



Research article

A methylomics-correlated nomogram predicts the recurrence free survival risk of kidney renal clear cell carcinoma

Xiuxian Zhu[†], Xianxiong Ma[†] and Chuanqing Wu^{*}

Department of Gastrointestinal Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[†] The authors contributed equally to this work.

^{*} **Correspondence:** Email: wucq2014@hust.edu.cn.

Abstract: *Background:* Various studies have suggested that the DNA methylation signatures were promising to identify novel hallmarks for predicting prognosis of cancer. However, few studies have explored the capacity of DNA methylation for prognostic prediction in patients with kidney renal clear cell carcinoma (KIRC). It's very promising to develop a methylomics-related signature for predicting prognosis of KIRC. *Methods:* The 282 patients with complete DNA methylation data and corresponding clinical information were selected to construct the prognostic model. The 282 patients were grouped into a training set (70%, n = 198 samples) to determine a prognostic predictor by univariate Cox proportional hazard analysis, least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression analysis. The internal validation set (30%, n = 84) and an external validation set (E-MTAB-3274) were used to validate the predictive value of the predictor by receiver operating characteristic (ROC) analysis and Kaplan–Meier survival analysis. *Results:* We successfully identified a 9-DNA methylation signature for recurrence free survival (RFS) of KIRC patients. We proved the strong robustness of the 9-DNA methylation signature for predicting RFS through ROC analysis (AUC at 1, 3, 5 years in internal dataset (0.859, 0.840, 0.817, respectively), external validation dataset (0.674, 0.739, 0.793, respectively), entire TCGA dataset (0.834, 0.862, 0.842, respectively)). In addition, a nomogram combining methylation risk score with the conventional clinic-related covariates was constructed to improve the prognostic predicted ability for KIRC patients. The result implied a good performance of the nomogram. *Conclusions:* we successfully identified a DNA methylation-associated nomogram, which was helpful in improving the prognostic predictive ability of KIRC patients.

Keywords: signature; DNA methylation; kidney renal clear cell carcinoma; recurrence free survival; nomogram

1. Introduction

Renal cell carcinoma (RCC) is a genitourinary type of carcinoma, accounting for more than 2% of cancer-associated deaths worldwide [1,2]. KIRC is the most common subtype of RCC [3]. Presently, surgical resection is the major therapy of KIRC patients, after which patients with KIRC still have a quite large risk of cancer recurrence and metastasis [4]. In fact, more than 30% of patients who are diagnosed with KIRC experience metastasis, leading to a poor five-year survival rate [5]. It is commonly recognized that the patients with metastatic KIRC have a very dismal prognosis mainly due to the failure of early diagnosis and resistance to chemoradiotherapy [6]. It is urgently needed to find sensitive and reliable biomarkers for individualized prediction of KIRC patients.

Previous studies have reported that specific molecules could serve as survival-associated KIRC hallmarks. For example, Chen et al. suggested that a 3-mRNA predictor was proved to be a useful model for prognostic evaluation and could improve personalized management of KIRC patients [7]. Guan et al. identified a biomarker in KIRC based on miRNA-seq and digital gene expression-seq data [8]. A study revealed the prognostic value of a long non-coding RNA signature in localized clear cell renal cell carcinoma [9]. A previous research reported that fibroblast activation protein predicted prognosis in KIRC [10]. In the process of exploring prognostic markers of cancer, accumulating studies reported that epigenetic modification could affect key cancer-related genes, which implied its significant role in carcinogenesis. For instance, a study suggested that epigenetic therapeutics served as a new weapon in the war against cancer [11]. A previous research revealed their study in the cancer epigenome for therapy [12]. A research reported the interacting cancer machineries: cell signaling, lipid metabolism, and epigenetics [13]. Presently, DNA methylation is significantly involved in epigenetic regulatory mechanism which is intensively studied. Alters in DNA methylation may impact gene expression and interplay with numerous feedback mechanisms [14]. Thus, DNA methylation is a key factor for cancer development and various studies have suggested that DNA methylation may function as a predictor for diagnosis and prognosis of cancer. For instance, Brock et al. revealed DNA methylation markers in stage I lung cancer [15]. Shen et al. suggested that seven-DNA methylation CpG-based prognostic signature coupled with gene expression may predict survival of oral squamous cell carcinoma [16]. Gündert et al. indicated that genome-wide DNA methylation analysis revealed a prognostic classifier for non-metastatic colorectal cancer [17]. It has been reported that DNA methylation profiling is highly accurate and reproducible even using small specimens and poor quality material [18]. Therefore, the analysis of DNA methylation is promising to uncover novel biomarkers for predicting prognosis of cancer. However, few studies have explored the ability of DNA methylation for prognostic prediction in patients with KIRC. It's very promising to build a methylomics-related signature for predicting prognosis of KIRC.

In present study, an integrative analysis was carried out for the prediction of KIRC patients' prognosis and a 9 DNA methylation signature-based classifier was developed using univariate, LASSO, and multivariate Cox regression model. Next, we used Kaplan-Meier analysis, ROC analysis to assess and validate the prediction accuracy of the classifier in internal validation set, external validation set and entire TCGA set. In addition, a high prognostic predicted power of our nomogram was proved based on the training set and internal validation set.

2. Materials and methods

2.1. DNA methylation information of KIRC patients

The KIRC DNA methylation information and relevant clinical information was

downloaded from TCGA database by using R TCGAbiolinks package [19]. The KIRC DNA methylation data was measured by using Illumina Infinium HumanMethylation450 BeadChip (450 k) arrays based on the manufacturer's instructions. E-MTAB-3274 was retrieved from ArrayExpress database using ArrayExpress package [20]. The DNA methylation levels were set as β values, computed as $M / (M + U + 100)$, with U standing for an unmethylated signal and M standing for a methylated signal. After removing the patients without survival data, the remaining 282 patients with complete DNA methylation data and clinical information were used for constructing prognostic model. The 282 patients were grouped into a training group (70%, $n = 198$ samples) to identify a prognostic classifier and an internal validation group (30%, $n = 84$) to validate the predictive performance of the classifier. On the other hand, the 102 KIRC patients of E-MTAB-3274 from ArrayExpress database were employed as an external validation dataset. LASSO COX regression analysis was employed to further determine the candidate methylation sites involved in KIRC patients' RFS. In addition, LASSO analysis was executed by using 1000 iterations on the basis of a publicly available R package "glmnet" [21].

2.2. Data processing, normalization and determination of differentially expressed methylation sites

Pre-processing of the raw data was implemented for the determination of a prognostic classifier of KIRC. First of all, Raw methylation data was filtered for selecting the probes in at least one sample. After that, the normalization of the data was executed with "betaqn" function of wateRmelon package [22]. Subsequently, the total KIRC samples were divided into two cohorts (recurrence and no recurrence cohort) in accordance to recurrence status. The standardized beta was transformed to M value on the basis of the formulation: $M = \log(\beta/(1-\beta))$. M value was applied for the elimination of the variance which resulted from multiple probes. Finally, M value was exploited for the determination of the differentially expressed methylation sites between recurrence and no recurrence cluster according to "dmpFinder" function of minfi package [23].

2.3. Generation of methylomics-related signature

In the training set, the differentially expressed methylation sites were analyzed by a univariate Cox proportional hazards regression analysis to screen the promising sites that were importantly associated with RFS of KIRC patients. Then, those sites with p -values < 0.05 in the univariate Cox proportional hazards regression model were used for the LASSO Cox regression analysis to extract candidate sites tightly related to KIRC patients' RFS. Finally, the multivariate Cox regression analysis was performed by the candidate sites for selecting the methylome-based predictor of KIRC patients' RFS. As a result, 9 DNA methylation sites were employed to develop the methylation-based prognostic predictor. Then a risk-score model was constructed by using the 9-DNA methylation signature to compute risk scores of all of the KIRC patients. According to this risk score tool, we separated patients into high and low-risk groups by using a cutoff risk score of the median risk score. We conducted ROC analysis to assess the power of the 9-DNA methylation signature. AUC value was calculated in ROC analysis to weigh the predicted accuracy of methylome-based predictor for KIRC patients' RFS with the "survivalROC" package [24], the larger of the AUC value, the better for the predicted accuracy of the signature. Kaplan-Meier survival was exploited to compare differences in RFS between high- and low-risk clusters and Kaplan-Meier curves were obtained across the "survival" package [25].

2.4. Gene set variation analysis (GSVA)

Single sample gene sets enrichment analysis (ssGSEA) was conducted based on TCGA KIRC mRNA dataset with GSVA package [26] to assess the 9-DNA methylation signature-relevant signaling pathways. The most important pathway positively correlated with risk score was measured. According to this risk score model, we separated patients into high and low-risk groups by using a cutoff risk score of the median risk score. A p value of < 0.05 was considered as statistically significant.

2.5. Construction of the nomogram

The univariate and multivariate Cox proportional hazard analysis were conducted via the methylation risk score and other clinic-related variables. Cox proportional hazard models was exploited to measure hazard ratios (HR) and corresponding 95% confidence interval (CI) of variables. Factors that were significant ($P < 0.05$) from the multivariate Cox proportional hazard analysis were applied for constructing a nomogram by using the 'rms' R package. The prognostic nomogram was developed in TCGA-KIRC dataset for predicting 1, 3, 5-year's RFS of KIRC patients, respectively. C-index, ROC and calibration plot and decision curve analysis (DCA) were applied to evaluate prognostic nomogram performance. The results of the nomogram were presented in the calibrate curve, and the 45° line refers to the ideal prediction.

3. Results

3.1. Clinical characteristics of the study populations

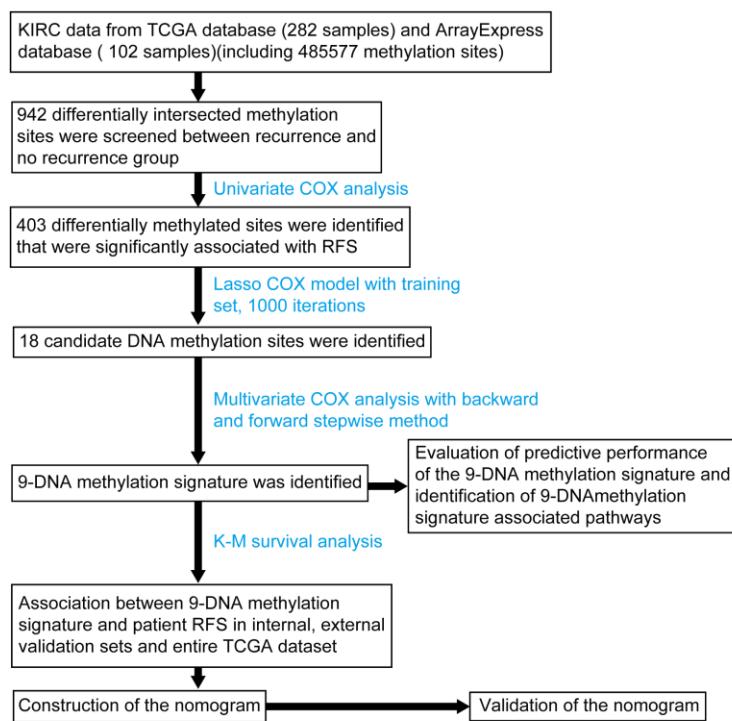


Figure 1. Flow chart of the present study.

Totally, 282 TCGA patients and 102 ArrayExpress database patients who were clinically and pathologically diagnosed as KIRC were included in this study. A flow chart of the study procedure was developed to summarize our study (Figure 1). The clinical feature of KIRC patients from TCGA dataset and ArrayExpress dataset was summarized in Table 1.

Table 1. Clinical characteristics of patients of the study.

Characteristics	Total (n=282)	Training (n=198)	dataset	Testing (n=84)	dataset
Gender					
FEMALE	104(36.88)	73(36.87)		31(36.9)	
MALE	178(63.12)	125(63.13)		53(63.1)	
Age					
≤ 60	132(46.81)	94(47.47)		38(45.24)	
> 60	150(53.19)	104(52.53)		46(54.76)	
Stage					
Stage I	147(52.13)	101(51.01)		46(54.76)	
Stage II	29(10.28)	19(9.6)		10(11.9)	
Stage III	68(24.11)	53(26.77)		15(17.86)	
Stage IV	36(12.77)	23(11.62)		13(15.48)	
Not Available	2(0.71)	2(1.01)			
M					
M0	238(84.4)	168(84.85)		70(83.33)	
M1	34(12.06)	22(11.11)		12(14.29)	
Not Available	10(3.55)	8(4.04)		2(2.38)	
T					
T1	149(52.84)	101(51.01)		48(57.14)	
T2	37(13.12)	24(12.12)		13(15.47)	
T3	93(32.98)	71(35.86)		22(26.19)	
T4	3(1.06)	2(1.01)		1(1.19)	
N					
N0	119(42.2)	85(42.93)		34(40.48)	
N1	8(2.84)	6(3.03)		2(2.38)	
Not Available	155(54.96)	107(54.04)		48(57.14)	
Grade					
G1	8(2.84)	7(3.54)		1(1.19)	
G2	122(43.26)	85(42.93)		37(44.05)	
G3	110(39.01)	80(40.4)		30(35.71)	
G4	38(13.48)	23(11.62)		15(17.86)	
Not Available	4(1.42)	3(1.52)		1(1.19)	
Cancer_Status					
TUMOR FREE	201(71.28)	144(72.73)		57(67.86)	
WITH TUMOR	59(20.92)	38(19.19)		21(25)	
Not Available	22(7.80)	16(8.08)		6(7.14)	
Laterality					
Bilateral	1(0.35)	1(0.51)			
Left	130(46.1)	92(46.46)		38(45.24)	
Right	151(53.55)	105(53.03)		46(54.76)	
<u>Hemoglobin_result</u>					

Continued on next page

Characteristics	Total (n=282)	Training (n=198)	dataset	Testing (n=84)	dataset
Elevated	4(1.42)	2(1.01)		2(2.38)	
Low	132(46.81)	86(43.43)		46(54.76)	
Normal	102(36.17)	75(37.88)		27(32.14)	
Not Available	44(16.60)	35(17.68)		9(10.71)	
Platelet_qualitative_result					
Elevated	14(4.96)	10(5.05)		4(4.76)	
Low	30(10.64)	24(12.12)		6(7.14)	
Normal	189(67.02)	125(63.13)		64(76.19)	
Not Available	49(17.38)	39(19.7)		10(11.9)	
Serum_calcium_result					
Elevated	4(1.42)	2(1.01)		2(2.38)	
Low	100(35.46)	65(32.83)		35(41.67)	
Normal	79(28.01)	60(30.3)		19(22.62)	
Not Available	99(35.11)	71(35.86)		28(33.33)	
White_cell_count_result					
Elevated	77(27.3)	56(28.28)		21(25)	
Low	3(1.06)	1(0.51)		2(2.38)	
Normal	150(53.19)	100(50.51)		50(59.52)	
Not Available	52(18.44)	41(20.71)		11(13.09)	
Race_list.race					
BLACK	OR	AFRICAN	47(16.67)	34(17.17)	13(15.48)
AMERICAN					
Not Available			3(1.06)	1(0.51)	2(2.38)
WHITE			232(82.27)	163(82.32)	69(82.14)

3.2. Determination of 9 methylation sites signature

We selected 942 differentially methylated sites between recurrence and no recurrence cohorts for univariate Cox proportional hazard regression analysis. As a result, 403 DNA methylation sites were proved to be strongly involved in KIRC patients' RFS ($P < 0.05$) (Table S1). After that the above 403 KIRC DNA methylation sites were projected to LASSO Cox regression model and 18 methylation sites were screened as the candidate prognostic sites (Figure 2) which were significantly related to RFS of KIRC patients. Those 18 candidate methylation sites were further analyzed by multivariable Cox analysis in the training set. Finally, 9 methylation sites ($P < 0.05$) were identified as the independent risk factors tightly associated with RFS of KIRC patients, including cg12009697, cg14207589, cg07990546, cg04094846, cg11132272, ca11955474, cg03010887, cg09217923, cg25206071. These 9 methylation sites were used for building a DNA methylation-based signature to predict the RFS of KIRC patients. Therefore, we constructed the risk score formula as follows: risk score = $2.625 \times \text{cg14207589} + 4.519 \times \text{cg09217923} + 2.615 \times \text{cg11132272} - 2.373 \times \text{cg12009697} + 2.751 \times \text{cg03010887} + 1.568 \times \text{cg07990546} + 1.344 \times \text{cg04094846} + 1.507 \times \text{cg11955474} + 2.065 \times \text{cg25206071}$. We discovered that, the hypermethylation levels of cg14207589, cg09217923, cg11132272, cg03010887, cg07990546, cg04094846, cg11955474, cg25206071 tended to have poor survival rates. Whereas the hypomethylation levels of cg12009697 tended to have poor survival rates. (Figures 3 and S1).

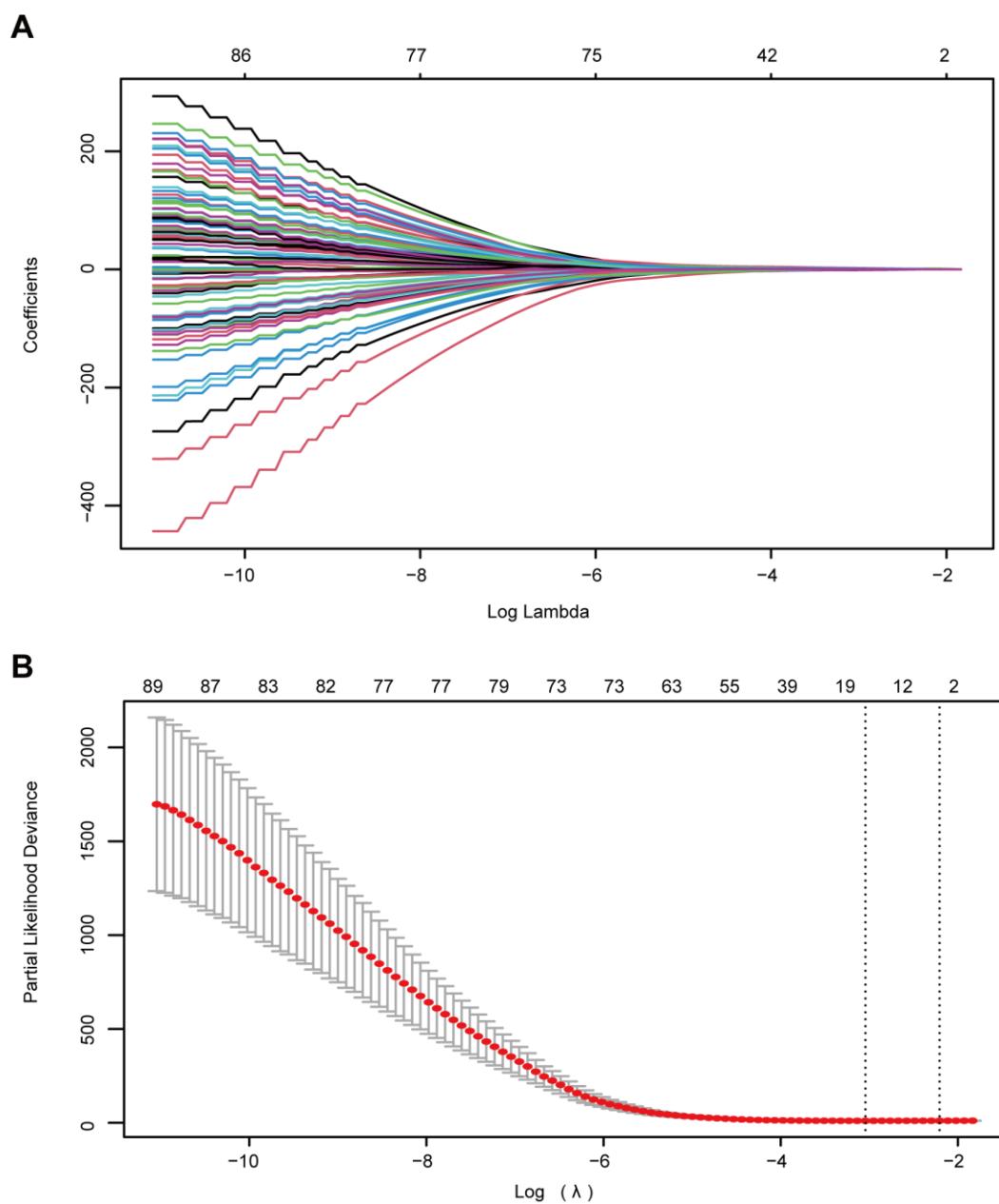


Figure 2. Candidate methylation sites selection using the LASSO Cox regression model. (A) 10-fold cross-validation for tuning parameter selection in the LASSO model using minimum criteria (the 1-SE criteria). (B) LASSO coefficient profiles of the 403 methylation sites. A coefficient profile plot was produced against log (lambda) sequence. Vertical line was implemented at the value selected by using 10-fold cross-validation, where optimal lambda resulted in 18 non-zero coefficients.

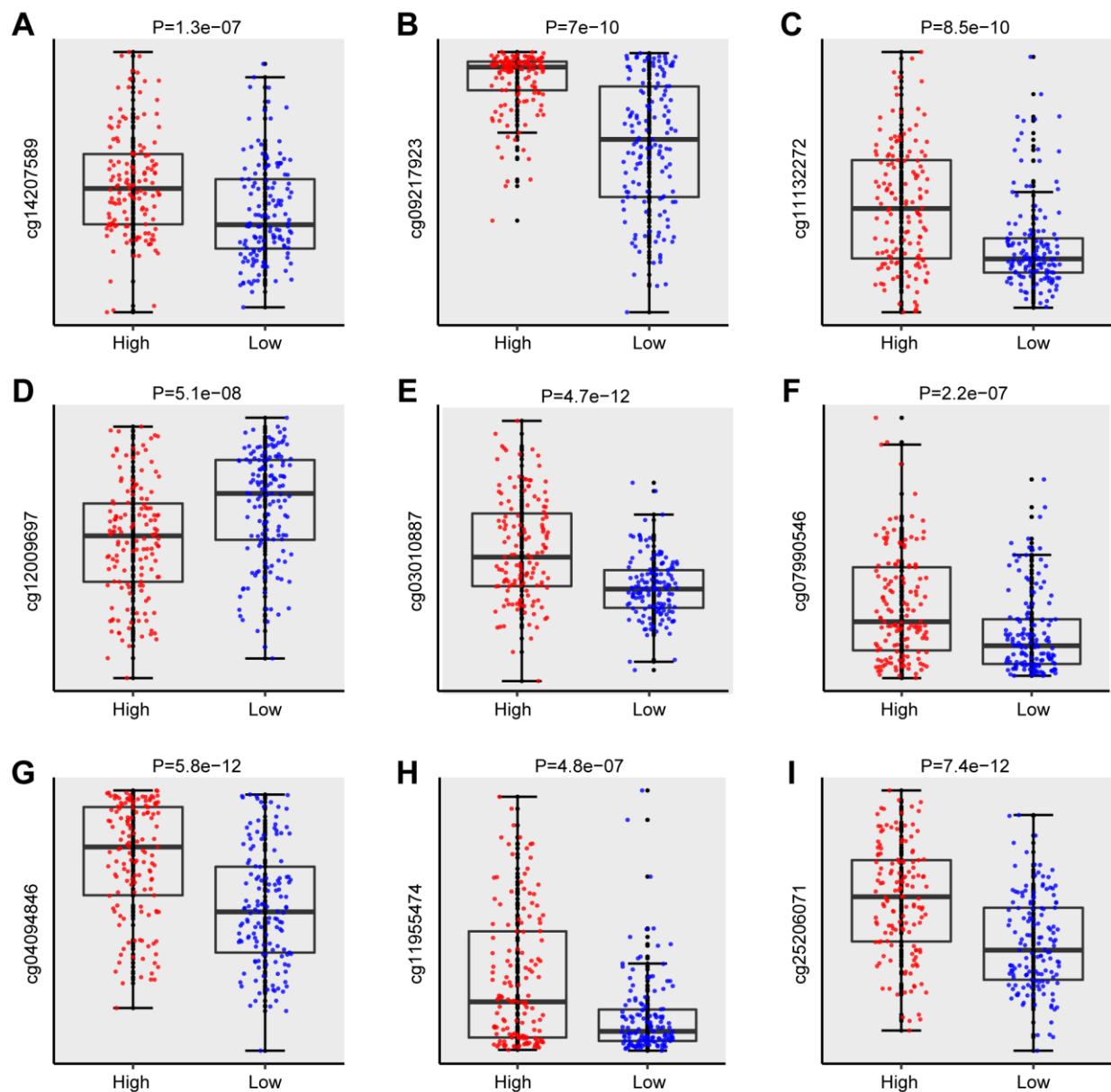


Figure 3. Boxplots of methylation β values against risk group in the entire TCGA dataset. “High Risk” and “Low Risk” stand for the high-risk and low-risk groups, respectively. The median risk score was applied as a cutoff. Y-axis stands for the β -value of 9-DNA methylation sites respectively. The differences between the 2 groups were measured by Mann-Whitney U test.

3.3. Interplays between 9 DNA methylation signature and KIRC patients' RFS in the internal validation, external validation and entire TCGA set

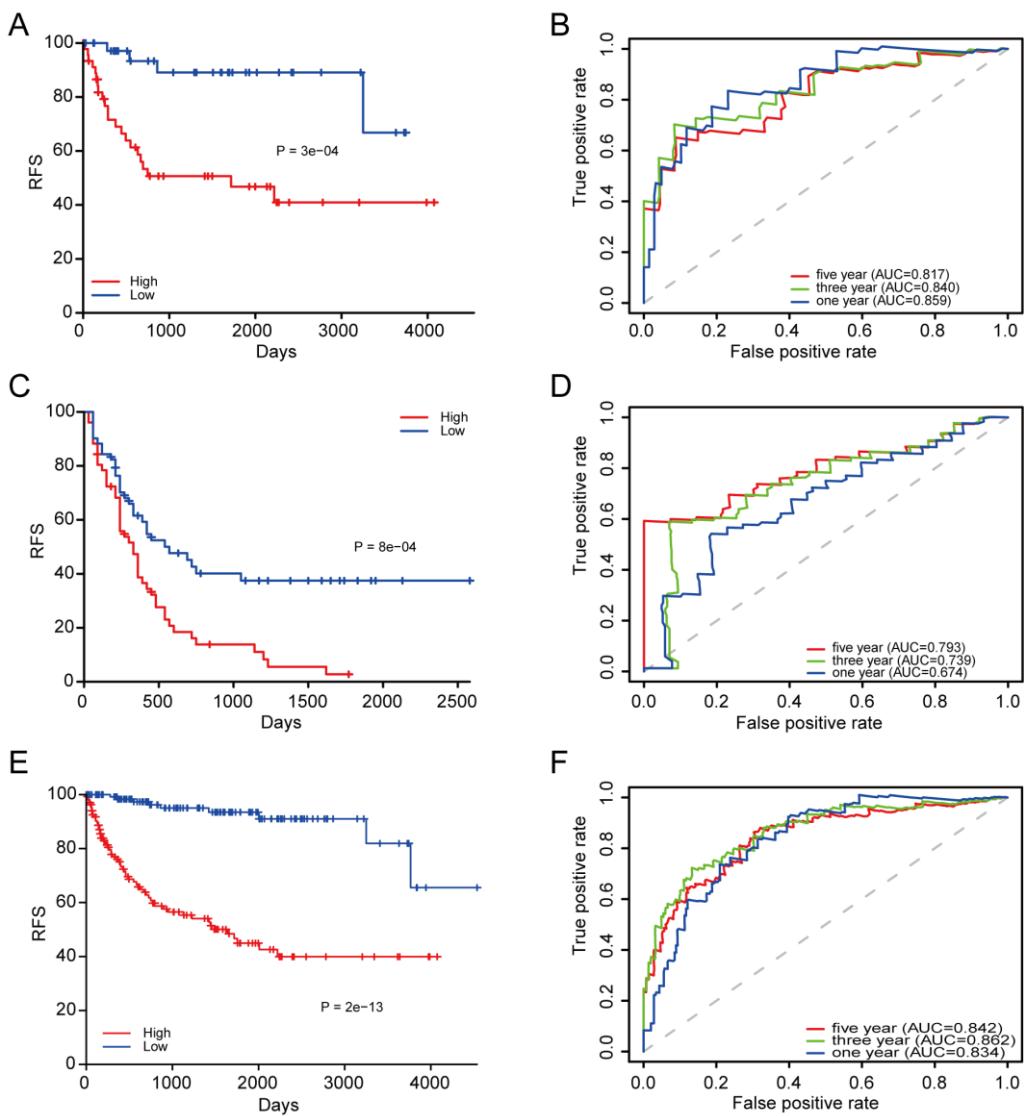


Figure 4. Kaplan-Meier and ROC analysis of patients with KIRC in the internal validation, external validation and entire TCGA set. (A, C, E) Kaplan-Meier analysis with two-sided log-rank test was used to estimate the differences in RFS between the low-risk and high-risk group KIRC patients. (B, D, F) 1-, 3-, 5-year ROC curves of the 9-DNA methylation signature were used to assess the value of predicting the RFS of KIRC patients. “High” and “Low” stood for the high risk score group and low risk score group, respectively. The median risk score was set as a cutoff.

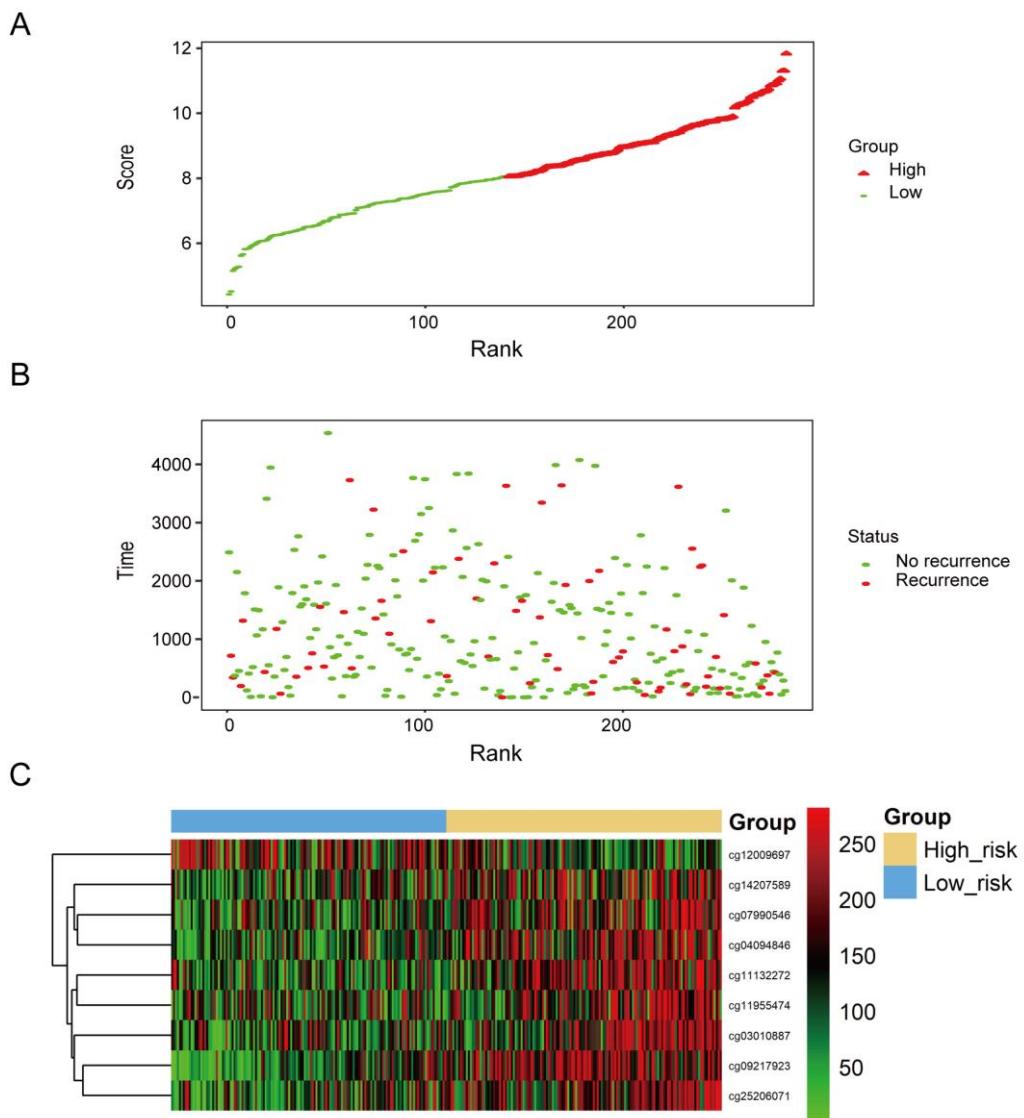


Figure 5. Methylation risk score analysis of 282 KIRC in the entire TCGA dataset. (A) Methylation risk score distribution against the rank of risk score. Median risk score served as the cut-off point. (B) Survival status of KIRC patients against the rank of risk score. (C) Heatmap of 9 methylation sites expression profiles of KIRC patients.

In order to analyze the predictive capability of this 9 DNA methylation-based signature, we divided the patients in each set into high and low-risk cohorts based on the median risk score. Kaplan-Meier survival analysis was exploited to examine the difference in RFS between the two cohorts. The patients in high-risk cohort generated a significantly unfavorable RFS in internal validation set ($p = 3e-04$) (Figure 4A), similar results were obtained in external validation set ($p = 8e-04$) (Figure 4C) and entire TCGA set ($p = 2e-13$) (Figure 4E). After that, the power of the 9-DNA methylation signature-based indicator for KIRC patients' RFS was explored by using a time-dependent ROC curve. The AUC of the biomarker at 1, 3, 5 years in internal set were 0.859, 0.840, 0.817, respectively (Figure 4B), in external validation set (0.674, 0.739, 0.793, respectively) (Figure 4D) and entire TCGA set (0.834, 0.862, 0.842, respectively) (Figure 4F). We concluded that the 9-

DNA methylation signature-related indicator yielded a great power for the prediction of KIRC patients' RFS.

Subsequently, the rank of the patients from entire TCGA set was carried out according to the risk scores (Figure 5A), and the dotplot of the total patients was performed according to their survival status (Figure 5B). We noted that the high-risk cohort generated a worse survival than that in the low-risk cohort. Figure 5C showed the heatmap of 9 methylation sites distribution according to risk score, which supported the previous observation (Figure S2). Then, we performed subgroup analysis by using a few clinical variables containing age, gender, anatomic site, grade and stage. The majority of sub-groups showed a good predictive power of the 9-DNA methylation signature for KIRC patients' RFS (Figures S3–S7).

3.4. Determination of the 9-DNA methylation signature-related biological pathways

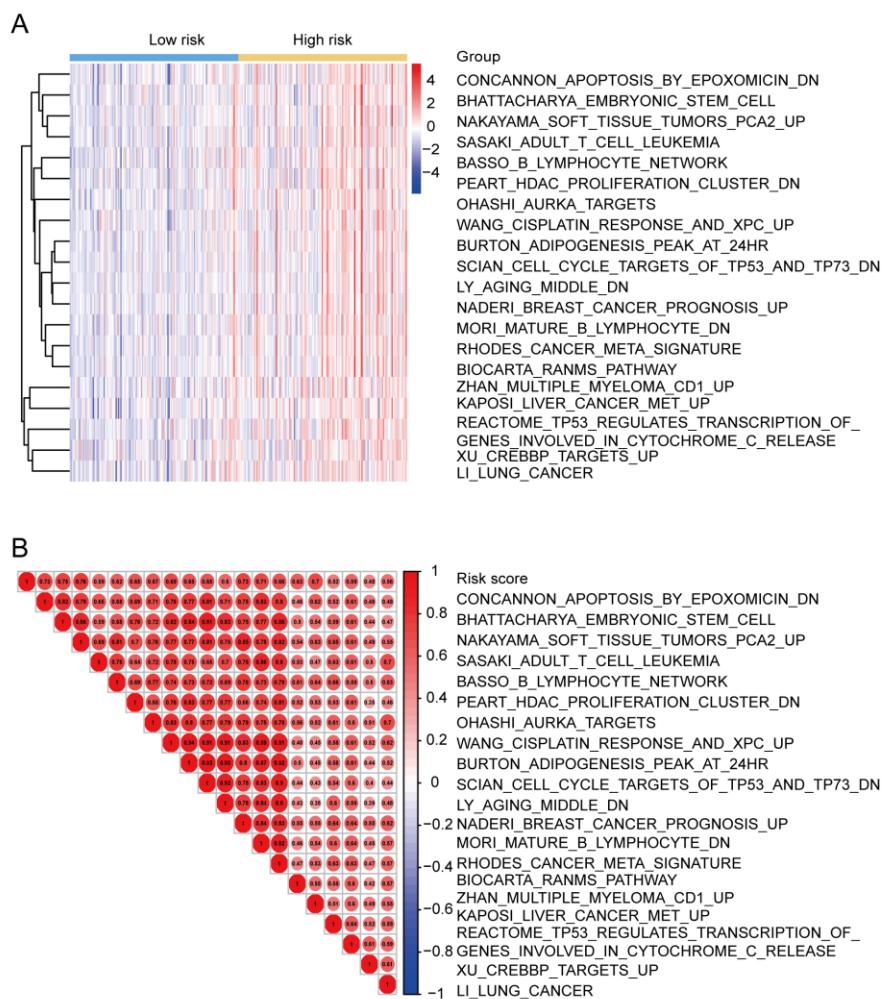


Figure 6. Identification of the 9 DNA methylation signature-relevant biological pathways. (A) Heatmap of top 20 enriched pathways associated with high risk group. (B) Association graph between risk scores and top 20 pathways.

We separated patients into high and low-risk groups by using a cutoff risk score of the median risk score. Figure 6A showed the top 20 enriched pathways that were more activated in the

high-risk cases than that in low-risk cases. The same trend between the enriched pathways and risk score was further addressed in Figure 6B, which proved a good correlation between the pathways and the risk score (Table S2).

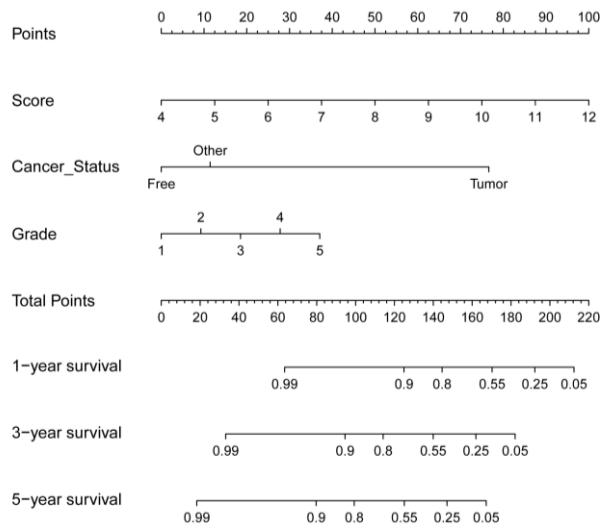


Figure 7. 9-DNA methylation-related nomogram for the prediction of KIRC patients' RFS. The nomogram was developed in the entire TCGA cohort based on the DNA methylation risk score, cancer status and grade.

3.5. Nomogram development

Table 2. Univariate and multivariate Cox regression analysis results in accordance to DNA methylation risk score as well as other clinical factors.

id	HR	HR.95 L	HR.95 H	pvalue	HR	HR.95 L	HR.95 H	pvalue
Cancer_Status	6.5596 82	4.5975 69	9.3591 68	3.30E- 25	4.3339 7	2.9315 38	6.4073 18	1.95E- 13
Score	2.7182 82	2.1972 52	3.3628 63	3.24E- 20	1.4694 38	1.1264 49	1.9168 62	0.0045 41
Grade	2.1965 78	1.6860 27	2.8617 31	5.53E- 09	1.3305 98	1.0250 83	1.7271 69	0.0318 65
platelet_qualitative_r esult	1.8093 24	1.1915 03	2.7474 99	0.0054 02	1.3384 6	0.8787 16	2.0387 41	0.1745 36
M	3.2791 85	2.4948 14	4.3101 64	1.68E- 17	1.3298 3	0.7804 02	2.2660 71	0.2945 44
T	2.7762 56	2.0779 11	3.7093 01	4.94E- 12	1.3279 95	0.6805 82	2.5912 71	0.4055 69
Stage	2.9309 31	2.3114 73	3.7164 19	6.90E- 49	1.1545 77	0.5554 06	2.3997 52	0.7002
White_cell_count_re sult	0.6543 28	0.4784 44	0.8948 71	0.0079 23	0.9649 21	0.6950 44	1.3395 88	0.8310 69
Serum_calcium_resu lt	1.4771 63	1.0922 62	1.9976 99	0.0113 11	1.0335 13	0.7454 13	1.4329 64	0.8432 71
Sex	1.5336	0.9034	2.6034	0.1132				

Continued on next page

id	HR	HR.95 L	HR.95 H	pvalue	HR	HR.95 L	HR.95 H	pvalue
Laterality	23 0.8488	22 0.6699	33 1.0754	39 0.1746				
Hemoglobin_result	2 0.8689	24 0.6679	9 1.1303	83 0.2951				
N	24 1.1983	48 0.7614	69 1.8858	56 0.4341				
Race	48 1.1029	75 0.7574	65 1.6059	47 0.6092				
	63 15	93 91						

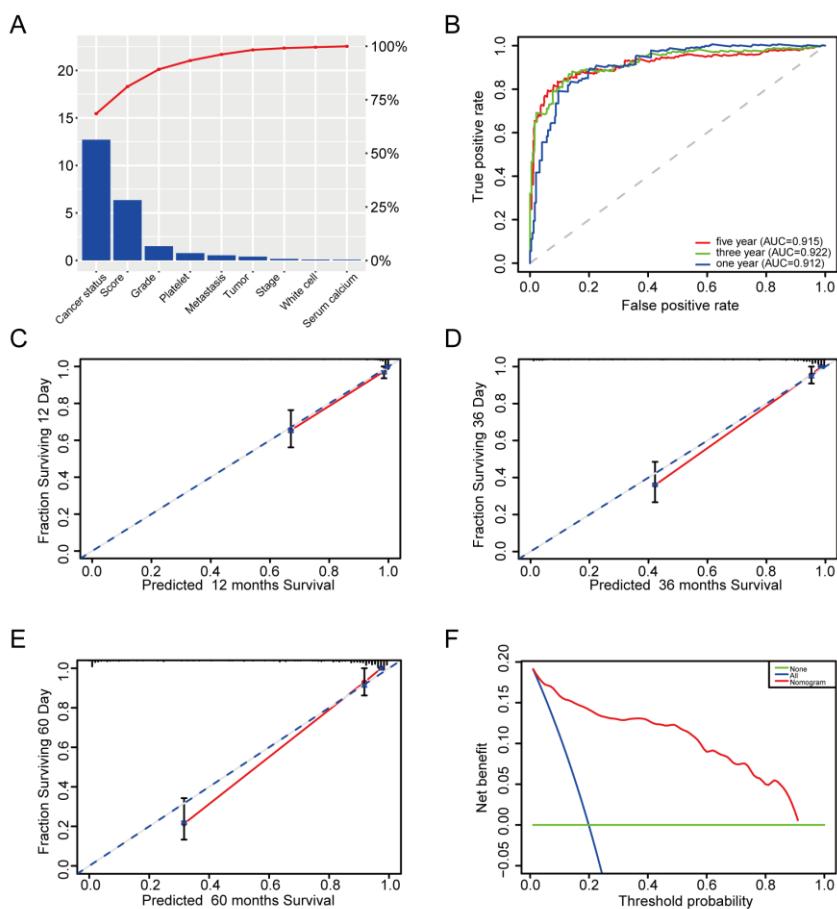


Figure 8. Validation of the DNA methylation gene-related nomogram in the entire TCGA dataset. (A) The higher the bar chart, the greater the percentage. (B) 1-, 3-, 5-year receiver operating characteristic curves for the metabolic gene-related nomogram. (C–E) referred to the 1-, 3-, 5-year nomogram calibration curves, respectively. The closer the dotted line fitted to the ideal line, the better the predictive value of the nomogram. (F) The DCA for the nomogram. The net benefit was plotted versus the threshold probability. The red line represented the nomogram. The blue line represented the treat-all and the green line represented the treat-none.

We first performed univariate and multivariate Cox tool according to a few clinic-related covariates. Hazard ratios (HRs) manifested that the 9-DNA methylation-related classifier was tightly related to KIRC patients' RFS ($P < 0.001$, HR 1.47, 95% CI 1.13–1.92) (Table 2 and Figure S8), manifesting that the 9-DNA methylation signature was an independent predictor of KIRC patients' RFS in the entire TCGA set. To further reinforce the prognostic value of the 9-DNA methylation-based signature for KIRC in a quantitative strategy. We developed a nomogram (Figure 7) that integrated the 9-DNA methylation-associated signature and the conventional clinic-associated covariates containing cancer status, grade. The significance between the 9-DNA methylation-based factor and the conventional clinic-associated covariates was described in Figure 8A. The value of the nomogram was weighed on the basis of C-index (0.802, 95%CI: 0.768–0.828), AUC (1, 3, 5-year: 0.912, 0.922, 0.915) (Figure 8B) and calibration plot (Figure 8C–E), proving a good value of the tool. In addition, DCA manifested that the nomogram created more crucial value of clinical utilization as predictor of KIRC patients' RFS than that in treat all or treat none group. Net benefit was available for KIRC patients in 3-year recurrent risks (Figure 8F). We concluded that our nomogram had a great value and may have potential for clinical application.

4. Discussion

The KIRC DNA methylation information and relevant clinical information from TCGA database was analyzed to unearth the predictor of KIRC patients' RFS. Eventually, 9 DNA methylation sites (cg12009697, cg14207589, cg07990546, cg04094846, cg11132272, ca11955474, cg03010887, cg09217923, cg25206071) were found to be strongly associated with RFS of KIRC patients. The above 9 DNA methylation sites were mapped to 4 genes (TLX2, OSBP, TAAR2, IL19). It is interesting to note that, previous researches suggested that most of these 4 genes were correlated with cancer, respectively. For instance, Liu et al. found that methylated COL4A1, COL4A2, TLX2, and ITGA4 demonstrated high accuracy for the detection of colorectal neoplasms in stool and they may be potentially valuable biomarkers for the detection of colorectal cancer [27]. Liao et al. demonstrated that TAT-OSBP-MKK6(E) was a novel artificially designed molecule, which induced apoptosis and selectively targeted human ovarian carcinoma HO8910 cells. Their study provided novel insights that may aid in the development of a new generation of anticancer drugs, which implied the importance of OSBP in cancer [28]. Hsing et al. suggested that upregulated IL-19 in breast cancer promoted tumor progression and affected clinical result [29]. The result manifested that the 3 of the 4 genes involved in the 9 sites played important roles in progression of cancer.

Multiple studies indicated that nomograms may strengthen prognostic predictive accuracy for cancer by integrating several clinical variables in a quantitative method. For example, Wang et al. constructed a nomogram to predict gleason sum upgrading of clinically diagnosed localized prostate cancer among Chinese patients [30]. Lee et al. built a prognostic nomogram to predict progression-free survival in patients with platinum-sensitive recurrent ovarian cancer [31]. Taşkın et al. found a nomogram with potential clinical application to predict lymph node metastasis in endometrial cancer patients diagnosed incidentally by postoperative pathological assessment [32]. However, nomograms for improving predictive ability of KIRC patients' RFS have been rarely reported. In this study, we aimed to develop a nomogram using both methylation risk score and several clinic-associated covariates for predicting 1, 3, 5-year RFS for KIRC patient from the entire TCGA set. The indicators of the C-index, AUC, DCA and calibration curve from the entire TCGA set exhibited that our nomogram had discriminative accuracy and was adopted as the preferred predictive model. Furthermore, we developed the nomogram as a predicator in a quantitative method for the prognosis

of KIRC patient, which could predict the accurate survival percentage of KIRC patients. In addition, the key virtues of the nomogram contained clinical applicability and convenient factors, which was readily available according to historical records. On the other hand, DNA methylation profiling is highly accurate and reproducible even using small specimens and poor quality material [18], suggesting that this methylomics-correlated nomogram method had a great advantage to predict prognosis of cancer patients in comparison to other signatures.

Growing researches manifested that LASSO Cox regression model can be employed to unearth markers of multiple cancers. For example, Connell et al. found a four-group urine risk classifier for predicting outcome in prostate cancer patients based on LASSO Cox regression analysis [33]. Zhang et al. identified a new eight-long noncoding RNA molecular signature for breast cancer survival prediction according to the LASSO method, univariate and multivariate Cox analyses [34]. Liu et al. identified an eight-lncRNA prognostic model for breast cancer using WGCNA network analysis and a Cox - proportional hazards model based on LASSO Cox regression model [35]. Du et al. constructed a diagnostic nomogram of platelet-based score models for hepatic alveolar echinococcosis and atypical liver cancer using LASSO Cox regression model [36]. A significant merit of LASSO is that it has sparse estimates of the regression coefficients which suggests that various components are precisely 0. In other words, LASSO can automatically remove unnecessary variables. LASSO has various preferred properties for regression models with numerous variables, and multiple efficient optimization algorithms are available for linear regression and generalized linear models [37]. To our knowledge, fewer studies have used LASSO regression model for the determination of a prognostic predictor of KIRC. We performed LASSO Cox regression analysis to select candidate methylation sites tightly related to KIRC patients' RFS for filtering the variables between univariate and multivariate Cox analysis, further improving prognostic predicted ability of 9-DNA methylation-associated signature.

There are still a few limitations of this study. Firstly, our study is a retrospective one and the prospective studies including a larger number of samples of various medical centers are needed to verify the outcomes. Secondly, more factors should be mapped into our external validation set to improve the predicted accuracy of the nomogram. Thirdly, the nomograms were developed by using retrospective data from TCGA database, which may be correlated with the potential hazard of selection bias.

5. Conclusions

We successfully identified a 9-DNA methylation-based signature, which was a useful prognostic predictor for KIRC patients' RFS. A nomogram combining methylation risk score and the conventional clinic-related covariates was helpful in improving the predicted ability of clinical prognosis for KIRC patients.

Acknowledgements

This work was supported by the National Nature Science Foundation of China (Grant Nos. 82073324).

Conflict of interest

The authors declare that they have no competing interests.

References

1. R. L. Siegel, K. D. Miller, A. Jemal, *Cancer Statistics, CA Cancer J. Clin.*, **67** (2017), 7–30.
2. G. Gandaglia, P. Ravi, F. Abdollah, A. E. Abd-El-Barr, A. Becker, I. Popa, et al., Contemporary incidence and mortality rates of kidney cancer in the United States, *Can. Urol. Assoc. J.*, **8** (2014), 247–252.
3. D. J. Sanchez, M. C. Simon, Genetic and metabolic hallmarks of clear cell renal cell carcinoma, *Biochem. Biophys. Acta Rev. Cancer*, **1870** (2018), 23–31.
4. P. Fisel, S. Kruck, S. Winter, J. Bedke, J. Hennenlotter, A. T. Nies, et al., DNA methylation of the SLC16A3 promoter regulates expression of the human lactate transporter MCT4 in renal cancer with consequences for clinical outcome, *Clin. Cancer Res.*, **19** (2013), 5170–5181.
5. P. Cairns, Renal cell carcinoma, *Cancer Biomarks*, **9** (2010), 461–473.
6. S. H. Rossi, T. Klatte, J. Usher-Smith, G. D. Stewart, Epidemiology and screening for renal cancer, *World J. Urol.*, **36** (2018), 1341–1353.
7. L. Chen, Y. Luo, G. Wang, K. Qian, G. Qian, C. L. Wu, et al., Prognostic value of a gene signature in clear cell renal cell carcinoma, *J. Cell Physiol.*, **234** (2019), 10324–10335.
8. L. Guan, J. Tan, H. Li, X. Jin, Biomarker identification in clear cell renal cell carcinoma based on miRNA-seq and digital gene expression-seq data, *Gene*, **647** (2018), 205–212.
9. L. Qu, Z. L. Wang, Q. Chen, Y. M. Li, H. W. He, J. J. Hsieh, et al., Prognostic Value of a Long Non-coding RNA Signature in Localized Clear Cell Renal Cell Carcinoma, *Eur. Urol.*, **74** (2018), 756–763.
10. J. I. Lopez, P. Errarte, A. Erramuzpe, R. Guarch, J. M. Cortes, J. C. Angulo, et al., Fibroblast activation protein predicts prognosis in clear cell renal cell carcinoma, *Hum. Pathol.*, **54** (2016), 100–105.
11. N. Ahuja, A. R. Sharma, S. B. Baylin, Epigenetic Therapeutics: A New Weapon in the War Against Cancer, *Ann. Rev. Med.*, **67** (2016), 73–89.
12. P. A. Jones, J. P. Issa, S. Baylin, Targeting the cancer epigenome for therapy, *Nat. Rev. Genet.*, **17** (2016), 630–641.
13. T. W. Grunt, Interacting Cancer Machineries: Cell Signaling, Lipid Metabolism, and Epigenetics, *Trends Endocrinol. Metab.*, **29** (2018), 86–98.
14. A. Nebbioso, F. P. Tambaro, C. Dell'Aversana, L. Altucci, Cancer epigenetics: Moving forward, *PLoS Genet.*, **14** (2018), e1007362.
15. M. V. Brock, C. M. Hooker, E. Ota-Machida, Y. Han, M. Guo, S. Ames, et al., DNA methylation markers and early recurrence in stage I lung cancer, *N. Eng. J. Med.*, **358** (2008), 1118–1128.
16. S. Shen, G. Wang, Q. Shi, R. Zhang, Y. Zhao, Y. Wei, et al., Seven-CpG-based prognostic signature coupled with gene expression predicts survival of oral squamous cell carcinoma, *Clin. Epigenet.*, **9** (2017), 88.
17. M. Gundert, D. Edelmann, A. Benner, L. Jansen, M. Jia, V. Walter, et al., Genome-wide DNA methylation analysis reveals a prognostic classifier for non-metastatic colorectal cancer (ProMCol classifier), *Gut*, **68** (2019), 101–110.
18. D. Capper, D. T. W. Jones, M. Sill, V. Hovestadt, D. Schrimpf, D. Sturm, et al., DNA methylation-based classification of central nervous system tumours, *Nature*, **555** (2018), 469–474.
19. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, et al., TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, *Nucleic Acids Res.*, **44** (2016), e71.

20. I. Merelli, P. Lio, L. Milanesi, NuChart: an R package to study gene spatial neighbourhoods with multi-omics annotations, *PLoS One*, **8** (2013), e75146.
21. S. Engebretsen, J. Bohlin, Statistical predictions with glmnet, *Clin. Epigenet.*, **11** (2019), 123.
22. R. Pidsley, Y. W. CC, M. Volta, K. Lunnon, J. Mill, L. C. Schalkwyk, A data-driven approach to preprocessing Illumina 450K methylation array data, *BMC Genom.*, **14** (2013), 293.
23. M. J. Aryee, A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, et al., Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays, *Bioinformatics*, **30** (2014), 1363–1369.
24. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinf.*, **12** (2011), 77.
25. G. De Angelis, R. De Angelis, L. Frova, A. Verdecchia, MIAMOD: a computer package to estimate chronic disease morbidity using mortality and survival data, *Comput. Methods Programs Biomed.*, **44** (1994), 99–107.
26. S. Hanzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.*, **14** (2013), 7.
27. X. Liu, J. Wen, C. Li, H. Wang, J. Wang, H. Zou, High-Yield Methylation Markers for Stool-Based Detection of Colorectal Cancer, *Dig. Dis. Sci.*, **65** (2020), 1710–1719.
28. H. Liao, J. L. Kang, W. Y. Jiang, C. Deng, J. Yuan, R. Shuai, Delivery of Constitutively Active Mutant MKK6(E) With TAT-OSBP Induces Apoptosis in Human Ovarian Carcinoma HO8910 Cells, *Int. J. Gynecol. Cancer*, **25** (2015), 1548–1556.
29. C. H. Hsing, H. C. Cheng, Y. H. Hsu, C. H. Chan, C. H. Yeh, C. F. Li, et al., Upregulated IL-19 in breast cancer promotes tumor progression and affects clinical outcome, *Clin. Cancer Res.*, **18** (2012), 713–725.
30. J. Y. Wang, Y. Zhu, C. F. Wang, S. L. Zhang, B. Dai, D. W. Ye, A nomogram to predict Gleason sum upgrading of clinically diagnosed localized prostate cancer among Chinese patients, *Chin. J. Cancer*, **33** (2014), 241–248.
31. C. K. Lee, R. J. Simes, C. Brown, S. Lord, U. Wagner, M. Plante, et al., Prognostic nomogram to predict progression-free survival in patients with platinum-sensitive recurrent ovarian cancer, *Br. J. Cancer*, **105** (2011), 1144–1150.
32. S. Taskin, Y. E. Sukur, B. Varli, K. Koyuncu, M. M. Seval, C. Ates, et al., Nomogram with potential clinical use to predict lymph node metastasis in endometrial cancer patients diagnosed incidentally by postoperative pathological assessment, *Arch. Gynecol. Obstet.*, **296** (2017), 803–809.
33. S. P. Connell, M. Hanna, F. McCarthy, R. Hurst, M. Webb, H. Curley, et al., A Four-Group Urine Risk Classifier for Predicting Outcome in Prostate Cancer Patients, *BJU Int.*, 2019.
34. Y. Zhang, Z. Li, M. Chen, H. Chen, Q. Zhong, L. Liang, et al., Identification of a New Eight-Long Noncoding RNA Molecular Signature for Breast Cancer Survival Prediction, *DNA Cell Biol.*, **38** (2019), 1529–1539.
35. Z. Liu, M. Li, Q. Hua, Y. Li, G. Wang, Identification of an eight-lncRNA prognostic model for breast cancer using WGCNA network analysis and a Coxproportional hazards model based on L1-penalized estimation, *Int. J. Mol. Med.*, **44** (2019), 1333–1343.
36. Q. Du, Y. Wang, S. Guan, C. Hu, M. Li, L. Zhou, et al., The diagnostic nomogram of platelet-based score models for hepatic alveolar echinococcosis and atypical liver cancer, *Sci. Rep.*, **9** (2019), 19403.

37. S. M. Kim, Y. Kim, K. Jeong, H. Jeong, J. Kim, Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography, *Ultrasonography*, **37** (2018), 36–42.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)