

# Construction of a Prognostic Model Based on Methylation-Related Genes in Patients with Colon Adenocarcinoma

ZhenDong Liu<sup>1</sup>, YuYang Xu<sup>1</sup>, Shan Jin<sup>2</sup>, Xin Liu<sup>1</sup>, BaoChun Wang<sup>1</sup>

<sup>1</sup>Department of General Surgery, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, Hainan Province, People's Republic of China; <sup>2</sup>Department of Anesthesiology, Second Affiliated Hospital of Hainan Medical University, Haikou, Hainan Province, People's Republic of China

Correspondence: BaoChun Wang, Department of General Surgery, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, 570311, People's Republic of China, Tel +86-13876180006, Email DrIance9@163.com

**Purpose:** Colon adenocarcinoma (COAD) is the second leading cause of death in the world, and the new incidence rate ranks third among all cancers. Abnormal DNA methylation is related to the occurrence and development of tumors. In this study, we aimed to identify genes associated with abnormal methylation in COAD.

**Methods:** COAD transcriptome data, methylation data and clinical information were downloaded from the TCGA database and GEO database. The differentially expressed genes (DEGs) and methylated genes (DMGs) were analyzed and identified in COAD. PCA analysis was applied to divide COAD into subtypes, and the survival and immune cell infiltration of each subtype were evaluated. Cox and LASSO analyses were performed to construct COAD risk model. GSEA was used to evaluate the enrichment pathways. The Kaplan–Meier was used to analyze the difference in survival. ROC curve was plotted to evaluate the accuracy of the model, and GSE17536 was used to verify the accuracy of the risk model. The risk model is combined with the clinicopathological characteristics of COAD patients to perform multivariate Cox regression analysis to obtain independent risk factors and draw nomograms.

**Results:** In total, 4564 DEGs and 1093 DMGs were screened, among which 298 were found to be overlapping genes. For 220 of these overlapping genes, the methylation was significantly negatively correlated to expression levels. An optimal signature from 4 methylated biomarkers was identified to construct the prognostic model.

**Conclusion:** Our study identified 4 methylated biomarkers in the COAD. Then, we constructed the risk model to provide a theoretical basis and reference value for the research and treatment of COAD.

**Keywords:** colorectal cancer, methylated genes, immune cell infiltration, WGCNA, prognosis

## Introduction

Colorectal cancer (CRC) is currently the second most lethal tumor in the world, the incidence of CRC has increased to the third among all cancers.<sup>1</sup> It is generally believed that environmental and chronic intestinal inflammation are the risk factors of CRC.<sup>2</sup> Most cases of CRC are considered as a result of stepwise accumulation of mutations, suppression of tumor suppressor genes and activation of oncogenes.<sup>3</sup> Colon adenocarcinoma (COAD) is the most common subtype of CRC. Significant progresses have been made in early diagnosis, surgery, radiotherapy and chemotherapy in the past decades;<sup>4</sup> however, the prognosis of colorectal cancer patients is still poor, and the 5-year survival rate is low. Therefore, diagnostic markers and therapeutic targets of COAD are needed to further explore to provide new methods of the treatment and intervention for patients.

Recently, epigenetic alterations have been presumed significant contributors to cancer development.<sup>5</sup> DNA methylation is the most widely studied epigenetic alteration in cancer, which can regulate gene expression, and involved in early stages of tumorigenesis. Epigenetic markers are thought to be interesting biomarkers for the diagnosis, prognosis and prediction of treatment response of cancer.<sup>5–7</sup> It is generally believed that epigenetic changes in CRC can be manifested earlier than genetic

changes. It makes the new epigenetic therapy an attractive treatment strategy by targeting to inhibit DNA methylation genes. However, there is insufficient evidence for the use of methylated genes in routine clinical practice for CRC currently, and further explorations are needed to discover more useful diagnostic biomarkers.

The purpose of this study was to find abnormal methylation genes in CRC and evaluate their prognostic value in colon cancer.

## Materials and Methods

### Patient Samples

Ten tumor tissues and their adjacent normal tissues were isolated and collected from patients who were diagnosed with COAD at the Department of General Surgery, Hainan General Hospital. Participants were enrolled upon obtaining their written consent. This study acquired the approval from the ethics committee of Hainan General Hospital (NO. 【2022】 719).

### Data Source

The clinical data, transcription and methylation profiles of 370 colon adenocarcinoma (COAD) patients were downloaded from TCGA database and were used as training set. GSE17536<sup>8</sup> was downloaded from GEO database and was used as external validation sets. A summary of clinical background data is presented in Table 1.

### Acquirement of Differentially Expressed Genes (DEGs) and Aberrantly Methylated Genes

The microarray data of methylation and expression were analyzed by R software after normalization.

**Table 1** Characteristics of COAD Patients in Our Study

	Training Set (n = 370)	GSE17536 (n = 177)
Age (years)		
<70	197	104
≥70	173	73
Gender		
Female	175	81
Male	195	96
T Stage		
T1	9	–
T2	65	–
T3	253	–
T4	42	–
TX	1	–
N Stage (pN)		
NO	223	–
N1	84	–
N2	63	–
NX	–	–
M Stage		
MO	273	–
M1	97	–
TNM Stage		
I	64	24
II	143	57
III	109	57
IV	54	39
X	–	–

R package “limma” was used to identify differentially expressed mRNAs between normal and COAD samples ( $|\log FC| > 1$ ,  $P_{\text{adjust}} < 0.05$ ).<sup>9</sup> Similarly, “limma” package was used to screen aberrantly methylated genes ( $|\log FC| > 0.2$ ,  $P_{\text{adjust}} < 0.05$ ). Thereafter, the differentially expressed mRNAs and the aberrantly methylated genes were intersected.

## Consensus Clustering in COAD Patients

Consensus clustering, which helped us determine the best grouping scheme, was performed using the ConsensusClusterPlus package in R.<sup>10</sup> The cumulative distribution function (CDF) reflects the value distribution of the consensus matrix for different  $k$  values. The optimal value of  $k$  was defined as the smallest increase in the area under the cumulative distribution function (CDF) curve. The heatmap corresponding to consensus clustering was generated using R heatmap package (<http://CRAN.R-project.org/package=pheatmap>).

## Survival Analyses

Kaplan–Meier plots were used to illustrate overall survival among COAD subtypes defined by DNA methylation profiles. Significant differences among the clusters were assessed using the Log rank test. Survival analysis was performed using the survival package in R. The relationship between clinical and biological characteristics with DNA methylation clustering was analysed using the chi-square test.

## Immune Cell Infiltration Analysis

We analyzed RNA-seq data from the 370 patients in CIBERSORTx,<sup>11</sup> then compared the degree of infiltration of 22 immune cells between the two clusters using the Wilcoxon test. Pearson correlations between immune cells were calculated using the Corrplot package.

The stromal scores, immune scores and estimate stores in TCGA COAD dataset were calculated using ESTIMATE algorithm to estimate the tumor purity for each COAD sample.

## GO Annotation and KEGG Pathway Enrichment Analysis

In order to know more about the molecular functions of DEGs in COAD, GO annotation and KEGG enrichment analysis were conducted with the aid of the “clusterProfiler” package.<sup>12</sup> GO analysis tends to explain gene function from three aspects: molecular function, biological process, and cellular components, while the KEGG pathway enrichment analysis is prone to describe gene function in the genomic and molecular levels and show the correlated genes ( $P_{\text{adjust}} < 0.05$ ).

## Establishment and Validation of the Prognostic Model

The DEGs from the TCGA-COAD dataset were subjected to univariate Cox regression analysis. Genes with  $P_{\text{adjust}} < 0.05$  were regarded as the genes correlated with COAD prognosis. Afterwards, Lasso Cox regression analysis was performed to select gene signatures in the TCGA-COAD dataset using the R packages “glmnet” (<https://cran.r-project.org/web/packages/glmnet/index.html>). Four powerful prognostic genes were selected through optimal value of lambda ( $\lambda$ ). A risk score was calculated for each sample, which was defined as a linear combination of expression values of individual normalized genes weighted by their estimated Cox model regression coefficients. Patients were divided into high-risk group and low-risk group according to the average risk score. Hallmarks and KEGG pathways significantly enriched in high-risk group were analyzed by GSEA. The Kaplan–Meier survival and ROC curves were drawn to evaluate the predictive value of the prognostic gene signature using the R packages “survival” and “survival ROC”, respectively. The risk model was tested in the external validation sets.

Thereafter, multivariate Cox regression analyses were performed to identify independent prognostic factors for COAD patients. A nomogram was constructed to predict 1-, 3-, and 5-year survival in COAD patients according to independent prognostic factors. Nomogram performance was assessed using calibration curves and ROC curves.

## RNA Extraction and cDNA Synthesis

Total RNAs of human cells were extracted using the Trizol reagent (Invitrogen, Carlsbad, CA, USA) according to the instructions of the manufacturer and treated with RQ1 DNase (Promega, Madison, WI, USA) to remove DNA. The

quality and quantity of the purified RNA were determined by measuring the absorbance at 260 nm and 280 nm (A260 and A280) using a SmartSpec Plus Spectrophotometer (Bio-Rad Laboratories, Inc., Hercules, CA, USA). RNA integrity was further verified by electrophoresis using a 1.5% agarose gel. All RNA samples were stored at  $-80^{\circ}\text{C}$  for future use. Reverse transcription reactions were carried out using ReverTra Ace qPCR RT Kit (TOYOBO Life Science, Shanghai, China), according to the manufacturer's instructions.

## Quantitative Real-Time PCR (qRT-PCR)

Expression levels of *Bal1*, *ERBB4*, *MPP2* and *PCDH9* gene were detected by qRT-PCR. For *Bal1*, the primers used were as follows: 5'-GTCTCCAGAACCACCACATCATTG-3' (forward), 5'-GCTCCTTCAATCTCTAACTCTGTCC-3' (reverse). For *ERBB4*, the primers used were as follows: 5'-TTTCGGGAGTTTGAGAATGG-3' (forward), 5'-GAAACTGTTTGCCCCCTGTA-3' (reverse). For *MPP2*, the primers used were as follows: 5'-GAGACGTTTCAGAGCCCTTG-3' (forward), 5'-ACCAGCTCCAGGTTGTTGTC-3' (reverse). For *PCDH9*, the primers used were as follows: 5'-ATTTGATAGAGAGCAGCAGAGTTCC-3' (forward), 5'-AGTTCAGCGTTCATTCCAGTGTC-3' (reverse). Actin was used as an internal reference control and its detection was performed with the following primers: Forward: 5'-TGGACTTCGAGCAAGAGATG-3' and reverse: 5'-GAAGGAAGGCTGGAAGAGTG-3'.

The qRT-PCR was performed on a Bio-Rad S1000 with Bestar SYBR GreenRT-PCR Master Mix (TOYOBO). PCR conditions consisted of denaturing at  $95^{\circ}\text{C}$  for 1 min, and 40 cycles of denaturing at  $95^{\circ}\text{C}$  for 15s followed by annealing and extension at  $60^{\circ}\text{C}$  for 30s. Relative gene expression was calculated using the Livak and Schmittgen  $2^{-\Delta\Delta\text{Ct}}$  method,<sup>13</sup> normalized with the reference gene Actin. PCR amplifications were performed in triplicate for each sample.

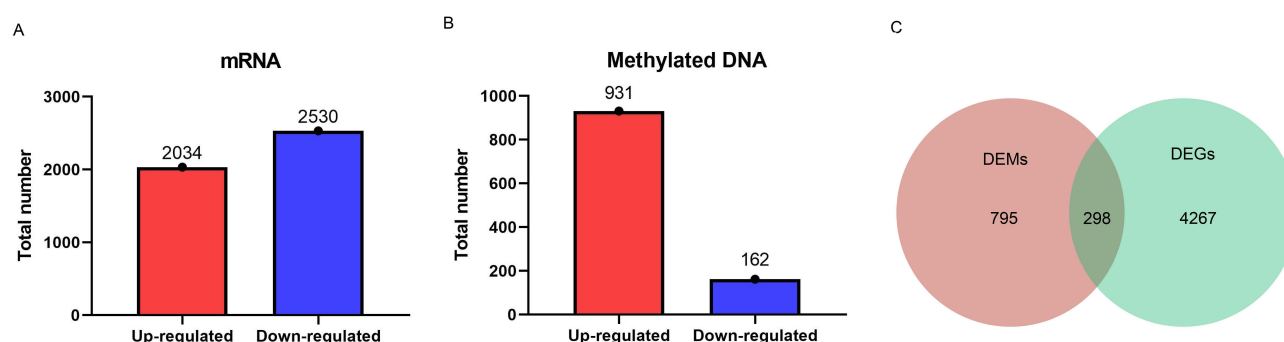
## Results

### Identification of the Differentially Expressed Genes and the Aberrantly Methylated Genes in COAD

Based on the mRNA expression data and the methylation data obtained from the TCGA-COAD dataset, a total of 4564 differentially expressed mRNAs (Figure 1A) and 1093 abnormal methylation sites (Figure 1B) were obtained. Thereafter, 298 genes were identified by intersecting differentially expressed mRNAs and the aberrantly methylated genes (Figure 1C). Among them, the expressions of 220 genes were negatively correlated with methylated level.

### Consensus Clustering of COAD Identified Distinct DNA Methylation Prognosis Subtypes with Different Prognosis and Immune Microenvironment

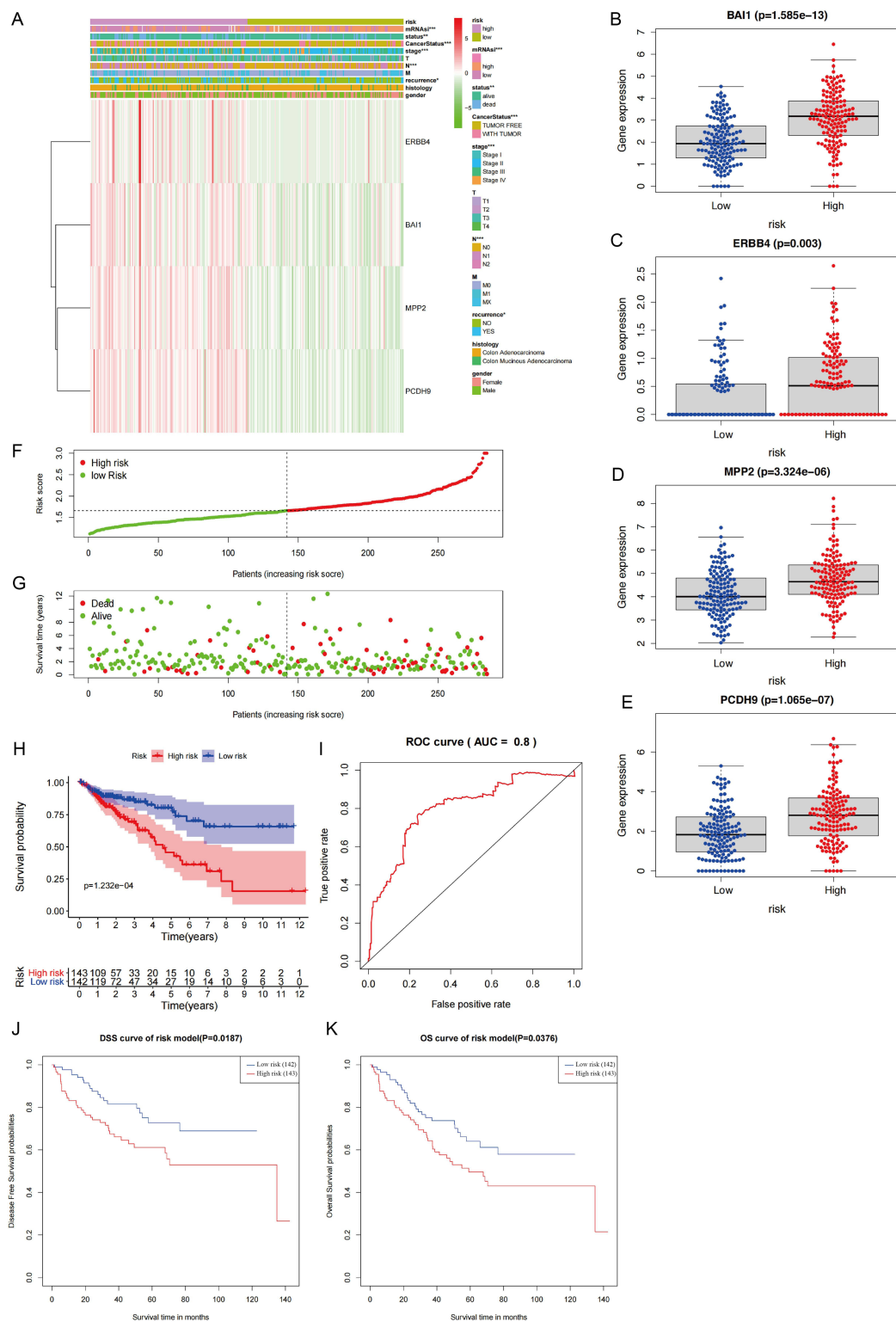
These 220 methylated DEGs were used for consensus clustering to obtain distinct DNA methylation prognostic molecular subtypes of COAD (Figure 2A–C). Finally, we created a heatmap of the consensus matrix when  $k = 2$  (Figure 2D). Meanwhile, principal component analysis (PCA) showed that patients in the cluster 2 were significantly



**Figure 1** Identification of mRNA expression data and the methylation data in the Cancer Genome Atlas (TCGA) database.

**Notes:** (A) Number of up- and downregulated DEGs. (B) Number of up- and downregulated DEMs. (C) Venn diagrams showing the number of genes in DEMs or DEGs or in both.





**Figure 2** Identification of the DNA methylation prognosis subtypes based on 220 methylated DEGs.

**Notes:** (A) Consensus clustering matrix for  $k = 2$ . (B) CDF plot of the consensus score ( $k = 2-9$ ). (C) Delta area curve of the sample. The Delta area curve of the consensus clustering indicates the relative change in area under the cumulative distribution function (CDF) curve for each category number  $k$  compared with  $k-1$ . The horizontal axis represents the category number  $k$ , and the vertical axis represents the relative change in area under the CDF curve. (D) Consensus matrix, the heat map of sample clustering when  $k = 2$ . (E) PCA based on clusters 1 and 2 results. (F) Kaplan-Meier analysis of survival rate for clusters 1 and 2. (G) Differential immune infiltrates in clusters 1 and 2 of TCGA-COAD dataset. The proportion of 22 immune infiltrates for each case is estimated using the ESTIMATE algorithm. (H) Correlation matrix followed by unsupervised hierarchical clustering in 22 immune subsets. Pearson correlation coefficients ( $R$ ) were calculated. Correlation coefficients were plotted with negative correlation (blue), positive correlation (red), and  $R = 0$  (white). Comparison of the (I) stromal, (J) immune and (K) estimate scores between clusters 1 and 2.

different from patients in the cluster 1 (Figure 2E). Kaplan–Meier survival analysis showed that the outcome of COAD patients in two clusters was significantly different (Figure 2F).

Comparing the degree of infiltration of 22 kinds of immune cells, we detected that the abundance of most immune cell types, except T cell CD4 naive, was significantly higher in cluster 2 than those in cluster 1 (Figure 2G). Monocytes, T cells gamma delta and eosinophils showed a positive correlation (Figure 2H). Also, the cluster 2 was found to have the higher stromal, immune and estimate scores (Figure 2I–K).

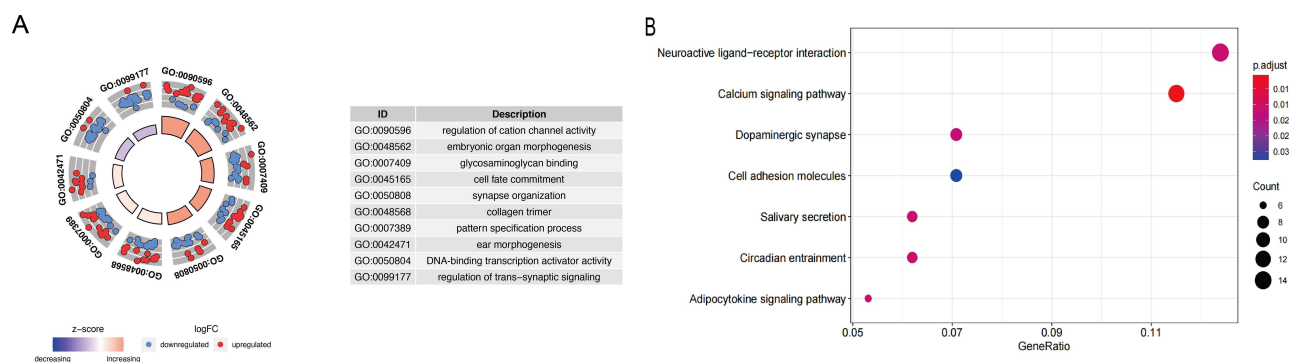
## GO Annotation and KEGG Pathway Enrichment Analysis

To gain more insight into the molecular mechanism of the 220 methylated DEGs underlying the initiation and progression of COAD, GO annotation and KEGG pathway enrichment analysis were carried out using the “clusterProfiler” package. The most enriched GO terms of 220 genes were regulation of cation channel activity, embryonic organ morphogenesis, and glycosaminoglycan binding (Figure 3A), and the KEGG analysis revealed that the genes were significantly enriched in the signaling pathways involved in neuroactive ligand–receptor interaction, calcium signaling pathway (Figure 3B).

## Establishment and Validation of the Prognostic Model

Firstly, the 220 methylated DEGs were subjected to univariate Cox regression analysis, and 29 genes with  $P_{\text{adjust}} < 0.05$  were identified to be significantly correlated with OS (Table 2). Then, lasso-penalized Cox analysis was performed to further narrow the mRNAs (Figure 4A and B). A 4-gene signature about BAI1, ERBB4, MPP2, and PCDH9 was identified based on the optimal value of  $\lambda$ . Spearman correlation analysis showed that 4 genes have strong correlation with each other (Figure 4C). The risk score was calculated as follows: The risk score =  $(0.008 \times \text{BAI1}) + (0.054 \times \text{ERBB4}) + (0.062 \times \text{MPP2}) + (0.085 \times \text{PCDH9})$ . The samples were classified into high-risk and low-risk groups according to the median risk score. Analyzing the correlation between clinical traits and risk scores, we found that the severity of traits is positively correlated with risk scores (Figure 4D–H).

Then, we analyzed the distribution of clinicopathological variables and the expression of four genes in the high- and low-risk groups (Figure 5A). The high-risk and low-risk groups were distinguishable in both the N stage and stage groups. The expression levels of 4 genes in the high-risk groups are higher than those in the low-risk groups (Figure 5B–E). Survival analysis showed that the lower survival outcome of COAD patients was related to higher risk score (Figure 5F–G). The Kaplan–Meier analysis showed a significant difference in the outcome of the patients between the high-risk group and the low-risk group (Figure 5H). The area under the ROC curve (AUC) for the model was 0.8 in TCGA (Figure 5I). In the GSE17536 cohort, Kaplan–Meier analysis also showed a significant difference of disease-specific survival (DSS) (Figure 5J) and overall survival (OS) (Figure 5K) in the two groups (high-risk and low-risk).



**Figure 3** Gene ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the 220 methylated DEGs.

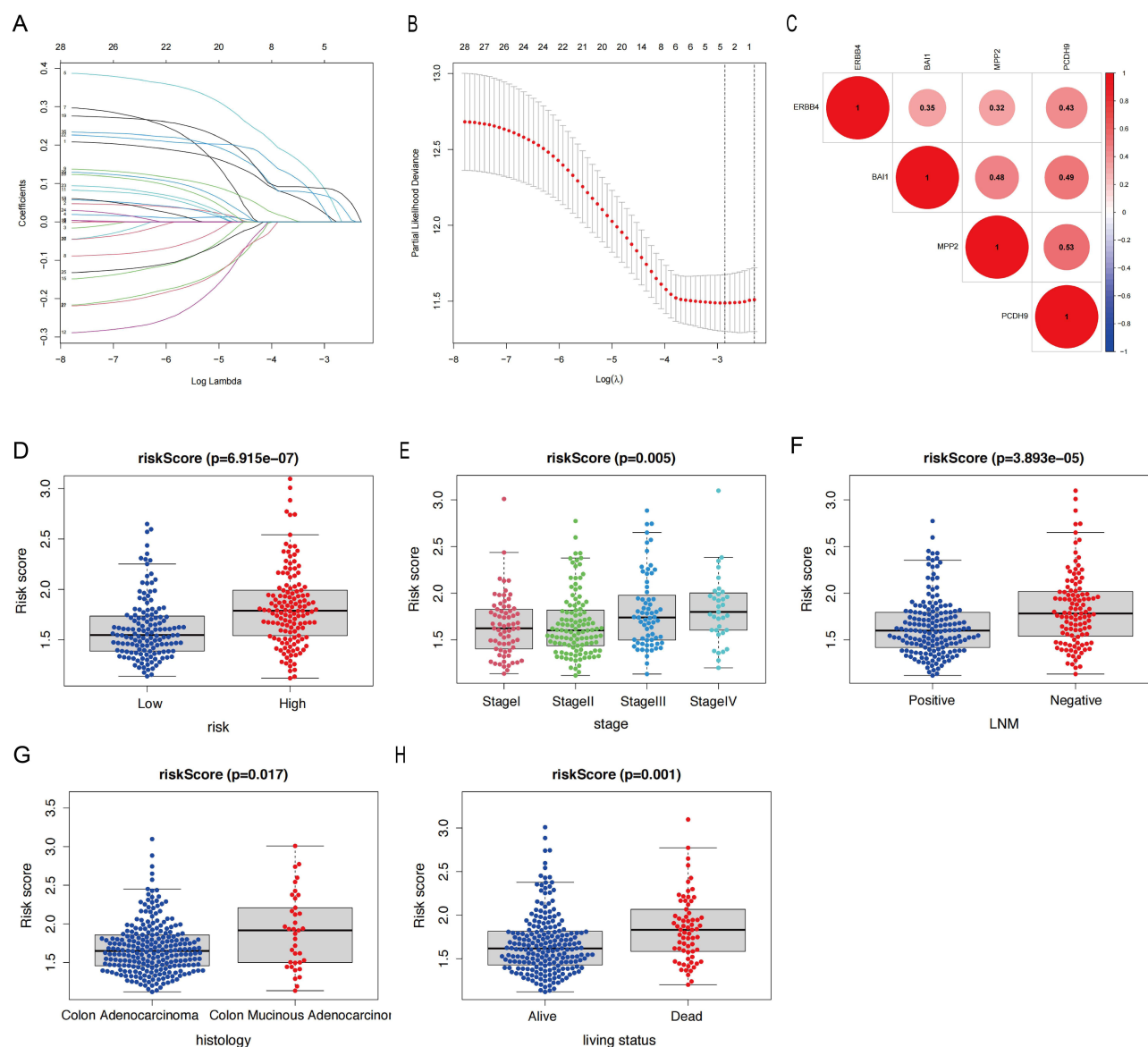
**Notes:** (A) GO annotation; (B) KEGG pathway enrichment.

**Table 2** Univariate Cox Regression Analysis of Overall Survival for 29 Methylated DEGs

ID	HR	p value
PCDH9	1.305635503	0.000416363
MPP2	1.389700255	0.000944191
SH3GL3	1.665780761	0.001690292
ERBB4	1.533967257	0.00238687
BAI1	1.339050065	0.002693955
DPY19L2	1.286991265	0.004343696
SPOCK3	1.269237521	0.008626801
GPR26	1.515352433	0.010395582
SUSD5	1.20318346	0.01462343
THBS4	1.094211413	0.014733892
SYPL2	1.23181808	0.016285655
CLDN11	1.18579082	0.016526704
PDE1B	1.206703294	0.023462408
LRRC7	1.287364371	0.023753827
MAP6	1.197648896	0.024951089
TNFAIP8L3	1.23212855	0.027177005
GAP43	1.139091339	0.030803494
HCG4	1.217710765	0.031797059
HAPLN4	1.250325133	0.036725183
CLEC4GPI	1.361106247	0.039393969
GRIN2A	1.188587806	0.039606636
P2RX2	1.442913531	0.041288082
SV2B	1.143512853	0.04381782
GSTM2	1.17805037	0.045839375
JPH4	1.218200089	0.046513808
PACSIN1	1.219576637	0.047952613
RNF180	1.209748005	0.048280595
KCNK12	1.260793438	0.048418966
TCEAL2	1.115879088	0.049142325

## GSEA Analysis

We analyzed the hallmarks and KEGG pathways enriched in the high-risk group by GSEA. Hallmarks such as adipogenesis, apical junction, hypoxia, myogenesis, and notch signaling were found to be significantly enriched in the high-risk group (Figure 6A). Additionally, KEGG pathways, including adherens junction, colorectal cancer, pancreatic cancer, ERBB, and

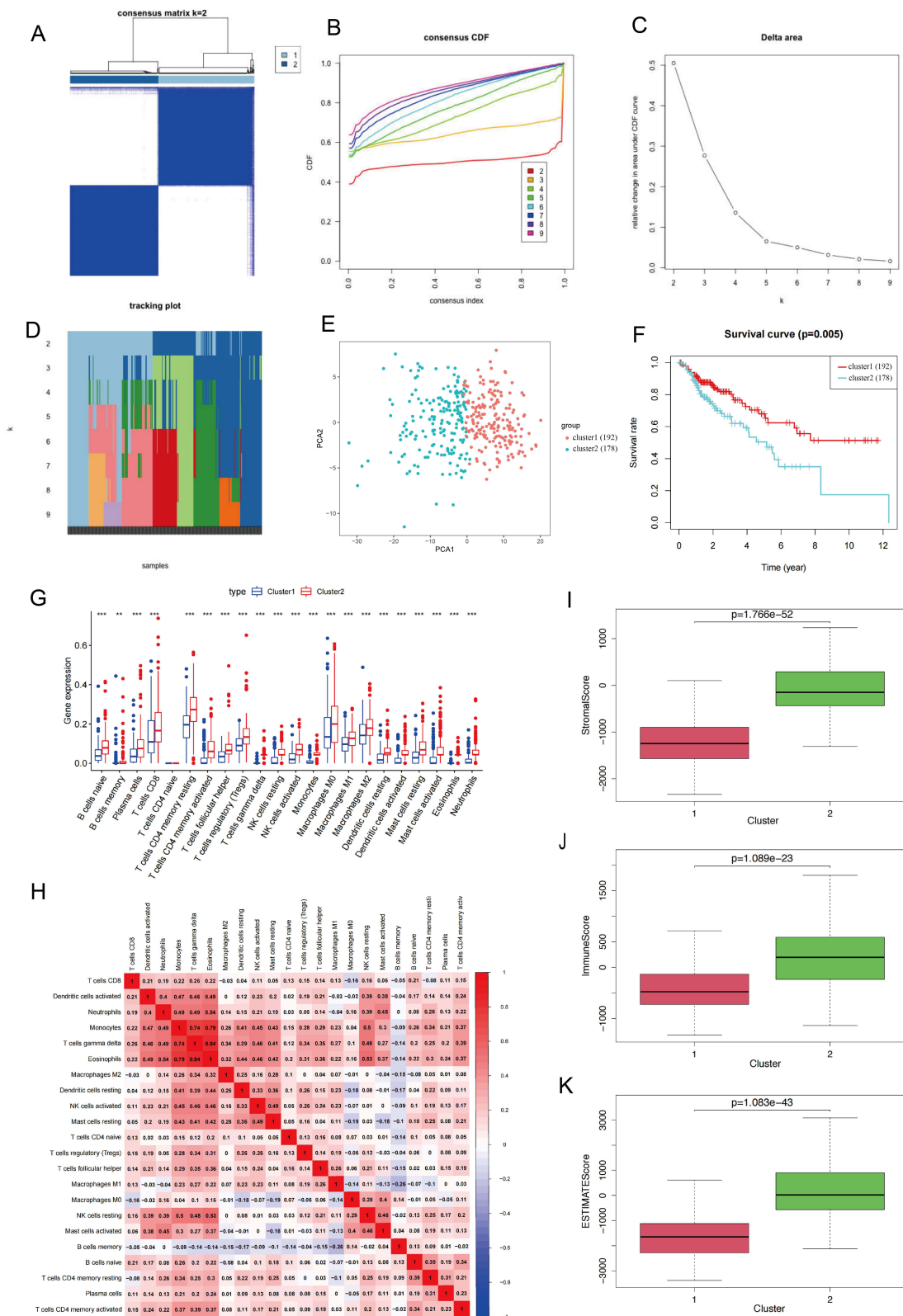


**Figure 4** Establishment of a prognostic model based on (A and B) LASSO regression analysis. (C) Identification of 4-gene signature about BAI1, ERBB4, MPP2, and PCDH9 by Spearman correlation analysis. (D–H) Correlation analysis of the 4-gene signature risk score with clinical characteristics.

WNT signaling pathway, were also significantly enriched in the high-risk group (Figure 6B). These findings suggest potential roles for the four prognostic genes in the progression, cell fate and immune responses of COAD.

## Construction of the Nomogram in COAD

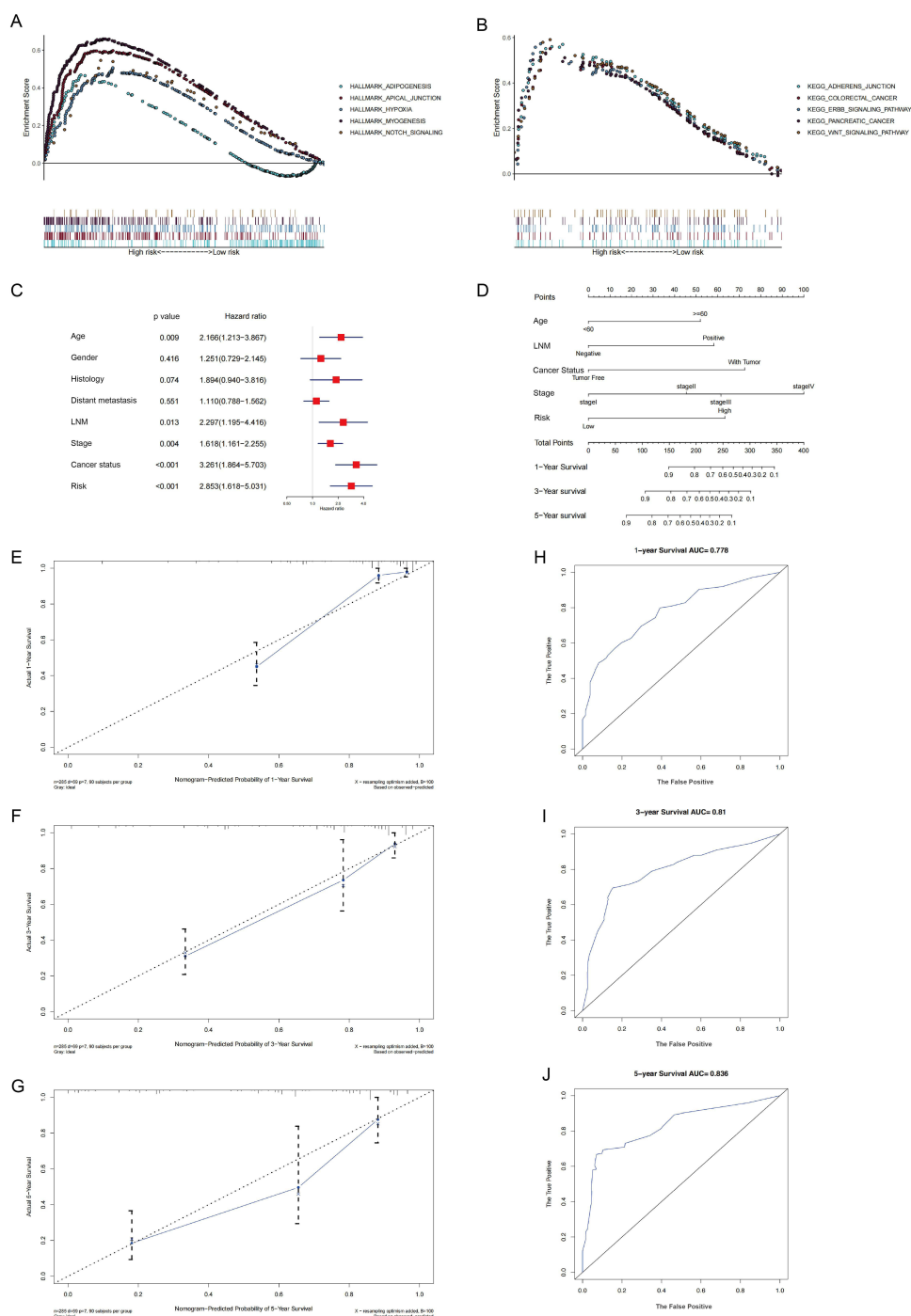
Next, we performed multivariate analyses to detect independent prognostic factors. Age, gender, histology, distant metastasis, lymph node metastasis (LNM), stage, cancer status and risk score were included into multivariate analysis. The result showed that age, LNM, stage, cancer status and risk score were significantly associated with prognosis (Figure 6C), indicating that they are independent prognostic factors in COAD. Thereafter, we constructed a nomogram to predict the 1-, 3-, and 5-year survival of COAD patients based on independent prognostic factors (Figure 6D). The calibration curves for 1-, 3-, and 5-year showed that the nomogram-predicted probability of survival was close to the actual survival (Figure 6E–G). ROC curves further demonstrated the high accuracy of the nomogram in predicting the survival of COAD patients that the AUC was 0.778, 0.81 and 0.836 for 1-, 3-, and 5-year (Figure 6H–J).



**Figure 5** Prognostic risk scores correlated with clinicopathological features.

**Notes:** (A) Heatmap and clinicopathological features of high- and low-risk groups. (B–E) Distribution of expression levels stratified by 4 genes. (F) Distribution of risk scores in high-risk group and low-risk group. Red point indicates case in high-risk group and green indicates low-risk case. (G) Distribution of survival status of patients in high-risk group and low-risk group. Green point represents alive and red point for death. \*\*P < 0.01, \*\*\*P < 0.001. (H) Survival curve of outcome. Red line depicts the survival of high-risk patients and blue line for low-risk patients. (I) ROC Curve for risk score. (J) DSS curve and (K) OS curve for COAD (GSE17536).

**Abbreviation:** AUC, area under curve.



**Figure 6** Gene Set Enrichment Analysis (GSEA) in high-risk patients and construction of the nomogram in COAD.

**Notes:** (A) The enriched Hallmarks in high-risk group. (B) The enriched KEGG pathways in high-risk group. (C) Multivariate analysis was performed for assessing risk factors and constructing prognostic nomogram. (D) The predicted 1-, 3-, and 5-year survival rates of COAD based on independent prognostic nomogram constructed using the risk score from prognostic signature and clinicopathological parameters. (E–G) The calibration plot for internal validation of the nomogram. (H–J) ROC curves of the nomogram for 1-, 3- and 5-year overall survival in COAD to evaluate the predictive performance of the nomogram.

## Discussion

More and more evidence indicated great exploration potential in the pathogenesis of malignant tumors.<sup>14</sup> The TCGA database covers 36 types of cancer and detailed clinical information of patients involved.<sup>15</sup> The GEO database provides



online tools and faster data updates. The combination of these tools provides us a more convenient and efficient method to explore various malignant tumors.

The usual treatment of COAD is a comprehensive treatment strategy consisting of surgery, neoadjuvant therapy, transformation therapy and postoperative adjuvant therapy. However, the 5-year survival rates of COAD patients remain unsatisfactory. It is necessary for us to identify novel biomarkers to improve the management of COAD.

In this study, we identified 220 aberrantly expressed methylation genes, and cluster analysis indicated that there was a significant difference in survival and immune infiltration between the two clusters of patients. The increasing expression level in cluster 2 group has positive correlations with CD8+ T cell, CD4+ Memory Resting T cell, M0 macrophage, M1 macrophage and M2 macrophage, which are associated with tumor immune microenvironment. The infiltration of immune cells could be a promising prognostic biomarker for colorectal cancer.<sup>16</sup> To further investigate the effect of methylation genes in biological processes and pathways, GO annotation and KEGG pathway enrichment analysis were performed. These 220 genes were enriched in signaling pathways like the regulation of cation channel activity, embryonic organ morphogenesis, glycosaminoglycan binding, cell fate commitment, synapse organization, collagen trimer, which were associated with the biological process of colonic epithelium and mucosa formation. Studies have revealed that cation channel activity and calcium signaling pathway were associated with the proliferation, migration, and survival of cancer cell.<sup>17–19</sup> Glycosaminoglycan could promote the growth of colorectal cancer.<sup>20</sup> In addition, DNA methylation and inhibition of gene expression are intimately related to the cellular differentiation process and colonic cancer formation.<sup>21,22</sup> The embryonic organ morphogenesis and cell fate commitment signaling pathways clustered by 220 methylated genes in our study were also corroborated, so aberrantly expressed methylation genes may have effect on colon cancer through biological process.

Then, single-factor regression and LASSO regression algorithm were used to find characteristic genes, and 4-gene (BAI1, PCDH9, ERBB4 and MPP2) signature was identified. These genes were reported to be closely related to the prognosis of colon cancer. As a number of brain-specific angiogenesis inhibitor (BAI) family, BAI1 is an orphan GPCR-like receptor abundantly expressed in normal brain and has been suggested to play an important role in angiostasis by acting as a mediator of p53 signaling.<sup>23–25</sup> It was reported that colorectal cancer was negatively correlated with BAI1 gene expression.<sup>23</sup> The expression of PCDH9 is downregulated in tumor samples compared with normal tissues and suggested to be a potential biomarker for predicting the prognosis of patients with cerebral glial tumors.<sup>26</sup> ERBB4 is over-expressed in colon cancer and enhances the survival and growth of cells driven by Ras and/or WNT signaling.<sup>27</sup> The mutation of ERBB4 confers the risk of colorectal cancer, and this result suggested that ERBB4 would be a potentially novel biomarker for CRC susceptibility.<sup>28,29</sup> MPP2 is a tumor suppressor gene and primarily participates in the adhesion of epithelial cell and immune infiltration.<sup>30,31</sup> The expression levels of MPP2 were high in CRC tissues.<sup>32</sup> It was reported that MPP2 is related to the 5-year survival of rate in colon cancer patients.<sup>33</sup> Studies have also constructed a prognostic risk model to predict overall survival in COAD, with MMP2 being the hub gene in both.<sup>31,32</sup> Similar results emerged in our study that low-risk group with BAI1, PCDH9, ERBB4 and MPP2 low-expression had a better overall survival outcome, although the expression of these 4 genes has no significant difference between cancerous and adjacent tissues (data not shown). In short, the methylation regulation mechanisms of these DEGs need to be further studied.

Of course, the specific role and mechanism of these genes in COAD need further experimental research. In our study, the potential mechanisms of four prognostic genes regulating colon cancer were analyzed by GSEA in high-risk group, and there were some hallmarks and signal pathways significantly enriched in high-risk group. Apical junction pathway is associated with metastatic potential, which could lead to aggressive clinical characteristics with shorter survival.<sup>34,35</sup> Epithelial–mesenchymal transition (EMT) plays an essential role in cancer development,<sup>36</sup> and previous reports showed that the apical junction pathway could increase metastatic potential through EMT as well as angiogenesis.<sup>34</sup> In addition, GSEA results in our study revealed enrichment of both WNT and adherens junction signaling pathways in high-risk group. WNT signaling affected the proliferation and tumor formation invasion and metastasis of colon cancer cells,<sup>37,38</sup> while adherens junction pathway involved in the canonical WNT signaling pathway.<sup>39</sup> Notch signaling contributes to EMT of colon cancer and correlated to colonic inflammation.<sup>40,41</sup> In summary, the abnormal expression of methylation genes may lead to the imbalance of these organisms which lead to worse survival.

Risk factors such as carcinoembryonic antigen (CEA) level, tumor site, TNM stage and other clinical pathological data have been reported related to prognosis for COAD, which have certain limitations as prognostic markers.<sup>42–44</sup> Therefore, there needs further exploration of promising prognostic parameters to improve the survival and quality of life for COAD patients. In view of the important role of DNA methylation in colon cancer, it is necessary to pay more attention to the potential predictive of methylation-related genes for COAD prognosis.<sup>45</sup> Several investigators have published related research. Xue et al identified 7 methylation-related genes as novel prognostic markers, which accurately classified patients into high- and low-risk groups with significantly different overall survival times.<sup>46</sup> Zhou et al found that YTHDF3, KIAA1429, ALKBH5 and METTL3 were significantly correlated with OS of colon cancer and recto-sigmoid cancer (RSC).<sup>47</sup> In our study, we constructed a four-gene risk prediction model, which combined age, LNM, stage, cancer status and risk score to facilitate personalized prediction. Time-dependent ROC indicated the high accuracy of the nomogram in predicting the survival of COAD patients.

In this study, we constructed a nomogram to predict COAD survival. However, there are still some limitations in our study. The prognostic symptom model needs to be further validated in other large independent samples to ensure the reliability. Functional experiments need to further reveal the possible mechanisms predicting the role of autophagy genes. In a word, we have provided a theoretical reference for the exploration of potential biomarkers for the diagnosis and prognosis of COAD in the future.

## Data Sharing Statement

The gene expression, methylation, and clinical data are available from NCBI Gene Expression Omnibus (GEO: GSE17536) <https://www.ncbi.nlm.nih.gov/geo/> and the cancer genome database (TCGA: COAD) <https://portal.gdc.cancer.gov>. The RT-qPCR data are available from the corresponding author upon request.

## Ethics Approval and Consent to Participate

The study was approved by the ethical committee of Hainan General Hospital (NO. 【2022】 719). Written informed consent was obtained from individual. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

## Acknowledgments

We would like to gratefully acknowledge Guangzhou Huayin Medical Laboratory Center. Ltd. for the kind technical help.

## Funding

This research was supported by the Natural Science Foundation of Hainan (Grants No. 823RC562).

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–249. doi:10.3322/caac.21660
2. Grady WM, Markowitz SD. The molecular pathogenesis of colorectal cancer and its potential application to colorectal cancer screening. *Dig Dis Sci.* 2015;60(3):762–772. doi:10.1007/s10620-014-3444-4
3. Long J, Guan P, Hu X, et al. Natural polyphenols as targeted modulators in colon cancer: molecular mechanisms and applications. *Front Immunol.* 2021;12:635484. doi:10.3389/fimmu.2021.635484
4. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet.* 2019;394(10207):1467–1480. doi:10.1016/S0140-6736(19)32319-0
5. Okugawa Y, Grady WM, Goel A. Epigenetic alterations in colorectal cancer: emerging biomarkers. *Gastroenterology.* 2015;149(5):1204–1225 e12. doi:10.1053/j.gastro.2015.07.011
6. Jung G, Hernandez-Illan E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat Rev Gastroenterol Hepatol.* 2020;17(2):111–130. doi:10.1038/s41575-019-0230-y
7. Muller D, Györfy B. DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. *Biochim Biophys Acta Rev Cancer.* 2022;1877(3):188722. doi:10.1016/j.bbcan.2022.188722

8. Pan JH, Zhou H, Cooper L, et al. LAYN is a prognostic biomarker and correlated with immune infiltrates in gastric and colon cancers. *Front Immunol*. 2019;10:6. doi:10.3389/fimmu.2019.00006
9. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007
10. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–1573. doi:10.1093/bioinformatics/btq170
11. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. *Methods Mol Biol*. 2020;2117:135–157. doi:10.1007/978-1-0716-0301-7\_7
12. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
13. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*. 2001;25(4):402–408. doi:10.1006/meth.2001.1262
14. Ben Ayed R, Rebai A. Tunisian table olive oil traceability and quality using SNP genotyping and bioinformatics tools. *Biomed Res Int*. 2019;2019:8291341. doi:10.1155/2019/8291341
15. Tomczak K, Czerwinski P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol*. 2015;19(1A):A68–77. doi:10.5114/wo.2014.47136
16. S X, Z C, F H, et al. ETV1 positively correlated with immune infiltration and poor clinical prognosis in colorectal cancer. *Front Immunol*. 2022;13:939806. doi:10.3389/fimmu.2022.939806
17. Villalobos C, Hernandez-Morales M, Gutierrez LG, Nunez L. TRPC1 and ORAI1 channels in colon cancer. *Cell Calcium*. 2019;81:59–66. doi:10.1016/j.ceca.2019.06.003
18. Liu X, Zhang P, Xie C, et al. Activation of PTEN by inhibition of TRPV4 suppresses colon cancer development. *Cell Death Dis*. 2019;10(6):460. doi:10.1038/s41419-019-1700-4
19. Li X, Wei R, Wang M, et al. MGP promotes colon cancer proliferation by activating the NF- $\kappa$ B pathway through upregulation of the calcium signaling pathway. *Mol Ther Oncolytics*. 2020;17:371–383. doi:10.1016/j.omto.2020.04.005
20. Makkar S, Riehl TE, Chen B, et al. Hyaluronic acid binding to TLR4 promotes proliferation and blocks apoptosis in colon cancer. *Mol Cancer Ther*. 2019;18(12):2446–2456. doi:10.1158/1535-7163.MCT-18-1225
21. R S, K TK. Gene regulatory networks and epigenetic modifications in cell differentiation. *IUBMB Life*. 2014;66(2):100–109. doi:10.1002/iub.1249
22. Islas JF, Quiroz-Reyes AG, Delgado-Gonzalez P, et al. Cancer stem cells in tumor microenvironment of adenocarcinoma of the stomach, colon, and rectum. *Cancers*. 2022;14(16):3948. doi:10.3390/cancers14163948
23. Fukushima Y, Oshika Y, Tsuchida T, et al. Brain-specific angiogenesis inhibitor 1 expression is inversely correlated with vascularity and distant metastasis of colorectal cancer. *Int J Oncol*. 1998;13(5):967–970. doi:10.3892/ijo.13.5.967
24. Zhu D, Hunter SB, Vertino PM, Van Meir EG. Overexpression of MBD2 in glioblastoma maintains epigenetic silencing and inhibits the antiangiogenic function of the tumor suppressor gene BAI1. *Cancer Res*. 2011;71(17):5859–5870. doi:10.1158/0008-5472.CAN-11-1157
25. Nam D-H, Park K, Suh YL, Kim J-H. Expression of VEGF and brain specific angiogenesis inhibitor-1 in glioblastoma: prognostic significance. *Oncol Rep*. 2004;11(4):863–869.
26. Wang C, Yu G, Liu J, et al. Downregulation of PCDH9 predicts prognosis for patients with glioma. *J Clin Neurosci*. 2012;19(4):541–545. doi:10.1016/j.jocn.2011.04.047
27. Williams CS, Bernard JK, Demory Beckler M, et al. ERBB4 is over-expressed in human colon cancer and enhances cellular transformation. *Carcinogenesis*. 2015;36(7):710–718. doi:10.1093/carcin/bgv049
28. Gao X, Zhang S, Zhu Z. Genetic variation of ErbB4 confers risk of colorectal cancer in a Chinese Han population. *Cancer Biomark*. 2014;14(6):435–439. doi:10.3233/CBM-140420
29. Rokavec M, Justenhoven C, Schroth W, et al. A novel polymorphism in the promoter region of ERBB4 is associated with breast and colorectal cancer risk. *Clin Cancer Res*. 2007;13(24):7506–7514. doi:10.1158/1078-0432.CCR-07-0457
30. Sakurai-Yageta M, Masuda M, Tsuboi Y, Ito A, Murakami Y. Tumor suppressor CADM1 is involved in epithelial cell structure. *Biochem Biophys Res Commun*. 2009;390(3):977–982. doi:10.1016/j.bbrc.2009.10.088
31. Tang L, Yu S, Zhang Q, et al. Identification of hub genes related to CD4(+) memory T cell infiltration with gene co-expression network predicts prognosis and immunotherapy effect in colon adenocarcinoma. *Front Genet*. 2022;13:915282. doi:10.3389/fgene.2022.915282
32. Di Z, Zhou S, Xu G, et al. Single-cell and WGCNA uncover a prognostic model and potential oncogenes in colorectal cancer. *Biol Proced Online*. 2022;24(1):13. doi:10.1186/s12575-022-00175-x
33. Huang C, Zhao J, Zhu Z. Prognostic nomogram of prognosis-related genes and clinicopathological characteristics to predict the 5-year survival rate of colon cancer patients. *Front Surg*. 2021;8:681721. doi:10.3389/fsurg.2021.681721
34. Takahashi H, Oshi M, Yan L, Endo I, Takabe K. Gastric cancer with enhanced apical junction pathway has increased metastatic potential and worse clinical outcomes. *Am J Cancer Res*. 2022;12(5):2146–2159.
35. Lodhi T, Song YP, West C, Hoskin P, Choudhury A. Hypoxia and its modification in bladder cancer: current and future perspectives. *Clin Oncol*. 2021;33(6):376–390. doi:10.1016/j.clon.2021.03.001
36. Zhang N, Ng AS, Cai S, Li Q, Yang L, Kerr D. Novel therapeutic strategies: targeting epithelial-mesenchymal transition in colorectal cancer. *Lancet Oncol*. 2021;22(8):e358–e368. doi:10.1016/S1470-2045(21)00343-0
37. Wang J, Cai H, Liu Q, et al. Cinobufacini inhibits colon cancer invasion and metastasis via suppressing wnt/beta-catenin signaling pathway and EMT. *Am J Chin Med*. 2020;48(3):703–718. doi:10.1142/S0192415X20500354
38. Yu J, Liu D, Sun X, et al. CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/beta-catenin signaling via transactivation of GSK-3 $\beta$  and Axin2 expression. *Cell Death Dis*. 2019;10(1):26. doi:10.1038/s41419-018-1263-9
39. Trojani A, Di Camillo B, Bossi LE, et al. Identification of a candidate gene set signature for the risk of progression in IgM MGUS to smoldering/symptomatic Waldenstrom Macroglobulinemia (WM) by a comparative transcriptome analysis of B cells and plasma cells. *Cancers*. 2021;13(8). doi:10.3390/cancers13081837
40. Pu Y, Song Y, Zhang M, et al. GOLM1 restricts colitis and colon tumorigenesis by ensuring Notch signaling equilibrium in intestinal homeostasis. *Signal Transduct Target Ther*. 2021;6(1):148. doi:10.1038/s41392-021-00535-1

41. Li X, Liu W, Geng C, et al. Ginsenoside Rg3 suppresses epithelial-mesenchymal transition via downregulating notch-Hes1 signaling in colon cancer cells. *Am J Chin Med*. 2021;49(1):217–235. doi:10.1142/S0192415X21500129
42. Fang L, Yang Z, Zhang M, Meng M, Feng J, Chen C. Clinical characteristics and survival analysis of colorectal cancer in China: a retrospective cohort study with 13,328 patients from southern China. *Gastroenterol Rep*. 2021;9(6):571–582. doi:10.1093/gastro/goab048
43. Lv MY, Chen XJ, Chen JG, et al. Nomogram for predicting overall survival time of patients with stage IV colorectal cancer. *Gastroenterol Rep*. 2022;10:goac072. doi:10.1093/gastro/goac072
44. Huang JH, Liu HS, Hu T, et al. Elevated preoperative CA125 is associated with poor survival in patients with metastatic colorectal cancer undergoing primary tumor resection: a retrospective cohort study. *Gastroenterol Rep*. 2022;10:goac020. doi:10.1093/gastro/goac020
45. Liu T, Li C, Jin L, Li C, Wang L. The prognostic value of m6A RNA methylation regulators in colon adenocarcinoma. *Med Sci Monit*. 2019;25:9435–9445. doi:10.12659/MSM.920381
46. Xue W, Wu X, Wang F, Han P, Cui B. Genome-wide methylation analysis identifies novel prognostic methylation markers in colon adenocarcinoma. *Biomed Pharmacother*. 2018;108:288–296. doi:10.1016/j.biopha.2018.09.043
47. Zhou W, Lin J, Li Z, Li M, Fan D, Hong L. The Cancer Genome Atlas (TCGA) based m6A methylation-related genes predict prognosis in rectosigmoid cancer. *Medicine*. 2022;101(51):e32328. doi:10.1097/MD.00000000000032328

## Cancer Management and Research

Dovepress

### Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/cancer-management-and-research-journal>