Cancer Cell International

PRIMARY RESEARCH

Open Access

DNA methylation-based classification and identification of renal cell carcinoma prognosis-subgroups

Wenbiao Chen[1], Jia Zhuang[3], Peizhong Peter Wang[4], Jingjing Jiang[1], Chenhong Lin[1], Ping Zeng[1], Yan Liang[1], Xujun Zhang[1], Yong Dai[2]* and Hongyan Diao[1]*

Abstract

Background: Renal cell carcinoma (RCC) is the most common kidney cancer and includes several molecular and histological subtypes with different clinical characteristics. The combination of DNA methylation and gene expression data can improve the classification of tumor heterogeneity, by incorporating differences at the epigenetic level and clinical features.

Methods: In this study, we identified the prognostic methylation and constructed specific prognosis-subgroups based on the DNA methylation spectrum of RCC from the TCGA database.

Results: Significant differences in DNA methylation profiles among the seven subgroups were revealed by consistent clustering using 3389 CpGs that indicated that were significant differences in prognosis. The specific DNA methylation patterns reflected differentially in the clinical index, including TNM classification, pathological grade, clinical stage, and age. In addition, 437 CpGs corresponding to 477 genes of 151 samples were identified as specific hyper/hypomethylation sites for each specific subgroup. A total of 277 and 212 genes corresponding to DNA methylation at promoter sites were enriched in transcription factor of GKLF and RREB-1, respectively. Finally, Bayesian network classifier with specific methylation sites was constructed and was used to verify the test set of prognoses into DNA methylation subgroups, which was found to be consistent with the classification results of the train set. DNA methylation-based classification can be used to identify the distinct subtypes of renal cell carcinoma.

Conclusions: This study shows that DNA methylation-based classification is highly relevant for future diagnosis and treatment of renal cell carcinoma as it identifies the prognostic value of each epigenetic subtype.

Keywords: Renal cell carcinoma, DNA methylation, Prognosis subgroups, Molecular subtypes

Background

Renal cell carcinoma is the most lethal urologic malignancies and accounts for 3% of all malignant tumors. In recent decades, the incidence of RCC has been rising due to the advances in ultrasound, computed tomography and magnetic resonance imaging used for the early detection of renal tumors. Although recent molecular targeting therapy has improved the overall survival of RCC patients, long-term prognosis remains precarious [1, 2]. Previous, pathological assessments determined the risk of recurrence and differentiation in RCC patients, thus guiding the diagnosis and treatment of recurrent diseases. The classification and diagnostic criteria of RCC are indeterminate and have a high interobserver bias. As a result, 30% of patients develop metastatic disease after treatment, with a median survival period of 1 year [3]. There are growing pieces of evidence that even on the

*Correspondence: daiyong22@aliyun.com; diaohy@zju.edu.cn
[1] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, 79 Qing Chun Road, Hangzhou, China
[2] Clinical Medical Research Center, Shenzhen People's Hospital, The Second Clinical Medical College of Jinan University, 1017 Dongmen North Road, Luohu District, Shenzhen, Guangdong, China
Full list of author information is available at the end of the article

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 2 of 14

basis of the same RCC histological subtype, there is an inconsistent representation of sets of histologically and molecularly heterogeneous diseases [4]. With further understanding of the morphology, immunohistochemistry, molecular and epidemiological features of RCC, the World Health Organization (WHO) revised the classification of renal cell tumors in 2016. It mainly focuses on the molecular pathological features of renal cell tumors [5]. To this end, a more detailed and comprehensive understanding of the molecular classification characteristics of RCC is required.

Currently, the analysis of molecular characteristics has shown that the RCC has obvious specificity. Since molecular subtype supplements histopathology, invaluable insights can be drawn from integrated classifications, which could refine and impact the future application of precision-focused, personalized clinical management of renal cell carcinoma [6]. Ricketts et al. performed a comprehensive genomic and phenotypic analysis on RCC in order to identify the common and specific molecular characteristics and laid the foundation for the development of subtype-specific treatment and management strategies for these cancer patients [7]. Wu et al. used an unsupervised Consensus Clustering algorithm to identify three distinct molecular subtypes of clear renal clear cell carcinoma (the most common subtype) on the basis of hierarchical clustering. Subtypes with poor prognosis were irregularly up-regulated in focal adhesion and cytoskeleton-related pathways, and the expression of core genes, including *LIMK1*, *COL5A1*, *MMP9* and *CCL26*, were negatively associated with the prognosis of the patients [8]. To produce a characteristic biomarker to determine the high and low risk of clear renal cell carcinoma, Samira et al. developed a 34 genes subtype predictor that divided 54 patients with metastatic RCC into good and poor follow-up groups. A model containing subtype-inclusive model was established to analyze the survival outcome of the patients [9, 10].

DNA methylation, one of the most common epigenetic modifications, plays a crucial role in the regulation of the structure and expression of genes and is involved in a variety of biological processes. In addition, changes in DNA methylation are inextricably linked to all aspects of cancer genomics and have been shown to be closely related to changes in the gene expression, sequence and copy number [11]. In addition, DNA methylation often occurs on the C (cytosine) of 5′-CpG-3′ and as a result generates 5-methyldeoxycytidine (5mC). DNA methylation has become an important research topic in epigenetics and epigenomics because of the close relationship between aberrant DNA methylation and physio-pathologic mechanisms underlying an array of human diseases, especially CpG

island methylation. CpG-rich regions (CpG islands) are found in about 40–50% of promoter regions of human genes. Abnormal CpG island hypermethylation of tumor suppressor genes and hypomethylation of oncogenes are vital in carcinogenesis [12, 13]. Therefore, as a promising molecular marker of RCC, abnormal DNA methylation appears in early detection, prognosis prediction, molecular classification and therapeutic targets [14]. Previous studies have shown that the increase in promoter hypermethylation frequency is associated with higher stage and grade in clear cell RCC, according to The Cancer Genome Atlas (TCGA) Research Network for cancer genome maps [15]. Ana et al. identified groups such as *OXR*, *MST1R* and *HOXA9* promoter methylation that positively identified renal cell tumors and differentiation of subtypes. In addition, the groups may improve risk stratification in patients with small renal masses, helping clinicians to determine the best treatment strategy [16]. Gabriel et al. found that the DNA methylation profiles of RCC subtypes can be divided into two main epigenetic clusters: one consists of clear-cell RCC, papillary RCC, mucinous and spindle cell carcinomas and translocation RCC; the other includes oncocytoma and chromophobe RCC [17]. Moreover, an epigenetic chart of the 56 genes that identify the ontogeny of renal cells can predict the results of clear cell RCC. However, their classification may not be detailed enough and has not yet been fully tested for specific sites related to each category.

In this study, our aim is to fully identify the entire DNA methylated profiles of RCC from the TCGA database in order to identify the biological and clinical-related subgroups. Our datasets and the accompanying classification schemes may help to identify new markers or molecular subtypes of RCC in order to accurately subdivide patients with renal cell carcinoma. Furthermore, the RCC classification based on integrated profiling combined with histology and DNA methylation profiling analysis provides a more accurate prediction of clinical behavior compared to the histology alone and provides higher risk assessment accuracy for individual patients. In addition to the identification of clinical-related groups and a new classification basis, our datasets can provide an in-depth understanding into the pathogenesis of RCC and guide accurate decision-making process during diagnosis and treatment.

Methods

Date preprocessing and DNA methylation loci in RCC

A total of 485 RCC samples containing DNA methylation data generated from the Illumina Infinium Human-Methylation450 Bead-Chip array of TCGA [18] were

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 3 of 14

downloaded by UCSC Cancer Browser [19] on June 25, 2018. RCC samples with more than 30 days of follow-up were selected to match the methylation profile. Ultimately, 307 RCC samples were selected for methylation analysis. Methylation level of probes with missing data in more than 70% of the samples were removed. Based on the discovery of cross-reactive probes and polymorphic CpGs by the Illumina Infinium HumanMethylation450 microarray, CpG sites of cross-reactive probes in the genome were removed. The other unidentified probes were estimated using the k-nearest neighbor (KNN) imputation approach in the sva R package [20]. Unstable genomic sites, which contained CpGs in sex chromosomes and single nucleotide polymorphisms were discarded. Finally, 208,022 CpG sites were obtained for further analysis. This study was approved by the Clinical Research Ethics Committee of College of Medicine, Zhejiang University.

Next, 307 RCC samples were randomly divided into two groups, a train group and a test group. In both groups, the age distribution, clinical stage, follow-up time and mortality were similar. Finally, 307 RCC samples were divided into 153 experimental samples in train group and 154 verified samples in test group.

Univariate COX model and multivariate COX proportional risk regression model analysis

To obtain a specific prognostic subgroup of molecular subtypes of RCC, CpG sits of DNA methylation that affect the survival were used as the classification feature. Firstly, based on the methylation level of each CpG site, the survival coxph function R package was used to establish a univariate COX proportional risk regression model. Next, significant CpG sites were selected for multivariate COX proportional risk regression model. Moreover, the TNM classification, pathological grading, and clinical staging were used as covariates to identify independent prognostic factors. For each CpG site, the formula for multivariate COX proportional risk regression model was:

$$h(t, x)_i = h0(t) \exp(\beta_{methylation} methylation_i$$
$$+ \ \beta_{TNM} TNM + \beta_{grade} grade + \beta_{stage} stage)$$

where $h(t, x)_i$ is a risk function associated with covariates at time t (TNM classification, pathological grade, clinical stage). Methylation$_i$ is the vector of the level of CpG$_i$ methylation in the sample, TNM, grade and stage represent their respective vector values. β stands for regression coefficient of methylation, TNM, grade, and stage. The *p* values of the Cox regression coefficients were calculated

and corrected by multiple test correction using the method of Benjamini and Hochberg false discovery rate.

Selection of molecular subtypes by consistent clustering

The K-means clustering algorithm in the ConcensusClusterPlus R packet [21] was used for consistent clustering to determine tumor subgroups based on the most variable CpG sites. The K-means clustering algorithm was designed to classify a pre-specified dataset into k clusters. This algorithm defined "consensus" clustering by measuring the stability of clustering results from a given clustering method applied to a random subset of data. At each iteration, 80% of the tumor samples were selected, the K-means algorithm and Euclidean square distance measure were used for groups of k ranging from 2 to 20. More than 100 iterations were carried out and the stability of each cluster was determined. The maximum number of clusters with at least 90% cluster consensus were chosen.

The optimal cluster number was determined by the cumulative distribution function and the delta area plot. We considered the optimal number of clusters should be that the consistency of the cluster was relatively high, the coefficient of variation was relatively low, and the area under the CDF curve did not increase significantly. The number of categories was selected based on the relative non-significant change in the area under CDF curve. The pheatmap R package was used to construct the consensus cluster heatmap.

Analysis of the clinical characteristics of molecular subtypes

The Kaplan–Meier method was used to construct the overall survival curve for RCC subgroups defined by DNA methylation spectrum and the statistical differences between clusters were determined by the log-rank test. R Bioconductor survival package was used for survival analysis, and Chi square test was used to comprehensively analyze the association between clinical and biological characteristics of DNA methylation clusters.

Identification of specific DNA methylation markers and transcription factor enrichment of specific methylation site annotated genes

To identify the molecular types of methylation-based RCC, the Epidiff software was applied to reveal specific CpG sites that defined specific DNA methylation. EpiDiff is an integrated software platform that provides bioinformatic analysis on quantitative differential chromatin modification region (QDCMR), quantitative differentially methylated region (QDMR), and quantitative differential expressed gene (QDEG). The free platform and free species nature of EpiDiff makes it suitable for unprecedented epigenome profiles described by high-throughput

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 4 of 14

experimental techniques using microarrays and next-generation sequencing [22]. To identify specific hypermethylation or hypomethylation of DNA CpG sites within specific RCC subgroups, QDMR software (http://bioinfo.hrbmu.edu.cn/qdmr) was used to quantify methylation differences and identify differentially methylated regions in multiple RCC samples [23]. Next, the significant CpG sites across the classified subgroups were used to find specific CpGs in each subgroup. For each subgroup, the average DNA methylation level for each sample with significant CpGs was calculated and the matrix with 3389×6 dimensions was input into the QDMR. The threshold of standard deviation parameter was set to 0.04.

DNA methylation at the promoter site can promote or inhibit gene expression. To explore the mechanism of the specific CpG sites, the transcription factor (TF) enrichment analysis of the corresponding genes at the promoter region of specific CpG sites was carried out by g:Profiler. G:Profiler is a simple, user-friendly Web interface with powerful visualization abilities to capture gene ontologies, pathways, or transcription factor binding sites, or even individual gene levels [24]. The genes corresponding to the TF enrichment promoters were displayed by Cytoscape.

Construction of Bayesian classifier and verification of test set data

In order to verify the distinguished ability of specific CpG sits, specific CpG sites which were identified by Epidiff software were constructed by Bayesian classifier. The samples in the test set were assigned to the corresponding subgroups using this classified model. The performance of the model was verified by tenfold cross-validation method, and the accuracy of the model was evaluated using accuracy rate. The pROC packet in R Program was used to establish a receiver Operation characteristic Curve (ROC). To verify the stability and reliability of the model, a prognostic model similar to the train group was used to assign a class label to 154 test samples.

Results

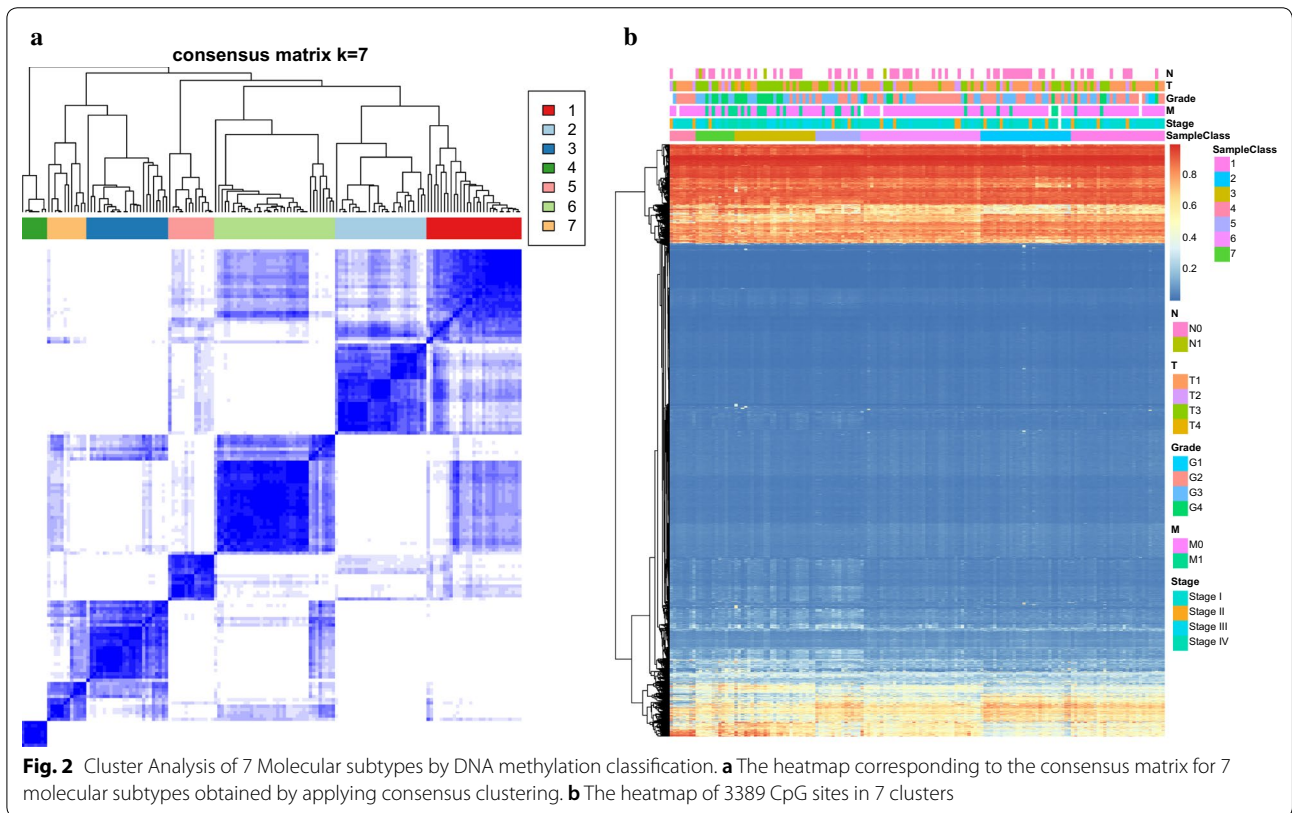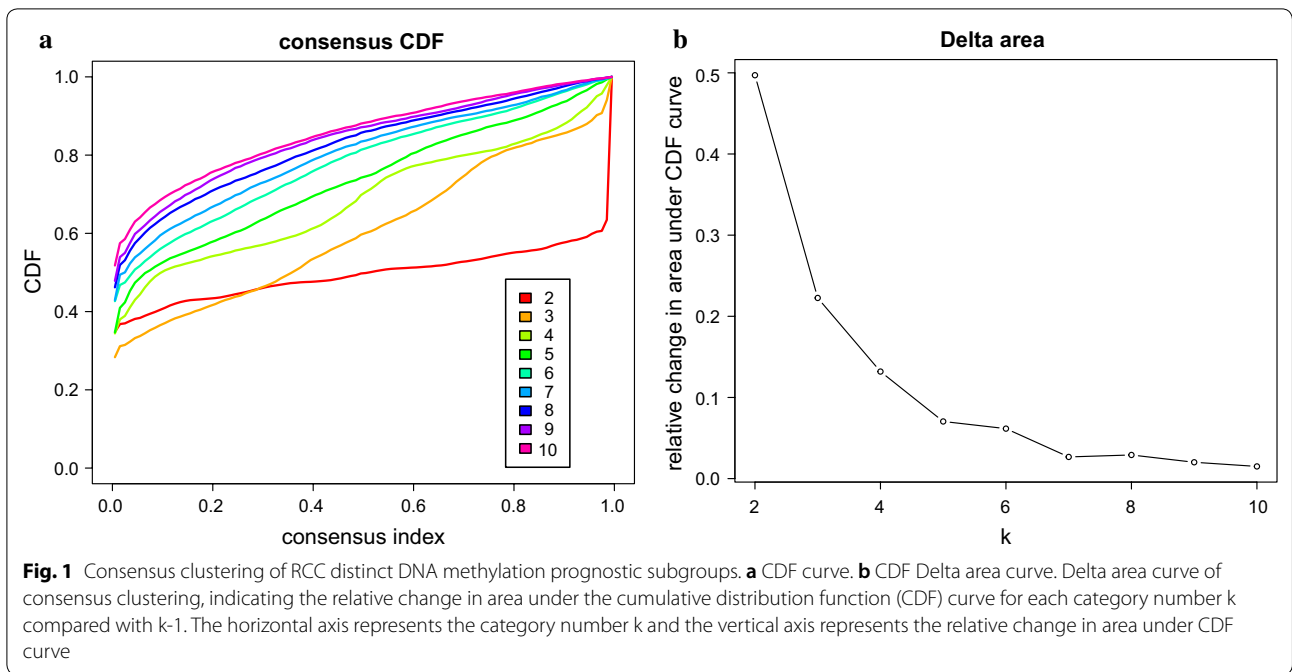### Characteristics of RCC samples evaluated for DNA methylation

The DNA methylation profiles of RCC samples from the TCGA database were used to cluster the RCC prognostic molecular subgroups. In total, 208,022 CpG sites were successfully screened for methylation in 307 RCC samples. For each of the CpG that sites in the train set, univariate Cox proportional hazards regression model was applied to each methylation site and the survival data recorded. 52,878 significant CpG sites were identified that could influence patient survival ($p < 0.05$). T ($p = 1.32483e-06$), N ($p = 0.00095$), M ($p = 1.32483e-06$), Stage ($p = 1.787979e-06$), and Grade

($p = 6.581784e-07$) were also found to be significant factors based on the univariate Cox proportional risk regression models analysis. Then, 52,878 significant CpG sites were introduced into the multivariate Cox proportional risk regression models, including TNM classification, pathological grade, clinical stage as covariates to investigate the independent prognostic factors. Finally, we got 3389 significant CpG sites for further prognosis subgroups analysis.

Consensus clustering of DNA methylation of RCC identified prognosis subgroups

In order to screen distinct DNA methylation prognostic molecular subtypes of RCC, consistent clustering of 3389 significant CpG sites were constructed. Then, the average cluster consensus and the coefficient of variation among clusters were calculated for each category number to obtain the optimal number of classes. According to the CDF curve, the clustering result is relatively stable when the cluster is 7 or 8 (Fig. 1a). To further observe the CDF delta area curve, it can be seen that cluster has a stable clustering result when it is selected as 7, as the area under the CDF curve begins to stabilize after 7 categories (Fig. 1b). Thus, 7 was regarded as the optimal number of categories for further analysis. According to the result of consistent clustering, we selected the stable clustering result of 7 to construct a consensus matrix graph, which was a better visualization tool in the assessment of the composition of clustering and number. The color-coded heatmap is seen to correspond to the consensus matrix, 153 RCC samples were assigned to 7 categories. Additionally, a color-coded heatmap characterized by blue blocks along the diagonal on a white background revealed a well-defined 7-block structure (Fig. 2a). Further, the heatmaps that corresponded to 3389 CpG sites were generated by the heatmap function with DNA methylation classification with TNM classification [25], pathological grade [26], clinical stage [27], and age [28] as the annotations. It can be seen that the abundance of most CpG sites in each sample is low. However, there were obvious differences found in DNA methylation profiles among these 7 categories (Fig. 2b).

The Kaplan–Meier plot showed the survival of RCC defined by the methylation-based consensus clustering had a significant difference amount the 7 clusters (Fig. 3a). Among them, clusters 1 and 2 had the best prognosis, whereas clusters 3 and 7 had the worst prognosis. Understanding the TNM classification, pathological grade, clinical stage, and age can aid clinicians in choosing the best treatment for patients in their current condition and predict the prognosis. Thus, we analyzed the distribution of TNM classification, pathological grade, clinical stage, and age in the prognosis of

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 5 of 14

Fig. 1 Consensus clustering of RCC distinct DNA methylation prognostic subgroups. **a** CDF curve. **b** CDF Delta area curve. Delta area curve of consensus clustering, indicating the relative change in area under the cumulative distribution function (CDF) curve for each category number k compared with k-1. The horizontal axis represents the category number k and the vertical axis represents the relative change in area under CDF curve

Fig. 2 Cluster Analysis of 7 Molecular subtypes by DNA methylation classification. **a** The heatmap corresponding to the consensus matrix for 7 molecular subtypes obtained by applying consensus clustering. **b** The heatmap of 3389 CpG sites in 7 clusters

7 molecular subtypes (Fig. 3b–g). Some of the DNA methylation subgroups reflected differently in the TNM classification, pathological grade, clinical stage and age
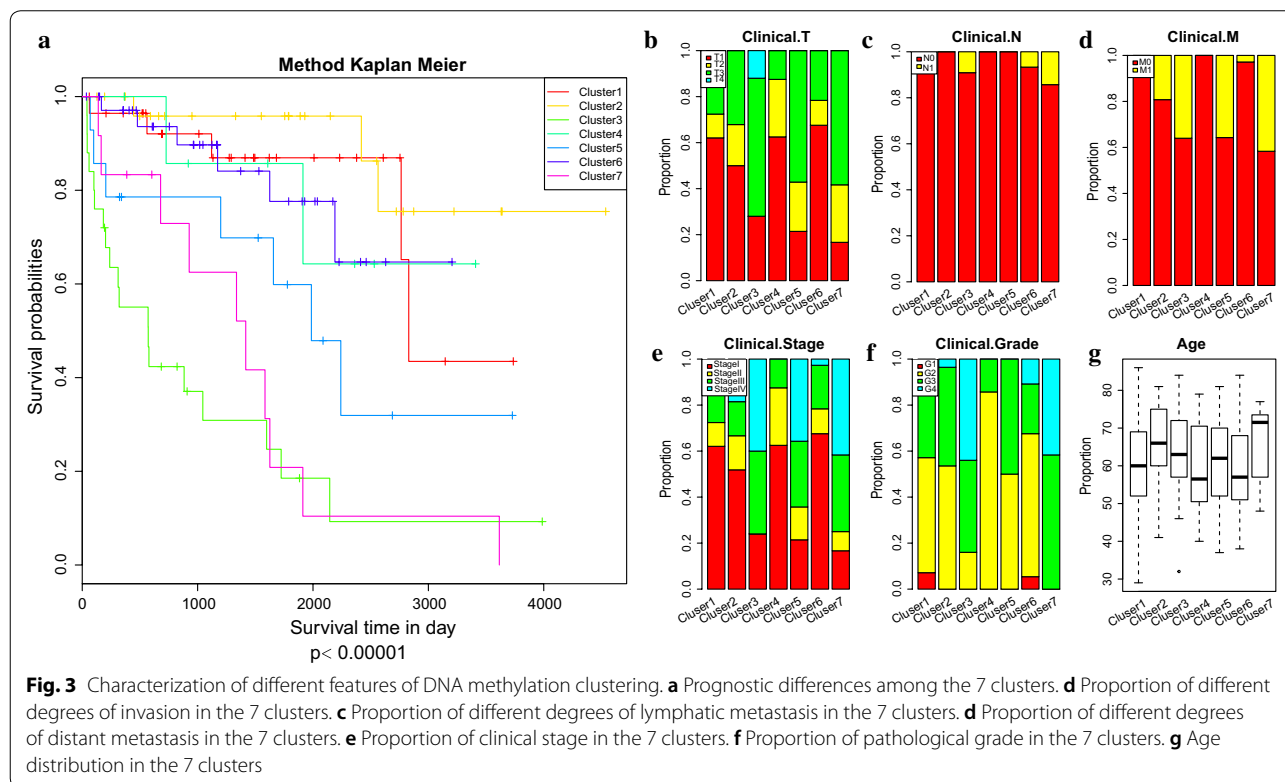
subtypes, whereas different DNA methylation subgroups were also found in the same TNM classification, pathological grade, clinical stage and age subtypes. These

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 6 of 14

results demonstrated that patients in different TNM classification, pathological grade, clinical stage and age subtypes share the same DNA methylation characteristics and specific subtypes of TNM classification, pathological grade, clinical stage and that age contained different DNA methylation characteristics. Also, we found that the DNA methylation status represents a more elaborate classification analysis for RCC and also that DNA methylation can serve as a commendable biomarker for RCC classification. For example, clusters 1 and 6 had low invasiveness (Fig. 3b), clusters 3 and 7 had high relevance with lymphatic metastasis and distant metastasis (Fig. 3c, d), and clusters 3 and 7 connected with high pathological grade and high clinical stage (Fig. 3e, f). This gave a detailed description of the reason for the clusters 3 and 7 having the worst prognosis, as clusters 3 and 7 were more likely to metastasize and develop into a higher degree of malignancy, even though they had the same underlying aetiology, such as DNA methylation abnormalities. Also, it might explain the reason for the low TNM classification (such as T1 in cluster 3, primary tumor which is limited to kidney, maximum diameter ≤ 7 cm) still had higher degree of malignancy (such as G4 in cluster 3, tumor with low differentiation or no differentiation) and had higher chance of deterioration (such as stage 4 in cluster 3, tumor disease in terminal stage), as DNA methylation changes in carcinogenesis. These results indicate that these DNA methylation profiles can help to understand the etiology of RCC and, most importantly, they demonstrated clinically applicable biomarkers for use in the early detection of kidney cancer.

Clinical and biological characteristics with DNA methylation clustering

Next, we investigated whether the DNA methylation clustering of RCC was relevant to clinical and biological characteristics. The clinical information of RCC was obtained from the TCGA Database. Then, we analyzed the clinical characteristics of tumor/normal, TNM classification, pathological grade, clinical stage, age, and gender in each group. The Chi square test was used to analyze the global associations between clinical and biological characteristics of DNA methylation clustering. Table 1 shows the tumor/normal, TNM classification, pathological grade, clinical stage, age was significantly different among the DNA methylation prognosis clusters ($p < 0.05$), but no significant difference was detected in gender ($p > 0.05$). This finding implied there was evidence of a strong association between the DNA methylation prognosis clustering and the clinical and biological characteristics.

Fig. 3 Characterization of different features of DNA methylation clustering. **a** Prognostic differences among the 7 clusters. **d** Proportion of different degrees of invasion in the 7 clusters. **c** Proportion of different degrees of lymphatic metastasis in the 7 clusters. **d** Proportion of different degrees of distant metastasis in the 7 clusters. **e** Proportion of clinical stage in the 7 clusters. **f** Proportion of pathological grade in the 7 clusters. **g** Age distribution in the 7 clusters

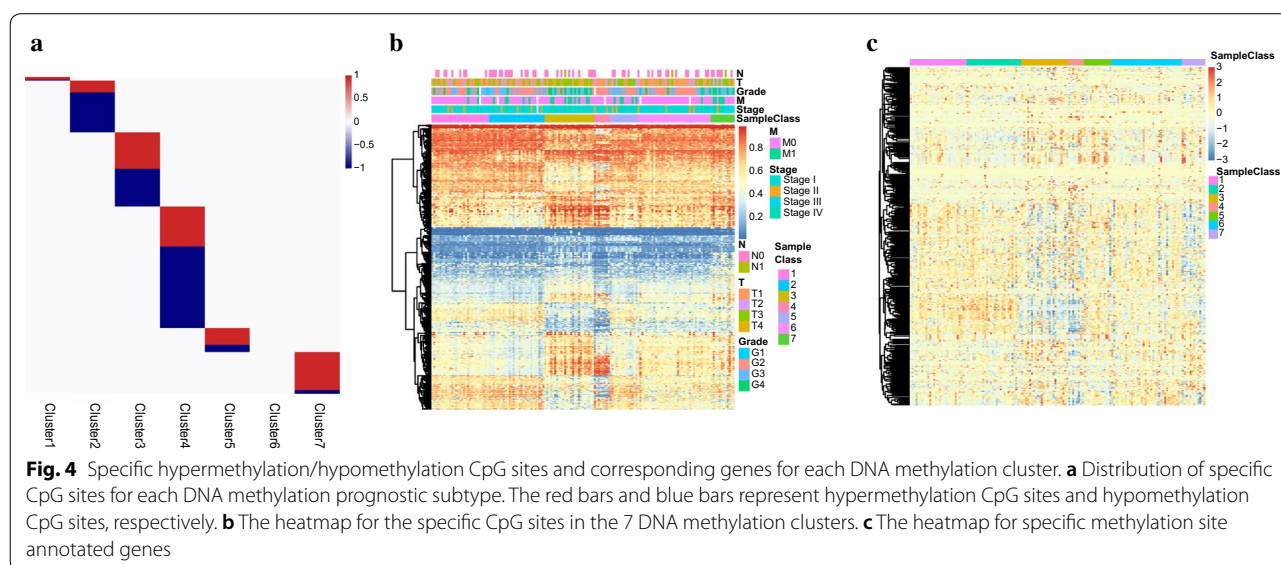## Identification of specific DNA methylation markers and corresponding genes

To find the specific DNA methylation, CpGs were specifically hypermethylated or hypomethylated within

Table 1 The associations between clinical and biological characteristics with DNA methylation clustering on Chi square test

Clinical attributes	Subclasses	P-value
Cancer_normal	Cancer	0.0003
	Normal	
T	T1	0.0019
	T2	
	T3	
	T4	
N	N0	0.0026
	N1	
M	M0	
	M1	
Stage	S1	0.0000
	S2	
	S3	
	S4	
Grade	G1	0.0000
	G2	
	G3	
	G4	
Age	Young (age < 60)	0.0292
	Old (age > 60)	
Sex	Male	0.1722
	Female	

particular RCC subgroups. We used EpiDiff software to identify the specific methylation sites and the QDMR software to quantify the methylation difference. Eventually, 437 of 3389 methylation sites were identified as cluster specific methylation sites. Specific hyper/hypomethylation CpG sites for each DNA methylation cluster are represented in Fig. 4a, which depicts cluster 4 having the largest number of specific methylations CpGs, most of which are hypomethylated. While others have a small number of specific methylations CpGs, most of them are hypermethylated. Nevertheless, no specific methylation CpGs was detected in cluster 6. Furthermore, the heatmap corresponding to the 437 specific methylation sites was constructed (Fig. 4b). The TNM classification, pathological grade and clinical stage, corresponding to 7 clusters of 437 specific methylation sites were evidently different, which suggested these were DNA methylation markers for the different subgroups in RCC.

We annotated these specific methylation sites for gene annotation and 437 specific methylation sites corresponding to 477 genes were identified. Besides, we explored the gene expression within the particular DNA methylation subgroups. Finally, the expression values in 151 samples of the train set for 436 of the 477 genes were obtained. Further, the heatmap for the expression profile of specific methylation site annotated genes was depicted (Fig. 4c). It can be seen that there were different expression patterns of these subgroups at the standard of expression profile, which indicated the specificities of these genes were consistent with the DNA methylation level and the gene expression level.

Fig. 4 Specific hypermethylation/hypomethylation CpG sites and corresponding genes for each DNA methylation cluster. **a** Distribution of specific CpG sites for each DNA methylation prognostic subtype. The red bars and blue bars represent hypermethylation CpG sites and hypomethylation CpG sites, respectively. **b** The heatmap for the specific CpG sites in the 7 DNA methylation clusters. **c** The heatmap for specific methylation site annotated genes

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 8 of 14

## TF enrichment of specific methylation site annotated genes

To observe the mechanism of the 437 specific methylation sites, the genes, corresponding to the specific methylation sites in the promoter region were subjected to TF enrichment analysis. It was found that these genes were enriched with two transcription factors, GKLF and RREB-1 with significant false discovery rate (FDR) values of 0.00555 and 0.00173, respectively (Fig. 5). Besides, 277 genes were enriched with these two TFs, of which 212 and 221 were enriched with GKLF and RREB-1, respectively. Previously, many researchers have revealed that GKIF is a putative tumor suppressor gene epigenetically silenced in RCC by promoter CpG methylation [29–31]. These outcomes suggest that the genes that correspond to the 437 specific methylation sites in the promoter region of enrichment of GKLF and RREB-1, have potential values of prognosis and therapy for RCC.

Validation of the prognosis prediction model

In order to verify the ability of discrimination of specific CpG sites for each prognosis subgroups in the set, we constructed a Bayesian classifier to determine the function of the prognosis prediction model, with 437 specific CpG sites as characters. The accuracy of the prognosis prediction model based on the train set was 81.7%. The area under the ROC curve reached 0.9455 (Fig. 6a). Then, we employed the prognosi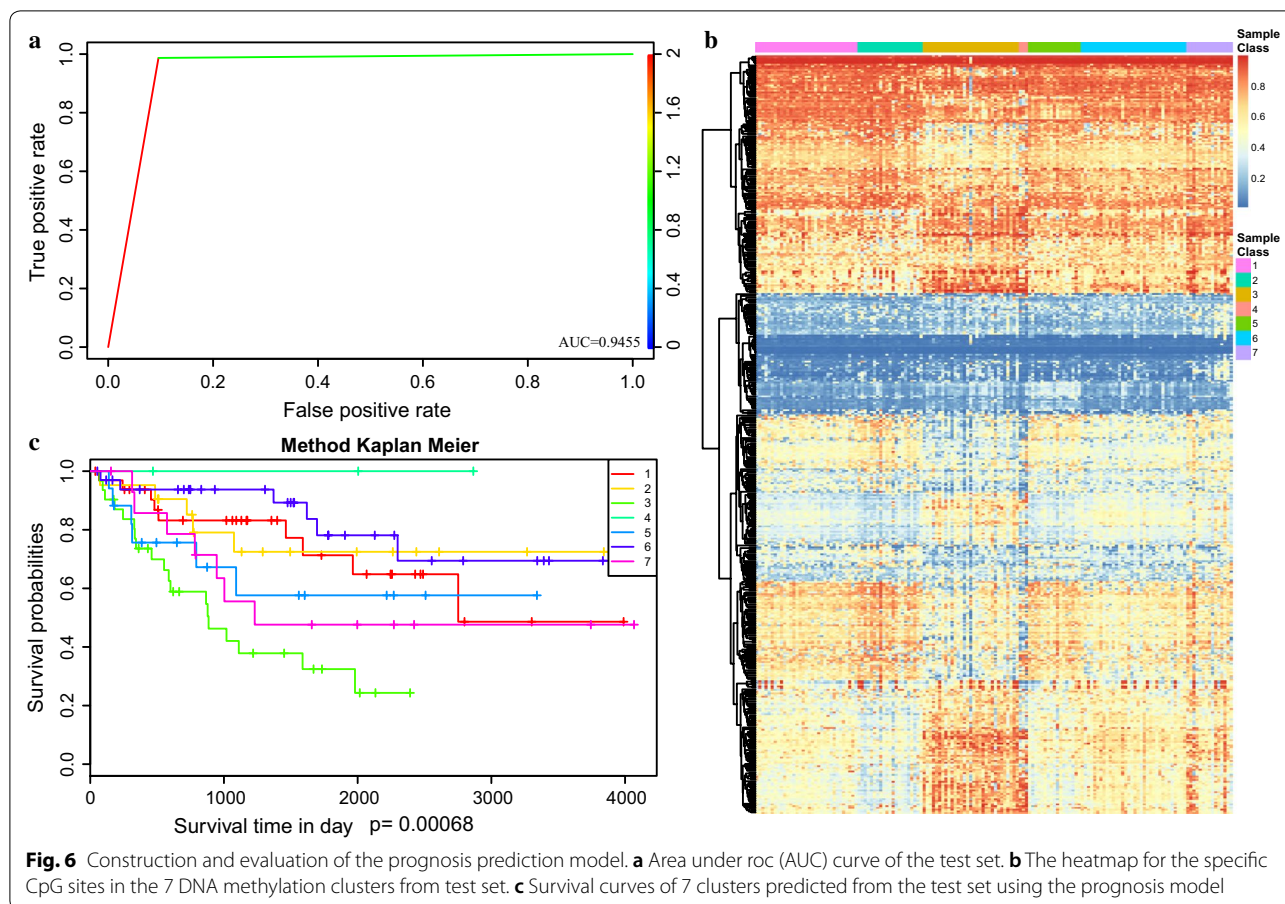s model to predict cases in the test set. We extracted the expression profile data of 437 specific CpG sites from the test set and put them into a prognosis model for verification. The number of statistics in the test set was assigned to a similar class corresponding to the train set. Also, the heatmap corresponding to the 437 specific CpG sites in the test set was generated by the heatmap function using the prognosis model constructed from the train set. The results showed that there were also significant differences in the methylation patterns between the 7 categories of the prognosis prediction model, which was consistent with the train set (Fig. 6b). Moreover, the survival analysis of the 7 clusters in the test set showed that they were a significant prognostic difference (p < 0.00068) (Fig. 6c). The prognosis of cluster 2 was significantly better than the others, which is consistent with the train set.

Next, we analyzed the distribution of TNM classification, pathological grade, clinical stage, and age in the prognosis of 7 molecular subtypes. A similar method as the train set was used and the relative consistent results were obtained (Fig. 7). In particular, clusters 1 and 4 had low invasiveness (Fig. 7a), clusters 3 and 6 had high relevance with lymphatic metastasis (Fig. 7b), clusters 3 and 7 had high relevance with distant metastasis (Fig. 7c), and clusters 3 and 7 connected with high pathological grade and high clinical stage (Fig. 7d, e). These results further illustrate the predictive accuracy of our model and the stability of its features.
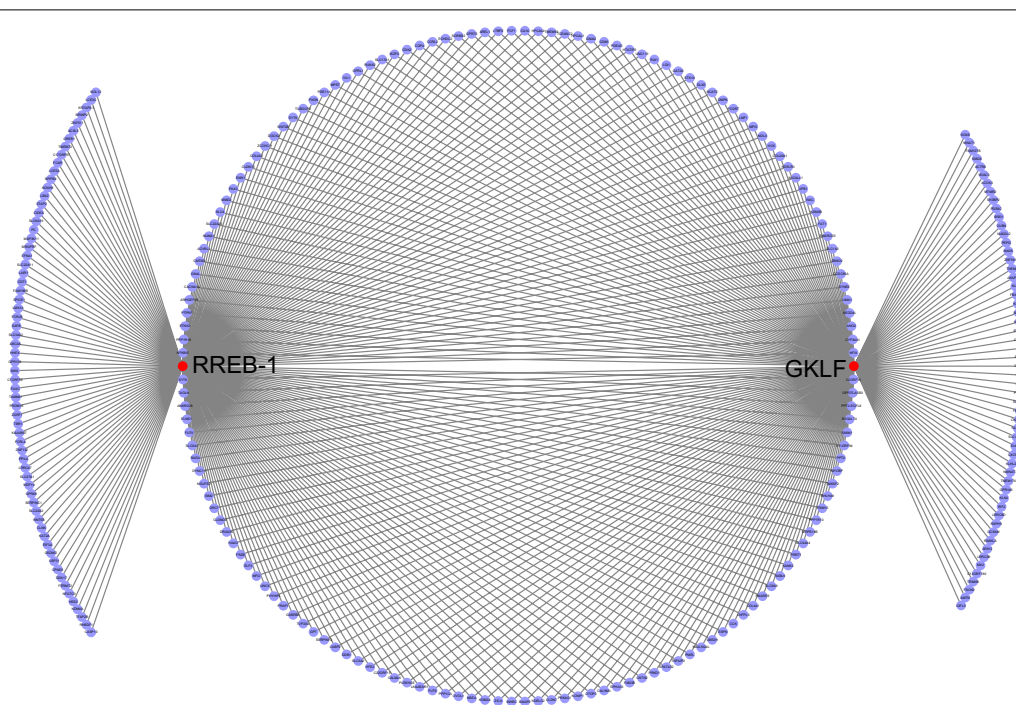
Fig. 5 TF enrichment analysis of genes in the promoter methylation region. The red nodes and purple node represent TF and genes, respectively

Fig. 6 Construction and evaluation of the prognosis prediction model. **a** Area under roc (AUC) curve of the test set. **b** The heatmap for the specific CpG sites in the 7 DNA methylation clusters from test set. **c** Survival curves of 7 clusters predicted from the test set using the prognosis model
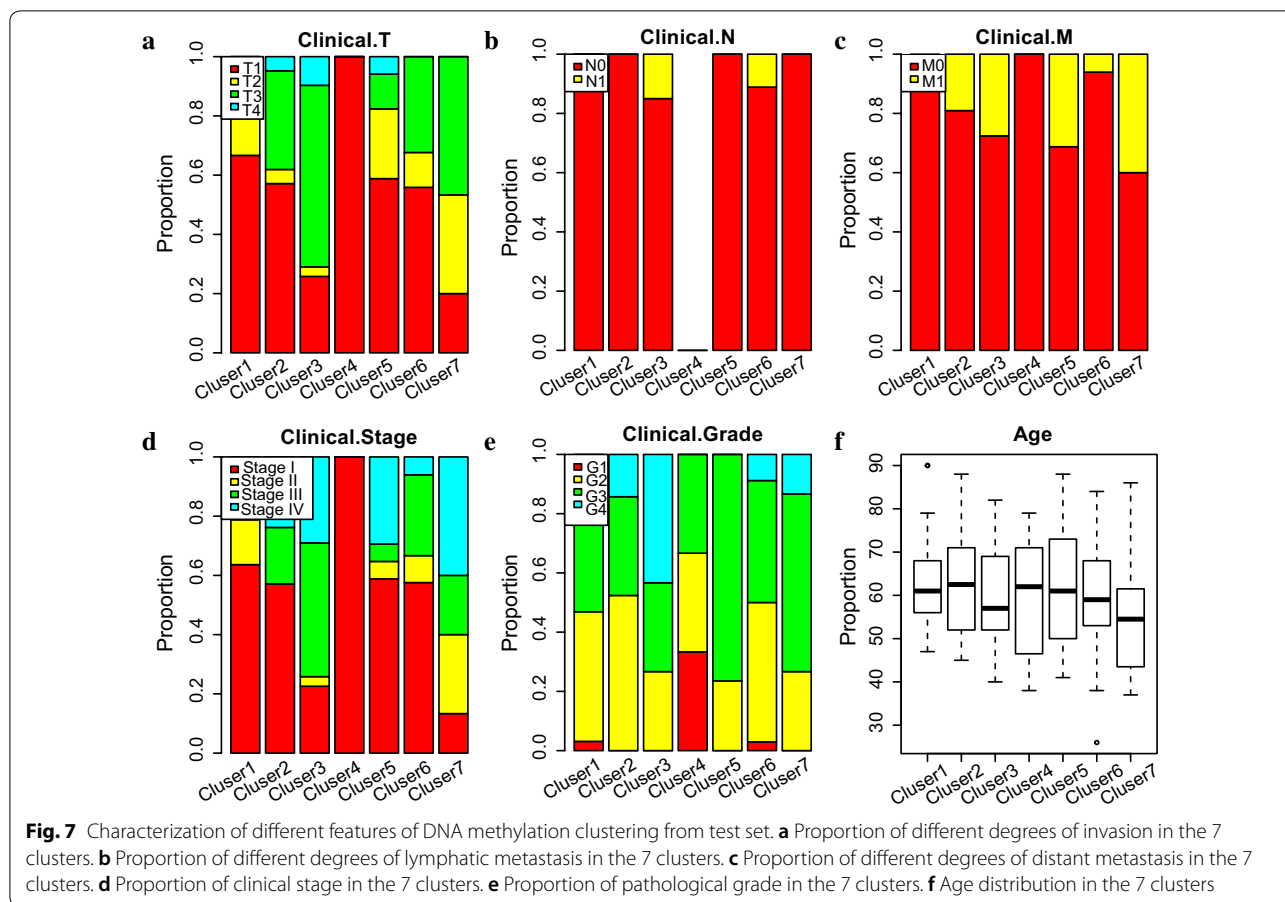
To examine whether the prognoses of clustering in the test set was similar to the corresponding clustering in the train set, we compared the prognostic relationship of 7 clusters between the test set and train set. There was no significant difference in the prognosis between the test set and the train set (Fig. 8). These results show that the cases classified or predicted to be in the same DNA methylation subgroups had the same prognosis.

Comparison between 7 prognosis subgroups and 34-gene classifier (ClearCode34)

34-gene classifier (ClearCode34), which was a authoritatively subtyped predictor for classifying clear cell RCC according to good risk (ccA) and poor risk (ccB) subtype for assigning clear cell RCC patient and built a subtype-inclusive model to analyze patient survival outcomes according to risk stratification [9, 32]. The ClearCode34 classifier was applied for RNA-sequencing of data from 380 nonmetastatic clear cell RCC samples from TCGA. In the meantime, a total of 199 RCC samples in our study were selected to compare 7 prognosis subgroups and ClearCode34 classification results (Additional file 1: Table S1). From which it can be seen that C l and C 2 are

mainly concentrated in ccA, whereas C 3, C 4, and C 7 mainly enriched in the ccB. ccA and ccB subtypes represented good prognosis and poor prognosis, respectively, which was consistent with our study that cluster 1 and cluster 2 have the best prognosis, whereas cluster 3 and cluster 7 have the worst prognosis (Fig. 3a). Interestingly, the ccA and ccB appeared in almost equal among patients in the C 6 subgroup, and therefore we compared the prognosis of ccA and ccB in the C 6 subgroup (Additional file 2: Figure S1). The results revealed that the prognosis of ccA was significantly better than the ccB, even for the C 6 subgroup in our study. Further, we compared the clinically important features of ccA and ccB in C 6, including TNM, stage, and grade. The findings revealed that there was significant difference in T 1 (p = 0.024), T 3 ($p = 0.008$), N 1 ($p = 0.038$), N X ($p = 0.003$), stage I ($p = 0.024$), and stage III ($p = 0.008$) between ccA and ccB in the C 6 subgroup (Additional file 3: Table S2). Moreover, most of the patients of ccA and ccB were distributed in T 1, N 0, stage 1, and grade 2, whereas less of the patients of ccA and ccB were distributed in T 2, N 1, stage 5, grade 1 in C 6 subgroup. The results were consistent with the analysis of the distribution of TNM

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 10 of 14

Fig. 7 Characterization of different features of DNA methylation clustering from test set. **a** Proportion of different degrees of invasion in the 7 clusters. **b** Proportion of different degrees of lymphatic metastasis in the 7 clusters. **c** Proportion of different degrees of distant metastasis in the 7 clusters. **d** Proportion of clinical stage in the 7 clusters. **e** Proportion of pathological grade in the 7 clusters. **f** Age distribution in the 7 clusters

classification, pathological grade, and clinical stage in the prognosis of 7 molecular subtypes (Fig. 3b, c, e, f). All these results demonstrate that our 7 prognosis subgroups and ClearCode34 classifier have a high degree of consistency with the RCC prognostic prediction and molecular classification.

Discussion

Alterations in DNA methylation play an important role in the development and progression of RCC as confirmed in recent studies. In addition, the DNA methylated profiles provide insights into the etiology of RCC, especially for clinicians to provide clinically viable biomarkers for early diagnosis and precise treatment of RCC [33, 34]. Therefore, the whole genome DNA methylation spectrum sequencing technique has been widely used to study tumor-specific methylation alterations. Moreover, many different histology and benign adjacent RCC tissues of human or experimental animals have been extensively examined for DNA methylation for diagnosis and treatment. However, its effectiveness is limited due to factors such as high cost and heavy analytical burden and large sample size. Owing to the TCGA database, a

publicly funded project has been prepared that aims at cataloging and discovering major carcinogenic genomic changes in more than 30 large cohorts of human tumors through large-scale genome sequencing and comprehensive multidimensional analysis [35]. In this study, 3389 significant CpG sites, corresponding to 307 RCC samples were obtained from TCGA for the classification analysis. Large sample sizes were used to provide an in-depth understanding of the etiology of RCC, supply proven biomarkers based on specific methylation of CpGs, and present a framework for the development of molecular subtypes based on DNA methylation for the detection of RCC patients.

RCC originates from the epithelial cells of the nephron, the functional unit of the kidney, and consists of several different molecular and histological subtypes. It has different clinical features, including therapeutic responses [36]. Therefore, it poses many challenges in the diagnosis and treatment of RCC, including the early diagnosis of malignant tumors, treatment selection, surgical benefits, side effects of chemotherapeutic drugs and prognosis assessment of patients. Since DNA methylation is one of the earliest molecular changes in cancer and is

Fig. 8 The survival curves of the same labelled clusters in the train set and test set
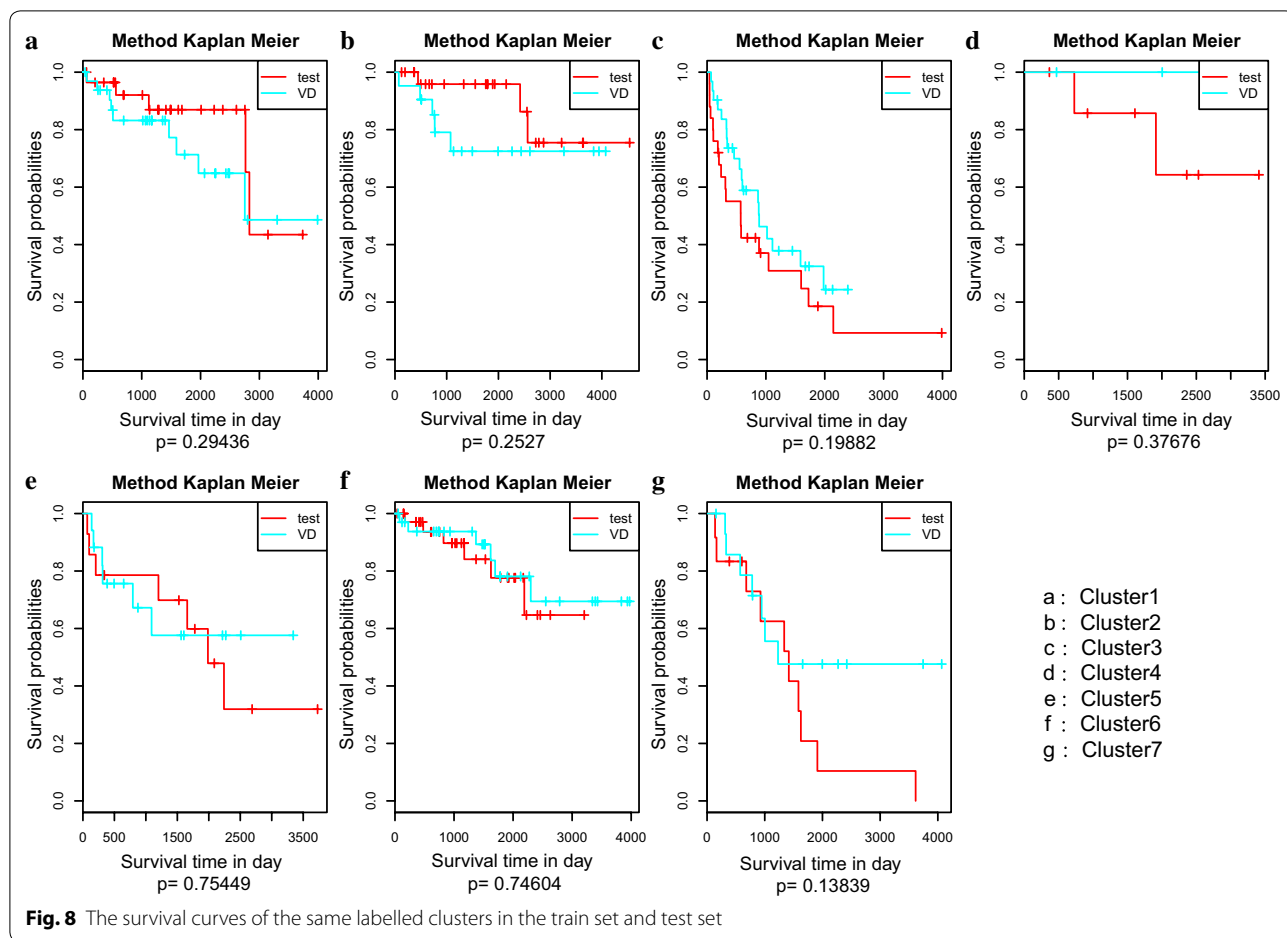
extensive and stable, its role in cancer biology has been extensively studied, including early diagnosis of renal cell carcinoma [37, 38]. Many studies have explored the role of DNA methylation in the etiology and progression of renal cell carcinoma. Clinical researchers have used DNA methylation as a biomarker for early diagnosis, prognosis and prediction, as well as the potential for DNA methylation to guide therapy [39, 40]. Therefore, we carried out this study to obtain a detailed classification of RCC epigenomes based on DNA methylation. Prognostic-related CpG sites were selected according to the classification of TNM, pathological grading, clinical staging, age, and survival analysis. In addition, according to the inherent molecular characteristics of DNA methylation, RCC can be divided into seven subtypes, which corresponds to the significant difference in clinical results and molecular subtypes. Thus, the classification scheme provided a molecular stratification suitable for individual tumors, which has the significance of influencing therapeutic decision-making and defined the biological mechanisms involved in the progression of RCC.

RCC is consists of a heterogeneous group of tumors, and according to the morphological and histological classification, RCC can be classified as clear cell carcinomas (60% of cases), papillary tumors, chromophobic tumors, oncocytomas, and collecting or Bellini duct tumors [41]. TNM stage and Fuhrman grade are still the most commonly used predictors of clinical prognosis in RCC patient. In addition, combined with comprehensive clinical systems such as the Mayo Clinic stage, size, grade and necrosis (SSIGN) score and the University of California Integrated Staging System, clinicians have improved the accuracy of prognosis in RCC patients [42]. However, RCC patients with different cell types originating from renal units, as well as their histological, molecular, genetic and clinical diversity, may have different results with similar clinical characteristics or comprehensive system scores. Therefore, accurate classification of RCC is crucial for prognostic risk stratification, targeted therapy selection, and gene detection and identification. Although the molecular profiling of RCC subtypes has been identified as a unique and diverse spectrum of changes that are used to improve our prognosis and

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 12 of 14

therapeutic effects in the initial stage, their development has been rapid [43, 44]. In this study, we used significant CpG sites to identify 7 distinct prognostic subgroups of RCC that seemed to predict disease-specific survival, as well as TNM classification and pathological grade, clinical stages, and age distribution of prognosis of 7 molecular subtypes. Therefore, the classification scheme provided a molecular stratification suitable for individual tumors, which significantly influenced treatment decision-making and accurate diagnosis. For example, an RCC patient was diagnosed as RCC at T1, which meant a primary tumor was limited to kidney, maximum diameter ≤ 7 cm. Based on the 7 molecular subtypes, when the significant CpG sites of patients were assigned to cluster 7, the patients were found to have high chances of developing lymph nodes, distant metastasis and had poor prognosis. This is necessary for early radical resection of RCC, to assess whether early lymph node dissection is needed to prevent metastasis. Whereas, when the significant CpG sites of patients were assigned to cluster 1, the tumors exhibited low degree of invasion and predicted a good prognosis. This result can prompt clinicians to re-evaluate the treatment of patients, whether to have radical total RCC or partial resection or close clinical follow-up, or chemotherapeutic therapy, which can help alleviate pain and benefit patients. In conclusion, our research on the 7 subgroups at the level of DNA methylation molecular can classify RCC more accurately and guide clinicians when diagnosing, treating and performing prognosis of different epigenetic subtypes.

Previous researches reported that CpG island methylation as an epigenetic form of gene regulation that disrupts the function of tumor suppressor genes or oncogenes and was shown to promote carcinogenesis [45]. Based on the DNA methylation and gene expression profiles in the tumor, we found 437 specific methylation sites corresponding to 477 genes, and they defined the particular DNA methylation subgroups of RCC. Besides, we found that the levels of DNA methylation and gene expression based on 7 subgroups were consistent, which suggested that the changes in methylation in some regions might be associated with RCC by changing the expression of corresponding genes. Moreover, a prognostic model characterized by 437 specific CpGs could be used to divide the test group into different prognostic clusters, which was consistent with the results of the classification. Hypomethylation and hypermethylation of DNA can affect the progression and prognosis of RCC, leading to overexpression of oncogenes and down-regulation of tumor suppressor genes, respectively [46]. On the other hand, abnormal DNA methylation was significantly associated with poor tumor differentiation, tumor invasion and poor prognosis in cancer patients [47]. In our study, we found

that cluster 1 had a small number of specific methylations CpGs and cluster 6 had no specific methylation CpGs, which is a reasonable explanation of clusters 1 and 6 having low invasiveness. Additionally, this might explain the reason for the cluster 1 having the best prognosis on survival analysis.

Increasing evidence has shown that epigenetic regulation, such as abnormal hypermethylation of CpG islands in promoters, is key to tumorigenesis [48]. In our study, the genes, corresponding to 437 specific methylation sites in the promoter region were considered for TF enrichment analysis. We found that the genes in the promoter methylation were enriched with TF of GKLF and RREB-1. Li et al. found that significant reduction or loss of GKLF in RCC was mainly due to promoter methylation abnormalities and that GKLF deletion was associated with the progression of RCC, relatively low overall survival, and disease-free survival [29]. Liu et al. found that RREB-1 regulated the transcription of the p53 gene, which is a tumor suppressor gene, through the core promoter element of p53 after genotoxic stress. The silencing of RREB-1 significantly lowers the expression of p53 mRNA and protein, which can lead to tumorigenesis [49]. Even though there is less literature focusing on the relationship between RCC and RREB-1, some mechanistic studies have suggested that the RREB-1 has a major role in tumor suppression of pancreatic cancer [50], bladder cancer [51], prostate cancer [52]. These outcomes imply that TFs of GKLF and RREB-1 promotes methylation in RCC biomarkers recognition, and highlights their diagnostic and prognostic values, and clinical application.

Conclusion

This study identified a new classification of RCC into 7 prognosis subgroups according to the DNA methylation data, indicating that molecular subtypes are independent prognostic factors of RCC. This provides a more detailed explanation of molecular subtypes, as a simple pathological classification. In addition, this classification will help to identify new biomarkers of RCC and provide more accurate subdivision of RCC. Changes in DNA methylation can be used as markers of molecular profiles associated with standard clinical features, which could significantly enhance the diagnosis of RCC and facilitate individualized treatment for patients. Furthermore, the specific CpG sites and corresponding genes in specific subgroups can be used as biomarkers for early diagnosis, accurate treatment and precise prognosis prediction. Finally, the framework proposed in this work can be used to study new classifications of molecular subtypes associated with specific tumors.

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 13 of 14

Additional files

Additional file 1: Table S1. The distribution of 199 RCC samples based on 7 prognosis subgroups and ClearCode34 classification.

Additional file 2: Figure S1. Prognostic difference between ccA and ccB in C 6 subgroup.

Additional file 3: Table S2. Analysis of clinical features of ccA and ccB in C 6 subgroup.

Abbreviations
RCC: renal cell carcinoma; WHO: World Health Organization; 5Mc: 5-methyl-deoxycytidine; TCGA: The Cancer Genome Atlas; KNN: k-nearest neighbours; QDMR: quantitative differentially methylated region; DMRs: differentially methylated regions; ROC: receiver operating characteristic curve.

Authors' contributions
Conceptualization, YD and HD; Interpretation of data, WC, CL, PZ, XZ, and YL; Formal analysis, PPW, WC and JZ; Funding acquisition, HD; Investigation, WC, JJ, CL, and JZ; Project administration, YD and HD; Visualization, JZ, PZ, and YL; Writing—original draft, WC; Writing—review and editing, PPW, YD and HD; Revision assistance, PPW. All authors read and approved the final manuscript.

Availability of data and materials
The datasets generated and analysed during the current study are available in the The Cancer Genome Atlas database.

Ethics approval and consent to participate
The study was approved by the Clinical Research Ethics Committee of College of Medicine, Zhejiang University.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, 79 Qing Chun Road, Hangzhou, China. [2] Clinical Medical Research Center, Shenzhen People's Hospital, The Second Clinical Medical College of Jinan University, 1017 Dongmen North Road, Luohu District, Shenzhen, Guangdong, China. [3] Department of Urinary Surgery, Puning People's Hospital, Puning People's Hospital Affiliated To Southern Medical University, 30 Liusha Avenue, Jieyang, Guangdong, China. [4] Division of Community Health and Humanities, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

References
1. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, Heng DY, Larkin J, Ficarra V. Renal cell carcinoma. Nat Rev Dis Primers. 2017;3(17009):9.
2. Calvo E, Schmidinger M, Heng DY, Grünwald V, Escudier B. Improvement in survival end points of patients with metastatic renal cell. Cancer Treat Rev. 2016;50:109–17.
3. Shingarev R, Jaimes EA. Renal cell carcinoma: new insights and challenges for a clinician scientist. Am J Physiol Renal Physiol. 2017;313(2):F145–54.
4. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. Bmj. 2014;10(349):4797.
5. Inamura K. Renal cell tumors: understanding their molecular patho-logical epidemiology and the 2016 WHO Classification. Int J Mol Sci. 2017;18(10):2195.
6. Hsieh JJ, Le V, Cao D, Cheng EH, Creighton CJ. Genomic classifications of renal cell carcinoma: a critical step towards the future application of personalized kidney cancer care with pan-omics precision. J Pathol. 2018;244(5):525–37.
7. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, Bowlby R, Gibb EA, Akbani R, Beroukhim R, et al. The cancer genome atlas com-prehensive molecular characterization of renal cell carcinoma. Cell Rep. 2018;23(12):032.
8. Wu P, Liu JL, Pei SM, Wu CP, Yang K, Wang SP, Wu S. Integrated genomic analysis identifies clinically relevant subtypes of renal clear cell carcinom. BMC Cancer. 2018;18(1):018–4176.
9. Brooks SA, Brannon AR, Parker JS, Fisher JC, Sen O, Kattan MW, Hakimi AA, Hsieh JJ, Choueiri TK, Tamboli P, et al. ClearCode34: a prognostic risk predictor for localized clear cell renal cell carcinoma. Eur Urol. 2014;66(1):77–84.
10. de Velasco G, Culhane AC, Fay AP, Hakimi AA, Voss MH, Tannir NM, Tamboli P, Appleman LJ, Bellmunt J, Kimryn Rathmell W, et al. Molecular Subtypes Improve prognostic value of international metastatic renal cell carcinoma database consortium prognostic model. Oncologist. 2017;22(3):286–92.
11. Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. Epigenomics. 2015;7(2):247–65.
12. Voisin S, Eynon N, Yan X, Bishop DJ. Exercise training and DNA methyla-tion in humans. Acta Physiol. 2015;213(1):39–59.
13. Ferreira HJ, Esteller M. CpG islands in cancer: heads, tails, and sides. Meth-ods Mol Biol. 2018;1766:49–80.
14. Shenoy N, Vallumsetla N, Zou Y, Galeas JN, Shrivastava M, Hu C, Susztak K, Verma A. Role of DNA methylation in renal cell carcinoma. J Hematol Oncol. 2015;8(88):015–0180.
15. Creighton CJ, Morgan M, Gunaratne PH, Wheeler DA, Gibbs RA, Robert-son A, Chu A, Beroukhim R, Cibulskis K, Signoretti S, et al. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43–9.
16. Pires-Luís AS, Costa-Pinheiro P, Ferreira MJ, Antunes L, Lobo F, Oliveira J, Henrique R, Jerónimo C. Identification of clear cell renal cell carcinoma and oncocytoma using a three-gene promoter methylation panel. J Transl Med. 2017;15(1):017–1248.
17. Malouf GG, Su X, Zhang J, Creighton CJ, Ho TH, Lu Y, Raynal NJ, Karam JA, Tamboli P, Allanick F, et al. DNA methylation signature reveals cell ontog-eny of renal cell carcinomas. Clin Cancer Res. 2016;22(24):6236–46.
18. Lee H, Palm J, Grimes SM, Ji HP. The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations. Genome Med. 2015;7(112):015–0226.
19. Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D, Zhu J. Explor-ing TCGA pan-cancer data at the UCSC Cancer Genomics Browser. Sci Rep. 2013;3:2652.
20. Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN classification with dif-ferent numbers of nearest neighbors. IEEE Trans Neural Netw Learn Syst. 2018;29(5):1774–85.
21. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–3.
22. Zhang Y, Su J, Yu D, Wu Q, Yan H. EpiDiff: entropy-based quantitative identification of differential epigenetic modification regions from epig-enomes. In: Conf Proc IEEE Eng Med Biol Soc, 2013. p. 6609585.
23. Zhang Y, Liu H, Lv J, Xiao X, Zhu J, Liu X, Su J, Li X, Wu Q, Wang F, et al. QDMR: a quantitative method for identification of differentially methyl-ated regions by entropy. Nucleic Acids Res. 2011;39(9):8.
24. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007;35(Web Server issue):193–200.

Chen *et al. Cancer Cell Int*   (2019) 19:185

Page 14 of 14

25. Martínez-Salamanca JI, Huang WC, Millán I, Bertini R, Bianco FJ, Carballido JA, Ciancio G, Hernández C, Herranz F, Haferkamp A, et al. Prognostic impact of the 2009 UICC/AJCC TNM staging system for renal cell carcinoma with venous extension. Eur Urol. 2011;59(1):120–7.

26. Ficarra V, Martignoni G, Maffei N, Brunelli M, Novara G, Zanolla L, Pea M, Artibani W. Original and reviewed nuclear grading according to the Fuhrman system: a multivariate analysis of 388 patients with conventional renal cell carcinoma. Cancer. 2005;103(1):68–75.

27. Lam JS, Shvarts O, Leppert JT, Figlin RA, Belldegrun AS. Renal cell carcinoma 2005: new frontiers in staging, prognostication and targeted molecular therapy. J Urol. 2005;173(6):1853–62.

28. Borgmann H, Musquera M, Haferkamp A, Vilaseca A, Klatte T, Shariat SF, Scavuzzo A, Jimenez Rios MA, Wolff I, Capitanio U, et al. Prognostic significance of Fuhrman grade and age for cancer-specific and overall survival in patients with papillary renal cell carcinoma: results of an international multi-institutional study on 2189 patients. World J Urol. 2017;35(12):1891–7.

29. Li H, Wang J, Xiao W, Xia D, Lang B, Yu G, Guo X, Guan W, Wang Z, Hu Z, et al. Epigenetic alterations of Kruppel-like factor 4 and its tumor suppressor function in renal cell carcinoma. Carcinogenesis. 2013;34(10):2262–70.

30. Song E, Ma X, Li H, Zhang P, Ni D, Chen W, Gao Y, Fan Y, Pang H, Shi T, et al. Attenuation of kruppel-like factor 4 facilitates carcinogenesis by inducing g1/s phase arrest in clear cell renal cell carcinoma. PLoS One. 2013;8(7):e67758.

31. Liu Y, Zhang C, Fan J, Xiao L, Yin B, Zhou L, Xia S. Comprehensive analysis of clinical significance of stem-cell related factors in renal cell cancer. World J Surg Oncol. 2011;9(121):1477–7819.

32. Haake SM, Brooks SA, Welsh E, Fulp WJ, Chen DT, Dhillon J, Haura E, Sexton W, Spiess PE, Pow-Sang J, et al. Patients with ClearCode34-identified molecular subtypes of clear cell renal cell carcinoma represent unique populations with distinct comorbidities. Urol Oncol. 2016;34:3.

33. Evelönn EA, Landfors M, Haider Z, Köhn L, Ljungberg B, Roos G, Degerman S. DNA methylation associates with survival in non-metastatic clear cell renal cell carcinoma. BMC Cancer. 2019;19(1):019–5291.

34. Yotani T, Yamada Y, Arai E, Tian Y, Gotoh M, Komiyama M, Fujimoto H, Sakamoto M, Kanai Y. Novel method for DNA methylation analysis using high-performance liquid chromatography and its clinical application. Cancer Sci. 2018;109(5):1690–700.

35. Zhang Z, Li H, Jiang S, Li R, Li W, Chen H. Bo X1: A survey and evaluation of Web-based tools/databases for variant analysis of TCGA data. Brief Bioinform. 2018;29:4956394.

36. Chen F, Zhang Y, Şenbabaoğlu Y, Ciriello G, Yang L, Reznik E, Shuch B, Micevic G, De Velasco G, Shinbrot E, et al. Multilevel genomics-based taxonomy of renal cell carcinoma. Cell Rep. 2016;14(10):2476–89.

37. Lee CJ, Evans J, Kim K, Chae H, Kim S. Determining the effect of DNA methylation on gene expression in cancer cells. Methods Mol Biol. 2014;1101:161–78.

38. Jung M, Ellinger J, Gevensleben H, Syring I, Lüders C, de Vos L, Pützer S, Bootz F, Landsberg J, Kristiansen G, et al. Cell-free SHOX2 DNA methylation in blood as a molecular staging parameter for risk stratification in renal cell carcinoma patients: a prospective observational cohort study. Clin Chem. 2019;65(4):559–68.

39. Wei JH, Haddad A, Wu KJ, Zhao HW, Kapur P, Zhang ZL, Zhao LY, Chen ZH, Zhou YY, Zhou JC, et al. A CpG-methylation-based assay to predict survival in clear cell renal cell. Nat Commun. 2015;6:8699.

40. Nam HY, Chandrashekar DS, Kundu A, Shelar S, Kho EY, Sonpavde G, Naik G, Ghatalia P, Livi CB, Varambally S, et al. Integrative epigenetic and gene expression analysis of renal tumor progression to metastasis. Mol Cancer Res. 2019;17(1):84–96.

41. Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, Kuczyk MA, Lam T, Marconi L, Merseburger AS, et al. EAU guidelines on renal cell carcinoma: 2014 update. Eur Urol. 2015;67(5):913–24.

42. Ficarra V, Novara G, Galfano A, Brunelli M, Cavalleri S, Martignoni G, Artibani W. The 'Stage, Size, Grade and Necrosis' score is more accurate than the University of California Los Angeles Integrated Staging System for predicting cancer-specific survival in patients with clear cell renal cell carcinoma. BJU Int. 2009;103(2):165–70.

43. Manley BJ, Hakimi AA. Molecular profiling of renal cell carcinoma: building a bridge toward clinical impact. Curr Opin Urol. 2016;26(5):383–7.

44. Yu C, Dai D, Xie J. Molecular subtype classification of papillary renal cell cancer using miRNA expression. Onco Targets Ther. 2019;12:2311–22.

45. Zhang Y, Zhang J. Identification of functionally methylated regions based on discriminant analysis through integrating methylation and gene expression data. Mol BioSyst. 2015;11(7):1786–93.

46. Arai E, Kanai Y. Genetic and epigenetic alterations during renal carcinogenesis. Int J Clin Exp Pathol. 2010;4(1):58–73.

47. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, Flagg K, Hou J, Zhang H, Yi S, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci USA. 2017;114(28):7414–9.

48. Lorincz MC, Schübeler D. Evidence for converging DNA methylation pathways in placenta and cancer. Dev Cell. 2017;43(3):257–8.

49. Liu H, Hew HC, Lu ZG, Yamaguchi T, Miki Y, Yoshida K. DNA damage signalling recruits RREB-1 to the p53 tumour suppressor promoter. Biochem J. 2009;422(3):543–51.

50. Costello LC, Levy BA, Desouki MM, Zou J, Bagasra O, Johnson LA, Hanna N, Franklin RB. Decreased zinc and downregulation of ZIP3 zinc uptake transporter in the development of pancreatic adenocarcinoma. Cancer Biol Ther. 2011;12(4):297–303.

51. Oxford G, Smith SC, Hampton G, Theodorescu D. Expression profiling of Ral-depleted bladder cancer cells identifies RREB-1 as a novel transcriptional Ral effector. Oncogene. 2007;26(50):7143–52.

52. Milon BC, Agyapong A, Bautista R, Costello LC, Franklin RB. Ras responsive element binding protein-1 (RREB-1) down-regulates hZIP1 expression in prostate cancer cells. Prostate. 2010;70(3):288–96.

Publisher's Note