N6-methyladenosine Methylation Related Immune Biomarkers Correlates With Clinicopathological Characteristics and Prognosis in Clear Cell Renal Cell Carcinoma

Hui Xiong  
Shandong University Affiliated Hospital: Shandong Provincial Hospital

Weiting Kang  
Shandong University Affiliated Hospital: Shandong Provincial Hospital

Qi Zhang  
Shandong University Affiliated Hospital: Shandong Provincial Hospital

Research Article

Keywords: clear cell renal cell carcinoma, N6-methyladenosine (m6A) methylation, Cox regression analysis, prognosis risk model

Posted Date: May 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-507507/v1

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

**Background:** This study aimed to explore N6-methyladenosine (m6A) methylation-related immune biomarkers and their clinical value in clear cell renal cell carcinoma (ccRCC).

**Methods:** The RNA-seq data and clinical phenotype of ccRCC were downloaded from TCGA database. Immune-related genes list was downloaded from InnateDB database. Correlation analysis, survival analysis, univariate and multivariate Cox regression analysis were used to investigate the prognostic independent m6A-related immune genes, followed by prognosis risk model establishment. Patients were divided into high/low risk groups, followed by survival analysis, clinical factors, immune checkpoint genes and gene set variation analysis in high-risk vs. low-risk group.

**Results:** Five prognostic independent m6A-related immune genes (PKHD1, IGF2BP3, RORA, FRK and MZF1) were identified. Low expression of PKHD1, RORA and FRK were associated with poor survival, while high expression of IGF2BP3 and MZF1 were associated with poor survival for ccRCC patients. Their expression showed correlations with multiple m6A genes. The risk model could stratify ccRCC patients into high/low risk group, and patients with high-risk were associated with short survival time. High-risk group had an high proportion of patients in tumor stage T3-T4, lymph node metastasis (N1) and distant metastasis (M1). Ten immune checkpoint genes were differentially expressed in high/low risk groups, such as PD1 and CTLA-4. The risk group could be an independent prognostic factor (HR=1.69, 95% CI 1.07-2.68, P=0.0246).

**Conclusion:** In this study, we developed a five genes risk model, which had independent prognostic value and associated with tumor stage, pathologic T/N/M and immune checkpoint expression in ccRCC.

Background

Renal cell carcinoma (RCC) is one of the most common malignancies of the human urinary system that originates from renal tubular epithelial cells, accounting for more than 80% of primary renal malignancies [1]. In United States, there were about 73,820 new cases diagnosed, and 14,770 deaths in 2019 [2]. Clear cell renal cell carcinoma (ccRCC) is the most common and most malignant pathological type of RCC, accounting for approximately 70%-80% of all RCC [3]. Due to its insidious onset, about 30% of ccRCC is diagnosed in advanced stage with a 5-year survival rate of about 11.7% [4]. The surgery is the conventional treatment for early stage ccRCC, however 10–20% of patients will relapse and metastasize after treatment [5]. Therefore, prolonging the survival time of patients with advanced renal cancer has become the mainstream of current study.

N6-methyladenosine (m6A) methylation is regarded as the most common internal modification of RNA in eukaryotic cells that has been found to implicate various fundamental cellular functions, including pre-mRNA splicing, translation regulation, 3’-end and microRNA processing, and nuclear transport, indicating associations with diverse pathophysiologies [6, 7]. For example, METTL3, a m6A methyltransferase installing m6A modification on target RNAs, has been demonstrated to accelerate tumor formation and progression of bladder cancer in a m6A-dependent manner by promoting the maturation of pri-microRNA221/222 [8]. Chen et al. identified the risk signature involving m6A methylation genes FTO, YTHDC1 and WTAP that taken part in the malignant progression of bladder cancer, and this risk signature showed independent prognostic value for patients with bladder cancer [9].

Additionally, it has been reported that m6A modification involves the regulation of the immune response and T cell homeostasis [10]. For example, the m6A modification was found to initiate naive T cells re-programming for proliferation and differentiation by inducing the mRNA degradation of suppressor of cytokine signaling genes [11]. Han et al. showed that loss of YTHDF1, a m6A effector, the mice displayed increased antigen-specific CD8 + T cell anti-tumor response and enhanced clinical responses to PD-L1 blockade, and YTHDF1 lacking in dendritic cells could elevate the cross-presentation of tumor neoantigens, which were vital for anti-tumor immune responses [12]. Su et al. suggested that inhibiting of m6A
demethylase FTO could promote the sensitizing of leukemia cells T cell cytotoxicity, and prevent the immune evasion triggered by hypomethylation agent [13]. However, the role of m6A modification related immune genes in ccRCC has not been investigated.

In this study, we identified several independent prognostic immune genes that showed strong expression correlations with m6A modification genes (methyltransferases, demethylases and effectors) based on the expression data and clinical phenotype of ccRCC in The Cancer Genome Atlas (TCGA) database. Then, a prognosis risk model was established on the basis of these genes, and the clinical value and predictive performance were further investigated. This study will provide potential therapeutic targets and prognostic markers for the treatment of ccRCC.

**Methods And Methods**

**Data acquisition and data preprocessing**

The RNA-seq FPKM expression data and clinical phenotype data of 602 kidney renal clear cell carcinoma samples in TCGA database (TCGA-KIRC) were downloaded, of which 531 samples were tumor samples. For genes annotation, the Ensembl_ID were converted into Symbol_ID based on the reference genome hg38 (version 22) in GENCODE database [14]. The mean value was selected when different Ensembl_ID mapping to the same Symbol_ID. Genes with expression value of 0 in more than 80% samples were removed. The analyses in this study were performed as the procedure in Figure S1.

**Identification of m6A-related immune genes**

The expression data of 21 m6A-related genes (methyltransferases: METTL3, METTL14, METTL15, WTAP, VIRMA, RBM15, RBM15B, KIAA1429, ZC3H13; demethylases: FTO, ALKBH5; effectors: RBMX, YTHDC1, YTHDC2, IGF2BP1, IGF2BP2, IGF2BP3, YTHDF1, YTHDF2, YTHDF3, HNRNPA2B1, HNRNPC) were retrieved from the KIRC tumor samples. In addition, the immune-related genes list that recorded in the ImmPort were downloaded from InnateDB database [15], then the expression data of these immune-related genes were also retrieved from the KIRC tumor samples. Then, correlation analysis was performed to calculate the Pearson correlation coefficient utilizing cor test in R. The m6A-related immune genes were identified by cut-off value as follows: absolute value of correlation coefficient > 0.6 and P < 0.001.

**Prognostic independent m6A-related immune genes**

After filtering samples with survival time less than 30 days, there remained 518 samples. Samples were divided into high and low expression group on the basis of median expression value of each m6A-related immune gene, followed by survival analysis utilizing Kaplan-Meier (KM) curve and log-rank test, and genes with P < 0.001 in survival analysis were selected. We randomly grouped patients into training-set (n = 259) and valid-set (n = 259) in a ratio of 5:5. In training-set, univariate Cox regression analysis was applied to explore genes which was associated with patient survival, and genes with log-rank P < 0.001 were included in multivariate Cox regression analysis. Prognostic independent m6A-related immune genes were screened by log-rank P < 0.001.

**Prognostic risk model**

Prognostic independent m6A-related immune genes (genes with log-rank P < 0.001 in multivariate Cox regression analysis) were utilized to establish prognostic risk model with the following formula: Risk score = Σ Coef genes × Exp genes. The Coef in formula refers to the coefficient β in multivariate Cox regression, while Exp refers to the expression value of each gene. Risk score was calculated for all samples, and patients were categorized into different risk groups according to the median of risk score, followed by survival analysis. This analysis was carried out in training-set (n = 259), valid-set (n = 259) and the whole set (n = 518).

**Clinical factors and immune checkpoint genes in high and low risk group**
Clinical factors, including pathologic T/N/M, tumor stage, age and gender were compared between high and low risk groups utilizing ggstatsplot (version: 0.5.0) in R package, and chi-square test was used to calculate the P value. Similarly, the difference on expression level of immune checkpoint genes were also explored, including PD1 (PDCD1), PD-L1 (CD274), CTLA-4, CD278 (ICOS), Tim3 (HAVCR2), LAG3, CD73, CD47, BTLA, TIGIT, myd1 (SIRPA), OX40 (TNFRSF4), 4-1BB (TNFRSF9) and B7-H4 (VTCN1).

**Gene set variation analysis (GSVA)**

In order to explore the differences on KEGG pathways between the high and low risk groups, GSVA was performed [16]. In brief, on the basis of the enrichment background (c2.cp.kegg.v7.2.symbols.gmt) in MSigDB v7.2 database [17], the enrichment scores of KEGG pathways in all samples were calculated to obtain a scoring matrix using GSVA (version: 1.36.2) in R package. Then, differential analysis was conducted for these KEGG pathways in high vs. low-risk groups utilizing Limma in R package. The significant results were selected with P < 0.05.

**Evaluation of prognostic risk model**

Univariate Cox regression analysis was carried out to screen factors (clinical factors and risk score) that impacted on survival of patients, in which factors with log-rank P < 0.05 were included in multivariate Cox regression analysis. Factors with log-rank P < 0.05 were considered to have independent prognostic value. Additionally, Normgram was utilized to evaluate the independent prognostic factors.

**Results**

**Identification of prognostic independent m6A-related immune genes**

After data preprocessing, we obtained expression data of 3788 immune genes from TCGA dataset. Then, the correlation analysis was carried out on the expression level between 3788 immune genes and the 21 m6A-related genes, and 856 co-expression pairs involving 458 immune genes and 19 m6A-related genes were obtained with the cut-off value of absolute value of correlation coefficient > 0.6 and P < 0.001.

Survival analysis was then performed for these 458 immune genes, and K-M curves showed that 195 immune genes were associated with survival of ccRCC patients. Univariate Cox regression analysis for these 195 immune genes revealed that 85 immune genes had significant impact on prognosis of ccRCC patients (Table S1). These 85 immune genes were then included in multivariate Cox regression analysis, and five genes were found to be independent prognosis immune genes, including PKHD1, IGF2BP3, RORA, FRK and MZF1 (Fig. 1A, Table S2). Figure 1B showed the correlation heatmap of the five immune genes with m6A-related genes. It could be seen that MZF1 showed strong positive correlation with METTL3, while showed strong negative correlation with KIAA1429 and YTHDF3. PKHD1 was negatively correlated with IGF2BP3 and IGF2BP2. Therefore, these five genes were considered as prognostic independent m6A-related immune genes. Survival analysis indicated that the low expression of PKHD1, RORA and FRK were associated with poor survival for ccRCC patients. While patients with low expression of IGF2BP3 and MZF1 showed long survival time than patients with high expression level (Fig. 2).

**Prognosis risk model**

Prognosis risk model based on these five genes was established, and risk score was calculated for all samples. Patients were categorized into different risk groups according to the median of risk score, followed by survival analysis. The K-M curve displayed that patients in high-risk group were associated with worse survival than patients in low-risk group in both training-set (n = 259), valid-set (n = 259) and the whole set (n = 518) (Fig. 3).

**Clinical factors and immune checkpoint genes in high and low risk group**
We further compared clinical factors between high-risk and low-risk groups. Except for age and gender, pathologic T, pathologic N, pathologic M and tumor stage were significant associated with risk score. High-risk group had a high proportion of patients in tumor stage III-IV and a high proportion of patients with invasiveness (pathologic T3 and T4), lymph node metastasis (N1) and distant metastasis (M1) (Fig. 4). In addition, the expression of multiple immune checkpoint genes showed significant difference between high-risk and low-risk groups (Fig. 5). For example, the expression of well-known immune checkpoint genes PD1 and CTLA-4 were significant higher in high-risk group.

**KEGG pathways in high and low risk group**

GSVA analysis found that 11 KEGG pathways were different between high-risk and low-risk groups, such as cell cycle, P53 signaling pathway and various metabolism pathways, including histidine/ butanoate/ propanoate/ ascorbate and aldarate metabolism (Figure S2, Table 1).

<table>
<thead>
<tr>
<th>Pathway</th>
<th>logFC</th>
<th>AveExpr</th>
<th>t</th>
<th>P.Value</th>
<th>adj.P.Val</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG_Cell cycle</td>
<td>0.035095001</td>
<td>0.340633</td>
<td>4.613654</td>
<td>4.98E-06</td>
<td>0.0092606</td>
<td>3.244114</td>
</tr>
<tr>
<td>KEGG_P53 signaling pathway</td>
<td>0.030019521</td>
<td>0.3599513</td>
<td>4.359866</td>
<td>1.57E-05</td>
<td>0.014572</td>
<td>2.152439</td>
</tr>
<tr>
<td>KEGG_Histidine metabolish</td>
<td>0.034572315</td>
<td>0.412188</td>
<td>3.812736</td>
<td>0.00154</td>
<td>0.00953346</td>
<td>-0.00446</td>
</tr>
<tr>
<td>KEGG_Oocyte meiosis</td>
<td>0.022081193</td>
<td>0.335807</td>
<td>3.530919</td>
<td>0.00451</td>
<td>0.01778489</td>
<td>-1.00885</td>
</tr>
<tr>
<td>KEGG_Pathogenic escherichia coli infection</td>
<td>0.028072745</td>
<td>0.3945073</td>
<td>3.51484</td>
<td>0.00478</td>
<td>0.01778489</td>
<td>-1.06394</td>
</tr>
<tr>
<td>KEGG_Butanoate metabolism</td>
<td>0.037942885</td>
<td>0.434312</td>
<td>3.378098</td>
<td>0.00784</td>
<td>0.02430781</td>
<td>-1.52277</td>
</tr>
<tr>
<td>KEGG_Steroid hormone biosynthesis</td>
<td>0.024499612</td>
<td>0.3601932</td>
<td>3.237615</td>
<td>0.001282</td>
<td>0.02993686</td>
<td>-1.97599</td>
</tr>
<tr>
<td>KEGG_Ascorbate and aldarate metabolism</td>
<td>0.038805757</td>
<td>0.4794297</td>
<td>3.236242</td>
<td>0.001288</td>
<td>0.02993686</td>
<td>-1.98033</td>
</tr>
<tr>
<td>KEGG_Valine leucine and isoleucine degradation</td>
<td>0.040939103</td>
<td>0.4557947</td>
<td>3.11005</td>
<td>0.001972</td>
<td>0.04075974</td>
<td>-2.3715</td>
</tr>
<tr>
<td>KEGG_Drug metabolism cytochrome P450</td>
<td>0.024330769</td>
<td>0.369035</td>
<td>3.020577</td>
<td>0.002646</td>
<td>0.04921436</td>
<td>-2.63977</td>
</tr>
<tr>
<td>KEGG_Propanoate metabolism</td>
<td>0.036001688</td>
<td>0.4484454</td>
<td>2.991018</td>
<td>0.002911</td>
<td>0.04922557</td>
<td>-2.72673</td>
</tr>
</tbody>
</table>

**Evaluation of prognostic risk model**

In univariate Cox regression analysis, pathologic T, pathologic N, pathologic M and risk group were associated with prognosis of ccRCC patients. These factors were included in multivariate Cox regression analysis, the results indicated that pathologic M (hazard ratio, HR = 2.50, 95% CI 1.53–4.08, P = 0.0002), pathologic T (HR = 2.18, 95% CI 1.37–3.46, P = 0.0010) and risk group (HR = 1.69, 95% CI 1.07–2.68, P = 0.0246) were independent prognostic factors (Fig. 6A, Table 2). These three independent prognostic factors were used to establish a nomogram, and it suggested that the nomogram could accurately predict the 1-, 2-, and 3-year survival probability for ccRCC patients (Fig. 6B).
Table 2
Univariate and multivariate Cox regression analysis for clinical factors.

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>Univariables cox</th>
<th>Multi-variables cox</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>lower.95</td>
</tr>
<tr>
<td>pathologic_M</td>
<td>4.228869</td>
<td>2.731769</td>
</tr>
<tr>
<td>pathologic_T</td>
<td>3.165791</td>
<td>2.075339</td>
</tr>
<tr>
<td>RiskGroup</td>
<td>2.523951</td>
<td>1.646848</td>
</tr>
<tr>
<td>pathologic_N</td>
<td>3.517107</td>
<td>1.865563</td>
</tr>
<tr>
<td>age</td>
<td>0.756154</td>
<td>0.494296</td>
</tr>
<tr>
<td>gender</td>
<td>1.040269</td>
<td>0.681729</td>
</tr>
</tbody>
</table>

HR, hazard ratio.

Discussion

Studies have demonstrated that m6A methylation implicates various fundamental cellular functions, and play crucial roles in the tumor formation and progression in different cancers [18, 19]. In addition, m6A modification has also been reported to act crucial roles in mediating the diversity and complexity of tumor immune microenvironment [20, 21]. In this study, we identified five prognostic independent m6A-related immune genes (PKHD1, IGF2BP3, RORA, MZF1 and FRK) in ccRCC based on the expression data in TCGA. Low expression of PKHD1, RORA and FRK were associated with poor survival, while high expression of IGF2BP3 and MZF1 were associated with poor survival for ccRCC patients. Their expression showed correlations with various m6A methylation genes.

IGF2BP3 (insulin-like growth factor 2 mRNA-binding protein 3), a m6A reader, IGF2BP3 had been reported to drive malignancy progression of ccRCC by stabilizing CDKN2B-AS1 which in turn to activate the expression of Ndc80 kinetochore complex component 2 [22]. Sun et al. suggested that IGF2BP3 could promote the G1/S transition and cell proliferation in ccRCC, and its high expression showed correlations with worse prognosis [23]. RORA encodes RAR related orphan receptor A, and the inhibition of RORA has been found to contribute to tumor proliferation and metastasis in gastric cancer [24] and endometrial cancer [25]. MZF1 (myeloid zinc finger 1) is an oncogenic transcription factor, which plays important roles in driving cancer malignancy and epithelial-mesenchymal transition, has been considered as a regulator of tumor invasion and metastasis [19]. FRK is a fyn related Src family tyrosine kinase. The inhibition of FRL was reported to implicated in promoting tumor proliferation and metastasis in RCC [26]. Jing et al. indicated that FRK could repress cell proliferation in ccRCC by the phosphorylation of PTEN, and its expression was decreased in ccRCC and associated with worse outcomes [27]. These findings were consistent with our results. Therefore, we suggested that these five genes were therapeutic targets and potential prognostic biomarkers in ccRCC.

Based on these five genes, a prognosis risk model was established, and the risk score could stratify ccRCC patients into high-risk and low risk group. Patients with high-risk were associated with short survival time than patients with low-risk. This further confirmed the prognosis value of the five genes-based risk model. Zhao et al. suggested that three-m6A related gene based prognostic risk model could precisely indicate the survival of patients in ccRCC [28]. Additionally, in order to investigate the possible reasons of worse outcomes for patients with high-risk, the differences on clinical factors and immune checkpoint genes were compared between high-risk and low-risk group. The results suggested that high-risk group had an high proportion of patients in tumor stage I-III and an high proportion of patients with invasiveness, lymph node metastasis and distant metastasis. In addition, ten immune checkpoint genes were found to show differential expression in high-risk and low-risk group, for example, PD1 and CTLA4 were highly expressed in high-risk in comparison with that of
low-risk. The blockade of PD-1 had been proved to be a therapeutic target in the immunotherapy of ccRCC patients, suggesting the associations between its expression and prognosis [29, 30]. It had been demonstrated that CTLA4 inhibitors could lengthen the overall survival in various tumors. CTLA4 was found to be elevated in ccRCC and its high expression showed strong correlations with tumor progression and worse outcomes [31]. Xiao et al. showed that high level of CTLA4 was associated with a obvious decreased survival time, pathologic stage and local recurrence, as well as negatively related to tumor purity [32]. These results might explain why patients with high risk had poor survival.

**Conclusion**

We identified five prognostic independent m6A-related immune genes (PKHD1, IGF2BP3, RORA, MZF1 and FRK) in ccRCC. The prognosis risk model-based on these five genes could predict survival for ccRCC patients, patients with high-risk scores were associated with short survival time. The risk score was closely associated with tumor stage and pathologic T/N/M and immune checkpoint expression, and the risk score had independent prognostic value for patients with ccRCC.

**Declarations**

**Acknowledgement**

The authors express thanks to the National Youth Fund of Natural Science Foundation (Grant Nos. 81902572) and The Shandong Provincial Plan for Development of Research in Key Areas (Grant Nos. 2019GSF107050).

**Funding**

The authors express thanks to the National Youth Fund of Natural Science Foundation (Grant Nos. 81902572) and The Shandong Provincial Plan for Development of Research in Key Areas (2019GSF107050).

**Declaration of Competing Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethics statement**

Not applicable

**Consent for publication**

Not applicable.

**Availability of data and materials**

The data that support the findings of this study are available on request from the corresponding author.

**Authors’ contributions**

Conception and design of the research and drafting the manuscript: Qi Zhang

Analysis and interpretation of data: Hui Xiong

Statistical analysis: Weiting Kang

Revision of manuscript for important intellectual content: Hui Xiong and Qi Zhang
All authors have read and approved the manuscript.

References


**Figures**
Figure 1

Five independent prognostic m6A-related immune genes. A, Forest plot shows the results of multivariate Cox regression analysis for the five genes; B, the correlation heatmap shows the correlations between the five immune genes with m6A-related genes. Blue represents negative correlations, and red represent positive correlations; * represent statistical significance.
Figure 2

Survival analysis of the five genes. Kaplan-Meier survival curves shows the prognosis value of the PKHD1, IGF2BP3, FRK, RORA and MZF1.

Figure 3

Training-set  Valid-set  Whole set
Survival analysis in high-risk and low-risk group. Survival curves shows the prognosis value of the risk score calculated by the prognosis risk model in training-set, valid-set and whole set.

Figure 4

Clinical factors in high and low risk group. Histogram shows the proportion of patients in high and low risk group by age, gender, tumor stage and pathologic T, pathologic N, pathologic M.
Figure 5

Immune checkpoint genes in high and low risk group. Box plots shows the difference on the expression of immune checkpoint genes in high and low risk group. * represent statistical significance.

Figure 6

Evaluation of prognostic risk model. A, Forest plot shows the results of multivariate Cox regression analysis for clinical factors and risk group; B, Nomogram established by the three independent prognostic factors for predicting 1-, 2-, and 3-year survival probability of ccRCC patients.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• FigureS1.tif
• FigureS2.tif
• SupplementalTables.doc