Contributions and prognostic performances of m7G RNA regulators in pancreatic adenocarcinoma

Qihang Yuan1,2, Jie Ren3, Xu Chen1,2, Yantong Dong1,2, Dong Shang1,2

1Department of General Surgery, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning116011, China; 2Clinical Laboratory of Integrative Medicine, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning116011, China; 3Department of Oncology, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning116011, China.

Pancreatic adenocarcinoma (PAAD) is the primary cause of mortality from endocrine cancer globally, bringing a burden to public health. Emerging evidence has suggested that the formation of PAAD is an extremely complicated biological process with multi-gene participation and multi-stage evolution.[1] Therefore, epigenetic or genomic instability may be responsible for the poor prognosis of PAAD patients. RNA methylation has emerged as a fundamental process in epigenetic regulation. Long non-coding RNAs (lncRNAs) have recently been shown to have the capacity to modulate the gene expression at several levels, such as post-transcriptional, transcriptional, and epigenetic.[2] Therefore, we systematically investigated contributions and prognostic performances of m7G RNA regulators in pan-cancer, especially in PAAD.

The mRNA expression profiles, copy number variation (CNV), and corresponding clinical characteristics of pan-cancer transcriptomes were obtained from The Cancer Genome Atlas (TCGA) database. For a pan-cancer overview about variations of m7G RNA regulators, CNV data from the TCGA database were analyzed. Additionally, pan-cancer analyses of differential mRNA expression and prognostic performances were conducted. All these analyses were conducted by R and TBtools and presented in the form of heat maps.

In the following section, we focused on PAAD for a deep and comprehensive study. The transcriptome data and clinical characteristics of PAAD and normal pancreas tissues were obtained from the TCGA, International Cancer Genome Consortium (ICGC), and Genotype-Tissue Expression (GTEx) project. After log2 transformation and batch correction, a total of 178 tumor samples and 171 normal samples in the TCGA and GTEx datasets, 82 tumor samples in the ICGC database were obtained for further analysis. m7G methylation-related regulators were obtained based on the following three datasets from the Molecular Signatures Database: GOMF_M7G_5_PPN_DIPHOSPHATASE_ACTIVITY, GOMF_RNA_7_METHYLGUANOSINE_CAP_BINDING, and GOMF_RNA_CAP_BINDING and previously published studies.[3] With complete expression data and survival data, 27 m7G methylation-related regulators were included for the following analysis. LncRNAs were separated from all genes based on the gtf file utilizing Perl. The transcriptome data of involved m7G methylation-related regulators and lncRNAs were collected, respectively. The Pearson correlation coefficients between m7G methylation-related regulators and lncRNAs were computed through the “cor.test” function in R. m7G methylation-related lncRNAs were selected with |correlation coefficients| >0.4 and P values <0.001. Differentially expressed genes (DEGs) were identified in PAAD samples and normal pancreatic tissues using the R package limma. m7G methylation-related regulators and m7G methylation-related lncRNAs associated with prognosis were determined by univariate Cox regression analysis. After the intersection, DEGs with prognostic values were obtained. Using the “ConsensusClusterPlus” package, PAAD samples in the TCGA cohort were divided into two clusters. Survival analysis was used to estimate the prognostic values of clusters using the Kaplan–Meier method. Then, we explored the correlation between clusters and clinicopathological traits, including age, gender, grade, stage, and survival status. Functional enrichment analyses of DEGs between the two clusters were conducted to provide insight into the potential mechanism of the above results.

The “caret” package in R was utilized to stratify the TCGA dataset of 177 PAAD patients into the train cohort (n = 89) and test1 cohort (n = 88) at a 1:1 ratio for subsequent model construction. The test2 cohort (n = 177) was assigned to all samples in the TCGA dataset, whereas the test3 cohort...
(n = 82) was assigned to all samples in the ICGC cohort. The least absolute shrinkage and selection operator (LASSO) regression model was utilized to develop a novel m7G-related prognostic signature (m7G-RPS) in the train cohort. The test1, test2, and test3 cohorts were employed to validate the predictive performance of m7G-RPS. The patients were stratified into two groups: high-risk and low-risk, with the median risk score serving as the cut-off point (risk score=\(\sum_{k=1}^{n} \exp(k \cdot b_k)\)). Similarly, Kaplan–Meier survival curves were then generated to demonstrate the prognostic significance of the model and receiver operating characteristic (ROC) curves of 1, 3, and 5 years were also plotted to verify the efficiency of predictive signatures by using the timeROC package. To test whether the m7G-RPS could become an independent prognostic indicator, we included the model risk score, age, gender, grade, and stage to perform univariate and multivariate cox regression analysis in the train, test1, and test2 datasets.

For in-depth investigation of immune components in the tumor microenvironment, the TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCPCOUNTER, XCELL, and EPIC algorithms were utilized to assess immune responses between high-risk and low-risk subgroups based on the signature, and the results were shown in the form of heat maps. The differential expression analysis of common immune checkpoint genes (ICG) in high-risk and low-risk subgroups was conducted and only the statistically significant results were shown.

To summarize and visualize the variation of m7G RNA regulators in various cancers, CNV results were visualized in the form of heat maps [Figure 1A–D]. According to contemporary cancer research, aberrant mRNA expression might indicate that the relevant gene has a high likelihood to perform a critical function in disease progression. For a visual exhibition, mRNA expression levels of m7G RNA regulators in pan-cancer were depicted in Figure 1C and D. What is more, univariable Cox regression analysis was implemented to explore the prognostic significance of m7G RNA regulators in pan-cancer [Figure 1E]. In the following specific analyses about PAAD, 277 m7G-related lncRNAs were selected based on Pearson correlation coefficients with \(P\) values <0.001 and |correlation coefficients| >0.4 [Supplementary Figure 1A, http://links.lww.com/CM9/B293]. Totally, 27 m7G RNA regulators and 277 m7G-related lncRNAs were merged as candidate genes. Among these candidate genes, 66 DEGs with prognostic values were identified for the construction of prognostic signature [Supplementary Figure 1B–D, http://links.lww.com/CM9/B293].

Next, PAAD samples were categorized into two subgroups with slower cumulative distribution function CDF growth rates that were distinct and non-overlapping [Supplementary Figure 2A, http://links.lww.com/CM9/B293]. The potential differences of cluster 1 and cluster 2 were investigated. Notably, samples in cluster 2 showed a lower survival rate than those in cluster 1 [Supplementary Figure 2B, http://links.lww.com/CM9/B293]. The clustering result was significantly associated with stage and grade, suggesting it was related to the malignancy of PAAD [Supplementary Figure 2C, http://links.lww.com/CM9/B293]. To have a better understanding of the clustering result and their functions, the DEGs (log2(FC) >2 and adjust. \(p < 0.05\)) were explored [Supplementary Figure 2D, http://links.lww.com/CM9/B293]. Gene ontology results indicated that DEGs were tightly related to the regulation of acute inflammatory response, signaling receptor activator activity, protein processing, cell–cell junction organization, endopeptidase activity, and cell–cell junction [Supplementary Figure 2E, http://links.lww.com/CM9/B293]. Kyoto Encyclopedia of Genes and Genomes results showed DEGs were mainly enriched in cell adhesion molecules, tyrosine metabolism, glutathione metabolism, extracellular matrix-receptor interaction pathway, wingless/integrated signaling pathway, and interleukin-17 signaling pathway [Supplementary Figure 2F, http://links.lww.com/CM9/B293].

After performing LASSO-Cox regression analysis in the train cohort, a novel m7G-RPS was established to predict the clinical outcomes of PAAD patients (risk score = 0.862 4343001335988443571858 expression of AP00523.2) [Supplementary Figure 3A–C, http://links.lww.com/CM9/B293]. The patients were split into high-risk and low-risk groups, with the train cohort’s median risk score serving as the cutoff criterion [Supplementary Figure 4A, http://links.lww.com/CM9/B293]. Patients with high risk scores were more likely to die, according to the distributions of risk scores and survival status. [Supplementary Figure 4B, http://links.lww.com/CM9/B293]. As illustrated in Supplementary Figure 4C, http://links.lww.com/CM9/B293, the expression of AL139147.1, AC105942.1, LINC01089, and AP00523.2 was aligned with the coefficients in the calculation formula. Similarly, the prognosis of the high-risk group was significantly worse than that of the low-risk group (\(P < 0.05\)) [Supplementary Figure 4D, http://links.lww.com/CM9/B293]. m7G-RPS can serve as an independent prognostic index on the basis of univariate and multivariate cox regression analyses [Supplementary Figure 4E and F, http://links.lww.com/CM9/B293]. In addition, the area under the curve of the time-dependent ROC curve demonstrated that the risk score had a satisfied prognostic value [Supplementary Figure 4G, http://links.lww.com/CM9/B293]. Test 1, test2, and test3 cohorts (ie, internal and external verification cohorts) yielded similar results [Supplementary Figures 5–7, http://links.lww.com/CM9/B293].

Furthermore, the discrepancies of immune response and ICG expression in low-risk and high-risk subgroups were explored [Supplementary Figure 8, http://links.lww.com/CM9/B293]. In training, test1, and test2 cohorts, the immune response showed a poorer status in the high-risk subgroup [Supplementary Figure 8A–C, http://links.lww.com/CM9/B293]. In the following analysis of ICGs, all ICGs with significantly different expression in low-risk and high-risk subgroups were shown in Supplementary Figure 8D–F, [Supplementary Figure 8D–F, http://links.lww.com/CM9/B293]. Notably, BLTA, FG1L, CD160, CD40LG, and CD200R1 were reduced while TNFSF4, CD276, and LDHA showed an upregulation in train, test1, and test2 cohorts.
### Funding
This study was supported by the Leading Talent Team of Support Program for High-Level Talent's Innovation of Dalian in 2019 (No. 2019RD11) and Key Projects of Overseas Training Foundation of the Higher Education Institutions of Liaoning Province, China (No. 2020GJWZD004).

### Conflicts of interest
None.

### References
1. Liu SL, Li S, Guo YT, Zhou YP, Zhang ZD, Li S, et al. Establishment and application of an artificial intelligence diagnosis system for pancreatic cancer with a faster region-based convolutional neural network. Chin Med J 2019;132:2795–2803. doi: 10.1097/cm9.000000000000544.
2. 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;12:2432–2448. doi: 10.1080/21655979.2021.1933868.
3. Tomikawa C. 7-Methylguanosine modifications in transfer RNA (tRNA). Int J Mol Sci 2018;19:4080. doi: 10.3390/ijms19124080.

How to cite this article: Yuan Q, Ren J, Chen X, Dong Y, Shang D. Contributions and prognostic performances of m7G RNA regulators in pancreatic adenocarcinoma. Chin Med J 2022;135:2101–2103. doi: 10.1097/CM9.0000000000002179