Prognostic Value of DNA Methylation-Driven Genes in Clear Cell Renal Cell Carcinoma: A Study Based on Methylation and Transcriptome Analyses

Maolin Hu,1,2 Jiangling Xie,3 Huiming Hou,2 Ming Liu,2 and Jianye Wang1,2

1Peking University Fifth School of Clinical Medicine, Beijing 100730, China
2Department of Urology, Beijing Hospital, National Center of Gerontology, Beijing 100730, China
3Department of Urology, Qianjiang Central Hospital of Chongqing Municipality, Chongqing 409000, China

Correspondence should be addressed to Ming Liu; liumingbjh@126.com and Jianye Wang; wangjianyebjyy@126.com

Received 20 May 2020; Revised 15 June 2020; Accepted 25 June 2020; Published 11 July 2020

Academic Editor: Mingjun Shi

Copyright © 2020 Maolin Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Few previous studies have comprehensively explored the level of DNA methylation and gene expression in ccRCC. The purpose of this study was to identify the key clear cell renal cell carcinoma- (ccRCC-) related DNA methylation-driven genes (MDG) and to build a prognostic model based on the level of DNA methylation. Methods. RNA-seq transcriptome data and DNA methylation data were obtained from The Cancer Genome Atlas. Based on the MethylMix algorithm, we obtain ccRCC-related MDG. The univariate and multivariate Cox regression analyses were employed to investigate the correlation between patient overall survival and the methylation level of each MDG. Finally, a prognosis risk score was established based on a linear combination of the regression coefficient derived from the multivariate Cox regression model ($\beta$) multiplied with the methylation level of the gene.

Results. 19 ccRCC-related MDG were identified. Three MDG (NCKAP1L, EVI2A, and BATF) were further screened and integrated into a prognostic risk score model, risk score = (3.710 * methylation level of NCKAP1L) + (−3.892 * methylation level of EVI2A) + (−3.907 * methylation level of BATF). The risk model was independent from conventional clinical characteristics as a prognostic factor for ccRCC (HR = 1.221, 95% confidence interval: 1.063–1.402, and $P = 0.005$). The joint survival analysis showed that the gene expression and methylation levels of the prognostic genes EVI2A and BATF were significantly related with prognosis. Conclusion. This study provided an important bioinformatics foundation for in-depth studies of ccRCC DNA methylation.

1. Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common pathological type of renal cancer [1, 2]. Approximately 350,000 new patients with RCC are confirmed globally every year, and RCC causes more than 140,000 deaths each year [3]. ccRCC has complex biological characteristics and is not sensitive to radiotherapy and chemotherapy; radical or partial nephrectomy is the main method for treating RCC [4, 5]. Although the surgical effect is definite, 20%–40% of patients experience local recurrence or distant metastasis after surgery [6], and the mortality rate is very high for advanced renal cell carcinoma [7]. In the last ten years, with the development of molecular biology, several ccRCC driver genes such as von Hippel–Lindau (VHL) and PBRM1 have been discovered [8, 9], and targeted drugs, such as tyrosine kinase inhibitors (TKIs) [10], have been developed to treat these advanced ccRCC patients; however, not all of these patients can benefit from these drugs. Therefore, the identification of new ccRCC-related driver genes and the establishment of risk models through bioinformatics analysis are very important for patient prognosis evaluation.

Epigenetic changes are thought to be closely related to cancer progression, and aberrant DNA methylation is one of the most crucial and routine epigenetic modifications [11]. DNA methylation is an epigenetic modification that plays a crucial role in regulating the growth, development, gene expression pattern, and stability of the genome without
changing the DNA sequence [12, 13]. Many studies in recent years have found that abnormal DNA methylation is closely related to tumorigenesis, development, and cell canceration [14–16]. In particular, alterations in DNA methylation can supply crucial information for the early diagnosis and prognosis of cancer [17–20].

Few previous studies have comprehensively explored the level of DNA methylation and gene expression in ccRCC. The purpose of this study was to identify the key ccRCC-related methylation-driven genes and to build a prognostic model based on the level of DNA methylation.

2. Materials and Methods

2.1. Data Acquisition and Integrative Analysis. In this study, RNA-seq transcriptome data and DNA methylation data were obtained from The Cancer Genome Atlas (TCGA) website (https://portal.gdc.cancer.gov/repository). Of them, the DNA methylation data was using the Illumina Infinium HumanMethylation450 platform and 27 platform, and beta values, ranged from 0 to 1, were quantified to indicate the levels of DNA methylation. Based on R software and packages [21], we analyzed the above data to obtain differentially methylated genes and differentially expressed genes. In addition, clinicopathological data including survival status, survival time, age, sex, the International Society of Urological Pathology (ISUP) grade, and American Joint Committee on Cancer (AJCC) stage were obtained from TCGA. Based on the MethylMix package [22], we performed an analysis by combining the differentially methylated genes and differentially expressed genes. MethylMix is an algorithm for identifying the hypermethylation and hypomethylation of genes in diseases [23]. MethylMix is based on a β-mixture model to identify the methylation status and compares this status with the normal DNA methylation state. The correlation was computed between the gene methylation level and the gene expression level. Finally, the exported result of MethylMix is methylation-driven genes (MDG).

2.2. MDG Set Enrichment Analysis. Gene functional enrichment analyses were conducted to discover the main biological characteristics of the MDG, including Gene Ontology (GO) analyses with molecular function, biological process, and cellular component analyses [24]. In our study, the biological process category was selected for GO analysis through the enrichGO function in the clusterProfiler package (version 3.6) and the original database was acquired from the "org.Hs.s.eg.db" package [25].

2.3. Definition of the MDG-Related Prognostic Model. We randomly divided the tumor samples obtained from TCGA into two cohorts: a training cohort (244 patients, Table S1) and a validation cohort (242 patients, Table S2). We used the training cohort to build the Cox regression risk model, and the validation cohort was adopted to verify the performance of the model. First, we performed univariate Cox regression analysis to identify six methylation-driven genes related to overall survival in patients by considering $P$ value < 0.01 as significant. Then, multivariate Cox regression analysis was used to build a prognostic risk model. The risk score was figured out using the regression coefficients ($\beta$) from the multivariate Cox regression model to weight the methylation values of the selected genes. Risk score = $\beta$gene (1) × methylation level of gene (1) + $\beta$gene (2) × methylation level of gene (2) + $\ldots$ + $\beta$gene ($n$) × methylation level of gene ($n$). According to the median risk score, the

### Table 1: DNA methylation-driven genes in ccRCC.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Normal me</th>
<th>Tumor me</th>
<th>logFC</th>
<th>$P$ value</th>
<th>Cor</th>
<th>Cor $P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1orf116</td>
<td>0.647716</td>
<td>0.896463</td>
<td>0.468882</td>
<td>3.76E-109</td>
<td>-0.6038</td>
<td>4.66E-54</td>
</tr>
<tr>
<td>ALDOC</td>
<td>0.315874</td>
<td>0.201405</td>
<td>-0.64925</td>
<td>1.32E-72</td>
<td>-0.44678</td>
<td>2.04E-27</td>
</tr>
<tr>
<td>NCKAP1L</td>
<td>0.822054</td>
<td>0.656117</td>
<td>-0.32528</td>
<td>1.06E-64</td>
<td>-0.35805</td>
<td>1.66E-17</td>
</tr>
<tr>
<td>EVL2A</td>
<td>0.854583</td>
<td>0.71327</td>
<td>-0.26077</td>
<td>1.95E-62</td>
<td>-0.5227</td>
<td>1.45E-38</td>
</tr>
<tr>
<td>CD2</td>
<td>0.74397</td>
<td>0.633319</td>
<td>-0.23231</td>
<td>1.57E-54</td>
<td>-0.46794</td>
<td>3.01E-30</td>
</tr>
<tr>
<td>BATF</td>
<td>0.696094</td>
<td>0.535991</td>
<td>-0.37707</td>
<td>3.59E-48</td>
<td>-0.43104</td>
<td>1.96E-25</td>
</tr>
<tr>
<td>CD96</td>
<td>0.703644</td>
<td>0.600698</td>
<td>-0.22821</td>
<td>2.19E-45</td>
<td>-0.35773</td>
<td>1.79E-17</td>
</tr>
<tr>
<td>RAB25</td>
<td>0.498379</td>
<td>0.687697</td>
<td>0.46453</td>
<td>2.40E-38</td>
<td>-0.38485</td>
<td>3.44E-20</td>
</tr>
<tr>
<td>IL20RB</td>
<td>0.886919</td>
<td>0.731896</td>
<td>-0.27716</td>
<td>1.13E-33</td>
<td>-0.50225</td>
<td>2.79E-35</td>
</tr>
<tr>
<td>TMEM173</td>
<td>0.514377</td>
<td>0.327888</td>
<td>-0.64963</td>
<td>4.07E-30</td>
<td>-0.36018</td>
<td>1.04E-17</td>
</tr>
<tr>
<td>CTSZ</td>
<td>0.478134</td>
<td>0.400661</td>
<td>-0.25503</td>
<td>7.68E-30</td>
<td>-0.32922</td>
<td>6.89E-15</td>
</tr>
<tr>
<td>MASP1</td>
<td>0.600411</td>
<td>0.503547</td>
<td>-0.25382</td>
<td>8.74E-26</td>
<td>-0.40218</td>
<td>4.61E-22</td>
</tr>
<tr>
<td>LAMA4</td>
<td>0.426073</td>
<td>0.346439</td>
<td>-0.2985</td>
<td>4.99E-22</td>
<td>-0.36743</td>
<td>2.04E-18</td>
</tr>
<tr>
<td>LTA</td>
<td>0.859977</td>
<td>0.795969</td>
<td>-0.11159</td>
<td>1.31E-18</td>
<td>-0.56031</td>
<td>3.21E-45</td>
</tr>
<tr>
<td>SLC9B2</td>
<td>0.43967</td>
<td>0.392664</td>
<td>-0.16312</td>
<td>3.45E-13</td>
<td>-0.38856</td>
<td>1.40E-20</td>
</tr>
<tr>
<td>FAM107A</td>
<td>0.465871</td>
<td>0.414634</td>
<td>-0.16809</td>
<td>1.77E-11</td>
<td>-0.32415</td>
<td>1.86E-14</td>
</tr>
<tr>
<td>SLC5A12</td>
<td>0.705801</td>
<td>0.634831</td>
<td>-0.15289</td>
<td>1.16E-08</td>
<td>-0.33355</td>
<td>2.90E-15</td>
</tr>
<tr>
<td>AL592211</td>
<td>0.44045</td>
<td>0.575433</td>
<td>0.38567</td>
<td>0.002048</td>
<td>-0.32193</td>
<td>2.87E-14</td>
</tr>
<tr>
<td>LRMP</td>
<td>0.534103</td>
<td>0.516217</td>
<td>-0.04914</td>
<td>0.039959</td>
<td>-0.33687</td>
<td>1.48E-15</td>
</tr>
</tbody>
</table>
Figure 1: Heat maps of ccRCC-related aberrant methylation-driven genes. (a) The hierarchical clustering heat map of ccRCC-specific methylation-driven mRNAs. (b) The hierarchical clustering heat map of the methylation level of ccRCC-specific methylation-driven genes. In the figure, red represents highly methylated genes and green represents low methylated genes between ccRCC and normal tissues.
training cohort was separated into the low-risk and high-risk groups. To assess the predictive power of the risk score, we performed receiver operating characteristic (ROC) curve analysis. The same method was adopted in the testing cohort to test the performance of the model.

2.4. Independence of the Risk Model from Traditional Clinical Features and Building a Nomogram. To verify whether the risk score was independent of other clinical variables in ccRCC patients (including age, sex, ISUP grade, and AJCC stage), the univariate and multivariate Cox regression analyses were performed on the entire TCGA cohort. According to results of multivariate Cox regression analysis, we constructed a nomogram with R package “rms” to assess the overall survival for ccRCC. The predictive performance was evaluated by the C index and calibration curve.

2.5. Joint Survival Analysis of the Gene Expression Levels and Methylation Levels of the MDG. To further identify the key genes related to the prognosis of ccRCC patients, based on the survival R package, a joint survival analysis was performed by combining the methylation levels of MDG with the corresponding gene expression levels.

2.6. Statistical Analysis. All statistical analyses were performed using R software. \( P < 0.05 \) was considered statistically significant. The MethylMix package was used to identify DNA methylation-driven genes. We performed the univariate and multivariate Cox regression analyses to determine whether the risk score had prognostic value independent of various clinicopathological characteristics. We conducted the log-rank test and Kaplan-Meier survival analysis to evaluate the predictive ability of the risk model.

3. Results

3.1. Data Analysis and Acquisition of DMG in ccRCC. All data were acquired from TCGA. The methylation data were downloaded from 542 cancer tissues and 357 noncancer tissues. In addition, the mRNA expression data were retrieved from 611 samples, consisting of 72 normal samples and 539 cancer samples. On the basis of the LIMMA software
Figure 4: Continued.
package, we analyzed the downloaded data to obtain differentially methylated genes and differentially expression genes. We performed an integrative analysis via the R package MethylMix, and the analysis required three datasets, including normal DNA methylation data, cancer DNA methylation data, and matched gene expression data (Tables S3, S4, and S5). $|\log FC| > 0$, $P < 0.05$, and $|\text{Cor}| > 0.3$ were adopted for screening MDG. Finally, 19 MDG were obtained (Table 1). Heat maps of the ccRCC-related aberrant MDG are shown in Figure 1.

### 3.2. MDG Set Enrichment Analysis in ccRCC

Gene Ontology analysis showed that MDG were mainly enriched in the regulation of interferon-gamma production, interferon-gamma production, regulation of the inflammatory response to antigenic stimuli, lymphocyte-mediated immunity, myeloid dendritic cell activation, regulation of lymphocyte-mediated immunity, negative regulation of interferon-gamma production, and negative regulation of lymphocyte-mediated immunity ($P < 0.05$, Figure 2).

Above MDG set enrichment analysis revealed that MDG were significantly linked to inflammatory response regulation and immune regulation.

### 3.3. Construction of the MDG-Related Prognostic Model

The training cohort was used to construct the risk model. Firstly, we conducted a univariate Cox regression analysis of the methylation level of nineteen MDG. We identified six genes (NCKAP1L, EVI2A, BATF, CD96, IL20RB, and CTSZ) related to overall survival in the training cohort by considering $P$ value $< 0.01$ as significant (Figure 3(a), $P$ value $< 0.01$). To better predict the relationships between the methylation levels of the methylation-driven genes and overall survival of ccRCC, we further analyzed them by stepwise multivariate Cox regression analysis. Finally, three genes were selected to build a predictive model (Figure 3(b), $P$ value $< 0.05$). The risk score was figured out using a linear combination of the methylation levels of the three selected MDG weighted by their specific regression coefficients ($\beta$). Risk score = $(3.710 \times \text{methylation level of NCKAP1L}) + (−3.892 \times \text{methylation level of EVI2A}) + (−3.907 \times \text{methylation level of BATF})$.

On the basis of the median risk score, the patients in the training cohort were classified as a high-risk group and a low-risk group. The risk score distribution of the patients according to the prognostic model is shown in Figure 4(a). Survival status scatter plots for the patients according to the prognostic model are shown in Figure 4(b), which shows that the high-risk subgroup contained a higher number of patients who died than the low-risk subgroup. We observed a significant difference in overall survival between the two groups (Figure 4(c), $P < 0.0001$), and the AUCs at three and five years were 0.719 and 0.689, respectively (Figures 4(d) and 4(e)).

### 3.4. Testing of the Methylation-Driven Gene-Related Prognostic Model

To confirm the performance of the model, the testing cohort was analyzed. First, we used the selected methylation-driven genes (NCKAP1L, EVI2A, and BATF) to compute the risk score of each patient in the validation cohort. Similar results were observed in the testing cohort, and the risk score distribution of the patients according to the prognostic model is shown in Figure 5(a). Survival status scatter plots for the patients according to the prognostic model are shown in Figure 5(b). Significant survival differences were observed in the testing cohort (Figure 5(c), $P < 0.0001$). The AUCs at three and five years were 0.674 and 0.659, respectively (Figures 5(d) and 5(e)).

### 3.5. Independence of the Prognostic Model from Other Clinical Characteristics and Building a Nomogram Based on the Risk Score and Other Clinical Characteristics

We conducted the univariate and multivariate Cox regression analyses to determine whether the risk score was a prognostic factor of ccRCC independent from traditional clinical features (including age,
Figure 5: Continued.
sex, ISUP grade, and AJCC stage). The results revealed that the risk score was independent of conventional clinical features in the entire TCGA cohort (Figure 6), with a hazard ratio (HR) of 1.221 (95% confidence interval: 1.063–1.402; \(P = 0.005\)).

Then, a nomogram predicting the overall survival in ccRCC was constructed based on the risk score and clinical characteristics (Figure 7(a)). The C index was 0.776. The calibration curve showed that the nomogram performed well (Figures 7(b) and 7(c)). Both the C index and the calibration curve suggested a good predictive performance.

### 3.6. MethylMix Model of Three Selected DNA MDG.

The MethylMix model of the three selected DNA MDG is shown in Figure 8. The correlation between DNA methylation and gene expression is visualized in Figures 9(a)–9(c), Cor<−0.3.

### 3.7. Joint Prognostic Assessment of MDG in ccRCC.

To further identify the key genes related to the prognosis of ccRCC patients, a joint survival analysis was conducted by combining the methylation levels of the MDG with the corresponding gene expression levels. The results are shown in Figures 9(d) and 9(e). The gene expression and methylation levels of the prognostic genes EVI2A and BATF were significantly related to prognosis.

### 4. Discussion

ccRCC is the most common pathological type of kidney cancer [26]. In-depth molecular pathogenesis research and the early detection of tumor prognostic biomarkers and specific driven-genes may have great significance for improving the prognosis of patients [27–29]. In recent years, with the deepening of epigenetic research, the epigenetic regulatory mechanisms of renal cancer, especially DNA methylation, are often used to predict the survival time of renal cancer, including DAB2IP [30], RCVRN [31], CRHBP [32], AR [33], and CDO1 [34]. However, the heterogeneity of ccRCC limits single-gene methylation in predicting ccRCC outcomes, so it is important to establish a multigene prediction model.

Related studies have shown that abnormal DNA MDG may cause transcriptional disorders, causing some gene expression mistakes and cell differentiation mistakes [35], and tumor suppressor genes and DNA repair genes were silenced due to hypermethylation. DNA methylation alterations at precancerous stages may determine tumor aggressiveness and patient prognosis [36]. Therefore, identifying abnormal DNA MDG can provide new insights for the risk assessment and prognosis of patients.

In this study, we identified key ccRCC-related MDG and constructed a prognostic model based on the level of DNA methylation. First, we screened nineteen methylation-driven genes by analyzing methylation and transcriptome data. Then, to discover the main biological characteristics of these methylation-driven genes, a series of gene functional enrichment analyses were performed. The gene functional enrichment analyses revealed that MDG were significantly linked to inflammatory response regulation and immune regulation (\(P < 0.05\), Figure 2).

In addition, a risk score-based prognostic model was constructed for these methylation-driven genes, and the predictability of this model was validated in the testing dataset. Three methylation-driven genes (NCKAP1L, EVI2A, and BATF) were identified and used to construct a prognostic risk model for ccRCC. We performed the univariate and multivariate Cox regression analyses on the entire TCGA cohort to verify that the risk model was a prognostic factor for ccRCC independent from other clinicopathological data (age, sex, ISUP grade, and AJCC stage) (HR = 1.221, 95% confidence interval: 1.063–1.402, and \(P = 0.005\)).
addition, we constructed a nomogram to predict overall survival in patients with ccRCC. The C index and calibration curve indicated that the predictive performance of the nomogram was good. Finally, we performed joint survival analysis by combining the methylation levels of the methylation-driven genes with the corresponding gene expression levels. The prognostic genes EVI2A and BATF were significantly related to prognosis.

The gene EVI2A, as the human homolog of mouse genes, may be associated with other proteins in the membrane as a part of a cell surface receptor complex [37]; EVI2A has been shown to be an oncogene [38]; EVI2A is highly expressed in oral tongue squamous cell carcinoma [39] and osteosarcoma [40] and is a risk factor for cancer prognosis. In this study, we found that EVI2A is one of the MDG of ccRCC, and its hypomethylation leads to poor prognosis (Figure 9(d)).

The BATF gene acts a pivotal part in the development of different types of cancer, including colon cancer, lymphoma, and multiple myeloma [41–43]. In addition, Feng et al. found that BATF acted as an oncogene in non-small-cell lung cancer [44]. In this study, the joint survival analyses revealed that the BATF gene was closely related with poor prognosis in patients with ccRCC (Figure 9(e)).

There are some limitations in this study. The specific mechanisms of EVI2A and BATF in ccRCC were not investigated in the previous literature. Therefore, future studies may focus on the molecular mechanisms underlying the potential interactions of EVI2A and BATF in ccRCC.
Figure 7: (a) Nomogram predicting 3- and 5-year overall survival for patients with ccRCC. The calibration curve for predicting patient survival at (b) 3 years and (c) 5 years in TCGA datasets. Nomogram-predicted probability of overall survival is plotted on the x-axis; actual overall survival is plotted on the y-axis.
Figure 8: MethylMix model of three selected DNA methylation-driven genes. The distribution maps show the methylation states of methylated genes. The histogram represents the distribution of methylation in tumor samples. The horizontal black bar demonstrates the distribution of methylation in the normal samples. The distribution of the methylation degree can be clearly seen from (a–c).
**Figure 9**: Correlation between DNA methylation and gene expression (a–c) and Kaplan-Meier survival curves for the integrative analysis of EVI2A and BATF (d, e).
5. Conclusion

In conclusion, we identified key ccRCC-related MDG and constructed a prognostic model based on the level of DNA methylation. In addition, we verified that the risk model was an independent prognostic factor for overall survival in ccRCC. Although further experimental verification is needed, this study provided an important bioinformatics foundation for in-depth studies of ccRCC DNA methylation.

Data Availability

The data included in the current study was available in TCGA database (https://cancergenome.nih.gov/).

Ethical Approval

Because the dataset in the current study was downloaded from TCGA, and data acquiring and application complied with TCGA publication guidelines and data access policies, additional approval by an ethics committee and consent to participate were not needed.

Conflicts of Interest

The authors declared that they have no conflict of interest.

Authors’ Contributions

MH and JX carried out the extraction of data and interpretation of results and wrote the initial manuscript. MH and HH performed the data analyses. ML and JW provided feedback on the manuscript and contributed to the study idea. All authors read and approved the final manuscript. Maolin Hu and Jiangling Xie contributed equally to this work.

Acknowledgments

This study was supported by the Beijing Hospital Clinical Research 121 Project (BJ-2018-090) and Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2018-I2M-1-002).

Supplementary Materials

Table S1 a training cohort. Table S2 a validation cohort. Table S3 normal DNA methylation data. Table S4 cancer DNA methylation data. Table S5 matched gene expression data. (Supplementary Materials)

References


