

RESEARCH

Open Access



# Identification of DNA methylation characteristics associated with metastasis and prognosis in colorectal cancer

Fang Qian<sup>1,2</sup>, Qiang Li<sup>1</sup>, Huidan Chang<sup>1</sup>, Kai Wei<sup>1</sup>, Xiaoyi Chen<sup>2\*</sup>, Tao Huang<sup>1\*</sup> and Yixue Li<sup>1,3,4,5,6\*</sup>

## Abstract

Colorectal cancer (CRC) is prone to metastasis and recurrence after surgery, which is one of the main causes for its poor treatment and prognosis. Therefore, it is essential to identify biomarkers associated with metastasis and recurrence in CRC. DNA methylation has a regulatory role in cancer metastasis, tumor immune microenvironment (TME), and prognosis and may be one of the most valuable biomarkers for predicting CRC metastasis and prognosis. We constructed a diagnostic model and nomogram that can effectively predict CRC metastasis based on the differential methylation CpG sites (DMCs) between metastatic and non-metastatic CRC patients. Then, we identified 17 DMCs associated with progression free survival (PFS) of CRC and constructed a prognostic model. The prognosis model based on 17 DMCs can predict the PFS of CRC with medium to high accuracy. The results of immunohistochemical analysis indicated that the protein expression levels of the genes involved in prognostic DMCs were different between normal and colorectal cancer tissues. According to the results of immune-related analysis, we found that the low-risk patients had better immunotherapy response. In addition, high risk scores were negatively correlated with high tumor mutation burden (TMB) levels, and patients with low TMB levels in the high-risk group had the worst PFS. Our work shows the clinical value of DNA methylation in predicting CRC metastasis and PFS, as well as their correlation with TME, immunotherapy, and TMB, which helps understand the changes of DNA methylation in CRC metastasis and improving the treatment and prognosis of CRC.

**Keywords** Colorectal cancer (CRC), DNA methylation, Tumor immune microenvironment (TME), Tumor metastasis, Prognosis

\*Correspondence:

Xiaoyi Chen  
chenxiaoyi@mailsucas.ac.cn  
Tao Huang  
huangtao@sinh.ac.cn  
Yixue Li

yxli@sinh.ac.cn

<sup>1</sup> Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup> Guoke Ningbo Life Science and Health Industry Research Institute, Ningbo 315000, China

<sup>3</sup> Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

<sup>4</sup> Guangzhou Laboratory, Guangzhou 510005, China

<sup>5</sup> School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>6</sup> Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai 200433, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Colorectal cancer (CRC) is a common and highly invasive tumor in the digestive tract, and its incidence rate and mortality are increasing yearly [1]. CRC is prone to metastasis or recurrence after conventional treatment (such as surgical resection), and the 5-year survival of CRC patients with metastasis is worse than that of primary cancer patients [2]. At present, systemic chemotherapy is often used to inhibit the growth and spread of cancer cells. However, it cannot eliminate potential disseminated cancer cells and only benefits patients for several months [3]. Cancer metastasis seriously affects the treatment and survival of CRC patients [4]. Therefore, it's urgent to understand the process of metastasis and screen new biomarkers that can predict the metastasis and prognosis of CRC patients.

Accumulative evidence shows that abnormal DNA methylation regulates cancer occurrence and progression [5]. The change of DNA methylation pattern occurs in the early stage of carcinogenesis and leads to the silencing of multiple tumor suppressor genes in CRC [6]. It is reported that DNA methylation may directly affect gene transcription to promote cancer transformation and tumor metastasis [7, 8]. In addition, several studies have been conducted to find abnormal methylation biomarkers based on DNA in plasma or feces to develop noninvasive diagnostic tools related to CRC [9]. Jin et al. developed a quantitative analysis method for DNA methylation markers to monitor CRC [10] effectively. Therefore, DNA methylation-related features may become a promising candidate for CRC biomarker development.

Many epigenetic studies have confirmed that DNA methylation plays a key regulatory role in inflammation, TME, and immunotherapy [11]. Xu et al. have shown that DNA methylation profiles can predict immunotherapy responses at the pan-cancer level [12]. Based on the 24 DNA methylation regulators in CRC, Yuan et al. distinguished 3 DNA methylation patterns with different TME and prognostic features [13]. Therefore, it is of great importance to investigate the correlation between DNA methylation with TME in guiding immunotherapy and improving the prognosis of CRC.

In this study, we elucidated the changes in DNA methylation during CRC metastasis and its correlation with prognosis and TME. Firstly, we identified differentially methylated CpG sites (DMCs) between CRC metastatic and non-metastatic groups. Subsequently, DMCs that can predict CRC transfer were identified. Then, we constructed a DNA methylation-related prognosis and nomogram model that can predict PFS in CRC patients to evaluate the clinical value of metastasis-related DMCs. In addition, the correlation of DNA methylation with TME and immunotherapy in CRC was determined by immune

infiltration correlation analysis, MSI analysis, and tumor mutation burden (TMB) analysis.

## Materials and methods

### Data acquisition and preprocessing

The 450 K DNA methylation array ( $n=289$ ), 27 K DNA methylation array ( $n=153$ ) of colorectal cancer (CRC) patients, and the corresponding clinical information and progression free survival data were obtained from the UCSC Xena database. 27 K DNA methylation array was used for the validation dataset of the prognosis model. Patients in the TCGA-COAD queue have received various treatment methods, including surgery, radiation therapy, drug therapy, and immunotherapy. CRC samples with progression free survival times greater than 0 and their corresponding data were used for subsequent analysis. 450 K DNA methylation array (GSE164811) of CRC patients was obtained from GEO database as the validation dataset of the diagnostic model. In addition, we downloaded tumor mutation burden (TMB) data related to CRC patients from TCGA database.

Then, quality control of CpG sites in the DNA methylation array was performed. First, remove the CpG locus located on the sex chromosome. Remove CpG sites null in more than 70% of the samples, and retain CpG sites where the transcription start site is 2 kb upstream to 0.5 kb downstream. The overlapping CpG sites in the 450 K DNA methylation array, 27 K DNA methylation array in TCGA, and 450 K DNA methylation array in GEO were reserved. Finally, 21,122 CpG sites were retained. Subsequently, KNN was used to impute missing values in DNA methylation array data. The "SVA" package was utilized to remove the batch effect between the three DNA methylation arrays.

### Differential analysis of methylation CpG sites

In this study, CRC patients were separate the tumor metastasis group and tumor non-metastasis group according to M (distant metastasis) stage and N (lymph node metastasis) stage. CRC patients with both M stage and N stage 0 ( $M=M0$  &  $N=N0$ ) were included in the tumor non-metastasis group, and the remaining CRC patients were included in the tumor metastasis group. In order to obtain differential methylation CpG sites (DMCs) in patients with and without metastasis, we used the "ChAMP" package to standardize and analyze the difference of 450 K DNA methylation array in TCGA. The methylation CpG sites with  $P$  values less than 0.01 were considered to be significantly different. These DMCs intersect with 21,122 methylation sites obtained by QC as the final DMCs.

### Functional enrichment analysis

We annotated the genes where DMCs are located. We conducted functional enrichment analysis to clarify the biological processes involved in developing CRC by DMCs. The “clusterProfiler” package was used to perform KEGG analysis. The condition of significant pathway is that its  $p$  value is less than 0.05. Next, we conducted Metascape analysis on the online Metascape platform (<https://metascape.org/>) for the genes where DMCs are located.

### Identification of DMCs associated with tumor metastasis in CRC

Firstly, univariate Cox regression analysis and KM test were conducted to evaluate the correlation between DMCs expression and PFS of CRC. DMCs with  $p$  values less than 0.05 were reserved for subsequent analysis. R packages “igraph” and “reshape2” were utilized to map the network of PFS-related CpG sites. The SVM-RFE (R package “e1071”) method was used to identify DMCs associated with metastasis in CRC. An SVM classifier based on the  $\beta$  values of DMCs was constructed to predict CRC metastasis. The random seed was set to “124579”. ROC curve was used to evaluate the accuracy of the diagnostic model constructed based on DMCs in predicting CRC metastasis. Based on the CRC metastasis-related DMCs obtained from SVM screening, we constructed a nomogram using the “rms” package. The calibration curve was utilized to evaluate the accuracy of the nomogram in predicting metastasis in CRC patients. Decision Curve Analysis (DCA) and Clinical Impact Curve were used to assess the performance of the nomogram.

### Construction of prognosis model and nomogram related to the PFS of CRC

To further identify CpG sites associated with PFS in CRC patients, LASSO-Cox regression analysis (“glmnet” package) was conducted on DMCs in the diagnostic model. Through 1000 iterations, the optimal penalty parameter  $\lambda$  of the model was determined. The risk score of each CRC patient was obtained based on the  $\beta$  values ( $CpG_i$ ) of the candidate CpG sites and their corresponding regression coefficients ( $coef_i$ ).  $riskScore = \sum_{i=1}^n (CpG_i * coef_i)$ ,  $n$  is the number of methylated CpG sites associated with PFS in CRC patients. Next, Kaplan-Meier analysis was performed to determine the difference in PFS between the two risk groups. The ROC curve was used to evaluate the accuracy of the constructed prognostic model in predicting 1- and 3-year PFS of CRC patients. Independent prognostic analysis was performed on the clinical

characteristics of CRC as well as the risk scores of prognostic models to identify factors that could independently predict PFS in CRC patients.

The “rms” package was conducted to construct a nomogram for predicting PFS of CRC patients. The calibration curve and ROC curve were drawn to evaluate the performance of the nomogram model in predicting the PFS of CRC patients.

### Immune microenvironment and microsatellite instability analysis

We utilized ssGSEA to obtain the scores of immune cells and immune-related pathways in CRC patients. In addition, we performed immune correlation analysis on the DNA methylation data (450 K) of TCGA-COAD using the HiTIMED algorithm. In order to determine the sensitivity of CRC patients in high- and low-risk groups to immunotherapy, we calculated the tumor immune dysfunction and exclusion (TIDE) score of the TCGA-COAD dataset from the TIDE online analysis platform (<http://tide.dfci.harvard.edu/>). The higher the TIDE score, the worse the immune response of patients. In addition, we conducted microsatellite instability (MSI) analysis to explore the correlation between risk score and MSI.

### Tumor mutation burden analysis

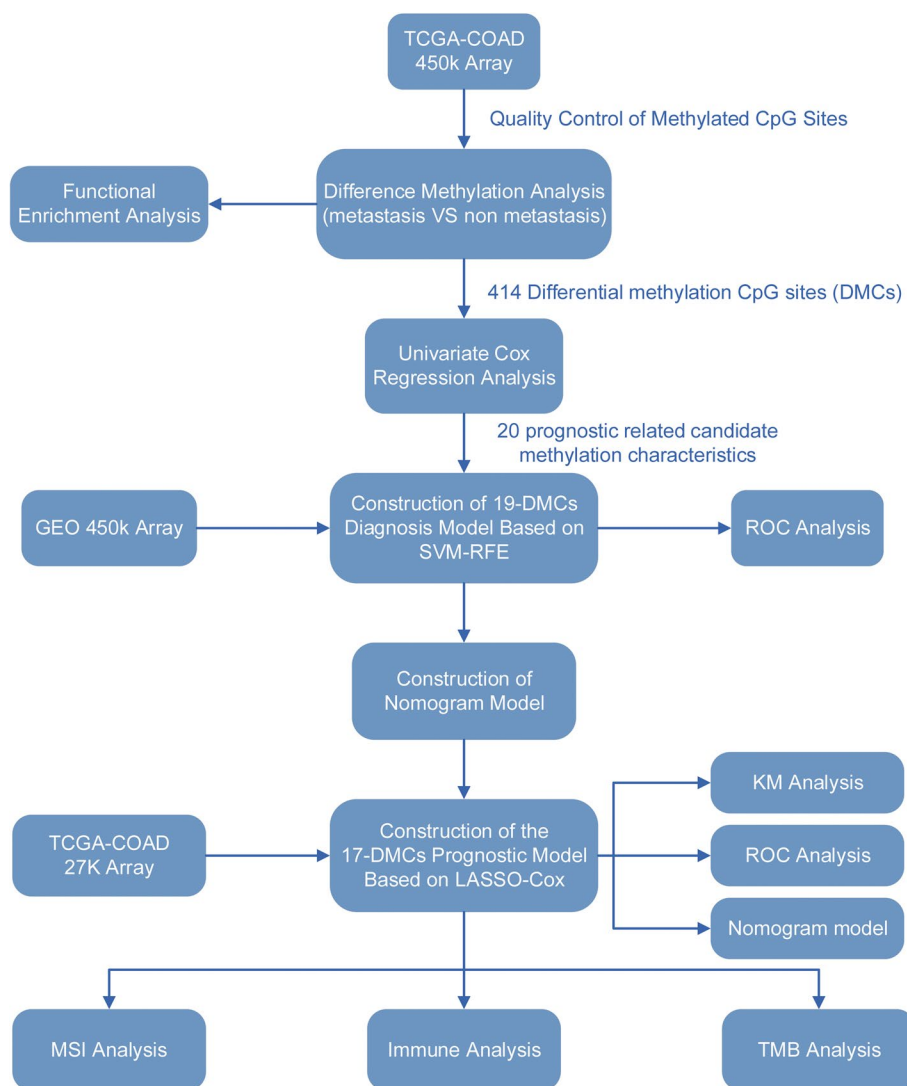
The “maftools” package was utilized to draw waterfall diagrams to show the TMB levels of two risk subgroups. In order to further determine the correlation between TMB and PFS of CRC patients, KM analysis was used to compare the difference of PFS of CRC patients between high- and low-TMB groups.

### Immunohistochemical analysis

To further elucidate the role of DMCs in the progression of CRC, we explored the expression at the protein level of genes in which prognostic related DMCs are located. We downloaded the immunohistochemical maps of normal and CRC tissues from the Human Protein Atlas (HPA) database.

### Statistical analysis

The data analysis and results visualization of this study was conducted on R (4.2.2). T-test or Wilcoxon-test was used to compare the differences between groups. Spearman was used for correlation analysis. In this study, a  $p$ -value less than 0.05 is statistically significant, and “\*” represents  $p < 0.05$ , “\*\*” represents  $p < 0.01$ , and “\*\*\*” represents  $p < 0.001$ . The workflow is shown in Fig. 1.



**Fig. 1** Workflow of the research

**Results**

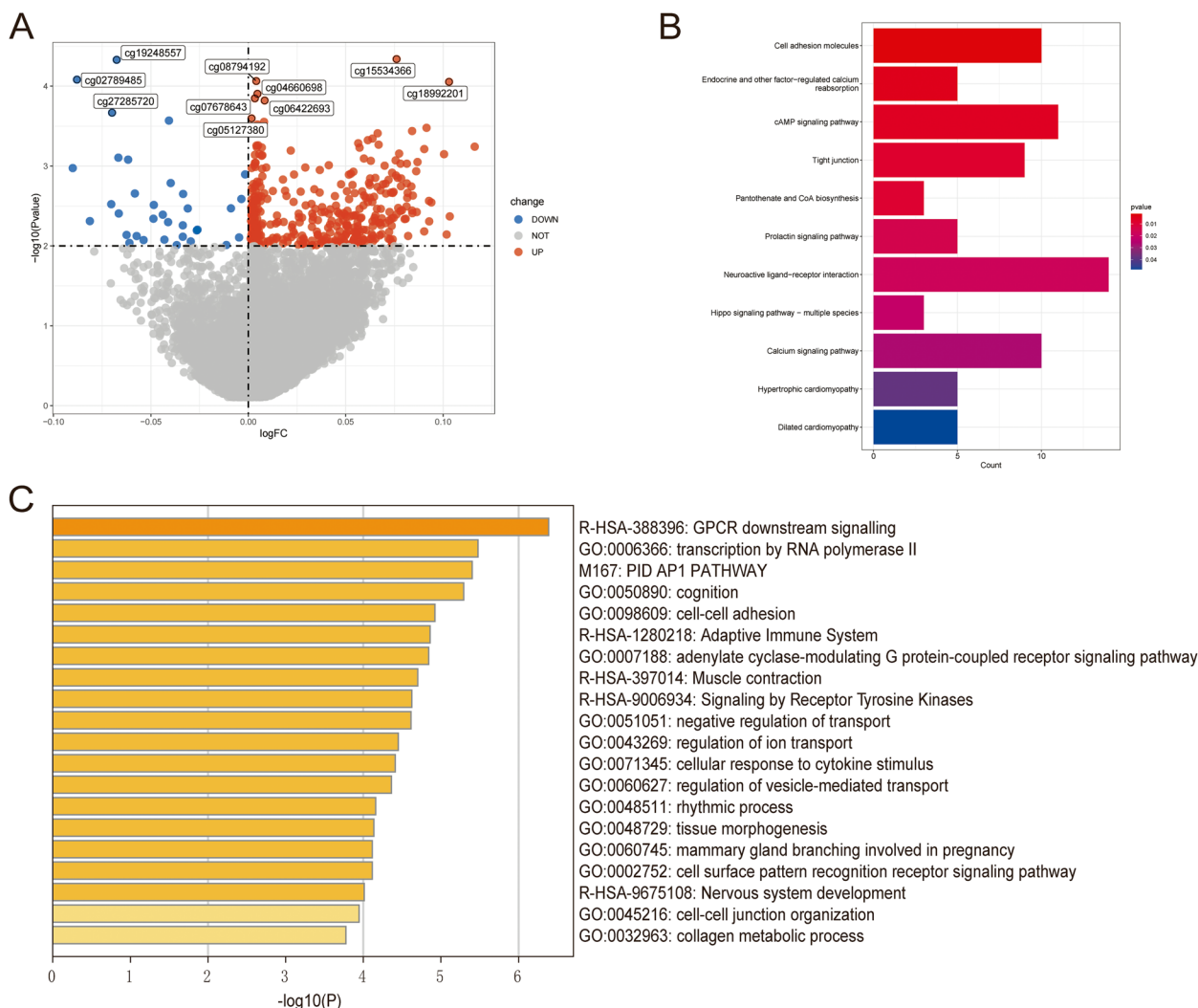
**Identification of differential methylated CpG sites**

According to TCGA-COAD 450 K methylation data, according to the *p*-value less than 0.01, there are 42,310 methylation CpG sites with differences between the metastatic and non-metastatic groups (Supplementary material 1). Then, 42,310 CpG sites obtained by the differential analysis were intersected with 21,122 CpG sites after quality control, and 414 CpG sites were used as DMCs for subsequent analysis. We found that the  $\beta$  values of DMCs were upregulated more in metastatic CRC patients. Among them, 380 DMCs had higher  $\beta$  values in the metastasis group than in the non-metastasis group, and 34 DMCs had lower  $\beta$  values in the metastasis group (Fig. 2A). In Fig. 2B, the genes where DMCs are located

were mainly participate in ligand-receptor interaction, Tight junction, Neuroactive cAMP signaling pathway, Cell adhesion molecules, and other biological pathways (Supplementary material 2). Metasape analysis showed that the genes where DMCs are located are mainly involved in GPCR downstream signalling, transcription by RNA polymerase II, PID AP1 PATHWAY, and cognition (Fig. 2C, Supplementary material 3).

**Methylation characteristics associated with CRC metastasis**

In order to screen CpG sites that could diagnose metastasis in CRC patients, we performed a comprehensive analysis. First, we utilised univariate Cox regression analysis and KM test to discern candidate DMCs associated with PFS in CRC patients. Among them, 186

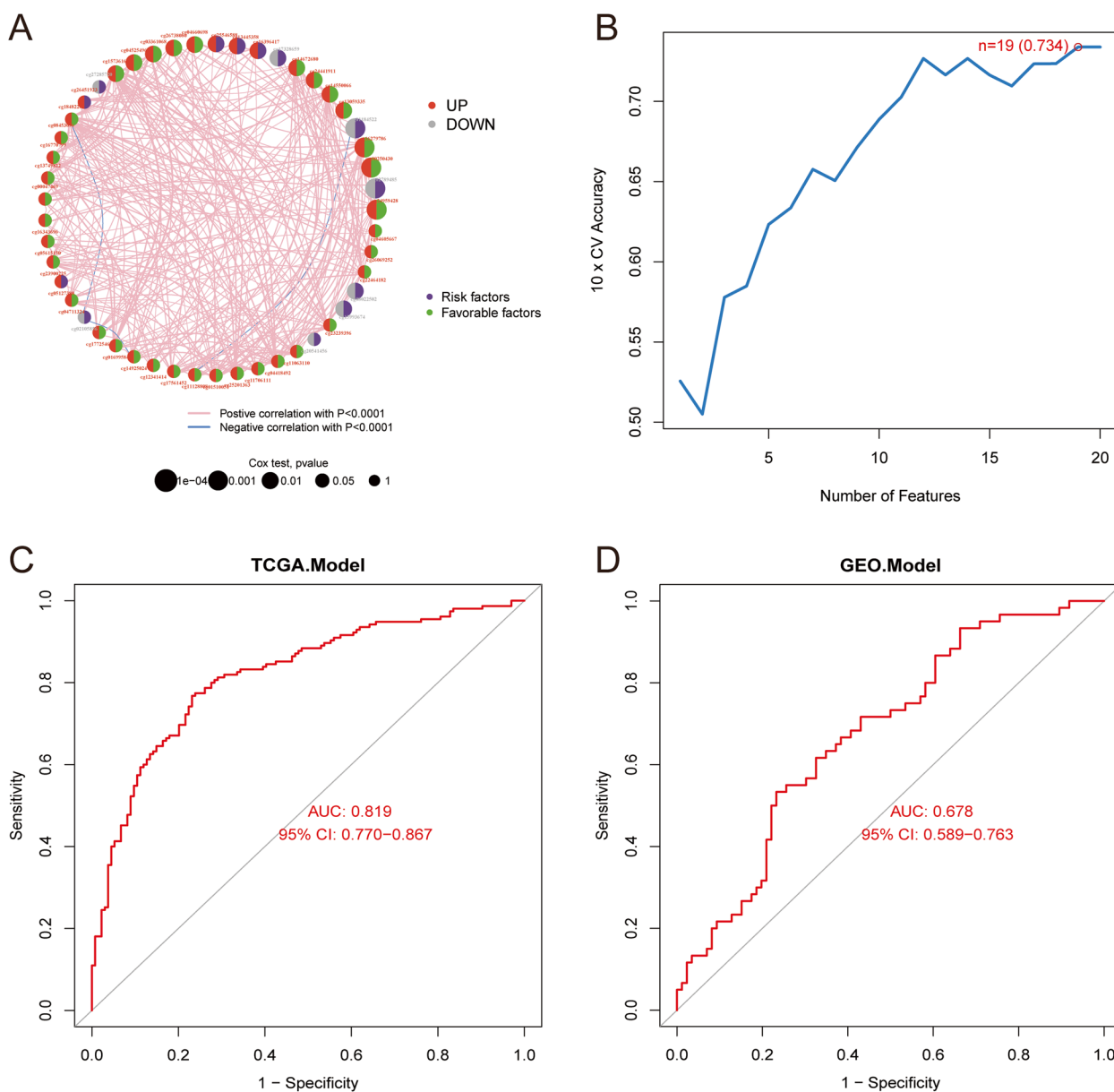


**Fig. 2** Identification of differential methylation CpG Sites (DMCs). **A** Volcano plot of DMCs between metastatic and non-metastatic CRC patients. The red represents the up-regulated methylation CpG Sites, blue represents the down-regulated methylation CpG Sites. **B** KEGG analysis of DMCs located genes. **C** Metascape analysis of DMCs located genes

DMCs with a *p*-value less than 0.05 in the KM test (Supplementary material 4). In Fig. 3A, we plotted the network of the top 50 CpG sites interactions, expression, and their association with PFS in CRC patients. DMCs with *p* values less than 0.05 in univariate Cox regression analysis and KM test were used to construct a diagnostic model for predicting metastasis of CRC patients, and a total of 20 DMCs were obtained. Subsequently, we used SVM-RFE to screen methylation signatures predicting metastasis in CRC patients based on the  $\beta$  values of 20 DMCs. According to the results of 10-fold cross-validation (Fig. 3B), the diagnostic model had the highest accuracy (0.734) when the number of methylation features was 19 (cg04660698, cg02789485, cg03361068, cg26738080, cg25546588, cg14550066,

cg08022502, cg17328659, cg01184522, cg15993674, cg24441911, cg04525496, cg14672680, cg13445358, cg15736165, cg16279786, cg16396417, cg00250430, and cg13059335). The results of ROC analysis declared that 19 DMCs diagnostic model had high accuracy in predicting the metastasis of CRC patients in the TCGA-COAD cohort (Fig. 3C, AUC=0.819) and the GEO cohort (Fig. 3D, AUC=0.678).

Then, we constructed a nomogram model based on the DMCs in the diagnostic model to predict the risk of metastasis in CRC patients (Fig. 4A). In Fig. 4B, the calibration curve showed that the error between the actual risk of CRC patients and the metastasis risk predicted by the nomogram is minor, suggesting that the nomogram based on 19-DMCs can predict the metastasis of CRC

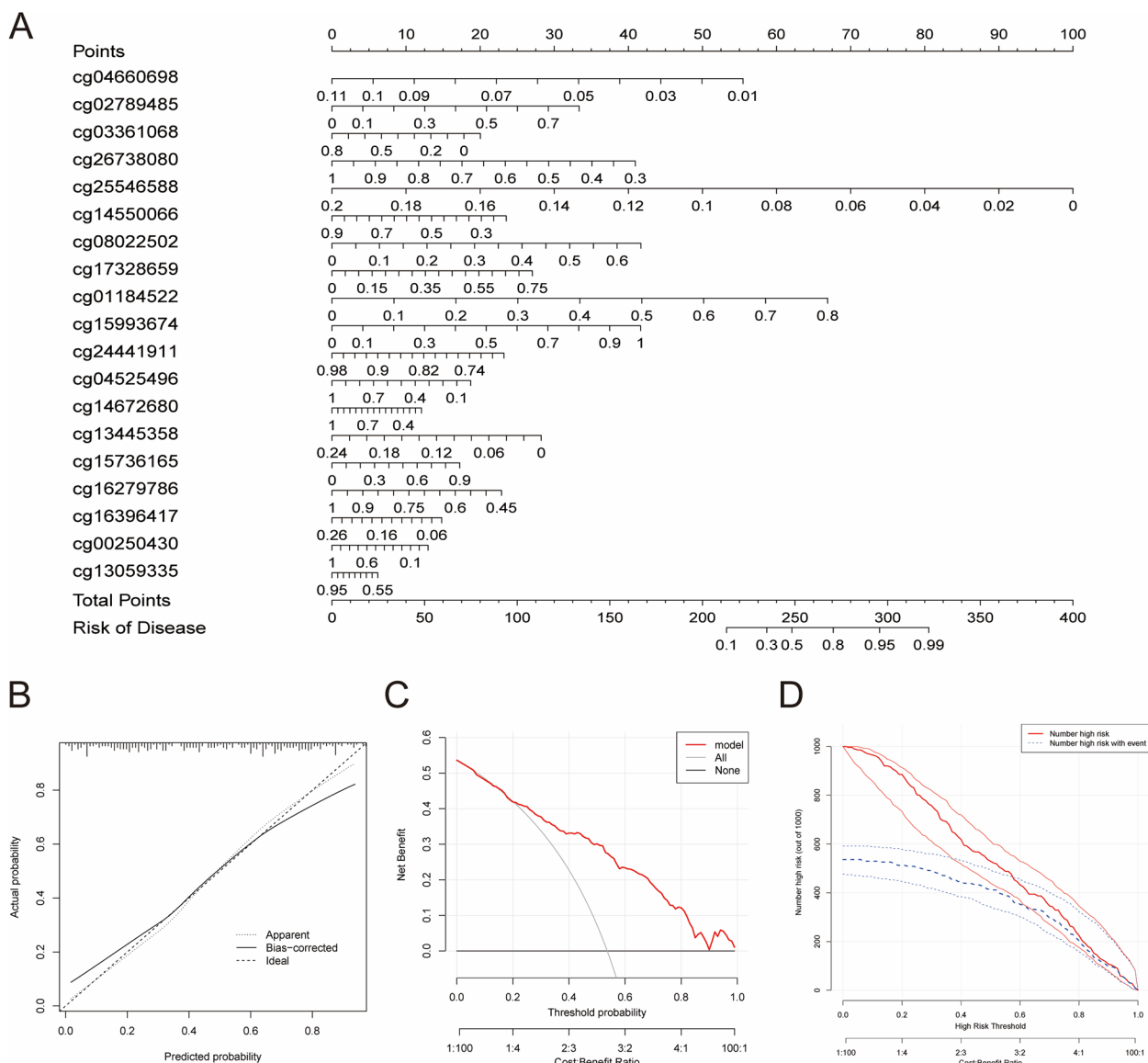


**Fig. 3** Construction of diagnostic model for predicting CRC metastasis. **A** Circos graph for univariate Cox regression analysis, which represents the correlation of DMCs  $\beta$  value and prognosis of CRC (red: up-regulated methylation CpG Site, grey: down-regulated methylation CpG Site, purple: risk factor; green: favorable factor) in the TCGA-COAD ( $p$ -values for Cox test:  $1e-04$  to 1). **B** Line plot of 10-fold cross-validation of the SVM-RFE algorithm for feature selection. ROC curves of diagnostic model in the TCGA cohort (**C**) and GEO cohort (**D**)

with high accuracy. DCA analysis showed that compared with the all or no patient intervention scheme, CRC patients would benefit more from using this nomogram to predict the probability of metastasis at a high risk threshold (0–1), which indicated that the clinical application of 19-DMCs nomograms has a higher impact (Fig. 4C). The clinical impact curve showed that the “Number high risk” curve and the “Number high risk with event” curve were close to 1 from 0.3, which

revealed that the nomogram had a good predictive ability (Fig. 4D). The results suggested that these 19 DMCs may play an essential role in the metastasis of CRC.

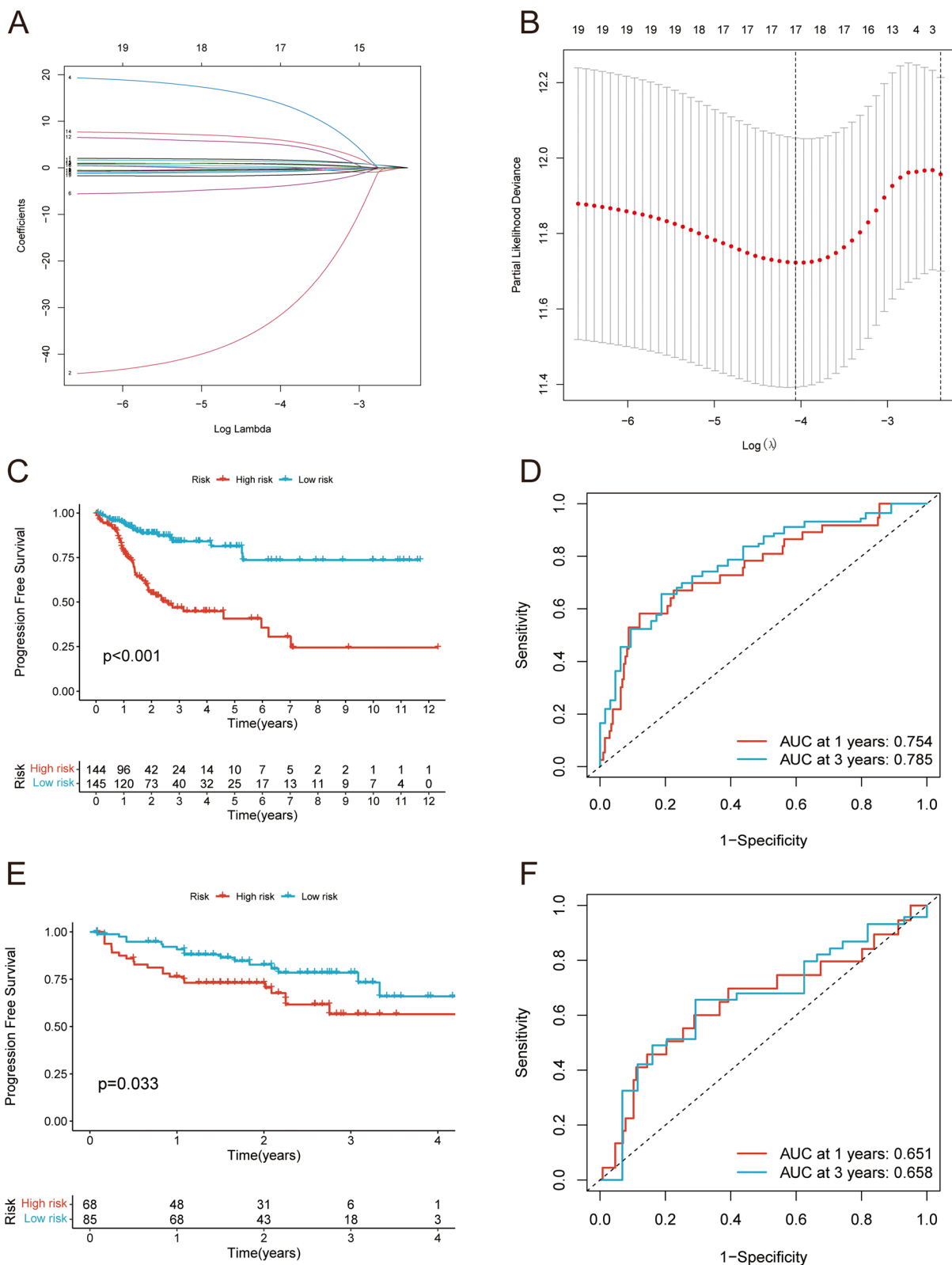
**Identification of PFS related methylation signatures in CRC**  
LASSO-Cox regression analysis identified methylation characteristics associated with PFS in CRC patients from 19 DMCs. 10-fold cross-validation was conducted for the model construction. When the number of methylation



**Fig. 4** Construction of a nomogram model for predicting CRC metastasis based on the TCGA cohort. **A** Nomogram to predict the metastasis of CRC. **B** Calibration curve to assess the predictive power of the nomogram model. **C** DCA curve and clinical impact curve (**D**) to evaluate the clinical value of the nomogram model

characteristics was 17,  $\lambda$  was the smallest, and the model was optimal (Fig. 5A-B). Therefore, these 17 DMCs as candidate methylation characteristics related to PFS in CRC patients. Based on the  $\beta$  values of these 17 DMCs and the corresponding regression coefficients, the following prognostic features were constructed:  $\text{riskScore} = (\text{cg}02789485 * 1.692) + (\text{cg}04660698 * - 32.388) + (\text{cg}00250430 * - 0.456) + (\text{cg}16,396,417 * 14.126) + (\text{cg}14,550,066 * - 0.655) + (\text{cg}24,441,911 * - 4.021) + (\text{cg}26,738,080 * - 0.403) + (\text{cg}04525496 * - 0.255) + (\text{cg}03361068 * - 0.834) + (\text{cg}01184522 * 1.356) + (\text{cg}25,546,588 * 5.063) +$

$(\text{cg}17,328,659 * 0.839) + (\text{cg}13,445,358 * 6.063) + (\text{cg}15,993,674 * 0.709) + (\text{cg}08022502 * 0.242) + (\text{cg}14,672,680 * - 0.269) + (\text{cg}13,059,335 * - 1.596)$ . CRC patients in the TCGA-COAD 450 K cohort (training dataset) and TCGA-COAD 27 K cohort (testing dataset) were divided into high-risk and low-risk groups based on the median risk score of methylation signatures. The KM analysis revealed that the PFS of patients in the high-risk group was significantly lower than that in the low-risk group in both the training dataset (Fig. 5C; Table 1) and the testing dataset (Fig. 5E). The ROC curve showed that the ROC



**Fig. 5** Establishment of a 17-DMCs signature. **A** and **B** LASSO Cox regression (with minimized lambda) of the DMCs. Survival curve showing different PFS between high- and low-risk groups in training dataset (**C**) and testing dataset (**E**). 1-year and 3-year ROC curves for training dataset (**D**) and testing datasets (**F**)



**Table 1** Clinical information of TCGA-COAD 450 K datasets

Characteristics	High-risk	Low-risk
n	144	145
Gender, n (%)		
Female	65 (22.5%)	68 (23.5%)
Male	79 (27.3%)	77 (26.6%)
Stage, n (%)		
Stage I	16 (5.5%)	27 (9.3%)
Stage III	56 (19.4%)	28 (9.7%)
Stage II	40 (13.8%)	72 (24.9%)
Stage IV	27 (9.3%)	13 (4.5%)
Unknow	5 (1.7%)	5 (1.7%)
T, n (%)		
T2	17 (5.9%)	25 (8.7%)
T4	28 (9.7%)	10 (3.5%)
T3	95 (32.9%)	106 (36.7%)
T1	3 (1%)	4 (1.4%)
Unknow	1 (0.3%)	0 (0%)
M, n (%)		
M0	87 (30.1%)	110 (38.1%)
MX	27 (9.3%)	20 (6.9%)
M1	27 (9.3%)	13 (4.5%)
Unknow	3 (1%)	2 (0.7%)
N, n (%)		
N0	63 (21.8%)	105 (36.3%)
N1	50 (17.3%)	23 (8%)
N2	31 (10.7%)	17 (5.9%)

AUC values of the prognostic model based on 17 DMCs for predicting 1-year and 3-year PFS of CRC patients in the training dataset were 0.754 and 0.785, respectively (Fig. 5D). The 1-year and 3-year ROC AUC values of the testing dataset were 0.651 and 0.658, respectively (Fig. 5F), which further verified the high accuracy of the prognostic model in predicting PFS of CRC patients. KM analysis showed that there was a significant difference in PFS between CRC patients in the high and low expression groups of 17DMCs (Fig. 6A-Q). All in all, the prognostic model based on 17 DMCs had good predictive value.

Then, we explored the correlation between the prognostic models and pathological factors in the training dataset. In Fig. 7A, Age ( $p < 0.01$ ), Stage ( $p < 0.001$ ), T ( $p < 0.05$ ), M ( $p < 0.05$ ), and N ( $p < 0.001$ ) were statistically different between the two risk subgroups. Among them, the Stage and N stages of CRC patients in the high-risk group were higher than those in the low-risk group. In addition, as the risk score increased, the Stage, T, M, and N stages also increased, implying that prognostic features may be associated with tumor enlargement and metastasis (Fig. 7B-D). In addition, we plotted the ROC curves of

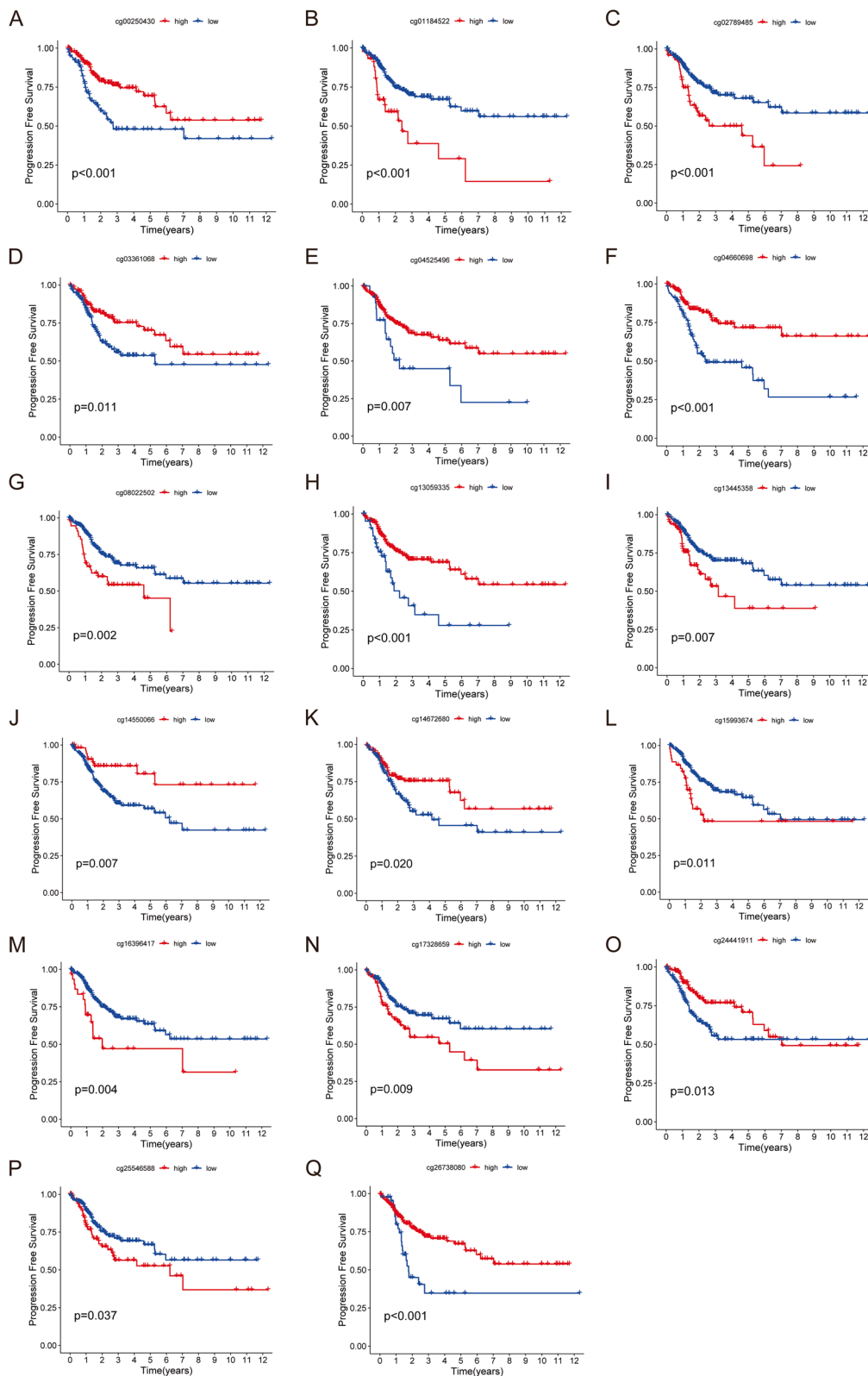
the prognostic model and clinicopathological features to predict the 1-, 3-, and 5-year PFS of CRC patients. The results showed that the prognostic model based on 17 DMCs had good prediction performance compared with clinicopathological features (Fig. 7F-H).

Furthermore, we explored the expression at the protein level of genes in which prognostic related DMCs are located. From the HPA database, we searched for immunohistochemical maps of 10 genes associated with prognostic DMCs. Immunohistochemical analysis showed that the protein expression of the genes involved in prognostic DMCs was different between normal and diseased tissues. (Fig. 8A-J). Compared with normal tissues, PAGR1 (cg04660698) (staining: Low, intensity: Weak), VCAN (cg04525496) (staining: Low, intensity: Moderate), IL15 (cg25546588) (staining: Low, intensity: Weak), DMRT2 (cg00250430) (staining: Low, intensity: Weak) and STUB1 (cg17328659) (staining: Medium, intensity: Moderate) were lower in CRC, while MZF1 (cg16396417) (staining: Medium, intensity: Moderate), RBP5 (cg24441911) (staining: High, intensity: Strong), ESPL1 (cg13445358) (staining: High, intensity: Strong), UNC45A (cg08022502) (staining: High, intensity: Strong) and TNNI2 (cg14672680) (staining: Medium, intensity: Moderate) were highly expressed in CRC. Immunohistochemical information for prognostic genes can be found in Supplementary Material 5. In addition, we analyzed the methylation levels of genes (GLIPR1L2, PAGR1, MZF1, NCR1, RBP5, TNNC1, VCAN, TRIM9, ZNF496, IL15, DMRT2, STUB1, ESPL1, PRNP, UNC45A, TNNI2, ADAMDEC1) 17 prognostic related methylation sites. We input the above 17 genes into the Ualcan database (<https://ualcan.path.uab.edu/>) to obtain their methylation level changes across different stages (Stage I-IV). The results showed that the 17 genes mentioned above showed significant differences between different stages (Supplementary Material 6), indicating a close correlation between the methylation levels of the 17 genes and disease progression in CRC. This further confirmed that the 17 prognostic related methylation sites may play an important role in CRC by affecting gene transcription or expression.

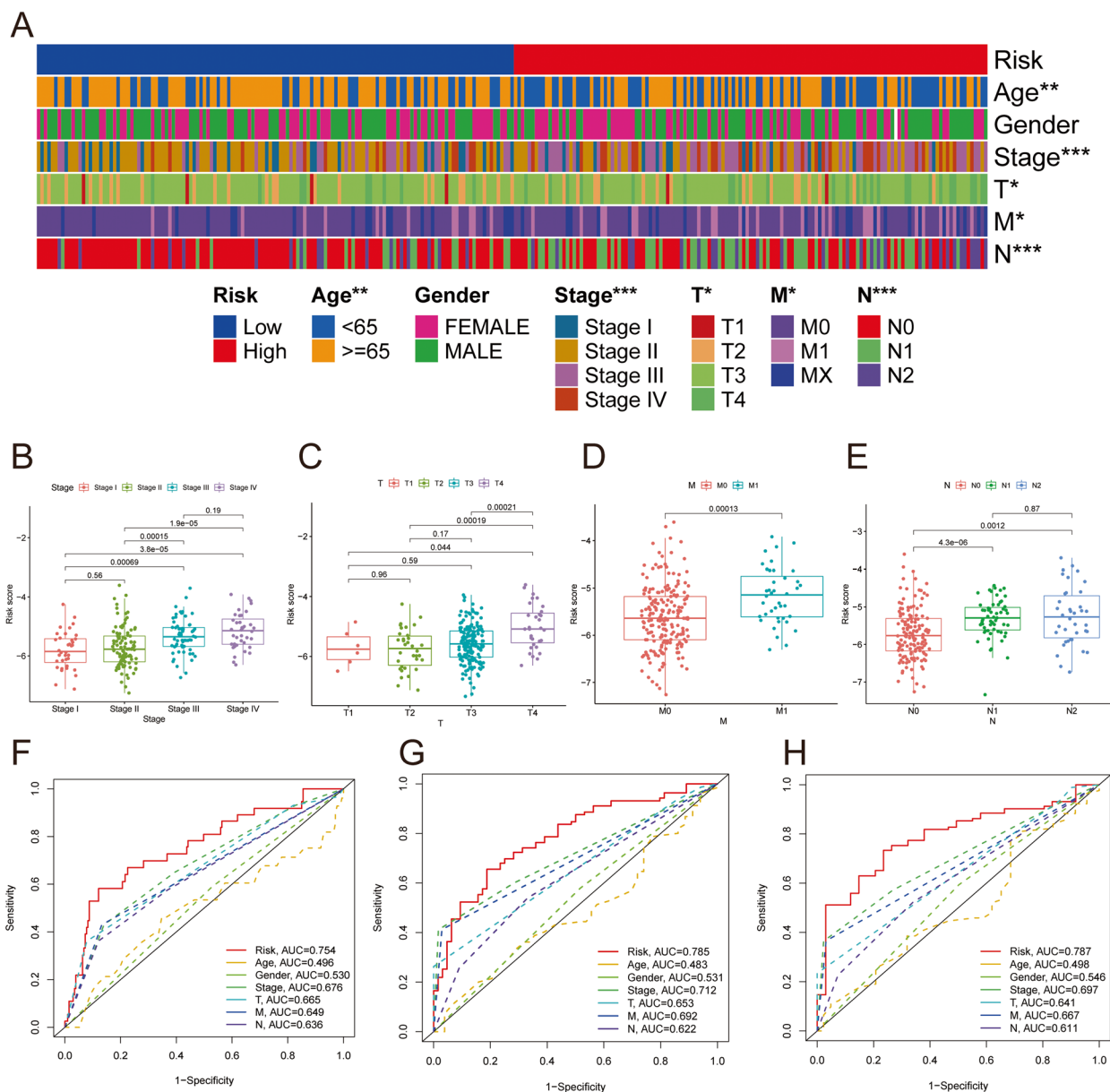
The above results further confirmed that prognosis related DMCs may play an important role in the pathological progression of CRC.

#### Nomogram model related to PFS of CRC

Univariate Cox and multivariate Cox regression analyses were performed on risk score and clinical characteristics in the training dataset to screen features that could independently predict PFS in CRC. Figure 9A and B (Supplementary Material 7) show that risk score, T, and M were independent predictors of PFS for CRC.



**Fig. 6** KM curve of 17-DMCs signature. **A-Q** cg00250430, cg01184522, cg02789485, cg03361068, cg04525496, cg04660698, cg08022502, cg13059335, cg13445358, cg14550066, cg14672680, cg15993674, cg16396417, cg17328659, cg24441911, cg25546588, and cg26738080



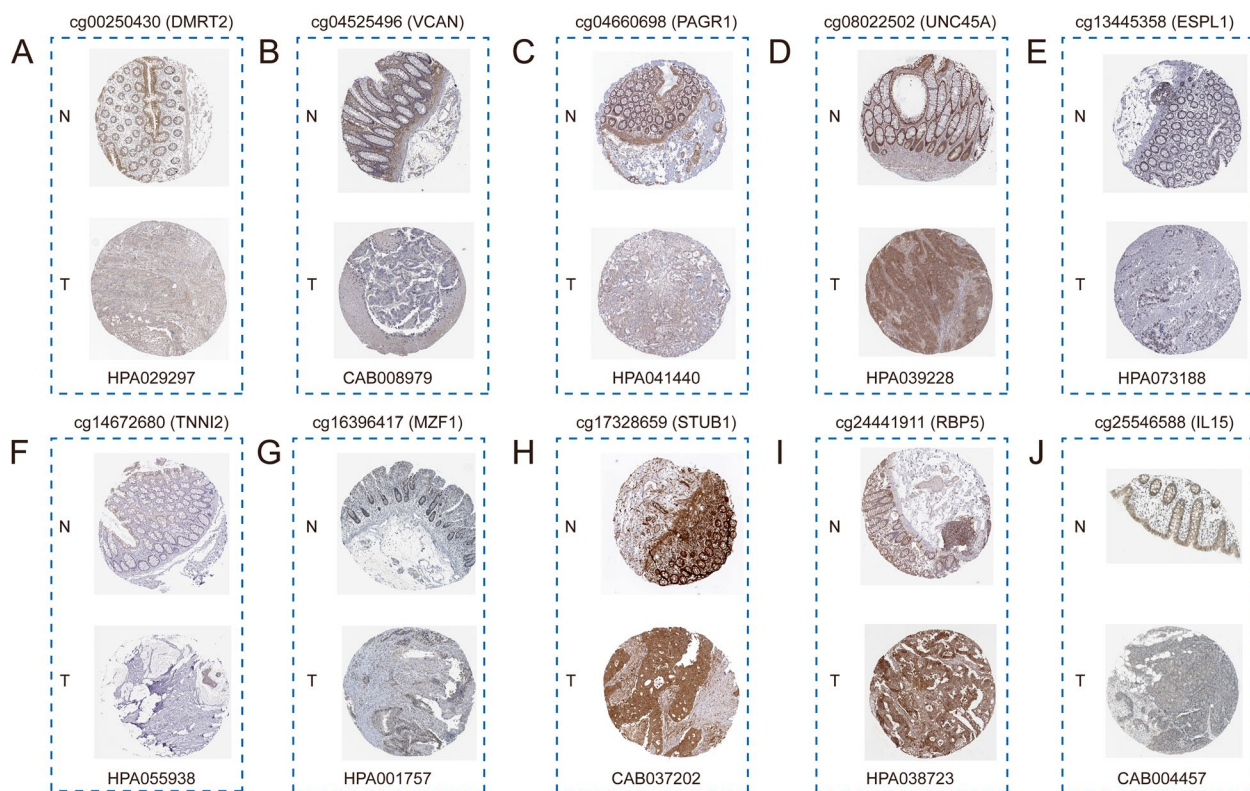
**Fig. 7** Correlation analysis between risk score and clinical indicators. **A** Heatmap for the connections between clinical indicators and the risk groups. The difference of risk score between different groups stratified by Age (**B**), Stage (**C**), T stage (**D**), and M stage (**E**). 1-year (**F**), 3-year (**G**) and 5-year (**H**) ROC curves of risk score and clinicopathological characteristics

Then, we constructed a nomogram model based on the risk score, T and M, to predict the PFS of CRC patients at 1-, 3-, and 5 years (Fig. 9C). To estimate the predictive performance of the nomogram, we plotted the ROC curve and calibration curve of the nomogram model. The ROC curve revealed that the nomogram model had high accuracy in predicting the PFS of CRC patients at 1 year (AUC=0.775), 3 years (AUC=0.825), and 5 years (AUC=0.809) (Fig. 9D). In addition, the calibration curve showed that the prediction results of the nomogram

model at 1 year and 5 years were in good agreement with the actual observed values (Fig. 9E-G).

**Immune differences between risk subgroups**

According to the results of ssGSEA, we found that the scores of immune cells and immune function were generally higher in the high-risk group than in the low-risk group. Figure 10A showed that the scores of B\_cells, DCs (Dendritic Cells), iDCs (Immature Dendritic Cells), Macrophages, Mast\_cells, Neutrophils, pDCs



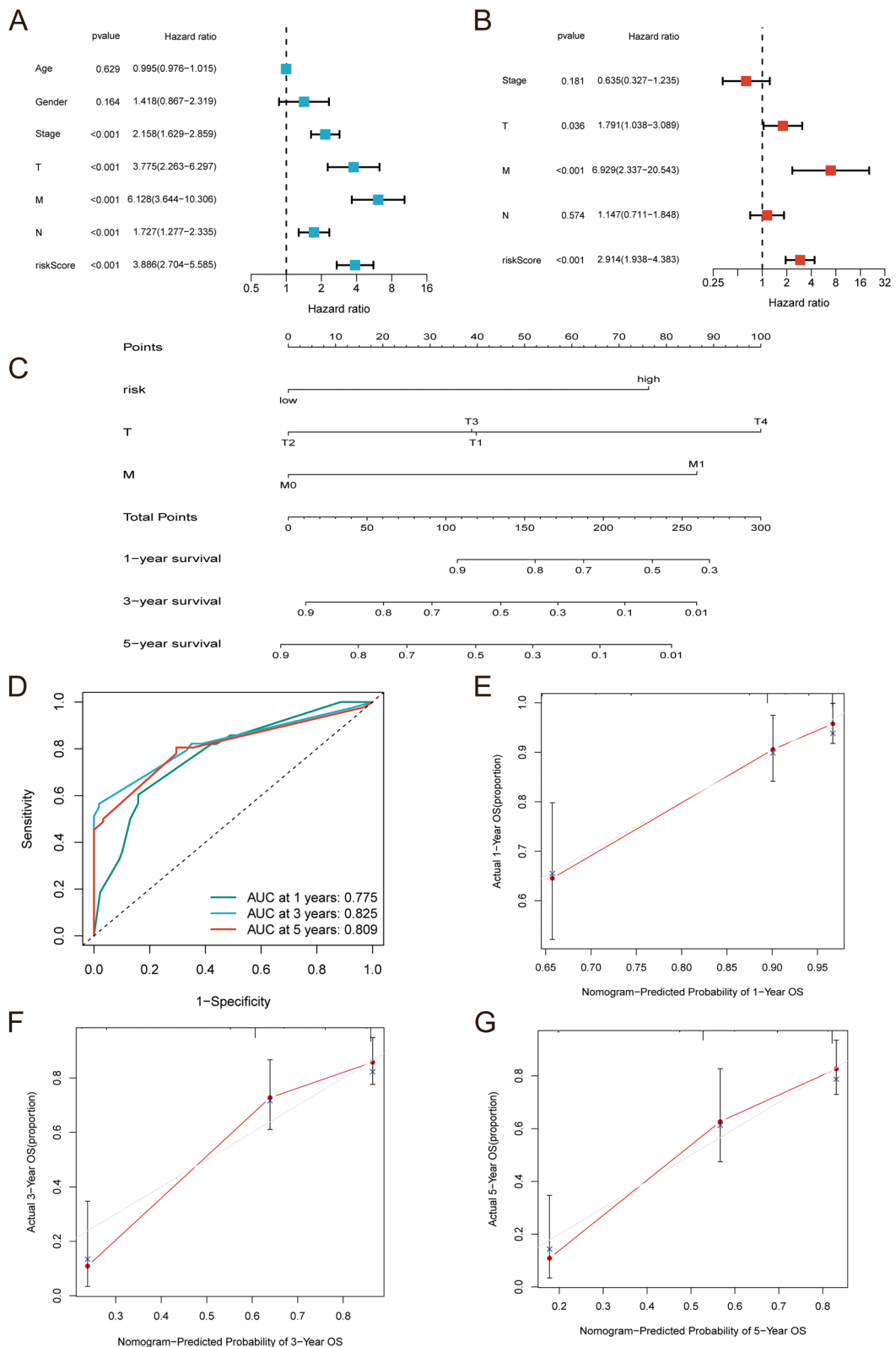
**Fig. 8** Immunohistochemistry analysis of genes where the prognostic DMCs was located. **A** cg00250430 (DMRT2). **B** cg04525496 (VCAN). **C** cg04660698 (PAGR1). **D** cg08022502 (UNC45A). **E** cg13445358 (ESPL1). **F** cg14672680 (TNNI2). **G** cg16396417 (MZF1). **H** cg17328659 (STUB1). **I** cg24441911 (RBP5). **J** cg25546588 (IL15). N represents normal cells, T represents tumor cells. Antibodies were marked below the box

(Plasmacytoid Dendritic Cells), T\_helper\_cells, Tumor Infiltrating Lymphocytes (TIL) and Regulatory T cells (Treg) were significantly different between the high-risk and low-risk groups. Figure 10B showed that the scores of APC\_co\_stimulation, CCR, HLA, T\_cell\_co-stimulation, and Type\_II\_IFN\_Reponse were significantly different between the two risk subgroups. The risk score was highly correlated with B\_cells, Macrophages, Mast\_cells, Neutrophils, pDCs, and Type\_II\_IFN\_Reponse. Next, we performed immune correlation analysis on DNA methylation data in the TCGA-COAD cohort using HiTIMED. The results showed that there was a significant difference in the immune scores of CD8nv and DC between the risk groups (Supplementary Material 8). Moreover, as the risk score increased, so did the score for immune cells or pathways (Fig. 10C-H). TIDE score can be used to screen patients suitable for immunotherapy. The TIDE score, the proportion of MSI, and the proportion of MSI-L in the high-risk group were significantly higher than those in the low-risk group (Fig. 10I-J). In addition, with the increase of risk score, MSI-L was significantly higher than MSI-H (Fig. 10K,  $p=0.021$ ). These results suggested that the prognosis of patients in the high-risk group and the response to immunotherapy are worse,

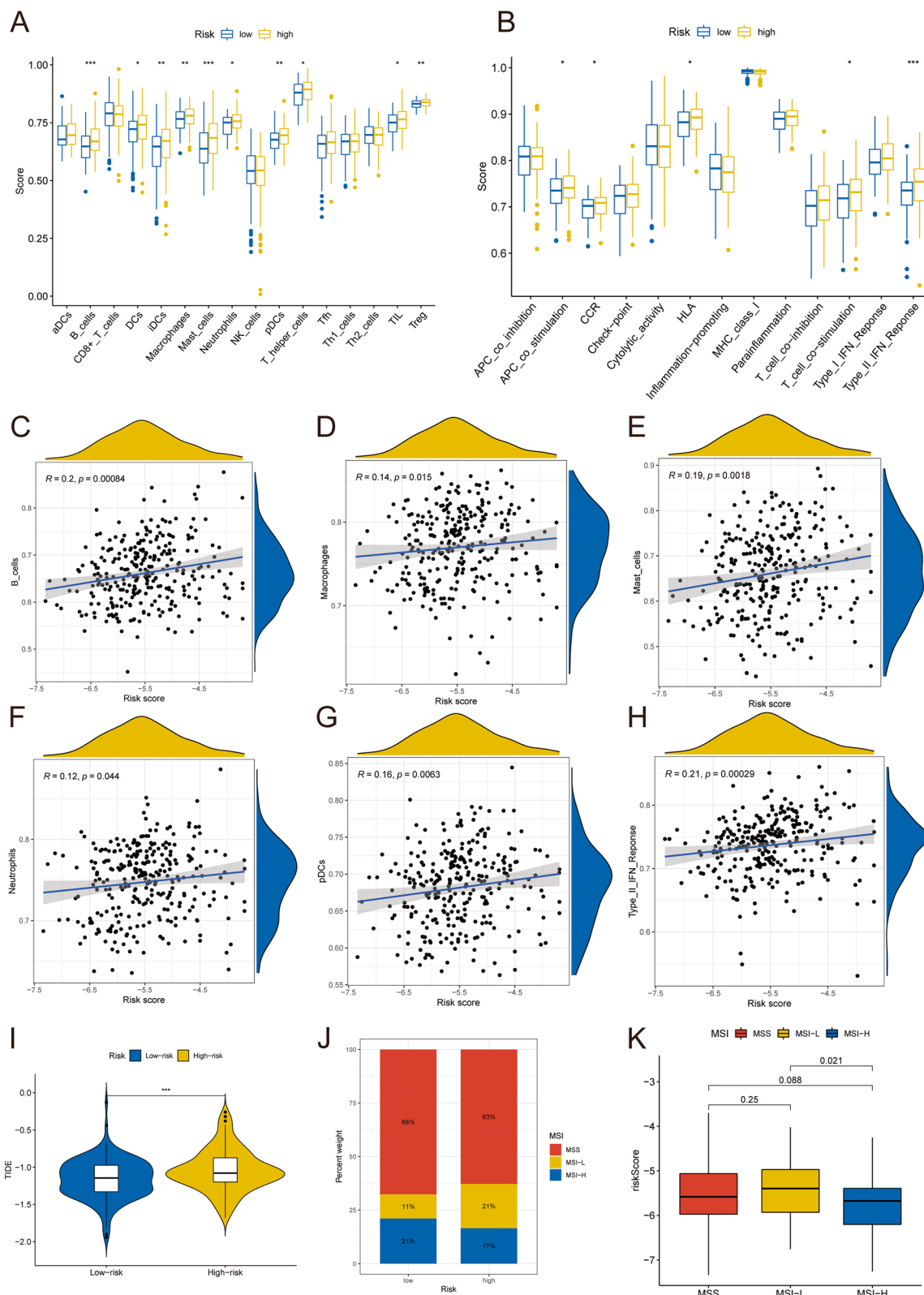
and immunotherapy helps to improve the prognosis of patients in the low-risk group.

#### Differences in TMB levels between risk subgroups

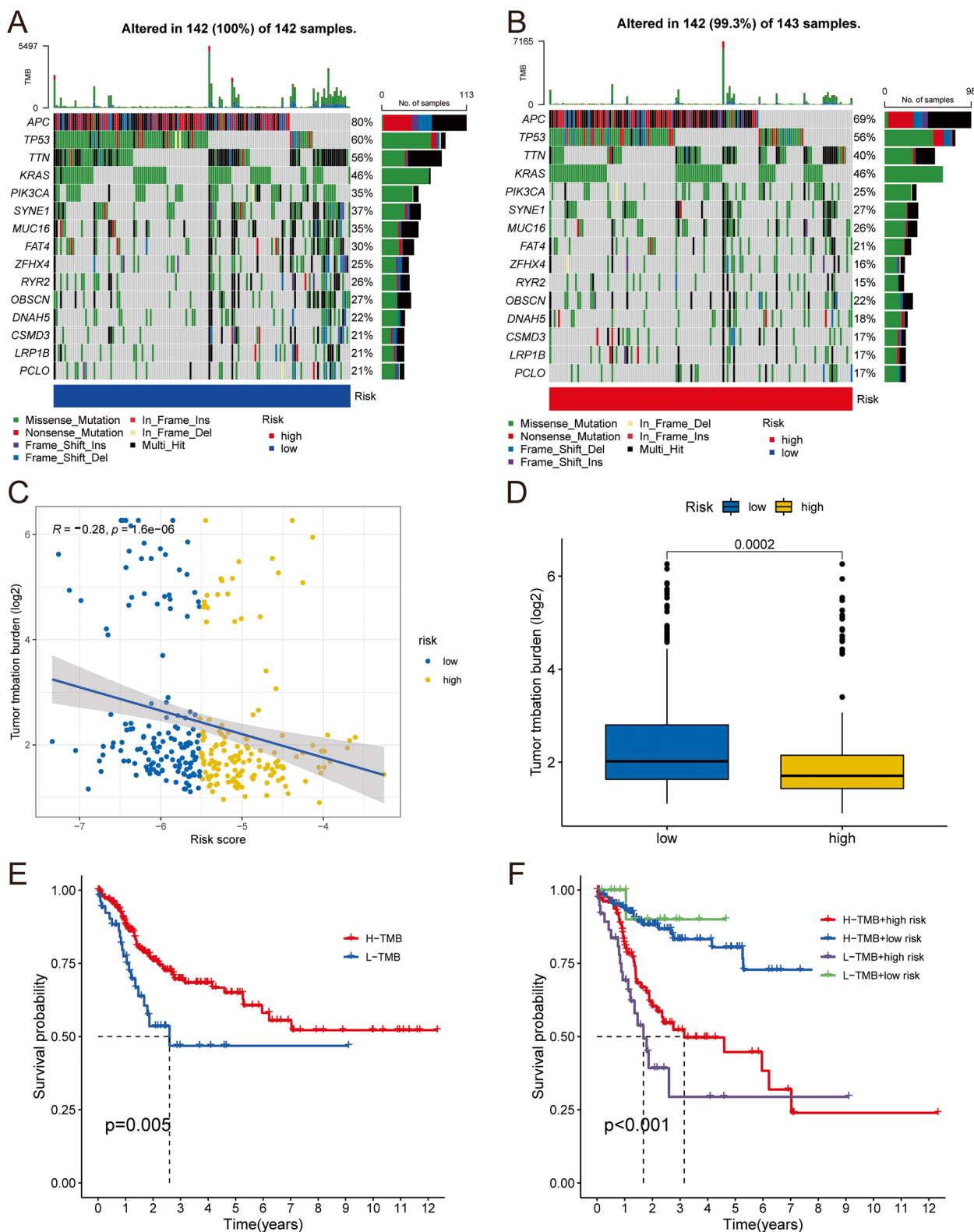
TMB levels are highly correlated with the prognosis of cancer patients. Therefore, we investigated the TMB differences between the high- and low-risk groups in the training dataset. Figure 11A and B revealed that the mutation rate of patients in the high-risk group was slightly lower than that in the low-risk group. Risk scores were negatively correlated with TMB levels (Fig. 11C), and the TMB level ( $\log_2$ ) of the low-risk group was significantly higher than that of the high-risk group ( $p=0.0002$ , Fig. 11D). KM analysis results revealed that patients in the H-TMB group had higher PFS than those in L-TMB (Fig. 11E), and patients with low TMB in the high-risk group had the worst PFS (Fig. 11F). Studies have found that patients with high TMB levels respond better to immunotherapy, suggesting that immunotherapy is less effective in improving outcomes in CRC patients. These results further confirmed that patients in the high-risk group have a poorer prognosis and sensitivity to immunotherapy than those in the low-risk group, and patients in the low-risk group are more suitable for immunotherapy.



**Fig. 9** Construction of nomograph model. The forest plot of univariate Cox analysis (**A**) and multivariate Cox analysis (**B**). **C** Nomogram to predict the 1-year, 3-year, and 5-year PFS of CRC patients. **D** ROC curve to assess the predictive power of the nomogram model. 1-year (**E**), 3-year (**F**) and 5-year (**G**) calibration curves of nomograph models



**Fig. 10** Comparison of immune microenvironment between two risk subgroups. Differences in the abundance of immune cells (**A**) and immune function (**B**) among risk subgroups. **C-H** Correlation between risk score and B\_cells, Macrophages, Mast\_cells, Neutrophils, pDCs, Type\_IL\_IFN\_Response respectively. **I** Differences in TIDE scores between high-risk and low-risk groups. **J** The proportion of MSS and MSI (MSI-H and MSI-L) in risk subgroups. **K** The difference of risk score among MSS subtype, MSI-H subtype and MSI-L subtype



**Fig. 11** Landscape of TMB between risk subgroups. Waterfall showing the top 20 mutated genes in low-(**A**) and high-risk group (**B**). **C** Correlation between risk score and TMB level. **D** Differences in TMB levels between risk subgroups. **E** Survival curve showing different PFS between high- and low-TMB groups in training dataset. **F** Survival curve of the comprehensive analyses between TMB and risk score

## Discussion

Abnormal DNA methylation can not only be used as a target for cancer therapy but also for disease diagnosis and prognosis prediction [14, 15]. In addition, changes in DNA methylation mainly occur in the early stage of carcinogenesis and cancer progression, which makes DNA methylation characteristics conducive to the early prevention and diagnosis of disease. Metastasis and recurrence are the key factors affecting the treatment and survival of CRC patients. DNA methylation features have great potential as biomarkers related to CRC progression and prognosis. Therefore, our study aimed to elucidate the landscape of DNA methylation during CRC metastasis and assess its clinical guiding value in CRC.

Firstly, we identified DMCs between CRC patients with and without metastasis. Most DMCs were found to have increased  $\beta$  values in CRC patients with metastasis compared to CRC patients without metastasis. It has been reported that many tumor suppressor genes are partially or wholly silenced due to hypermethylation of their CpG sites [16], suggesting that the hypermethylation level of CpG sites may play a vital role in the occurrence and progression of CRC metastasis. The results of enrichment analysis showed that the genes where DMCs are located were mainly enriched in Neuroactive ligand-receptor interaction, cAMP signaling pathway, transcription by RNA polymerase II, cognition, and cell-cell adhesion. In the process of cancer BMS, Neuroactive ligand-receptor interaction plays a prominent role in adapting to the environment of target organs [17]. cAMP signaling pathway can regulate the ability of colon cancer metastasis [18]. Guo et al. found that Tetraspanin CO-029 participated in cancer metastasis in the digestive system by regulating cell-cell adhesion [19]. The biological pathways obtained from the enrichment of genes where DMCs are located related to the progression of CRC, suggesting that DMCs may also play an important role in CRC.

In order to evaluate the prognostic ability of DMCs, univariate Cox regression analysis and KM test were performed for DMCs, and a total of 20 candidate DMCs related to PFS in CRC patients were screened. Based on the PFS-related candidate DMCs, 19 DMCs were selected by the SVM-RFE algorithm for the construction of the diagnostic model and nomogram model. ROC analysis showed that the 19-DMCs diagnostic model could predict the metastasis of CRC patients with high accuracy. The calibration curve, DCA analysis, and clinical impact curve showed that the 19-DMCs nomogram model had good clinical guiding value in predicting CRC metastasis. These results indicated that 19 DMCs may play a key role in CRC metastasis. Based on 19 DMCs, LASSO-Cox regression analysis was utilized to construct 17 DMCs (cg02789485, cg04660698, cg00250430, cg16396417, cg14550066, cg24441911, cg26738080,

cg04525496, cg03361068, cg01184522, cg25546588, cg17328659, cg13445358, cg15993674, cg08022502, cg14672680, and cg13059335). The ROC curve showed that the prognostic model could predict 1-year and 3-year PFS of CRC patients with medium-high accuracy. Compared with the actual observed values, the nomogram model can better predict the 1-year and 5-year PFS of CRC patients.

The alteration of CpG sites has an important effect on gene expression. We annotated 17 prognostic DMCs genes: cg02789485 (GLIPR1L2), cg04660698 (PAGR1), cg16396417 (MZF1), cg14550066 (NCR1), cg24441911 (RBP5), cg26738080 (TNNC1), cg04525496 (VCAN), cg03361068 (TRIM9), cg01184522 (ZNF496), cg25546588 (IL15), cg00250430 (DMRT2), cg17328659 (STUB1), cg13445358 (ESPL1), cg15993674 (PRNP), cg08022502 (UNC45A), cg14672680 (TNNI2), cg13059335 (ADAMDEC1). GLIPR1L2 is a novel target gene of tumor suppressor gene p53, and p53 mutations occur in 60% of CRC [20, 21]. PAGR1 can affect fat generation by regulating C/EBP- $\beta$  and C/EBP- $\delta$  [22]. Multiple studies have shown that fat is one of the risk factors for CRC [23]. Studies have found that MZF1 can promote the proliferation of tumor cells in CRC and inhibit cancer progression through apoptosis [24]. In mouse models, NCR1-mediated IFN- $\gamma$  production leads to increased expression of FN1 in tumors, thereby altering the primary tumor structure and reducing tumor metastasis [25]. Wan et al. found that deletion of RBP5-mediated protein delayed tumor progression in a mouse model of cholangiocarcinoma [26]. Studies have shown that TNNC1 is a promising biomarker for metastasis of ovarian and tongue cancers [27, 28]. The upregulation of VCAN mediated by INHBA can promote the migration and proliferation of cancer cells in CRC [29]. Cui et al. found that IFN- $\gamma$  produced by NK cells could affect the proliferation of CRC cells through the regulation of IL-15 [30]. In some CRC samples, the proportion of CHIP (aka STUB1) increased and was associated with poor survival [31]. Studies have found that, compared with normal samples, PRNP expression is up-regulated in CRC and is an independent prognostic factor for 3-year survival of CRC [32]. A large amount of evidence has revealed that the expression of UNC45A in cancer cells is related to the proliferation and metastasis of solid tumors [33]. Through comprehensive biogenic analysis, Li et al. found that TNNI2 was identified as a prognostic biomarker for CRC. In addition, the model based on 7 genes including TNNI2 can predict CRC metastasis (liver or lung), with AUC of 0.933 [34]. Macartney-Coxson et al. identified ADAMDEC1 as a candidate gene associated with CRC liver metastasis, whose mRNA and protein expression decreased during the occurrence and progression of CRC [35]. In summary, 17 prognostic-related DMCs genes



play an important role in the occurrence and progression of CRC, which further verifies that 17 prognostic-related DMCs may also play a key role in CRC.

DNA methylation is essential for the interaction between tumors and immune cells [36]. In this study, we found that the scores of immune cells and immune-related pathways in the high-risk group were generally higher than those in the low-risk group. Multiple immune cells or immune-related pathways, such as Macrophages, Mast\_cells, pDCs, and Type\_II\_IFN\_Reponse, have significant differences between the two risk subgroups. In addition, as the risk score increased, so did the score of immune cells or immune-related pathways. The interaction between the extracellular trap of macrophages and colon cancer cells promotes CRC invasion [37]. Neutrophils and their releases have been associated with the progression and metastasis of various cancers, and targeting neutrophils in extracellular traps may be an effective strategy to inhibit the metastasis of colorectal cancer [38]. Type\_II\_IFN\_Reponse is involved in cancer's immune response regulation mechanism and is related to the growth and migration ability of cancer cells [39]. We also found that TIDE scores were higher in high-risk patients than in low-risk patients. Furthermore, patients with high TMB had higher PFS than those with low TMB, and TMB levels decreased in CRC patients as risk scores increased. These results indicated that the low-risk patients had better immunotherapy response.

There are some limitations to this study. We need more CRC 450 K DNA methylation array data to validate the DNA methylation-related prognostic model. In addition, the treatment methods of CRC patients have some potential impacts on their prognosis, including whether the treatment methods are effective for patients, the interaction effects between multiple treatment methods, genetic variations of patients, and their sensitivity to treatment methods. Therefore, there is an urgent need for more precise personalized treatment to help patients choose more effective treatment methods and effectively improve their prognosis.

## Conclusion

In short, we identified DNA methylation biomarkers associated with CRC metastasis and constructed a 19 DMCs correlation diagnostic model and a line graph model that can accurately predict CRC metastasis. We also identified DNA methylation biomarkers associated with PFS in CRC patients and constructed a 17 DMCS-related prognostic model that predicted PFS in CRC patients with moderate to high accuracy. In addition, we elucidate the association of TME, MSI, and TMB with DNA methylation-related prognostic models. Our results may provide new biomarkers for predicting metastasis in CRC patients and PFS.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-024-01898-4>.

Supplementary Material 1.  
Supplementary Material 2.  
Supplementary Material 3.  
Supplementary Material 4.  
Supplementary Material 5.  
Supplementary Material 6.  
Supplementary Material 7.  
Supplementary Material 8.

## Acknowledgements

We acknowledge the TCGA, GEO and TIMER databases for free use.

## Research involving human participants and/or animals

Not applicable.

## Disclosure of potential conflicts of interest

Not applicable.

## Authors' contributions

Fang Qian wrote the main manuscript text, Kai Wei prepared Fig. 1, Fang Qian and Li Qiang prepared Figs. 2 and 3, Fang Qian and Huidan Chang prepared Figs. 4 and 5, Fang Qian prepared Figs. 6, 7, 8, 9, 10 and 11. Xiaoyi Chen, Tao Huang, and Yixue Li checked the manuscript. All authors reviewed the manuscript.

## Funding

This work was supported by National Key R&D Program of China (2018YFA0900700, 2022YFF1203202), Self-supporting Program of Guangzhou Laboratory (SRPG22-007), R&D Program of Guangzhou National Laboratory, Grant No. GZNL2024A01002, Science and Technology Project of Yunnan Province (202103AQ100002), II Phase External Project of Guoke Ningbo Life Science and Health Industry Research Institute, University of Chinese Academy of Sciences (2020YJY0217), Strategic Priority Research Program of Chinese Academy of Sciences (XDB38050200, XDB38040202, XDA26040304).

## Availability of data and materials

All data can be obtained in TCGA and GEO databases.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors consent to the publication of this study.

### Competing interests

The authors declare no competing interests.

Received: 22 September 2023 Accepted: 30 April 2024

Published online: 10 May 2024

## References

- Jia SN, Han YB, Yang R, Yang ZC. Chemokines in colon cancer progression. *Semin Cancer Biol.* 2022;86(Pt 3):400–7. <https://doi.org/10.1016/j.semcan.2022.02.007>. Epub 2022 Feb 7. PMID: 35183412.
- Zeng X, Ward SE, Zhou J, Cheng ASL. Liver Immune Microenvironment and Metastasis from Colorectal Cancer-Pathogenesis and therapeutic

- perspectives. *Cancers (Basel)*. 2021;13(10):2418. <https://doi.org/10.3390/cancers13102418>. PMID: 34067719; PMCID: PMC8156220.
3. Tauriello DV, Calon A, Lonardo E, Batlle E. Determinants of metastatic competency in colorectal cancer. *Mol Oncol*. 2017;11(1):97–119. <https://doi.org/10.1002/1878-0261.12018>. Epub 2017 Jan 3. PMID: 28085225; PMCID: PMC5423222.
  4. Li X, Zhang Q, Zhao L, Jiang L, Qi A, Wei Q, Song X, Wang L, Zhang L, Zhao Y, Lv X, Wei M, Zhao L. A combined four-mRNA signature Associated with Lymphatic metastasis for prognosis of Colorectal Cancer. *J Cancer*. 2020;11(8):2139–49. <https://doi.org/10.7150/jca.38796>. PMID: 32127941; PMCID: PMC7052913.
  5. Sui J, Li YH, Zhang YQ, Li CY, Shen X, Yao WZ, Peng H, Hong WW, Yin LH, Pu YP, Liang GY. Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *Int J Oncol*. 2016;49(5):2023–36. <https://doi.org/10.3892/ijo.2016.3716>. Epub 2016 Sep 30. PMID: 27826625.
  6. Yang Z, Jones A, Widschwendter M, Teschendorff AE. An integrative pan-cancer-wide analysis of epigenetic enzymes reveals universal patterns of epigenomic deregulation in cancer. *Genome Biol*. 2015;16(1):140. <https://doi.org/10.1186/s13059-015-0699-9>. PMID: 26169266; PMCID: PMC4501092.
  7. Casalino L, Verde P. Multifaceted roles of DNA methylation in Neoplastic Transformation, from Tumor suppressors to EMT and metastasis. *Genes (Basel)*. 2020;11(8):922. <https://doi.org/10.3390/genes11080922>. PMID: 32806509; PMCID: PMC7463745.
  8. Gutierrez A, Demond H, Brebi P, Ili CG. Novel methylation biomarkers for Colorectal Cancer Prognosis. *Biomolecules*. 2021;11(11):1722. <https://doi.org/10.3390/biom11111722>. PMID: 34827720; PMCID: PMC8615818.
  9. Coppedè F, Lopomo A, Spisni R, Migliore L. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *World J Gastroenterol*. 2014;20(4):943–56. <https://doi.org/10.3748/wjg.v20.i4.943>. PMID: 24574767; PMCID: PMC3921546.
  10. Jin S, Zhu D, Shao F, Chen S, Guo Y, Li K, Wang Y, Ding R, Gao L, Ma W, Lu T, Li D, Zhang Z, Cai S, Liang X, Song H, Ji L, Li J, Zheng Z, Jiang F, Wu X, Luan J, Zhang H, Yang Z, Cantor CR, Xu C, Ding C. Efficient detection and post-surgical monitoring of colon cancer with a multi-marker DNA methylation liquid biopsy. *Proc Natl Acad Sci U S A*. 2021;118(5):e2017421118. <https://doi.org/10.1073/pnas.2017421118>. PMID: 33495330; PMCID: PMC7865146.
  11. Luo H, Ye M, Hu Y, Wu M, Cheng M, Zhu X, Huang K. DNA methylation regulator-mediated modification patterns and tumor microenvironment characterization in glioma. *Aging*. 2022;14(19):7824–50. <https://doi.org/10.18632/aging.204291>. Epub 2022 Sep 21. PMID: 36152044; PMCID: PMC9596205.
  12. Xu B, Lu M, Yan L, Ge M, Ren Y, Wang R, Shu Y, Hou L, Guo H. A Pan-cancer analysis of predictive methylation signatures of response to Cancer Immunotherapy. *Front Immunol*. 2021;12:796647. <https://doi.org/10.3389/fimmu.2021.796647>. PMID: 34956232; PMCID: PMC8695566.
  13. Yuan S, Gao Y, Xia Y, Wang Z, Wang X. DNA methylation regulator-mediated modification pattern defines tumor microenvironment immune infiltration landscape in colon cancer. *Front Genet*. 2022;13:1008644. <https://doi.org/10.3389/fgene.2022.1008644>. PMID: 36276973; PMCID: PMC9582351.
  14. Yamashita K, Hosoda K, Nishizawa N, Katoh H, Watanabe M. Epigenetic biomarkers of promoter DNA methylation in the new era of cancer treatment. *Cancer Sci*. 2018;109(12):3695–706. <https://doi.org/10.1111/cas.13812>.
  15. Szejniuk WM, Robles AI, McCulloch T, Falkmer U G I, Røe OD, Roe OD. Epigenetic predictive biomarkers for response or outcome to platinum-based chemotherapy in non-small cell lung cancer, current state-of-art. *Pharmacogenomics J*. 2019;19(1):5–14. <https://doi.org/10.1038/s41397-018-0029-1>.
  16. Shenoy N, Vallumsetla N, Zou Y, Galeas JN, Shrivastava M, Hu C, Susztak K, Verma A. Role of DNA methylation in renal cell carcinoma. *J Hematol Oncol*. 2015;8:88. <https://doi.org/10.1186/s13045-015-0180-y>. PMID: 26198328; PMCID: PMC4511443.
  17. Zhang L, Fan M, Napolitano F, Gao X, Xu Y, Li L. Transcriptomic analysis identifies organ-specific metastasis genes and pathways across different primary sites. *J Transl Med*. 2021;19(1):31. <https://doi.org/10.1186/s12967-020-02696-z>. PMID: 33413600; PMCID: PMC7791985.
  18. Fujishita T, Kojima Y, Kajino-Sakamoto R, Mishiro-Sato E, Shimizu Y, Hosoda W, Yamaguchi R, Taketo MM, Aoki M. The cAMP/PKA/CREB and TGFβ/SMAD4 Pathways Regulate Stemness and Metastatic Potential in Colorectal Cancer Cells. *Cancer Res*. 2022;82(22):4179–4190. <https://doi.org/10.1158/0008-5472.CAN-22-1369>. PMID: 36066360.
  19. Guo Q, Xia B, Zhang F, Richardson MM, Li M, Zhang JS, Chen F, Zhang XA. Tetraspanin CO-029 inhibits colorectal cancer cell movement by deregulating cell-matrix and cell-cell adhesions. *PLoS ONE*. 2012;7(6):e38464. <https://doi.org/10.1371/journal.pone.0038464>. Epub 2012 Jun 5. PMID: 22679508; PMCID: PMC3367972.
  20. Ren C, Ren CH, Li L, Goltsov AA, Thompson TC. Identification and characterization of RTVP1/GLIPR1-like genes, a novel p53 target gene cluster. *Genomics*. 2006;88(2):163–72. <https://doi.org/10.1016/j.ygeno.2006.03.021>. Epub 2006 May 22. PMID: 16714093.
  21. Nakayama M, Oshima M. Mutant p53 in colon cancer. *J Mol Cell Biol*. 2019;11(4):267–76. <https://doi.org/10.1093/jmcb/mjy075>. PMID: 30496442; PMCID: PMC6487790.
  22. Lee JE, Cho YW, Deng CX, Ge K. MLL3/MLL4-Associated PAGR1 regulates adipogenesis by Controlling induction of C/EBPβ and C/EBPδ. *Mol Cell Biol*. 2020;40(17):e00209–20. <https://doi.org/10.1128/MCB.00209-20>. PMID: 32601106; PMCID: PMC7431048.
  23. Ocvirk S, Wilson AS, Appolonia CN, Thomas TK, O'Keefe SJD. Fiber, Fat, and Colorectal Cancer: New Insight into Modifiable Dietary Risk Factors. *Curr Gastroenterol Rep*. 2019;21(11):62. <https://doi.org/10.1007/s11894-019-0725-2>. PMID: 31792624.
  24. Liu S, Sima X, Liu X, Chen H. Zinc finger proteins: functions and mechanisms in Colon cancer. *Cancers (Basel)*. 2022;14(21):5242. <https://doi.org/10.3390/cancers14215242>. PMID: 36358661; PMCID: PMC9654003.
  25. Glasner A, Levi A, Enk J, Isaacson B, Viukov S, Orlanski S, Scope A, Neuman T, Enk CD, Hanna JH, Sexl V, Jonjic S, Seliger B, Zitvogel L, Mandelboim O. NKp46 Receptor-Mediated Interferon-γ Production by Natural Killer Cells Increases Fibronectin 1 to Alter Tumor Architecture and Control Metastasis. *Immunity*. 2018;48(1):107–119.e4. <https://doi.org/10.1016/j.immuni.2017.12.007>. Epub 2018 Jan 9. Erratum in: *Immunity*. 2018;48(2):396–398. PMID: 29329948.
  26. Wan ZH, Jiang TY, Shi YY, Pan YF, Lin YK, Ma YH, Yang C, Feng XF, Huang LF, Kong XN, Ding ZW, Tan YX, Dong LW, Wang HY. RPB5-Mediating protein promotes Cholangiocarcinoma Tumorigenesis and Drug Resistance by competing with NRF2 for KEAP1 binding. *Hepatology*. 2020;71(6):2005–22. <https://doi.org/10.1002/hep.30962>. Epub 2020 Feb 20. PMID: 31541481.
  27. Yang X, Wu K, Li S, Hu L, Han J, Zhu D, Tian X, Liu W, Tian Z, Zhong L, Yan M, Zhang C, Zhang Z. MFAP5 and TNNC1: potential markers for predicting occult cervical lymphatic metastasis and prognosis in early stage tongue cancer. *Oncotarget*. 2017;8(2):2525–35. <https://doi.org/10.18632/oncotarget.12446>. PMID: 27713166; PMCID: PMC5356821.
  28. Yin JH, Elumalai P, Kim SY, Zhang SZ, Shin S, Lee M, Chung YJ. TNNC1 knockout reverses metastatic potential of ovarian cancer cells by inactivating epithelial-mesenchymal transition and suppressing F-actin polymerization. *Biochem Biophys Res Commun*. 2021;547:44–51. <https://doi.org/10.1016/j.bbrc.2021.02.021>. Epub 2021 Feb 13. PMID: 33592378.
  29. Guo J, Liu Y. INHBA promotes the proliferation, migration and invasion of colon cancer cells through the upregulation of VCAN. *J Int Med Res*. 2021;49(6):3000605211014998. <https://doi.org/10.1177/03000605211014998>. PMID: 34130530; PMCID: PMC8212385.
  30. Cui F, Qu D, Sun R, Zhang M, Nan K. NK cell-produced IFN-γ regulates cell growth and apoptosis of colorectal cancer by regulating IL-15. *Exp Ther Med*. 2020;19(2):1400–6. <https://doi.org/10.3892/etm.2019.8343>. Epub 2019 Dec 18. PMID: 32010315; PMCID: PMC6966233.
  31. Ruckova E, Muller P, Nenutil R, Vojtesek B. Alterations of the Hsp70/Hsp90 chaperone and the HOP/CHIP co-chaperone system in cancer. *Cell Mol Biol Lett*. 2012;17(3):446–58. <https://doi.org/10.2478/s11658-012-0021-8>. Epub 2012 Jun 5. PMID: 22669480; PMCID: PMC6275590.
  32. Antonacopoulou AG, Grivas PD, Skarlas L, Kalofonos M, Scopa CD, Kalofonos HP. POLR2F, ATP6V0A1 and PRNP expression in colorectal cancer: new molecules with prognostic significance? *Anticancer Res*. 2008;28(2B):1221–7. PMID: 18505059.
  33. Eisa NH, Jilani Y, Kainth K, Redd P, Lu S, Bougrine O, Abdul Sater H, Patwardhan CA, Shull A, Shi H, Liu K, Elsherbiny NM, Eissa LA, El-Shishtawy MM, Horuzsko A, Bollag R, Mailhe N, Roig J, Korkaya H, Cowell JK, Chadli A. The co-chaperone UNC45A is essential for the expression of mitotic kinase NEK7 and tumorigenesis. *J Biol Chem*. 2019;294(14):5246–60. <https://doi.org/10.1074/jbc.RA118.006597>. Epub 2019 Feb 8. PMID: 30737284; PMCID: PMC6462532.

34. Li W, Guo L, Tang W, Ma Y, Wang X, Shao Y, Zhao H, Ying J. Identification of DNA methylation biomarkers for risk of liver metastasis in early-stage colorectal cancer. *Clin Epigenetics*. 2021;13(1):126. <https://doi.org/10.1186/s13148-021-01108-3>. PMID: 34108011; PMCID: PMC8190869.
35. Macartney-Coxson DP, Hood KA, Shi HJ, Ward T, Wiles A, O'Connor R, Hall DA, Lea RA, Royds JA, Stubbs RS, Rooker S. Metastatic susceptibility locus, an 8p hot-spot for tumour progression disrupted in colorectal liver metastases: 13 candidate genes examined at the DNA, mRNA and protein level. *BMC Cancer*. 2008;8:187. <https://doi.org/10.1186/1471-2407-8-187>. PMID: 18590575; PMCID: PMC2488356.
36. Cao J, Yan Q. Cancer Epigenetics T, Immunity. *Immunotherapy Trends Cancer*. 2020;6(7):580–92. <https://doi.org/10.1016/j.trecan.2020.02.003>. Epub 2020 Mar 31. PMID: 32610068; PMCID: PMC7330177.
37. Chen T, Wang Y, Nan Z, Wu J, Li A, Zhang T, Qu X, Li C. Interaction between macrophage extracellular traps and Colon cancer cells promotes Colon Cancer Invasion and correlates with unfavorable prognosis. *Front Immunol*. 2021;12:779325. <https://doi.org/10.3389/fimmu.2021.779325>. PMID: 34925357; PMCID: PMC8671452.
38. Khan U, Chowdhury S, Billah MM, Islam KMD, Thorlacius H, Rahman M. Neutrophil Extracellular traps in Colorectal Cancer Progression and Metastasis. *Int J Mol Sci*. 2021;22(14):7260. <https://doi.org/10.3390/ijms22147260>. PMID: 34298878; PMCID: PMC8307027.
39. Di Franco S, Turdo A, Todaro M, Stassi G. Role of type I and II interferons in Colorectal Cancer and Melanoma. *Front Immunol*. 2017;8:878. <https://doi.org/10.3389/fimmu.2017.00878>. PMID: 28798748; PMCID: PMC5526853.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.