Analysis of the prognostic significance and potential mechanisms of lncRNAs associated with m6A methylation in papillary thyroid carcinoma

Yinde Huang

The First Affiliated Hospital of China Medical University

Xin Li

The First Affiliated Hospital of China Medical University

Wenbin Chen

The First Affiliated Hospital of China Medical University

Yuzhen He

The First Affiliated Hospital of China Medical University

Song Wu

The First Affiliated Hospital of China Medical University

Xinyang Li

The First Affiliated Hospital of China Medical University

Bingchen Hou

The First Affiliated Hospital of China Medical University

Shiyue Wang

The First Affiliated Hospital of China Medical University

Yuchen He

The First Affiliated Hospital of China Medical University

Han Jiang

The First Affiliated Hospital of China Medical University

Yu Lun

The First Affiliated Hospital of China Medical University

Jian Zhang (✉ jianzhang@cmu.edu.cn)

The First Affiliated Hospital of China Medical University <https://orcid.org/0000-0003-4448-0774>

Research Article

Keywords: m6A, Papillary thyroid carcinoma, lncRNA, prognostic biomarker, PD-L1

Posted Date: August 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-807935/v1>

Abstract

Background : m6A methylation-related long non-coding RNAs (lncRNAs) play a significant role in the progression of various tumors and can be used as prognostic markers. However, whether m6A-related lncRNAs also play the same function as prognostic markers in papillary thyroid carcinoma (PTC) remains unclear.

Methods : Consensus cluster analysis was performed to divide PTC samples obtained from The Cancer Genome Atlas database into two clusters according to the expression of m6A-related lncRNAs. Then, the least absolute shrinkage and selection operator (LASSO) regression analysis was performed to create and verify a prognostic model. Furthermore, the relationship among risk scores, clusters, programmed death-ligand 1 (PD-L1), tumor microenvironment (TME), clinicopathological characteristics, immune infiltration, immune checkpoint, and tumor mutation burden (TMB) was analyzed. In addition, a nomogram was created, and subsequently, the drug sensitivity of lncRNAs in the prognostic model was analyzed. Finally, the relationship between these lncRNAs and prognosis in pan-cancer was investigated.

Results: The prognosis, RAS, BRAF, M, and TME were found to be different in two clusters. The prognostic model included three lncRNAs: PSMG3-AS1 , BHLHE40-AS1 , and AC016747.3 . The risk score was associated with clusters, PD-L1, tumor microenvironment, clinicopathological characteristics, immune cell infiltration, immune checkpoint, and TMB, and thus, risk score was confirmed as useful prognostic indicators. Differentially expressed lncRNAs are involved in many malignancies and can be identified as cancer prognostic makers.

Conclusion : According to our research, we can regard m6A-related lncRNAs involved in the procession of PTC as a biomarker of PFS for PTC patients, and pan-cancer.

Introduction

The incidence of thyroid cancer, the most common endocrine cancer, has been continuously increasing in recent decades [1, 2]. Papillary thyroid carcinoma (PTC) is the most common type of thyroid cancer with an excellent prognosis [3–5]. However, approximately 30% of patients with PTC have a poor prognosis due to local recurrence or distant metastasis [6, 7]. Individuals with poorer prognoses should be crucially identified, and the mechanism of PTC development should be deeply understood.

Post-transcription modifications have become a significant regulator of many physiological processes and disease development, which attracts researchers' interest in cancer studies [8]. A previous study has shown that N (6)-methyladenosine (m6A) is involved in the most common internal (non-cap) alteration of the messenger RNA (mRNA) in all higher eukaryotes, which are essential for cell viability and development [9]. There are three types of m6A regulators: the methyltransferase complex (MTC) often known as “writers” that catalyzes m6A; demethylase, commonly known as “eraser” that removes the m6A; and the RNA reader protein identifies m6A, attaches to the RNA, and performs the necessary activities. In cancer etiology and progression, writers, erasers, and readers are crosslinked [10, 11].

Researchers show great interest in long non-coding RNAs (lncRNAs) after understanding their functions. lncRNAs, the non-coding transcripts, are involved in at least 200 nucleotides [12]. However, their functions are still mostly unclear. lncRNAs may function as mRNA expression control via translational regulation, histone changes, or posttranscriptional regulation [13]. They can also impact numerous tumor cell characteristics (survival, proliferation, and migration) by gene control [12]. Aberrant lncRNA expression is involved in tumor aggressiveness and procession in various malignancies, including PTC. At this point, researchers have concluded that m6A, a necessary part of the complex lncRNA process, functions on tumor development [14]. However, the mechanism and prognostic value of m6A-related lncRNAs in PTC remains unknown.

Transcriptome data of PTC were obtained from a public source (TCGA-THCA database) and divided into two groups according to the consensus cluster analysis based on the expression level of m6A-related lncRNAs. To further accurately predict the progression-free survival (PFS) of patients with PTC, the prognostic model of m6A-related lncRNAs was developed and verified. The relationship among, programmed death-ligand 1 (PD-L1), tumor microenvironment (TME), risk score, clinicopathological characteristics, immune cell infiltration, tumor mutation burden (TMB), and clusters were also analyzed. Subsequently, to explore the predicted probability of the PFS in patients with PTC, a nomogram was drawn, and the drug sensitivity of three lncRNAs was analyzed in the prognostic model. Finally, the relationship among three lncRNAs and the prognosis in pan-cancer was also analyzed.

Materials And Methods

Data collections

Relevant genetic alterations and pathways that influence the tumor genesis, progression, differentiation, and metastasis [15] can be investigated because of a tremendous amount of genomic and clinical data available in the Cancer Genome Atlas (TCGA). The TCGA-THCA dataset contains 58 adjacent nontumorous tissues (N) and 510 patients with PTC (T) and 487 with simple nucleotide variation data (VarScan), which are used to calculate the TMB of patients with PTC. The University of California Santa Cruz (UCSC) genome database (<https://xenabrowser.net>; accessed 12 July 2021) provided the clinical data of 501 patients with PTC [16, 17]. Patients with incomplete follow-up information were excluded.

To broaden the scope of our study, 23 traditional regulators were included based on previous studies: writers (*METTL3*, *METTL14*, *METTL16*, *WTAP*, *VIRMA* [*KIA1499*], *ZC3H13*, *RBM15*, and *RBM15B*), readers (*YTHDC1*, *YTHDC2*, *YTHDF1*, *YTHDF2*, *YTHDF3*, *HNRNPC*, *FMR1*, *LRPPRC*, *HNRNPA2B1*, *IGFBP1*, *IGFBP2*, *IGFBP3*, and *RBMX*), and erasers (*FTO* and *ALKBH5*) [18-20].

Screening out lncRNAs associated with m6A methylation

After extracting expressions of 23 m6A regulators from the TCGA-THCA dataset, co-expression analysis between m6A regulators and m6A-related lncRNAs is performed using the “limma” package. Based on the “|correlation coefficient| of ≥ 0.5 ” and “*p*-value of < 0.001 ,” co-expression analysis was conducted and

regulators of m6A methylation-related lncRNAs were selected using the Pearson correlation test. The expression information of the co-expression network for m6A RNA methylation-related lncRNA was also achieved in the “igraph” R package. m6A-related lncRNAs associated with the PFS of patients with PTC were screened using univariate COX progression analysis, and lncRNAs with $p < 0.01$ were included in the follow-up study.

Consensus Clustering Analysis

In the TCGA-THCA cohort, patients with PTC were divided according to clusters and examined the biological features of m6A-related lncRNAs using the R package “ConsensusClusterPlus.” Then, the PFS of clusters were estimated using the Kaplan-Meier analysis. The probable functionalities and downstream access of two clusters were observed using the Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) to evaluate the composition of the TME including the estimate score (stromal-immune comprehensive score), immune score (immune content), stromal score (stromal content). The R package “corrplot” was used to explore the correlation between the expression levels of PD-L1 and m6A-related lncRNAs.

Construction and validation of the prognostic model

After evenly dividing samples collected from TCGA-THCA into training and testing datasets, the least absolute shrinkage and selection operator (LASSO) regression analysis method was performed using the “glmnet” package [21] in the R4.1.0 to develop the prognostic model of m6A-related lncRNAs for PTC samples in the training dataset. The risk score of the LASSO regression model can be calculated as follows [22]:

$$\text{Risk score} = \sum_{i=1}^n \beta_i * E_i$$

where β_i represented the lncRNA coefficient and E_i represented the lncRNA expression. The risk score was calculated using this method in the training, testing, and TCGA-THCA datasets. Three datasets were divided into the high- and low-risk groups according to the median risk score. The area under the curve (AUC) was used to analyze the predictive power of signatures using the receiver operating characteristic (ROC) curves. The “stats” R package with the “prcomp” function was used to show principal component analysis (PCA) for the lncRNA expression. t-Distributed Stochastic Neighbor Embedding (t-SNE) was also performed using the “Rtsne” R package to explore variation from different groups in terms of their distribution. Heatmaps were drawn to display the distribution of clinicopathological characteristics between the high- and low-risk groups.

Analysis of the value of prognostic models

To assess the usefulness of risk score as independent prognostic variables, Cox regression models were used with other clinical characteristics (age [≥ 55 vs. < 55 years], gender [male vs. female], American Joint

Committee on Cancer [AJCC] stage [III+IV vs. I+II], AJCC M [M1 vs. M0], AJCC N [N1 vs. N0], AJCC T [T3+4 vs. T1+2], BRAF [mutated vs. wild-type], RAS [mutated vs. wild-type], and risk score (high vs. low)]. A stratification study was conducted to determine whether risk score retained their capacity in analyzing PFS among different subgroups according to the order stratified analysis based on clinicopathological characteristics, which further evaluates the prognostic capability. A heatmap was created using the R package “pheatmap” that depicted the correlation between clinicopathological variables and groups. Given the significance of checkpoint inhibitor-based immunotherapies, differences in 47 immune checkpoint expressions were explored further between the two subgroups, and the connection among these lncRNAs was shown (47 immune checkpoints are provided in **Supplementary List 1**). Subsequently, we investigated the correlation among risk scores, immune cells, or TMB using the Pearson correlation analysis.

Sensitivity Analysis of Chemotherapy

In the cancer research community, a database, known as Cellminer, was developed to determine the integration and analysis of molecular and pharmacological data for the NCI-60 malignant cell lines collected by the United States Developmental Therapeutics Program. More than 100,000 chemical compounds and natural items will be screened by the National Cancer Institute (since 1990) (<https://discover.nci.nih.gov/cellminer/home.do>). The Pearson correlation analysis was used to evaluate the connection between prognostic lncRNAs expression and drug sensitivity. The correlation study was used to explore the effectiveness of 263 medicines that have been authorized by the Food and Drug Administration or are currently undergoing clinical studies.

Prognosis-related lncRNA's pan-cancer value promotion

The Gene Expression Profiling Interactive Analysis 2 (GEPIA2) (<http://gepia2.cancer-pku.cn/#index>), which can integrate the TCGA and Genotype-Tissue Expression Project (GTEx) database, was used to analyze the pan-cancer value about prognostic-related lncRNAs [23]. Expression levels of three prognostic-related lncRNAs were first investigated in 33 tumors and normal tissues (p -value <0.05 , $|\log FC| >0.1$, differential methods: limma). We also explored the relationship between lncRNA and the prognosis of 33 tumors.

Statistical analysis

The Kaplan-Meier survival analysis (PFS considered as the end-point) was performed using the “survival” package in R. The chi-square test was performed to explore differences in clinical parameters of the high- and low-risk groups in the training, testing, and overall TCGA-THCA datasets. The Pearson's correlation analysis was performed to assess related variables. A pair of independent t-test samples was used to compare the results in the Mann-Whitney test to explore differences between high- and low - risk samples. A log-rank test was used to calculate the p-values for PFS. Statistical analyses were performed and visualized in R version 4.1.0, IBM SPSS Statistics for Windows version 25.0 (IBM Corp., Armonk, NY, USA), and GraphPad Prism version 8.0 (GraphPad Software Inc., La Jolla, CA, USA).

Results

Analyzing the consensus cluster based on median expression levels of m6A-related lncRNAs in patients with PTC

We firstly found that 630 lncRNAs co-expressed with 15 m6A methylation regulators and then used them to visualize a co-expression network (**Supplementary Figure 1**). Red nodes show the m6A RNA methylation regulators, and blue nodes display m6A-related lncRNAs.

To further investigate the relationship between lncRNAs and the PFS of patients with PTC, we found prognostic-related lncRNAs (**Figure 2A, B**) using the univariate COX analysis and collected 21 lncRNAs (*AC010168.2*, *CAPN10-DT*, *PSMG3-AS1*, *AL662844.4*, *AC010761.3*, *AC093752.3*, *LMNTD2-AS1*, *AL031705.1*, *CD44-AS1*, *U73166.1*, *BHLHE40-AS1*, *MZF1-AS1*, *AL360181.2*, *AC138956.1*, *AL359878.1*, *AC002128.1*, *AC084125.2*, *AC139795.2*, *AC016747.3*, *AL355987.4*, and *AC138956.2*) for subsequent studies.

Based on the expression of 21 m6A-related lncRNAs, the TCGA-THCA cohort was divided into distinct clusters using the “ConsensusClusterPlus” R package (**Figure 2C, D, E**). Regarding 2 as the consensus matrix’s k-value, the least crossover was obtained from all PTC samples. The difference in PFS was explored between two clusters using the Kaplan-Meier technique. PFS was significantly longer in cluster 1 (PFS, $p = 0.005$) as compared to cluster 2 (**Figure 2F**). After comparing the clinicopathological characteristics of two clusters and analyzing their connection, RAS, BRAF, and AJCC M stage were all significantly related to our cluster analysis (**Figure 2G**). The percentage of the 22 immune cell types in two clusters was analyzed using the CIBERSORT. CD8 T cells, CD4 memory activated T cells, follicular helper T cells, macrophages M1, and activated dendritic cells were significantly different between the two clusters (**Figure 2H**). The TME was examined using the ESTIMATE, stromal, and immune scores and found significant differences in the two groups (**Figure 2I, J, K**). Considering the important function of PD-L1 in tumor treatment, a correlation was observed among the PD-L1 and 21 m6A-related lncRNAs (**Figure 2L**) and the expression of PD-L1 between PTC samples and normal tissues in clusters 1 and 2 (**Supplementary Figure 2A, B**). The expression of PD-L1 was negatively correlated with 18 lncRNAs. Statistically significant changes in PD-L1 expression were observed between tumor and healthy control. However, compared to cluster 2, expression levels of PD-L1 expression were higher in cluster 1.

Identification and verification of the prognostic model of m6A-related lncRNAs

Next, the effectiveness of the prognostic model was evaluated. After randomly dividing the entire TCGA-THCA dataset into two groups (training dataset [$n = 252$] and testing dataset [$n = 249$]), a prognostic model was obtained in the training dataset using the LASSO regression analysis based on the expression of 21 lncRNAs (**Supplementary Figure 3A, B**). A total of three lncRNAs (*PSMG3-AS1*, *BHLHE40-AS1*, and *AC016747.3*) were included in the prognostic model. The risk score of patients in the training, testing, and entire TCGA-THCA dataset were calculated by the prognostic model. The following prognostic model formula was used: risk score = $(0.264992826432626 \times \text{PSMG3-AS1 expression level}) +$

($-0.777201837590316 \times BHLHE40\text{-}AS1$ expression level) + ($1.34421059388075 \times AC016747.3$ expression level). According to the median risk score, patients in all three datasets were divided into high- and low-risk groups. The associations among the risk score, prognostic status, and expressive signatures of three m6A-related lncRNAs in the training, testing, and entire TCGA-THCA cohorts are shown (**Supplementary Figure 4A-I**). The PFS between the two groups was further analyzed, with significant differences, whether in the training dataset ($p = 0.004$, **Figure 3A**), the testing dataset ($p = 0.001$, **Figure 3B**), or the entire TCGA-THCA dataset ($p < 0.001$, **Figure 3C**). As shown in the time-dependent ROC curves, significant differences were found between the high- and low-risk groups. AUC analyses for 1-, 2-, 3-year PFS prediction according to the lncRNA signature were 0.747, 0.626, and 0.722 in the training dataset (**Figure 3D**); 0.709, 0.651, and 0.771 in the testing dataset (**Figure 3E**); 0.733, 0.634, and 0.739 in the entire TCGA-THCA dataset (**Figure 3F**). PCA and t-SNE analyses displayed that patient in the three datasets were dispersed in two methods (**Figure 3G-L**). Thus, the risk score is relatively accurate in determining the PFS of patients with PTC because it is greatly distinct among patients.

Univariate and multivariate Cox analyses were performed to examine the association between the risk score and PFS. A risk score was found as a risk factor for PFS in all three datasets, as demonstrated by the univariate cox analysis results (training dataset: hazard ratio [HR] = 2.767, 95% confidence interval [CI] = 1.891-4.050, $p < 0.001$, **Figure 3M**; testing dataset: HR = 6.817, 95%CI = 2.375-19.561, $p < 0.001$, **Figure 3N**; entire TCGA-TCHA dataset: HR = 3.147, 95%CI = 2.260-4.383, $p < 0.001$, **Figure 3O**). In the multivariate Cox regression analysis, the risk score remained a significant predictor of PFS even considering the effects of the other factors (training dataset: HR = 2.849, 95%CI = 1.702-4.771, $p < 0.001$, **Figure 3P**; testing dataset: HR = 40.147, 95%CI = 3.020-533.628, $p = 0.005$, **Figure 3Q**; entire TCGA-THCA dataset: HR = 2.458, 95%CI = 1.593-3.794, $p < 0.001$, **Figure 3R**).

Stratification analysis based on clinicopathological characteristics

The prognostic profile associated with m6A-mediated lncRNAs might be used to predict PFS in both high- and low-risk groups. A stratified analysis on clinicopathological variables showed appropriate results, including age (≤ 55 years vs. > 55 years), gender (female vs. male), AJCC stage (stage $\boxtimes + \boxtimes$ vs. $\boxtimes + \boxtimes$), AJCC T stage (T $\boxtimes + \boxtimes$ vs. T $\boxtimes + \boxtimes$), AJCC M stage (M0 vs. M1), AJCC N stage (N0 vs. N1), BRAF (wild-type vs. mutated), and RAS (wild-type vs. mutated) ($p < 0.05$, **Figure 4**). The Kaplan-Meier method of survival analysis has been performed, and results have shown that the high-risk group has a shorter PFS than the low-risk group in the number of clinical strata, except for patients aged < 55 years, and patients with RAS mutation (**Figure 4A**). This additional proof confirmed our hypothesis on its significance.

The correlation between risk score and clinicopathology characteristics, cluster analysis results, or immune scores

When comparing high- with low-risk samples from the TCGA-THCA dataset with clinicopathological factors, immune scores, and cluster analysis results, expression levels of *PSMG3-AS1* and *AC016747.3* were found to be higher in the high-risk group (**Figure 5A**), indicating high-risk lncRNAs. Conversely, the expression level of *BHLHE40-AS1* in low-risk samples was higher than in high-risk samples, indicating a

low-risk lncRNA. A higher-risk value can be observed in the RAS-mutated (**Figure 5B**), BRAF wild-type (**Figure 5C**), N0 (**Figure 5D**), age ≥ 55 years (**Figure 5E**), ImmuneScore low (**Figure 5F**), and cluster 2 (**Figure 5G**). Then, we found differences in the expression levels of 30 immune checkpoints between high- and low-risk samples (**Figure 5H**).

Correlation between risk score with immunocytes and tumor mutation burden and prognostic nomogram

After performing the CIBERSORT analysis, Pearson correlation was used to analyze the correlation between risk scores and 22 immune cells. A positive correlation can be observed between regulating T cells (Tregs) and risk score (**Figure 6C**), whereas the correlation between resting dendritic cell (**Figure 6A**) or resting mast cell (**Figure 6B**) and risk scores was negative. To quantitatively predict the patient's prognosis, a nomogram was created based on clinicopathological characteristics and risk scores to accurately predict the patient's PFS (**Figure 6E**).

Sensitivity analysis of chemotherapy

Two prognosis-related lncRNAs (*PSMG3-AS1*, *BHLHE40-AS1*) were found to correspond to 11 drugs (**Figure 7**). As the *PSMG3-AS1* expression increases, the sensitivity of cancer cells to these drugs (mitomycin C, mitoxantrone, tegafur, idarubicin, gemcitabine, fulvestrant, and teniposide) increases. As the *PSMG3-AS1* expression increases, the resistance of cancer cells to the drug (bosutinib) increases. The sensitivity of cancer cells to epirubicin, mitoxantrone, and doxorubicin and their resistance to allopurinol increase while increasing the *BHLHE40-AS1* expression levels. Remarkably, as the expression of these two lncRNAs increases, the sensitivity of cancer cells to mitoxantrone increases (**Figure 7C, H**).

Prognosis-related lncRNA's pan-cancer value promotion

We found that these three lncRNAs are differentially expressed in most tumors (**Figure 8A, C, E**). The results show that the overall survival rate is significantly different with the increasing *PSMG3-AS1* and *BHLHE40-AS1* expression levels (**Figure 8B, D**). A significant difference in disease-free survival rate was observed between patients with low- and high expressions of *AC016747.3* (**Figure 8F**).

Discussion

The role of post-transcriptional modification in life activities has increasingly received attention [8]. Among the many RNA modifications, m6A is the most common type of modification and is involved in the occurrence and development of different tumors [14, 24]. Some studies have reported that METTL3 inhibits the progression of thyroid papillary carcinoma by affecting neutrophil infiltration through m6A /c-Rel/IL-8 [25] and regulating m6A methylation on TCF1 [26]. Another study has found that the protein content of FTO correlated with lymph node metastasis and tumor grade is an independent prognostic factor of thyroid cancer grade [27]. Some studies have shown that m6A-methylation-related regulators can predict the prognosis of patients with PTC using bioinformatics methods [28–30]. Therefore, the formation and prognosis of patients with PTC can be evaluated with m6A methylation. The function of

lncRNAs in different cancers has greatly increased the interest of researchers. Thus, studies have claimed that m6A is a key step in the lncRNA process that plays a role in the pathological process of tumors [31], and we can infer the prognosis of patients through the expression of m6A-methylation-related lncRNAs, such as pancreatic adenocarcinoma [32], ovarian cancer [33], hepatocellular carcinoma [34], triple-negative breast cancer [35], primary head and neck squamous cell carcinoma [14], kidney renal clear cell carcinoma [36], metastatic skin cutaneous melanoma [37], and lower-grade glioma [38]. No studies have shown the possibility of predicting the prognosis of patients with PTC by m6A-related lncRNAs. Therefore, we conducted this study using Pearson's method to obtain 630 lncRNAs co-expressed with m6A genes based on the expression levels of 15 m6A genes.

We performed the univariate COX analysis to screen out 21 lncRNAs associated with prognosis and draw forest maps and heat maps. The results showed that 20 lncRNAs are risk factors for prognosis, and one lncRNA is a protective factor for prognosis. According to the expression of 21 lncRNAs, TCGA-THCA cohort were divided into two clusters by consensus clustering method. The results show that patients in cluster2 have a worse prognosis, with higher expression of 21 lncRNAs.

To further study the function of these lncRNAs, we analyzed fractions of infiltrating immune cells between two clusters by CIBERSORT, and the results showed that the content of CD8 T cells, CD4 memory active T cells, follicular helper T cells, macrophages M1, and dendritic cells activated were statistically different, indicating that the immune infiltration between the two clusters was different. Therefore, the ESTIMATE analysis method was used to further explore the tumor microenvironment between the two clusters. Cluster 2 was also found to show a lower estimate score, matrix score, and immune score, which emphasized the role of these lncRNAs in the TME. Considering the importance of PD-L1 during the tumor immunity process [39–42], a correlation between PD-L1 and 21 lncRNAs was observed. The results showed that PD-L1 was negatively correlated with several lncRNAs, and most of the lncRNAs are positively related.

Two datasets (a training group and a testing group) were created by a random method and obtained a prognostic model containing three m6A-related lncRNAs (*PSMG3-AS1*, *BHLHE40-AS1*, and *AC016747.3*) by the LASSO regression analysis in the training group, all of which can be used to further increase the feasibility of our prediction.

Previous reports have indicated that *PSMG3-AS1* sponging miR-613 to upregulate SphK1 increases the growth of non-small cell lung cancer cells [43]; and sponging miR-143-3p to increase the proliferation and migration of breast cancer cells [44]; and can function on the migration, proliferation, and invasion of cervical squamous cell carcinoma [45], lung adenocarcinoma [46], hepatocellular carcinoma [47], renal cell carcinoma [48], and breast cancer cells; can distinguish glioblastoma and sarcoidosis [49]; and can potentially predict breast cancer [50] and ovarian cancer [51]. Some studies have shown that early breast cancer progresses by increasingly expressed *BHLHE40-AS1* functioning on IL-6/STAT3 signaling [52]. *AC016747.3* has not been reported, and these three lncRNAs have not been investigated in PTC; therefore, follow-up *in vivo* and *in vitro* experiments on these lncRNAs in PTC is necessary.

Furthermore, the prognostic model was verified in the entire TCGA-THCA dataset and the testing group. Based on the prognostic model, the risk score of patients was calculated. According to the median risk value, we divided patients with PTC into high- and low-risk groups. Our analytical results show that patients from high-risk groups have worse PFS than low-risk groups (training dataset: $p = 0.004$; testing dataset: $p = 0.001$; TCGA-THCA: $p < 0.001$); ROC curve, PCA, and t-SNE analysis displayed that the prognostic model can be used to predict the patient's PFS to a certain extent. We got an analytical result according to the univariate and multivariate COX analyses: risk score is a prognostic factor independent of other clinical traits for patients with PTC. In the stratified analysis based on clinicopathological characteristics, the prognosis of patients in different groups is also statistically different. The analytical results also showed that some significant difference was observed in the risk scores among several clinicopathological characteristics (RAS, BRAF, N, age, immune score, and cluster) and is correlated with three types of immune cells (resting dendritic cell, resting mast cell, and regulatory T cells). These results increase our understanding of the function of m6A-related lncRNAs in PTC and provide directions for further experiments. To further expand the clinical value of our study, we developed a nomogram to predict the PFS of patients with PTC, performed a drug sensitivity analysis among these three lncRNAs, and found that with the expression of two lncRNAs increased sensitivity or resistance to certain chemotherapeutic drugs. Notably, as the expression level of these two lncRNAs increases, the sensitivity of tumor cells to mitoxantrone increases. Moreover, these three lncRNAs also have a certain prognostic value in pan-cancer.

This study has some limitations that should be considered. First, it is the data source from the TCGA-THCA cohort used to analyze and verify the prognostic significance of m6A-related lncRNAs. To further verify the prognostic model, we still lack more central data. Second, only the correlation between the expression of these three lncRNAs and clinicopathological characteristics, immune checkpoints, drug sensitivity, or pan-cancer was considered; however, we did not analyze pathways and mechanisms of their effects in in vivo and in vitro experiments.

Conclusion

In summary, m6A-related lncRNAs of patients with PTC were obtained based on public databases and constructed and verified a prognostic model. The three m6A-related lncRNAs in the prognostic model can be used as prognostic markers for patients with PTC. In addition, the functional analysis also provides a new perspective in future studies on the role and mechanism of m6A-related lncRNAs in patients with PTC.

Abbreviations

lncRNAs: long non-coding RNAs; PTC: papillary thyroid carcinoma; LASSO: least absolute shrinkage and selection operator; TCGA: The Cancer Genome Atlas database; PD-L1: programmed death-ligand 1, TME: tumor microenvironment; TMB: tumor mutation burden; m6A:N (6)-methyladenosine; mRNA: messenger RNA; UCSC: University of California Santa Cruz; PFS: progression-free survival; ESTIMATE: Estimation of

Stromal and Immune cells in Malignant Tumor tissues using Expression data; AUC: area under the curve; ROC: receiver operating characteristic; PCA: principal component analysis; t-SNE: t-Distributed Stochastic Neighbor Embedding; AJCC: American Joint Committee on Cancer; GEPIA2: Gene Expression Profiling Interactive Analysis 2; GTEx: Genotype-Tissue Expression Project database; HR: hazard ratio; CI: confidence interval.

Declarations

Acknowledgements

We appreciate the TCGA database for allowing access to their data.

Authors' contributions

YDH created the analytical strategies and drafted the paper. XL verified the data analysis method. XYL, YL, XL, YZH, SW, SYW, YCH, and HJ performed data analyses. JZ designed and wrote the paper. Each author equally contributed to the article, and all of them approved the final draft that was eventually published.

Funding

The National Natural Science Foundation of China sponsored this study (81600602).

Availability of data and materials

The data of current study are available from TCGA dataset.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors agree for publication.

Competing interests

The writers indicated that their work was not influenced by competing financial interests or personal connections.

References

1. Pellegriti G, Frasca F, Regalbuto C, Squatrito S, Vigneri R. Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors. *J Cancer Epidemiol.* 2013;2013:965212.
2. Rajabi S, Dehghan MH, Dastmalchi R, Jalali Mashayekhi F, Salami S, Hedayati M. The roles and role-players in thyroid cancer angiogenesis. *Endocr J.* 2019;66:277–93.
3. Schneider DF, Chen H. New developments in the diagnosis and treatment of thyroid cancer. *CA Cancer J Clin.* 2013;63:374–94.
4. Fagin JA, Wells SA. Jr.: Biologic and Clinical Perspectives on Thyroid Cancer. *N Engl J Med.* 2016;375:1054–67.
5. Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer.* 2013;13:184–99.
6. Lamartina L, Grani G, Durante C, Filetti S. Recent advances in managing differentiated thyroid cancer. *F1000Res.* 2018;7:86.
7. Davies L, Morris LG, Haymart M, Chen AY, Goldenberg D, Morris J, Ogilvie JB, Terris DJ, Netterville J, Wong RJ, Randolph G. AMERICAN ASSOCIATION OF CLINICAL ENDOCRINOLOGISTS AND AMERICAN COLLEGE OF ENDOCRINOLOGY DISEASE STATE CLINICAL REVIEW: THE INCREASING INCIDENCE OF THYROID CANCER. *Endocr Pract.* 2015;21:686–96.
8. He L, Li H, Wu A, Peng Y, Shu G, Yin G. Functions of N6-methyladenosine and its role in cancer. *Mol Cancer.* 2019;18:176.
9. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature.* 2014;505:117–20.
10. Deng X, Su R, Weng H, Huang H, Li Z, Chen J. RNA N(6)-methyladenosine modification in cancers: current status and perspectives. *Cell Res.* 2018;28:507–17.
11. Panneerdoss S, Eedunuri VK, Yadav P, Timilsina S, Rajamanickam S, Viswanadhapalli S, Abdelfattah N, Onyeagucha BC, Cui X, Lai Z, et al. Cross-talk among writers, readers, and erasers of m(6)A regulates cancer growth and progression. *Sci Adv.* 2018;4:eaar8263.
12. Ramilowski JA, Yip CW, Agrawal S, Chang JC, Ciani Y, Kulakovskiy IV, Mendez M, Ooi JLC, Ouyang JF, Parkinson N, et al. Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Res.* 2020;30:1060–72.
13. Huang Q, Yan J, Agami R. Long non-coding RNAs in metastasis. *Cancer Metastasis Rev.* 2018;37:75–81.
14. Feng ZY, Gao HY, Feng TD. Immune Infiltrates of m(6)A RNA Methylation-Related lncRNAs and Identification of PD-L1 in Patients With Primary Head and Neck Squamous Cell Carcinoma. *Front Cell Dev Biol.* 2021;9:672248.
15. Cancer Genome Atlas Research Network. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis

- project. *Nat Genet.* 2013 Oct;45(10):1113–20. doi:10.1038/ng.2764. PMID: 24071849; PMCID: PMC3919969.
16. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* 2020;38:675–8.
 17. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 2019;47:D853-d858.
 18. Tang J, Wan Q, Lu J. The prognostic values of m6A RNA methylation regulators in uveal melanoma. *BMC Cancer.* 2020;20:674.
 19. Tian S, Lai J, Yu T, Li Q, Chen Q. Regulation of Gene Expression Associated With the N6-Methyladenosine (m6A) Enzyme System and Its Significance in Cancer. *Front Oncol.* 2020;10:623634.
 20. He PC, He C. m(6) A RNA methylation: from mechanisms to therapeutic potential. *Embo j.* 2021;40:e105977.
 21. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33:1–22.
 22. Huang R, Li G, Wang Z, Hu H, Zeng F, Zhang K, Wang K, Wu F. Identification of an ATP metabolism-related signature associated with prognosis and immune microenvironment in gliomas. *Cancer Sci.* 2020;111:2325–35.
 23. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 2019;47:W556-w560.
 24. Liu X, Wang P, Teng X, Zhang Z, Song S. Comprehensive Analysis of Expression Regulation for RNA m6A Regulators With Clinical Significance in Human Cancers. *Front Oncol.* 2021;11:624395.
 25. He J, Zhou M, Yin J, Wan J, Chu J, Jia J, Sheng J, Wang C, Yin H, He F. METTL3 restrains papillary thyroid cancer progression via m(6)A/c-Rel/IL-8-mediated neutrophil infiltration. *Mol Ther.* 2021;29:1821–37.
 26. Wang K, Jiang L, Zhang Y, Chen C. Progression of Thyroid Carcinoma Is Promoted by the m6A Methyltransferase METTL3 Through Regulating m(6)A Methylation on TCF1. *Onco Targets Ther.* 2020;13:1605–12.
 27. Tian R, Zhang S, Sun D, Bei C, Li D, Zheng C, Song X, Chen M, Tan S, Zhu X, Zhang H. **M6A Demethylase FTO Plays a Tumor Suppressor Role in Thyroid Cancer.** *DNA Cell Biol* 2020.
 28. Hou J, Shan H, Zhang Y, Fan Y, Wu B. m(6)A RNA methylation regulators have prognostic value in papillary thyroid carcinoma. *Am J Otolaryngol.* 2020;41:102547.
 29. Wang X, Fu X, Zhang J, Xiong C, Zhang S, Lv Y. Identification and validation of m(6)A RNA methylation regulators with clinical prognostic value in Papillary thyroid cancer. *Cancer Cell Int.* 2020;20:203.

30. Xu N, Chen J, He G, Gao L, Zhang D. Prognostic values of m6A RNA methylation regulators in differentiated Thyroid Carcinoma. *J Cancer*. 2020;11:5187–97.
31. Lan Y, Liu B, Guo H. The role of M(6)A modification in the regulation of tumor-related lncRNAs. *Mol Ther Nucleic Acids*. 2021;24:768–79.
32. Yuan Q, Ren J, Li L, Li S, Xiang K, Shang D. Development and validation of a novel N6-methyladenosine (m6A)-related multi- long non-coding RNA (lncRNA) prognostic signature in pancreatic adenocarcinoma. *Bioengineered*. 2021;12:2432–48.
33. Zheng J, Guo J, Cao B, Zhou Y, Tong J. Identification and validation of lncRNAs involved in m6A regulation for patients with ovarian cancer. *Cancer Cell Int*. 2021;21:363.
34. Li L, Xie R, Lu G. **Identification of m6A methyltransferase-related lncRNA signature for predicting immunotherapy and prognosis in patients with hepatocellular carcinoma.** *Biosci Rep* 2021, 41.
35. Wu J, Cai Y, Zhao G, Li M. A ten N6-methyladenosine-related long non-coding RNAs signature predicts prognosis of triple-negative breast cancer. *J Clin Lab Anal*. 2021;35:e23779.
36. Yu J, Mao W, Sun S, Hu Q, Wang C, Xu Z, Liu R, Chen S, Xu B, Chen M. Identification of an m6A-Related lncRNA Signature for Predicting the Prognosis in Patients With Kidney Renal Clear Cell Carcinoma. *Front Oncol*. 2021;11:663263.
37. Huang S, Lyu S, Gao Z, Zha W, Wang P, Shan Y, He J, Li Y. m6A-Related lncRNAs Are Potential Biomarkers for the Prognosis of Metastatic Skin Cutaneous Melanoma. *Front Mol Biosci*. 2021;8:687760.
38. Tu Z, Wu L, Wang P, Hu Q, Tao C, Li K, Huang K, Zhu X. N6-Methyladenosine-Related lncRNAs Are Potential Biomarkers for Predicting the Overall Survival of Lower-Grade Glioma Patients. *Front Cell Dev Biol*. 2020;8:642.
39. Ishida Y, Agata Y, Shibahara K, Honjo T. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *Embo j*. 1992;11:3887–95.
40. Francisco LM, Sage PT, Sharpe AH. The PD-1 pathway in tolerance and autoimmunity. *Immunol Rev*. 2010;236:219–42.
41. Asano T, Meguri Y, Yoshioka T, Kishi Y, Iwamoto M, Nakamura M, Sando Y, Yagita H, Koreth J, Kim HT, et al. PD-1 modulates regulatory T-cell homeostasis during low-dose interleukin-2 therapy. *Blood*. 2017;129:2186–97.
42. Sun Y, Tan J, Miao Y, Zhang Q. The role of PD-L1 in the immune dysfunction that mediates hypoxia-induced multiple organ injury. *Cell Commun Signal*. 2021;19:76.
43. Shen W, Hong X, Jin C, Xi Y. lncRNA PSMG3AS1 promotes proliferation of non-small cell lung cancer cells by sponging miR-613 to upregulate SphK1. *Cell Cycle*. 2021;20:829–38.
44. Cui Y, Fan Y, Zhao G, Zhang Q, Bao Y, Cui Y, Ye Z, Chen G, Piao X, Guo F, et al. Novel lncRNA PSMG3-AS1 functions as a miR-143-3p sponge to increase the proliferation and migration of breast cancer cells. *Oncol Rep*. 2020;43:229–39.

45. Man S, Li X, Zhu W. miR-4417 targets lncRNA PSMG3-AS1 to suppress cell invasion and migration in cervical squamous cell carcinoma. *Oncol Lett.* 2021;22:502.
46. Yue N, Ye M, Zhang R, Guo Y. MiR-449b-5p targets lncRNA PSMG3-AS1 to suppress cancer cell proliferation in lung adenocarcinoma. *BMC Pulm Med.* 2020;20:152.
47. Zhang J, Huang J, Chen W, Hu Z, Wang X. miR-143-3p Targets lncRNA PSMG3-AS1 to Inhibit the Proliferation of Hepatocellular Carcinoma Cells. *Cancer Manag Res.* 2020;12:6303–9.
48. Cheng G, Liu D, Liang H, Yang H, Chen K, Zhang X. A cluster of long non-coding RNAs exhibit diagnostic and prognostic values in renal cell carcinoma. *Aging.* 2019;11:9597–615.
49. Chen L, Wang G, Xu Z, Lin K, Mu S, Pan Y, Shan M. Overexpression of lncRNA PSMG3-AS1 Distinguishes Glioblastomas from Sarcoidosis. *J Mol Neurosci.* 2020;70:2015–9.
50. Gong PJ, Shao YC, Huang SR, Zeng YF, Yuan XN, Xu JJ, Yin WN, Wei L, Zhang JW. Hypoxia-Associated Prognostic Markers and Competing Endogenous RNA Co-Expression Networks in Breast Cancer. *Front Oncol.* 2020;10:579868.
51. Li N, Zhan X. Identification of clinical trait-related lncRNA and mRNA biomarkers with weighted gene co-expression network analysis as useful tool for personalized medicine in ovarian cancer. *Epma j.* 2019;10:273–90.
52. DeVaux RS, Ropri AS, Grimm SL, Hall PA, Herrera EO, Chittur SV, Smith WP, Coarfa C, Behbod F, Herschkowitz JI. Long noncoding RNA BHLHE40-AS1 promotes early breast cancer progression through modulating IL-6/STAT3 signaling. *J Cell Biochem.* 2020;121:3465–78.

Figures

Figure 1

Data analysis flowchart.

Figure 2

Consensus cluster analysis. (A) The forest plot of differentially expressed lncRNAs for prognosis; (B) The box plot of differentially expressed lncRNAs for prognosis; (C) Consensus CDF curve; (D) A relative increase in the CDF graph area under the curve for $k = 2-9$; (E) The TCGA-THCA dataset was divided into cluster 1 and cluster 2 with $k = 2$; (F) Progression-free survival probability between the two clusters; (G) Heatmap of clinicopathological characteristics between two clusters; (H) Violin plot of immune cell

infiltration between two clusters; (I) Box plot of ESTIMATE scores between two clusters; (J) Box plot of Immune scores between two clusters; (K) Box plot of Stromal scores between two clusters; (L) Correlation between PD-L1 and 21 lncRNAs.

Figure 3

(A) The K-M curve for PFS between the high- and low-risk groups in the training dataset. (B) The K-M curve for PFS between the high- and low-risk groups in the testing dataset. (C) K-M curve for PFS between

the high- and low-risk groups in the TCGA-THCA dataset. (D) ROC curve for predicting PFS in the training dataset. (E) ROC curve to predict the PFS in the testing dataset. (F) ROC curve to predict PFS in the TCGA-THCA dataset. (G) PCA plot in the training dataset. (H) PCA plot in the testing dataset. (I) PCA plot in the TCGA-THCA dataset. (J) t-SNE analysis of the training dataset. (K) t-SNE analysis of the testing dataset. (L) t-SNE analysis of the TCGA-THCA dataset. (M-R) Results of univariate and multivariate cox analyses indicate that the risk score is an independent prognostic indicator for patients with PTC. (M) Univariate cox analysis in the training dataset. (N) Univariate cox analysis in the training dataset. (O) Univariate cox analysis in the TCGA-THCA dataset. (P) Multivariate cox analysis in the training dataset. (Q) Multivariate cox analysis in the testing dataset. (R) Multivariate cox analysis in the TCGA-THCA dataset.

Figure 4

Stratified analysis of patient prognosis based on risk scores in different clinicopathological characteristics in the TCGA-THCA dataset. (A, B) Age (<55 vs. ≥55 years). (C, D) Gender (female vs.

male). (E, F) Stage (stage I+II vs. III+IV). (G, H) M (M0 vs. M1). (I, J) N (N0 vs. N1). (K, L) T (T I+II vs. III+IV). (M, N) BRAF (wild type vs. mutated). (O, P) RAS (wild type vs. mutated).

Figure 5

Correlation between risk scores and clinicopathological characteristics, immune scores, and immune checkpoints in the TCGA-THCA dataset. (A) Heatmap of the expression of m6A-related lncRNAs in different risk groups and clinicopathological characteristics. (B-G) Box plot of risk scores classified according to clinicopathological characteristics and immune score. (H) Expression of immune checkpoints between high- and low-risk groups.

Figure 7

Scatter plot of m6A-related lncRNAs expression and drug sensitivity.

Figure 8

The value of m6A-related lncRNAs in pan-cancer. (A) Comparison of PSMG3-AS1 expression in 33 cancers and paracancerous tissues based on GEPIA2. (B) Comparing the overall survival (OS) rate of high- and low-PSMG3-AS1 groups in 33 cancers based on GEPIA2. (C) Comparison of BHLHE40-AS1 expression in 33 cancers and paracancerous tissues based on GEPIA2. (D) Comparing the OS rate of high- and low-BHLHE40-AS1 groups in 33 cancers based on GEPIA2. (E) Comparison of AC016747.3

expression in 33 cancers and paracancerous tissues based on GEPIA2. (F) Comparing the disease-free survival (RFS) rate of high- and low AC016747.3 groups in 33 cancers based on GEPIA2.

Supplementary Files

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

- [SupplementaryFigure1.tif](#)
- [SupplementaryFigure2.tif](#)
- [SupplementaryFigure3.tif](#)
- [SupplementaryFigure4.tif](#)
- [SupplementaryList1.txt](#)