



A methylation-based prognostic model predicts survival in patients with colorectal cancer

Shanyue Tan^{#^}, Weiwei Gui[#], Sumeng Wang, Chongqi Sun, Xian Xu, Lingxiang Liu

Department of Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Contributions: (I) Conception and design: L Liu; (II) Administrative support: L Liu; (III) Provision of study materials or patients: X Xu; (IV) Collection and assembly of data: S Wang, C Sun; (V) Data analysis and interpretation: S Tan, W Gui; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Lingxiang Liu. Department of Oncology, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China. Email: llxlu@163.com.

Background: To construct a model that could effectively predict the prognosis of colorectal cancer (CRC) by searching for methylated-differentially expressed genes (MDEGs).

Methods: We identified MDEGs through four databases from Gene Expression Omnibus (GEO) and annotated their functions via bioinformatics analysis. Subsequently, after adjusting for gender, age, and grading, multivariate Cox hazard analysis was utilized to select MDEGs interrelated with the prognosis of CRC, and LASSO analysis was utilized to fit the prediction model in the training set. Furthermore, another independent dataset was harnessed to verify the effectiveness of the model in predicting prognosis.

Results: In total, 252 hypomethylated and up-regulated genes and 132 hypermethylated and down-regulated genes were identified, 27 of which were correlated with the prognosis of CRC, and a 10-gene prognostic model was established after LASSO analysis. The overall survival rate could be effectively grouped into different risks by the median score of this model in the training set [risk ratio (HR) =2.27, confidence interval (95% CI), 1.69–3.13, $P=8.15\times 10^{-8}$], and the validity of its effect in predicting prognosis in CRC was verified in the validation dataset (HR =1.75, 95% CI, 1.15–2.70, $P=9.32\times 10^{-3}$).

Conclusions: Our model could effectively predict the overall survival rate of patients with CRC and provides potential application guidelines for its clinically personalized treatment.

Keywords: Colorectal cancer (CRC); methylation; prognostic model; overall survival time

Submitted Jun 08, 2021. Accepted for publication Jul 29, 2021.

doi: 10.21037/jgo-21-376

View this article at: <https://dx.doi.org/10.21037/jgo-21-376>

Introduction

Colorectal cancer (CRC) is the third leading cause of cancer-related mortality worldwide, ranking third in men and second in women among the most common cancers. Recent research shows that in the United States, about 53,200 people die of CRC each year, accounting for about 8% of all cancer-related mortality (1). Surgery is the most effective treatment for early CRC, while chemotherapy

drugs commonly used include oxaliplatin, fluorouracil, and irinotecan (2). In addition, tumor targeted therapy, which is based on the gene mutation of tumor cells for precise targeting is an emerging treatment (3). However, the efficacy of conventional therapies is limited. Although the incidence of CRC has decreased in recent years, a high mortality rate remains. On the one hand, most patients are already at an advanced stage when they are diagnosed

[^] ORCID: 0000-0001-8467-9951.

Table 1 Summary of data sets used

Datasets	Normal (n)	Tumor (n)	Platform
Methylation dataset			
GSE25062	29	125	Illumina HM27
GSE17648	22	22	Illumina HM27
Expression dataset			
GSE24514	34	15	Affymetrix U133 Plus 2
GSE21510	25	123	Affymetrix U133 Plus 2
Training dataset			
GSE39582		556	Affymetrix U133 Plus 2
Validation dataset			
TCGA		469	Illumina HiSeqV2

TCGA, The Cancer Genome Atlas.

with CRC, and the effect of treatment is much worse than that at early stage. On the other hand, there is a lack of personalized prognostic guidance to achieve a better precision treatment. So it is necessary to find new, effective, and reliable biomarkers for improving risk assessment and guiding individualized treatment for the disease.

A growing number of recent studies have shown that in addition to genetic mutation, epigenetic variations including DNA methylation, histone modification (acetylation and methylation), and non-coding RNA-mediated transcriptional regulation play an important role in the occurrence and development of CRC (4,5). Especially for methylation, they can inhibit a variety of tumor suppressor factors by hypermethylation their promoter regions, which ultimately leads to the occurrence of tumors (6,7). Abnormal DNA methylation influences CRC through the regulation and control of the expression of cancer-related genes (8,9). Some researchers suggest that as DNA methylation usually occurs in early cancer, it can provide biomarkers for early-stage cancer detection. Recently, hypermethylated tumor-suppressing genes and hypomethylated tumor-promoting genes have been found to be related to the positive transcriptional regulation of oncogenes in multiple cell processes (10,11), and numerous studies have found them in CRC. Some drugs targeting DNA methylation been used in the treatment of CRC such as 5-azacitidine, decitabine and zebularine, and achieved a great therapeutic effect (12).

However, as far as we know, no previous studies have focused on the function of methylated-differentially

expressed genes (MDEGs) in predicting the prognosis of CRC genome-wide. To establish a new predictive model based on MDEGs, we used R to conduct an overall analysis of the data derived from a research cohort from the Gene Expression Omnibus (GEO). The results may help predict prognosis in patients with CRC and benefit personalized treatment for the disease. We present the following article in accordance with the TRIPOD checklist (available at <https://dx.doi.org/10.21037/jgo-21-376>).

Methods

Data processing

As shown in *Table 1*, all the data sets and clinical information were derived from GEO (<https://www.ncbi.nlm.nih.gov/>) and TCGA (<https://cancergenome.nih.gov/>). From these, GSE24514 and GSE21510 were utilized to select the differentially expressed genes (DEGs), GSE25062 and GSE17648 were utilized to select differentially methylated genes (DMGs), and the methylation differential expression genes (MDEGs) were obtained by overlapping of the two. Furthermore, genes related to the prognosis of CRC were selected in GSE39582, which is referred to as the training set. Further, we also created our prediction model based on this. The CRC data of TCGA served as a validation set to verify the efficacy of the prognosis model. GSE24514 contains 34 cases of normal tissues and 15 cases of CRC tissues, while GSE21510 consists of 25 cases of normal tissues and 123 cases of CRC tissues and GSE39582 contains

556 cases of CRC patients. The GPL570 platform (Affymetrix Human Genome U133 plus 2.0 Array) was used to analyze gene expression profiles, and a robust multi-array averaging algorithm was used to preprocess the raw data generated by the Affymetrix platform. GSE25062 possesses 29 cases of normal tissue and 125 cases of CRC tissue, while GSE17648 consists of 22 pairs of CRC and adjacent normal tissue. The expression profiles of methylation were analyzed by GPL8490 (Illumina Human Methylation27BeadChip) platform, and the gene expression level of TCGA was determined by Illumina HiSeqV2 sequencing with standardized counting. The numerical value of gene expression of all sequencing platforms was transformed by Log₂ for subsequent analysis. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Identification of MDEGs

DEGs and DMGs were selected by means of Student's *t*-test as described above. The false discovery rate (FDR) correction of P value was performed by Benjamini-Hochberg method to reduce the high false positive rate caused by multiple comparisons and the screening criteria were FDR <0.05 (13). DMEGs refer to differential expression in tissues as well as differentially methylated genes. The hypomethylated and up-regulated genes were obtained by means of overlapping high expression gene sets and hypomethylated gene sets, and hypermethylated and down-regulated genes were obtained by the same method.

Function annotation of MDEGs

The Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of MDEGs were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) to explore their biological functions. The biological functions, biochemical process, and subcellular localization of MDEGs were roughly explored by GO enrichment analysis, and KEGG pathway enrichment was used to determine which pathways of MDEGs were principally enriched. The standard of significant enrichment was P<0.05.

Establishing a prognostic prediction model based on MDEGs

The prognostic model was constructed in the training set.

Multivariate Cox hazard analysis was used to preliminarily select genes that highly correlated with the overall survival time of CRC, and LASSO (least absolute shrinkage and selection operator) regression was used to filter variables and reduce the complexity of the model. While variable screening involves placing all variables into the model, we chose to place limited variables for fitting to obtain a better performance parameter. LASSO is based on the penalty method to select variables and compresses the original coefficients, while the originally small coefficients are directly compressed to 0, so that the variables corresponding to these coefficients are regarded as invalid variables. The complexity of the linear model is directly related to the number of variables, so controlling it can avoid overfitting. By multiplying the expression value of each selected gene and its corresponding risk coefficient and adding these together, we established a risk prediction model. This formula was then executed in the training set to calculate the risk score of each person suffering from CRC. The median of the risk score was used as the segmentation point, which saw patients with scores higher than the median regarded as high-risk groups, and those with scores lower than the median regarded as low-risk groups.

Statistical methods

All analyses were carried out in R (Version 3.6.1). After adjusting for gender, age and stage, multivariate cox regression analysis was used to explore the predictive effect of the risk prognostic model on CRC, and the risk ratios (HRs) and 95% confidence intervals (95% CIs) were also executed. Kaplan-Meier method was employed to draw the survival curve, and the log-rank test was performed to calculate the P value, which was statistically significant at P<0.05.

Results

Screening of MDEGs

The screening conditions for each step of MDEGs were FDR <0.05, and the detailed process for screening is illustrated in *Figure 1*. Firstly, 5,555 and 18,756 genes differentially expressed in tumor tissues were screened from GSE24514 and GSE21510, respectively; 4,967 DEGs were obtained by overlapping the results of these two datasets while filtering genes with inconsistent expression directions. Thereinto, 2,515 genes were highly expressed while 2,452 genes were lowly expressed in CRC tissues. Secondly,

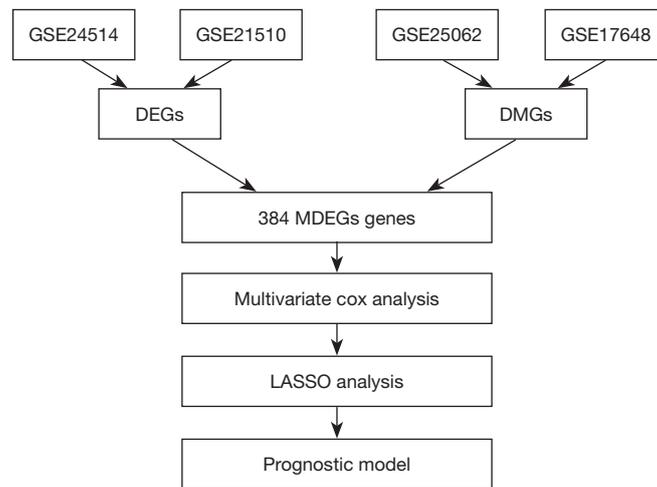


Figure 1 Flow chart of constructing our prognostic risk model. DEGs, differentially expressed genes; DMGs, differentially methylated genes; MDEGs, methylation-differential expression genes.

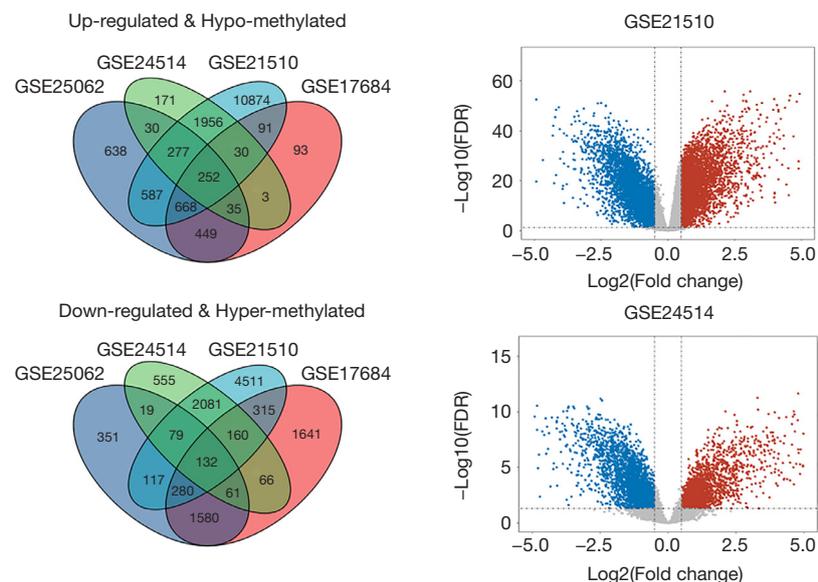


Figure 2 Venn diagram of the expression datasets and the methylation datasets. Volcano of the two expression datasets. Blue represents down-regulated genes and red represents up-regulated genes (FDR < 0.05). FDR, false discovery rate.

5,441 and 5,473 methylation differentially expressed genes were screened from GSE25062 and GSE17648, respectively; 3,457 DMGs were obtained by overlapping the results of these two datasets while filtering genes with inconsistent expression directions similarly, there were 2,053 hypermethylated genes and 1,404 hypomethylated genes among them. Finally, by associating RNA expression with DNA methylation, we identified 384 MDEGs, including 252 hypomethylated-highly expressed genes and

132 hypermethylated-lowly expressed genes (Figure 2). In order to test whether the FDR value was reasonable and to further visualize it, we constructed two representative volcanos based on the expression spectra of GSE24514 and GSE21510 (Figure 2).

Function prediction of MDEGs

The biological functions of MDEGs were annotated by

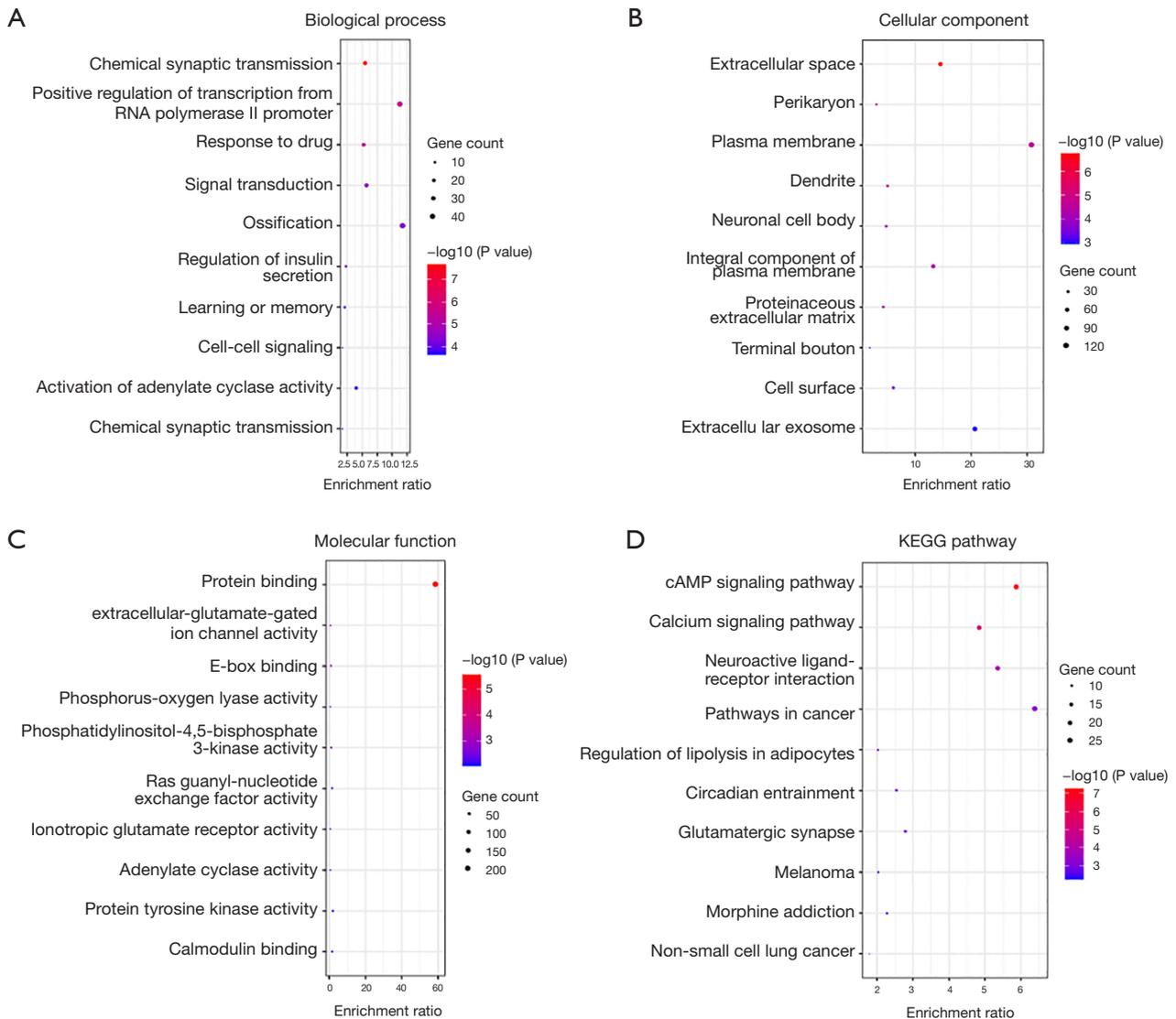


Figure 3 Bubble chart of the top 10 results of gene enrichment analysis. (A) Biological process enrichment; (B) cytological component enrichment; (C) molecular biological function enrichment; (D) KEGG pathway enrichment. KEGG, Kyoto Encyclopedia of Genes and Genomes.

bioinformatics methods including KEGG enrichment analysis and GO enrichment analysis, the latter which is composed of three parts primarily: Molecular Function, Biological Process, and Cellular Component. The data of enrichment analysis were derived from the DAVID database, and as shown in *Figure 3*, the enrichment results were visualized by bubble plot. The biological processes of MDEGs were primarily enriched in “chemical synaptic transmission”, “signal transduction”, and “positive regulation of transcription from RNA

polymerase II promoter”. As for the cellular component, MDEGs enriched most in “plasma membrane”, followed by “extracellular space”. In terms of molecular function, most MDEGs were enriched in “protein binding”, indicating that MDEGs may play critical biological roles by regulating the transcription of CRC. Furthermore, KEGG enrichment analysis speculated that MDEGs were mainly enriched in “cAMP signaling pathway”, “neuroactive ligand-receptor interaction”, and “pathways in cancer”.

Table 2 Multivariate COX regression analysis of the prognostic model in the training set and validation set

Variables	GSE39582			TCGA		
	HR	95% CI	P	HR	95% CI	P
Age (>65 vs. ≤65)	1.75	1.29–2.37	3.32×10^{-4}	3.08	1.91–4.98	4.20×10^{-6}
Risk (high-risk vs. low-risk)	2.27	1.69–3.13	8.15×10^{-8}	1.75	1.15–2.70	9.32×10^{-3}
Gender (male vs. female)	1.50	1.12–2.01	5.07×10^{-3}	1.00	0.66–1.53	9.92×10^{-1}
Stage	1.96	1.60–2.39	6.69×10^{-11}	2.04	1.59–2.63	3.20×10^{-8}

Establishment of CRC prognosis model

The variable selection and final establishment of our prognosis model were performed in the training set, and multivariate Cox regression analysis and LASSO regression analysis were utilized in turn to analyze the prognosis efficacy of 384 MDEGs. Firstly, after adjusting gender, age, and staging, multivariate Cox regression analysis explored the correlation between MDEGs and the overall survival time of CRC. With P value less than or equal to 0.05 as the standard, 27 genes were preliminary screened to relation to the prognosis of CRC. Subsequently, by using R for LASSO regression analysis to control the number of variables in the model, we further identified key genes for predicting prognosis. After multiple cross-validation, we determined that when the number of variables was 10, the parameter λ of the model was optimal. Thus, combined with the best regression coefficient under λ , a prognosis predictive model for CRC consisting of 10 genes was established, and the calculation formula for our risk score is listed below: risk score = $(0.033 \times \text{SPP1} + 0.048 \times \text{CLDN1} + 0.052 \times \text{COLEC12} + 0.062 \times \text{PTPRZ1} + 0.083 \times \text{SCARA3} + 0.108 \times \text{EDAR} + 0.136 \times \text{SOCS3} + 0.185 \times \text{SYNGR1} + 0.453 \times \text{ARMCX4} + 0.478 \times \text{MMP16})$.

Prognostic analysis of the prognostic risk model in the training set

The risk scores for each patient in the training set were calculated using our prognosis predicting model. Patients were arranged in descending order according to the risk score value, and those whose risk score was higher than the median were placed as a high-risk group, and those whose score was lower than the median were placed as a low-risk group. Multivariate Cox regression analysis was used to explore the relationship between different risk groups and overall survival of CRC, and included gender,

age, and grading as covariates. The results showed that the prognostic risk model had an independent predictive effect on the prognosis of CRC (HR =2.27, 95% CI, 1.69–3.13, $P=8.15 \times 10^{-8}$) (Table 2). Furthermore, we utilized log rank test to detect whether the overall survival of CRC between the two risk groups was statistically different, and Kaplan-Meier method was used to draw the survival curve. As shown in Figure 4A, the prognosis of patients in the high-risk group was worse than that in the low-risk group ($P<0.0001$). Moreover, people in the high-risk group had a higher mortality rate compared to those in the low-risk group, which was 45.33% and 21.72%, respectively (Figure 4B). Figure 4C displays the risk score distribution, survival status distribution, and heatmap of 10 gene expression profiles of each patient in the training set.

Verification in the validation set

The efficacy verification of the prognosis model was carried out in the TCGA database, which contained the gene transcription and prognosis information of 469 CRC patients. Based on the calculation formula of our risk model, we calculated the risk score of each patient in the validation set, and divided patients into different risk groups considering the median of the risk score. Multivariate cox regression analysis in the validation set confirmed that the risk group was an independent predictor of the overall survival of CRC (HR =1.75, 95% CI, 1.15–2.70, $P=9.32 \times 10^{-3}$) (Table 2). The survival curve indicated that there was a large distinction in the prognosis between the different risk groups, and the overall survival time of the high-risk group was shorter ($P=0.0013$) (Figure 5A). In addition, the mortality rate of the low-risk group was 18.48%, while that of the high-risk group was 28.91% (Figure 5B). The risk score distribution of each person in the validation set is displayed as well as the distribution of survival status to distinguish mortality changes in different

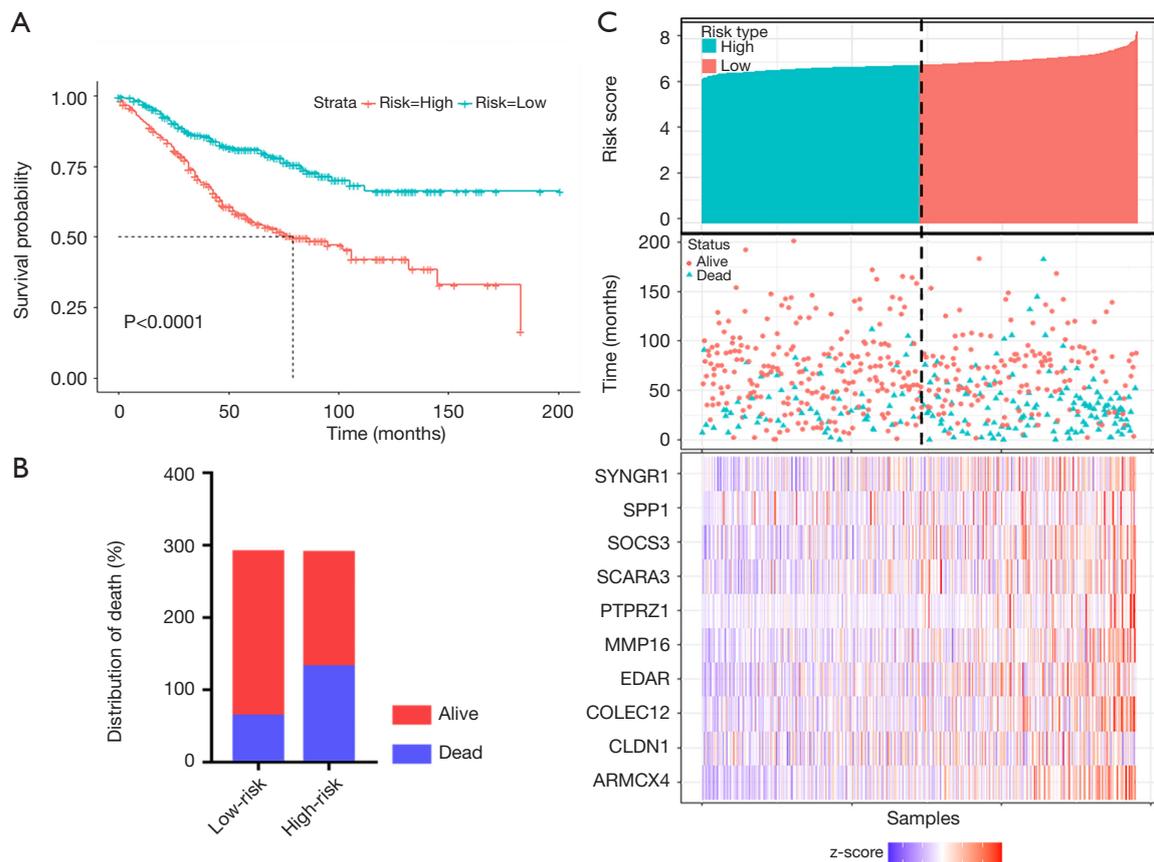


Figure 4 Prognostic model established in the training set. Patients were divided into high-risk groups and low-risk groups according to the median risk score. (A) Kaplan-Meier survival curve of different risk groups between the training set; (B) survival distribution of patients in the high-risk group and the low-risk group; (C) distribution of risk value and survival status and heat map of the 10 genes contained in our risk model.

risk groups intuitively. The expression of the corresponding 10 genes of our model is exhibited by the heat map, which shows the prognosis model can effectively classify patients in the validation set into a high-risk group and low-risk group (Figure 5C).

Discussion

DNA methylation is a common epigenetic modification regulating gene expression (14). Methylation is an unattractive small modification on the C-site of DNA sequence, which sees a methyl group added to the 5th carbon atom of the cytosine ring. This can cause the development of, metastasis, and deterioration of tumors in various ways (15): (I) the high frequency of deamination of cytosine in methylated CpG island dinucleotides

to thymine, resulting in gene mutations; (II) tumor suppressor genes and DNA repair genes are silenced due to hypermethylation; (III) oncogenes are activated due to lower methylation levels; (IV) the decrease in the overall methylation level of the genome causes the activation of transposons and repetitive sequences, resulting in a decrease in chromosome stability (16). The importance of gene methylation in genetics, including carcinogenesis, has been researched in a variety of cancers including that of the lung, breast, liver and CRC. While DNA methylation patterns could potentially provide a reference for treatment options and serve as prognostic biomarkers for cancers (17), there has been no exploration of the prognostic value of MDEGs for CRC patients in the clinical setting. This study is the first to establish a risk prognosis model based on MDEGs in CRC genome-wide.

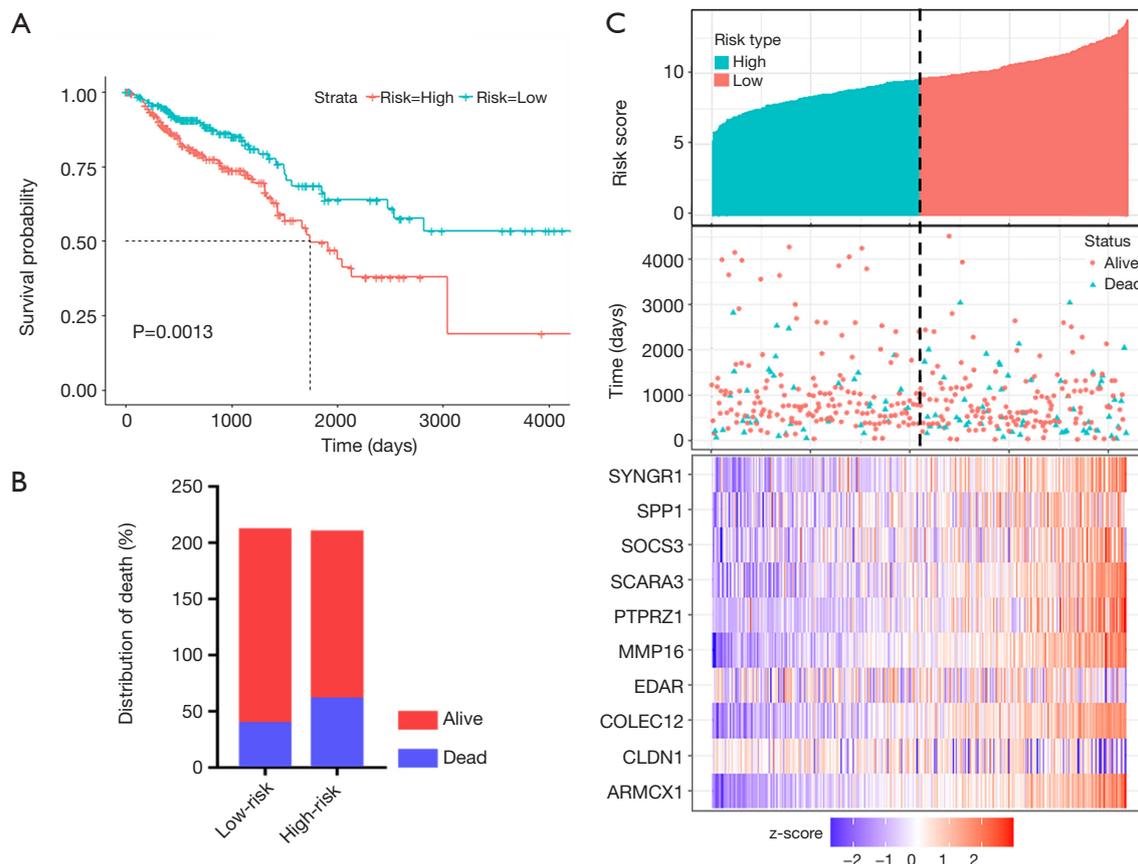


Figure 5 Test of the prognostic model in the validation set. (A) Kaplan-Meier overall survival curves; (B) the survival distribution of patients in different risk groups; (C) the distribution of the risk value in the validation set of each patient, and the corresponding survival status and heat map of the expression of our 10 genes.

We used the gene expression and methylation expression data set in GEO to screen out 252 hypomethylated and highly-expressed genes and 132 hypermethylated and lowly-expressed genes in CRC tissue, totaling 384 MDEGs. We also conducted a bioinformatics analysis of MDEGs to predict their possible biological functions and found “chemical synaptic transmission”, “signal transduction”, and “positive regulation of transcription from RNA polymerase II promoter” were the primary biological process in which MDEGs were most enriched in. A study in the journal of Nature in 2019 showed that brain metastasis from breast cancer were caused by interfering with synaptic conduction, thereby accelerating its growth and lethal rate (18). Dysregulation of physiological signal transduction is the basis of tumorigenesis, and multiple signal transductions (19,20) play a critical role in the occurrence and development of CRC. In recent years, Drugs targeting

signal transduction such as curcumin, bortezomib have already made great contributions to the treatment of CRC (21,22). The reduction of differentiation ability and the enhancement of proliferation ability are characteristic of tumor cells, and a study on acute leukemia by researchers at Yale University found that excessive cell proliferation could directly lead to cancer (23). As MDEGs in cellular components are mainly enriched in the cytoplasm, there is speculation that they may regulate tumor progression by participating in the post-transcriptional modification of coding genes. The most enriched molecular biological function of MDEGs is protein binding, and a variety of proteins can bind to specific genes to regulate gene expression, including BAHCC1 protein, which can mediate gene silencing and tumorigenesis by binding to H3K27me3 (24). The cAMP signaling pathway is one of the five intracellular communication pathways. After

receiving extracellular signals, the G protein-coupled receptor conducts the signal to change the level of the second messenger cAMP and transmit extracellular signals to intracellular, and previous studies have confirmed its relationship with CRC. Activating the cAMP pathway in CRC can also inhibit tumor angiogenesis and inhibit tumor growth (25).

Following the prognostic analysis of 384 genes, we screened 10 MDEGs including SPP1, CLDN1, COLEC12, PTPRZ1, SCARA3, EDAR, SOCS3, SYNGR1, ARMCX4, and MMP16 to construct a prognostic risk model. It is worth mentioning that specific pathological type of colorectal cancer is lacking in GEO. We did not perform stratified analysis in colorectal cancer base on it, so we cannot sure this study have high prognostic value for all pathological types of colorectal cancer. Previous studies have reported that almost all of the above-mentioned genes play a potential regulatory role in the process of tumor carcinogenesis, including SPP1, which is up-regulated in CRC and is associated with its poor prognosis. SPP1 can also promote the metastasis of CRC by promoting epithelial-mesenchymal transition (26,27). Research using 50 pairs of CRC tissues revealed that the rs17501976 polymorphism on CLDN1 could decrease the risk of CRC in a Chinese population (28). The expression of PTPRZ1 is closely related to the KRAS mutation in CTC, which plays a vital role and exists widely in its occurrence (29). EDAR, which is up-regulated in CRC and has been confirmed to be a component of the Wnt/ β -catenin signaling pathway, can also stimulate the proliferation of CRC *in vitro* (30). *In vivo* experiments have indicated SOCS3 is poorly expressed in CRC tissues, and its lower expression is likely to indicate lymph node metastasis and a worse clinical prognosis. Furthermore, overexpression of SOCS3 in CRC cells could inhibit cell proliferation, migration, and invasion (31). Genome-wide association studies have found that the G allele mutation of single nucleotide mutation site rs6509 can down-regulate the expression of SYNGR1 and reduce the risk of ovarian cancer (32). Silencing the expression of MMP16 can restrain the migration and invasion of colon cells, and the up regulation of MMP16 is highly correlated with aggressive behavior in patients and poor survival in CRC (33). The aforementioned studies provide a theoretical basis for the prediction of the risk model constructed by us on the prognosis of CRC. Our findings provide new insights into the treatment of CRC and have the potential to be transformed into clinical drugs for the treatment of colorectal cancer.

Conclusions

Our research combined gene mRNA expression and DNA methylation modification to screen out MDEGs. We further established a CRC prognostic risk model consisting of 10 key genes, which could effectively divide patients with CRC into high-risk and low-risk groups. Our prognostic model plays a role in assessing the prognosis of CRC, and may provide a principle for individualized clinical treatment.

Acknowledgments

Funding: This study was partly supported by the Natural Science Foundation of Jiangsu Province (BK20181371 to LL) and National Natural Science Foundation of China (Grant No.82003094 to CS).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://dx.doi.org/10.21037/jgo-21-376>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/jgo-21-376>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Institutional ethical approval and informed consent were waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal

- cancer statistics, 2020. *CA Cancer J Clin* 2020;70:145-64.
2. Gustavsson B, Carlsson G, Machover D, et al. A review of the evolution of systemic chemotherapy in the management of colorectal cancer. *Clin Colorectal Cancer* 2015;14:1-10.
 3. Franke AJ, Skelton WP, Starr JS, et al. Immunotherapy for Colorectal Cancer: A Review of Current and Novel Therapeutic Approaches. *J Natl Cancer Inst* 2019;111:1131-41.
 4. Ushijima T, Asada K. Aberrant DNA methylation in contrast with mutations. *Cancer Sci* 2010;101:300-5.
 5. Calcagno DQ, Gigeck CO, Chen ES, et al. DNA and histone methylation in gastric carcinogenesis. *World J Gastroenterol* 2013;19:1182-92.
 6. Kim MS, Lee J, Sidransky D. DNA methylation markers in colorectal cancer. *Cancer Metastasis Rev* 2010;29:181-206.
 7. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* 2011;8:686-700.
 8. Tse JWT, Jenkins LJ, Chionh F, et al. Aberrant DNA Methylation in Colorectal Cancer: What Should We Target? *Trends Cancer* 2017;3:698-712.
 9. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. *Gastroenterology* 2015;149:1204-1225.e12.
 10. Bi H, Liu Y, Pu R, et al. CHST7 Gene Methylation and Sex-Specific Effects on Colorectal Cancer Risk. *Dig Dis Sci* 2019;64:2158-66.
 11. Gu S, Lin S, Ye D, et al. Genome-wide methylation profiling identified novel differentially hypermethylated biomarker MPPED2 in colorectal cancer. *Clin Epigenetics* 2019;11:41.
 12. Jung G, Hernández-Illán E, Moreira L, et al. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat Rev Gastroenterol Hepatol* 2020;17:111-30.
 13. Gossmann A, Zille P, Calhoun V, et al. FDR-Corrected Sparse Canonical Correlation Analysis With Applications to Imaging Genomics. *IEEE Trans Med Imaging* 2018;37:1761-74.
 14. Kanai Y, Hirohashi S. Alterations of DNA methylation associated with abnormalities of DNA methyltransferases in human cancers during transition from a precancerous to a malignant state. *Carcinogenesis* 2007;28:2434-42.
 15. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell* 2012;150:12-27.
 16. Köhler F, Rodríguez-Paredes M. DNA Methylation in Epidermal Differentiation, Aging, and Cancer. *J Invest Dermatol* 2020;140:38-47.
 17. Lee JJ, Geli J, Larsson C, et al. Gene-specific promoter hypermethylation without global hypomethylation in follicular thyroid cancer. *Int J Oncol* 2008;33:861-9.
 18. Zeng Q, Michael IP, Zhang P, et al. Synaptic proximity enables NMDAR signalling to promote brain metastasis. *Nature* 2019;573:526-31.
 19. Koni M, Pinnarò V, Brizzi MF. The Wnt Signalling Pathway: A Tailored Target in Cancer. *Int J Mol Sci* 2020;21:7697.
 20. Xu H, Liu L, Li W, et al. Transcription factors in colorectal cancer: molecular mechanism and therapeutic implications. *Oncogene* 2021;40:1555-69.
 21. Cheng X, Xu X, Chen D, et al. Therapeutic potential of targeting the Wnt/ β -catenin signaling pathway in colorectal cancer. *Biomed Pharmacother* 2019;110:473-81.
 22. Patel M, Horgan PG, McMillan DC, et al. NF- κ B pathways in the development and progression of colorectal cancer. *Transl Res* 2018;197:43-56.
 23. Chen X, Burkhardt DB, Hartman AA, et al. MLL-AF9 initiates transformation from fast-proliferating myeloid progenitors. *Nat Commun* 2019;10:5767.
 24. Fan H, Lu J, Guo Y, et al. BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat Genet* 2020;52:1384-96.
 25. Wang S, Zhang Z, Qian W, et al. Angiogenesis and vasculogenic mimicry are inhibited by 8-Br-cAMP through activation of the cAMP/PKA pathway in colorectal cancer. *Onco Targets Ther* 2018;11:3765-74.
 26. Xu C, Sun L, Jiang C, et al. SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. *Biomed Pharmacother* 2017;91:1167-77.
 27. Choe EK, Yi JW, Chai YJ, et al. Upregulation of the adipokine genes ADIPOR1 and SPP1 is related to poor survival outcomes in colorectal cancer. *J Surg Oncol* 2018;117:1833-40.
 28. Chen JJ, Zhong M, Dou TH, et al. rs17501976 polymorphism of CLDN1 gene is associated with decreased risk of colorectal cancer in a Chinese population. *Int J Clin Exp Med* 2015;8:1247-52.
 29. Laczmanska I, Karpinski P, Gil J, et al. High PTPRQ Expression and Its Relationship to Expression of PTPRZ1 and the Presence of KRAS Mutations in Colorectal Cancer Tissues. *Anticancer Res* 2016;36:677-81.
 30. Wang B, Liang Y, Chai X, et al. Ectodysplasin A receptor (EDAR) promotes colorectal cancer cell proliferation via regulation of the Wnt/ β -catenin signaling pathway. *Exp Cell Res* 2020;395:112170.
 31. Chu Q, Shen D, He L, et al. Prognostic significance of

- SOCS3 and its biological function in colorectal cancer. *Gene* 2017;627:114-22.
32. Yodsurang V, Tang Y, Takahashi Y, et al. Genome-wide association study (GWAS) of ovarian cancer in Japanese predicted regulatory variants in 22q13.1. *PLoS One* 2018;13:e0209096.
33. Wu S, Ma C, Shan S, et al. High expression of matrix metalloproteinases 16 is associated with the aggressive malignant behavior and poor survival outcome in colorectal carcinoma. *Sci Rep* 2017;7:46531.

Cite this article as: Tan S, Gui W, Wang S, Sun C, Xu X, Liu L. A methylation-based prognostic model predicts survival in patients with colorectal cancer. *J Gastrointest Oncol* 2021;12(4):1590-1600. doi: 10.21037/jgo-21-376