*Original Article*
# DNA methylation regulator-based signature for predicting clear cell renal cell carcinoma prognosis

Wenliang Gong[2*], Runan Yang[1*], Zhenmeng Wang[1]

[1]Department of Anesthesiology, Eastern Hepatobiliary Surgery Hospital, Naval Medical University (Second Military Medical University), Shanghai 200438, China; [2]Department of Urology, Shanghai Changhai Hospital, Naval Medical University (Second Military Medical University), Shanghai 200433, China. *Equal contributors.

**Abstract:** Objectives: To investigate the role of DNA methylation regulators in the prognosis of clear cell renal cell carcinoma (ccRCC) and to construct a DNA methylation regulator-based signature for predicting patient outcome. Methods: Data from the TCGA dataset were downloaded and analyzed to identify differentially expressed DNA methylation regulators and their interaction as well as correlation. Consensus clustering was used to establish groups of ccRCC with distinct clinical outcomes. A prognostic signature based on two sets of DNA methylation regulators was established and validated in an independent cohort. Results: Our analysis revealed that the expression levels of DNMT3B, MBD1, SMUG1, DNMT1, DNMT3A, TDG, TET3, MBD2, UHRF2, MBD3, UHRF1, and TET2 were significantly upregulated in ccRCC samples, while UNG, ZBTB4, TET1, ZBTB38, and MECP2 were markedly downregulated. UHRF1 was identified as a hub gene in the DNA methylation regulator interaction network. Significant differences were found regarding overall survival, gender, tumor status, and grade between ccRCC patients in the two risk groups. The prognostic signature, based on two sets of DNA methylation regulators, was an independent prognostic indicator, and these findings were validated in an external, independent cohort. Conclusions: The study provides evidence that DNA methylation regulators play a significant role in the prognosis of ccRCC and the developed DNA methylation regulator-based signature could effectively predict patient outcome.

## Introduction

Renal cell carcinoma (RCC) is a common malignancy of the genitourinary system, affecting more than 400,000 people worldwide each year [1]. There are several pathologic types of RCC, with clear cell renal cell carcinoma (ccRCC) being the most common type, accounting for approximately 75% of all primary renal cancer [2]. Although early-stage renal cancer is amenable to radical surgical treatment, 20-40% of patients experience disease progression after surgery. The 2-year survival rate for advanced or metastatic renal cancer is less than 20%, and the 5-year survival rate is only 12% [3, 4]. Moreover, drug resistance or non-response to the treatment also contribute to the poor prognosis of patients with ccRCC [3]. Therefore, it is crucial to identify biomarkers that can accurately and reliably predict the prognosis of ccRCC to improve the treatment outcome.

Although TNM pathological stage is the most important factor affecting the prognosis of patients with renal cancer [3], other factors such as, histologic grade, constitutional symptoms, sarcomatoid features, tumor thrombus level, perinephric or sinus fat invasion, the existence of tissue necrosis in the tumor, as well as biochemical abnormalities and changes are also related to the prognosis of renal cancer [5, 6]. Thus, various prognostic models have been developed and externally validated by combining independent prognostic factors [7-9]. However, the predictive power of these models is suboptimal as the clinical outcomes vary significantly between patients with ccRCC at identical TNM pathological stages [10]. Hence, identifying novel and reliable prognostic molecular signatures is vital for selecting the most appropriate therapeutic strategies and ameliorating the unfavorable prognosis of patients with ccRCC.

Cells in an organism contain the same genomic sequence, but they display a wide morphologic and functional diversity. It has been shown that epigenetic changes are responsible for this heterogeneity. Similarly, during tumorigenesis, epigenetic modifications occur such that DNA methylation modulate the chromatin structure and regulate gene expression and cancer cell proliferation [11, 12]. The process of DNA methylation is the covalent addition of the methyl group to the DNA sequence, for example, at the 5' carbon of the cytosine ring. This modification has been extensively studied in several diseases, particularly in cancer including renal cell carcinoma. The role of this modification in early detection, cancer progression, accurate diagnosis, and as biomarkers for the response to treatment has been the focus of many studies [13-19]. For example, Kubiliūtė et al. performed DNA methylation analysis in the whole genome of eleven pairs of ccRCC and noncancerous renal tissues (NRT) and found that the methylation frequency in the entire genome was significantly higher in ccRCC tissues than in NRT (33-60% vs 0-11%). Furthermore, numerous adverse clinicopathologic data were found to be associate with the hypermethylation of ZNF677 and PCDH8 in ccRCC tissues [20].

Hypermethylation on the promoter of oncogenes is a key event during tumorigenesis. Therefore, the addition and removal of methyl groups at the appropriate time and location play a crucial role in suppressing tumor gene mutations and hence cancer progression. Thus, DNA methylation is precisely controlled. A recent study performed a genome-wide CRISPR-Cas9 knockout screening in human embryonic stem cells (ESCs) to discover DNA methylation regulators. It successfully identified not only known methylation regulators such as KDM2B, TDG, and TET1 but also the functionally unknown gene QSER1, which cooperates with TET1 to prevent the chromatin binding of two DNA methyltransferases (DNMT3B and DNMT3A), thereby interfering with DNA methylation [21].

As we know, in the initial stages of ccRCC development, genetic and epigenetic alterations are gradually acquired, resulting in uncontrollable tumor cell growth [2]. Hence, understanding the molecular triggers of ccRCC tumorigenesis is crucial for the prevention and treatment of ccRCC. Although aberrant DNA methylation modifications is known to induce various tumor types, its role in ccRCC is not yet well understood. In particular, DNA methylation regulators and their relationship to the prognosis of ccRCC remain to be determined.

**Materials and methods**

*Collection of ccRCC datasets*

The clinical data in the TCGA database were downloaded from the Genomic Data Commons (GDC) (https://portal.gdc.cancer.gov/), where all TCGA RNA-seq transcriptome data are also available. The quantification and normalization of these data were performed with the Partek expectation maximization (EM) algorithm. In total, 542 ccRCC samples and 72 normal samples were analyzed.

*Analysis of the differential expression of DNA methylation regulator and data preprocessing*

The Edge R package was used to screen DNA methylation regulators between cancer specimens and normal control specimens. The false discovery rate (FDR) and adjusted *P*-values were determined using the classical Benjamini-Hochberg procedure. Meanwhile, log2FC > 1 was chosen as the cutoff threshold. In this study, instead of investigating DNA methylation, we studied the expression of genes that regulate DNA methylation since DNA methylation can function differently depending on the genomic context. We focused on twenty currently known DNA methylation regulators, including three erasers (TET1, TET2, TET3), three writers (DNMT1, DNMT3A, DNMT3B), and 14 readers (ZBTB4, ZBTB33, ZBTB38, UHRF1, UHRF2, MBD1, MBD2, MBD3, MBD4, MECP2, UNG, SMUG1, TDG, NTHL1) [22].

*PPI network construction and correlation analysis*

To identify interactions among the DNA methylation regulators, the STRING database (version 11.5 https://string-db.org/) was employed to analyze the protein-protein interaction (PPI) among DNA methylation regulators. Hub gene network illustration was visualized by utilizing

Cytoscape software (version 3.10.0 https://cytoscape.org). A Pearson correlation heatmap was used to display the association among different DNA methylation regulators.

*Analysis of consensus clusters*

To reveal the prognostic value of these DNA methylation regulator, samples from TCGA ccRCC dataset was clustered into two groups based on the consensus expression of the DNA methylation regulator in R using "Consensus Cluster Plus". Kaplan-Meier and log-rank tests were used to calculate the difference in overall survival (OS) between the two groups. A Chi-square test was performed to compare the distribution of T, M, gender, age, grade, and stage between the different groups.

*Generation and prediction of prognostic signatures*

Survival analysis in R was used to evaluate the relationship between DNA methylation regulators and the overall survival in the samples from TCGA ccRCC dataset. Genes with hazard ratios (HRs) exceeding 1 were considered high-risk, while genes with HRs less than 1 were considered protective. A two-set of prognostic gene signatures was identified. The Akaike information criterion (AIC) and multivariate Cox regression analysis were implemented for optimal model selection. For each patient, a risk score was calculated by adding the scores of each regulator together, which was done by multiplying the expression by a coefficient. The samples in the TCGA ccRCC dataset were stratified into low-risk and high-risk groups based on the median risk score. Finally, we used the Kaplan-Meier estimator and the 2-sided log-rank test to evaluate the differences in OS between the high-risk and low-risk patients. This prognostic model was assessed using a receiver operating characteristic curve (ROC). The clinicopathologic data were compared using the Chi-square test. Differences were visualized using heatmaps created with the Pretty Heatmaps R package. To identify independent prognostic factors for the TCGA ccRCC cohort, both univariate and multivariate Cox regression analyses were performed. Further evaluation was conducted on the survival difference between the high-risk and low-risk groups stratified by gender, stage, grade, and age.

*Patient recruitment in the validation cohort*

The validation study using patient samples from our own hospital was approved by the Ethics Committee of the Naval Medical University Shanghai Changhai Hospital. A total of 285 ccRCC patients with survival data at follow-up from the Urology Department of Shanghai Changhai Hospital Kidney Cancer Specialized Database were included in this study. The participants in this independent validation cohort had the same inclusion and exclusion criteria as used the discovery cohort. The patient inclusion criteria were: (1) Clinically diagnosed with renal cancer with indications for partial nephrectomy or radical nephrectomy, no significant contraindications to surgery; (2) Pathologically diagnosed again with ccRCC postoperatively and was verified by an experienced pathologist. The detailed clinicopathologic information of this cohort was summarized in Table S1. Written informed consent was obtained from all participants for the use of their frozen tissue samples.

*Quantitative real-time PCR*

Total RNA was extracted from tissue specimens byRNAiso Plus (Takara Bio, Tokyo, Japan) according to the manufacturer's instruction. PrimeScript™ RT Master Mix (Perfect Real Time) (Takara Bio, Tokyo, Japan) was used to synthesize the complementary DNAs (cDNAs). The amplification of cDNAs was conducted with TB Green® Premix Ex Taq™ (Takara Bio, Tokyo, Japan) using the QuantStudio™ 6 Flex Real-Time PCR System (Thermo Fisher Scientific, MA, USA). Normalization of gene expression against GAPDH and measurement of the relative expression levels of DNA methylation regulators were performed using the $2^{-\Delta\Delta Ct}$ method. The standard procedure for two-step PCR amplification was used: one cycle of 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, and 60°C for 20 s, followed by a melt curve run. The primer sequences and other regulators are shown in Table S2 and Figure S1.

*Prognostic signature validation*

Patients in the validation dataset were divided into high- and low-risk groups based on the risk scores and the median scores derived from linear prediction. The difference in OS between these two groups was then calculated, and the

clinicopathologic data in the validation cohort were also evaluated. Univariate and multivariate Cox regression analyses were conducted to assess risk scores. Clinicopathological data were used to stratify high-risk and low-risk groups.

*Statistical analysis*

The expression levels of DNA methylation regulators in ccRCC were compared across different WHO grades using one-way ANOVA, while t-tests were used to compare the expression levels based on age, gender, status, and stage. Patients were divided into two groups according to the consensus expression of DNA methylation regulators or classified into high- and low-risk groups based on the median risk score derived from the risk signature. Chi-square tests were employed to compare gender distribution, WHO grade, and status between the two risk groups. Prognostic analysis was performed using the Kaplan-Meier method to generate survival curves, and log-rank tests were used to determine the differences between groups. Hazard ratios (HRs) for DNA methylation regulators and DNA methylation regulator pattern-related genes were calculated utilizing the univariate Cox regression model. Statistical significance was set at $P<0.05$ for all two-sided tests. R 3.6.2 (https://www.r-project.org), SPSS 26 (SPSS Inc., Chicago, IL), and Prism 9 (GraphPad Software Inc., LaJolla, CA) were used for all statistical analyses.

## Results

*Twenty DNA methylation regulators were differentially expressed between ccRCC and control samples*

As shown in **Figure 1A**, the heatmap illustrated the expression level or pattern of the DNA methylation regulators in ccRCC and normal control samples. Positive colors near red stood for those that were highly expressed and positively correlated, while negative colors near blue indicated lowly expressed and negatively correlated. The expression levels of DNMT3B (***), MBD1 (***), SMUG1 (***), DNMT1 (***), DNMT3A (***), TDG (***), TET3 (***), MBD2 (***), UHRF2 (***), TET2 (*), MBD3 (***) and UHRF1 (***) were significantly upregulated in tumor samples compared to normal control samples, while tumor samples had

much lower levels of UNG (***), ZBTB4 (***), TET1 (**), ZBTB38 (*) and MECP2 (***). There was no significant difference between normal and cancer samples in NTHL1, MBD4, or ZBTB33 expression (**Figure 1B**).

*The interaction and correlation among twenty DNA methylation regulators*

The interaction among the twenty DNA methylation regulators was displayed in **Figure 2A**.

UHRF1 appeared to be the hub gene highly connected with neighboring genes within a given module of the interaction network among the differential gene expression (DEGs) according to the three centrality indicators: degree strength, betweenness centrality, and eigenvector centrality. The result of the Cytoscape network was further supported by a correlation analysis (**Figure 2B**). Similarly, UHRF1 was correlated to the other seven DNA methylation regulators, with DNMT1 being the most related (r=0.65) (**Figure 2C**).

*ccRCC patients with different clinical outcomes were clustered according to the expression of DNA methylation regulator*

Using the expression similarity of twenty DNA methylation regulators, k=2 was used to divide the patients in the ccRCC dataset into two clusters, cluster 1 and cluster 2 (**Figure 3A-C**). Patients in cluster 1 had a significantly shorter OS than those in cluster 2 (P=0.026) (**Figure 3D**). Moreover, we evaluated the association between the expression of the DNA methylation regulators and clinicopathologic features in ccRCC and found a significant difference between clusters 1 and 2 in rank (*), gender (*), and status (**), while there was no significant difference in other data such as stage and age (**Figure 3E**).

*Prognostic signature identification*

The TCGA ccRCC dataset was analyzed using univariate Cox regression to identify DNA methylation regulators related to OS. The results demonstrated that MECP2 (*P<0.001*), ZBTB4 (*P<0.001*), TET2 (*P=0.03*), MBD3 (*P=0.04*), DNMT3A (*P=0.02*), UHRF1 (*P<0.001*), SMUG1 (*P<0.001*), DNMT3B (*P<0.001*) and TDG (*P< 0.001*) were significantly correlated with OS (**Figure 4A**). To further confirm the reliability of
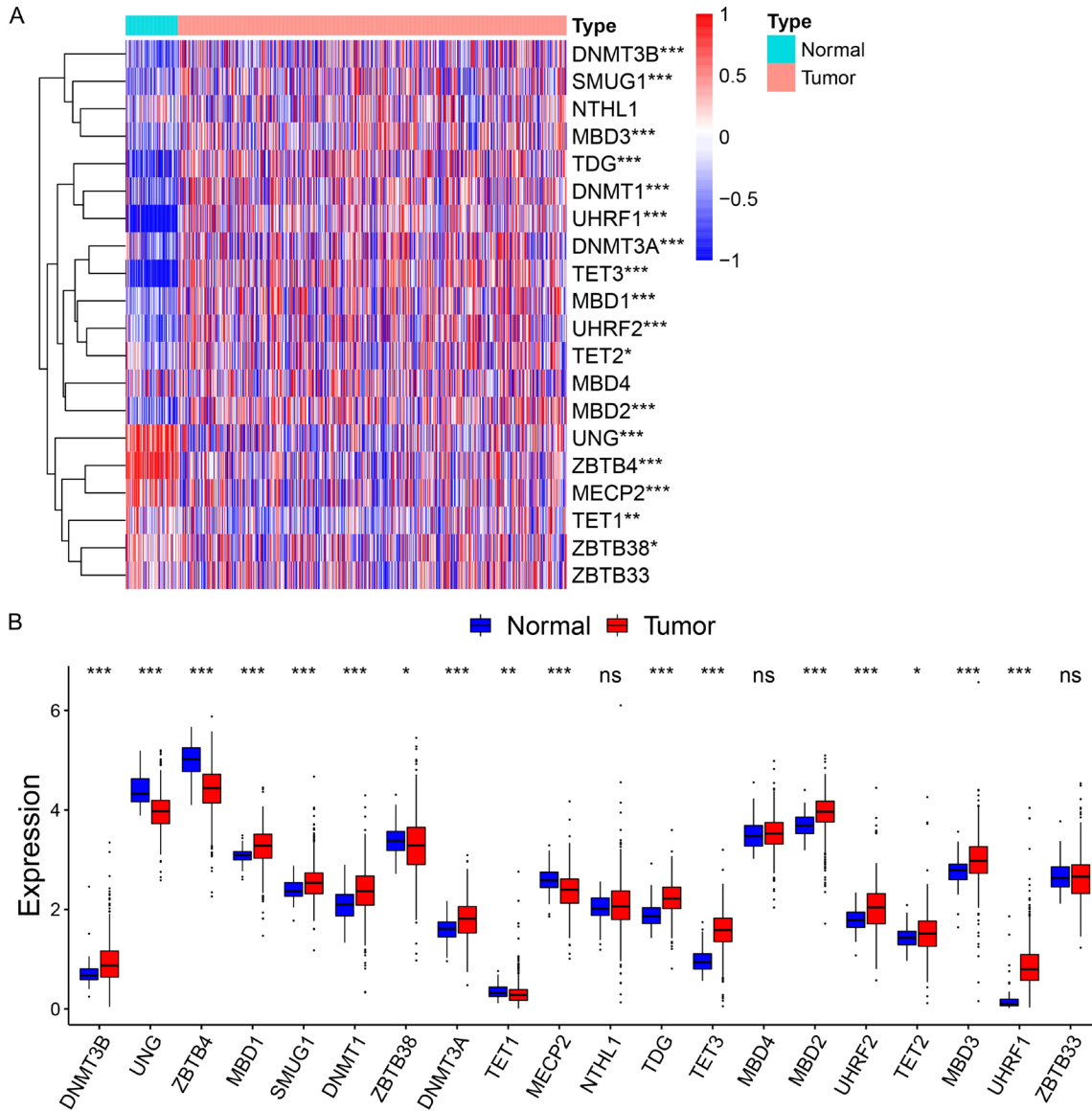
**Figure 1.** Expression levels of DNA methylation regulators in the TCGA clear cell renal cell carcinoma (ccRCC) dataset. A. The expression profiles of DNA methylation regulators were visualized as a heatmap. Red indicates higher expression, and blue indicates lower expression. B. Significantly different expression of DNA methylation regulators between tumor and normal tissue.

the prediction model, we examined the expression of the significant DNA methylation regulators in 64 pairs of ccRCC clinical samples with OS by quantitative RT-PCR. Among the significant DNA methylation regulator modifiers associated with OS, only UHRF1, TDG, DNMT3B, MBD3, MECP2, ZBTB4, and TET2 significantly differed in the clinical sample (**Figure 4B**). Furthermore, MECP2 (HR=0.42, 95% CI=0.28-0.63), ZBTB4 (HR=0.49, 95% CI=0.37-0.65), and TET2 (HR=0.55, 95% CI=0.37-0.81) were a group of protective genes with HRs less than

1, while TDG (HR=2.70, 95% CI=1.69-4.32), DNMT33B (HR=2.09, 95% CI=1.58-2.78), UHRF1 (HR=1.92, 95% CI=1.53-2.41) and MBD3 (HR=1.44, 95% CI=1.15-1.97) were risk genes with HRs more than 1 (**Figure 4A**).

Moreover, a two-set of regulators was chosen to establish the prognostic signature based on quantitative RT-PCR and HRs results. The coefficients were obtained from the LASSO algorithm to calculate the risk score for each patient using the formula: total risk score * (# of mark-
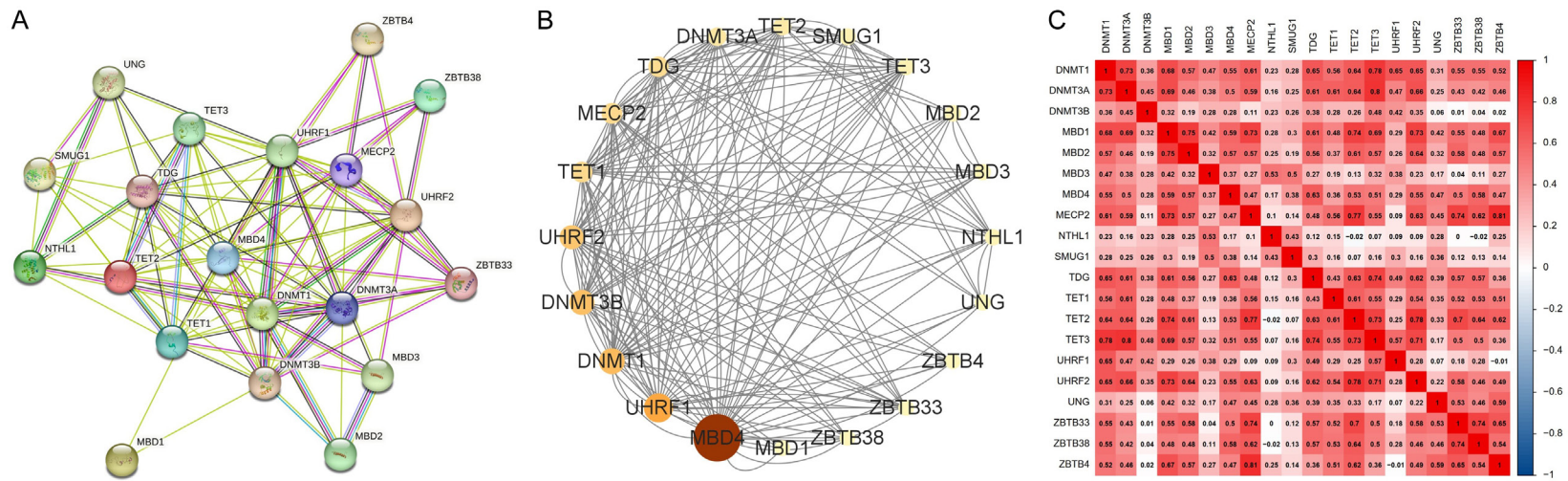
**Figure 2.** Correlations and interactions among DNA methylation regulators. A. A PPI network was constructed to evaluate the interaction among DNA methylation regulators. B. Cytoscape network quantified the regulator-regulator interactions. C. The Pearson correlation heatmap was used to determine the correlation among DNA methylation regulators.
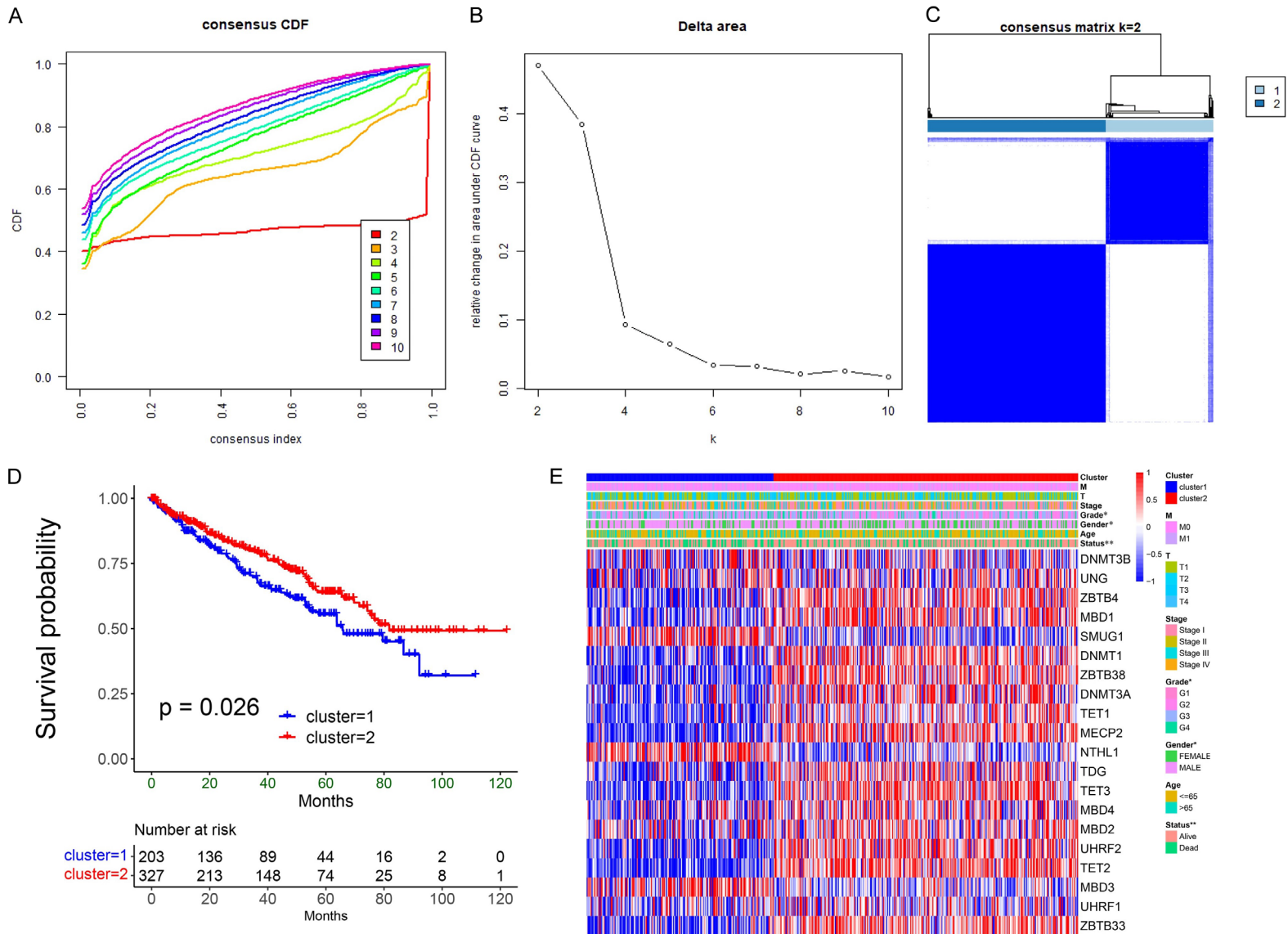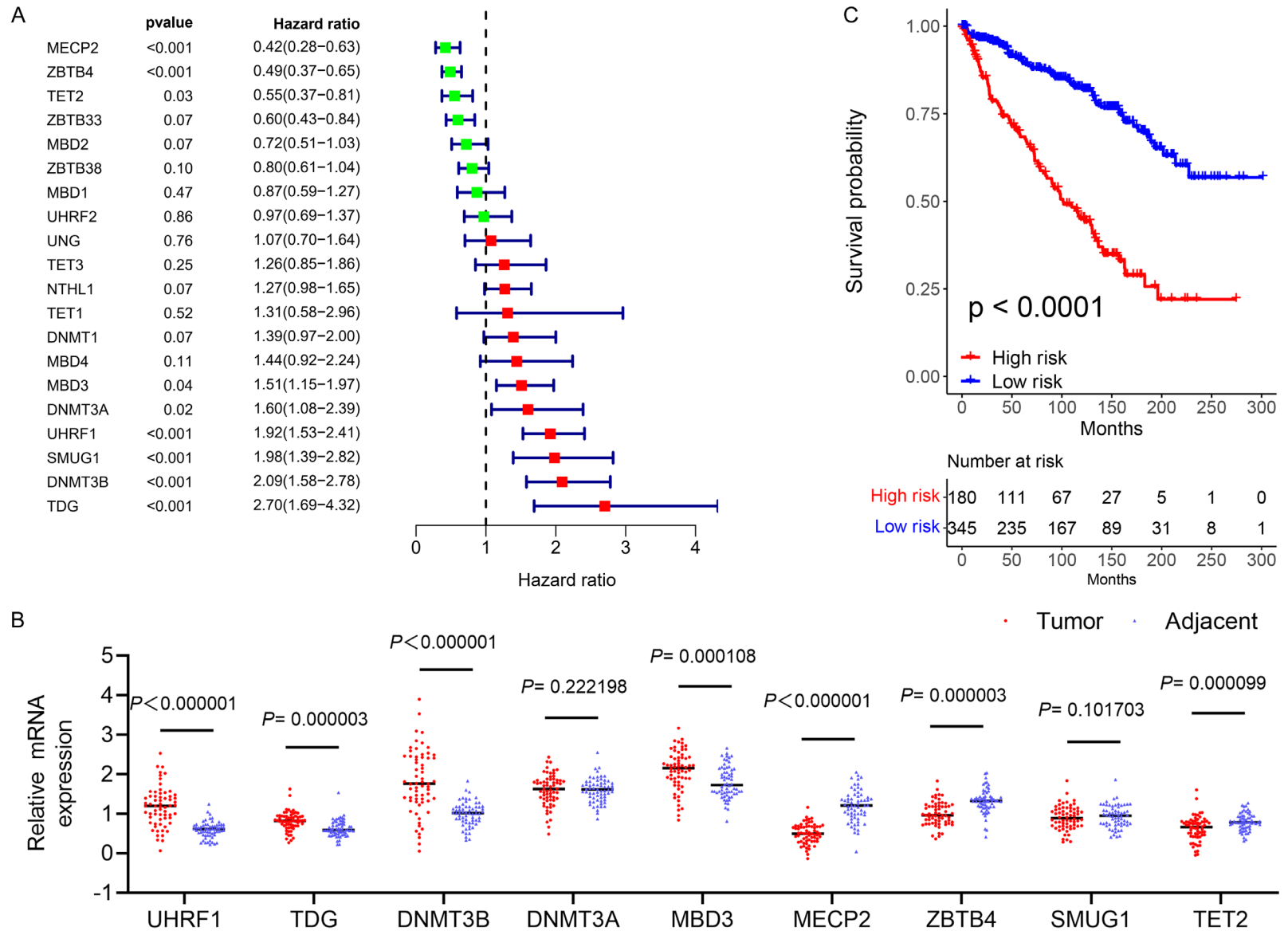
**Figure 3.** Overall survival (OS) and grade of patients in the two different clusters in TCGA ccRCC cohort. A. Consensus clustering cumulative distribution function (CDF) for k=2 to 10. B. Relative change in area under CDF curve for k=2 to 10. C. The TCGA ccRCC cohort was divided into two distinct clusters when k=2. D. The OS of patients in cluster 1 was significantly shorter than in cluster 2. E. Significant difference in grade, gender, and status between clusters 1 and 2.

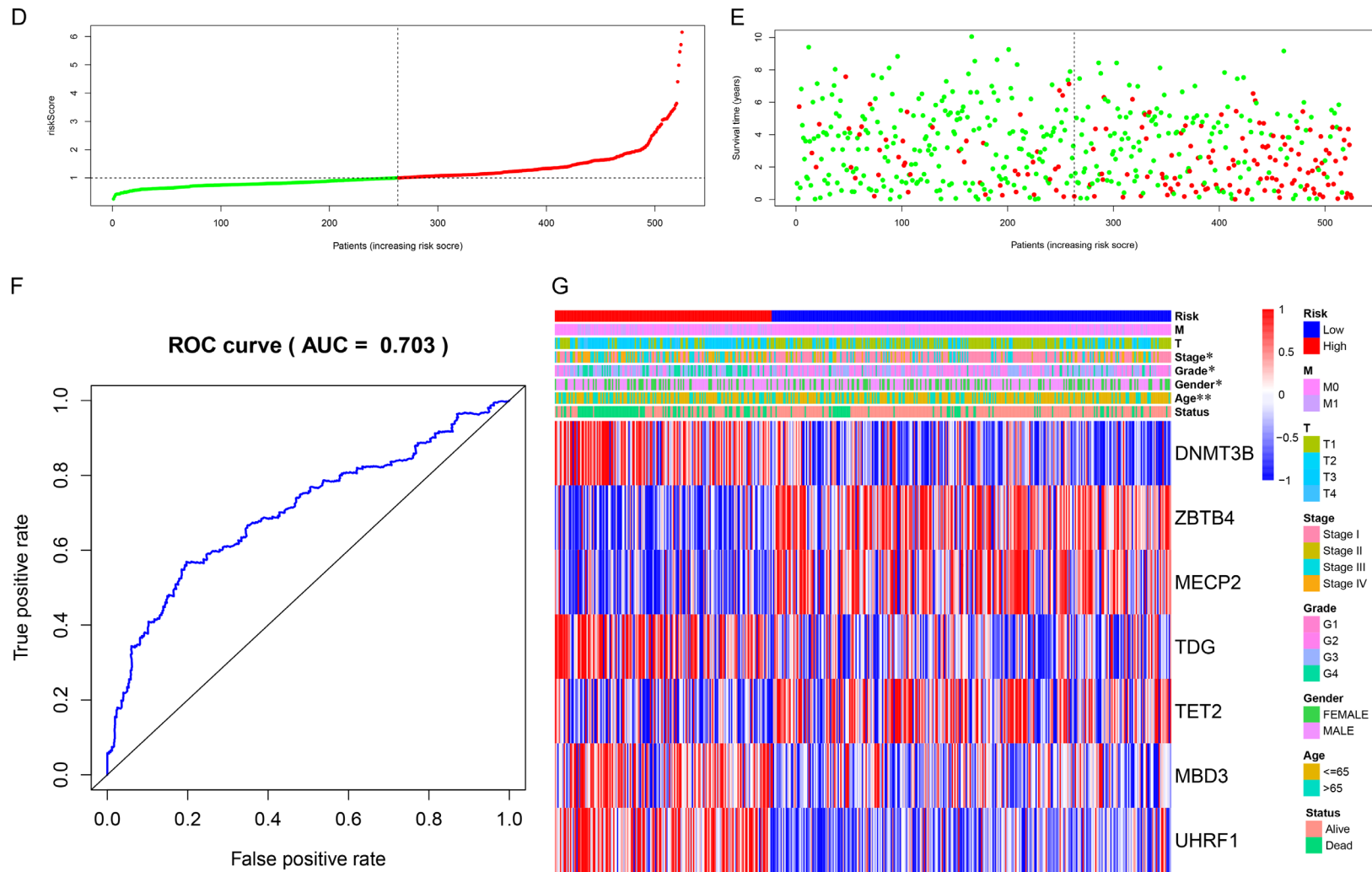DNA methylation regulator-based prognosis for ccRCC

**Figure 4.** Construction of the prognostic signature based on data from the TCGA ccRCC cohort. A. Univariate Cox regression analysis of the DNA methylation regulators to identify genes significantly correlated with overall survival (OS). B. qRT-PCR analysis of DNA methylation regulators in 64 pairs of ccRCC clinical samples. C. The OS of patients was much shorter in the high-risk group than in the low-risk group. D. The distributions of risk scores. E. The distributions of risk scores and OS status. The green and red dots indicate the alive and dead status, respectively. F. An ROC curve was used to evaluate the prediction efficiency of the prognostic signature. G. No significant differences in gender, grade, and stage were found between the high- and low-risk groups.
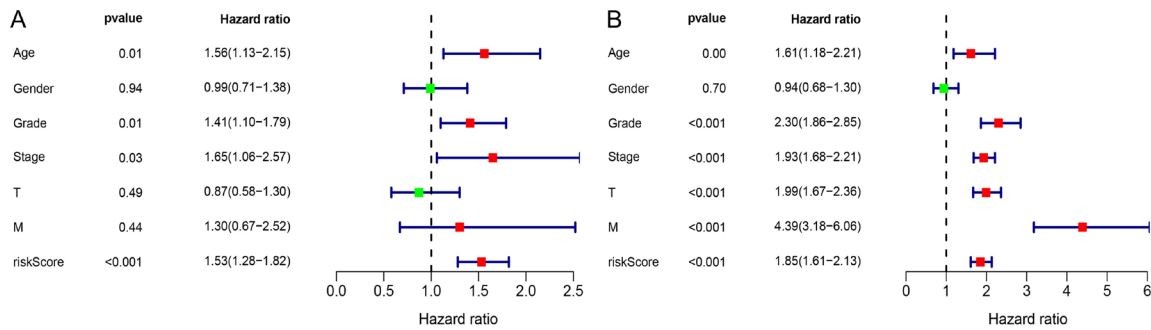
**Figure 5.** Identification of independent prognostic factors in the TCGA ccRCC cohort. A. Univariate Cox regression analysis of the risk score and clinicopathologic data to identify the indicators that significantly correlated with overall survival (OS). B. Multivariate Cox regression analysis of the risk score and clinicopathologic data to reveal the independent prognostic factors.

ers tested/sum of individual regulator risk scores). A total of 525 ccRCC cases were divided into 2 groups, high- and low-risk groups. The survival curve displayed that the patients with in the high-risk group had a significantly shorter OS than those in the low-risk group (P<0.0001) (**Figure 4C**). **Figure 4D** shows the distributions of the two sets of gene signature-based risk scores. The risk scores as well as OS status were distributed (**Figure 4E**). Importantly, the prognostic signature model exhibited satisfactory prediction efficiency with an area under the curve value of 0.703 (**Figure 4F**). As shown in **Figure 4G**, two-set of DNA methylation regulators were expressed in the high- and low-risk groups, and there was a significant difference between the two risk groups in stage (*), grade (*), gender (*), and age (**).

*Prognostic signature-based risk score was an independent prognostic factor for TCGA ccRCC*

The prognostic signature-based risk scores were rated utilizing univariate and multivariate Cox regression analyses. After excluding samples with unknown age or gender, 525 patients were included for the subsequent analysis. The univariate Cox regression analysis indicated that the age (P=0.01, HR=1.56, 95% CI=1.13-2.15), grade (P=0.01, HR=1.41, 95% CI=1.10-1.79), stage (P=0.03, HR=1.65, 95% CI=1.06-2.57) and risk score (P<0.001, HR=1.53, 95% CI=1.28-1.82) were all significantly associated with the OS (**Figure 5A**). Furthermore, the age (P<0.001, HR=1.61, 95% CI=1.18-2.21), grade (P<0.001, HR=2.30, 95% CI=1.86-2.85), stage (P<0.001, HR=1.93, 95% CI=1.68-2.21), T (P<0.001, HR=1.99, 95% CI=1.67-2.36), M (P<0.001, HR=4.39, 95% CI=3.18-6.06) and

risk score (P<0.001, HR=1.85, 95% CI=1.61-2.13) were identified as independent prognostic factors by the Cox multivariate regression model (**Figure 5B**). In addition, we tested the intrinsic value of different clinicopathological data. We found a significantly shorter OS in high-risk group than in the low-risk group for patients younger than 65 years (P<0.0001), patients older than 65 years (P<0.0001), female patients (P<0.0001), male patients (P<0.0001), patients at G1-G2 (P=0.0024), patients at G3-G4 (P<0.0001), patients at stage I-II (P=0.03), and patients at the stage III-IV (P<0.0001) (**Figure 6A-H**).

*Prognostic signature validation*

The validated cohort included 285 patients with ccRCC. A cutoff value was used to classify 86 patients into a high-risk group, and 199 patients were categorized into a low-risk group. The OS was significantly shorter in the high-risk group than in the low-risk group (P<0.0001) (**Figure 7A**). **Figure 7B** and **7C** show the risk scores, OS, and OS status distributions. The high- and low-risk groups differed significantly in several clinicopathologic values such as T (*), stage (**), gender (*), and age (*) (**Figure 7D**). Univariate Cox regression analysis revealed that stage (P<0.001, HR=3.51, 95% CI=2.04-6.03) was significantly associated with OS (**Figure 8A**). Furthermore, multivariate Cox regression analysis showed that stage (P<0.001, HR=2.74, 95% CI=1.99-3.77), T (P<0.001, HR=2.25, 95% CI=1.62-3.12), and risk score (P<0.001, HR=1.25, 95% CI=1.11-1.41) were independent prognostic indicators (**Figure 8B**). As shown in **Figure 9A-H**, the OS
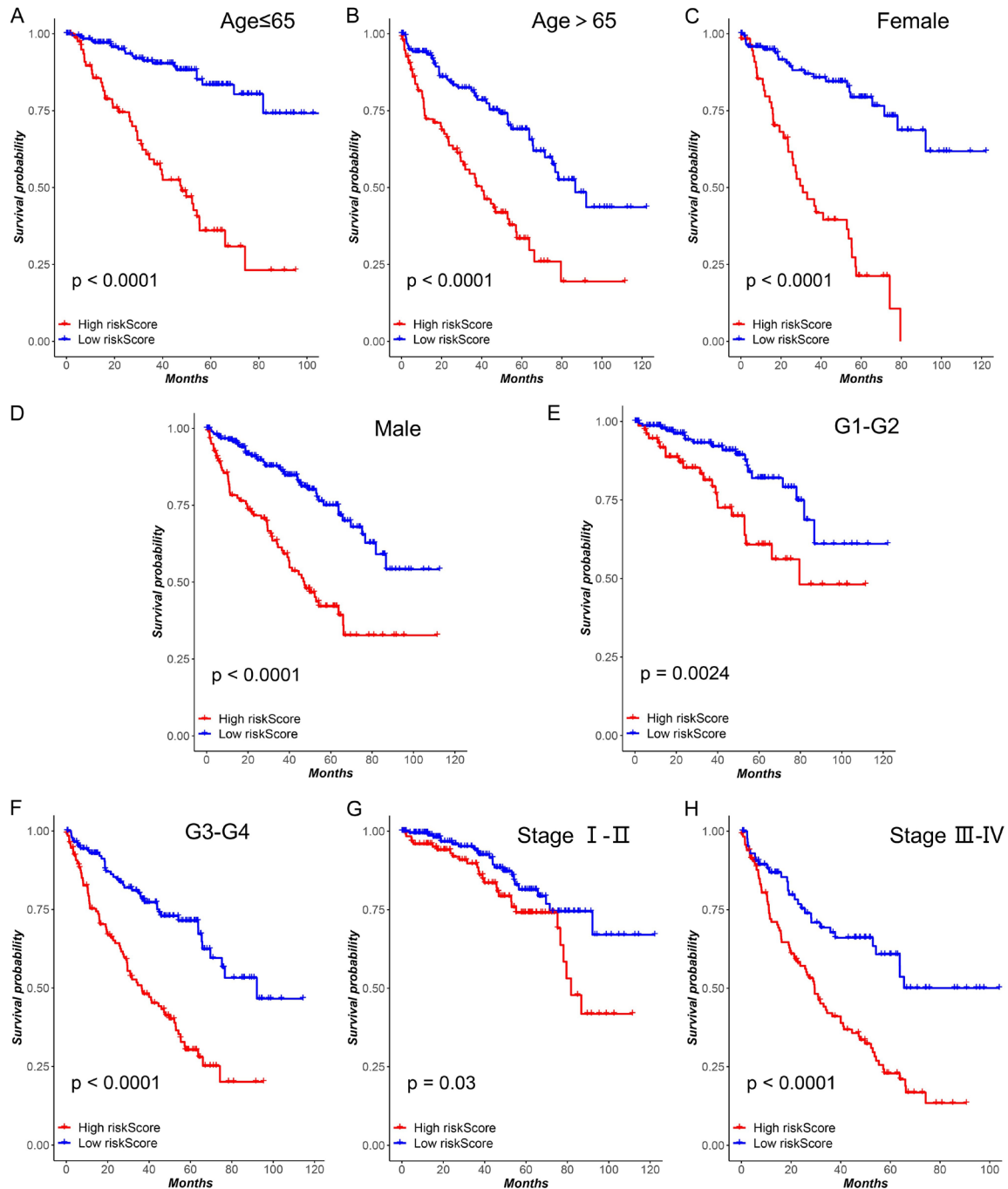
**Figure 6.** Survival difference between the high- and low-risk groups stratified by clinicopathologic data in the TCGA ccRCC cohort. A, B. The difference in overall survival (OS) between high- and low-risk groups stratified by age. C, D. The difference in OS between high- and low-risk groups stratified by gender. E, F. The difference in OS between high- and low-risk groups stratified by grade. G, H. The difference in OS between high- and low-risk groups stratified by stage.

rate was significantly shorter in the high-risk group compared to that of the low-risk group for patients younger than age 65 ($P$=0.00027), older than age 65 ($P$=0.061), female patients ($P$=0.012), male patients ($P$=0.00022), pati-

ents at the T3-T4 ($P$=0.025), patients at the stage I-II ($P$=0.035) or those at the stage III-IV ($P$=0.046). However, no significant difference in OS was observed for T1-T2 patients ($P$=0.35) between the high-risk and low-risk groups.
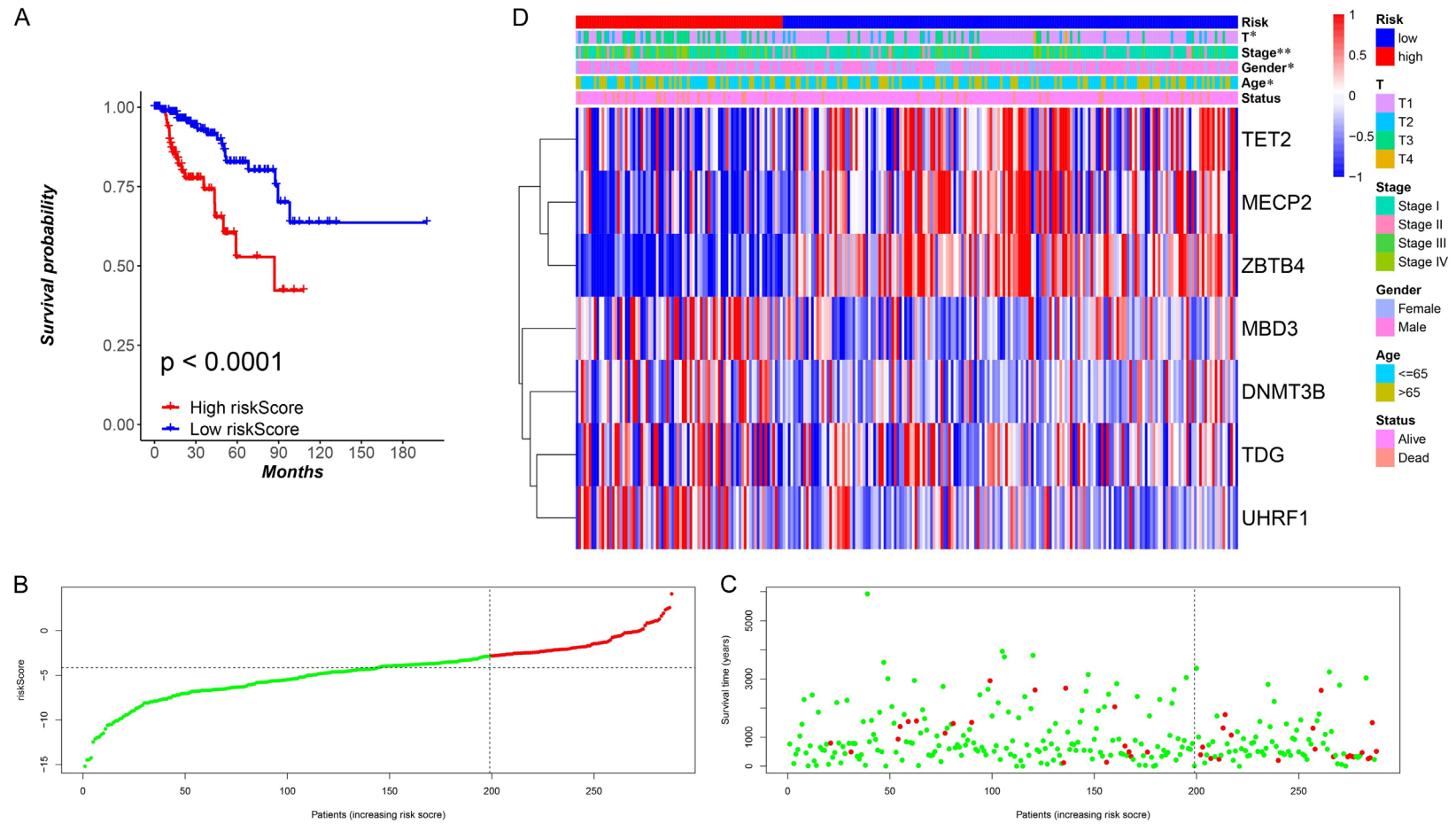
**Figure 7.** Validation of the prognostic signature in an independent ccRCC cohort. A. The ccRCC patients in the high-risk group had a significantly shorter overall survival (OS) than those in the low-risk group. B. Distributions of risk scores. C. Distributions of risk scores and OS status. D. Significant differences in T, stage, gender, and age between the high- and low-risk groups.
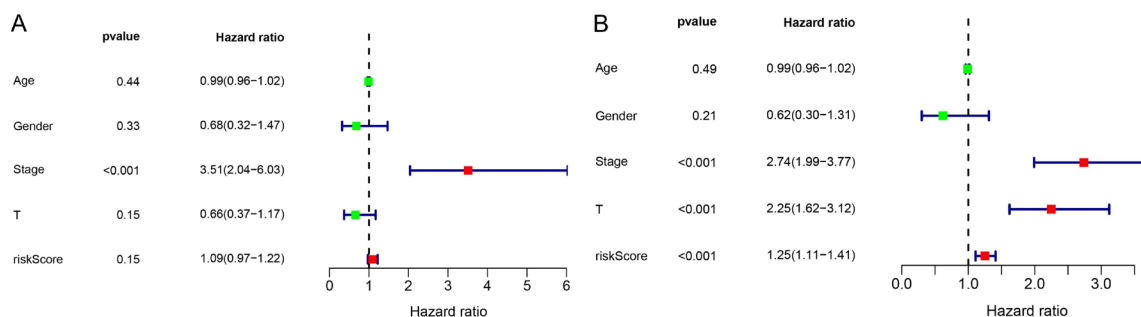
**Figure 8.** Identification of independent prognostic factors in the validated ccRCC cohort. A. Univariate Cox regression analysis of the risk score and clinicopathologic data to identify the significantly associated indicators with overall survival (OS). B. Multivariate Cox regression analysis of the risk score and clinicopathologic data to reveal the independent prognostic factors.

## Discussion

In this study, we found that 17 out of the 20 foremost DNA methylation regulators were either highly expressed or expressed at low levels in TCGA ccRCC datasets, indicating involvement of these important regulators in the oncogenesis or in the prognosis of patients with ccRCC. Our further unsupervised consensus clustering of DNA methylation regulators defined 2 ccRCC subgroups. The OS, tumor grade, gender, and status were dramatically distinct between the 2 ccRCC subgroups, implying that an unfavorable prognosis of ccRCC is associated with the expression level of DNA methylation regulator.

Significantly, we established seven genetic risk profiles, including TDG, DNMT3B, UHRF1, MBD3, MECP2, ZBTB4, and TET2, and demonstrated that they could robustly predict the prognosis of ccRCC patients from different independent cohorts. Our predictive model revealed that MECP2, ZBTB4, and TET2 expression were positively associated with ccRCC prognosis, suggesting that they might act as tumor suppressors.

Consistent with the findings from previous studies, epigenetic regulators, including DNA methylation regulators, are aberrantly expressed in diverse types of cancer. For example, low expression of ZBTB4 is related to a decreased latency of recurrence in breast and prostate cancers [23]. ZBTB4-/- mice were more vulnerable to DMBA/TPA-induced skin carcinogenesis [24].

Recently, Luo et al. discovered that MECP2 depletion in colorectal cancer (CRC) cells sig-

nificantly inhibited stem cell populations, as well as suppressing the migration and invasion of the CRC cells in vitro and metastasis in vivo [25]. However, in our bioinformatic analysis and validation experiments, we found that MECP2 was downregulated in ccRCC, suggesting that the same DNA methylation regulators might have distinct roles in different cancers. A further in-depth investigation is required to elucidate the underlying molecular mechanisms of these different functions.

Moreover, we found DNMT3B, TDG, UHRF1, and MBD3 were highly expressed in ccRCC and negatively associated with prognosis. As described above, DNA methylation is catalyzed by DNMTs which include DNMT1, DNMT2, DNMT3A and DNMT3B. Distinct from other members of DNMTs, DNMT2 lacks methylation transferase activity in vitro, while DNMT3B, DNMT3A, and DNMT1 are the main ones that exert methyltransferase activity in vivo [26, 27]. As an oncogene, DNMT3B is frequently upregulated in tumors and is associated with the downregulation of its targets. In some cancer types, its overexpression is an adverse prognostic marker [28].

Similarly, aberrant TDG expression is observed in many malignant diseases or their disease models, including esophageal cancer [29], rectal cancer [30], and melanoma [31]. The TDG-mediated imbalance of the demethylation pathway increases the risk of gene mutation, which, in turn, increases the susceptibility of cells to various risk factors, thereby leading to genomic instability and tumor formation. Several studies have discovered that decreased gene expression levels of TDG or decreased TDG protein
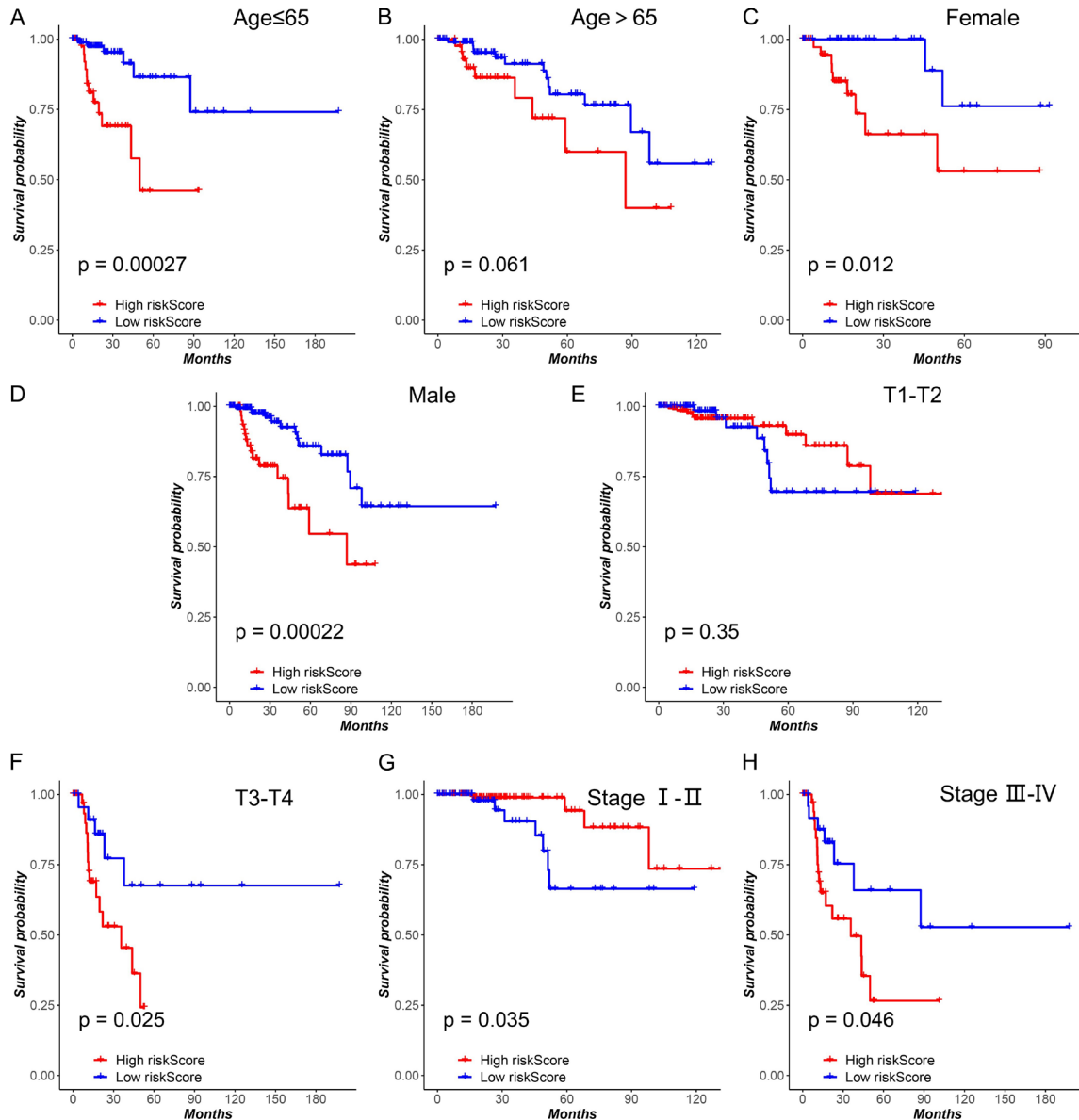
**Figure 9.** Survival differences between high- and low-risk groups stratified by clinicopathologic data in the validated ccRCC cohort. A, B. The difference in overall survival (OS) between the high- and low-risk groups stratified by age. C, D. The difference in OS between high- and low-risk groups stratified by gender. E, F. The difference in OS between high- and low-risk groups stratified by grade. G, H. The difference in OS between high- and low-risk groups stratified by stage.

activity may increase the risk of cancer development.

At present, information on the roles of these genes in tumorigenesis is still scarce. Nevertheless, a few researchers recently reported that UHRF1 expression was significantly higher in RCC tumor tissues than in normal control tissue [32], which is consistent with our findings. The downregulation of UHRF1 expression induced by shRNA in RCC cell lines can cause decreased cell viability, attenuated cell migration and invasion, and incremental apoptosis. In support of these in vitro findings, knockdown of UHRF1 in a RCC xenograft models also significantly inhibited tumor growth. Furthermore, UHRF1 was shown to recruit HDAC1 to the TXNIP promoter, regulating the deacetylation of histone H3K9 and leading to the repression of TXNIP expression.

The development of ccRCC is a complex and multi-step process resulting from the interplay of mutations, such as VHL, epigenetic regulation, abnormal angiogenesis, and the aberrant expression of immune-related genes [7]. This complexity limits the use of single or a few genes to predict disease risk. Furthermore, the accuracy and reliability of disease prediction are significantly influenced by the quality and quantity of gene expression and other molecular events. In our study, we constructed a robust risk profile of MECP2, ZBTB4, TET2 and TDG, DNMT3B, UHRF1, and MBD3 using the TCGA ccRCC dataset, which showed excellent performance in predicting the clinical outcome of ccRCC. These findings provide important evidence for future investigations into the involvement of DNA methylation or DNA methylation regulators in ccRCC.

However, the limited sample number may have led to bias and reduce the accuracy of the prediction models. As technology advances and research progresses, including experimental validation, the expression of these DNA methylation regulators holds great potential to improve the accuracy and reliability for the prognosis of patients with ccRCC.

## Conclusion

In this study, we investigated the crucial role of DNA methylation regulators in the progression of ccRCC. More importantly, we established a robust prognostic signature which was further validated in independent patient cohorts. This DNA methylation regulators-related signature could accurately and reliably predict a poor clinical outcome, indicating that it could serve as a promising biomarker for monitoring ccRCC development and providing useful information for therapeutic decision making.

## Acknowledgements

## Disclosure of conflict of interest

None.

Address correspondence to: Zhenmeng Wang, Department of Anesthesiology, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, 225 Changhai Road, Yangpu District, Shanghai 200438, China. E-mail: wzm11998@163.com

## References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Chen F, Zhang Y, Senbabaoglu Y, Ciriello G, Yang L, Reznik E, Shuch B, Micevic G, De Velasco G, Shinbrot E, Noble MS, Lu Y, Covington KR, Xi L, Drummond JA, Muzny D, Kang H, Lee J, Tamboli P, Reuter V, Shelley CS, Kaipparettu BA, Bottaro DP, Godwin AK, Gibbs RA, Getz G, Kucherlapati R, Park PJ, Sander C, Henske EP, Zhou JH, Kwiatkowski DJ, Ho TH, Choueiri TK, Hsieh JJ, Akbani R, Mills GB, Hakimi AA, Wheeler DA and Creighton CJ. Multilevel genomics-based taxonomy of renal cell carcinoma. Cell Rep 2016; 14: 2476-2489.

[3] Diaz-Montero CM, Rini BI and Finke JH. The immunology of renal cell carcinoma. Nat Rev Nephrol 2020; 16: 721-735.

[4] Singh D. Current updates and future perspectives on the management of renal cell carcinoma. Life Sci 2021; 264: 118632.

[5] Turajlic S, Swanton C and Boshoff C. Kidney cancer: the next decade. J Exp Med 2018; 215: 2477-2479.

[6] Chowdhury N and Drake CG. Kidney cancer: an overview of current therapeutic approaches. Urol Clin North Am 2020; 47: 419-431.

[7] Bui TO, Dao VT, Nguyen VT, Feugeas JP, Pamoukdjian F and Bousquet G. Genomics of clear-cell renal cell carcinoma: a systematic review and meta-analysis. Eur Urol 2022; 81: 349-361.

[8] Ghatalia P, Gordetsky J, Kuo F, Dulaimi E, Cai KQ, Devarajan K, Bae S, Naik G, Chan TA, Uzzo R, Hakimi AA, Sonpavde G and Plimack E. Prognostic impact of immune gene expression signature and tumor infiltrating immune cells in localized clear cell renal cell carcinoma. J Immunother Cancer 2019; 7: 139.

[9] Zhou QH, Li KW, Chen X, He HX, Peng SM, Peng SR, Wang Q, Li ZA, Tao YR, Cai WL, Liu RY and Huang H. HHLA2 and PD-L1 co-expression pre-

dicts poor prognosis in patients with clear cell renal cell carcinoma. J Immunother Cancer 2020; 8: e000157.

[10] Chipollini J, Abel EJ, Peyton CC, Boulware DC, Karam JA, Margulis V, Master VA, Zargar-Shoshtari K, Matin SF, Sexton WJ, Raman JD, Wood CG and Spiess PE. Pathologic predictors of survival during lymph node dissection for metastatic renal-cell carcinoma: results from a multicenter collaboration. Clin Genitourin Cancer 2018; 16: e443-e450.

[11] Vaissiere T, Sawan C and Herceg Z. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutat Res 2008; 659: 40-48.

[12] Kanwal R and Gupta S. Epigenetic modifications in cancer. Clin Genet 2012; 81: 303-311.

[13] Kerachian MA, Azghandi M, Mozaffari-Jovin S and Thierry AR. Guidelines for pre-analytical conditions for assessing the methylation of circulating cell-free DNA. Clin Epigenetics 2021; 13: 193.

[14] Zhou C, Ye M, Ni S, Li Q, Ye D, Li J, Shen Z and Deng H. DNA methylation biomarkers for head and neck squamous cell carcinoma. Epigenetics 2018; 13: 398-409.

[15] Kelly H, Benavente Y, Pavon MA, De Sanjose S, Mayaud P and Lorincz AT. Performance of DNA methylation assays for detection of high-grade cervical intraepithelial neoplasia (CIN2+): a systematic review and meta-analysis. Br J Cancer 2019; 121: 954-965.

[16] Li R, Yang YE, Yin YH, Zhang MY, Li H and Qu YQ. Methylation and transcriptome analysis reveal lung adenocarcinoma-specific diagnostic biomarkers. J Transl Med 2019; 17: 324.

[17] Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, Lin Y, Wang D, Yang X, Zheng Y, Sang X and Zhao H. DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. Theranostics 2019; 9: 7251-7267.

[18] Chakravarthy A, Furness A, Joshi K, Ghorani E, Ford K, Ward MJ, King EV, Lechner M, Marafioti T, Quezada SA, Thomas GJ, Feber A and Fenton TR. Pan-cancer deconvolution of tumour composition using DNA methylation. Nat Commun 2018; 9: 3220.

[19] Chen W, Zhuang J, Wang PP, Jiang J, Lin C, Zeng P, Liang Y, Zhang X, Dai Y and Diao H. DNA methylation-based classification and identification of renal cell carcinoma prognosis-subgroups. Cancer Cell Int 2019; 19: 185.

[20] Kubiliute R, Zukauskaite K, Zalimas A, Ulys A, Sabaliauskaite R, Bakavicius A, Zelvys A, Jankevicius F and Jarmalaite S. Clinical significance of novel DNA methylation biomarkers for renal clear cell carcinoma. J Cancer Res Clin Oncol 2022; 148: 361-375.

[21] Dixon G, Pan H, Yang D, Rosen BP, Jashari T, Verma N, Pulecio J, Caspi I, Lee K, Stransky S, Glezer A, Liu C, Rivas M, Kumar R, Lan Y, Torregroza I, He C, Sidoli S, Evans T, Elemento O and Huangfu D. QSER1 protects DNA methylation valleys from de novo methylation. Science 2021; 372: eabd0875.

[22] Meng Q, Lu YX, Ruan DY, Yu K, Chen YX, Xiao M, Wang Y, Liu ZX, Xu RH, Ju HQ and Qiu MZ. DNA methylation regulator-mediated modification patterns and tumor microenvironment characterization in gastric cancer. Mol Ther Nucleic Acids 2021; 24: 695-710.

[23] Kim K, Chadalapaka G, Pathi SS, Jin UH, Lee JS, Park YY, Cho SG, Chintharlapalli S and Safe S. Induction of the transcriptional repressor ZBTB4 in prostate cancer cells by drug-induced targeting of microRNA-17-92/106b-25 clusters. Mol Cancer Ther 2012; 11: 1852-1862.

[24] Roussel-Gervais A, Naciri I, Kirsh O, Kasprzyk L, Velasco G, Grillo G, Dubus P and Defossez PA. Loss of the methyl-CpG-binding protein ZBTB4 alters mitotic checkpoint, increases aneuploidy, and promotes tumorigenesis. Cancer Res 2017; 77: 62-73.

[25] Luo D and Ge W. MeCp2 promotes colorectal cancer metastasis by modulating ZEB1 transcription. Cancers (Basel) 2020; 12: 758.

[26] Cui D and Xu X. DNA methyltransferases, DNA methylation, and age-associated cognitive function. Int J Mol Sci 2018; 19: 1315.

[27] Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat Rev Genet 2018; 19: 81-92.

[28] Bishop KS and Ferguson LR. The interaction between epigenetics, nutrition and the development of cancer. Nutrients 2015; 7: 922-947.

[29] Li WQ, Hu N, Hyland PL, Gao Y, Wang ZM, Yu K, Su H, Wang CY, Wang LM, Chanock SJ, Burdett L, Ding T, Qiao YL, Fan JH, Wang Y, Xu Y, Shi JX, Gu F, Wheeler W, Xiong XQ, Giffen C, Tucker MA, Dawsey SM, Freedman ND, Abnet CC, Goldstein AM and Taylor PR. Genetic variants in DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population. Carcinogenesis 2013; 34: 1536-1542.

[30] Vasovcak P, Krepelova A, Menigatti M, Puchmajerova A, Skapa P, Augustinakova A, Amann G, Wernstedt A, Jiricny J, Marra G and Wimmer K. Unique mutational profile associated with a loss of TDG expression in the rectal cancer of a patient with a constitutional PMS2 deficiency. DNA Repair (Amst) 2012; 11: 616-623.

[31] Mancuso P, Tricarico R, Bhattacharjee V, Cosentino L, Kadariya Y, Jelinek J, Nicolas E, Einarson M, Beeharry N, Devarajan K, Katz RA, Dorjsuren DG, Sun H, Simeonov A, Giordano A, Testa JR, Davidson G, Davidson I, Larue L, Sobol RW, Yen TJ and Bellacosa A. Thymine DNA glycosylase as a novel target for melanoma. Oncogene 2019; 38: 3710-3728.

[32] Jiao D, Huan Y, Zheng J, Wei M, Zheng G, Han D, Wu J, Xi W, Wei F, Yang AG, Qin W, Wang H and Wen W. UHRF1 promotes renal cell carcinoma progression through epigenetic regulation of TXNIP. Oncogene 2019; 38: 5686-5699.

**Table S1.** The clinical information of the validated ccRCC cohort from Changhai Hospital Kidney Cancer Specialized Disease Database

| Clinicopathological features | Number |
| --- | --- |
| Age | |
| ≤65 | 198 (69.47%) |
| >65 | 87 (30.53%) |
| Gender, n (%) | |
| Male | 186 (65.26%) |
| Female | 99 (34.74%) |
| Pathological diagnosis | |
| clear cell renal cell carcinoma | 285 (100%) |
| Tumor grade | |
| G1 | 130 (45.61%) |
| G2 | 117 (41.05%) |
| G3 | 28 (9.82%) |
| G4 | 10 (3.52%) |
| TNM stage | |
| Stage I | 166 (58.24%) |
| Stage II | 99 (34.73%) |
| Stage III | 18 (6.32%) |
| Stage IV | 2 (0.07%) |

**Table S2.** Primer information for DNA methylation regulators

| Species | Gene | primer | Sequence (5'->3') | Length/nt | Tm/°C | Product length/bp |
| --- | --- | --- | --- | --- | --- | --- |
| Homo | DNMT3B | F1 | CTGGATGTTTGAGAATGTTGTAGC | 24 | 60.06 | 90 |
| | | R1 | GCATCAATCATCACTGGATTACAC | 24 | 60.63 | |
| | MBD1 | F2 | ACTGTGGAATCAGCTTCTCAGG | 22 | 60.82 | 119 |
| | | R2 | CACACGCTTAAACATTCTCTGTTC | 24 | 60.22 | |
| | SMUG1 | F3 | CACTGTTTTGTCCACAATCTATGC | 24 | 60.81 | 111 |
| | | R3 | ACAGATCCCAAGAAGCTGTTCTC | 23 | 61.14 | |
| | DNMT1 | F4 | AAACCTCAGGAAGAGTCTGAAAGA | 24 | 59.93 | 102 |
| | | R4 | TCGTTCTCTGGATGTAACTCTACG | 24 | 59.84 | |
| | DNMT3A | F5 | CCTGCGGTGATCTCCAAGTC | 20 | 63.13 | 90 |
| | | R5 | CTCACTCCGCTTCTCCAAGT | 20 | 59.60 | |
| | TDG | F6 | AACCATTTTTGGAAGTGTTTGTTT | 24 | 60.06 | 97 |
| | | R6 | ATCCAATACCATACTTCCCTGGTA | 24 | 59.87 | |
| | TET3 | F7 | ACCTCTTAAGTACCTGGACACACC | 24 | 59.86 | 120 |
| | | R7 | GGACCTTCATCTTTCTCCACTATT | 24 | 59.07 | |
| | MBD2 | F8 | GCGATGTCTACTACTTCAGTCCAA | 24 | 59.85 | 106 |
| | | R8 | GTTCTGAAGTCAAAACTGCTGAGA | 24 | 60.10 | |
| | UHRF2 | F9 | TGGACAGTGTACCCTCTACGTCTA | 24 | 60.11 | 101 |
| | | R9 | TAGAGTACCGCTTTCTGGGTATTC | 24 | 60.05 | |
| | TET2 | F10 | GAGAAATCATGGAAGAAAGGTTTG | 24 | 60.33 | 111 |
| | | R10 | ACTTAGCAATAGGACATCCCTGAG | 24 | 60.04 | |
| | MBD3 | F11 | GGCCACAGGGATGTCTTTTACTAT | 24 | 61.72 | 108 |
| | | R11 | GAAGTCGAAGGTGCTCAGGTC | 21 | 61.35 | |
| | UHRF1 | F12 | TCATCAGAGAGGACAAGAGCAAC | 23 | 60.95 | 111 |
| | | R12 | CCTCCACTTTACTCAGGAACAACT | 24 | 60.09 | |

| | | | | | |
|---|---|---|---|---|---|
| UNG | F13 | TTGCAGAAGAAAGAAAGCATTACA | 24 | 60.40 | 99 |
| | R13 | GGATGACAACCTTCACATCTTTTA | 24 | 59.42 | |
| ZBTB4 | F14 | CTATTGTGAGAAAGTGTTTGCTCTG | 25 | 59.54 | 113 |
| | R14 | AGTAAGTGACAAAGGTCTCCCAAC | 24 | 59.97 | |
| TET1 | F15 | ACATAAGATAAGGGCAGTGGAAAA | 24 | 60.24 | 123 |
| | R15 | ACGGTCTCAGTGTTACTCCCTAAG | 24 | 60.11 | |
| ZBTB38 | F16 | TATGAAAATGCACGAGAAAACAGT | 24 | 60.06 | 121 |
| | R16 | TCAGGGTGTTTACAGTATCTTGGA | 24 | 59.93 | |
| MECP2 | F17 | AAGCTTAAGCAAAGGAAATCTGG | 23 | 60.25 | 108 |
| | R17 | GAAGTACGCAATCAACTCCACTTT | 24 | 60.90 | |
| NTHL1 | F18 | ATGAGGAACAAAAAGGATGCAC | 22 | 60.36 | 101 |
| | R18 | GTCTGCTTGATGTATTTCACCTTG | 24 | 60.06 | |
| MBD4 | F19 | GTACTTTATCAGCCCACAAGGACT | 24 | 59.96 | 91 |
| | R19 | TGGCTTAAGAGAAGTCTCTCCATT | 24 | 59.92 | |
| ZBTB33 | F20 | ACAAGGAAGTGAAAAATTGTTGGT | 24 | 60.18 | 106 |
| | R20 | AGAACTGACATTTGGTGGTGTAGA | 24 | 59.96 | |
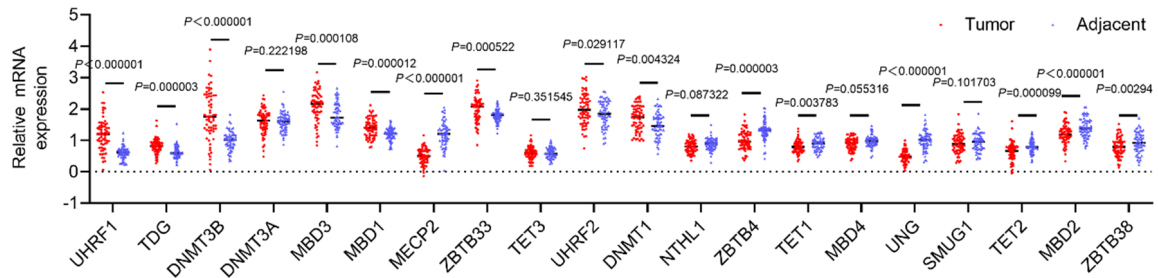


Figure S1. qRT-PCR of DNA methylation regulators in 64 pairs of ccRCC clinical samples.