

## Research Article

# A Risk Score Model Based on Nine Differentially Methylated mRNAs for Predicting Prognosis of Patients with Clear Cell Renal Cell Carcinoma

Jingmin Zhou, Guanghua Liu, Xingcheng Wu, Zhien Zhou, Jialin Li, and Zhigang Ji 

Department of Urology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, No. 1 Shuaifuyuan Hutong, Dongcheng District, Beijing 100730, China

Correspondence should be addressed to Zhigang Ji; [jingminzhou@163.com](mailto:jingminzhou@163.com)

Received 3 May 2020; Revised 22 July 2020; Accepted 1 January 2021; Published 15 January 2021

Academic Editor: Hubertus Himmerich

Copyright © 2021 Jingmin Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Purpose.** DNA methylation alterations play important roles in initiation and progression of clear cell renal cell carcinoma (ccRCC). In this study, we attempted to identify differentially methylated mRNA signatures with prognostic value for ccRCC. **Methods.** The mRNA methylation and expression profiling data of 306 ccRCC tumors were downloaded from The Cancer Genome Atlas (TCGA) to screen differentially methylated lncRNAs and mRNAs (DMLs and DMMs) between bad and good prognosis patients. Uni- and multivariable Cox regression analyses and LASSO Cox-PH regression analysis were used to select prognostic lncRNAs and mRNAs. Corresponding risk scores were calculated and compared for predictive performance in the training set using Kaplan-Meier OS and ROC curve analyses. The optimal risk score was then identified and validated in the validation set. Function enrichment analysis was conducted. **Results.** This study screened 461 DMMs and 63 DMLs between good prognosis and bad prognosis patients, and furthermore, nine mRNAs and six lncRNAs were identified as potential prognostic molecules. Compared to nine-mRNA status risk score model, six-lncRNA methylation risk score model, and six-lncRNA status risk score model, the nine-mRNA methylation risk score model showed superiority for prognosis stratification of ccRCC patients in the training set. The prognostic ability of the nine-mRNA methylation risk score model was validated in the validation set. The nine prognostic mRNAs were functionally associated with neuroactive ligand receptor interaction and inflammation-related pathways. **Conclusion.** The nine-mRNA methylation signature (DMRTA2, DRGX, FAM167A, FGGY, FOXI2, KRTAP2-1, TCTEX1D1, TTBK1, and UBE2QL1) may be a useful prognostic biomarker and tool for ccRCC patients. The present results would be helpful to elucidate the possible pathogenesis of ccRCC.

## 1. Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common histological subtype of renal cell carcinoma which accounts for more than 90% of cancers in the kidney [1]. ccRCC is characterized by high rates of progression and mortality, and patients have a dire prognosis [2]. Traditional clinical predictors of outcome such as clinical stages, grade, and necrosis could not achieve satisfactory results for ccRCC patients with similar clinical features but different outcomes [3]. Therefore, novel sensitive prognostic biomarkers for ccRCC are demanded.

Aberrant DNA methylation is a critical factor for initiation and progression of cancer [4]. Associations of DNA methylation with prognosis of ccRCC patients are gaining attention. For instance, based on genome-wide CpG methylation profiling, Wei et al. develop a five-CpG-based classifier as a reliable tool for survival prediction of ccRCC patients [5]. Using The Cancer Genome Atlas (TCGA) data, a recent study by Chen et al. finds that a four-gene DNA methylation signature is tightly related to prognosis of patients with kidney renal clear cell carcinoma and is able to serve as a prognostic predictor [6]. Moreover, a promoter methylation classifier panel of 172 differentially methylated CpGs is reported to be

capable of classifying prognosis of nonmetastatic ccRCC patients [7]. Long noncoding RNAs (lncRNAs) modulate gene expression at various levels and facilitate cancer phenotypes via interaction with other cellular macromolecules, such as DNA, RNA, and protein [8, 9]. Important implications of lncRNAs have been demonstrated for tumorigenesis of ccRCC [10]. A large number of lncRNAs gain DNA methylation in ccRCC, and several lncRNAs with promoter methylation are related to outcome of patients [11]. However, aberrantly methylated lncRNAs which can function as prognostic biomarkers for ccRCC patients remain elusive.

In this paper, we tried to identify potential mRNA methylation signature and lncRNA methylation signature for survival prediction of ccRCC patients through a comprehensive analysis of mRNA and lncRNA methylation data downloaded from TCGA dataset, and the predictive performances of methylation or status risk scores were compared based on these prognostic lncRNAs or mRNAs to select the best prognostic model for ccRCC.

## 2. Materials and Methods

**2.1. Datasets and Preprocessing.** Methylation and gene expression data of 309 ccRCC tumor tissue samples and 24 normal control tissue samples were downloaded from TCGA repository using Illumina Infinium Human Methylation 450 BeadChip platform and Illumina HiSeq 2000 RNA Sequencing platform. Patients without complete survival information were excluded from the study, and the resulting 306 patients with ccRCC tumor tissue samples were used as the training set in the current study. E-MTAB-3274 [12] dataset consisting of methylome data of 107 ccRCC samples was obtained from EBI ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) based on Illumina Infinium Human Methylation 450 BeadChip platform. Of these samples, 102 samples had corresponding survival information and were used as the validation set.

According to probe sites and ID information, we identified lncRNAs and mRNAs from TCGA set and E-MTAB-3274 dataset based on HUGO Gene Nomenclature Committee (HGNC) [13] database (<http://www.genenames.org/>) that enrolls 4112 lncRNAs and 19201 protein coding genes. As a result, 13919 mRNAs and 1028 lncRNAs were identified.

**2.2. Differential Methylation Analysis between Good and Bad Prognosis Samples.** In the training set, differentially methylated lncRNAs and mRNAs (DMLs and DMMs) were identified between good prognosis patients (being alive with OS time lasting for more than 60 months) and bad prognosis patients (being dead with OS time lasting for less than 36 months). Significant DMLs and DMMs were defined as those with  $FDR < 0.05$  and  $|\log_2 FC| > 0.263$ . Next, two-way hierarchical clustering analysis of these significant DMLs and DMMs was carried out based on centered Pearson correlation algorithm.

**2.3. Correlation Analysis between DNA Methylation and Expression of Significant DMLs and DMMs.** We analyzed

relationships of overall methylation and expression levels of preselected significant DMLs and DMMs through calculating Pearson correlation coefficient (PCC) and Spearman correlation coefficient (SCC) [14]. Subsequently, PCC of methylation and expression levels of each individual DML or DMM was computed. These DMLs and DMMs with significant negative correlations ( $p$  value  $< 0.05$ ) were selected to be applied in further analysis.

**2.4. Prognostic Risk Scoring Model Building.** In order to construct risk scoring models for predicting survival in ccRCC patients, we employed a step-by-step strategy to identify the prognostic DMLs and DMMs in the training set. Firstly, we performed univariable Cox regression analysis to assess the association between methylation levels of the aforementioned DMLs and DMMs with patients' OS time. The resulting lncRNAs and mRNAs were considered significant if their  $p$  values  $< 0.05$ . Secondly, the significant lncRNAs and mRNAs were chosen to undergo a multivariable Cox regression analysis. The significant RNAs ( $p$  value  $< 0.05$ ) were determined to be independent predictors of prognosis and were then used to fit a L1-penalized (LASSO) Cox-PH regression [15] model with estimation of optimal lambda value through performing 1,000 cross-validations. The optimal feature combination was selected using LASSO regularized regression algorithm in penalized package. CV and lambda values were obtained by feature filtering. The small lambda value usually resulted in small penalty term value as well as over fitting phenomenon; therefore, we chose the largest lambda value which is corresponding to the lambda parameter value. Eventually, the resulting optimal set of predictive lncRNAs and mRNAs was applied to build risk scoring models based on their methylation levels or status together with the estimated multivariable Cox regression coefficients. Methylation status of an individual lncRNA or mRNA was defined according to optimal cutoff point of methylation value (Monte-Carlo  $p$  value  $< 0.05$ ) which was determined based on patients' survival through performing X-Tile Bio-Informatics Tool [16] (high expression, expression status = 1; low expression, expression status = 0). We utilized the following equations to quantify methylation or status risk scores for survival prediction:

$$\text{Status risk score} = \sum \beta \text{RNA}_n \times \text{status RNA}_n,$$

$$\text{Methylation risk score} = \sum (\beta \text{RNA}_n \times \text{methylation RNA}_n).$$

$\beta \text{RNA}_n$  suggests regression coefficient of  $\text{RNA}_n$ ; status  $\text{RNA}_n$  and methylation  $\text{RNA}_n$  suggest methylation status and values of  $\text{RNA}_n$ , respectively.

Risk score was estimated for each patient using above formula. With median risk score as cutoff point, the training set or the validation set was dichotomized into a high-risk group and a low-risk group.

**2.5. Statistical Analysis.** Kaplan-Meier survival plots as well as log-rank test were used to analyze difference in OS time between two risk groups. Receiver operating characteristic (ROC) curve was plotted to compare sensitivity and specificity of different risk scores in predicting OS time of patients. Based on the training set, associations of clinical features

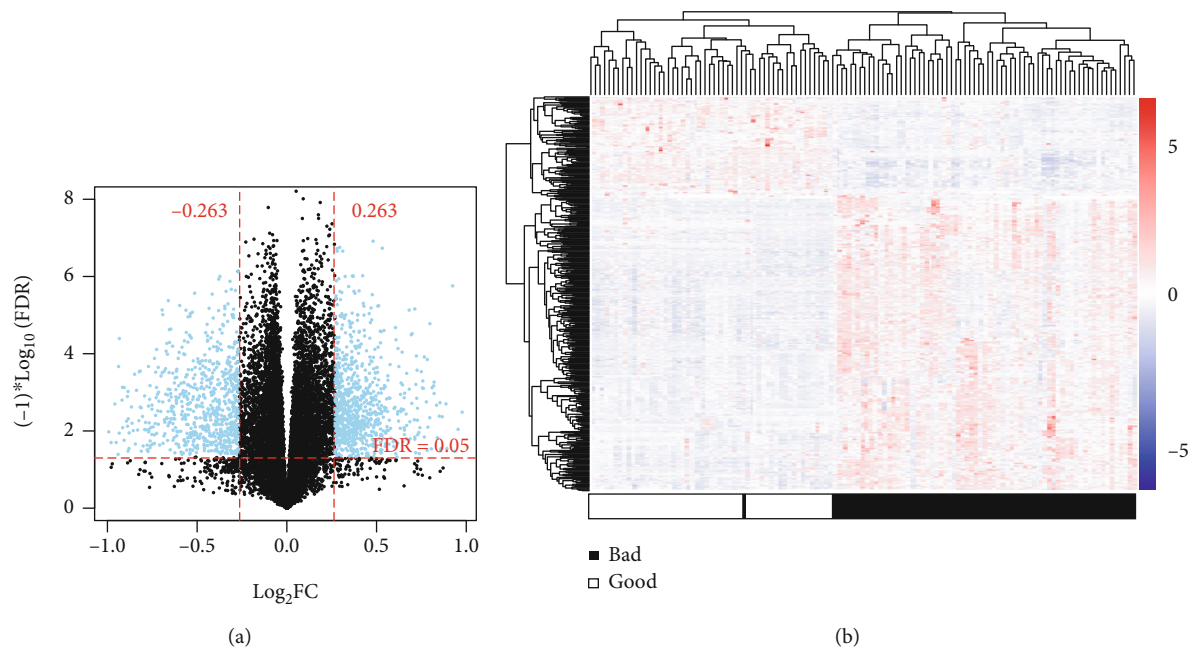


FIGURE 1: Graphical display of differentially methylated RNAs between good prognosis and bad prognosis samples. (a) Volcano plot of effect size ( $\log_2FC$ ) –  $\log_{10}(FDR)$  of the identified differentially methylated RNAs. (b) Two-way hierarchical clustering heatmap for the identified differentially methylated RNAs.

and risk score status with patients' OS were calculated using uni- and multivariable Cox regression analyses. Nomogram was built combining the optimal risk score status with prognostic clinical features. Calibration plots were used to evaluate predictive capability of the nomogram. Different packages of the R software (version 3.4.1) were used to implement all these bioinformatics and statistical analyses in the study: limma [17] package for differential methylation analysis and differential expression analysis; pheatmap [18] package (<https://cran.r-project.org/web/packages/pheatmap/index.html>) for hierarchical clustering analysis; survival package (version 2.41-1, <http://bioconductor.org/packages/survival/>) for uni- and multivariable Cox regression analyses and Kaplan-Meier OS analysis; cor.test (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.test.html>) function for calculation of PCC and SCC values; pROC package for ROC curve analysis; penalized package (version 0.9.50) for LASSO Cox-PH regression model; rms package [19] (version 5.1-2, <https://cran.r-project.org/web/packages/rms/index.html>) for nomogram analysis; gene set enrichment analysis (GSEA) [20] software (<http://software.broadinstitute.org/gsea/index.jsp>) for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Significance was defined as  $p$  value  $< 0.05$ .

**2.6. Functional Enrichment Analysis.** According to the optimal risk score, the training set was divided into a high-risk group and a low-risk group. Between the two risk groups, differentially expressed genes (DEGs) were identified based on the threshold of  $FDR < 0.05$  and  $|\log_2FC| > 0.5$  and then underwent KEGG pathway enrichment analysis. Significant signaling pathways were selected in the present study.

### 3. Results

**3.1. Detecting DMLs and DMMs between Good Prognosis and Bad Prognosis Patients.** The training set consisted of 68 bad prognosis patients and 52 good prognosis patients. As shown in Figures 1(a) and 1(b), we found a total of 524 DMRs ( $FDR < 0.05$  and  $|\log_2FC| > 0.263$ ) between the good prognosis patients and the bad prognosis patients, including 461 DMMs and 63 DMLs. Furthermore, the 461 DMMs were composed of 119 hypomethylated mRNAs and 342 hypermethylated mRNAs, and the 63 DMLs were comprised of 34 hypomethylated lncRNAs and 29 hypermethylated lncRNAs. Results of two-way hierarchical clustering analysis for these patients were shown in a heatmap (Figure 1(b)). Based on methylation value, ccRCC patients were categorized into two groups.

**3.2. Correlation between DNA Expression and Methylation.** For the total of 524 identified DMRs, the overall methylation level was negatively correlated with the overall expression level ( $PCC = -0.1909$ ,  $p$  value =  $1.252E^{-04}$ ;  $SCC = -0.1456$ ,  $p$  value =  $3.589E^{-04}$ ). A significant negative correlation between methylation and expression of each individual DMR was observed for 222 DMRs, consisting of 27 lncRNAs and 195 mRNAs. These 222 DMRs were selected for further analysis.

**3.3. Identifying a Nine-mRNA Methylation Signature and a Six-lncRNA Methylation Signature for Survival Prediction.** Using data in the training set, we performed univariable Cox regression analysis of the 222 DMRs. Totally 124 resulting DMRs were significantly correlated with patients' OS time ( $p$  value  $< 0.05$ ), including 113 mRNAs and 11 lncRNAs.

TABLE 1: Information of the six-lncRNA signature and the nine-mRNA signature.

Type	ID	Locus	Coefficient	HR	95% CI	Value	Cutoff
lncRNA	FAM138B	cg08267300	1.3497	4.436	1.388-6.461	$1.31 \times 10^{-3}$	0.6
	HCG11	cg27490387	0.2787	2.181	1.158-4.108	$2.20 \times 10^{-2}$	0.10
	KIAA0087	cg25962015	-2.9759	0.043	0.003-0.532	$7.10 \times 10^{-3}$	0.72
	MIAT	cg23345500	1.2621	1.408	1.064-3.084	$4.66 \times 10^{-2}$	0.04
	PSORS1C3	cg26504835	-1.6876	0.047	0.006-0.350	$1.41 \times 10^{-3}$	0.32
	SNHG11	cg27314048	-1.1996	0.111	0.013-0.988	$2.47 \times 10^{-2}$	0.85
mRNA	DMRTA2	cg27265170	0.4475	3.839	1.078-4.366	$4.54 \times 10^{-2}$	0.23
	DRGX	cg17494438	-1.1456	0.038	0.020-0.718	$2.91 \times 10^{-2}$	0.47
	FAM167A	cg02771117	-0.2809	8.842	3.902-10.04	$1.42 \times 10^{-2}$	0.41
	FGGY	cg27571329	1.5940	4.524	3.678-5.565	$6.08 \times 10^{-4}$	0.69
	FOXI2	cg26115633	0.0342	2.546	1.086-5.971	$4.43 \times 10^{-2}$	0.55
	KRTAP2-1	cg24624629	0.1612	1.385	1.195-5.194	$7.66 \times 10^{-3}$	0.35
	TCTEX1D1	cg25339566	0.5241	2.711	1.670-4.396	$2.03 \times 10^{-2}$	0.10
	TTBK1	cg27363327	0.4388	0.152	0.084-0.275	$4.60 \times 10^{-3}$	0.12
	UBE2QL1	cg26692294	0.5770	0.063	0.012-0.342	$1.29 \times 10^{-2}$	0.37

HR: hazard ratio; CI: confidence interval.

Furthermore, 31 DMRs were determined to be independent prognostic factors in multivariable Cox regression analysis ( $p$  value  $< 0.05$ ), consisting of 24 mRNAs and 7 lncRNAs. Consequently, the 31 DMRs were used to fit a LASSO Cox-PH regression model. Using the optimal lamda value, as shown in Table 1 and supplement Figure 1 (Figure S1), we retrieved an optimal panel of 9 predictive mRNAs (lamda = 0.51184, CVL = -546.8306; DMRTA2, DRGX, FAM167A, FGGY, FOXI2, KRTAP2-1, TCTEX1D1, TTBK1, and UBE2QL) and an optimal panel of 6 predictive lncRNAs (lamda = 0.03557, CVL = -557.1598; FAM138B, HCG11, KIAA0087, MIAT, PSORS1C3, and SNHG11).

Based on methylation levels or status of the 9 prognostic mRNAs or the 6 prognostic lncRNAs and their multivariable Cox regression coefficients, we designed formulas for calculating methylation or status risk scores as follows:

$$\text{mRNA status risk score} = (0.447514) * \text{status}_{\text{DMRTA2}} + (-1.145573) * \text{status}_{\text{DRGX}} + (-0.280914) * \text{status}_{\text{FAM167A}} + (1.593956) * \text{status}_{\text{FGGY}} + (0.034273) * \text{status}_{\text{FOXI2}} + (0.161233) * \text{status}_{\text{KRTAP2-1}} + (0.524107) * \text{status}_{\text{TCTEX1D1}} + (0.43882) * \text{status}_{\text{TTBK1}} + (0.577028) * \text{status}_{\text{UBE2QL1}}$$

$$\text{lncRNA status risk score} = (1.349587) * \text{status}_{\text{FAM138B}} + (0.278684) * \text{status}_{\text{HCG11}} + (-2.97586) * \text{status}_{\text{KIAA0087}} + (1.262154) * \text{status}_{\text{MIAT}} + (-1.687644) * \text{status}_{\text{PSORS1C3}} + (-1.199576) * \text{status}_{\text{SNHG11}}$$

$$\text{mRNA methylation risk score} = (0.447514) * \text{meth}_{\text{DMRTA2}} + (-1.145573) * \text{meth}_{\text{DRGX}} + (-0.280914) * \text{meth}_{\text{FAM167A}} + (1.593956) * \text{meth}_{\text{FGGY}} + (0.034273) * \text{meth}_{\text{FOXI2}} + (0.161233) * \text{meth}_{\text{KRTAP2-1}} + (0.524107) * \text{meth}_{\text{TCTEX1D1}} + (0.43882) * \text{meth}_{\text{TTBK1}} + (0.577028) * \text{meth}_{\text{UBE2QL1}}$$

$$\text{lncRNA methylation risk score} = (1.349587) * \text{meth}_{\text{FAM138B}} + (0.278684) * \text{meth}_{\text{HCG11}} + (-2.97586) * \text{meth}_{\text{KIAA0087}} + (1.262154) * \text{meth}_{\text{MIAT}} + (-1.687644) * \text{meth}_{\text{PSORS1C3}} + (-1.199576) * \text{meth}_{\text{SNHG11}}$$

Nine-mRNA methylation risk score, nine-mRNA status risk score, six-lncRNA methylation risk score, and six-lncRNA status risk score were calculated for each patient in the training set, respectively. With median risk score as threshold, the training set was divided into a high-risk group and a low-risk group. For each risk score, difference in OS time was significant between the two risk groups according to results of Kaplan-Meier survival analysis and log-rank test (Figures 2 and 3). The nine-mRNA methylation risk score displayed markedly smaller  $p$  value than the other three risk scores (nine-mRNA methylation risk score, HR (95%CI) = 3.968 (2.497 - 6.305) and  $p$  value =  $3.109E^{-10}$ ; six-lncRNA methylation risk score, HR (95%CI) = 2.489 (1.614 - 3.839) and  $p$  value =  $1.984E^{-05}$ ; nine-mRNA status risk score, HR (95%CI) = 3.867 (2.429 - 6.158) and  $p$  value =  $9.245E^{-10}$ ; six-lncRNA status risk score, HR (95%CI) = 2.262 (1.495 - 3.424) and  $p$  value =  $7.318E^{-05}$ , Table 2). Validation was performed in the validation set to assess predictive performance of the four risk scores. Only the nine-mRNA methylation risk score could dichotomize the validation set into two risk groups with significantly different OS time (nine-mRNA methylation risk score, HR(95%CI) = 1.723 (1.087 - 2.730) and  $p$  value =  $1.915E^{-02}$ ; six-lncRNA methylation risk score, HR(95%CI) = 1.491 (0.944 - 2.356) and  $p$  value =  $8.507E^{-02}$ ; nine-mRNA status risk score, HR(95%CI) = 1.538 (0.967 - 2.445) and  $p$  value =  $6.722E^{-02}$ ; six-lncRNA status risk score, HR(95%CI) = 1.307 (0.825 - 2.070) and  $p$  value =  $2.528E^{-01}$ , Table 2).



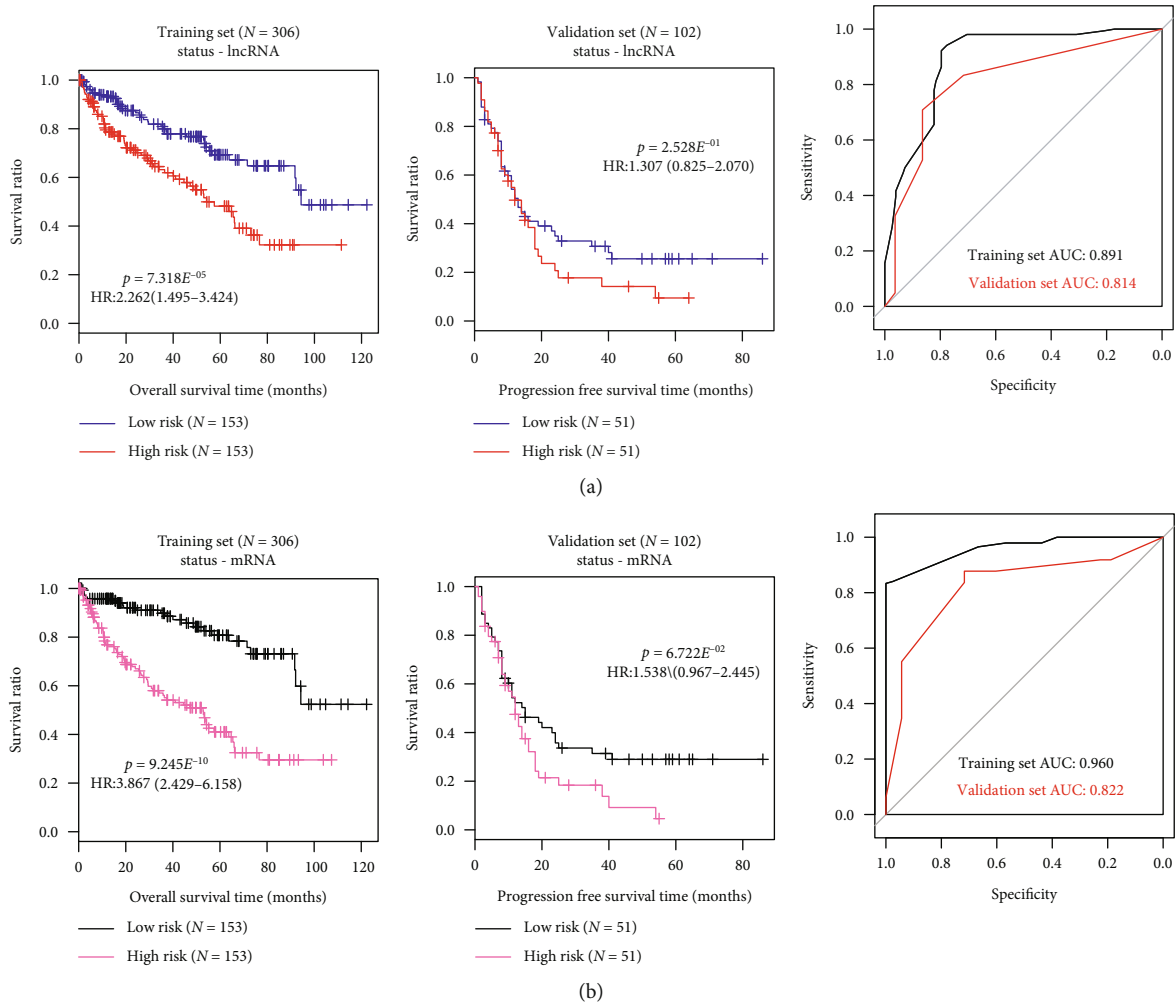


FIGURE 2: Kaplan-Meier survival and ROC curves for (a) six-lncRNA status risk score model and (b) nine-mRNA status risk score model in the training set and the validation set.

Results of ROC curve analysis were depicted in Figures 2 and 3 and Table 3. AUC values of nine-mRNA methylation risk score for the training set and the validation set were 0.987 and 0.884, respectively, which are higher compared to the other risk scores. All above observations suggest that the nine-mRNA methylation risk score performs better than other risk scores for survival prediction of ccRCC patients.

**3.4. Characterization of Four Independent Prognostic Features and Construction of a Composite Nomogram.** The univariable Cox regression analysis in the training set indicated that age, histologic grade, pathologic N, pathologic T, pathologic stage, hemoglobin, platelet qualitative, and nine-mRNA methylation risk score were significantly related to patients' OS time ( $p$  value  $< 0.05$ , Table 4). Furthermore, histologic grade (HR = 1.470, 95%CI = 1.076 – 2.009, and  $p$  value =  $1.56E^{-02}$ ), pathologic stage (HR = 1.989, 95%CI = 1.101 – 3.592, and  $p$  value =  $2.26E^{-02}$ ), platelet qualitative (HR = 0.690, 95%CI = 0.501 – 0.949, and  $p$  value =  $2.28E^{-02}$ ), and nine-mRNA methylation risk score model (HR = 2.574, 95%CI = 1.149 – 5.767, and  $p$  value =  $2.17E^{-04}$ )

were found to be independent predictive factors of prognosis in multivariable Cox regression analysis (Table 4, Figure 4). It revealed that prognostic value of the nine-mRNA methylation risk score model was independent of clinical factors. To integrate the nine-mRNA methylation risk score with the three prognostic clinical variables, a composite nomogram was built (Figure 5(a)). Calibration curves for 5-year survival probability showed sound agreement between predicted and actual probabilities of 5-year survival (Figure 5(b)).

**3.5. Functional Implications of the Nine Prognostic mRNAs.** To investigate functional characteristics of the nine prognostic mRNAs in ccRCC biology, we performed functional enrichment analysis on the identified DEGs between the high-risk group and the low-risk group of the training set classified by the nine-mRNA methylation risk score. We found 715 DEGs between the two risk groups, consisting of 190 downregulated genes and 525 upregulated genes (Figures 6(a) and 6(b)). These genes were significantly involved in cytokine-cytokine receptor interaction, neuroactive ligand receptor interaction, toll-like receptor signaling pathway, and cell receptor signaling pathway (Table 5).

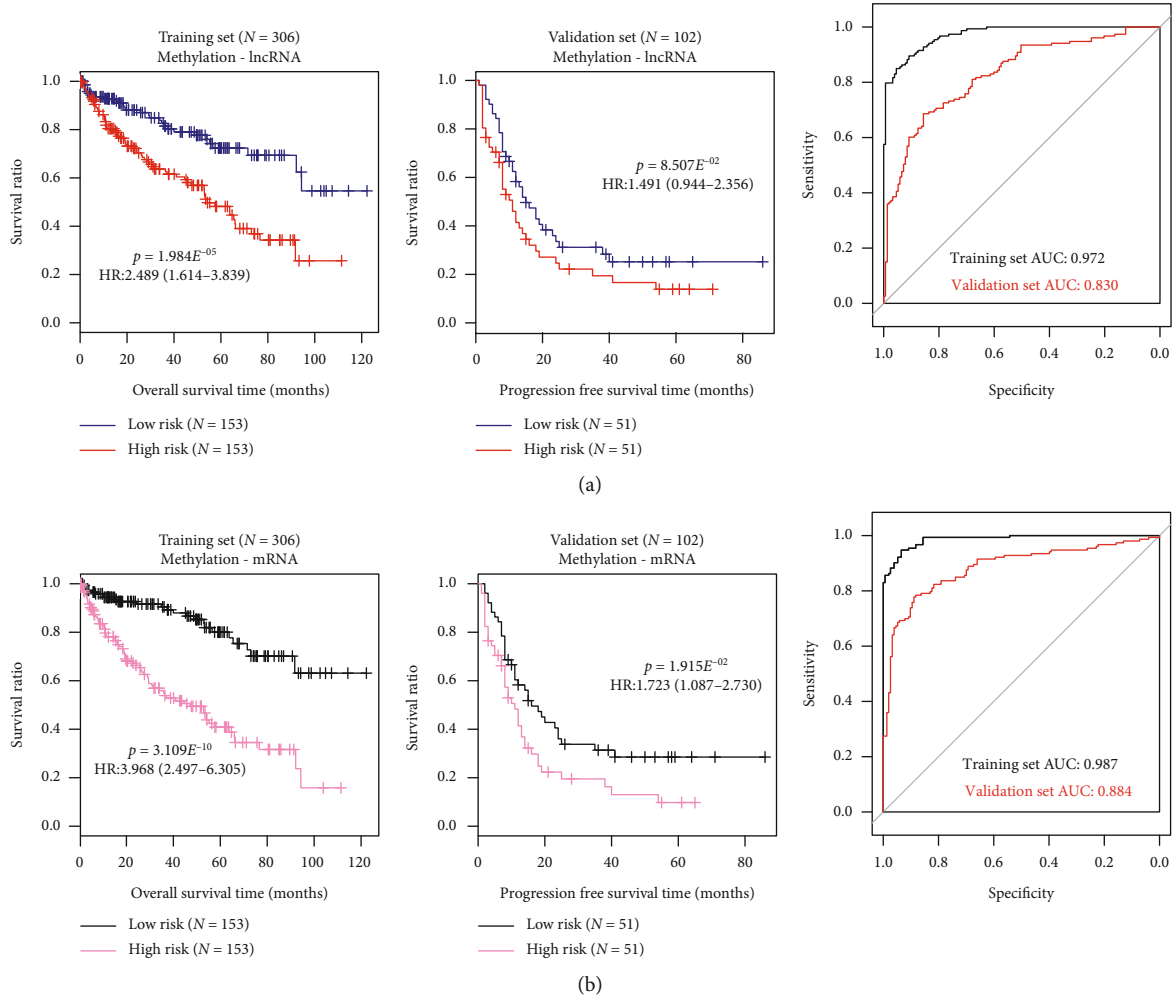


FIGURE 3: Kaplan-Meier survival and ROC curves for six-lncRNA methylation risk score model and nine-mRNA methylation risk score model in the training set and the validation set. (a) Six-lncRNA status risk score; (b) nine-mRNA status risk score.

TABLE 2: Results of Kaplan-Meier OS analysis for four risk scores.

Type	Training set		Validation set	
	<i>p</i> value	HR (95% CI)	<i>p</i> value	HR (95% CI)
Six-lncRNA status risk score	$7.318 \times 10^{-5}$	2.262 (1.495-3.424)	$2.528 \times 10^{-1}$	1.307 (0.825-2.070)
Nine-mRNA status risk score	$9.245 \times 10^{-10}$	3.867 (2.429-6.158)	$6.722 \times 10^{-2}$	1.538 (0.967-2.445)
Six-lncRNA methylation risk score	$1.984 \times 10^{-5}$	2.489 (1.614-3.839)	$8.507 \times 10^{-2}$	1.491 (0.944-2.356)
Nine-mRNA methylation risk score	$3.109 \times 10^{-10}$	3.968 (2.497-6.305)	$1.915 \times 10^{-2}$	1.723 (1.087-2.730)

HR: hazard ratio; CI: confidence interval.

TABLE 3: AUC values of ROC curves for four risk scores.

Type	Training set	Validation set
Six-lncRNA status risk score	0.891	0.814
Nine-mRNA status risk score	0.960	0.822
Six-lncRNA methylation risk score	0.972	0.830
Nine-mRNA methylation risk score	0.987	0.884

## 4. Discussion

Considering the prognostic potential of DNA methylation for ccRCC, the present work conducted a genome-wide analysis of mRNAs and lncRNAs methylation profiling of ccRCC samples from TCGA systematically, with a view to discover novel and reliable prognostic biomarkers. There were 461 DMMs and 63 DMLs between good prognosis and bad prognosis patients. Among which, 24 DMMs and 7 DMLs were

TABLE 4: Uni- and multivariable Cox regression analyses in the training set.

Clinical characteristics	Training set (N = 306)	Univariable Cox			Multivariable Cox		
		HR	95% CI	p value	HR	95% CI	p value
Age (years, mean ± SD)	64.57 ± 11.79	1.777	1.171-2.697	6.21 × 10 <sup>-3</sup>	1.352	0.846-2.158	2.07 × 10 <sup>-1</sup>
Gender (male/female)	200/106	1.141	0.746-1.743	5.43 × 10 <sup>-1</sup>	-	-	-
Neoplasm histologic grade (G1/G2/G3/G4/-)	7/130/118/47/4	2.783	2.088-3.710	3.45 × 10 <sup>-13</sup>	1.470	1.076-2.009	1.56 × 10 <sup>-2</sup>
Pathologic M (M0/M1/-)	229/52/25	4.720	3.115-7.153	7.77 × 10 <sup>-16</sup>	1.034	0.403- 2.657	1.36 × 10 <sup>-1</sup>
Pathologic N (N0/N1/-)	130/9/167	2.617	1.030-6.651	3.57 × 10 <sup>-2</sup>	0.462	0.0985-2.171	3.28 × 10 <sup>-1</sup>
Pathologic T (T1/T2/T3/T4)	151/40/107/8	2.101	1.678-2.630	5.38 × 10 <sup>-12</sup>	0.753	0.456-1.243	2.65 × 10 <sup>-1</sup>
Pathologic stage (I/II/III/IV)	148/30/73/55	2.135	1.774-2.568	8.88 × 10 <sup>-16</sup>	1.989	1.101-3.592	2.26 × 10 <sup>-2</sup>
Hemoglobin (elevated/low/normal/-)	5/151/104/46	0.423	0.276-0.646	5.54 × 10 <sup>-5</sup>	0.466	0.269-1.087	3.86 × 10 <sup>-1</sup>
Platelet qualitative (elevated/low/normal/-)	18/32/202/54	0.568	0.425-0.759	8.22 × 10 <sup>-5</sup>	0.690	0.501- 0.949	2.28 × 10 <sup>-2</sup>
White cell count (elevated/low/normal/-)	84/4/161/57	1.156	0.910-1.468	2.33 × 10 <sup>-1</sup>	-	-	-
Serum calcium (elevated/low/normal/-)	5/112/90/99	1.017	0.658-1.570	9.40 × 10 <sup>-1</sup>	-	-	-
Methylation RS model (high/low)	153/153	4.713	2.898-.665	6.50 × 10 <sup>-12</sup>	2.574	1.149-5.767	2.17 × 10 <sup>-4</sup>
Vital status (dead/alive)	97/209	-	-	-	-	-	-
Overall survival time (months, mean ± SD)	35.21 ± 28.73	-	-	-	-	-	-

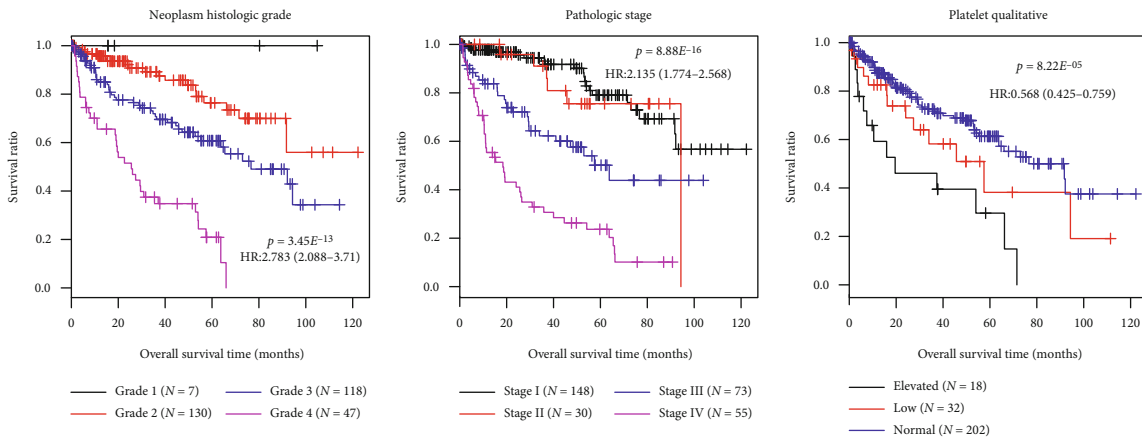


FIGURE 4: Kaplan-Meier survival curves for the training set classified by neoplasm histologic grade, pathologic stage, and platelet qualitative. (a) Neoplasm histologic grade; (b) pathologic stage; (c) platelet qualitative. Patients of the training set are separated by neoplasm histologic grade, pathologic stage, or platelet qualitative into different subgroups, separately. OS time of different subgroups is compared and analyzed using Kaplan-Meier survival analysis.

identified to be independent predictive factors by uni- and multivariable Cox regression analyses. Eventually, a nine-mRNA methylation signature and a six-lncRNA methylation signature were determined by LASSO Cox-PH regression model. Four prognostic prediction systems were established, and four risk score models were generated, including nine-mRNA methylation risk score model, nine-mRNA status risk score model, six-lncRNA methylation risk score model, and six-lncRNA status risk score model. The nine-mRNA methylation risk score model performed better than other DML-based risk scores in predicting survival outcome of ccRCC patients. This study suggests that the nine-mRNA methylation risk score is reliable and powerful for classifying ccRCC patients according to prognosis.

Our study identified a prognostic nine-mRNA methylation signature, comprising of DMRTA2, DRGX, FAM167A, FGGY, FOXI2, KRTAP2-1, TCTEX1D1, TTBK1, and UBE2QL1. DMRTA2, a member of the DMRT family, is involved in central nervous system development [21]. Methylation of DMRTA2 is associated with survival of patients with triple-negative breast cancer [22]. DRGX encodes dorsal root ganglia homeobox protein that is a transcription factor implicated in embryonic development of the nervous system. There is evidence that DGRX may be related to serum concentration of monocyte chemoattractant protein-1, an inflammation biomarker, in subjects treated with fenofibrate, an anti-inflammatory and triglyceride-lowering drug [23]. FAM167A, also namely C8orf13, belongs to the FAM167

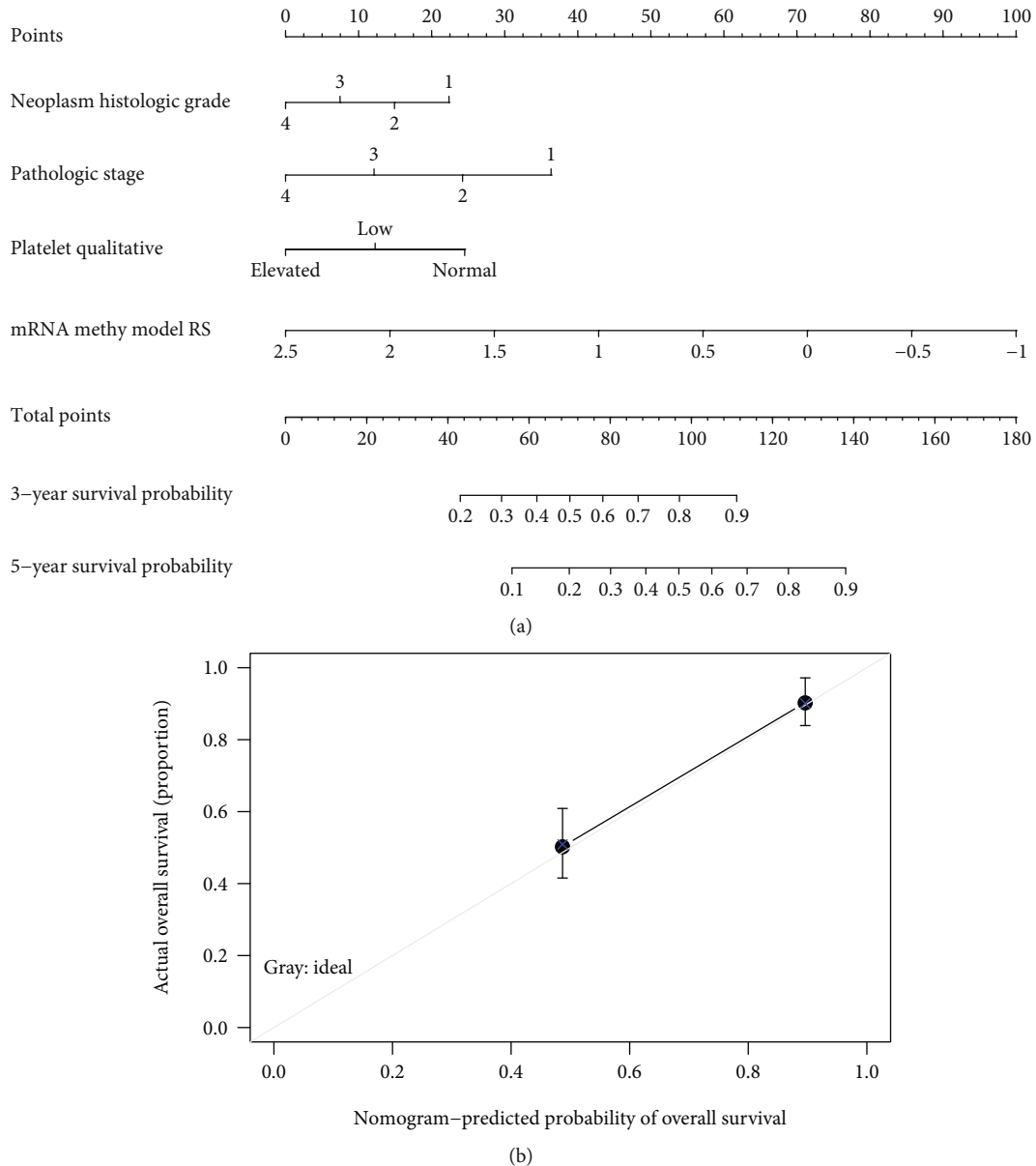


FIGURE 5: Nomogram for predicting survival of ccRCC patients. (a) The nomogram combines nine-mRNA methylation model, neoplasm histologic grade, pathologic stage, and platelet qualitative. (b) Calibration plots for predicting 5-year OS. Grey line, ideal agreement between the actual and predicted probabilities of 5-year OS; solid line, the actual nomogram.

(SEC) family. The polymorphisms of FAM167A-BLK region are linked to susceptibility of autoimmune diseases [24, 25]. Low expression of FAM167A is related to longer survival time of ccRCC patients [26]. FGGY encodes protein FGGY carbohydrate kinase domain containing proteins that phosphorylate carbohydrates [27]. FGGY is significantly upregulated during neurogenic skeletal muscle atrophy [28]. FOX family members play roles in regulating cell growth and differentiation [29], and FOXI2 expression is associated with survival of ccRCC patients [30]. KRTAP proteins are major components of hair shaft, playing crucial roles in human hair shaft keratinization [31]. KRTAP2-1 shows an association with chemosensitivity of colorectal cells to oxaliplatin [32]. TCTEX1D family members contain a conserved domain similar to C-terminus of TCTEX1, with little information

regarding their biological function [33]. TCTEX1D1 is found to be hypermethylated and downregulated in endometrial cancer [34]. Tau Tubulin Kinase 1 encoded by TTBK1 is implicated in regulating microtubule assembly and stabilization. TTBK1 phosphorylation activity is linked to several neurodegenerative diseases like Alzheimer's disease [35]. UBE2QL1 protein is related to protein ubiquitination. There is evidence that UBE2QL1 is a candidate renal tumor suppressor gene [36]. Aberrant methylation of UBE2QL1 has been found in HBV-related hepatocellular carcinoma [37]. As far as we know, for the first time, the nine differentially methylated mRNAs are reported to be of prognostic value for ccRCC patients.

The training set of this study was dichotomized by the nine-mRNA methylation risk score model into a high-risk



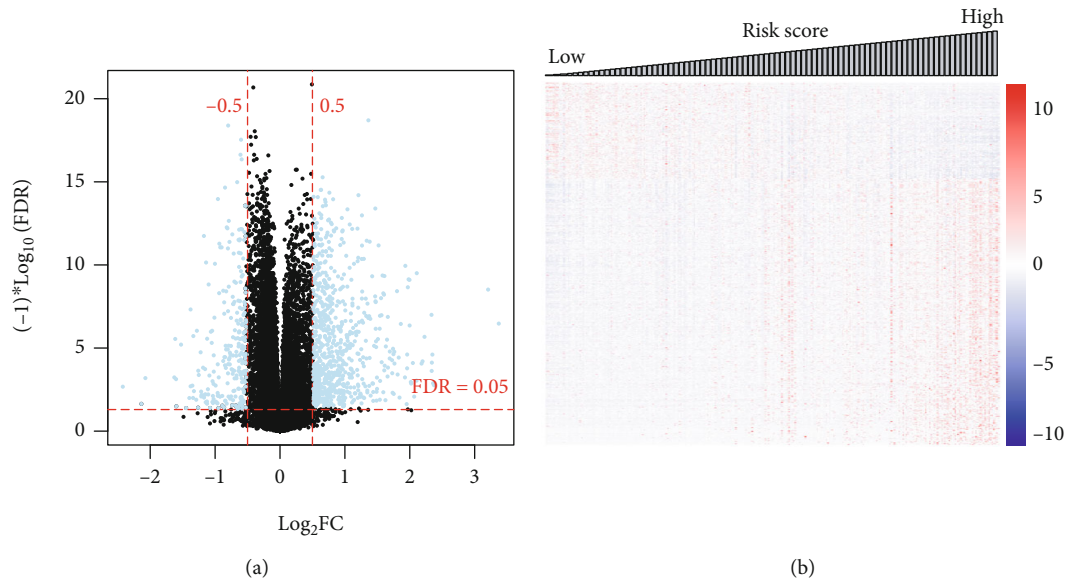


FIGURE 6: Identification of DEGs between the high-risk and low-risk groups. (a) Volcano plot of effect size ( $\log_2FC$ ) –  $\log_{10}(FDR)$  of the DEGs. (b) Risk score distribution of these DEGs expression profiles.

TABLE 5: Significant enriched KEGG pathways.

Name	Size	ES	NES	NOM $p$ value	Enriched genes
KEGG cytokine-cytokine receptor interaction	18	0.5690	1.9390	0	IL6, IL1R2, IL2RA, CCL13, IL20RB, CXCL13, CXCR3, IL21R, TNFRSF17, TNFSF14, TNFRSF9, TNFRSF18, INHBE, IFNG, XCL2, LTA, CD27, XCL1
KEGG neuroactive ligand receptor interaction	11	-0.4944	-1.7734	0.0108	GABRB3, CHRM3, VIPR1, GRIK3, PLG, CALCR, GABRQ, AVPR2, TACR1, AGTR1, F2RL3
KEGG toll-like receptor signaling pathway	2	0.8721	1.5058	0.0188	LBP, IL6
KEGG B cell receptor signaling pathway	3	0.7853	1.5119	0.0448	CD79A, CD72, CARD11

ES: enrichment score; NES: normalized enrichment score; NOM  $p$  value: nominal  $p$  value; size: the number of genes enriched in one pathway.

group and a low-risk group. Function enrichment analysis showed that the DEGs between the two risk groups were significantly enriched in cytokine-cytokine receptor interaction, neuroactive ligand receptor interaction, toll-like receptor signaling pathway, and cell receptor signaling pathways. Mounting evidences demonstrate that neuroactive ligand receptor interaction pathway is an important pathway for ccRCC [38, 39]. Toll-like receptors are critical participants in inflammation responses. Cancer-related inflammation has been accepted as an important hallmark of cancer [40]. These findings reveal that the nine prognostic mRNAs may be involved in neuroactive ligand receptor interaction and inflammation-related pathways. Further functional annotation of these mRNAs is needed to deepen our understanding of their biological implications in ccRCC prognosis.

It should be noted that this study is an extensive bioinformatics study based on published data. The results of these studies should also be further validated *in vitro* or *in vivo* models. We hope that these useful clues will help other researchers to carry out relevant research.

## 5. Conclusion

We report and validate predictive value of a nine-mRNA methylation risk score model for risk stratification of ccRCC patients. The nine-mRNA methylation signature, including DMRTA2, DRGX, FAM167A, FGGY, FOXI2, KRTAP2-1, TCTEX1D1, TTBK1, and UBE2QL1, may be a useful prognostic biomarker for ccRCC patients. Further experimental investigations are required to confirm this prognostic signature.

## Data Availability

All data used and/or analyzed in this study are available from TCGA database (<https://portal.gdc.cancer.gov>) or the EBI Array database (<https://www.ebi.ac.uk/arrayexpress/>).

## Conflicts of Interest

The authors declare that they have no competing interests.

## Supplementary Materials

Figure S1: the lambda parameter curve (left) and coefficient distribution chart (right) of (a) lncRNA and (b) mRNA. The lambda parameter curve was gained basing on cross-validation likelihood screening. The horizontal axis and vertical axis represent the different values of lambda and CVL, respectively, and the intersection of red dotted lines indicates the value of lambda parameters when CVL reaches the maximum value. The coefficient distribution charts for optimal prognosis were screened basing on Cox-PH model using L1-penalized regularized regression algorithm. (*Supplementary Materials*)

## References

- [1] J. J. Hsieh, M. P. Purdue, S. Signoretti et al., "Renal cell carcinoma," *Nature reviews Disease primers*, vol. 3, p. 17009, 2017.
- [2] B. Ljungberg, K. Bensalah, S. Canfield et al., "EAU guidelines on renal cell carcinoma: 2014 update," *European Urology*, vol. 67, no. 5, pp. 913–924, 2015.
- [3] R. Zigeuner, G. Hutterer, T. Chromecki et al., "External validation of the Mayo Clinic stage, size, grade, and necrosis (SSIGN) score for clear-cell renal cell carcinoma in a single European centre applying routine pathology," *European Urology*, vol. 57, no. 1, pp. 102–111, 2010.
- [4] M. Klutstein, D. Nejman, R. Greenfield, and H. Cedar, "DNA methylation in cancer and aging," *Cancer Research*, vol. 76, no. 12, pp. 3446–3450, 2016.
- [5] J. H. Wei, A. Haddad, K. J. Wu et al., "A CpG-methylation-based assay to predict survival in clear cell renal cell carcinoma," *Nature Communications*, vol. 6, no. 1, p. 8699, 2015.
- [6] G. Chen, Y. Wang, L. Wang, and W. Xu, "Identifying prognostic biomarkers based on aberrant DNA methylation in kidney renal clear cell carcinoma," *Oncotarget*, vol. 8, p. 5268, 2017.
- [7] E. A. Evelönn, M. Landfors, Z. Haider et al., "DNA methylation associates with survival in non-metastatic clear cell renal cell carcinoma," *BMC Cancer*, vol. 19, p. 65, 2019.
- [8] S. Cheetham, F. Gruhl, J. Mattick, and M. Dinger, "Long non-coding RNAs and the genetics of cancer," *British Journal of Cancer*, vol. 108, p. 2419, 2013.
- [9] E.-D. Kim and S. Sung, "Long noncoding RNA: unveiling hidden layer of gene regulatory networks," *Trends in Plant Science*, vol. 17, pp. 16–21, 2012.
- [10] X. Liu, Y. Hao, W. Yu et al., "Long non-coding RNA emergence during renal cell carcinoma tumorigenesis," *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*, vol. 47, pp. 735–746, 2018.
- [11] G. Malouf, J. Zhang, N. M. Tannir et al., "Charting DNA methylation of long non-coding RNA in clear-cell renal cell carcinoma," *Journal of Clinical Oncology*, vol. 33, p. 432, 2015.
- [12] B. Beuselinck, S. Job, E. Becht et al., "Molecular subtypes of clear cell renal cell carcinoma are associated with sunitinib response in the metastatic setting," *Clinical Cancer Research: an official journal of the American Association for Cancer Research*, vol. 21, no. 6, pp. 1329–1339, 2015.
- [13] M. W. Wright, "A short guide to long non-coding RNA gene nomenclature," *Human Genomics*, vol. 8, no. 1, p. 7, 2014.
- [14] K. H. Zou, K. Tuncali, and S. G. Silverman, "Correlation and simple linear regression," *Radiology*, vol. 227, pp. 617–622, 2003.
- [15] S. Huang, C. Yee, T. Ching, H. Yu, and L. X. Garmire, "A novel model to combine clinical and pathway-based transcriptomic information for the prognosis prediction of breast cancer," *PLoS Computational Biology*, vol. 10, no. 9, article e1003851, 2014.
- [16] C. Pan, X. Wang, W. Chen et al., "Reevaluation of glypican-3 as a prognostic marker in HCC using X-tile software," *Medical Oncology*, vol. 32, no. 1, p. 359, 2015.
- [17] M. E. Ritchie, B. Phipson, Y. H. Di Wu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, p. e47, 2015.
- [18] L. Yang, L. Xu, Q. Wang, M. Wang, and G. An, "Dysregulation of long non-coding RNA profiles in human colorectal cancer and its association with overall survival," *Oncology Letters*, vol. 12, pp. 4068–4074, 2016.
- [19] Z. Zhang and M. W. Kattan, "Drawing nomograms with R: applications to categorical outcome and survival data," *Annals of Translational Medicine*, vol. 5, no. 10, p. 211, 2017.
- [20] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [21] M. Poulain, N. Frydman, S. Tourpin et al., "Involvement of doublesex and mab-3-related transcription factors in human female germ cell development demonstrated by xenograft and interference RNA strategies," *Molecular Human Reproduction*, vol. 20, no. 10, pp. 960–971, 2014.
- [22] C. Stirzaker, E. Zotenko, J. Z. Song et al., "Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value," *Nature Communications*, vol. 6, p. 5899, 2015.
- [23] S. Aslibekyan, E. K. Kabagambe, M. R. Irvin et al., "A genome-wide association study of inflammatory biomarker changes in response to fenofibrate treatment in the Genetics of Lipid Lowering Drug and Diet Network," *Pharmacogenetics and Genomics*, vol. 22, no. 3, pp. 191–197, 2012.
- [24] R.-h. Song, Q. Li, X. Jia, Q.-m. Yao, B. Wang, and J.-a. Zhang, "Polymorphisms of FAM167A-BLK region confer risk of autoimmune thyroid disease," *DNA and Cell Biology*, vol. 37, no. 11, pp. 932–940, 2018.
- [25] S. Chen, W. Wu, J. Li et al., "Single nucleotide polymorphisms in the FAM167A-BLK gene are associated with polymyositis/dermatomyositis in the Han Chinese population," *Immunologic Research*, vol. 62, no. 2, pp. 153–162, 2015.
- [26] Y. Wang, Y. Wang, and F. Liu, "A 44-gene set constructed for predicting the prognosis of clear cell renal cell carcinoma," *International Journal of Molecular Medicine*, vol. 42, pp. 3105–3114, 2018.
- [27] Y. Zhang, O. Zagnitko, I. Rodionova, A. Osterman, and A. Godzik, "The FGGY carbohydrate kinase family: insights into the evolution of functional specificities," *PLoS Computational Biology*, vol. 7, no. 12, article e1002318, 2011.
- [28] B. De Las Casas and D. Waddell, "FGGY carbohydrate kinase domain containing is upregulated during neurogenic skeletal muscle atrophy," *The FASEB Journal*, vol. 30, pp. 1009–1016, 2016.

- [29] M. Katoh, M. Igarashi, H. Fukuda, H. Nakagama, and M. Katoh, "Cancer genetics and genomics of human FOX family genes," *Cancer Letters*, vol. 328, pp. 198–206, 2013.
- [30] Z. Jia, F. Wan, Y. Zhu et al., "Forkhead-box series expression network is associated with outcome of clear-cell renal cell carcinoma," *Oncology Letters*, vol. 15, no. 6, pp. 8669–8680, 2018.
- [31] H. Fujikawa, A. Fujimoto, M. Farooq, M. Ito, and Y. Shimomura, "Characterization of the human hair keratin-associated protein 2 (KRTAP2) gene family," *Journal of Investigative Dermatology*, vol. 132, no. 7, pp. 1806–1813, 2012.
- [32] Y.-J. Zhang, A.-J. Li, Y. Han, L. Yin, and M.-B. Lin, "Inhibition of Girdin enhances chemosensitivity of colorectal cancer cells to oxaliplatin," *World Journal of Gastroenterology*, vol. 20, p. 8229, 2014.
- [33] A. A. Gholkar, S. Senese, Y.-C. Lo et al., "Tctex1d2 associates with short-rib polydactyly syndrome proteins and is required for ciliogenesis," *Cell Cycle*, vol. 14, no. 7, pp. 1116–1125, 2015.
- [34] C. Men, H. Chai, X. Song, Y. Li, H. Du, and Q. Ren, "Identification of DNA methylation associated gene signatures in endometrial cancer via integrated analysis of DNA methylation and gene expression systematically," *Journal of Gynecologic Oncology*, vol. 28, no. 6, 2017.
- [35] V. Nozal and A. Martinez, "Tau Tubulin Kinase 1 (TTBK1), a new player in the fight against neurodegenerative diseases," *European Journal of Medicinal Chemistry*, vol. 161, pp. 39–47, 2018.
- [36] N. C. Wake, C. J. Ricketts, M. R. Morris et al., "UBE2 QL 1 is disrupted by a constitutional translocation associated with renal tumor predisposition and is a novel candidate renal tumor suppressor gene," *Human Mutation*, vol. 34, no. 12, pp. 1650–1661, 2013.
- [37] C. Ye, R. Tao, Q. Cao et al., "Whole-genome DNA methylation and hydroxymethylation profiling for HBV-related hepatocellular carcinoma," *International Journal of Oncology*, vol. 49, pp. 589–602, 2016.
- [38] X. Liu, J. Wang, and G. Sun, "Identification of key genes and pathways in renal cell carcinoma through expression profiling data," *Kidney and Blood Pressure Research*, vol. 40, no. 3, pp. 288–297, 2015.
- [39] W. Yang, K. Yoshigoe, X. Qin et al., "Identification of genes and pathways involved in kidney renal clear cell carcinoma," *BMC Bioinformatics*, vol. 15, no. S17, pp. 1–10, 2014.
- [40] Y. Chang, H. An, L. Xu et al., "Systemic inflammation score predicts postoperative prognosis of patients with clear-cell renal cell carcinoma," *British Journal of Cancer*, vol. 113, p. 626, 2015.