

Research Article

# A novel DNA methylation-based model that effectively predicts prognosis in hepatocellular carcinoma

 Xiang-Yong Hao<sup>1,2,\*</sup>, An-Qiang Li<sup>1,2</sup>, Hao Shi<sup>1,2</sup>, Tian-Kang Guo<sup>1,2</sup>, Yan-Fei Shen<sup>3</sup>, Yuan Deng<sup>1,2</sup>, Li-Tian Wang<sup>1,2</sup>, Tao Wang<sup>1,2,\*</sup> and Hui Cai<sup>1,2,\*</sup>

<sup>1</sup>Department of General Surgery, Gansu Provincial Hospital, Lanzhou, China; <sup>2</sup>Key Laboratory of Molecular Diagnosis and Precision Treatment of Surgical Tumors in Gansu Province, Lanzhou, China; <sup>3</sup>Department of Medical Management, Gansu Provincial Hospital, Lanzhou, China

**Correspondence:** Xiang-Yong Hao (haoxy2007@lzu.edu.cn; haoxy-2008@163.com) or Tao Wang (wtdoctor831030@sina.com) or Hui Cai (goodcaihui@163.com)



**Purpose:** To build a novel predictive model for hepatocellular carcinoma (HCC) patients based on DNA methylation data.

**Methods:** Four independent DNA methylation datasets for HCC were used to screen for common differentially methylated genes (CDMGs). Gene Ontology (GO) enrichment, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were used to explore the biological roles of CDMGs in HCC. Univariate Cox analysis and least absolute shrinkage and selection operator (LASSO) Cox analysis were performed to identify survival-related CDMGs (SR-CDMGs) and to build a predictive model. The importance of this model was assessed using Cox regression analysis, propensity score-matched (PSM) analysis and stratification analysis. A validation group from the Cancer Genome Atlas (TCGA) was constructed to further validate the model.

**Results:** Four SR-CDMGs were identified and used to build the predictive model. The risk score of this model was calculated as follows: risk score = (0.01489826 × methylation level of *WDR69*) + (0.15868618 × methylation level of *HOXB4*) + (0.16674959 × methylation level of *CDKL2*) + (0.16689301 × methylation level of *HOXA10*). Kaplan–Meier analysis demonstrated that patients in the low-risk group had a significantly longer overall survival (OS; log-rank *P*-value = 0.00071). The Cox model multivariate analysis and PSM analysis identified the risk score as an independent prognostic factor (*P* < 0.05). Stratified analysis results further confirmed this model performed well. By analyzing the validation group, the results of receiver operating characteristic (ROC) curve analysis and survival analysis further validated this model.

**Conclusion:** Our DNA methylation-based prognosis predictive model is effective and reliable in predicting prognosis for patients with HCC.

\*These authors contributed equally to this work, and should be considered co-corresponding authors.

Received: 17 November 2020  
Revised: 22 February 2021  
Accepted: 25 February 2021

Accepted Manuscript online:  
26 February 2021  
Version of Record published:  
10 March 2021

## Introduction

Hepatocellular carcinoma (HCC) is one of the most prevalent malignant tumors worldwide, with its incidence increasing over the last 10 years [1]. Current TNM staging systems and Barcelona Clinic Liver Cancer staging systems were constructed based on the anatomical features of the tumor and clinic-pathological information [2,3]. These approaches remain the most commonly used staging systems for patients with HCC. Therefore, the prognoses of patients with HCC vary significantly even among those with similar clinical staging. Therefore, more prognostic and survival-related factors of patients with HCC must be identified and applied to improve prognostic accuracy.

DNA methylation is an important epigenetic modification that is involved in the regulation of gene expression and genome stability [4]. There is mounting evidence that aberrant DNA methylation in promoter regions contributes to the carcinogenesis, progression and prognosis of many tumors [5–13]. For example, promoter methylation of *P16* (*CDKN2A* or *Ink4a*) may directly inactivate gene transcription and drive cancer metastasis [11]. Besides, DNA methylation in the promoter regions of *IRAK3* is clearly associated with the tumor stage and clinical outcomes of HCC patients [10]. Moreover, the development of next-generation sequencing and microarrays further improves our understanding of DNA methylation associated with tumors.

Here, we aimed to identify important prognosis-related methylation signatures. Furthermore, we constructed a composite predictive model for HCC patients using CpG methylation data.

## Materials and methods

### Retrieval of DNA methylation data

Four independent DNA methylation datasets for HCC examined by HumanMethylation450 Bead Chip (Illumina) were downloaded from public databases. Three of the datasets (GSE54503, GSE77269 and GSE56588 datasets) were obtained from the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>), while the fourth dataset (from the Cancer Genome Atlas [TCGA], <https://cancergenome.nih.gov/>) contained corresponding clinical prognostic information. Each dataset contained information on more than ten cancer tissues and ten adjacent non-cancerous tissues. For the TCGA dataset, non-HCCs and patients without DNA methylation information were removed; 365 HCC samples and 50 non-tumor samples were retained. The methylation levels of each probe are expressed as a  $\beta$ -value.

### Identification of common differentially methylated genes

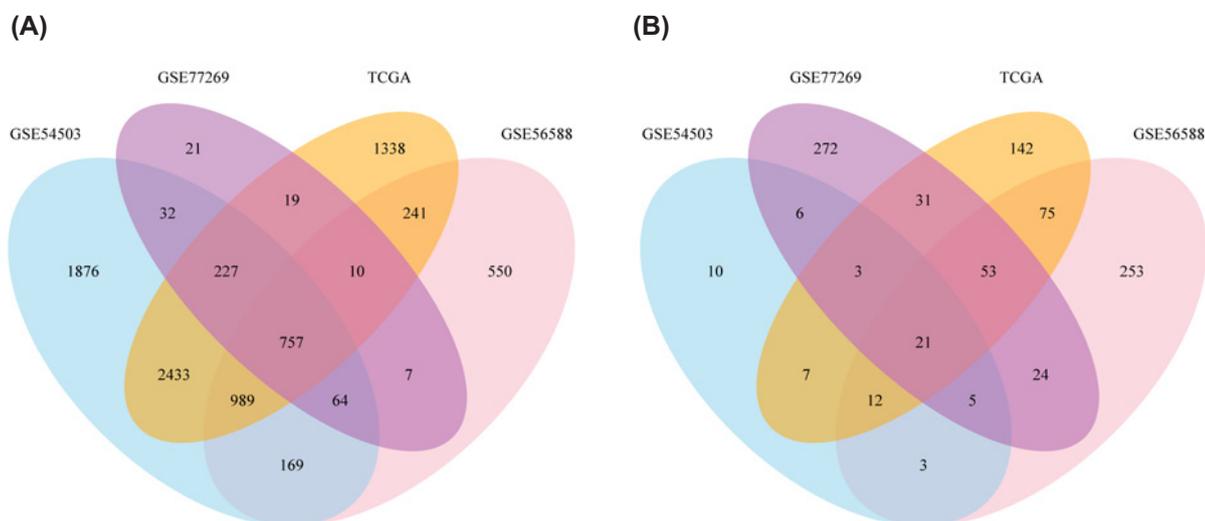
In the four datasets, we used the Chip Analysis Methylation Pipeline Bio-conductor package to identify differentially methylated genes (DMGs) between HCC and adjacent non-cancerous tissues [14]. Only DNA methylation at CpG islands within the promoter region of the gene was screened for analysis. The promoter region indicates loci residing within the Tss1500 (200–1500 bps upstream of the transcription start site) and Tss200 (200 bps upstream of the transcription start site) of known genes [15]. DNA methylation sites on sex chromosomes, probes containing single nucleotide polymorphisms as reported by Zhou et al. [16] and probes aligning to multiple locations as identified by Nordlund et al. [17] were removed. In addition, probes with  $\beta$  values in less than 95% of samples were removed. Then, NAs were replaced by the k-nearest neighbor imputation procedure [18,19]. Only probes with adjusted  $P < 0.05$  and a  $|\log(\text{fold change})| \geq 1$  were regarded as differentially methylated probes (DMPs). The genes mapped by DMPs were confirmed as DMGs. DMGs that were common across the four datasets (common differentially methylated genes, CDMGs) were selected for further analysis.

### Bioinformatics analysis for CDMGs

CDMGs, including common hyper-methylated genes and common hypo-methylated genes obtained from the four independent datasets, were subjected to Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis via DAVID 6.8 (<https://david.ncifcrf.gov/>) to investigate their activities in HCC [20,21]. The STRING database (<https://string-db.org/>) and Cytoscape 3.6.0 were used to build a protein–protein interaction (PPI) network of CDMGs. The STRING database is a precomputed resource used to investigate protein regulatory networks in different diseases [22–24]. The PPI score (confidence) was set as 0.4.

### Identification of survival-related CDMGs and construction of the predictive model

After removing patients with detectable residual tumors (R1 and R2) and patients who received adjuvant treatment (including radiation therapy, chemotherapy and embolization treatment), 296 HCC samples in the TCGA dataset were retained to screen for survival-related methylated markers and to construct a predictive model. Univariate Cox regression analysis was used to identify risk markers among the CDMGs that were clearly associated with the overall survival (OS) of HCC. The results were considered significant when  $P < 0.05$ . The risk markers from univariate Cox analysis were further analyzed using the least absolute shrinkage and selection operator (LASSO) Cox regression model for survival-related CDMG (SR-CDMG) selection and classifier construction. LASSO Cox analysis is a regularized Cox regression approach that is suitable for analyzing high-dimensional data of small sample size [25]. In this way, it was possible to select a subset of variables that best predict OS while reducing the overfitting of data [26]. Ten-fold cross-validation with minimum criteria was applied to identify the optimal parameter  $\lambda$  in the LASSO



**Figure 1. CDMGs in four datasets**

(A) Venn diagram showing 757 common hyper-methylated genes in HCC samples. (B) Venn diagram showing 21 common hypo-methylated genes in HCC samples.

Cox model [27,28]. Using this model, the regression coefficients of most features were shrunk towards zero, allowing the most useful predictors (representing SR-CDMGs) to be identified. Then, the SR-CDMGs with non-zero coefficient were used to build a predictive model for HCC. A risk score for each patient was computed according to the summation of the methylation levels ( $\beta$  value) multiplied by the corresponding coefficients from the LASSO Cox analysis.

We divided patients into two groups according to the median values of SR-CDMGs and the risk score. Kaplan–Meier survival curves were used to analyze the correlation of SR-CDMGs and the risk score with OS. In addition, univariate and multivariate-adjusted Cox regression analysis, propensity score-matched (PSM) analysis and stratification analysis were performed to estimate the importance of the predictive model in the prognosis of HCC.

Moreover, in order to further validate the predictive model that we constructed, HCC samples with DNA methylation information and clinical prognostic information in the TCGA dataset were randomly selected to construct a validation group. Then the predictive model was verified by receiver operating characteristic (ROC) curve and survival analysis. Area under the ROC curve (AUC) values for 1-, 3-, and 5-year survival of the predictive model were also calculated.

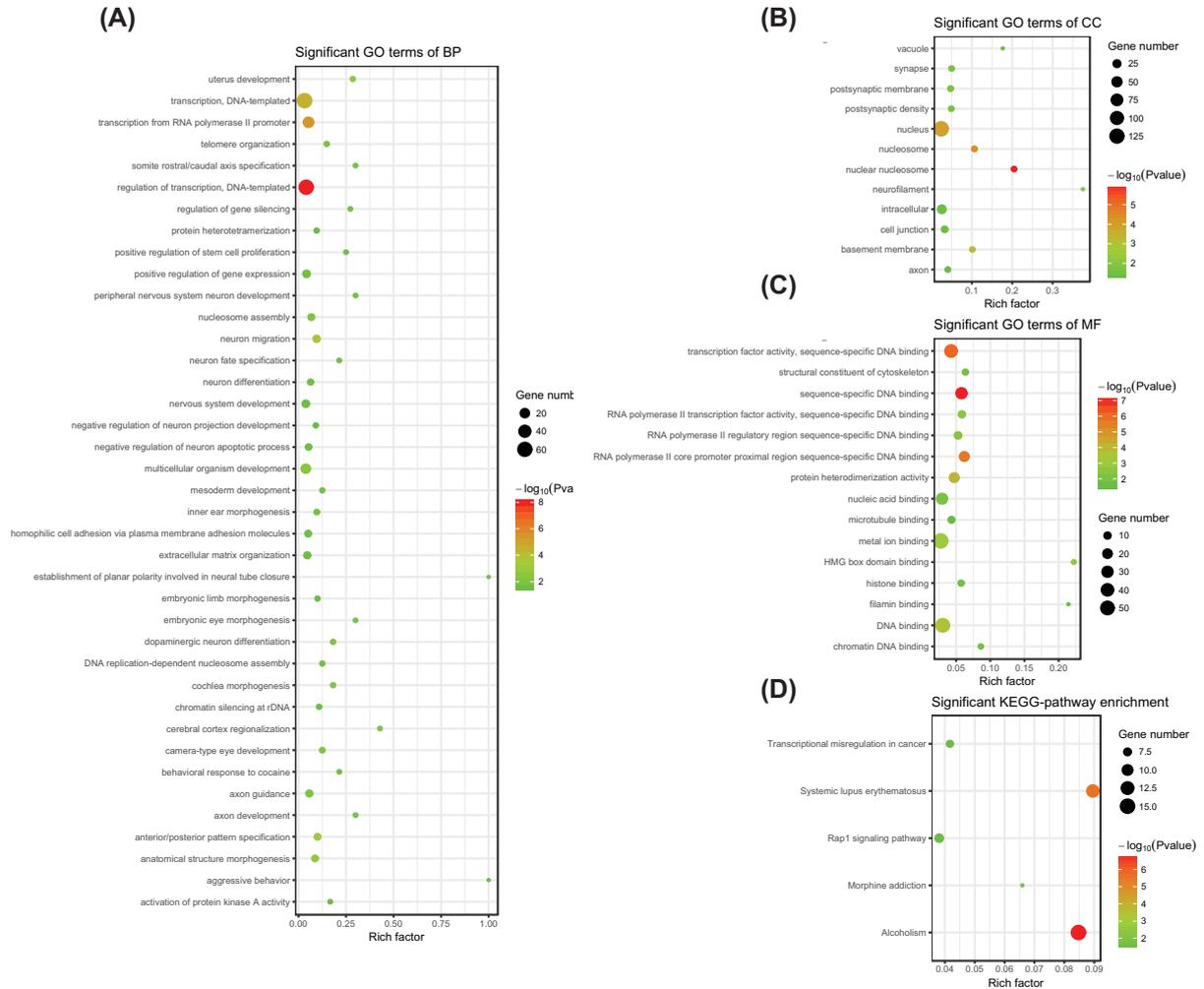
## Statistical analysis

R software 3.3.4 (<https://www.r-project.org/>) was used to perform the statistical analysis and graphic visualization. The results were considered statistically significant at a two-tailed  $P$  value  $<0.05$ . Differences in continuous variables between the two groups were analyzed using the non-parametric Mann–Whitney U test or Student's  $t$  test [29]. DNA methylation data from HumanMethylation450 Bead Chip were processed with the R package 'ChAMP' [14]. LASSO Cox analysis was performed with the R package 'glmnet' [30,31]. Cox regression analysis was carried out with the package 'survival' [32]. The R package 'MatchIt' was used for PSM analysis [33]. The ROC curve analyses were performed using the R package pROC [34].

## Results

### Identification of CDMGs and bioinformatics analysis

To improve the reliability of our results, four independent datasets were used to recognize differentially methylated markers of tumor vs adjacent non-cancerous tissues. A total of 778 CDMGs (Additional file 1: Supplementary Table S1) from the four datasets were regarded as credible, including 757 hyper-methylated genes (Figure 1A) and 21 hypo-methylated genes (Figure 1B) in HCC.



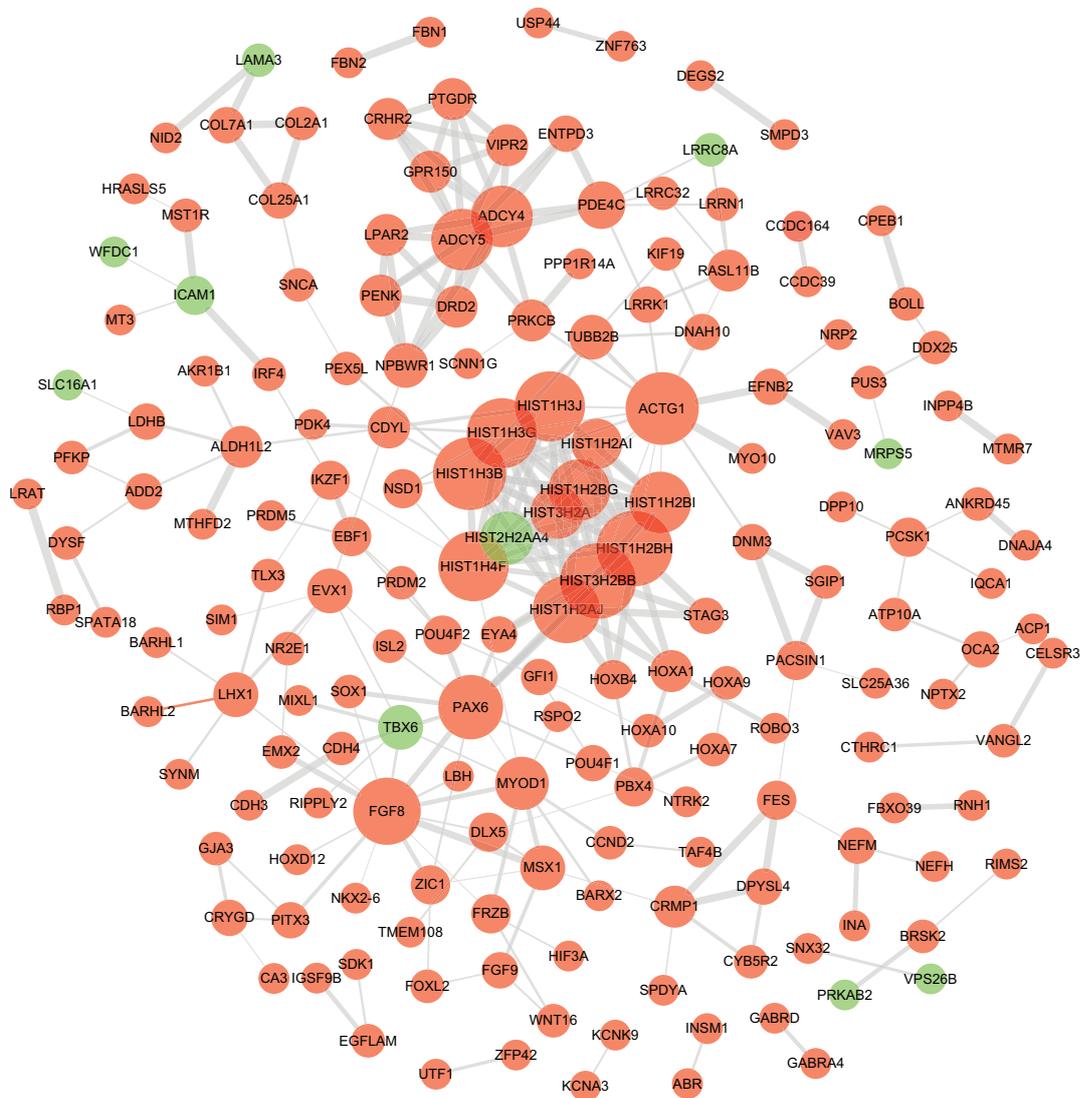
**Figure 2.** GO enrichment analysis and KEGG pathway enrichment analysis of CDMGs

(A) Significant GO terms for biological processes. (B) Significant GO terms for cellular components. (C) Significant GO terms for molecular functions. (D) Significant KEGG pathways. Terms and pathways with  $P < 0.05$  were here considered significant.

To explore the biological roles of the 778 CDMGs, GO enrichment analysis and KEGG pathway analysis were performed. Significant GO terms and pathways ( $P < 0.05$ ) are shown in Figure 2. The main terms of the biological processes include the regulation of tissue development, morphogenesis and transcription (Figure 2A). Terms related to the synapse and nucleosome were significantly enriched in the cellular component (Figure 2B). The main terms related to molecular functions were clearly related to DNA binding (Figure 2C). In addition, pathways involving transcriptional misregulation in cancer, Rap1 signaling, systemic lupus erythematosus, morphine addiction and alcoholism were enriched in KEGG analysis (Figure 2D). To evaluate the interactive relationships among all CDMGs and to elucidate the molecular mechanisms underlying HCC, a PPI network with 180 nodes and 299 edges was constructed (Figure 3).

## Identification of SR-CDMGs and construction of the predictive model

For the 778 candidate CDMGs, univariate Cox regression analysis was conducted to screen significant risk markers associated with OS in the TCGA dataset. Seventy-two CDMGs with  $P < 0.05$  were identified (Additional file 2: Supplementary Table S2). To identify SR-CDMGs among these 72 CDMGs, LASSO Cox analysis was performed. The parameter  $\log(\lambda) = -2.418524$  ( $\lambda = 0.08905296$ ), chosen by the ten-fold cross-validation method with minimum

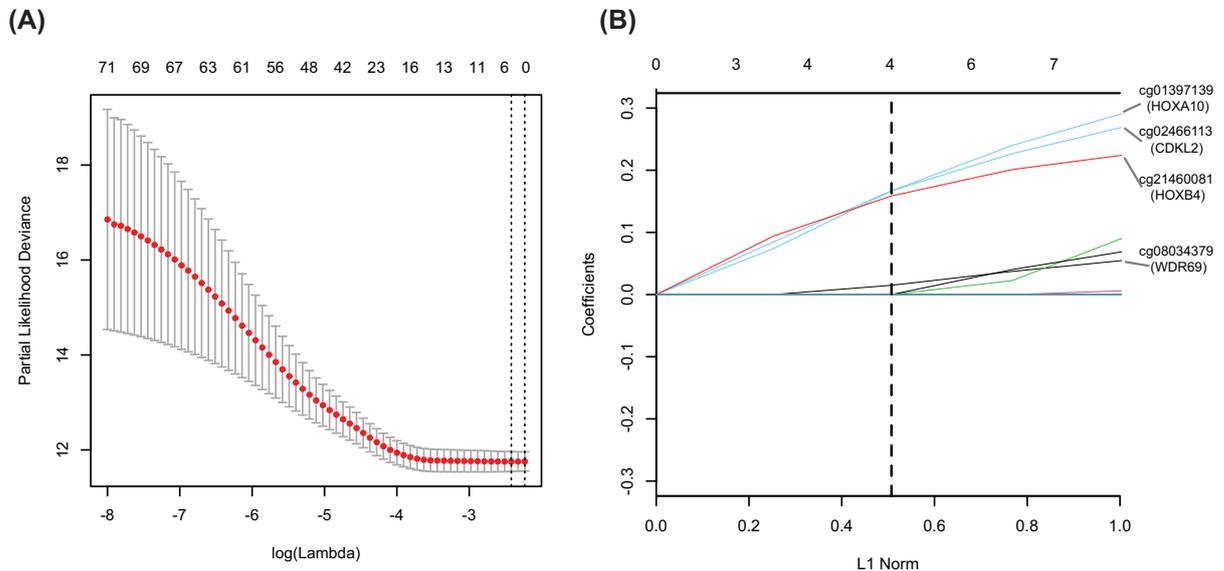


**Figure 3. PPI network of CDMGs**

After removing the unconnected nodes, 180 nodes and 299 edges were retained. Red nodes represent common hyper-methylated genes, and blue nodes represent common hypo-methylated genes. Node size is proportional to the node's degree. Line thickness indicates the strength of data support. The PPI score was set at 0.4.

criteria, produced the best value (Figure 4A). Four genes with non-zero coefficients were identified as SR-CDMGs (*HOXA10*, *CDKL2*, *HOXB4* and *WDR69*) (Figure 4B); all four genes had higher methylation levels in HCC samples (Additional file 3: Supplementary Figure S1A–D). Based on the median values of the methylation levels ( $\beta$  values) of these four SR-CDMGs, patients in the TCGA dataset were divided into hyper-methylated and hypo-methylated groups. All SR-CDMGs were closely related to poor OS (log-rank  $P$ -value  $< 0.05$ ) (Figure 5A–D).

To build an individualized predictive model for HCC based on the four SR-CDMGs, a risk score for each patient in the TCGA dataset was computed by summing the methylation levels multiplied by the corresponding coefficients from the LASSO Cox analysis: risk score =  $(0.01489826 \times \text{methylation level of } WDR69) + (0.15868618 \times \text{methylation level of } HOXB4) + (0.16674959 \times \text{methylation level of } CDKL2) + (0.16689301 \times \text{methylation level of } HOXA10)$



**Figure 4. Four SR-CDMGs selected through LASSO Cox analysis**

(A) Ten-fold cross-validation for selection of the parameter  $\lambda$ . The solid vertical lines represent partial likelihood deviance  $\pm$  standard error (SE). The two dotted vertical lines were drawn at the optimal values using minimum criteria (right) and 1-SE criteria (left). The parameter  $\lambda = 0.08905296$  ( $\log(\lambda) = -2.418524$ ) was chosen using minimum criteria. (B) LASSO coefficients of the four SR-CDMGs. The dotted vertical line was drawn at the  $\lambda$  value chosen using minimum criteria. The upper axis represents the number of non-zero coefficients at each  $\lambda$  (the degrees of freedom for the LASSO model). The L1 norm represents the summation of absolute non-zero coefficients at each  $\lambda$ . The vertical axis represents the values of non-zero coefficients at each  $\lambda$ . The LASSO coefficients of the four SR-CDMGs *HOXA10*, *CDKL2*, *HOXB4* and *WDR69* were 0.16689301, 0.16674959, 0.15868618 and 0.01489826, respectively.

*HOXA10*). The distributions of the risk score, survival status and the methylation level of the four SR-CDMGs are presented in Figure 6A. Patients with higher risk scores generally had lower survival rates than with patients with low risk scores. Patients were divided into low-risk and high-risk groups according to the median value of the risk score. Kaplan–Meier analysis demonstrated that patients in the low-risk group had a significantly longer OS than those in the high-risk group (log-rank  $P$ -value = 0.00071) (Figure 6B).

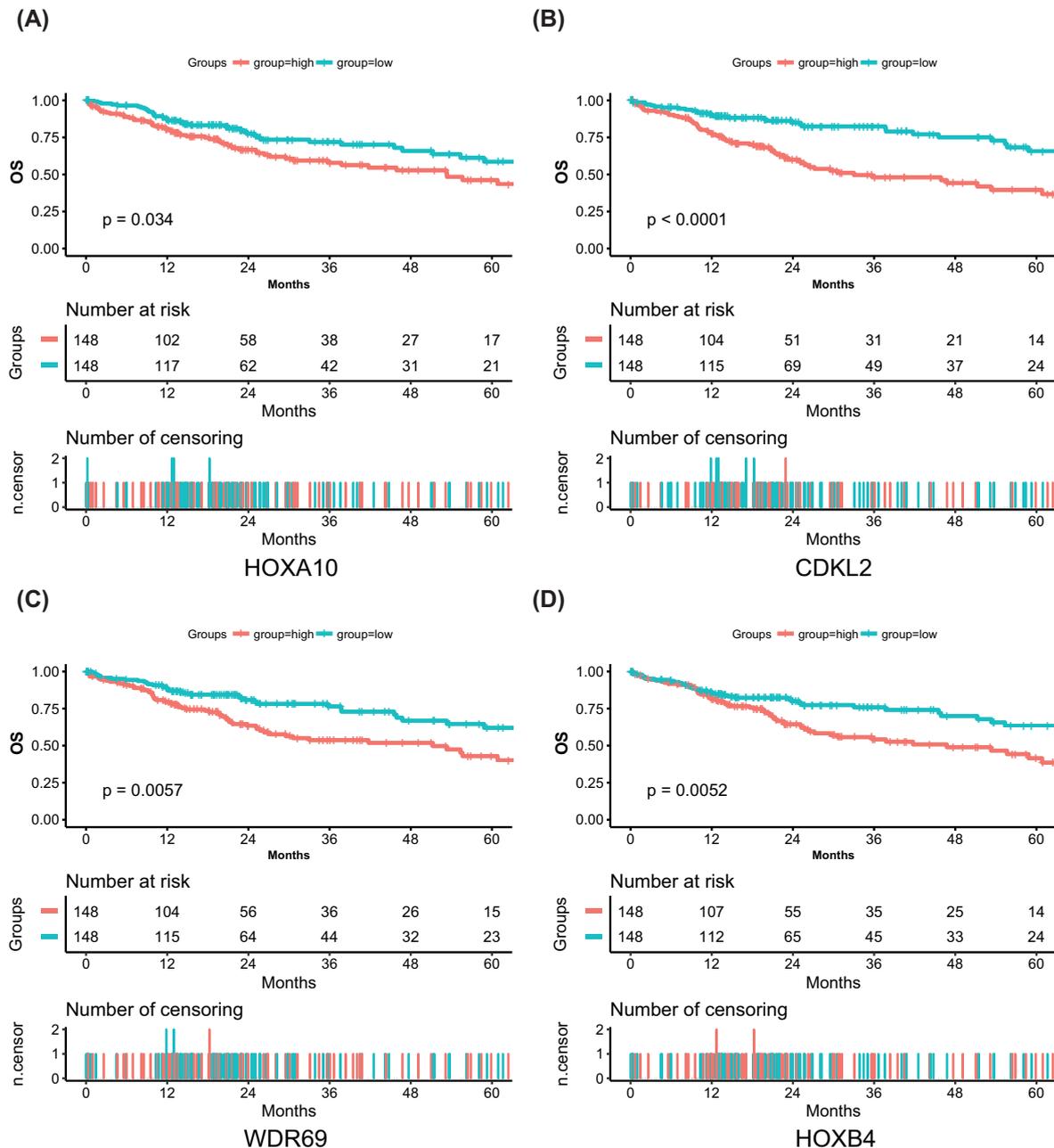
### Cox regression analysis and PSM analysis of the predictive model

Univariate Cox analysis and Cox model multivariate analysis were used to examine the prognostic factors for OS in HCC patients (Table 1). The unadjusted univariate Cox model showed that (i) the risk score ( $P < 0.001$ ), (ii) the American Joint Committee on Cancer (AJCC) stage ( $P = 0.008$ , II, III and IV vs I) and (iii)  $\alpha$ -fetoprotein (AFP) levels ( $P = 0.015$ ) were statistically related to the prognosis of HCC. A multivariate Cox model adjusted for important clinical factors (such as sex, age and race) showed that the risk score remained a significant independent risk factor for OS ( $P < 0.05$ ).

To reduce the confounding effects of the outcomes of the statistical comparisons between high- and low-risk score groups, PSM was performed to balance all known variables. The baseline characteristics of the two groups are shown in Table 2. Univariate Cox and multivariate-adjusted Cox regression analyses after PSM support the result that the risk score was an independent risk factor associated with poor prognosis (Table 3, all  $P$ -values of the risk scores were under 0.05).

### Stratification analysis of the predictive model

Stratification analysis was performed using important clinical variables to further estimate the influence of the predictive model in the prognosis of HCC. The results showed that patients with high-risk scores had shorter OS in some HCC subgroups (Figure 7, Additional file 4: Supplementary Figure S2), including male patients (2.083 [95% CI, 1.227–3.537],  $P = 0.007$ ), patients  $\leq 60$  years (2.750 [95% CI, 1.432–5.283],  $P = 0.002$ ), patients with risk factors (2.391 [95% CI, 1.417–4.034],  $P = 0.001$ ), AJCC stage I patients (3.336 [95% CI, 1.631–6.822],  $P = 0.001$ ), patients



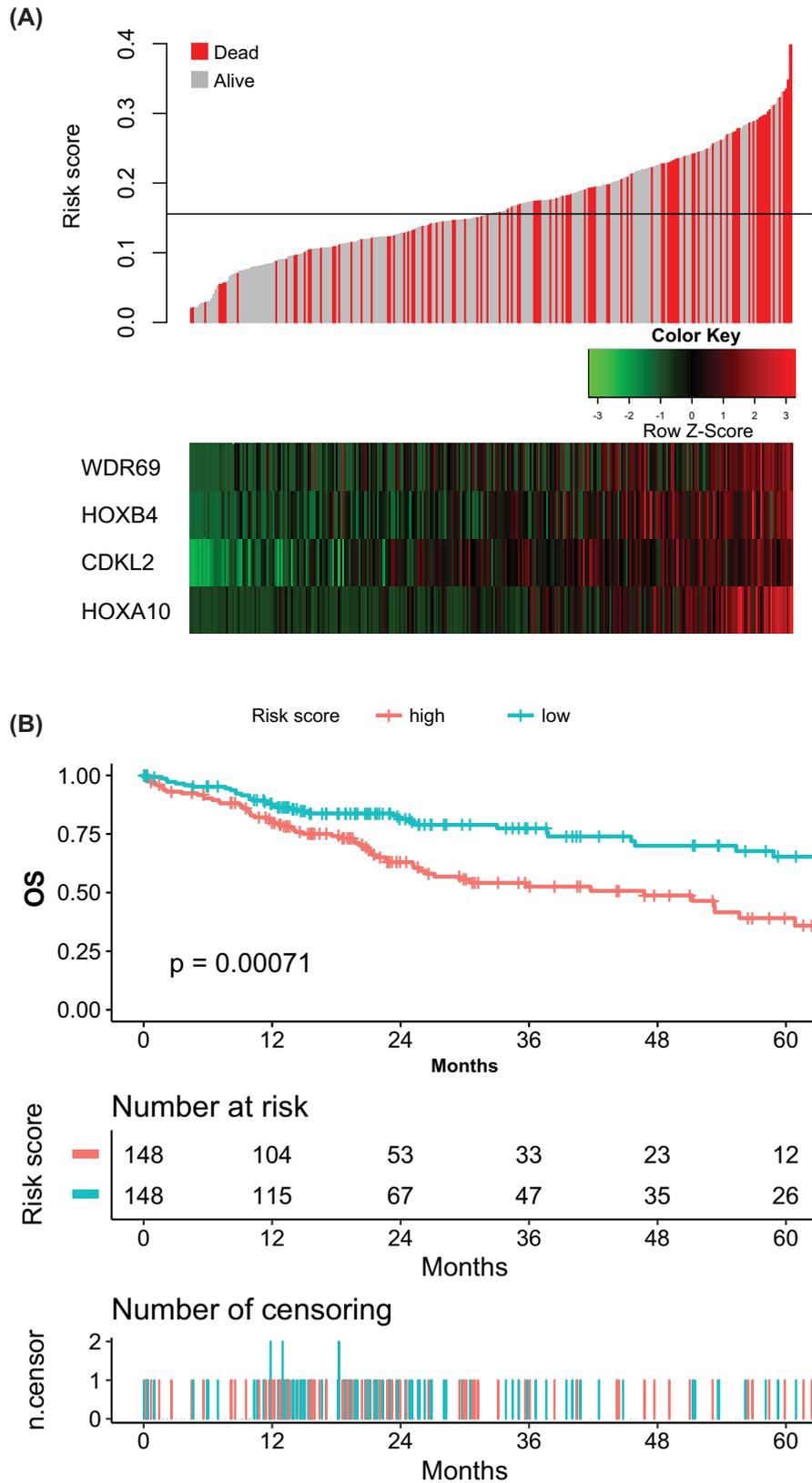
**Figure 5. Kaplan–Meier survival analysis for the four SR-CDMGs**

(A) Survival curve of *HOXA10*. (B) Survival curve of *CDKL2*. (C) Survival curve of *WDR69*. (D) Survival curve of *HOXB4*. Two-hundred and ninety-six patients from the TCGA were divided into a hyper-methylated and hypo-methylated groups according to the median value of the four SR-CDMGs. *P*-values were calculated with the log-rank test.

without vascular invasion (3.495 [95% CI, 1.895–6.445],  $P < 0.001$ ) and patients with low AFP levels (4.262 [95% CI, 1.692–10.740],  $P = 0.002$ ). The race and histological grade of patients did not influence the performance of the model.

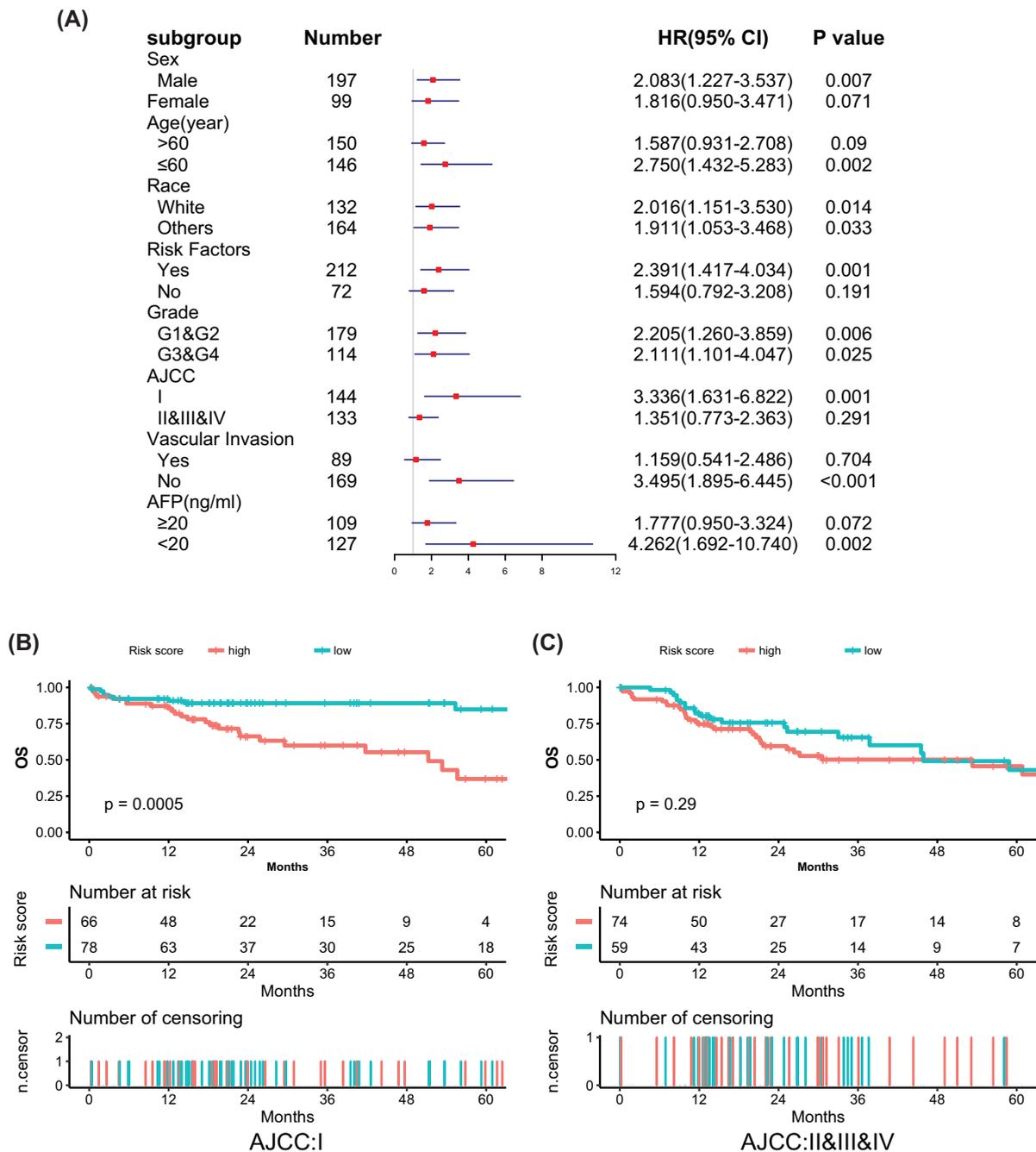
## Validation of the predictive model

A validation group was constructed by randomly selecting 148 HCC samples in the TCGA dataset, and ROC curve and survival analyses were used to further validate the prognosis predictive model. The results confirm that our predictive model is effective in the prognosis of HCC (Additional file 5: Supplementary Figure S3).



**Figure 6. Survival analysis of the predictive model**

(A) Distribution of the risk score of the predictive model, the survival status of patients and the methylation level of the four SR-CD-MGs. (B) Kaplan-Meier survival curve of the predictive model. Patients were divided into low-risk and high-risk groups according to the median value of the risk score. *P*-values were calculated with the log-rank test.



**Figure 7. Stratification analysis of the predictive model**

**(A)** Univariate COX regression analysis of the predictive model in different subgroups stratified by important clinical variables. **(B,C)** Kaplan–Meier survival analysis of the predictive model in subsets of different AJCC stage patients with HCC (log-rank test). Other: Asian and black patients. Risk factors: hepatitis B, hepatitis C, hemochromatosis, cirrhosis, alcohol consumption, non-alcoholic fatty liver disease or  $\alpha$ -1 antitrypsin deficiency.

**Table 1** Cox regression analysis of risk factors associated with OS in TCGA

Variables	Unadjusted		Adjusted 1		Adjusted 2	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Risk score						
Low						
High	1.997 (1.328–3.003)	<b>&lt;0.001</b>	2.400 (1.312–4.390)	<b>0.005</b>	2.54 (1.485–4.347)	<b>&lt;0.001</b>
Sex						
Female						
Male	0.803 (0.537–1.199)	0.283	1.072 (0.535–2.148)	0.845		
Age (years)						
≤60						
>60	1.200 (0.810–1.780)	0.363	1.359 (0.717–2.577)	0.348		
Race						
Other <sup>‡</sup>						
White	1.363 (0.903–2.057)	0.141	1.876 (0.925–3.801)	0.081		
Risk factors <sup>†</sup>						
No						
Yes	0.666 (0.435–1.019)	0.061	1.063 (0.523–2.159)	0.866		
Grade						
G1 and G2						
G3 and G4	1.140 (0.762–1.704)	0.525	1.298 (0.697–2.416)	0.411		
AJCC						
I						
II and III and IV	1.775 (1.165–2.706)	<b>0.008</b>	1.088 (0.555–2.133)	0.806	1.261 (0.772–2.061)	0.354
Vascular invasion						
No						
Yes	1.405 (0.884–2.233)	0.150	1.168 (0.566–2.408)	0.674		
AFP (ng/ml)						
<20						
≥20	1.890 (1.163–3.073)	<b>0.010</b>	2.020 (1.075–3.798)	<b>0.029</b>	1.551 (0.937–2.567)	0.088
History of other malignancy						
No						
Yes	1.232 (0.620–2.449)	0.552	1.863 (0.670–5.185)	0.233		
Family history of cancer						
No						
Yes	1.114 (0.732–1.696)	0.615	0.897 (0.483–1.665)	0.730		

Adjusted 1: Adjusted covariates include all the indicators above.

Adjusted 2: Adjusted covariates include the prognostic factors from unadjusted univariate COX analysis.

<sup>\*</sup>Represents Asian and black patients.

<sup>†</sup>Represents patients with the history of hepatitis B, hepatitis C, hemochromatosis, cirrhosis, alcohol consumption, non-alcoholic fatty liver disease or  $\alpha$ -1 antitrypsin deficiency.

The values in bold denote that the difference between the two groups is statistically significant ( $P < 0.05$ ).

## Discussion

DNA methylation is a common mechanism of epigenetic regulation in eukaryotic organisms ranging from fungi to mammals [35]. Increased methylation of tumor-associated genes is an early event in many tumors [36,37]. HCC is one of the most common solid tumors; its incidence is increasing, and it has high mortality rates worldwide. Therefore, it is urgent to identify reliable prognostic markers and construct a robust risk model to predict HCC outcomes. In this study, we developed an effective DNA methylation-based prognosis predictive model.

To screen genes with different levels of methylation between HCC and non-tumor tissues, we first analyzed the DNA methylation data in multiple independent datasets to significantly enhance the reliability of our results. GO enrichment analysis and KEGG pathway enrichment analysis were performed to investigate the biological roles of CDMGs. As shown in Figure 2, CDMGs were clearly enriched in pathways involved in transcriptional misregulation in cancer and in Rap1 signaling. The important roles of these two pathways in the development and progression of

**Table 2** Baseline characteristics of the patients in the PSM cohort stratified by risk score (high vs low group)

Variable	PSM cohort		P
	Risk score: High (n=89)	Risk score: Low (n=89)	
Age (years), mean (SD)	60.31 (12.08)	58.42 (11.67)	0.288
Sex			0.263
Male	56 (62.9)	64 (71.9)	
Female	33 (37.1)	25 (28.1)	
Race			0.540
White	38 (42.7)	33 (37.1)	
Other	51 (57.3)	56 (62.9)	
History of other malignancy			1.000
Yes	7 (7.9)	7 (7.9)	
No	82 (92.1)	82 (92.1)	
Family history of cancer			0.108
Yes	34 (38.2)	23 (25.8)	
No	55 (61.8)	66 (74.2)	
Risk factors			0.112
Yes	63 (70.8)	73 (82.0)	
No	26 (29.2)	16 (18.0)	
Grade			0.880
G1 and G2	49 (55.1)	47 (52.8)	
G3 and G4	40 (44.9)	42 (47.2)	
AJCC			0.759
I	52 (58.4)	55 (61.8)	
II and III and IV	37 (41.6)	34 (38.2)	
Vascular invasion			0.339
Yes	33 (37.1)	26 (29.2)	
No	56 (62.9)	63 (70.8)	
AFP (ng/ml)			0.450
≥20	42 (47.2)	36 (40.4)	
<20	47 (52.8)	53 (59.6)	

HCC have been reported in many studies [38–41]. Therefore, bioinformatics analysis could enhance our understanding of the molecular and pathological mechanisms underlying HCC.

Furthermore, we identified four SR-CDMGs (*HOXA10*, *CDKL2*, *HOXB4* and *WDR69*) that were significantly associated with the survival of HCC patients. It has been reported that the homeobox (HOX) transcription factors play critical roles in many physiological and pathological processes [42,43]. Previous studies have reported that *HOXA10* contributes to tumorigenesis and progression in HCC [44–46]. In addition, it has been reported that the combination of *HOXA9* and *HOXA10* promoter CpG islands could predict the survival of breast cancer patients [47]. *HOXB4* has been reported to contribute to tumorigenesis in many cancers [48–51]. Moreover, *HOXB4* methylation could be a potential marker in human cervical cancer [52] and in extrahepatic cholangiocarcinoma [49]. Hyper-methylation of the *CDKL2* promoter has been suggested to play an important role in hepatocarcinogenesis and to be a potentially valuable biomarker for HCC [53,54].

*WDR69* is a member of the WD-repeat domain (WDR) family. Although its association with cancer has not been reported, an increasing number of members of the WDR family have been identified as involved in tumorigenesis and cancer progression, including (i) *WDR34*, which may be used as a prognosis predictor in breast cancer and may provide a novel target for the treatment of breast cancer [55], (ii) *WDR54*, which serves as an oncogene in colorectal cancer and may be a potential prognostic marker and therapeutic target [56], (iii) *WDR1*, which plays a predictive role in nodal metastasis for primary breast cancer [57], and (iv) *WDR34*, which inhibits oral squamous cell carcinoma progression [58]. In the present study, *WDR69* methylation was significantly negatively correlated with the survival of HCC patients. Furthermore, after analyzed the clinical and pathological features of HCC patients in the TCGA dataset, the results showed that the *WDR69* methylation level was higher in HCC tumor tissues with high histological grade, AJCC stage, and AFP levels (Additional file 6: Supplementary Figure S4). Therefore, the molecular functions of *WDR69* in HCC deserve to be further exploited. In summary, the potential molecular biological mechanism of the SR-CDMGs *HOXA10*, *CDKL2*, *HOXB4* and *WDR69* further supports the reliability of our results.

**Table 3** Cox regression analysis of risk factors associated with OS in TCGA after PSM analysis

Variables	Unadjusted		Adjusted 1		Adjusted 2	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Risk Score						
Low						
High	2.815 (1.554–5.097)	<b>&lt;0.001</b>	2.536 (1.355–4.749)	<b>0.004</b>	2.481 (1.364–4.513)	<b>0.003</b>
Sex						
Female						
Male	0.576 (0.329–1.01)	0.054	1.102 (0.546–2.222)	0.787		
Age (years)						
≤60						
>60	1.567 (0.895–2.744)	0.116	1.37 (0.712–2.634)	0.346		
Race						
Other*						
White	2.168 (1.222–3.847)	<b>0.008</b>	2.049 (0.994–4.223)	0.052	2.056 (1.153–3.667)	<b>0.015</b>
Risk factors†						
No						
Yes	0.742 (0.396–1.393)	0.353	1.133 (0.547–2.349)	0.736		
Grade						
G1 and G2						
G3 and G4	1.264 (0.726–2.198)	0.408	1.248 (0.663–2.351)	0.492		
AJCC						
I						
II and III and IV	1.18 (0.674–2.066)	0.561	0.961 (0.475–1.943)	0.912		
Vascular invasion						
No						
Yes	1.411 (0.775–2.569)	0.26	1.283 (0.607–2.711)	0.514		
AFP (ng/ml)						
<20						
≥20	2.015 (1.146–3.542)	<b>0.015</b>	2.047 (1.079–3.886)	<b>0.028</b>	1.954 (1.105–3.457)	<b>0.021</b>
History of other malignancy						
No						
Yes	1.944 (0.76–4.97)	0.165	2.053 (0.725–5.814)	0.176		
Family history of cancer						
No						
Yes	1.307 (0.746–2.291)	0.349	0.846 (0.451–1.587)	0.602		

Adjusted 1: Adjusted covariates include all the indicators above.

Adjusted 2: Adjusted covariates include the prognostic factors from unadjusted univariate COX analysis.

\*Represents Asian and black patients.

†Represents patients with the history of hepatitis B, hepatitis C, hemochromatosis, cirrhosis, alcohol consumption, non-alcoholic fatty liver disease or  $\alpha$ -1 antitrypsin deficiency.

The values in bold format denote that the difference between the two groups is statistically significant ( $P < 0.05$ ).

A previous study reported that altered DNA methylation patterns could be among the first detectable neoplastic changes associated with tumorigenesis [36,37]. DNA methylation alteration is frequently observed in HCC and is known to play important roles in carcinogenesis and diagnosis. Therefore, many studies have focused on DNA methylation in HCC, and several HCC methylation-based models have been proposed.

In terms of diagnostic prediction models, Cheng et al. identified six HCC-specific hyper-methylated sites (*NEBL* [cg23565942], *FAM55C* [cg21908638, cg11223367 and cg03509671], *GALNT3* [cg05569109], and *DSE* [cg11481534]) as potential diagnostic biomarkers, and the combination of these six sites showed a high sensitivity and a high specificity for the diagnosis of HCC [59]. Kisiel et al. identified a methylated DNA marker panel (*HOXA1*, *EMX1*, *AK055957*, *ECE1*, *PFKP* and *CLEC11A*), which was proven to accurately detect HCC by plasma tests [60]. Zhao et al. identified three DMGs, i.e., *ZNF300*, *SLC22A20* and *SHISA7*, which are potential markers for the early detection of HCC [61]. Zhong et al. reported that a panel of promoter methylations in *RNA5SP38*, *IL21* and *SDC4P* can

serve in the diagnosis of HCC [62]. The study by Wu et al. supported the idea that methylation markers (cg10272601 in *WNK2*, cg12680131 in *TPO* and cg22511877 in *MYT1L*) in white blood cell DNA can serve as biomarkers of HCC susceptibility [63].

In terms of diagnostic and prognostic prediction model, Xu et al. successfully constructed a diagnostic prediction model using a 10-methylation marker panel (*BMPRIA*, *PSD*, *ARHGAP25*, *KLF3*, *PLAC8*, *ATXN1*, Chr 6:170, Chr 6:3, *ATAD2* and Chr 8:20) and a prognostic prediction model with an 8-methylation marker panel (*SH3PXD2A*, *C11orf9*, *PPFIA1*, Chr 17:78, *SERPINB5*, *NOTCH3*, *GRHL2* and *TMEM8B*) after screening for circulating tumor DNA methylation markers in HCC [64]. However, this study was limited by a relatively short follow-up period. Long et al. constructed prognostic and diagnostic models consisting of two DNA methylation-driven genes (*SPP1* and *LCAT*) [65] and Lu et al. showed that a combination of multiple methylation markers (*APC*, *COX2*, *RASSF1A* and miR-203) could accurately identify hepatitis B virus (HBV)-related HCC from patient plasma samples and exhibited diagnostic and prognostic potential [66], but the inclusion and exclusion criteria of HCC patients were not reported in these two papers.

In terms of prognostic prediction model, Li et al. developed a model to predict prognosis for HCC patients based on six methylation-driven genes (*SNHG6*, *S100P*, *DCDC2*, *LIME1*, *FMO3* and *GPR171*) [67]. However, again, the inclusion and exclusion criteria of HCC patients were not reported. Dong et al. constructed a DNA methylation-based survival prediction model for HCC based on 134 methylation sites [68]. However, due to the high number of methylated markers, this model is not convenient for clinical application. Qiu et al. reported that a three-CpG island signature (*SCAND3*, *SGIP1* and *PI3*) is useful in predicting recurrence for patients with early-stage HBV-related HCC [69]. However, it needs to be confirmed that it can be applied in patients with advanced-stage disease.

In our study, after excluding patients with detectable residual tumors (R1 and R2) and patients who received adjuvant treatment (including radiation therapy, chemotherapy and embolization treatment), we identified four SR-CDMGs and constructed a composite predictive model for HCC patients. The Kaplan–Meier curve showed that the OS were significantly different between different risk groups of HCC patients. Moreover, by Cox multivariate analysis and PSM analysis, we identified the risk score as an independent prognostic factor for poor prognosis, even after adjusting for important clinical variables. Furthermore, stratified analysis results showed that our model performed well in predicting the outcome of some HCC subgroups (mainly including male patients, patients aged  $\leq 60$  years, AJCC stage I patients, patients without vascular invasion and patients with low AFP levels). Besides, we constructed a validation group by randomly selecting 148 HCC samples in the TCGA dataset; the results of ROC curve analysis and survival analysis further validated our predictive model. In addition, because this model only uses a small number of prognostic markers, it is practical and easy to use for patient assessment and therapeutic decision making. In summary, we developed a relatively objective and effective DNA methylation-based predictive model that could provide more accurate and useful information in the prognosis of HCC.

In the future, we plan to further investigate the roles of these four methylated markers in HCC carcinogenesis. In particular, the molecular functions and mechanisms of *WDR69* in HCC deserve to be further exploited. Finally, this DNA methylation-based prognosis predictive model needs to be further validated.

## Conclusion

We identified four methylated biomarkers that are significantly associated with HCC survival, which might serve as potential therapeutic targets for HCC. Our DNA methylation-based prognosis predictive model is effective and reliable in the prediction of prognosis in HCC, which could be useful for patient stratification in clinical practice, facilitate clinical therapeutic decision-making and therefore improve the prognosis for patients with HCC.

## Data Availability

The data included in the current study are available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo>) and the TCGA database (<https://cancergenome.nih.gov/>).

## Competing Interests

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

## Funding

This work was supported by the Science Foundation for Young Scientists and the Science and Technology Planning Project of Gansu Province [grant number 18JR3RA058]; the Research Projects of Gansu Provincial Hospital [grant numbers 19SYPYB-7, 19SYPYB-6, 18GSSY3-1, 18GSSY4-2]; and the Natural Science Foundation of Gansu Province [grant number 18JR3RA054].

## Author Contribution

X.-Y.H., T.W., and H.C. designed and supervised the project. A.-Q.L., H.S., Y.-F.S., T.-K.G., Y.D., and L.-T.W. performed data collection, pre-processing, analysis and visualization. X.-Y.H. wrote the manuscript. T.W., A.-Q.L., H.S., Y.-F.S., and H.C. revised the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank LetPub ([www.letpub.com](http://www.letpub.com)) for its linguistic assistance during the preparation of this manuscript

## Abbreviations

AFP,  $\alpha$ -fetoprotein; AJCC, American Joint Committee on Cancer; CDMG, common differentially methylated gene; DMP, differentially methylated probe; GO, Gene Ontology; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PPI, protein-protein interaction; PSM, propensity score-matched; ROC, receiver operating characteristic; SR-CDMG, survival-related CDMG; TCGA, the Cancer Genome Atlas; WDR, WD-repeat domain.

## References

- Maluccio, M. and Covey, A. (2012) Recent progress in understanding, diagnosing, and treating hepatocellular carcinoma. *CA Cancer J. Clin.* **62**, 394–399, <https://doi.org/10.3322/caac.21161>
- Minagawa, M., Ikai, I., Matsuyama, Y., Yamaoka, Y. and Makuuchi, M. (2007) Staging of hepatocellular carcinoma: assessment of the Japanese TNM and AJCC/UICC TNM systems in a cohort of 13,772 patients in Japan. *Ann. Surg.* **245**, 909–922, <https://doi.org/10.1097/01.sla.0000254368.65878.da>
- Bruix, J., Reig, M. and Sherman, M. (2016) Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. *Gastroenterology* **150**, 835–853, <https://doi.org/10.1053/j.gastro.2015.12.041>
- Corso-Diaz, X., Jaeger, C., Chaitankar, V. and Swaroop, A. (2018) Epigenetic control of gene regulation during development and disease: A view from the retina. *Prog. Retin. Eye Res.* **65**, 1–27, <https://doi.org/10.1016/j.preteyeres.2018.03.002>
- Draht, M.X., Riedl, R.R., Niessen, H., Carvalho, B., Meijer, G.A., Herman, J.G. et al. (2012) Promoter CpG island methylation markers in colorectal cancer: the road ahead. *Epigenomics* **4**, 179–194, <https://doi.org/10.2217/epi.12.9>
- Ng, J.M. and Yu, J. (2015) Promoter hypermethylation of tumour suppressor genes as potential biomarkers in colorectal cancer. *Int. J. Mol. Sci.* **16**, 2472–2496, <https://doi.org/10.3390/ijms16022472>
- Lorincz, M.C. and Schubeler, D. (2017) Evidence for converging DNA methylation pathways in placenta and cancer. *Dev. Cell* **43**, 257–258, <https://doi.org/10.1016/j.devcel.2017.10.009>
- Tao, Y.F., Fang, F., Hu, S.Y., Lu, J., Cao, L., Zhao, W.L. et al. (2015) Hypermethylation of the GATA binding protein 4 (GATA4) promoter in Chinese pediatric acute myeloid leukemia. *BMC Cancer* **15**, 756, <https://doi.org/10.1186/s12885-015-1760-5>
- Zhu, C., Utsunomiya, T., Ikemoto, T., Yamada, S., Morine, Y., Imura, S. et al. (2014) Hypomethylation of long interspersed nuclear element-1 (LINE-1) is associated with poor prognosis via activation of c-MET in hepatocellular carcinoma. *Ann. Surg. Oncol.* **21**, S729–S735, <https://doi.org/10.1245/s10434-014-3874-4>
- Kuo, C.C., Shih, Y.L., Su, H.Y., Yan, M.D., Hsieh, C.B., Liu, C.Y. et al. (2015) Methylation of IRAK3 is a novel prognostic marker in hepatocellular carcinoma. *World J. Gastroenterol.* **21**, 3960–3969, <https://doi.org/10.3748/wjg.v21.i13.3960>
- Cui, C., Gan, Y., Gu, L., Wilson, J., Liu, Z., Zhang, B. et al. (2015) P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. *Genome Biol.* **16**, 252, <https://doi.org/10.1186/s13059-015-0819-6>
- Grasse, S., Lienhard, M., Frese, S., Kerick, M., Steinbach, A., Grimm, C. et al. (2018) Epigenomic profiling of non-small cell lung cancer xenografts uncover LRP12 DNA methylation as predictive biomarker for carboplatin resistance. *Genome Med.* **10**, 55, <https://doi.org/10.1186/s13073-018-0562-1>
- Zhang, J., Zhang, J., Xu, S., Zhang, X., Wang, P., Wu, H. et al. (2018) Hypoxia-induced TPM2 methylation is associated with chemoresistance and poor prognosis in breast cancer. *Cell. Physiol. Biochem.* **45**, 692–705, <https://doi.org/10.1159/000487162>
- Morris, T.J., Butcher, L.M., Feber, A., Teschendorff, A.E., Chakravarthi, A.R., Wojdacz, T.K. et al. (2014) ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430, <https://doi.org/10.1093/bioinformatics/btt684>
- Bacos, K., Gillberg, L., Volkov, P., Olsson, A.H., Hansen, T., Pedersen, O. et al. (2016) Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat. Commun.* **7**, 11089, <https://doi.org/10.1038/ncomms11089>
- Zhou, W., Laird, P.W. and Shen, H. (2017) Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, e22
- Nordlund, J., Backlin, C.L., Wahlberg, P., Busche, S., Berglund, E.C., Eloranta, M.L. et al. (2013) Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* **14**, r105, <https://doi.org/10.1186/gb-2013-14-9-r105>
- Zhang, S., Wang, Y., Gu, Y., Zhu, J., Ci, C., Guo, Z. et al. (2018) Specific breast cancer prognosis-subtype distinctions based on DNA methylation patterns. *Mol. Oncol.* **12**, 1047–1060, <https://doi.org/10.1002/1878-0261.12309>
- Vila, A.M., de Arrilucea, P.R., Pelaez, E.C. and Gago-Martinez, A. (2010) Development of a new magnetic beads-based immunoprecipitation strategy for proteomics analysis. *J. Proteomics* **73**, 1491–1501, <https://doi.org/10.1016/j.jprot.2010.02.015>

- 20 Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. et al. (2000) Gene ontology: tool for the unification of biology. *Gene Ontol. Consortium. Nat. Genet.* **25**, 25–29, <https://doi.org/10.1038/75556>
- 21 Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34, <https://doi.org/10.1093/nar/27.1.29>
- 22 Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguéz, P. et al. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568, <https://doi.org/10.1093/nar/gkq973>
- 23 Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J. et al. (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412–D416, <https://doi.org/10.1093/nar/gkn760>
- 24 von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261, <https://doi.org/10.1093/nar/gkg034>
- 25 Wei, J.H., Haddad, A., Wu, K.J., Zhao, H.W., Kapur, P., Zhang, Z.L. et al. (2015) A CpG-methylation-based assay to predict survival in clear cell renal cell carcinoma. *Nat. Commun.* **6**, 8699, <https://doi.org/10.1038/ncomms9699>
- 26 Yang, Y.T., Chiu, Y.C., Kao, C.J., Hou, H.A., Lin, C.C., Tsai, C.H. et al. (2018) The prognostic significance of global aberrant alternative splicing in patients with myelodysplastic syndrome. *Blood Cancer J.* **8**, 78, <https://doi.org/10.1038/s41408-018-0115-2>
- 27 Lohavanichbutr, P., Mendez, E., Holsinger, F.C., Rue, T.C., Zhang, Y., Houck, J. et al. (2013) A 13-gene signature prognostic of HPV-negative OSCC: discovery and external validation. *Clin. Cancer Res.* **19**, 1197–1203, <https://doi.org/10.1158/1078-0432.CCR-12-2647>
- 28 Lao, J., Chen, Y., Li, Z.C., Li, Q., Zhang, J., Liu, J. et al. (2017) A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* **7**, 10353, <https://doi.org/10.1038/s41598-017-10649-8>
- 29 Mainio, A., Hakko, H., Niemela, A., Koivukangas, J. and Rasanen, P. (2005) Depression and functional outcome in patients with brain tumors: a population-based 1-year follow-up study. *J. Neurosurg.* **103**, 841–847, <https://doi.org/10.3171/jns.2005.103.5.0841>
- 30 Haddad, A.Q., Luo, J.H., Krabbe, L.M., Darwish, O., Gayed, B., Youssef, R. et al. (2017) Prognostic value of tissue-based biomarker signature in clear cell renal cell carcinoma. *BJU Int.* **119**, 741–747, <https://doi.org/10.1111/bju.13776>
- 31 Lindahl, L.M., Besenbacher, S., Rittig, A.H., Celis, P., Willerslev-Olsen, A., Gjerdrum, L.M.R. et al. (2018) Prognostic miRNA classifier in early-stage mycosis fungoides: development and validation in a Danish nationwide study. *Blood* **131**, 759–770, <https://doi.org/10.1182/blood-2017-06-788950>
- 32 Moreno-Betancur, M., Sadaoui, H., Piffaretti, C. and Rey, G. (2017) Survival analysis with multiple causes of death: extending the competing risks model. *Epidemiology* **28**, 12–19, <https://doi.org/10.1097/EDE.0000000000000531>
- 33 Parikh, N.D., Marshall, V.D., Singal, A.G., Nathan, H., Lok, A.S., Balkrishnan, R. et al. (2017) Survival and cost-effectiveness of sorafenib therapy in advanced hepatocellular carcinoma: an analysis of the SEER-Medicare database. *Hepatology* **65**, 122–133, <https://doi.org/10.1002/hep.28881>
- 34 Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C. et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77, <https://doi.org/10.1186/1471-2105-12-77>
- 35 Chen, T. (2011) Mechanistic and functional links between histone methylation and DNA methylation. *Prog. Mol. Biol. Transl. Sci.* **101**, 335–348, <https://doi.org/10.1016/B978-0-12-387685-0-00010-X>
- 36 Baylin, S.B. and Jones, P.A. (2016) Epigenetic determinants of cancer. *Cold Spring Harb. Persp. Biol.* **8**, <https://doi.org/10.1101/cshperspect.a019505>
- 37 Baylin, S.B. and Jones, P.A. (2011) A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734, <https://doi.org/10.1038/nrc3130>
- 38 Fu, L., Yao, T., Chen, Q., Mo, X., Hu, Y. and Guo, J. (2017) Screening differential circular RNA expression profiles reveals hsa\_circ\_0004018 is associated with hepatocellular carcinoma. *Oncotarget* **8**, 58405–58416, <https://doi.org/10.18632/oncotarget.16881>
- 39 Mo, S.J., Hou, X., Hao, X.Y., Cai, J.P., Liu, X., Chen, W. et al. (2018) EYA4 inhibits hepatocellular carcinoma growth and invasion by suppressing NF-kappaB-dependent RAP1 transactivation. *Cancer Commun. (Lond.)* **38**, 9, <https://doi.org/10.1186/s40880-018-0276-1>
- 40 Zha, Y., Gan, P., Yao, Q., Ran, F.M. and Tan, J. (2014) Downregulation of Rap1 promotes 5-fluorouracil-induced apoptosis in hepatocellular carcinoma cell line HepG2. *Oncol. Rep.* **31**, 1691–1698, <https://doi.org/10.3892/or.2014.3033>
- 41 Lowdon, R.F. and Wang, T. (2017) Epigenomic annotation of noncoding mutations identifies mutated pathways in primary liver cancer. *PLoS ONE* **12**, e0174032, <https://doi.org/10.1371/journal.pone.0174032>
- 42 MacPherson, L., Anokye, J., Yeung, M.M., Lam, E.Y.N., Chan, Y.C., Weng, C.F. et al. (2020) HBO1 is required for the maintenance of leukaemia stem cells. *Nature* **577**, 266–270, <https://doi.org/10.1038/s41586-019-1835-6>
- 43 Antonchuk, J., Sauvageau, G. and Humphries, R.K. (2002) HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. *Cell* **109**, 39–45, [https://doi.org/10.1016/S0092-8674\(02\)00697-9](https://doi.org/10.1016/S0092-8674(02)00697-9)
- 44 Shao, M., Yang, Q., Zhu, W., Jin, H., Wang, J., Song, J. et al. (2018) LncHOXA10 drives liver TICs self-renewal and tumorigenesis via HOXA10 transcription activation. *Mol. Cancer* **17**, 173, <https://doi.org/10.1186/s12943-018-0921-y>
- 45 Zhang, Y., Chen, J., Wu, S.S., Lv, M.J., Yu, Y.S., Tang, Z.H. et al. (2019) HOXA10 knockdown inhibits proliferation, induces cell cycle arrest and apoptosis in hepatocellular carcinoma cells through HDAC1. *Cancer Manag. Res.* **11**, 7065–7076, <https://doi.org/10.2147/CMAR.S199239>
- 46 Xiao, Z.D., Jiao, C.Y., Huang, H.T., He, L.J., Zhao, J.J., Lu, Z.Y. et al. (2014) miR-218 modulate hepatocellular carcinoma cell proliferation through PTEN/AKT/PI3K pathway and HoxA10. *Int. J. Clin. Exp. Pathol.* **7**, 4039–4044
- 47 Park, S.M., Choi, E.Y., Bae, M., Choi, J.K. and Kim, Y.J. (2017) A long-range interactive DNA methylation marker panel for the promoters of HOXA9 and HOXA10 predicts survival in breast cancer patients. *Clin. Epigenetics* **9**, 73, <https://doi.org/10.1186/s13148-017-0373-z>
- 48 Wang, H., Jia, X.H., Chen, J.R., Yi, Y.J., Wang, J.Y., Li, Y.J. et al. (2016) HOXB4 knockdown reverses multidrug resistance of human myelogenous leukemia K562/ADM cells by downregulating P-gp, MRP1 and BCRP expression via PI3K/Akt signaling pathway. *Int. J. Oncol.* **49**, 2529–2537, <https://doi.org/10.3892/ijo.2016.3738>
- 49 Shu, Y., Wang, B., Wang, J., Wang, J.M. and Zou, S.Q. (2011) Identification of methylation profile of HOX genes in extrahepatic cholangiocarcinoma. *World J. Gastroenterol.* **17**, 3407–3419, <https://doi.org/10.3748/wjg.v17.i29.3407>

- 50 Shen, J., Wang, S., Zhang, Y.J., Kappil, M.A., Chen Wu, H., Kibriya, M.G. et al. (2012) Genome-wide aberrant DNA methylation of microRNA host genes in hepatocellular carcinoma. *Epigenetics* **7**, 1230–1237, <https://doi.org/10.4161/epi.22140>
- 51 Zhang, X.B., Beard, B.C., Trobridge, G.D., Wood, B.L., Sale, G.E., Sud, R. et al. (2008) High incidence of leukemia in large animals after stem cell gene therapy with a HOXB4-expressing retroviral vector. *J. Clin. Invest.* **118**, 1502–1510, <https://doi.org/10.1172/JCI34371>
- 52 Barba-de la Rosa, A.P., Briones-Cerecero, E., Lugo-Melchor, O., De León-Rodríguez, A., Santos, L., Castelo-Ruelas, J. et al. (2012) Hox B4 as potential marker of non-differentiated cells in human cervical cancer cells. *J. Cancer Res. Clin. Oncol.* **138**, 293–300, <https://doi.org/10.1007/s00432-011-1081-2>
- 53 Zhou, Y., Qiu, X.P., Li, Z.H., Zhang, S., Rong, Y., Yang, G.H. et al. (2019) Clinical significance of aberrant cyclin-dependent kinase-like 2 methylation in hepatocellular carcinoma. *Gene* **683**, 35–40, <https://doi.org/10.1016/j.gene.2018.10.009>
- 54 Zheng, Y., Huang, Q., Ding, Z., Liu, T., Xue, C., Sang, X. et al. (2018) Genome-wide DNA methylation analysis identifies candidate epigenetic markers and drivers of hepatocellular carcinoma. *Brief. Bioinform.* **19**, 101–108
- 55 Hu, D.J., Shi, W.J., Yu, M. and Zhang, L. (2019) High WDR34 mRNA expression as a potential prognostic biomarker in patients with breast cancer as determined by integrated bioinformatics analysis. *Oncol. Lett.* **18**, 3177–3187, <https://doi.org/10.3892/ol.2019.10634>
- 56 Yuan, Y., Qi, G., Shen, H., Guo, A., Cao, F., Zhu, Y. et al. (2019) Clinical significance and biological function of WD repeat domain 54 as an oncogene in colorectal cancer. *Int. J. Cancer* **144**, 1584–1595, <https://doi.org/10.1002/ijc.31736>
- 57 Park, K., Lee, J.H., Han, H.H., Mun, S.G., Kang, S., Cha, Y.J. et al. (2017) Nodal metastasis signatures in breast cancer. *Pathol. Res. Pract.* **213**, 680–687, <https://doi.org/10.1016/j.prp.2016.11.017>
- 58 Yamamoto, J.I., Kasamatsu, A., Okubo, Y., Nakashima, D., Fushimi, K., Minakawa, Y. et al. (2018) Evaluation of tryptophan-aspartic acid repeat-containing protein 34 as a novel tumor-suppressor molecule in human oral cancer. *Biochem. Biophys. Res. Commun.* **495**, 2469–2474, <https://doi.org/10.1016/j.bbrc.2017.12.138>
- 59 Cheng, J., Wei, D., Ji, Y., Chen, L., Yang, L., Li, G. et al. (2018) Integrative analysis of DNA methylation and gene expression reveals hepatocellular carcinoma-specific diagnostic biomarkers. *Genome Med.* **10**, 42, <https://doi.org/10.1186/s13073-018-0548-z>
- 60 Kisiel, J.B., Dukek, B.A., Kanipakam R., V.S.R., Ghaz, H.M., Yab, T.C., Berger, C.K. et al. (2019) Hepatocellular carcinoma detection by plasma methylated DNA: discovery, Phase I Pilot, and Phase II clinical validation. *Hepatology* **69**, 1180–1192, <https://doi.org/10.1002/hep.30244>
- 61 Zhao, Y., Xue, F., Sun, J., Guo, S., Zhang, H., Qiu, B. et al. (2014) Genome-wide methylation profiling of the different stages of hepatitis B virus-related hepatocellular carcinoma development in plasma cell-free DNA reveals potential biomarkers for early detection and high-risk monitoring of hepatocellular carcinoma. *Clin. Epigenetics* **6**, 30, <https://doi.org/10.1186/1868-7083-6-30>
- 62 Zhong, D. and Cen, H. (2017) Aberrant promoter methylation profiles and association with survival in patients with hepatocellular carcinoma. *Oncotargets Ther.* **10**, 2501–2509, <https://doi.org/10.2147/OTT.S128058>
- 63 Wu, H.C., Shen, J., Yang, H.I., Tsai, W.Y., Chen, C.J. and Santella, R.M. (2016) Blood DNA methylation markers in prospectively identified hepatocellular carcinoma cases and controls from Taiwan. *World J. Hepatol.* **8**, 301–306, <https://doi.org/10.4254/wjh.v8.i5.301>
- 64 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>
- 65 Long, J., Chen, P., Lin, J., Bai, Y., Yang, X., Bian, J. et al. (2019) DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. *Theranostics* **9**, 7251–7267, <https://doi.org/10.7150/thno.31155>
- 66 Lu, C.Y., Chen, S.Y., Peng, H.L., Kan, P.Y., Chang, W.C. and Yen, C.J. (2017) Cell-free methylation markers with diagnostic and prognostic potential in hepatocellular carcinoma. *Oncotarget* **8**, 6406–6418, <https://doi.org/10.18632/oncotarget.14115>
- 67 Li, G.X., Ding, Z.Y., Wang, Y.W., Liu, T.T., Chen, W.X., Wu, J.J. et al. (2019) Integrative analysis of DNA methylation and gene expression identify a six epigenetic driver signature for predicting prognosis in hepatocellular carcinoma. *J. Cell. Physiol.* **234**, 11942–11950, <https://doi.org/10.1002/jcp.27882>
- 68 Dong, R.Z., Yang, X., Zhang, X.Y., Gao, P.T., Ke, A.W., Sun, H.C. et al. (2019) Predicting overall survival of patients with hepatocellular carcinoma using a three-category method based on DNA methylation and machine learning. *J. Cell. Mol. Med.* **23**, 3369–3374, <https://doi.org/10.1111/jcmm.14231>
- 69 Qiu, J., Peng, B., Tang, Y., Qian, Y., Guo, P., Li, M. et al. (2017) CpG methylation signature predicts recurrence in early-stage hepatocellular carcinoma: results from a multicenter study. *J. Clin. Oncol.* **35**, 734–742, <https://doi.org/10.1200/JCO.2016.68.2153>