathway and NOS1 pathway, whereas normal samples were enriched in sialic acid metabolism and response to metal ion pathways (**Figure 3D**).

Interaction network between m6A-related lncRNAs and mRNAs

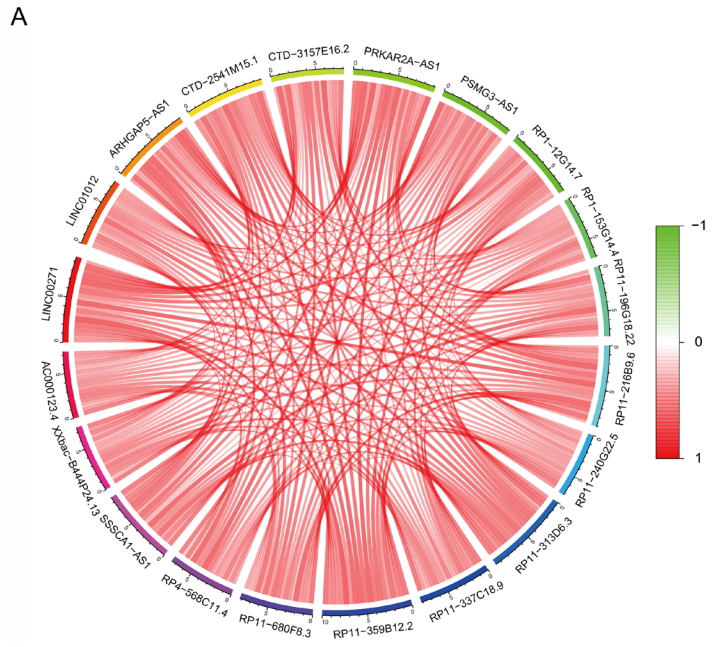
The Pearson's correlation coefficient was used to evaluate the correlation between the expression of lncRNAs and mRNAs. The absolute value of the correlation coefficient > 0.85 and *P* value < 0.05 were considered the screening criteria for the correlation between molecules. A total of 595 pairs of lncRNA-mRNA interactions were obtained, including 37 lncRNAs and 440 mRNAs. To further explore the function of lncRNAs in this network, we conducted GO enrichment analysis on related mRNAs and showed that the biological process is mainly enriched in ubiqutin-like protein transferase activity and histone binding; in cell components, it is enriched in nuclear speck and spindle; in molecular function, it is enriched in RNA splicing (**Figure 4**; **Table 4**).

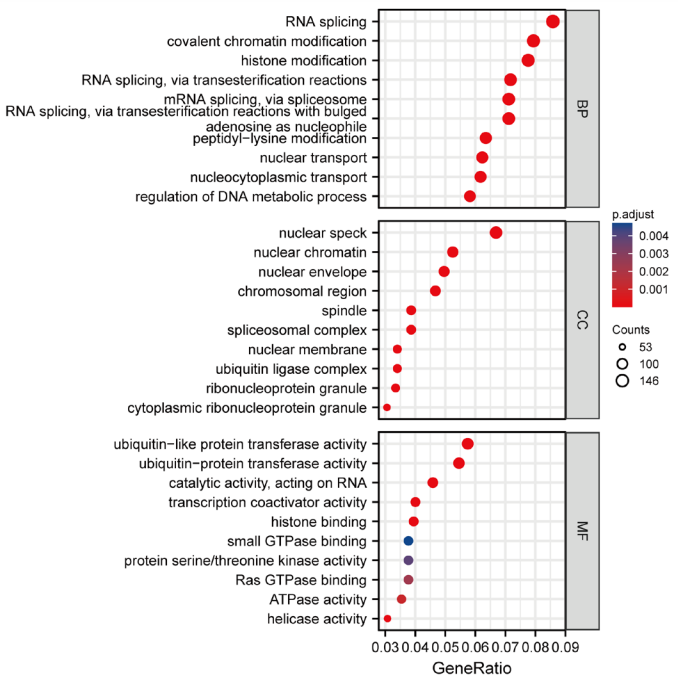
Construction and validation of prognostic model based on m6A-related mRNAs

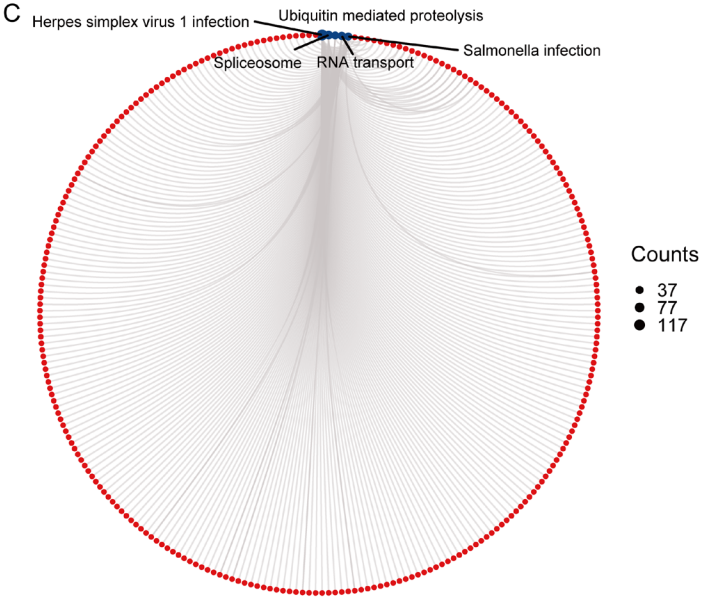
The complete prognostic data from 497 patients with THCA were divided to training set (n = 298) and validation set (n = 199) (**Table 2**) at a 6:4 ratio. Using single factor Cox regression and lasso Cox regression analyses, a prognostic model of 12 mRNA molecules was constructed on the training set (**Figure 5A**, **5B**). For each patient in the training set, validation set, and entire dataset, the risk score was calculated using the following risk formula:

BRD8*1.77078 + CCDC82* - 0.65499 + EPG5*0.89343 + KHNYN*0.19688 + LCMT2* - 0.15309 + MFAP1* - 0.00118 + SNIP1* - 0.78969 + SRRM2*0.17337 + STX16*0.00351 + TAF3* - 0.41327 + ZBTB49* - 0.94961 + ZFAND4* - 0.44280.

Using the survival data of the training set and the surv_cutpoint function of the survminer package, 2.7461 was used as the cut-off value to distinguish between the high-risk (n = 172) and low-risk groups (n = 126). This cut-off value was also used to distinguish between the high and low risk groups of the validation set and the entire dataset (**Figure 5D-F**). The results of the Kaplan-Meier curve analysis showed that the prognosis of the low-risk group was significantly better than that of the high-risk group ($P < 0.05$). We also calculated the AUC value of the training, validation, and the entire dataset for the 1-, 3- and 5-year progression-free survival prediction, which showed 0.84, 0.69, 0.74; 0.77, 0.69, 0.72; 0.78, 0.83, 0.75, respectively (**Figure 5G-L**). The mRNA model accurately predicted the prognosis of patients with THCA. The correlation between the mRNA prognostic model and clinical characteristics was evaluated using a box plot to show the difference in the


A
CTD-3157E16.2
PRKAR2A-AS1
PSMG3-AS1
RP1-12G14.7
RP1-153G14.4
RP11-196G18.22
RP11-216B9.6
RP11-240G22.5
RP11-313D6.3
RP11-337C18.9
RP11-359B12.2
RP11-680F8.3
RP4-568C11.4
SSSCA1-AS1
XX-bac-BMAP24.13
AC000123.4
LINC00271
LINC01012
ARHGAP5-AS1
CTD-2541M15.1
1
0
-1


B
RNA splicing
covalent chromatin modification
histone modification
RNA splicing, via transesterification reactions
mRNA splicing, via spliceosome
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
peptidyl−lysine modification
nuclear transport
nucleocytoplasmic transport
regulation of DNA metabolic process
BP
nuclear speck
nuclear chromatin
nuclear envelope
chromosomal region
spindle
spliceosomal complex
nuclear membrane
ubiquitin ligase complex
ribonucleoprotein granule
cytoplasmic ribonucleoprotein granule
CC
ubiquitin−like protein transferase activity
ubiquitin−protein transferase activity
catalytic activity, acting on RNA
transcription coactivator activity
histone binding
small GTPase binding
protein serine/threonine kinase activity
Ras GTPase binding
ATPase activity
helicase activity
MF
0.03 0.04 0.05 0.06 0.07 0.08 0.09
GeneRatio
p.adjust
0.004
0.003
0.002
0.001
Counts
53
100
146


C
Herpes simplex virus 1 infection
Ubiquitin mediated proteolysis
Salmonella infection
Spliceosome
RNA transport
Counts
37
77
117

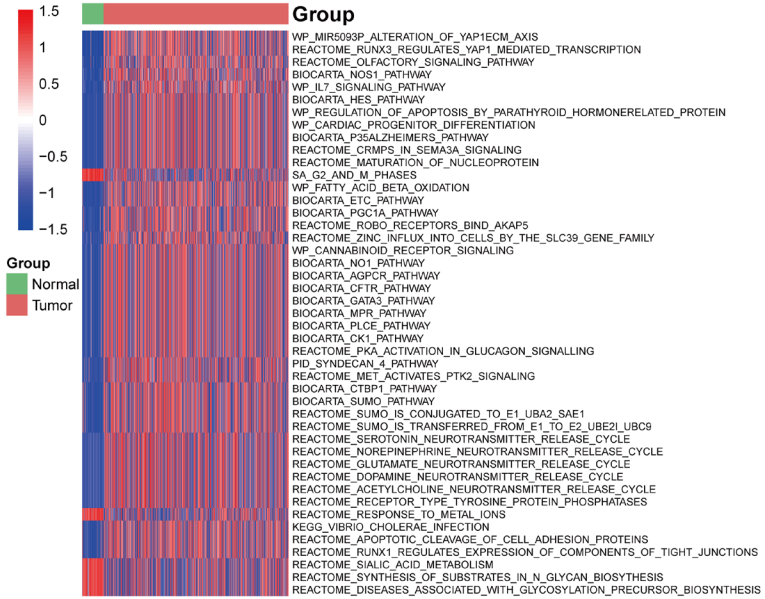

D
1.5
1
0.5
0
−0.5
−1
−1.5
Group
Normal
Tumor
Group
WP_MIR5093P_ALTERATION_OF_YAP1ECM_AXIS
REACTOME_RUNX3_REGULATES_YAP1_MEDIATED_TRANSCRIPTION
REACTOME_OLFACTORY_SIGNALING_PATHWAY
BIOCARTA_NOS1_PATHWAY
WP_IL7_SIGNALING_PATHWAY
BIOCARTA_HES_PATHWAY
WP_REGULATION_OF_APOPTOSIS_BY_PARATHYROID_HORMONERELATED_PROTEIN
WP_CARDIAC_PROGENITOR_DIFFERENTIATION
BIOCARTA_P35ALZHEIMERS_PATHWAY
REACTOME_CRMPS_IN_SEMA3A_SIGNALING
REACTOME_MATURATION_OF_NUCLEOPROTEIN
SA_G2_AND_M_PHASES
WP_FATTY_ACID_BETA_OXIDATION
BIOCARTA_ETC_PATHWAY
BIOCARTA_PGC1A_PATHWAY
REACTOME_ROBO_RECEPTORS_BIND_AKAP5
REACTOME_ZINC_INFLUX_INTO_CELLS_BY_THE_SLC39_GENE_FAMILY
WP_CANNABINOID_RECEPTOR_SIGNALING
BIOCARTA_NO1_PATHWAY
BIOCARTA_AGPCR_PATHWAY
BIOCARTA_CFTR_PATHWAY
BIOCARTA_GATA3_PATHWAY
BIOCARTA_MPR_PATHWAY
BIOCARTA_PLCE_PATHWAY
BIOCARTA_CK1_PATHWAY
REACTOME_PKA_ACTIVATION_IN_GLUCAGON_SIGNALLING
PID_SYNDECAN_4_PATHWAY
REACTOME_MET_ACTIVATES_PTK2_SIGNALING
BIOCARTA_CTBP1_PATHWAY
BIOCARTA_SUMO_PATHWAY
REACTOME_SUMO_IS_CONJUGATED_TO_E1_UBA2_SAE1
REACTOME_SUMO_IS_TRANSFERRED_FROM_E1_TO_E2_UBE2I_UBC9
REACTOME_SEROTONIN_NEUROTRANSMITTER_RELEASE_CYCLE
REACTOME_NOREPINEPHRINE_NEUROTRANSMITTER_RELEASE_CYCLE
REACTOME_GLUTAMATE_NEUROTRANSMITTER_RELEASE_CYCLE
REACTOME_DOPAMINE_NEUROTRANSMITTER_RELEASE_CYCLE
REACTOME_ACETYLCHOLINE_NEUROTRANSMITTER_RELEASE_CYCLE
REACTOME_RECEPTOR_TYPE_TYROSINE_PROTEIN_PHOSPHATASES
REACTOME_RESPONSE_TO_METAL_IONS
KEGG_VIBRIO_CHOLERAE_INFECTION
REACTOME_APOPTOTIC_CLEAVAGE_OF_CELL_ADHESION_PROTEINS
REACTOME_RUNX1_REGULATES_EXPRESSION_OF_COMPONENTS_OF_TIGHT_JUNCTIONS
REACTOME_SIALIC_ACID_METABOLISM
REACTOME_SYNTHESIS_OF_SUBSTRATES_IN_N_GLYCAN_BIOSYTHESIS
REACTOME_DISEASES_ASSOCIATED_WITH_GLYCOSYLATION_PRECURSOR_BIOSYNTHESIS

Figure 3. Functional enrichment analysis of m6A-related mRNAs. A. Correlation between m6A-lncRNA prognostic model molecules; B. GO enrichment analysis results of m6A-related mRNAs (m6A-mRNAs); BP, Biological Process; CC, Cellular Components; MF, Molecular Function; C. KEGG enrichment analysis of m6A-related mRNAs, the red dots represent related genes in the KEGG enrichment pathway; D. GSVA analysis of m6A-related mRNAs.

Table 2. Patient characteristics of mRNA prognostic model in the TCGA training and validation datasets

Characteristics	Train (n = 298)	Test (n = 199)	*P*-value
Age (mean (sd))	48.08±16.1	46.19±15.38	0.30
Gender (%)			1.00
Female	81 (27.18)	54 (27.14)	
Male	217 (72.82)	145 (72.86)	
T_category (%)			0.19
T1	108 (36.24)	62 (31.16)	
T2	88 (29.53)	75 (37.69)	
T3	84 (28.19)	56 (28.14)	
T4	17 (5.7)	5 (2.51)	
TX	1 (0.34)	1 (0.5)	
N_category (%)			0.65
N0	32 (10.74)	17 (8.54)	
N1	133 (44.63)	95 (47.74)	
NX	133 (44.63)	87 (43.72)	
M_category (%)			0.26
M0	161 (54.03)	120 (60.3)	
M1	130 (43.62)	77 (38.69)	
MX	7 (2.35)	2 (1.01)	
UICC_stage (%)			0.45
I	162 (54.36)	118 (59.3)	
II	71 (23.83)	40 (20.1)	
III	29 (9.73)	23 (11.56)	
IV	36 (12.08)	18 (9.05)	
Papillary Carcinoma Type (%)			0.10
Classical/usual	6 (2.01)	3 (1.51)	
Follicular (≥ 99 follicular patterned)	221 (74.16)	131 (65.83)	
Tall Cell (≥ 50 tall cell features)	49 (16.44)	51 (25.63)	
Other, specify	22 (7.38)	14 (7.04)	

risk scores between different tumor grades, age, and gender. The risk score of the high-grade tumors was significantly higher than that of low-grade tumors, although no significant correlation between the risk score and age and gender was observed (P > 0.05) (**Figure 5J-L**).

Biological characteristics of prognostic models

The correlation between the prognostic model of m6A-lncRNA and m6A-mRNA and the tumor microenvironment and clinicopathological characteristics was explored. The entire TCGA dataset was compared using the risk model scores; the distinguishing value of high and low risk scores was the same as that reported in section 3.6. **Figure 6A-D** shows the overall landscape of gene mutations in the high and low mRNA (lncRNA) risk scores; the rate of BRAF mutations in the high-risk group was significantly higher (lower) than that in the low-risk group.

Immune infiltration plays an important role in tumor formation and progression, and successful clinical applications of immunosuppressants also highlight the role of intratumor immune infiltration in tumor treatment. There were significant differences in CD8+ T cells, regulatory T cells (Tregs), γδ-T cells, M0 macrophages, and activated dendritic cells in the high- and low-risk groups of the mRNA model (P < 0.05) (**Figure 6E**). Furthermore, the high- and low-risk populations of the lncRNA model showed significant differences in Naïve CD4+ T cells, follicular T helper cells, γδ-T cells, monocytes, activated mast cells, and eosinophils (P < 0.05) (**Figure 6F**). The tumor microenvironment score can also reflect the degree of infiltration of immune cells and interstitial cells in tumor tissues. The results showed that the mesenchymal score of the high-mRNA risk group was significantly lower than that of the low-risk group, whereas no difference in the immune score between the two groups was observed (**Figure 6G**). This shows that the intratumoral stroma content of the tumor tissues of the high-mRNAs risk score group is significantly lower than that of the low-risk group. However, no significant correlation between the lncRNA risk score and the intersti-

Table 3. GO and KEGG enrichment analysis of m6A-related mRNAs

GO Category	Term	Description	Count	Adjust *P*
BP	GO:0008380	RNA splicing	146	7.46E-39
BP	GO:0016569	covalent chromatin modification	135	8.01E-32
BP	GO:0016570	histone modification	132	5.21E-32
BP	GO:0000375	RNA splicing, via transesterification reactions	122	1.81E-33
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	121	1.86E-33
BP	GO:0000398	mRNA splicing, via spliceosome	121	1.86E-33
BP	GO:0018205	peptidyl-lysine modification	108	1.36E-23
BP	GO:0051169	nuclear transport	106	1.56E-27
BP	GO:0006913	nucleocytoplasmic transport	105	2.47E-27
BP	GO:0051052	regulation of DNA metabolic process	99	3.42E-16
MF	GO:0019787	ubiquitin-like protein transferase activity	99	1.48E-15
MF	GO:0004842	ubiquitin-protein transferase activity	94	3.17E-15
MF	GO:0140098	catalytic activity, acting on RNA	79	8.27E-09
MF	GO:0003713	transcription coactivator activity	69	9.39E-09
MF	GO:0042393	histone binding	68	1.18E-18
MF	GO:0017016	Ras GTPase binding	65	0.002296
MF	GO:0004674	protein serine/threonine kinase activity	65	0.003889
MF	GO:0031267	small GTPase binding	65	0.004715
MF	GO:0016887	ATPase activity	61	0.001116
MF	GO:0004386	helicase activity	53	1.27E-13
CC	GO:0016607	nuclear speck	116	1.24E-29
CC	GO:0000790	nuclear chromatin	91	4.78E-17
CC	GO:0005635	nuclear envelope	86	8.49E-10
CC	GO:0098687	chromosomal region	81	2.72E-14
CC	GO:0005681	spliceosomal complex	67	1.46E-22
CC	GO:0005819	spindle	67	1.77E-08
CC	GO:0000151	ubiquitin ligase complex	59	7.90E-09
CC	GO:0031965	nuclear membrane	59	4.92E-08
CC	GO:0035770	ribonucleoprotein granule	58	1.79E-12
CC	GO:0036464	cytoplasmic ribonucleoprotein granule	53	9.51E-11
KEGG	hsa05168	Herpes simplex virus 1 infection	117	5.36E-22
KEGG	hsa03040	Spliceosome	50	3.51E-15
KEGG	hsa03013	RNA transport	43	1.37E-07
KEGG	hsa05132	Salmonella infection	38	0.016608
KEGG	hsa04120	Ubiquitin mediated proteolysis	37	4.37E-08

tial score and immune score of the tumor microenvironment was observed (P > 0.05) (**Figure 6J**). Both the mutational burden and instability of microsatellites in tumors were closely associated with the infiltration of immune cells in tumor tissues. Several studies have shown that the higher the mutation load in tumor tissue and the more unstable the microsatellite instability, the stronger the immunoreactivity in tumor tissue, the higher the probability of successful application of immunosuppressive agents. Patients with high mRNA risk scores have significantly higher tumor mutation burdens (**Figure 6H**). However, there was no significant difference in tumor mutation burden between the lncRNA high and low risk groups (**Figure 6K**). The mRNA and lncRNA risk scores were not correlated with microsatellite instability (**Figure 6I**, **6L**).

Identification of independent prognostic factors

In the TCGA training set, single-factor and multifactor Cox regression analyses were used to

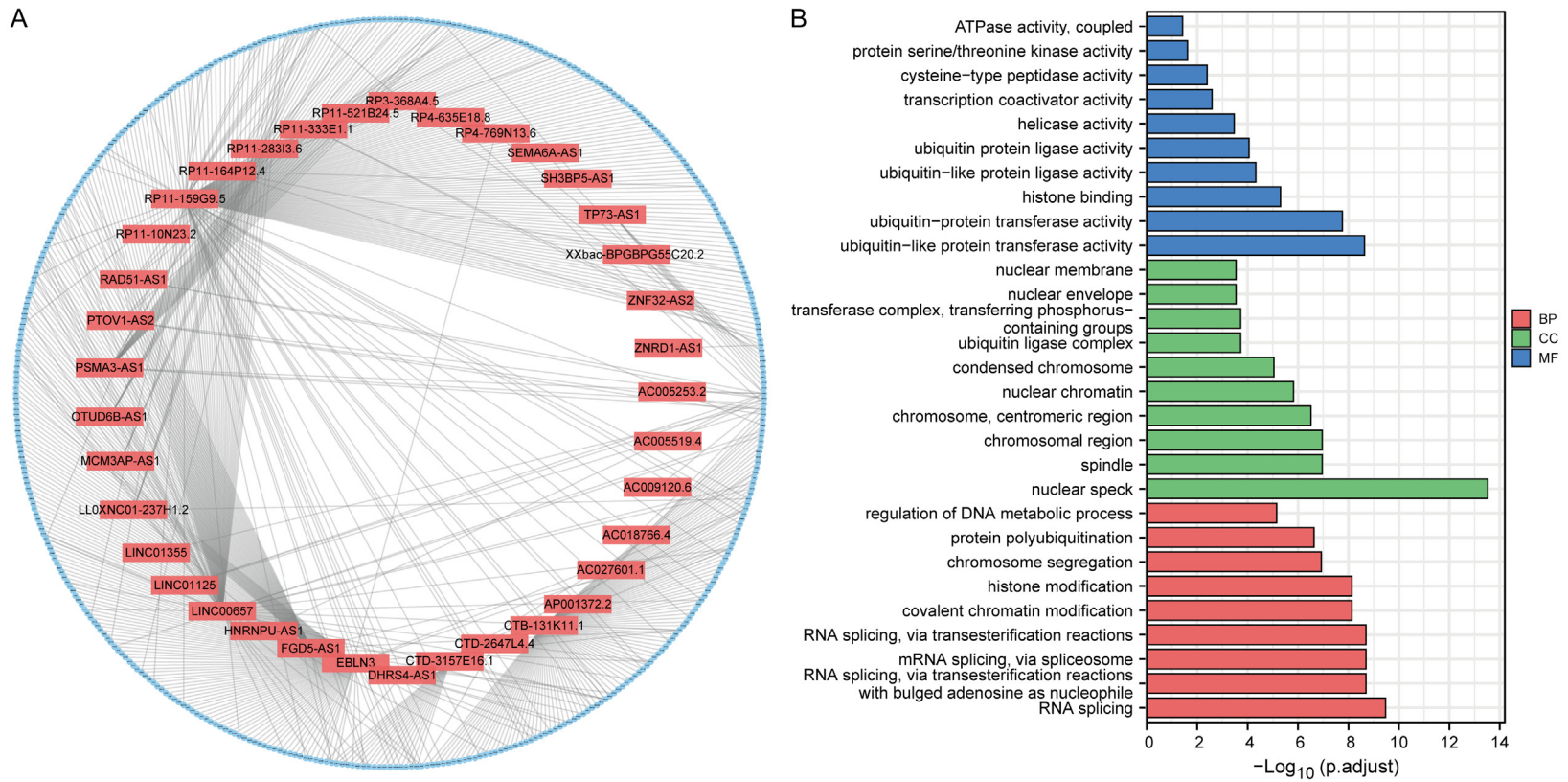


Figure 4. Interaction network between m6A-related lncRNA and m6A-related mRNA. A. Interaction network between m6A-related lncRNAs and m6A-related mRNAs; B. GO enrichment results of mRNAs in the interaction network; BP, Biological Process; CC, Cellular Components; MF, Molecular Function.

Table 4. GO enrichment analysis of mRNAs within mRNA-lncRNA interaction network

GO Category	Term	Description	Count	Adjust *P*
BP	GO:0008380	RNA splicing	40	3.37E-10
BP	GO:0016569	covalent chromatin modification	37	7.33E-09
BP	GO:0016570	histone modification	36	7.33E-09
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	34	2.01E-09
BP	GO:0000398	mRNA splicing, via spliceosome	34	2.01E-09
BP	GO:0000375	RNA splicing, via transesterification reactions	34	2.01E-09
BP	GO:0051052	regulation of DNA metabolic process	29	7.03E-06
BP	GO:0007059	chromosome segregation	28	1.19E-07
BP	GO:0000209	protein polyubiquitination	27	2.32E-07
MF	GO:0019787	ubiquitin-like protein transferase activity	36	2.32E-09
MF	GO:0004842	ubiquitin-protein transferase activity	33	1.75E-08
MF	GO:0004674	protein serine/threonine kinase activity	21	0.024184
MF	GO:0042393	histone binding	20	4.95E-06
MF	GO:0061659	ubiquitin-like protein ligase activity	20	4.76E-05
MF	GO:0003713	transcription coactivator activity	20	0.002593
MF	GO:0061630	ubiquitin protein ligase activity	19	8.85E-05
MF	GO:0004386	helicase activity	15	0.000341
MF	GO:0042623	ATPase activity, coupled	15	0.038239
MF	GO:0008234	cysteine-type peptidase activity	14	0.004058
CC	GO:0016607	nuclear speck	40	2.99E-14
CC	GO:0005819	spindle	28	1.10E-07
CC	GO:0098687	chromosomal region	28	1.10E-07
CC	GO:0000790	nuclear chromatin	27	1.52E-06
CC	GO:0005635	nuclear envelope	25	0.000289
CC	GO:0000775	chromosome, centromeric region	20	3.07E-07
CC	GO:0000793	condensed chromosome	19	9.01E-06
CC	GO:0000151	ubiquitin ligase complex	19	0.000189
CC	GO:0031965	nuclear membrane	19	0.000289
CC	GO:0061695	transferase complex, transferring phosphorus-containing groups	18	0.000189

determine whether the mRNA prognosis model risk score, age, gender, and tumor grade were independent prognostic factors. Both risk score and tumor grade were independent prognostic factors of the training set (**Table 5**) (risk score, HR (hazard ratio): 2.86; 95% CI (confidence interval): 1.97-4.13; P < 0.001; tumor grade, HR: 1.35; 95% CI: 0.98-1.86, P < 0.1).

Construction and validation of prognostic nomogram

The risk score and tumor grade data based on 12 mRNAs were independent prognostic factors for patients in the THCA training set. The nomogram constructed by integrating these two factors was used to predict the 1-, 3-, and 5-year progression-free survival (**Figure 7A**). The correction curve of the training set showed that the nomogram's prediction of the 1- and 5-year progression-free survival was significantly better than the 3-year progression-free survival (**Figure 7B**). The nomogram showed high prognostic accuracy in the validation set and the entire dataset (**Figure 7C**, **7D**). According to the decision curve analysis curve (Figure S2), the nomogram could provide the highest net benefit compared to the tumor grade and risk score.

Construction and validation of lncRNA combined mRNA molecular model

A total of 497 thyroid cancer patients with complete prognostic data were included in this study, and they were divided into training set (N = 298) and validation set (N = 199) according to the ratio of 6:4 (**Table 2**). A prognostic model

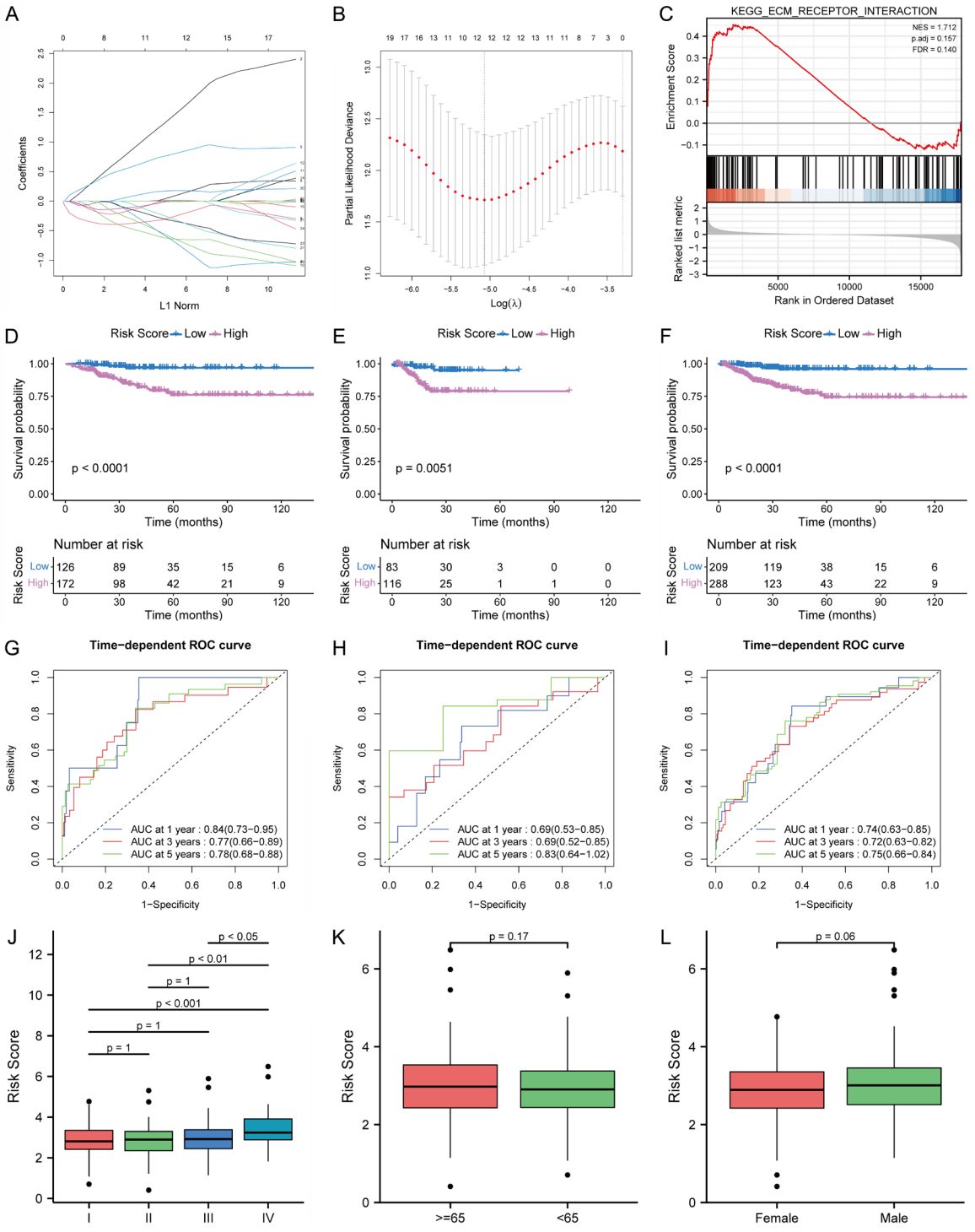


Figure 5. Construction and verification of the prognostic model of m6A-related mRNAs. A. Lasso regression coefficients of 27 genes in the THCA training set; B. In lasso regression, based on the minimum lambda value, the most suitable gene set for constructing the model was screened; C. Gene set enrichment analysis results in the high-risk group of the m6A-related mRNAs prognostic model; D-F. Kaplan-Meier survival analysis of the prognostic model of m6A-related mRNAs in the training set, validation set, and the entire dataset; G-I. The ROC curve of the prognostic model of m6A-related mRNAs in the training set, validation set, and the entire dataset; J-L. Correlation analysis between the prognostic model of m6A-related mRNAs and tumor grade, age, and sex.

containing 12 lncRNA molecules was constructed in the training set using univariate Cox regression and LASSO Cox regression analysis (**Figure 8A**, **8B**). For each patient in the train-

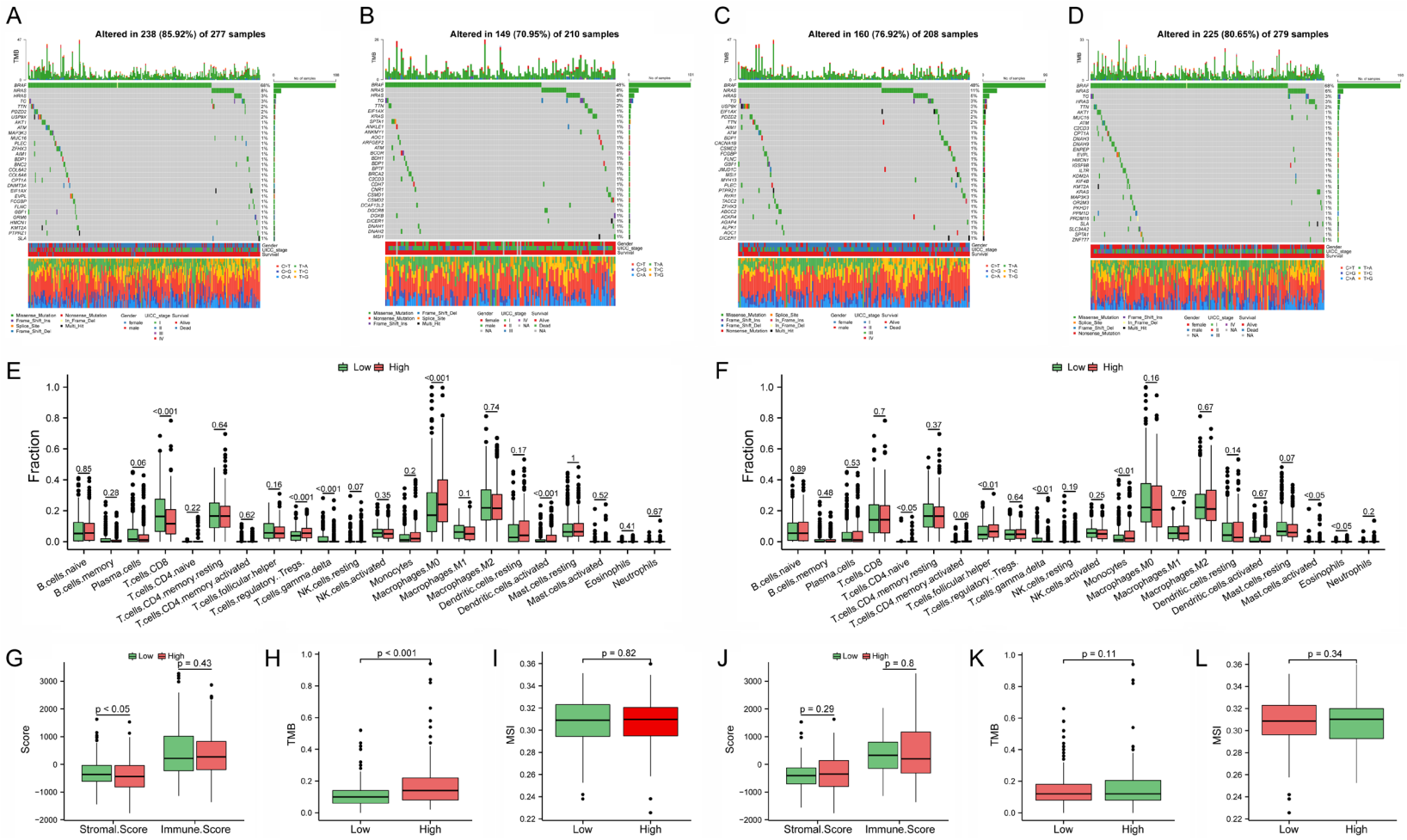


Figure 6. Relationship between m6A-mRNA and m6A-lncRNA prognostic models and gene mutation, tumor microenvironmental immune infiltration, tumor mutation burden (TMB), and microsatellite instability (MSI). A. Panoramic view of mutations in high-risk groups of mRNA risk model; B. Panoramic view of mutations in the low-risk population of the mRNA risk model; C. Panoramic view of mutations in high-risk groups of the lncRNA risk model; D. Panoramic view of mutations in the low-risk population of the lncRNA risk model; E. Correlation between the high and low scores of the mRNA risk model and the infiltration of immune cells in the tumor tissue; F. Correlation between the high and low scores of the lncRNA risk model and the infiltration of immune cells in the tumor tissue; G. Correlation between the high and low scores of the mRNA risk model and the immune score and interstitial score; H. Correlation between the high and low scores of the mRNA risk model and TMB; I. Correlation between the high and low scores of the mRNA risk model and MSI; J. Correlation between the high and low scores of the lncRNA risk model and the immune score and interstitial score; K. Correlation between the high and low scores of the lncRNA risk model and TMB; L. Correlation between the high and low scores of the lncRNA risk model and MSI.

Table 5. Univariate and multivariate Cox regression to identify independent prognostic predictor in the TCGA training cohort

Characteristics	Progression-free survival in TCGA Training Cohort				
	Number of patients	Univariate Analysis		Multivariate Analysis	
		HR (95% CI)	*P*-Value	HR (95% CI)	*P*-Value
Age (≥ 65 vs < 65)	298	1.61 (0.69-3.75)	0.266	--	--
Gender (Male vs Female)	298	1.41 (0.66-2.99)	0.374	--	--
Risk.Score	298	3.34 (2.37-4.71)	< 0.001	2.86 (1.97-4.13)	< 0.001
UICC stage	298	1.76 (1.3-2.39)	< 0.001	1.35 (0.98-1.86)	0.062

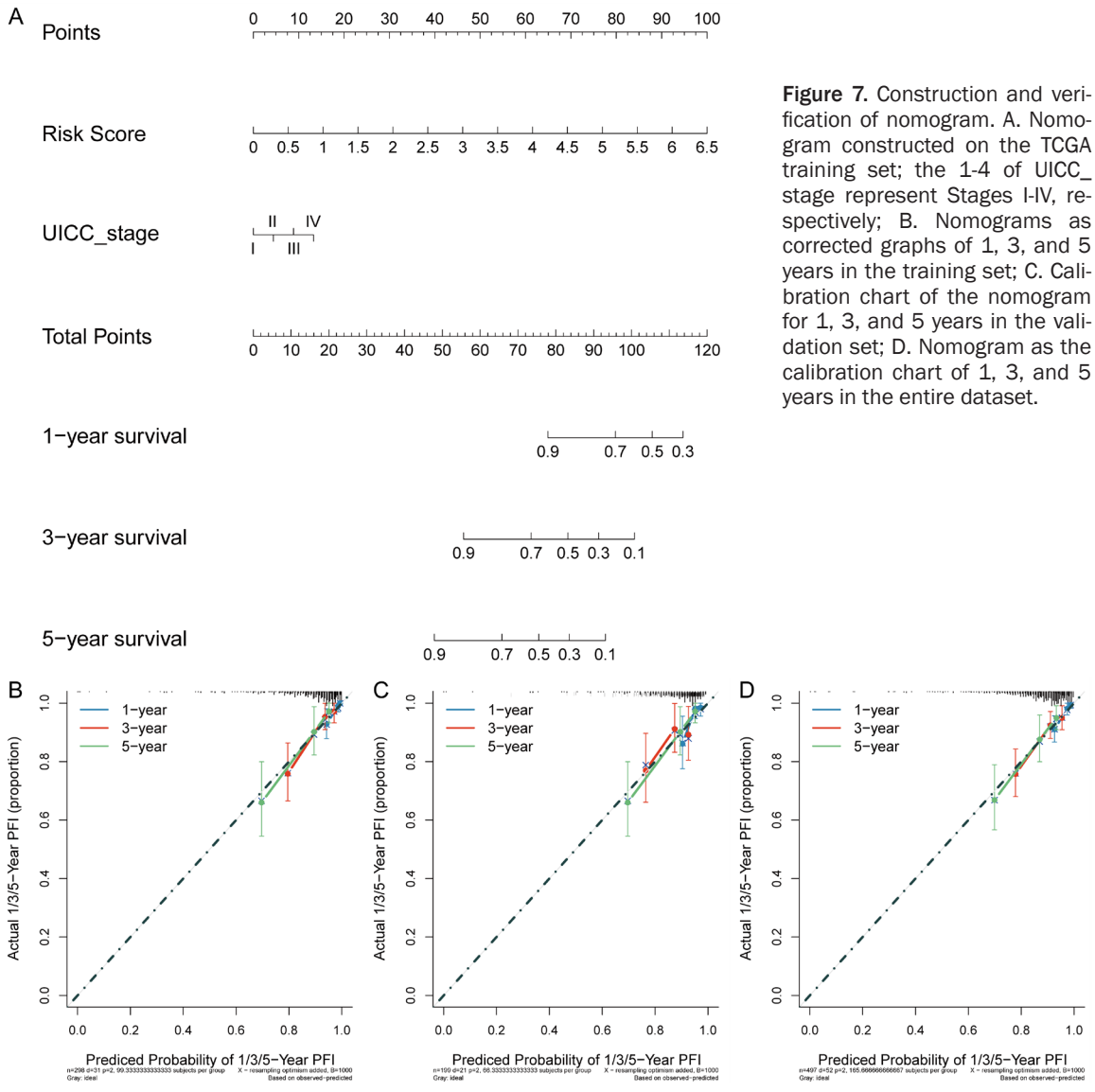


Figure 7. Construction and verification of nomogram. A. Nomogram constructed on the TCGA training set; the 1-4 of UICC_ stage represent Stages I-IV, respectively; B. Nomograms as corrected graphs of 1, 3, and 5 years in the training set; C. Calibration chart of the nomogram for 1, 3, and 5 years in the validation set; D. Nomogram as the calibration chart of 1, 3, and 5 years in the entire dataset.

ing set, validation set and the entire dataset, the risk score is calculated based on the following risk formula: LENG8-AS1* 0.45213 - LINC00271*0.42282 - LINC01257* 0.59266 + RP1-283E3.8*0.36323 - RP11-141M1.4*0.05587 - RP11-175019.4*0.16684 + RP11-196G18.22*0.46581 - RP11-216-B9.6*0.79439 - RP11-251G23.5*0.50873 + RP11-33B1.4*0.129 - 777023.2*0.02021 - RP5-991G20.1*0.00003. Using the survival data from the training set and the surv_cutpoint function of the "survminer" package, we

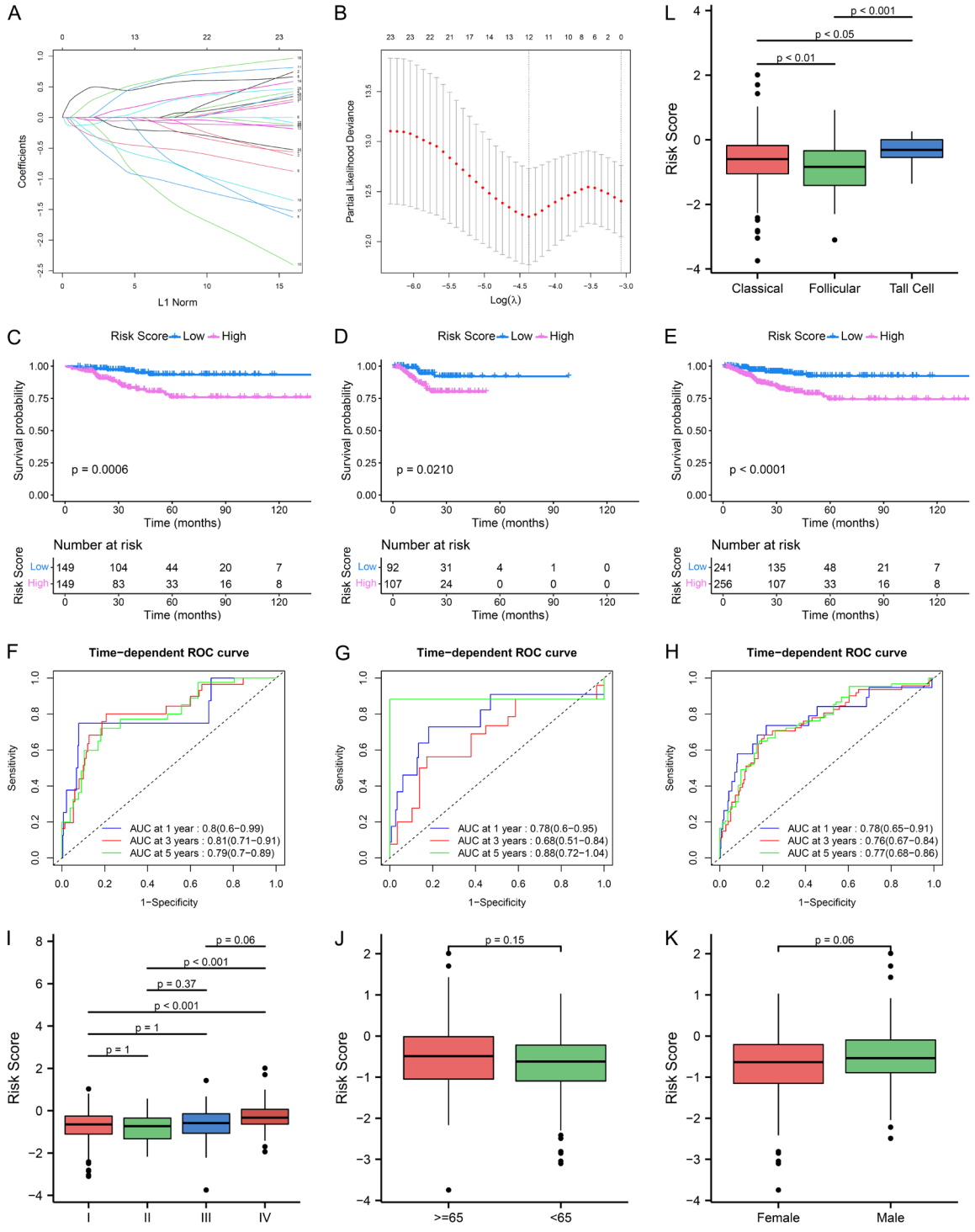


Figure 8. Construction and validation of lncRNA combined mRNA molecular model. A. Lasso regression coefficients of 26 genes in the THCA training set; B. Screen out the optimal gene set for building the model based on the minimum lambda value in LASSO regression; C-E. Kalan-Meier survival analysis of m6A-lncRNA combined mRNA prognostic model in training set, validation set and the whole dataset; F-H. Time-ROC curves of m6A-lncRNA combined mRNA prognosis model in training set, validation set and the whole dataset; I-L. Correlation analysis of m6A-lncRNA combined with mRNA prognosis model and tumor grade, age, gender and pathological grade.

define -0.6314 as the cutoff value to distinguish between high risk (N = 149) and low risk (N = 149) groups, which will also serve as validation Discriminatory value between high and

low risk groups for the set and the entire dataset. According to the Kaplan-Meier curve analysis results in **Figure 8C-E**, the prognosis of the low-risk group was significantly better than that of the high-risk group ($P < 0.05$). In addition, to further evaluate the accuracy of the lncRNA combined mRNA model for prognosis prediction, we calculated the area under the ROC curve (AUC). The AUCs for 1-year, 3-year and 5-year progression-free survival were 0.80 vs 0.78 vs 0.78; 0.81 vs 0.68 vs 0.76; 0.79 vs 0.88 vs 0.77 in the training set, validation set, and the entire dataset, respectively. It can be seen that the lncRNA combined mRNA model has very good accuracy in predicting the prognosis of thyroid cancer patients (**Figure 8F-H**). We further evaluated the correlation between the lncRNA combined mRNA model and clinicopathological factors. In this study, the differences in risk scores among different tumor grades, different ages, different genders, and different pathological subtypes were shown in box plots. The results showed that the risk scores of high-grade tumors were significantly higher than those of low-grade tumors, and different histopathological subtypes There was also a significant difference in the risk score between the two groups, but there was no significant correlation between the risk score and age and gender ($P > 0.05$) (**Figure 8I-L**).

Discussion

The incident of thyroid cancer has a rapidly increased, thereby posing significant clinical, economic, and psychological burdens [1, 2]. Overall, the prognosis of THCA is relatively good [5]; however, some patients whose tumors are highly aggressive and more likely to recur [6] often undergo a second operation, which may lead to post-surgical complications such as recurrent laryngeal nerve and parathyroid injuries [36-38]. In addition, postoperative patients require thyroid stimulating hormone suppression therapy, which can lead to potential side effects such as osteoporosis and atrial fibrillation [39, 40]. Therefore, it is essential to better identify and evaluate the prognosis of patients with THCA.

Traditional clinicopathological parameters such as tumor-node-metastasis staging can predict the mortality associated with THCA, yet it is difficult to accurately estimate the risk of recurrence [41-43]. The American Thyroid Association recurrence risk stratification can predict the recurrence risk of thyroid cancer, although its accuracy, which does not reflect the biological progress of papillary thyroid carcinoma, requires further improvement [44, 45]. Accurate prediction of the prognosis of patients with THCA will help to better identify patients with high risk of recurrence. In addition, extensive surgery, I^{131} treatment, and thyroid stimulating hormone suppression therapy for low-risk patients will be avoided, which will improve the quality of life of patients.

Previous studies have reported that the expression of m6A regulators and lncRNA varies among different types of tumors, thereby playing a key regulatory role [7, 10, 46] and affecting the prognosis of cancer patients. In this study, most m6A RNA methylation regulatory genes underwent expression changes in THCA; therefore, we analyzed the correlation between m6A-related lncRNA and mRNA and the clinicopathological characteristics of THCA, hence constructed a risk model for m6A-related lncRNA and m6A-related mRNA molecules. In addition, these prognostic genes may play an important role in the progression of papillary thyroid carcinoma and represent potential targets for suppressing recurrence and metastasis.

In this study, complete prognostic data from 497 patients with THCA were analyzed using single-factor Cox regression and lasso Cox regression analyses to construct a prognostic model of 20 lncRNAs, among which LINC00271 is associated with thyroid cancer [47-50]. Buishand et al. [48] reported that adrenocortical tumors have a distinct, long, non-coding RNA expression profile and LINC00271 is downregulated in malignancy. Ma et al. [50] showed that LINC00271 expression was significantly downregulated in papillary thyroid carcinoma versus adjacent normal tissues ($P < 0.001$). LINC00271 was also found to be an independent risk factor for extrathyroidal extension, lymph node metastasis, advanced tumor stage III/IV, and recurrence in multivariate analyses [50]. The lncRNA ARHGAP5-AS1 in the prediction model constructed in this study is associated with breast and gastric cancers [51, 52].

The role of the other lncRNAs in tumor development has not been reported. Compared to the

previous prognostic models of THCA [53, 54], the prognostic model developed in this study is novel and has potential clinical applications. Our findings will help to identify potential prognostic lncRNAs and mRNAs regulated by m6A, providing new insights into evaluating the risk of THCA patients.

The nomogram of this study combined with m6A-related lncRNA and mRNA prediction models and clinicopathological parameters can provide clinicians with a convenient and accurate method to assess the prognosis of patients with THCA. The graphical scoring system is e patient friendly and helps with treatment planning to achieve personalized.

This study has some limitations. The main source of clinical information for our dataset is the TCGA database where most patients were from either Europe or the United States; thus, caution should be exercised when extending the results to the patients from other regions. The establishment and verification of the nomogram are also based on the TCGA database; therefore, in future studies, external datasets should be used to verify clinical information and gene expression data. Finally, our prognostic model and nomogram must be validated in prospective studies.

In conclusion, we revealed the expression and prognostic value of m6A-related lncRNAs and mRNAs in THCA. Our study established an m6A-related tumor prognosis prediction nomogram, which can reliably predict the prognosis of patients with THCA. These findings will expand our understanding of m6A modification in THCA and provide insights into the prognosis and treatment strategies of THCA.

Acknowledgements

This work was supported by grants from the Science and technology Research Special project of Sichuan provincial Administration of Traditional Chinese Medicine (Nos. 2021ZD011) and Ren-song Yue's TCM studio.

Disclosure of conflict of interest

None.

Address correspondence to: Rensong Yue, Department of Endocrinology, Hospital of Chengdu University of Traditional Chinese Medicine, No. 39-41, Shi-er-qiao Road, Jinniu District, Chengdu 610075, Sichuan Province, P. R. China. Tel: +86-028-87783481; E-mail: songrenyue@cdutcm.edu.cn

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020; 70: 7-30.
[2] Cao M, Li H, Sun D and Chen W. Cancer burden of major cancers in China: a need for sustainable actions. Cancer Commun (Lond) 2020; 40: 205-210.
[3] Mariotto AB, Yabroff KR, Shao Y, Feuer EJ and Brown ML. Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst 2011; 103: 117-128.
[4] Miller KD, Fidler-Benaoudia M, Keegan TH, Hipp HS, Jemal A and Siegel RL. Cancer statistics for adolescents and young adults, 2020. CA Cancer J Clin 2020; 70: 443-459.
[5] Cabanillas ME, McFadden DG and Durante C. Thyroid cancer. Lancet 2016; 388: 2783-2795.
[6] Carling T and Udelsman R. Thyroid cancer. Annu Rev Med 2014; 65: 125-137.
[7] Ma S, Chen C, Ji X, Liu J, Zhou Q, Wang G, Yuan W, Kan Q and Sun Z. The interplay between m6A RNA methylation and noncoding RNA in cancer. J Hematol Oncol 2019; 12: 121.
[8] Huang H, Weng H and Chen J. m(6)A modification in coding and non-coding RNAs: roles and therapeutic implications in cancer. Cancer Cell 2020; 37: 270-288.
[9] He L, Li H, Wu A, Peng Y, Shu G and Yin G. Functions of N6-methyladenosine and its role in cancer. Mol Cancer 2019; 18: 176.
[10] Sun T, Wu R and Ming L. The role of m6A RNA methylation in cancer. Biomed Pharmacother 2019; 112: 108613.
[11] Song P, Feng L, Li J, Dai D, Zhu L, Wang C, Li J, Li L, Zhou Q, Shi R, Wang X and Jin H. β-catenin represses miR455-3p to stimulate m6A modification of HSF1 mRNA and promote its translation in colorectal cancer. Mol Cancer 2020; 19: 129.
[12] Ransohoff JD, Wei Y and Khavari PA. The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol 2018; 19: 143-157.
[13] Quinn JJ and Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 2016; 17: 47-62.
[14] Tan YT, Lin JF, Li T, Li JJ, Xu RH and Ju HQ. LncRNA-mediated posttranslational modifications and reprogramming of energy metabolism in cancer. Cancer Commun (Lond) 2021; 41: 109-120.

[15] Chan JJ and Tay Y. Noncoding RNA: RNA regulatory networks in cancer. Int J Mol Sci 2018; 19: 1310.
[16] Bhan A, Soleimani M and Mandal SS. Long noncoding RNA and cancer: a new paradigm. Cancer Res 2017; 77: 3965-3981.
[17] Fang Y and Fullwood MJ. Roles, functions, and mechanisms of long non-coding RNAs in cancer. Genomics Proteomics Bioinformatics 2016; 14: 42-54.
[18] Yuan J, Song Y, Pan W, Li Y, Xu Y, Xie M, Shen Y, Zhang N, Liu J, Hua H, Wang B, An C and Yang M. LncRNA SLC26A4-AS1 suppresses the MRN complex-mediated DNA repair signaling and thyroid cancer metastasis by destabilizing DDX5. Oncogene 2020; 39: 6664-6676.
[19] Liu H, Deng H, Zhao Y, Li C and Liang Y. LncRNA XIST/miR-34a axis modulates the cell proliferation and tumor growth of thyroid cancer through MET-PI3K-AKT signaling. J Exp Clin Cancer Res 2018; 37: 279.
[20] Lei H, Gao Y and Xu X. LncRNA TUG1 influences papillary thyroid cancer cell proliferation, migration and EMT formation through targeting miR-145. Acta Biochim Biophys Sin (Shanghai) 2017; 49: 588-597.
[21] Coker H, Wei G and Brockdorff N. m6A modification of non-coding RNA and the control of mammalian gene expression. Biochim Biophys Acta Gene Regul Mech 2019; 1862: 310-318.
[22] Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE and Jaffrey SR. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat Methods 2015; 12: 767-772.
[23] Yang X, Zhang S, He C, Xue P, Zhang L, He Z, Zang L, Feng B, Sun J and Zheng M. METTL14 suppresses proliferation and metastasis of colorectal cancer by down-regulating oncogenic long non-coding RNA XIST. Mol Cancer 2020; 19: 46.
[24] Arguello AE, DeLiberto AN and Kleiner RE. RNA Chemical proteomics reveals the N(6)-methyladenosine (m(6)A)-regulated protein-RNA interactome. J Am Chem Soc 2017; 139: 17249-17252.
[25] Chen XY, Zhang J and Zhu JS. The role of m(6) A RNA methylation in human cancer. Mol Cancer 2019; 18: 103.
[26] Li Y, Xiao J, Bai J, Tian Y, Qu Y, Chen X, Wang Q, Li X, Zhang Y and Xu J. Molecular characterization and clinical relevance of m(6)A regulators across 33 cancer types. Mol Cancer 2019; 18: 137.
[27] Liu S, Li Q, Chen K, Zhang Q, Li G, Zhuo L, Zhai B, Sui X, Hu X and Xie T. The emerging molecular mechanism of m(6)A modulators in tumorigenesis and cancer progression. Biomed Pharmacother 2020; 127: 110098.
[28] Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G and Noushmehr H. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016; 44: e71.
[29] Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33: 1-22.
[30] Scrucca L, Santucci A and Aversa F. Competing risk analysis using R: an easy guide for clinicians. Bone Marrow Transplant 2007; 40: 381-387.
[31] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT and Lin CY. CytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014; 8 Suppl 4: S11.
[32] Yu G, Wang LG, Han Y and He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16: 284-287.
[33] Hänzelmann S, Castelo R and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013; 14: 7.
[34] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB and Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013; 4: 2612.
[35] Chen B, Khodadoust MS, Liu CL, Newman AM and Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol 2018; 1711: 243-259.
[36] Yu T, Wang FL, Meng LB, Li JK and Miao G. Early detection of recurrent laryngeal nerve damage using intraoperative nerve monitoring during thyroidectomy. J Int Med Res 2020; 48: 300060519889452.
[37] Schneider M, Dahm V, Passler C, Sterrer E, Mancusi G, Repasi R, Gschwandtner E, Fertl E, Handgriff L and Hermann M. Complete and incomplete recurrent laryngeal nerve injury after thyroid and parathyroid surgery: characterizing paralysis and paresis. Surgery 2019; 166: 369-374.
[38] Lynch J and Parameswaran R. Management of unilateral recurrent laryngeal nerve injury after thyroid surgery: a review. Head Neck 2017; 39: 1470-1478.
[39] Biondi B and Cooper DS. Thyroid hormone suppression therapy. Endocrinol Metab Clin North Am 2019; 48: 227-237.
[40] Biondi B and Wartofsky L. Treatment with thyroid hormone. Endocr Rev 2014; 35: 433-512.

[41] Manzardo OA, Cellini M, Indirli R, Dolci A, Colombo P, Carrone F, Lavezzi E, Mantovani G, Mazziotti G, Arosio M and Lania AGA. TNM 8th edition in thyroid cancer staging: is there an improvement in predicting recurrence? Endocr Relat Cancer 2020; 27: 325-336.

[42] Lee J, Lee SG, Kim K, Yim SH, Ryu H, Lee CR, Kang SW, Jeong JJ, Nam KH, Chung WY and Jo YS. Clinical value of lymph node ratio integration with the 8(th) edition of the UICC TNM classification and 2015 ATA risk stratification systems for recurrence prediction in papillary thyroid cancer. Sci Rep 2019; 9: 13361.

[43] Lamartina L, Grani G, Arvat E, Nervo A, Zatelli MC, Rossi R, Puxeddu E, Morelli S, Torlontano M, Massa M, Bellantone R, Pontecorvi A, Montesano T, Pagano L, Daniele L, Fugazzola L, Ceresini G, Bruno R, Rossetto R, Tumino S, Centanni M, Meringolo D, Castagna MG, Salvatore D, Nicolucci A, Lucisano G, Filetti S and Durante C. 8th edition of the AJCC/TNM staging system of thyroid cancer: what to expect (ITCO#2). Endocr Relat Cancer 2018; 25: L7-L11.

[44] Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM and Wartofsky L. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid 2016; 26: 1-133.

[45] Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL and Tuttle RM. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid 2009; 19: 1167-1214.

[46] Liu ZX, Li LM, Sun HL and Liu SM. Link between m6A modification and cancers. Front Bioeng Biotechnol 2018; 6: 89.

[47] Liu H, Zhang M, Shi M, Zhang T, Zhang Z, Cui Q, Yang S and Li Z. A survival-related competitive endogenous RNA network of prognostic lncRNAs, miRNAs, and mRNAs in Wilms tumor. Front Oncol 2021; 11: 608433.

[48] Buishand FO, Liu-Chittenden Y, Fan Y, Tirosh A, Gara SK, Patel D, Meerzaman D and Kebebew E. Adrenocortical tumors have a distinct, long, non-coding RNA expression profile and LINC00271 is downregulated in malignancy. Surgery 2020; 167: 224-232.

[49] Murugan AK, Munirajan AK and Alzahrani AS. Long noncoding RNAs: emerging players in thyroid cancer pathogenesis. Endocr Relat Cancer 2018; 25: R59-R82.

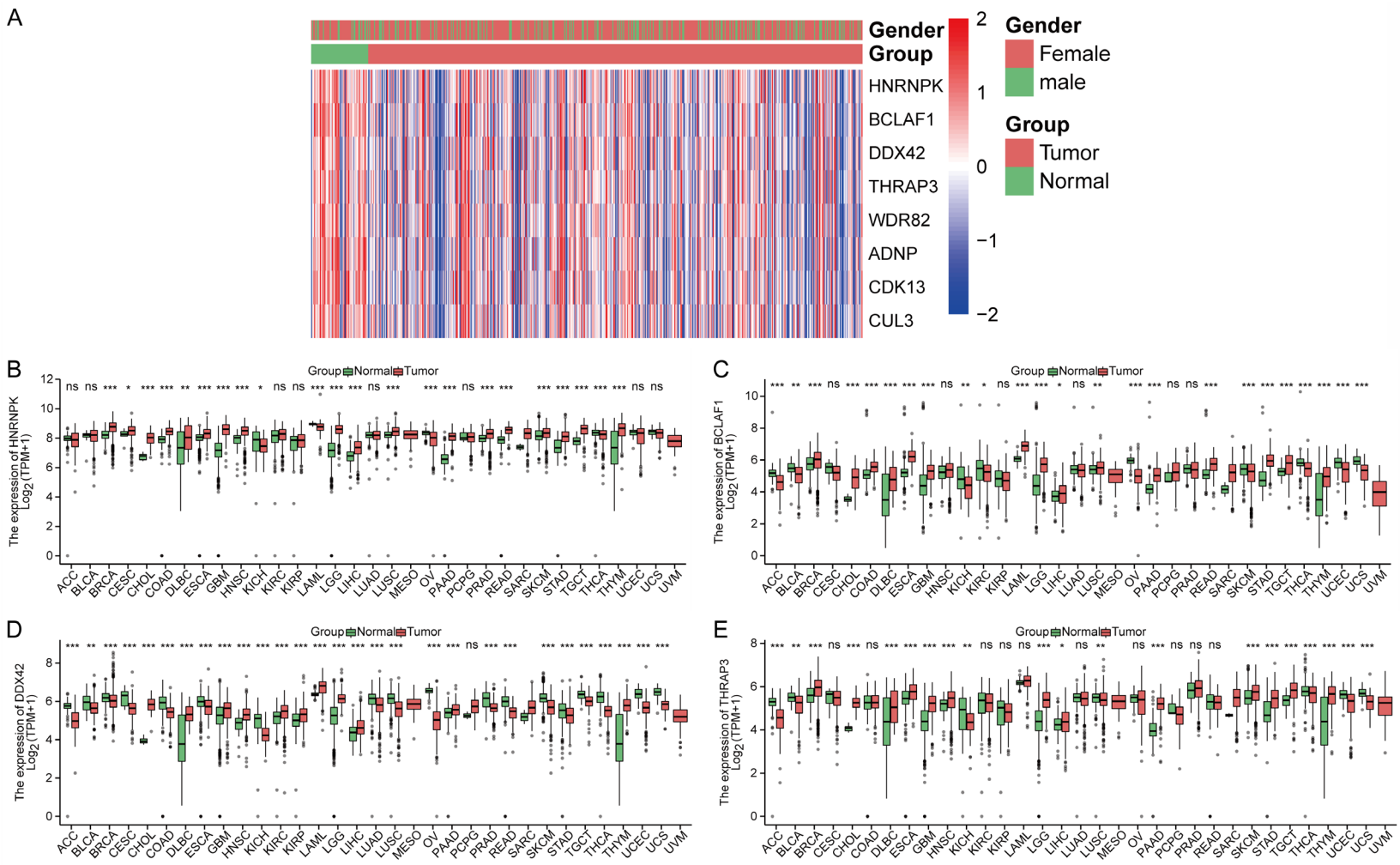
[50] Ma B, Liao T, Wen D, Dong C, Zhou L, Yang S, Wang Y and Ji Q. Long intergenic non-coding RNA 271 is predictive of a poorer prognosis of papillary thyroid cancer. Sci Rep 2016; 6: 36973.

[51] Wang CL, Li JC, Zhou CX, Ma CN, Wang DF, Wo LL, He M, Yin Q, He JR and Zhao Q. Long noncoding RNA ARHGAP5-AS1 inhibits migration of breast cancer cell via stabilizing SMAD7 protein. Breast Cancer Res Treat 2021; 189: 607-619.

[52] Zhu L, Zhu Y, Han S, Chen M, Song P, Dai D, Xu W, Jiang T, Feng L, Shin VY, Wang X and Jin H. Impaired autophagic degradation of lncRNA ARHGAP5-AS1 promotes chemoresistance in gastric cancer. Cell Death Dis 2019; 10: 383.

[53] Wu M, Yuan H, Li X, Liao Q and Liu Z. Identification of a five-gene signature and establishment of a prognostic nomogram to predict progression-free interval of papillary thyroid carcinoma. Front Endocrinol (Lausanne) 2019; 10: 790.

[54] Gu Y and Hu C. Bioinformatic analysis of the prognostic value and potential regulatory network of FOXF1 in papillary thyroid cancer. Biofactors 2019; 45: 902-911.


A
Gender
Group
HNRNPK
BCLAF1
DDX42
THRAP3
WDR82
ADNP
CDK13
CUL3
2
1
0
−1
−2
Gender
Female
male
Group
Tumor
Normal
B
C
D
E
Group Normal Tumor
The expression of HNRNPK Log2 (TPM+1)
The expression of BCLAF1 Log2 (TPM+1)
The expression of DDX42 Log2 (TPM+1)
The expression of THRAP3 Log2 (TPM+1)
ACC BLCA BRCA CESC CHOL COAD DLBC ESCA GBM HNSC KICH KIRC KIRP LAML LGG LIHC LUAD LUSC MESO OV PAAD PCPG PRAD READ SARC SKCM STAD TGCT THCA THYM UCEC UCS UVM

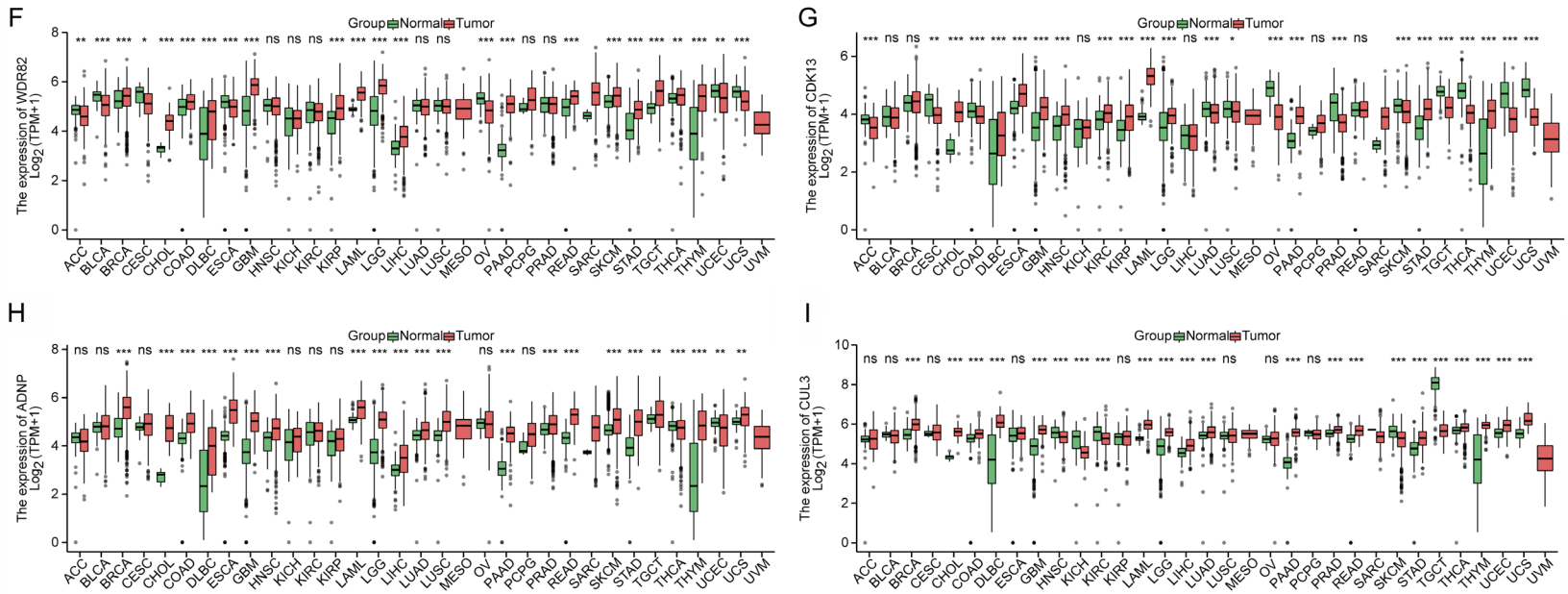
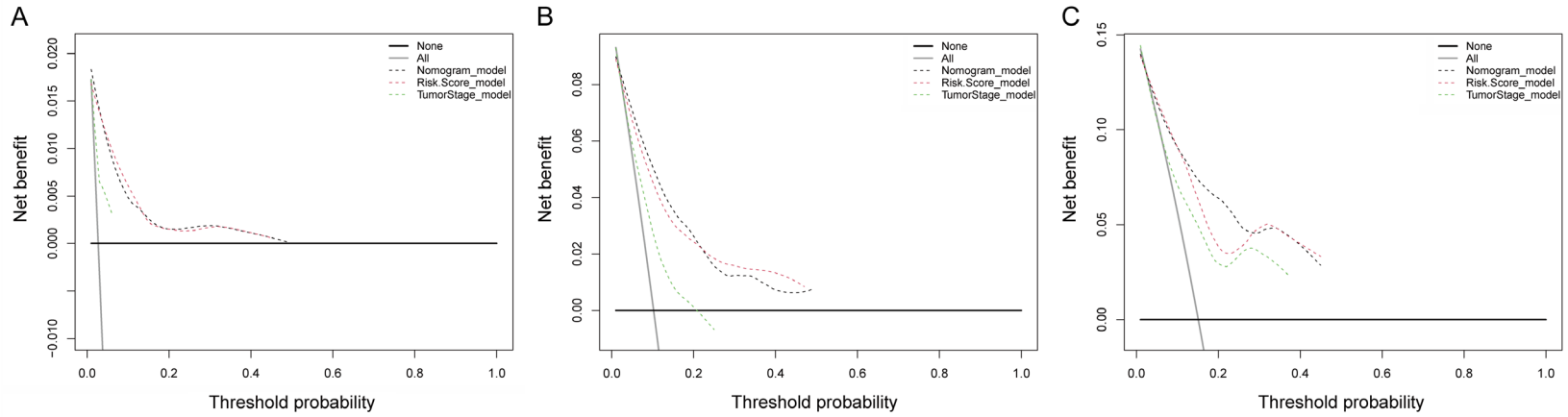


Figure S1. Difference in the expression of eight key m6A-mRNA molecules in pan-cancer tumor tissues. A. Heat map display of the expression levels of eight key m6A-mRNA molecules between thyroid cancer and adjacent tissues; B-I. Results of differential expression of HNRNPK, BCLAF1, DDX42, THRAP3, WDR82, ADNP, CDK13, and CUL3 molecules between 33 pan-cancer tumor tissues and corresponding 31 normal tissues; *P < 0.05; **P < 0.01; ***P < 0.001.



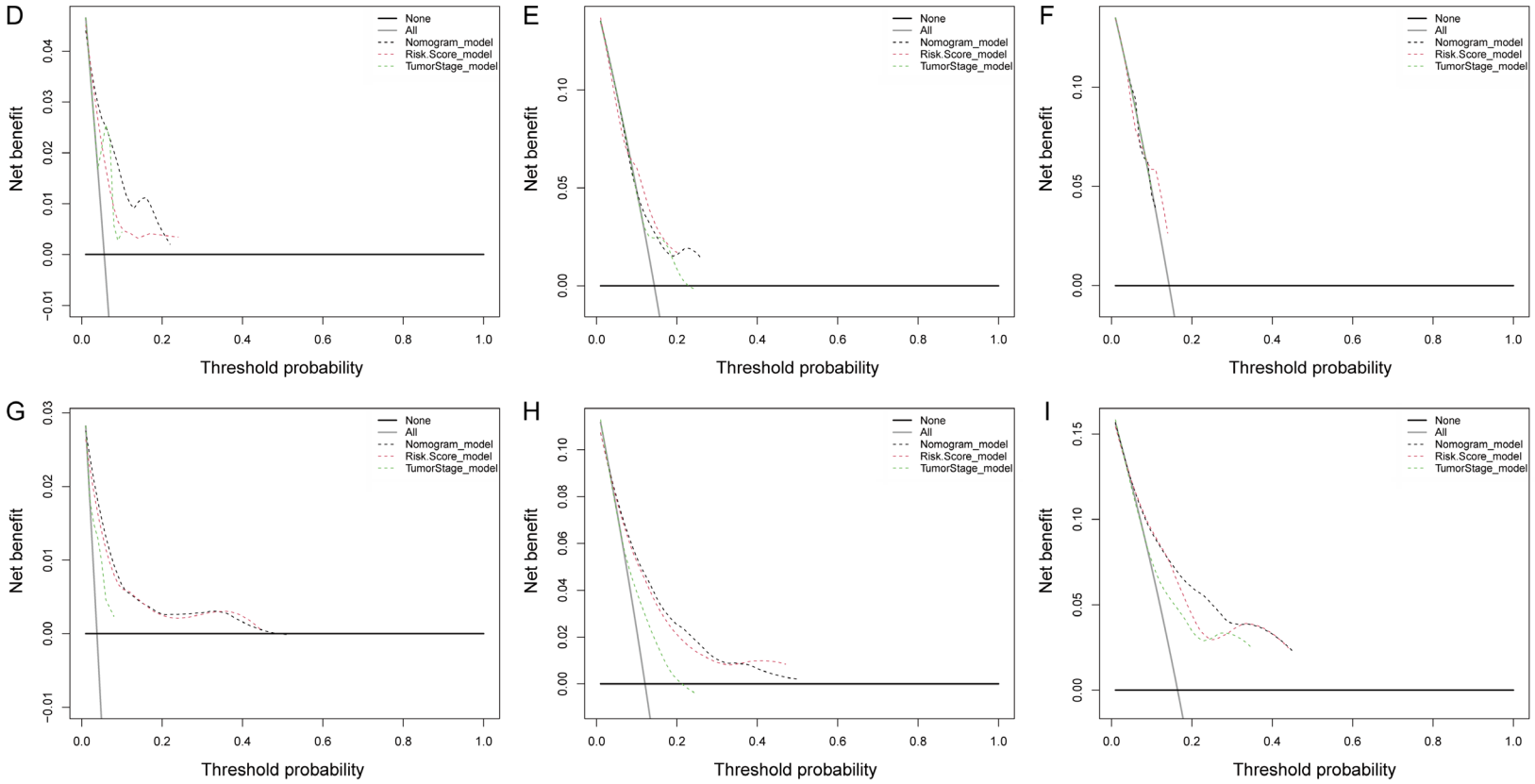


Figure S2. DCA curve of the nomogram model in the TCGA training set, validation set, and the entire TCGA dataset. A-C. DCA images of 1-, 3-, and 5-year prognostic prediction results in the training set; D-F. DCA images of 1-, 3-, and 5-year prognosis prediction results in the validation set; G-I. DCA images of 1-, 3-, and 5-year prognostic prediction results in the entire dataset.