

Research Article

Novel prognostic prediction model constructed through machine learning on the basis of methylation-driven genes in kidney renal clear cell carcinoma

Weihao Tang¹, Yiling Cao² and  Xiaoke Ma³

¹Chengdu Foreign Language School, International Department, Class AP-1, Grade 2019, Chengdu, Sichuan, China; ²West China School of Medicine, Sichuan University, Chengdu, Sichuan, China; ³School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, China

Correspondence: Xiaoke Ma (xkma@xidian.edu.cn)



Kidney renal clear cell carcinoma (KIRC) is a common tumor with poor prognosis and is closely related to many aberrant gene expressions. DNA methylation is an important epigenetic modification mechanism and a novel research target. Thus, exploring the relationship between methylation-driven genes and KIRC prognosis is important. The methylation profile, methylation-driven genes, and methylation characteristics in KIRC was revealed through the integration of KIRC methylation, RNA-seq, and clinical information data from The Cancer Genome Atlas. The Lasso regression was used to establish a prognosis model on the basis of methylation-driven genes. Then, a trans-omics prognostic nomogram was constructed and evaluated by combining clinical information and methylated prognosis model. A total of 242 methylation-driven genes were identified. The Gene Ontology terms of these methylation-driven genes mainly clustered in the activation, adhesion, and proliferation of immune cells. The methylation prognosis prediction model that was established using the Lasso regression included four genes in the methylation data, namely, *FOXI2*, *USP44*, *EVI2A*, and *TRIP13*. The areas under the receiver operating characteristic curve of 1-, 3-, and 5-year survival rates were 0.810, 0.824, and 0.799, respectively, in the training group and 0.794, 0.752, and 0.731, respectively, in the testing group. An easy trans-omics nomogram was successfully established. The C-indices of the nomogram in the training and the testing groups were 0.8015 and 0.8389, respectively. The present study revealed the overall perspective of methylation-driven genes in KIRC and can help in the evaluation of the prognosis of KIRC patients and provide new clues for further study.

Introduction

Kidney renal clear cell carcinoma (KIRC) is the most common type of renal carcinoma and accounts for approximately 80–90% of all kidney tumors [1]. KIRC is characterized by a high risk of metastasis and the main cause of death from kidney tumors [2,3]. With the application of medical imaging, such as computed tomography and ultrasound, the early discovery of tumor is improved. However, despite substantial advances in its diagnosis and treatment, the mechanisms underlying KIRC are not fully understood, and the prognosis is still poor [4].

DNA methylation is an important epigenetic modification to regulate gene expression. In some promoter-related cytosine–phosphate–guanine (CpG) sites, DNA methylation often contributes to the silencing of downstream genomic regions [1]. Studies show that tumor suppressor genes can be inhibited by hypermethylation, whereas oncogenes can be activated by hypomethylation [5,6]. Many studies

Received: 11 May 2020
Revised: 24 June 2020
Accepted: 06 July 2020

Accepted Manuscript online:
07 July 2020
Version of Record published:
21 July 2020

have explored the abnormal DNA methylation in cervical [6] and breast [7] cancers, which may be possible markers for diagnoses, therapeutic targets, and prognosis [5,8]. The DNA methylation in KIRC has been studied for several years [9]. The suppression of *Wnt* antagonists through DNA methylation plays an important role in the proliferation of many RCC cell lines and patient tumor samples [10]. The DNA methylation of several genes, such as *PARVG*, *PLCB2*, and *RAC2*, are correlated with RCC prognosis [11]. The hypermethylation of cystatin 6 and *LAD1* appears to be associated with poor prognosis and poor response to some antiangiogenic therapies in RCC [12].

With the rapid development of computer technology, an increasing number of biomarkers of various tumors are identified using bioinformatics analysis, and many tumor-associated databases are established [13]. The Cancer Genome Atlas (TCGA) is a public database that includes 33 cancer types and matched clinical data [14]. Studies reveal many genes, such as *CEP55*, *ACAA1*, *ACADSB*, *ALDH6A1*, and *AUH*, that may serve as potential diagnostic and prognostic biomarkers for KIRC [13,15]. The methylation-driven gene is differentially expressed in the control and the disease groups, and the methylation status is negatively correlated with corresponding gene expression value. At present, systemic analysis on methylation-driven genes in KIRC remains limited. In addition, no multivariate prognostic model based on methylation-driven genes especially a clinical prediction model combining DNA methylation data and clinical data exists.

In the present study, we combined the TCGA–KIRC clinical, methylation, and expression data to identify the DNA methylation profiling in KIRC, discover the prognosis-related methylation genes and position, establish methylation prognostic prediction model, and construct an easy trans-omics prognostic nomogram. The findings of the present study can provide valuable data resources for KIRC in clinical and molecular study in the future.

Methods

Preparation of dataset

The KIRC methylation data, RNA-seq expression counts, and clinical information were downloaded from the TCGA website (<https://cancergenome.nih.gov/>). After combining the methylation data, the methylation positions and gene names were matched, and the average methylation level of each gene was calculated (normal samples = 160, tumor samples = 325). The R package DEseq2 (version 1.20.0) was used to standardize the RNA-seq expression counts and obtain the gene expression value matrix. The follow-up time, survival state, age, gender, and tumor stage were extracted from clinical information and used for subsequent analysis. A total of 317 tumor samples were obtained by matching the methylation data, gene expression value, and clinical information matrices in accordance with the sample number. The samples with a survival period of less than 30 days were removed, and a total of 294 tumor samples were obtained for survival analysis.

Differential methylation analysis

The R package limma (version 3.36.5) was used to calculate the differential methylation between the tumor and the normal groups. The fold change value of methylation levels between the tumor and the normal groups was calculated and then standardized through logarithm. Wilcox test was used for the statistical analysis of methylation data and adjusted. *P* value < 0.05 was used as a cutoff standard. The R package pheatmap (version 1.0.10) was used to visualize differential methylation.

Calculation of methylation-driven genes

The matched gene expression values matrix, the DNA methylation values of the tumor group, and the DNA methylation values of the normal group were input into the R package MethylMix (version 2.10.2) to obtain the methylation-driven gene. The Wilcox test was used for statistical analysis, and *P* value was adjusted using the Bonferroni method. Adjusted *P* value < 0.05 was used as cutoff standard.

Gene Ontology (GO) analysis

The GO analysis for all methylation-driven genes was performed using the R package cluster profiler (version 3.8.1) based on org.Hs.eg.dbdatabase (version 3.6.0). The parameters of enrichGO analysis function were: ont = “BP,” pAdjustMethod = “BH,” pvalueCutoff = 0.05. The GO analysis results were visualized using the enrichplot (version 1.2.0) and the GOplot package (version 1.0.2).

Construction of the methylation prognostic prediction model by using the Lasso regression

The clinical survival information and methylation data of driven genes were integrated in accordance with the sample name. Survival analysis and result visualization were performed using the R package survival (version 2.43-1) and the SurfMiner package (version 0.4.3). The “Handout” method was used for model construction and testing. Briefly, all samples were divided into the training (70%) and the testing (30%) groups by using the hierarchical random method through the “createDataPartition” function in the machine learning R package “caret” (version 6.0-81) to keep each categorical variable of the data in the subset consistent with the original proportion of the overall data, thereby ensuring that the data distribution of the training and the testing groups is consistent. The methylation prediction model in the training group was established as follows. The cv.glmnet function of the machine learning package glmnet (version 2.0-16) was used to calculate the Lasso rank. The glmnet function was used to calculate the Lasso by using the Cox multivariate regression model. The model with the best performance and the least number of independent variables was selected using the best lambda value. Samples were divided into high- and low-risk groups, and the survival rate was further compared using this methylation prognostic prediction model. $P < 0.05$ was regarded as the cutoff standard. The predictive efficiency was evaluated using the area under the receiver operating characteristic curve (AUC). This methylation prognostic prediction model was also validated in the testing group.

Correlation analysis of methylation-driven genes and methylation positions

In accordance with the name of selected methylation-driven genes in the previous methylation prognostic prediction model, the multiple methylation position details of these genes were extracted from the TCGA methylation dataset. The Rcorr function in the Hmisc package (version 4.1-1) was used to calculate the correlation between the expression value of methylation-driven genes and methylation positions. The significant correlations were screened using the absolute value of correlation coefficient > 0.4 and the P value of correlation test < 0.05 . These correlations were visualized using the Cytoscape software (version 3.7.1).

Construction of trans-omics prognostic nomogram

The clinical information, such as age, gender, and tumor stage, combined with the score of methylation prognostic model was used to construct the trans-omics prognostic prediction model. The machine learning package caret (version 6.0-81) was used to divide all samples through random stratification. Approximately 70% and 30% of the samples were used as the training and the testing groups, respectively. The rms package (version 5.1-2) was used to build the Cox proportional hazards model, draw the nomogram, and calculate the c-index to evaluate the efficiency of the trans-omics prediction model in the training and the testing groups. In both groups, the machine learning calibration method was used to calculate and draw the calibration curve to evaluate the model's accuracy.

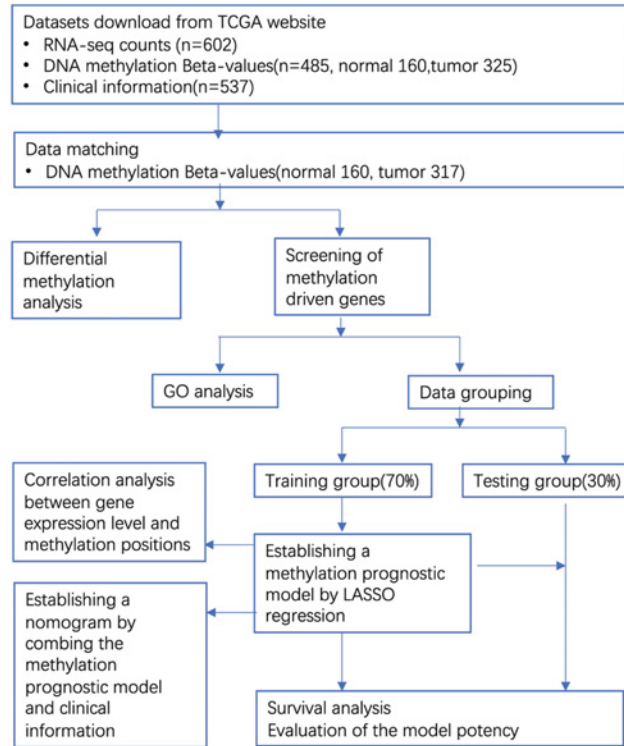
Results

Differential methylation gene analysis

Figure 1A shows the flow diagram of the present study. We obtained the related RNA-seq counts ($n=602$), DNA methylation β -values ($n=485$, normal = 160, tumor = 325), and clinical information ($n=537$) from TCGA. The methylation of some samples was not tested, and the methylation and the expression data did not have one-to-one correspondence. After matching the methylation data with the expression data in accordance with the sample number, we obtained DNA methylation samples ($n=477$, normal = 160, tumor = 317) to screen methylation-driven genes. Then, we constructed and verified the methylation prognostic model by using machine learning.

A total of 134 differential methylation genes with $|\log_2 \text{FC}| > 1$ were adjusted. P values < 0.05 are shown in the hierarchical clustering heat map (Figure 1B). The methylation levels of the normal and the tumor groups were significantly different, and the hypomethylation and the hypermethylation genes were distinguished evidently by hierarchical clustering. Results also showed that the hypomethylation and the hypermethylation genes contained protein coding genes, such as *CECR6*, *RNF180*, *CXCL1P*, *HIST3H2A*, and *ZNF492*, and noncoding genes, such as *HOXB13-AS1_2*, *RP11-343J3.2*, *DGUOK-AS1*, *AC009506.1*, and *AL021918.2*, which indicated that the extent of methylation was universal for genes.

(A)



(B)

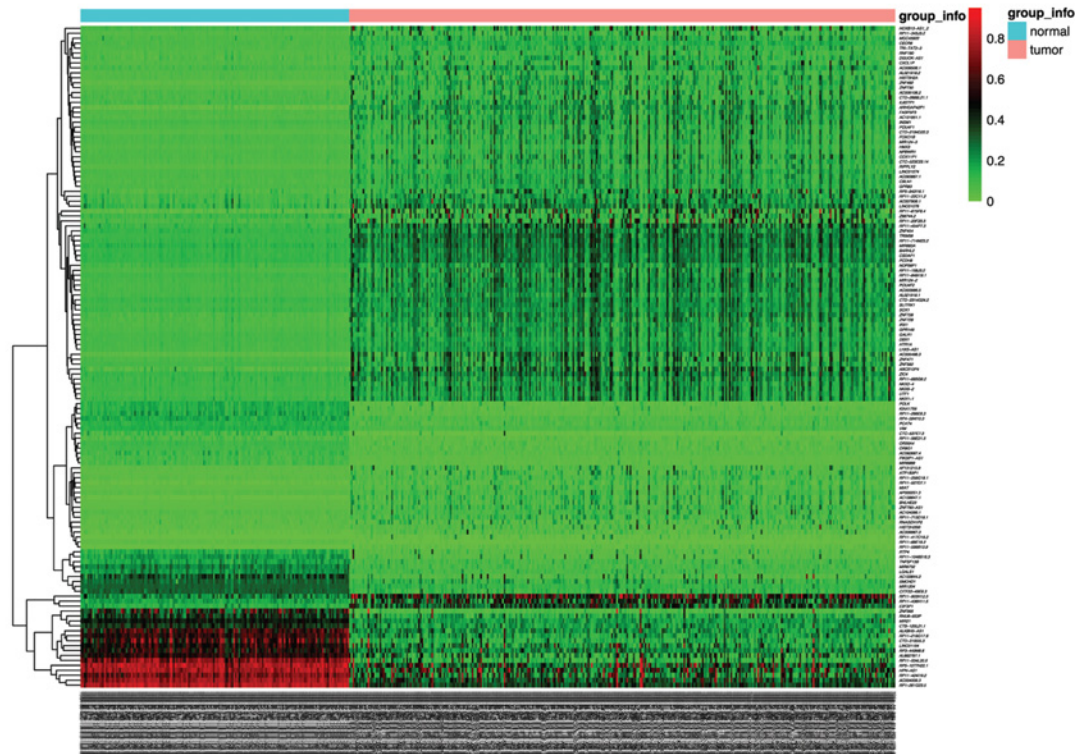


Figure 1. Workflow and hierarchical clustering of differential methylation genes

(A) Workflow of the experiments. (B) Heat map of differential methylation genes between tumor group and normal group. Genes with $|\log_2FC| < 1$ and adjusted. P value < 0.05 is shown in the map. Top band is grouping information (red for tumor group, blue for normal group); left is clustering tree; bottom is sample number; right is the name of the methylation gene.

Methylation-driven gene discovery and GO analysis

A total of 242 methylation-driven genes were screened using the MethylMix package. The detailed information of all methylation-driven genes is listed in Supplementary Table S1. The top 10 genes with the highest absolute correlation coefficient between methylation state and gene expression value are shown in Figure 2A. These genes were *XIST*, *CCDC8*, *KRTCAP3*, *SMIM3*, *DCAF4L2*, *ZNF471*, *ALDOC*, *LGALS12*, *VEGFA*, and *AQP1*. The correlation coefficients were between -0.681 and -0.875 and showed a significant difference ($P < 0.05$).

The GO analysis of the 242 methylation-driven genes showed that in biological process, 242 methylation-driven genes was mainly clustered in the activation of immune cells (GO:1903039: positive regulation of T-cell activation), the adhesion of immune cells (GO:0050671: positive regulation of leukocyte cell–cell adhesion), and the proliferation of immune cells (GO:0050870: positive regulation of T-cell proliferation; GO:0010951: positive regulation of mononuclear cell proliferation) (Figure 2B). These biological processes were closely related to tumorigenesis and development.

Figure 2C–J shows the typical gene methylation pattern of the top 10 genes, and Figure 2A,C shows the methylation density distribution of the *LGALS12* gene, which was hypomethylated in the tumor group with simple methylation component. Figure 2D shows the correlation coefficient ($r = -0.706$, $P = 3.295 \times 10^{-49}$) between the expression value of *LGALS12* and its methylation level. *AQP1* also had low methylation in the tumor group but had two components with different distributions. The correlation coefficient between the expression value and the methylation state was -0.681 (P value = 1.724×10^{-44} ; Figure 2E,F). Figure 2G shows that *CCDC8* was a hypermethylation gene with a single methylation component in the tumor group. Figure 2H shows the correlation coefficient ($r = -0.804$, $P = 5.893 \times 10^{-73}$) between the expression value of *CCDC8* and the methylation level. The *KRTCAP3* gene was also hypermethylated in the tumor group with two components in different distribution. The correlation coefficient between the expression value and the methylation level was -0.765 ($P = 2.95 \times 10^{-62}$, Figure 2I,J).

Construction and evaluation of the methylation prognostic prediction model by using machine learning

Samples with survival period less than 30 days were removed, and the remaining samples were stratified to obtain the training ($n=206$) and the testing ($n=88$) groups. First, the Lasso method was used to screen the variables and establish the survival prognosis model in the training group. As shown in Figure 3A,B, the Lasso logistic regression was performed in 242 methylation-driven genes. Certain coefficients were accurately reduced to 0 by forcing the total absolute value of the regression coefficients to be less than the constant value, and the most powerful prognostic predictors were selected. The survival prognosis model included four gene methylation data, namely, *FOXI2*, *USP44*, *EVI2A*, and *TRIP13* with Lasso coefficients of 1.7373362, 0.4491624, -1.7746901 , and -3.2954915 , respectively.

We used the present model to score and divide the training group into high- and low-risk groups in accordance with the best segmentation point. Compared with the survival rate of the two groups, the Kaplan–Meier (K–M) curve showed that the survival rate of the high-risk group was significantly lower than that of the low-risk group ($P < 0.001$, Figure 3C). The ROC curve was used to evaluate the prediction efficiency. The areas of 1-, 3-, and 5-year survival rates were 0.810, 0.824, and 0.799, respectively. Results indicated that this model had a good prediction ability for the training group (Figure 3D).

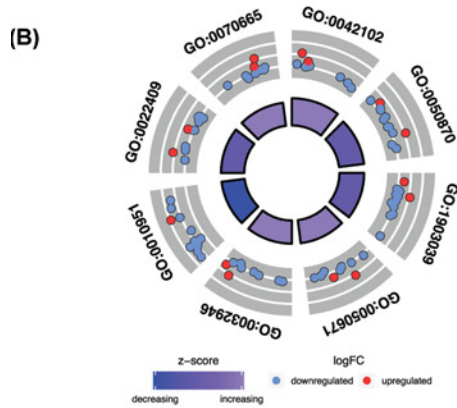
Similarly, we used the same model to score the testing group and compared the survival rate of the high- and the low-risk groups. The K–M curve also showed that the survival rate of the high-risk group was significantly lower than that of the low-risk group ($P = 0.001$, Figure 3E). The ROC curve indicated that the 1-, 3-, and 5-year survival rates were 0.794, 0.752, and 0.731, respectively. The prediction ability of the model was also desirable for the testing group.

Analysis of methylation positions of the key genes in the prediction model

We further investigated the methylation and the main methylation positions of the four key genes in the prediction model. Results showed that *USP44* was a hypermethylation gene in the tumor group (Figure 4A), and the most related methylation positions were cg22538054, cg23982858, cg22802813, and cg17368254 (Figure 4B). *FOXI2* was also a hypermethylation gene (Figure 4C), and the most related methylation positions were cg19509778, cg24718722, cg08829841, and cg26115633 (Figure 4D). For the low methylation gene (Figure 4E,G), the most relevant methylation positions of *EVI2A* were cg2332595 and cg22473770 (Figure 4F), and the most relevant methylation positions of *TRIP13* were cg03637066, cg11421768, and cg12705693 (Figure 4H). These selected methylation positions will contribute to further identify the role of these genes in KIRC development.

(A)

Gene	Normal methylation	Tumor methylation	Log(FC)	Adjusted.P.value	Cor_coefficient	Cor_test P-value
<i>XIST</i>	0.834	0.811	-0.039	2.46e-02	-0.875	4.12e-101
<i>CCDC8</i>	0.339	0.454	0.423	2.23e-25	-0.804	5.89e-73
<i>KRTCAP3</i>	0.450	0.575	0.354	1.09e-20	-0.765	2.95e-62
<i>SMIM3</i>	0.494	0.435	-0.185	7.28e-17	-0.765	3.13e-62
<i>DCAF4L2</i>	0.884	0.860	-0.040	3.01e-39	-0.722	2.24e-52
<i>ZNF471</i>	0.073	0.213	1.544	2.34e-45	-0.719	9.29e-52
<i>ALDOC</i>	0.427	0.362	-0.239	7.86e-29	-0.710	6.33e-50
<i>LGALS12</i>	0.570	0.412	-0.466	2.43e-64	-0.706	3.29e-49
<i>VEGFA</i>	0.487	0.364	-0.419	8.22e-54	-0.689	6.68e-46
<i>AQP1</i>	0.506	0.358	-0.500	3.01e-36	-0.681	1.72e-44



ID	Description
GO:0042102	positive regulation of T cell proliferation
GO:0050870	positive regulation of T cell activation
GO:1903039	positive regulation of leukocyte cell-cell adhesion
GO:0050671	positive regulation of lymphocyte proliferation
GO:0032946	positive regulation of mononuclear cell proliferation
GO:0010951	negative regulation of endopeptidase activity
GO:0022409	positive regulation of cell-cell adhesion
GO:0070665	positive regulation of leukocyte proliferation

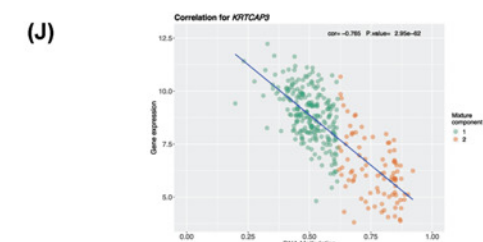
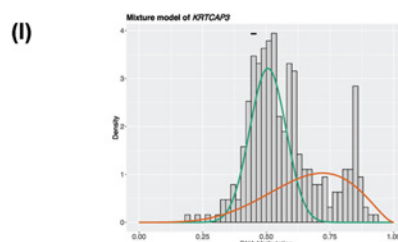
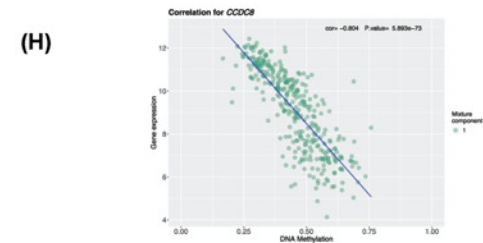
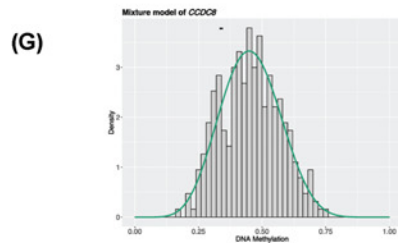
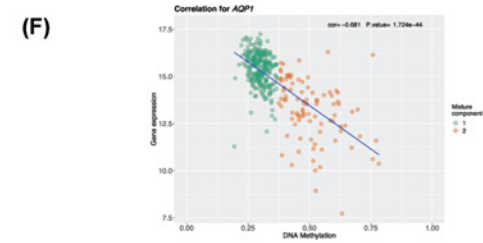
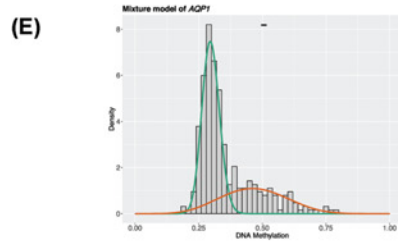
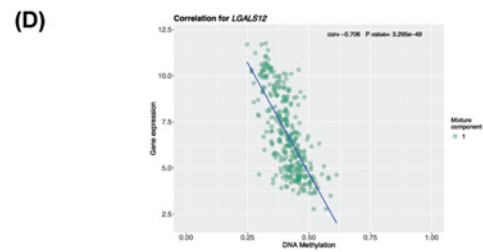
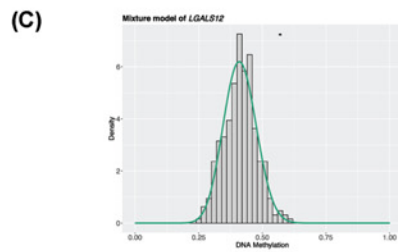


Figure 2. Methylation-driven genes

(A) Top 10 genes with the highest absolute correlation between the methylation level and expression value. (B) GO analysis of all methylation-driven genes. The left figure shows the Z-score and gene distribution of the top eight GO terms, and the right figure shows the name of the top eight GO terms. (C and D) Density map of the representative hypomethylation gene *LGALS12* (a single component distribution pattern) and the scatter map of correlation with gene expression value; the black short line in panel (C) represents the average methylation level of the normal group. (E and F) Density map of the representative hypomethylation gene *AQP1* (two-component distribution pattern) and the scatter map of correlation with gene expression value; the black short line in panel (E) represents the average methylation level of the normal group. (G and H) Density map of the representative hypermethylation gene *CCDC8* (a single component distribution pattern) and the scatter map of correlation with gene expression value; the black short line in panel (G) represents the average methylation level of the normal group. (I and J) Density map of the representative hypermethylation gene *KRTCAP3* (two-component distribution pattern) and the scatter map of correlation with gene expression value; the black short line in panel (I) represents the average methylation level of the normal group.

Easy trans-omics prognostic nomogram combined with methylation prediction model and clinical information

Tumor prognosis was closely related to clinical stage and other indicators. We combined several key clinical indicators, such as age, gender, tumor stage, and methylation model score, to create a simple prognostic nomogram and establish a comprehensive and easy prediction system. Figure 5A shows that the training group had the largest weight of the tumor stage, same extent of age and methylation prognosis score, and small gender effect. Through the score of these indicators, the 1-, 3-, and 5-year survival rates were easily obtained by calculating the total scores and the corresponding survival rate. The C-indices of the training and the testing groups were 0.8015 and 0.8389, respectively, which indicated that the prediction system was effective and would be verified in other cases. The nomogram displayed high levels of accuracy to predict the 1-, 3-, and 5-year overall survival of patients with KIRC in the training and the testing groups, as shown in the calibration curves (Figure 5B–G). Therefore, the prognostic nomogram works accurately for patients with KIRC.

Discussion

In our study, we have found many differential methylation genes and positions. Methylation occurs in coding and noncoding genes. For coding genes, DNA methylation has been observed in the suppression of specific tumor suppressor genes [16]. The methylation loss of *CDKN2A* and *CDKN2B* is demonstrated in myeloma patients [16]. The hypomethylation of *SALL4*, a member of the zinc-finger transcription factor gene family, is responsible for the aberrant expression of *SALL4* in B-ALL [17]. Another level of methylation is the influence of noncoding genes on protein expression regulation. microRNAs and ncRNAs may have important roles [13]. The methylation-mediated loss of miR-34b/c can up-regulate its oncogenic target genes [18,19]. The hypermethylation of mir-127 and mir-125b-1 in breast cancer is closely associated with tumor metastasis [20]. The methylation of noncoding genes has an important prospect, and their direct or indirect regulation with gene expression needs to be studied. Our study also indicates that the methylation of coding and noncoding genes extensively exists in KIRC.

We have used the negative correlation between the gene methylation level and expression value to select methylation-driven genes via the R package MethylMix, a generally accepted method [21,22]. Otherwise, establishing the corresponding relationship between methylation and expression value is difficult. In accordance with this criterion, more than 200 methylation-driven genes are identified. Their functions are concentrated on immunoregulation, such as the activation, proliferation, and adhesion of immune cells, which are closely related to the mechanism of tumorigenesis and development. KIRC has a high level of tumor-infiltrating immune cells [23]. The activated CD8+T cells are associated with the prognosis of many cancers, including RCC, and the infiltrating CD4+ T cells can regulate RCC cell proliferation by modulating [24]. Our GO results suggest that these methylation-driven genes can be significant in immune infiltration in KIRC.

In addition, we have found that the methylation of some genes has different distribution patterns in different samples, indicating that different subtypes may be present in KIRC. Previous studies have reported that several genes, such as *NTHL1*, *ZCCHC24*, and *SNX1b*, have different methylation status in different subtypes of breast cancer [25]. These features need further study, and these genes may become potential clinical biomarkers for KIRC subtyping. We have calculated the average methylation level of the genes in the present study. However, the methylation

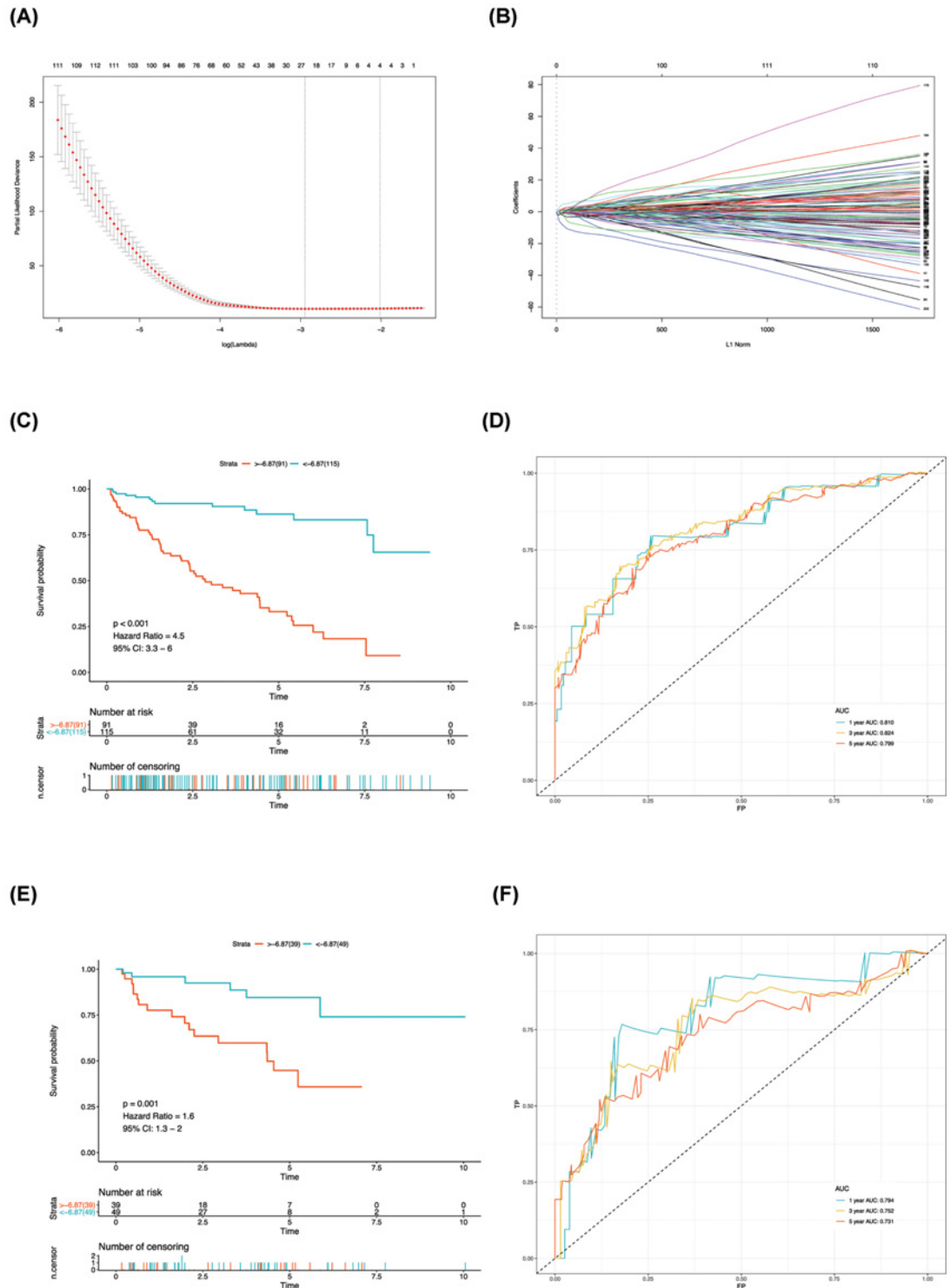


Figure 3. Establishment and validation of methylation prognostic prediction model via machine learning

(A) Selection of the tuning parameter (λ) in the Lasso model. The Y-axis represents partial likelihood deviance, and the lower X-axis represents the $\log(\lambda)$. Numbers along the upper X-axis represent the average number of predictors. Red dots indicate average deviance values for each model with a given λ , where the model provides its best fit to data. (B) Lasso coefficient profiles of the survival-associated methylations. The gray dotted vertical line is the value selected using 10-fold cross-validation in A. (C) This part shows the Kaplan–Meier (K–M) curves for high- and low-risk groups in the training set. (D) ROC curves for patients under training set at 1, 3, and 5 years. (E) This part shows the K–M curves for high- and low-risk groups in the testing set. (F) ROC curves for patients under testing set at 1, 3, and 5 years. AUC, area under the ROC curve.

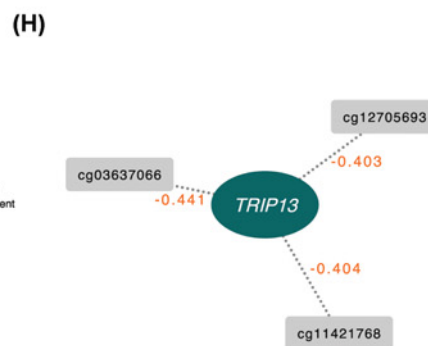
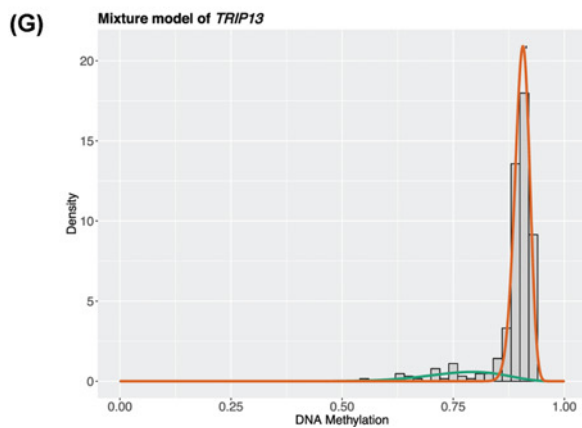
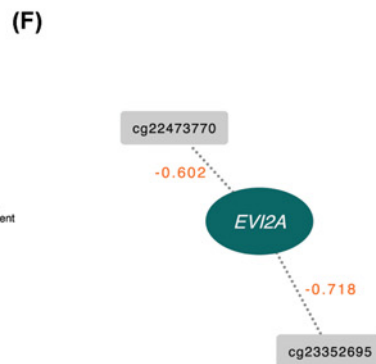
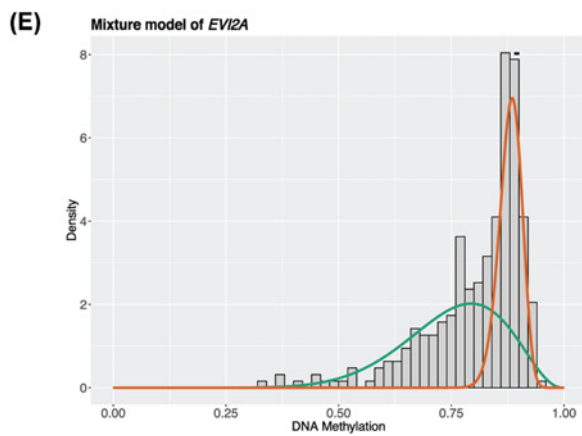
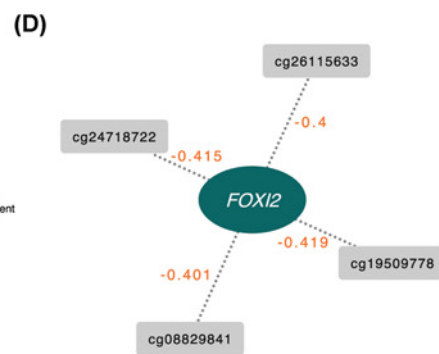
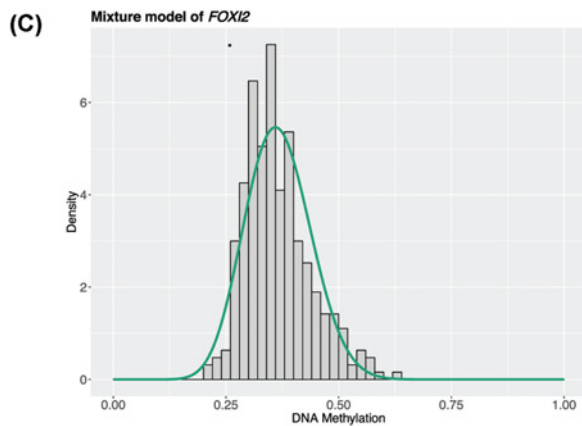
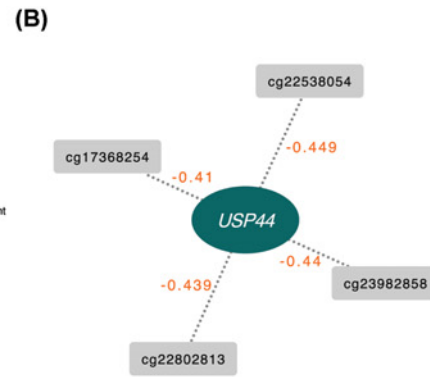
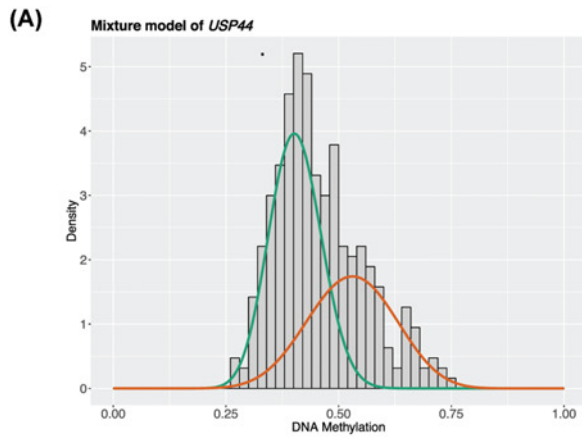


Figure 4. Correlation analysis of genes and methylation position in the prognosis model

(A) Methylation density map of *USP44* gene. (B) Correlation analysis of *USP44* gene expression level and methylation position (the value on the dotted line is the correlation coefficient, correlation test $P < 0.05$). (C) Methylation density map of *FOXI2* gene. (D) Correlation analysis of *foxi2* gene expression level and methylation site position (the value on the dotted line is the correlation coefficient, correlation test $P < 0.05$). (E) *EVI2A* gene methylation density map. (F) Correlation analysis of *EVI2A* gene expression level and methylation position (the value on the dotted line is the correlation coefficient, correlation test $P < 0.05$). (G) Methylation density map of *TRIP13* gene. (H) Correlation analysis of *TRIP13* gene expression level and methylation position (the value on the dotted line is the correlation coefficient, correlation test $P < 0.05$).

levels of different methylation positions in the same gene are different. Our analysis also provides several methylation positions with the largest correlation among key genes in the model. Further investigation on these positions can provide evidence for single-gene methylation.

The Lasso regression, a kind of machine learning method, is suitable for the multivariate selection. This method shrinks the regression coefficients, thereby effectively selecting important predictors and the established model that can compress the number of variables as much as possible [26,27]. In our study, we have used the selected methylation-driven genes as variables and established a multivariate prognostic model by using the machine learning method. The model only contains four methylation genes and is relatively simple. These genes and their methylation have not been reported in KIRC. For the functions of these genes, *USP44* is related to proliferation, migration and invasion, induced apoptosis, and arrested cell cycle in the G2/M phase in the established glioma cell lines [28]. *FOXI2* methylation may be associated with increased risk of oral and colorectal cancers [29,30]. *EVI2A* hypomethylation may be associated with head and neck squamous cell carcinomas [31]. *TRIP13* plays an indispensable role in cell progression, contributing to tumorigenesis and drug resistance [32]. The functions of these genes are closely related to tumors, further indicating that they may play important roles in KIRC. By using these gene methylation statuses as score criterion, we can easily distinguish high- and low-risk patients through this model. The model's predictive efficiency (i.e. AUC) is also satisfactory, which is verified in the testing group. The change in the DNA methylation pattern may be a good indicator of tumorigenesis and development. Several studies have mentioned the clinical application of multivariate predictive model through DNA methylation. Studies have reported the methylation profiles of HCC tumor DNA and successfully constructed a diagnostic prediction model to predict prognosis and survival rate [33]. Another study has also constructed a survival prognosis model by using the DNA methylation signature in ovarian serous cystadenocarcinoma, which shows high sensitivity and specificity to predict the prognostic survival of patients [34]. Our results demonstrate that the model shows a superior performance for prediction and has potential clinical application.

Tumor prognosis cannot be determined using a single factor. We combined the methylation prognostic prediction model with the clinical information of patients and constructed an easy and effective trans-omics prognostic nomogram to further study the clinical prediction model. Trans-omics integrates clinical and molecular multiomics [35]. Many research studies combine biomarkers and clinical indicators to construct the clinical model. In lung cancer, serum biomarkers ProGRP, CEA, SCC, and CYFRA21-1 are combined with clinical information to construct patient and nodule risk models [36]. A prognostic prediction model reveals that *C1QTNF3* is a promising biomarker for prostate cancer [37]. Our trans-omics prognostic nomogram is accurate, simple, and easy to apply. We have constructed this novel nomogram in KIRC, which has great value for the prediction of survival rate.

Recently, studies report a few prognostic models related to methylation in KIRC. The model reported by Xu et al. is based on five methylated CpG sites without trans-omics data [38]. In Guang Chen's study, the top 15% of the nodes in the enriched pathway networks are selected to screen the methylated genes related to prognosis. However, in the process of variable screening, some important information may be missed. Moreover, the model's performance has not been evaluated [39]. Evelönn et al. has built a model from local patients' CpG methylation data combined with CNV data and only used TCGA data for validation. This model has many complicated parameters, which may lead to limitation for clinical application [1]. The above models are completely different from the model established in our study. The model built by Hu et al. is similar to ours [40]. However, methylation-driven genes are not used for variable screening. Seven genes are incorporated in their model, whereas only four genes are incorporated in our model. Moreover, nonsilent mutations in *VHL* are incorporated in their model, which may lead to increased difficulty of practice. The factor of gender is not incorporated in their model. The AUC values of their model in the testing dataset are only 0.677, 0.66, and 0.71 (1-, 3-, and 5-year survival rates, respectively), whereas the AUC values of our model are 0.794, 0.752, and 0.731 (1-, 3-, and 5-year survival rates, respectively). Therefore, the model discovered in the present study is simple and efficient.

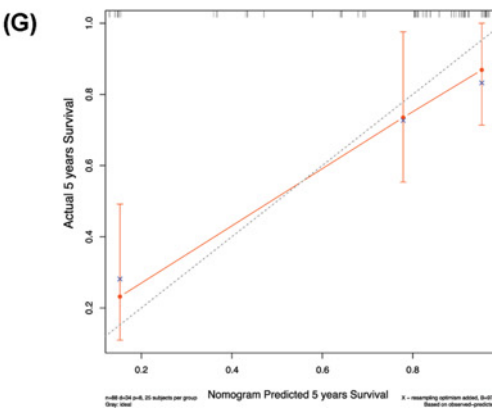
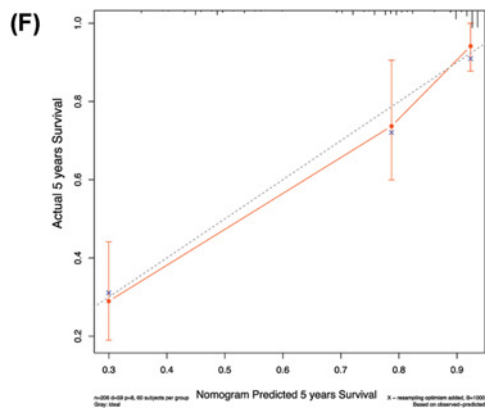
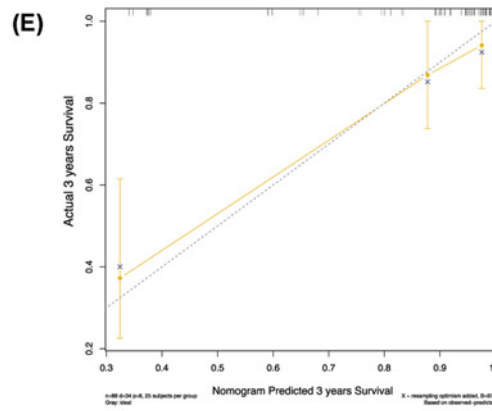
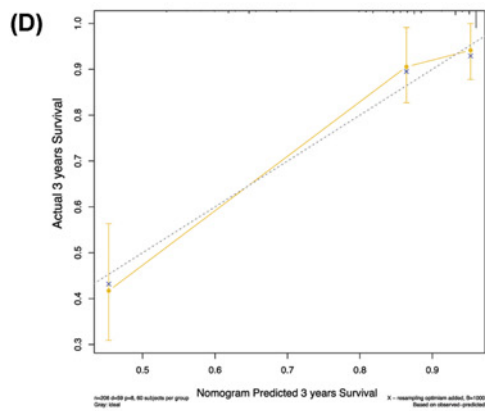
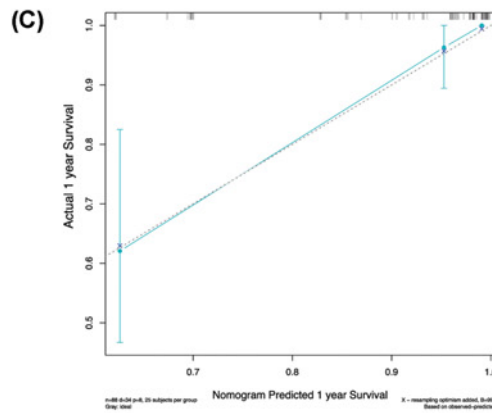
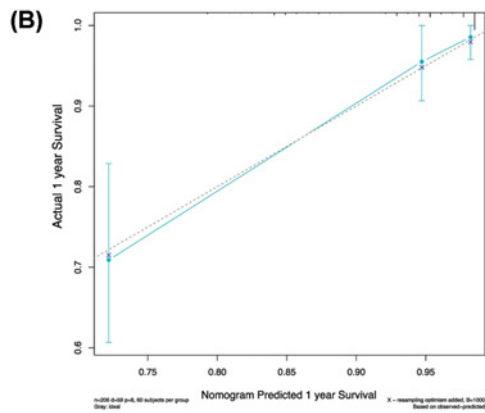
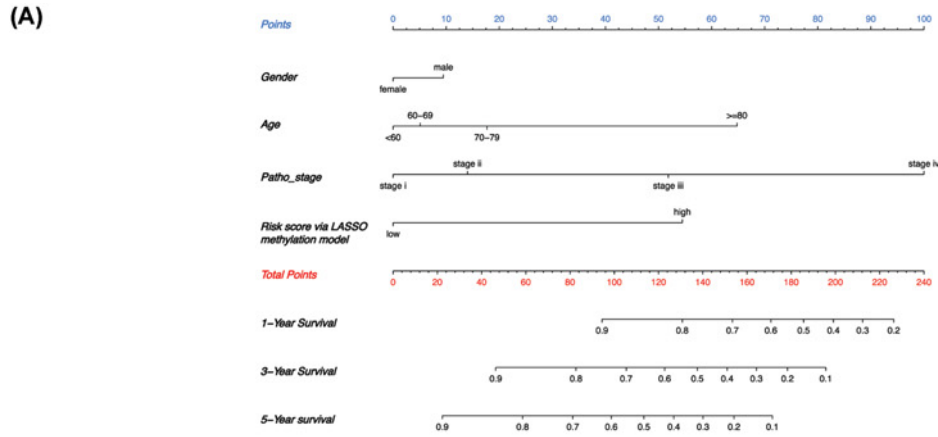


Figure 5. Combining the score of methylation-driven gene prognostic model and clinical information to construct nomogram for prognosis prediction

(A) Nomogram of overall survival (OS) in KIRC. The top points are the scale used for scoring; clinical indexes and methylation-driven gene prognosis model scores can correspond to the top scale scores. Add the scores of each predictor to obtain the total score. The total score on the total point scale is located, and then the coordinates corresponding on the probability scale at the bottom of the figure are used to obtain 1-, 3-, and 5-year probability of survival. (B–G) Calibration curves of the nomogram in the training and testing group of KIRC. (B) Prediction of 1-year OS in the training group of KIRC. (C) Prediction of 1-year OS in the testing group of KIRC. (D) Prediction of 3-year OS in the training group of KIRC. (E) Prediction of 3-year OS in the testing group of KIRC. (F) Prediction of 5-year OS in the training group of KIRC. (G) Prediction of 5-year OS in the testing group of KIRC.

Conclusion

Several limitations were reported in the present study. First, the external data were not available for further verification. Second, the analysis results should be supported and verified by clinical and experimental tests. In general, the present study was the first to reveal the overall perspective of methylation-driven genes in KIRC. We used the machine learning method to establish a multivariate methylation prognostic prediction model and combined with clinical information to build the trans-omics prognostic nomogram. The model discovered in the present study is novel, simple, and efficient. These results can help in the accurate evaluation of the prognosis of KIRC patients and provide new clues and data resources for the further study of the pathogenesis and the development of the disease.

Data Availability

All data generated or analyzed during this study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) [grant number 61502363].

Author Contribution

X.M designed the experiments; W.T and X.M analyzed the data; C.Y and W.T wrote the manuscript. All authors read and approved the final manuscript.

Abbreviations

CpG, Cytosine–phosphate–guanine; KIRC, Kidney renal clear cell carcinoma; K-M, Kaplan–Meier; OS, Overall survival; TCGA, The Cancer Genome Atlas.

References

- Evelonn, E.A., Landfors, M., Haider, Z., Kohn, L., Ljungberg, B., Roos, G. et al. (2019) DNA methylation associates with survival in non-metastatic clear cell renal cell carcinoma. *BMC Cancer* **19**, 65, <https://doi.org/10.1186/s12885-019-5291-3>
- Gong, J., Maia, M.C., Dizman, N., Govindarajan, A. and Pal, S.K. (2016) Metastasis in renal cell carcinoma: Biology and implications for therapy. *Asian J. Urol.* **3**, 286–292, <https://doi.org/10.1016/j.ajur.2016.08.006>
- Gu, T. and Zhao, X. (2019) Integrating multi-platform genomic datasets for kidney renal clear cell carcinoma subtyping using stacked denoising autoencoders. *Sci. Rep.* **9**, 16668, <https://doi.org/10.1038/s41598-019-53048-x>
- Cheng, G., Liu, D., Liang, H., Yang, H., Chen, K. and Zhang, X. (2019) A cluster of long non-coding RNAs exhibit diagnostic and prognostic values in renal cell carcinoma. *Aging* **11**, 9597–9615, <https://doi.org/10.18632/aging.102407>
- Ruik, Y., Imanaka, Y., Sato, F., Shimizu, K. and Tsujimoto, G. (2010) Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics* **11**, 137, <https://doi.org/10.1186/1471-2164-11-137>
- Li, C., Ke, J., Liu, J. and Su, J. (2019) DNA methylation data-based molecular subtype classification related to the prognosis of patients with cervical cancer. *J. Cell. Biochem.* **121**, 2713–2724, <https://doi.org/10.1002/jcb.29491>
- Tao, C., Luo, R., Song, J., Zhang, W. and Ran, L. (2019) A seven-DNA methylation signature as a novel prognostic biomarker in breast cancer. *J. Cell. Biochem.* **121**(3), 2385–2393 [10.1002/jcb.29461](https://doi.org/10.1002/jcb.29461)
- Jensen, S.O., Ogaard, N., Orntoft, M.W., Rasmussen, M.H., Bramsen, J.B., Kristensen, H. et al. (2019) Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer—a clinical biomarker discovery and validation study. *Clin. Epigenetics* **11**, 158, <https://doi.org/10.1186/s13148-019-0757-3>

- 9 Lasseigne, B.N. and Brooks, J.D. (2018) The Role of DNA Methylation in Renal Cell Carcinoma. *Mol. Diagn. Ther.* **22**, 431–442, <https://doi.org/10.1007/s40291-018-0337-9>
- 10 Shenoy, N., Vallumsetta, N., Zou, Y., Galeas, J.N., Shrivastava, M., Hu, C. et al. (2015) Role of DNA methylation in renal cell carcinoma. *J. Hematol. Oncol.* **8**, 88, <https://doi.org/10.1186/s13045-015-0180-y>
- 11 Guang Chen, Y.W., Wang, L. and Xu, W. (2017) Identifying prognostic biomarkers based on aberrant DNA methylation in kidney renal clear cell carcinoma. *Oncotarget* **8**, 5268–5280, <https://doi.org/10.18632/oncotarget.14134>
- 12 Peters, I., Dubrowskaja, N., Abbas, M., Seidel, C., Kogosov, M., Scherer, R. et al. (2014) DNA methylation biomarkers predict progression-free and overall survival of metastatic renal cell cancer (mRCC) treated with antiangiogenic therapies. *PLoS ONE* **9**, e91440, <https://doi.org/10.1371/journal.pone.0091440>
- 13 Zhou, L., Liu, S., Li, X., Yin, M., Li, S. and Long, H. (2019) Diagnostic and prognostic value of CEP55 in clear cell renal cell carcinoma as determined by bioinformatics analysis. *Mol. Med. Rep.* **19**, 3485–3496
- 14 Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O. et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**, pl1, <https://doi.org/10.1126/scisignal.2004088>
- 15 Zhang, B., Wu, Q., Wang, Z., Xu, R., Hu, X., Sun, Y. et al. (2019) The promising novel biomarkers and candidate small molecule drugs in kidney renal clear cell carcinoma: Evidence from bioinformatics analysis of high-throughput data. *Mol. Genet. Genomic. Med.* **7**, e607, <https://doi.org/10.1002/mgg3.607>
- 16 Wong, K.Y. and Chim, C.S. (2015) DNA methylation of tumor suppressor protein-coding and non-coding genes in multiple myeloma. *Epigenomics* **7**, 985–1001, <https://doi.org/10.2217/epi.15.57>
- 17 Ueno, S., Lu, J., He, J., Li, A., Zhang, X., Ritz, J. et al. (2014) Aberrant expression of SALL4 in acute B cell lymphoblastic leukemia: mechanism, function, and implication for a potential novel therapeutic target. *Exp. Hematol.* **42**, 307e8–316e8, <https://doi.org/10.1016/j.exphem.2014.01.005>
- 18 Toyota, M., Suzuki, H., Sasaki, Y., Maruyama, R., Imai, K., Shinomura, Y. et al. (2008) Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. *Cancer Res.* **68**, 4123–4132, <https://doi.org/10.1158/0008-5472.CAN-08-0325>
- 19 Tanaka, N., Toyooka, S., Soh, J., Kubo, T., Yamamoto, H., Maki, Y. et al. (2012) Frequent methylation and oncogenic role of microRNA-34b/c in small-cell lung cancer. *Lung Cancer* **76**, 32–38, <https://doi.org/10.1016/j.lungcan.2011.10.002>
- 20 Pronina, I.V., Loginov, V.I., Burdenny, A.M., Fridman, M.V., Senchenko, V.N., Kazubskaya, T.P. et al. (2017) DNA methylation contributes to deregulation of 12 cancer-associated microRNAs and breast cancer progression. *Gene* **604**, 1–8, <https://doi.org/10.1016/j.gene.2016.12.018>
- 21 Lu, T., Chen, D., Wang, Y., Sun, X., Li, S., Miao, S. et al. (2019) Identification of DNA methylation-driven genes in esophageal squamous cell carcinoma: a study based on The Cancer Genome Atlas. *Cancer Cell Int.* **19**, 52, <https://doi.org/10.1186/s12935-019-0770-9>
- 22 Wang, J., Zhang, Q., Zhu, Q., Liu, C., Nan, X., Wang, F. et al. (2020) Identification of methylation-driven genes related to prognosis in clear-cell renal cell carcinoma. *J. Cell. Physiol.* **235**, 1296–1308, <https://doi.org/10.1002/jcp.29046>
- 23 Xia, Y., Zhang, Q., Zhen, Q., Zhao, Y., Hao, Y., Zhang, Y. et al. (2017) Negative regulation of tumor-infiltrating NK cell in clear cell renal cell carcinoma patients through the exosomal pathway. *Oncotarget* **8**, 37783–37795, <https://doi.org/10.18632/oncotarget.16354>
- 24 Zhang, S., Zhang, E., Long, J., Hu, Z., Peng, J., Liu, L. et al. (2019) Immune infiltration in renal cell carcinoma. *Cancer Sci.* **110**, 1564–1572, <https://doi.org/10.1111/cas.13996>
- 25 Chen, L., Zeng, T., Pan, X., Zhang, Y.H., Huang, T. and Cai, Y.D. (2019) Identifying Methylation Pattern and Genes Associated with Breast Cancer Subtypes. *Int. J. Mol. Sci.* **20**, 4269, <https://doi.org/10.3390/ijms20174269>
- 26 Raita, Y., Goto, T., Faridi, M.K., Brown, D.F.M., Camargo, Jr, C.A. and Hasegawa, K. (2019) Emergency department triage prediction of clinical outcomes using machine learning models. *Crit. Care* **23**, 64, <https://doi.org/10.1186/s13054-019-2351-7>
- 27 Shu, J., Wen, D., Xi, Y., Xia, Y., Cai, Z., Xu, W. et al. (2019) Clear cell renal cell carcinoma: Machine learning-based computed tomography radiomics analysis for the prediction of WHO/ISUP grade. *Eur. J. Radiol.* **121**, 108738, <https://doi.org/10.1016/j.ejrad.2019.108738>
- 28 Zou, Y., Qiu, G., Jiang, L., Cai, Z., Sun, W., Hu, H. et al. (2017) Overexpression of ubiquitin specific proteases 44 promotes the malignancy of glioma by stabilizing tumor-promoter securin. *Oncotarget* **8**, 58231–58246, <https://doi.org/10.18632/oncotarget.16447>
- 29 Foy, J.P., Pickering, C.R., Papadimitrakopoulou, V.A., Jelinek, J., Lin, S.H., William, Jr, W.N. et al. (2015) New DNA methylation markers and global DNA hypomethylation are associated with oral cancer development. *Cancer Prev. Res. (Phila.)* **8**, 1027–1035, <https://doi.org/10.1158/1940-6207.CAPR-14-0179>
- 30 Mitchell, S.M., Ross, J.P., Drew, H.R., Ho, T., Brown, G.S., Saunders, N.F. et al. (2014) A panel of genes methylated with high frequency in colorectal cancer. *BMC Cancer* **14**, 54, <https://doi.org/10.1186/1471-2407-14-54>
- 31 Poage, G.M., Houseman, E.A., Christensen, B.C., Butler, R.A., Avissar-Whiting, M., McClean, M.D. et al. (2011) Global hypomethylation identifies Loci targeted for hypermethylation in head and neck cancer. *Clin. Cancer Res.* **17**, 3579–3589, <https://doi.org/10.1158/1078-0432.CCR-11-0044>
- 32 Lu, S., Qian, J., Guo, M., Gu, C. and Yang, Y. (2019) Insights into a Crucial Role of TRIP13 in Human Cancer. *Comput. Struct. Biotechnol. J.* **17**, 854–861, <https://doi.org/10.1016/j.csbj.2019.06.005>
- 33 Xu, R.H., Wei, W., Krawczyk, M., Wang, W., Luo, H., Flagg, K. et al. (2017) Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* **16**, 1155–1161, <https://doi.org/10.1038/nmat4997>
- 34 Guo, W., Zhu, L., Yu, M., Zhu, R., Chen, Q. and Wang, Q. (2018) A five-DNA methylation signature act as a novel prognostic biomarker in patients with ovarian serous cystadenocarcinoma. *Clin. Epigenetics* **10**, 142, <https://doi.org/10.1186/s13148-018-0574-0>
- 35 Wang, X. (2018) Clinical trans-omics: an integration of clinical phenomes with molecular multiomics. *Cell Biol. Toxicol.* **34**, 163–166, <https://doi.org/10.1007/s10565-018-9431-3>
- 36 Yang, D., Zhang, X., Powell, C.A., Ni, J., Wang, B., Zhang, J. et al. (2018) Probability of cancer in high-risk patients predicted by the protein-based lung cancer biomarker panel in China: LCBP study. *Cancer* **124**, 262–270, <https://doi.org/10.1002/cncr.31020>

- 37 Hou, Q., Bing, Z.T., Hu, C., Li, M.Y., Yang, K.H., Mo, Z. et al. (2018) RankProd Combined with Genetic Algorithm Optimized Artificial Neural Network Establishes a Diagnostic and Prognostic Prediction Model that Revealed C1QTNF3 as a Biomarker for Prostate Cancer. *EBioMedicine* **32**, 234–244, <https://doi.org/10.1016/j.ebiom.2018.05.010>
- 38 Xu, L., He, J., Cai, Q., Li, M., Pu, X. and Guo, Y. (2020) An effective seven-CpG-based signature to predict survival in renal clear cell carcinoma by integrating DNA methylation and gene expression. *Life Sci.* **243**, 117289, <https://doi.org/10.1016/j.lfs.2020.117289>
- 39 Chen, G., Wang, Y., Wang, L. and Xu, W. (2017) Identifying prognostic biomarkers based on aberrant DNA methylation in kidney renal clear cell carcinoma. *Oncotarget* **8**, 5268–5280, <https://doi.org/10.18632/oncotarget.14134>
- 40 Hu, F., Zeng, W. and Liu, X. (2019) A Gene Signature of Survival Prediction for Kidney Renal Cell Carcinoma by Multi-Omic Data Analysis. *Int. J. Mol. Sci.* **20**, <https://doi.org/10.3390/ijms20225720>