BMC Cancer

RESEARCH

Open Access

DNA methylation analysis of the SDC2, SEPT9 and VIM genes in fecal DNA for colorectal cancer diagnosis

Yue Liu[1], Hongbo Ming[1], Lizhi Xu[1], Lizhen Li[1], Qi Liu[1], Jinyin Zhao[1], Cundi Zhong[2*] and Hongzhi Li[1*]

Abstract

Background Colorectal cancer is one of the most common cancers worldwide. DNA methylation sites may serve as a new gene signature for colorectal cancer diagnosis. The search for representative DNA methylation sites is urgently needed. This study aimed to systematically identify a methylation gene panel for colorectal cancer diagnosis via tissue and fecal samples.

Methods A total of 181 fecal and 50 tumor tissue samples were collected. They were obtained from 83 colorectal cancer patients and 98 healthy subjects. These samples were evaluated for DNA methylation of 9 target genes via quantitative bisulfite next-generation sequencing. We employed the rank-sum test to screen the colorectal cancer-specific methylation sites in the tissue and fecal cohorts. A data model was subsequently constructed and validated via the dedicated validation dataset.

Results Compared with the fecal and negative control samples, the colorectal cancer tissue samples presented significantly higher methylation rates for all the selected gene sites. The methylation rates of the tissue and preoperative fecal samples showed the same high and low rates at the same sites. After screening, a panel of 29 loci in the SDC2, SEPT9, and VIM genes proved to be reliable biomarkers for colorectal cancer diagnosis in fecal samples. Logistic regression models were then constructed and validated using this panel. The sensitivity of the model was 91.43% (95% CI = [89.69, 93.17]), the specificity was 100% (95% CI = [100,100]), and the AUC value is 99.31% (95% CI = [99,99.62]). The diagnostic accuracy of the model for stage I and stage II colorectal cancer was 100% (11/11) and 91.3% (21/23), respectively. Overall, this study confirms that the gene locus panel and the model can be used to diagnose colorectal cancer effectively through feces.

Conclusions Our study identified a set of key methylation sites for colorectal cancer diagnosis from fecal samples, highlighting the importance of using tissue and fecal samples to accurately assess DNA methylation levels to screen for methylation sites, and developing an effective diagnostic model for colorectal cancer.

Keywords Colorectal cancer, Targeted sequencing, DNA methylation analysis, Logistic regression

[†]Yue Liu and Hongbo Ming contributed equally to this work and should be considered co-first authors.

*Correspondence:
Cundi Zhong
zhongcundi@dmu.edu.cn
Hongzhi Li
lihzh@gentalker.com
Full list of author information is available at the end of the article

Background

The global incidence of colorectal cancer (CRC) continues to increase, with concurrent increases in mortality and morbidity rates [1]. The main reason for the poor prognosis is that the early symptoms of CRC are not obvious and are easily ignored. This may cause the patient to miss the optimal timing for treatment [2]. The combination of targeted early stage detection measures and corresponding therapeutic interventions can significantly contribute to reducing mortality [3, 4]. The genesis and advancement of CRC are fundamentally underpinned by a spectrum of genetic and epigenetic alterations within colonic epithelial cells [5]. Decreased expression or silencing of repressor genes caused by DNA promoter methylation appears to be a common event in CRC carcinogenesis [6]. Second, DNA methylation mutations can be reliably detected in the feces of CRC patients because of their strong chemical and biological stability and have been used to diagnose CRC [7]. Although the immunochemical fecal occult blood test (IFOBT) is a method commonly used in clinical practice, it screens CRC via specific immunodetection of human hemoglobin [8]. While the hemoglobin used in testing lacks stability, one needs to ensure that it comes from a tumor bleed rather than another disease disease [9]. IFOBT has a high false-positive rate in CRC detection and has difficulty distinguishing adenomas from carcinomas [10]. In addition, although colonoscopy and tissue biopsy are the gold standards for detecting CRC [11], colonoscopy is invasive and prone to bowel perforation and bleeding. Including minor complications, approximately 23% of patients have been reported to experience complications, including minor complications [12]. Thus, compliance with colonoscopy is poor [13]. The detection of DNA methylation in fecal samples has become a favorable choice for CRC diagnosis because of its convenient sampling, noninvasiveness, high sensitivity and specificity [14].

Many studies have shown that the detection of DNA methylation mutations in specific genes (BMP3, NDRG4, SDC2, SFRP2, TFPI2 and VIM) in feces is a promising approach for CRC diagnosis [15–18]. However, research on the integration of multiple fecal methylation gene mutation regions to locate the optimal gene site accurately in CRC diagnosis is lacking [19]. In our study, we combined methylation analysis and high-throughput sequencing technology. Nine genes were chosen from the Methy Cancer database (http://methycancer.psych.ac.cn/). These genes have been shown to be significantly associated with the development and progression of CRC. We analyzed tissue and fecal samples from CRC patients with methylation variants and calculated the methylation rate of each site to further accurately identify DNA hypermethylation sites. Finally, we constructed a model to determine the robustness of the selected sites for CRC detection to find the best biomarkers for CRC diagnosis.

Method

Clinical samples

All the subjects were re-examined and confirmed by professional pathologists for histopathological diagnosis. Anonymous numerical codes were employed to manage all the data pertaining to the study's human subjects carefully.

Collection and storage of samples

The DNA samples used in this study were obtained from fresh-frozen tissues and feces. All of the specimens were collected at the Second Affiliated Hospital of Dalian Medical University from March 2021 to May 2022. The tumor tissues were derived from patients who underwent receiving CRC resection. Feces were collected from patients with CRC and healthy volunteers. Approximately 5 g of feces was collected and stored in 50 mL tubes with 15 mL of preservative buffer (0.5 mol/L Tris, 0.15 mol/L EDTA, and 10 mmol/L NaCl, pH 9.0). These samples were immediately stored at -80 °C after collection.

We obtained feces from a subset of 83 CRC patients as well as their tumor tissue from some of them (Table 1). The feces of 98 healthy control subjects were also collected. Patients were verified through histology or colonoscopy. Healthy controls had to fulfill the following

Table 1 Clinicodemographic characteristics of the CRC patients enrolled in the study

pathological stage	Patients number(male/female)	Patients by age number				Median age
		< 50	50–59	60–69	≥ 70	
Stage I	13(9/4)	1	6	4	2	61
Stage II	31(17/14)	6	10	9	6	58
Stage III	36(25/11)	7	10	16	3	58
Stage IV	3(3/0)	0	1	2	0	62
Total	83(54/29)	14	27	31	11	60

requirements to be included in the study: (1) no notable medical history of chronic illnesses or other diseases such as cancer; (2) were willing to undergo a colonoscopy prior to fecal collection, and the results were normal.

Isolation of tissue and fecal genomic DNA

Tissue genomic DNA was extracted from fresh-frozen tissue specimens via a QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. With the help of a shaker, the feces were homogenized in preservation buffer. Fecal DNA was extracted via a Fine Mag fecal DNA extraction Kit (GENFINE BIOTECH (BEIJING) Co., Ltd.). To prevent DNA deterioration, repeated freezing and thawing of the feces were avoided. The DNA concentration was measured via a NanoDrop 2000 (Thermo Scientific, MA, USA). For the creation of sequencing libraries, DNA with a yield greater than 5 ng and no discernible genomic DNA contamination was used.

Bisulfite treatment and DNA purification

An EZ DNA Methylation-Gold Kit (ZYMO Research, CA, USA) was used for bisulfite treatment and DNA purification. Briefly, genomic DNA was treated with sodium bisulfite for 2.5 h at 65 °C after being denatured for 10 min at 98 °C. After DNA was added to a spin column, desulfonation buffer was added, and the mixture was then incubated for 20 min at room temperature. Then bisulfite-converted DNA was purified and eluted with 20 μL distilled water. The extracted DNA was either used immediately or stored at -20 °C. The Qubit® fluorescent dye method was used for quality control (QC) of the extracted DNA.

Selection of genes

We selected a set of nine common hypermethylated genes in CRC from the Methy Cancer database (http://methycancer.psych.ac.cn/), namely BMP3, SFRP2, SEPT9, VIM, SDC2, WNT2, MGMT, CDKN2A (P16), and NDRG4, and the detection region of each gene is shown in Supplement 1. BMP3 and SFRP2 were confirmed to be associated with CRC in thirty-one studies on the basis of fecal gene methylation [20]. SEPT9 methylation has been shown to be a plasma biomarker for CRC [21]. VIM, SDC2, WNT2 and MGMT are commonly used biomarkers for CRC and have been validated in numerous studies [22, 23]. CDKN2A was hypermethylated in 38% of CRC cases in previous studies [24]. NDRG4 is a candidate oncogene for CRC and its expression is often inactivated by promoter methylation [25].

Prediction of CpG islands and primer design

We identified cytosine-phosphate-guanine (CpG) islands by using genomic sequences stored in the Ensembl genome browser (http://www.ensembl.org/index.html). MethPrimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) was used to identify CpG and the optimal primers [26].

Targeted bisulfite sequencing

The library was prepared in two steps: First, multiplex polymerase chain reaction (PCR) amplification was carried out for target enrichment. Second, a second round of amplification with a low number of cycles was performed to allow the barcoding of the template-specific amplicons obtained from the first amplification step. Template -specific bisulfite amplicon libraries were generated with tagged primers via Phusion U DNA polymerase (Thermo Fisher, cod. F555L) (the primer sequence is available in Supplement 2). The amplification products were purified via Agencourt AMPure XP beads (Agencourt-Beckman Coulter, cod. A63881) was quantified with a Quantus™ fluorometer (Promega, cod. E6150), and then barcoding was performed using the Nextera™ index kit as previously described. Finally, the Illumina HiSeq 2000 sequencing platform was used for PE150 sequencing.

Data filtering and DNA methylation level calculation

Cutadapt 3.4 was used to cut the adapter sequences from the raw sequence reads. The reads generated by bisulfite sequencing were fed into Bismark software along with options. We used the Bismark_genome_preparation module with Bowtie2 2.4.4 to create preconverted versions of the reference, and we used the Bismark module's default option to align the read files to the reference genome. To verify reliability, we gathered the genomic locations for which we could identify > 90% of the methylation level with a supporting depth > 10 among those that were extracted via Bismark_methylation_extractor.

Statistical analysis

To identify methylation gene loci related to the preoperative feces of CRC patients, we divided preoperative fecal samples and negative samples into positive and negative classes, respectively. Every site's methylation rate was calculated, and each site was subjected to a Wilcoxon signed-rank test. Sites with P values less than 0.05 were used to choose features. Additionally, we highlighted the absolute methylation change for each site's distinguishing power [27]. The following standards were used to find DNA methylation biomarkers unique to CRC:

1) The same trend was observed for CRC tissue and fecal samples compared with normal controls.
2) The mean methylation rate in CRC feces samples was significantly different from that in normal samples (adjusted $P < 0.05$), and the absolute shift was substantial (> 5%)

Multidimensional scaling (MDS) in the Sklearn 0.24.2 package was used to characterize selected gene loci in a Python 3.7 environment. The MDS results allowed us to classify the preoperative feces and negative samples, as well as observe the differences and similarities in methylation levels between the two sample types. Tenfold cross-validation was used to accurately assess our classification model's performance accurately during training and testing. We use the logistic regression algorithm from the Sklearn 0.24.2 library in Python 3.7 to construct the model, and the dataset is randomly divided into 10 subsets or folds, with an 8:2 ratio for the training and validation sets. This process was repeated 10 times. We calculated the statistical parameters of accuracy, including specificity, sensitivity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV), after each model was trained and tested. Ultimately, the model's performance metrics are the average of these metrics over ten training and validation sessions. Furthermore, we generated receiver operating characteristic (ROC) curves and computed the area under the curve (AUC) value to evaluate the classification performance of the model. All the data are presented as the means and standard deviations (SDs). Model accuracy statistics (Table 2) and calculations of the methylation rates are described below.

Specificity: The proportion of true negative class samples that were correctly identified by the classifier.

$$Specificity = d/(b + d)$$

Sensitivity: The proportion of true class samples that were correctly identified by the classifier.

$$Sensitivity = a/(a + c)$$

$$Methylated\ rate = Methylation/(Methylation + Unmethylated)$$

Table 2 Confusion matrix

		Standard method		total
		positive	negative	
candidate methods	positive	a	c	a+c
	negative	b	d	b+d
total		a+b	c+d	a+b+c+d

where, "Methylation" represents the number of methylated sites and "Unmethylated" represents the number of unmethylated sites.

Development and validation of the CRC fecal methylation diagnostic model

Based on the methylation rates of tissue, preoperative stool, and control samples and the principles of locus screening mentioned above, we ended up with 29 CRC-specific methylation loci (Supplement 3). The fecal methylation dataset consisted of healthy controls and CRC patients was then divided into training and validation cohorts at random. For the training cohort, we constructed CRC diagnostic models using the methylation profiles of these 29 loci as covariates in logistic regression models.

Results

Patient and sample characteristics

To study colorectal cancer-specific DNA methylation sites, we collected 181 feces specimens (98 normal feces and 83 CRC feces) and 50 tissue sections. After DNA extraction and library construction, a total of 215 DNA samples passed QC and underwent DNA methylation second-generation sequencing (NGS). Sixteen samples failed extraction and construction of gene libraries's QC, including 4 cases of low DNA yield and 12 cases of low library yield. And another 3 samples did not meet sequencing QC. Ultimately, 212 samples (96 normal feces, 67 colorectal cancer feces, and 49 CRC tissues) were included in the analysis to discover CRC-specific DNA methylation biomarkers. To explore the clinical application of the detection of DNA methylation sites in the diagnosis of CRC, we used preoperative feces from CRC patients and feces from healthy subjects for model development and validation (Fig. 1).

Characterization of methylation sites specific to CRC

Methylation rates were calculated for all selected CpG sites in tissue, fecal, and control samples, and the line graphs are shown in Fig. 2. Our observations showed a consistent trend in methylation rates for most loci between tissue and feces. Specifically, when a site exhibited a higher methylation rate in tissue, the same trend was observed in feces. Similarly, when a locus exhibited a lower methylation rate in tissue, a similar pattern was observed in feces. In general, the average methylation rate in tissues was higher than that in feces at almost all selected locus, which was consistent with our expectation that shed tumor cells could be detected in feces after defecation. After significance testing and calculation of absolute differences in methylation rates, 4 CpG sites in the SDC2 gene, 19 CpG

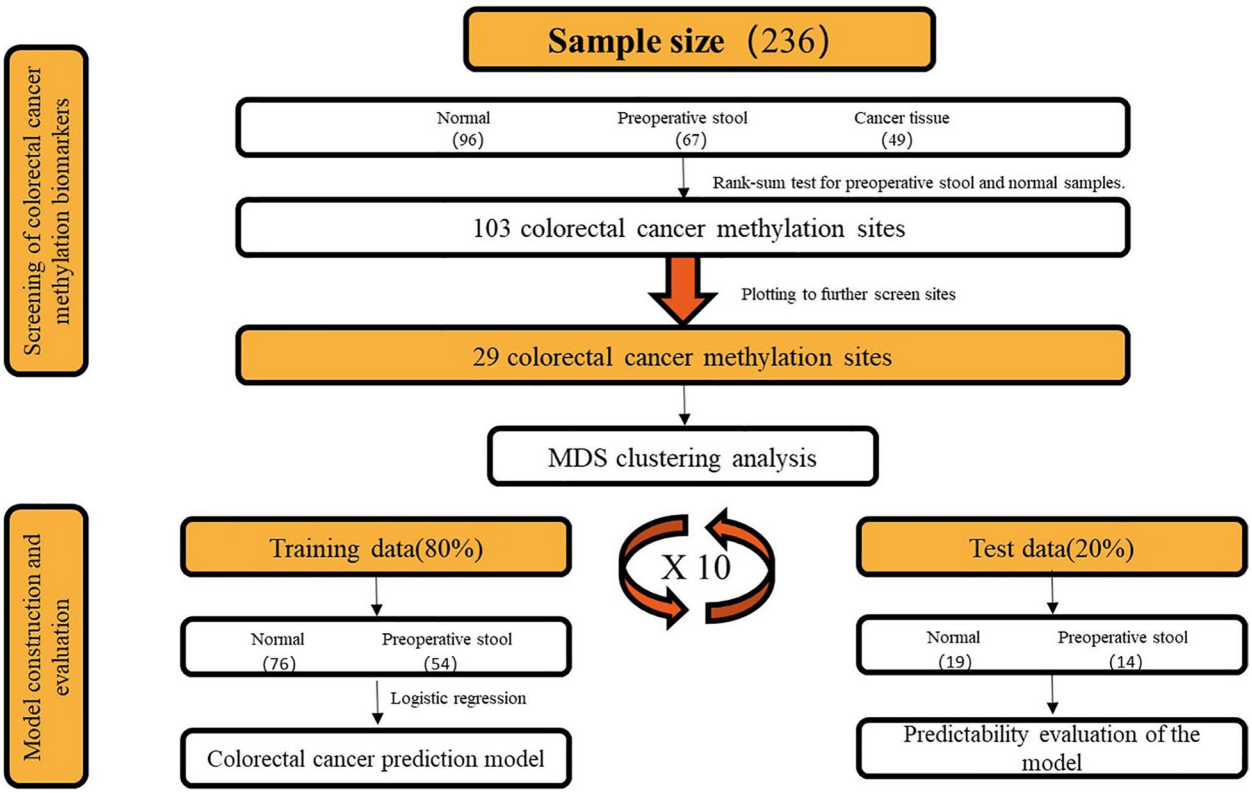

Fig. 1 The study workflow chart. In the DNA methylation sequencing phase, 96 normal fecal samples, 67 preoperative CRC fecal samples and 49 tissue samples were collected for next-generation sequencing (NGS) and analyzed. A rank-sum test was used to screen the CRC-specific methylation sites in the tissue and fecal cohorts, which led to the discovery of 103 DNA methylation sites. The results were subsequently plotted to screen sites, and 29 methylation sites were ultimately identified. Twenty-nine CRC-specific methylation sites were used in the training cohort, which were then further confirmed in the validation cohort

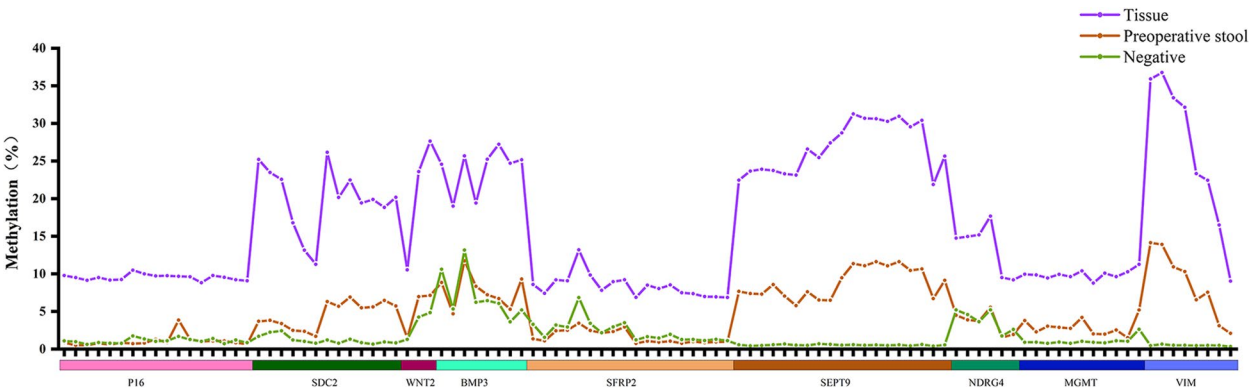

Fig. 2 Methylation rates of genes in tissues, fecal and controls. All genes and their methylation sites are represented on the abscissa; the ordinate is the methylation rate; the points on the broken lines represent the means of all samples

sites in the SEPT9 gene, and 6 CpG sites in the VIM gene met the screening criteria. A total of 29 methylation sites were identified for the construction of the CRC fecal detection model.

Development and validation of the CRC methylation diagnostic model

After screening for CRC-specific methylation sites in feces, 95 normal and 68 CRC fecal samples were randomly assigned to the training and validation cohorts

respectively. CpG loci of selected genes SEPTIN9, SDC2 and VIM were analyzed in the sample database to construct diagnostic model. The analysis showed that for these three genes, 100% of CRC tissue samples were hypermethylated in at least one gene, and the probability of hypermethylation in at least one of the genes was also 100% in fecal samples. Notably, in both tissue and fecal samples, these three genes were hypermethylated in patients and hypomethylated in controls (Fig. 3).

A logistic regression model for the diagnosis of CRC based on DNA methylation levels in fecal samples was developed based on these 29 methylation sites. The formula for the regression model was $\pi(x) = \frac{1}{1+e^{-w^T x+b}}$ (x: methylation rate of gene loci; w: coefficient; b: intercept). The relevant parameters are shown in Supplement 4 of the Supplementary file 1. In the training set, the model had a sensitivity of 94. 28% (95% CI = [91.55,97.01]), a specificity of 100% (95% CI = [100,100]), an AUC value of 100% (95% CI = [100,100]), a PPV value of 100% (95% CI = [100,100]), and an NPV value of 96. 11% (95% CI = [94.29,97.93]). In the test set, the model had a sensitivity of 91. 43% (95% CI = [89.69, 93.17]), a specificity of 100% (95% CI = [100,100]), a PPV value of 100% (95% CI = [100,100]), and an NPV value of 94.02% (95% CI = [79.50,100]). We also analyzed the diagnostic accuracy of the model for stage I and stage II CRC. For patients with stage I, the accuracy of the model was 100% (11/11), and the accuracy for stage II was 91.3% (21/23). Figure 4 shows the ROC curve for the validation set of the model with an AUC value of 99.31% (95% CI = [99,99.62]). The accuracy stats for the model in training and test set are shown in Supplement 5 of the Supplementary file1. Unsupervised hierarchical clustering of CRC tumor, fecal and control samples using these specific DNA methylation biomarkers showed

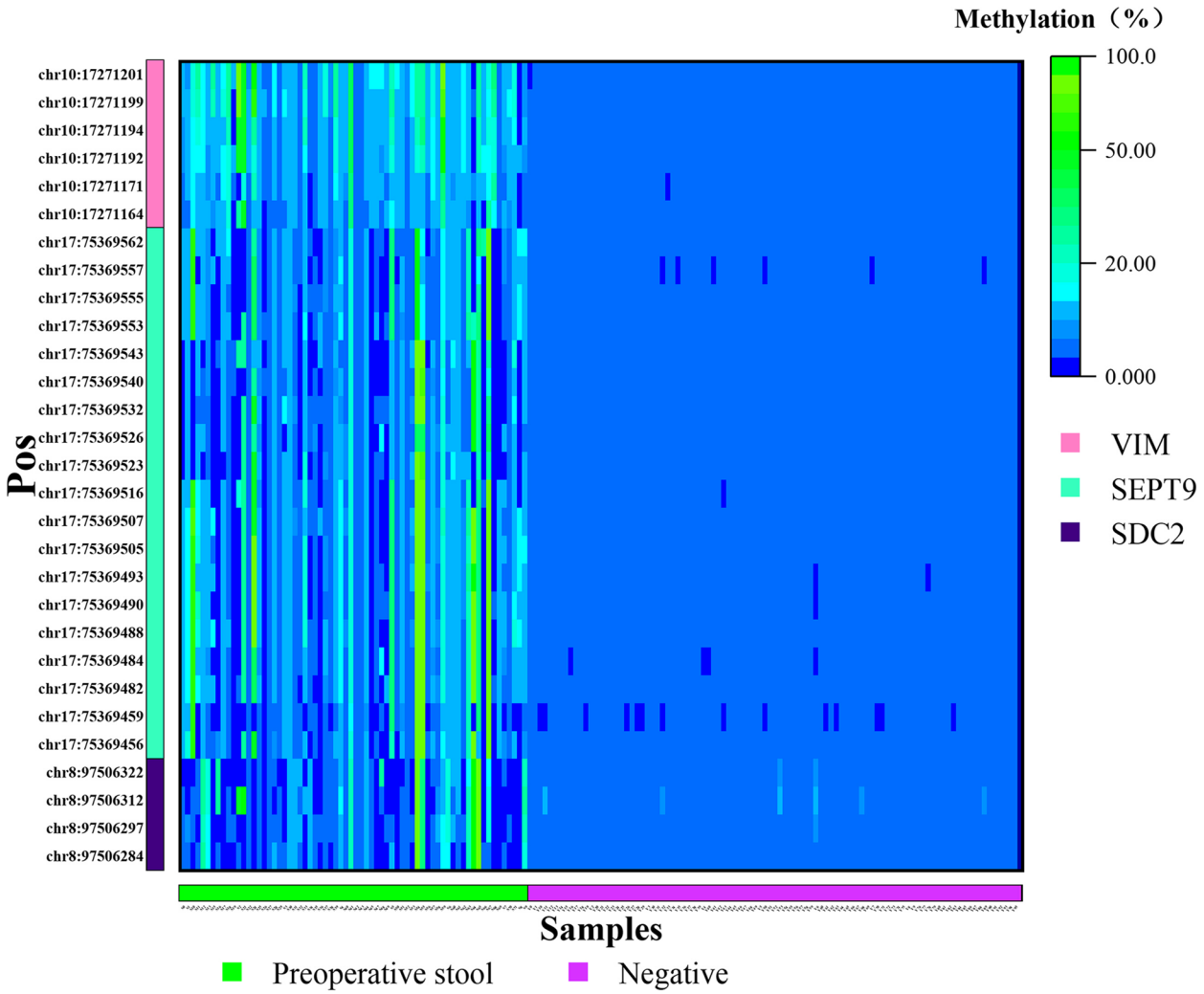

Fig. 3 Unsupervised hierarchical clustering of the 29 CRC-specific DNA methylation sites. The sites are from SEPTIN9, SDC2 and VIM in 49 CRC tissue samples, 67 CRC fecal samples and controls

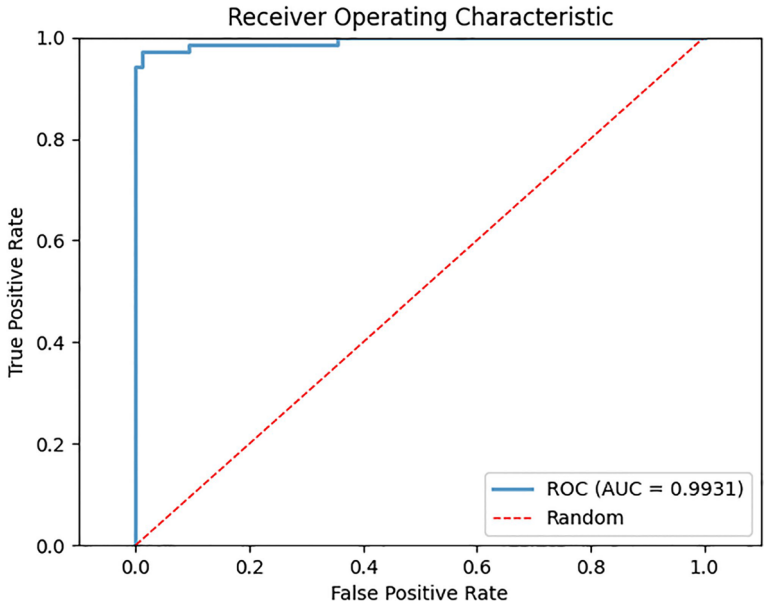

Fig. 4 Receiver operating characteristic curve. The model was cross-validated tenfold via CRC ($N=14$) and normal ($N=19$) fecal samples

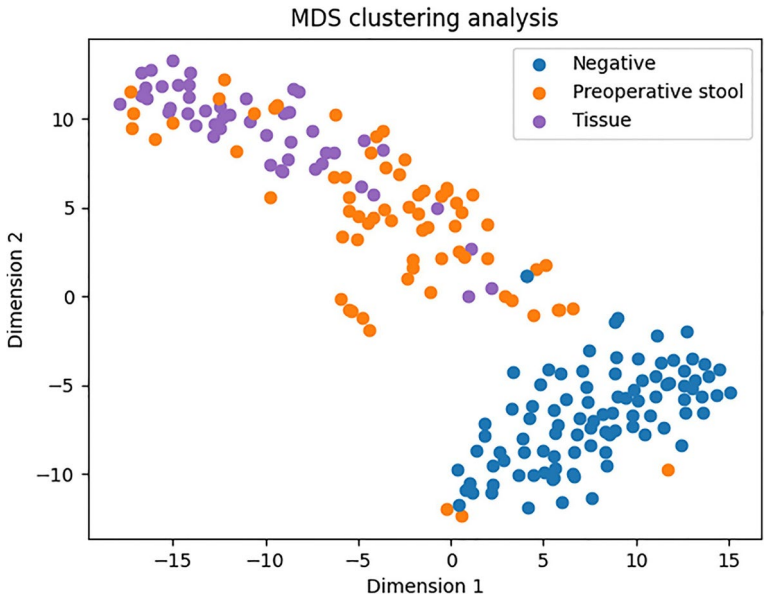

Fig. 5 Unsupervised hierarchical clustering of CRC-specific DNA methylation sites in tissue, fecal and control samples

that the biomarkers can distinguish between three groups of samples (Fig. 5). These findings suggest that our screening method and constructed model for CRC methylation sites has the potential to be a robust method for CRC diagnosis.

Discussion

DNA methylation patterns play crucial roles in establishing stable gene expression profiles [28]. It is recognized as a significant biomarker for CRC diagnosis, offering promise in reducing CRC mortality rates [29]. Although numerous DNA methylation biomarkers for CRC detection have been identified, the precise selection of CpG sites has not been thoroughly considered, with only a few DNA methylation panels being tested in fecal samples. In our study involving 236 CRC patients and healthy controls, DNA methylation testing and model construction were conducted on fecal and tissue samples. A panel of 29 CpG sites in the SDC2, SEPT9, and VIM genes from fecal samples was identified, and a model was developed

to demonstrate its potential as a biomarker for CRC diagnosis. This research serves as a valuable resource for identifying DNA methylation biomarkers in fecal samples for CRC diagnosis and highlights their potential clinical application.

The results demonstrate that our model effectively distinguishes CRC patients from normal individuals. In the verification cohort, the model showed an overall sensitivity of 91.43% and specificity of 100% in detecting CRC. These results are superior to those of the detection of SDC2 gene methylation alone in feces (sensitivity of 77.0%, specificity of 98.1%) [30], VIM gene methylation alone in feces (sensitivity of 72.5%, specificity of 86.9%) [31], and the combined detection of SEPT9 and SDC2 gene methylation (sensitivity of 89.1%, specificity of 90.8%) [14]. The excellent performance of this model can be attributed to the combined methylation assessment of the SDC2, SEPT9, and VIM genes, which enhances the labeling signal and improves sensitivity and specificity. Hypermethylation of any of these three genes was considered indicative of increased CRC risk during model development. Previous studies have also suggested that incorporating multiple gene methylation markers can increase detection accuracy [32]. In addition, one-to-one screening for CRC hypermethylation sites via tissue and fecal samples is more compelling. This screening method involves the collection of both tumor tissue and fecal samples from CRC patients. Owing to the difficulty in sample collection, this method is rarely used. Even so, this method overcomes the effect of tumor molecular heterogeneity to some extent and is able to accurately screen for CpG sites with the highest predictive value among hotspot genes. Thus, the process reduces the weakening of the sensitivity and specificity of the assay by sites of no predictive value.

In this study, most of the methylated sites of the SDC2, SEPT9 and VIM genes were hypermethylated in tumor tissues and CRC feces, whereas they were hypomethylated in control fecal samples (Fig. 3). With an AUC value of 99.31% (Fig. 4), the model constructed from their gene combinations effectively discriminated between CRC patients and healthy subjects (Fig. 5). The SEPT9 gene is a tumor suppressor gene. Aberrant DNA methylation reduces its transcriptional activity, which in turn leads to dysregulation of gene expression and aberrant physiological functions, which may cause cancer. The SEPT9 promoter V2 region is methylated in CRC tissues but rarely in normal colonic mucosa [33]. SDC2 encodes a membrane protein that plays a role in cell proliferation and migration. Oh et al. [34] examined the methylation of the SDC2 gene in CRC tissues and normal tissues and reported that hypermethylation was present only in tumor tissues, with a probability of 100%. Cytoskeletal proteins encoded by the VIM gene are thought to be involved in cancer invasion and metastasis. VIM promoter methylation in fecal samples has high sensitivity and specificity for CRC detection [35]. Methylation of the three selected genes is rarely detected in normal individuals, and These genes have great potential for diagnosing CRC. Although the WNT2, BMP3 and NDRG4 genes were hypermethylated in CRC tissues in this study, they were poorly distinguished between CRC fecal samples and healthy control samples (Fig. 2). Previous studies have shown that NDRG4 methylation to has high sensitivity and specificity in fecal and urine samples (positivity rates of 72.6 and 76.2%, respectively [36]), whereas the average methylation rate of this gene in feces was slightly lower in this study. The decrease in detection results may be related to differences in assay methods and patient groups. The genes associated with prognosis in this study, SFRP2 and P16, were methylated at approximately 10% in CRC tissues and were not well distinct in both CRC feces and healthy controls (Fig. 2). It is hypothesized that these two genes have poor potential for prognostic testing in CRC using feces.

CRC tumors grow in specific locations and CRC cells are constantly shed into the lumen of the colon and released directly into the feces, which lays the foundation for the use of fecal DNA as a screening method for CRC [37]. Blood tests are also important for early screening of CRC. Since apoptotic and necrotic CRC cells release free DNA into the bloodstream, blood testing is feasible. However, DNA testing of serum samples tends to exhibit high background noise due to other diseases. Some biomarkers (e.g., methylation of the SFRP2 gene) can not only indicate CRC but also mammary cancer which will affect the specificity of the test. Human DNA extracted from feces, the other hand, originates directly from tubular lumen shedding is more likely to be derived from CRC than from metastases or other primary tumors. It has also been found that markers released from cell shedding of cancer cells in the intestines may invade the bloodstream before blood vessels [38]. Therefore, feces testing is more accurate for detecting CRC. In our study, for the diagnosis of stage I and stage II CRC, the diagnostic model built utilizing the chosen loci had an accuracy of 100% (11/11) and 91.3% (21/23), respectively. The total number of samples may have an impact on the accuracy discrepancy. In addition, the modeling approach of logistic regression was shown to have good performance [39]. Previous studies have shown that logistic regression modeling methods can improve the potential for CRC risk prediction [40], and thus the choice of modeling

method may also be responsible for the better performance of the selected gene methylation sites and models in CRC detection.

The results of the study of the screened gene loci and models were shown to be able to diagnose CRC, however they performed poorly in predicting the cancer's stage. The average accuracy of the five validation results using the confusion matrix is only 27.3% (95% CI = [10.41,44.20]). These findings are included in the Supplementary file1 (Supplement 6 and Supplement 7). At the time of sample collection, there were more patients in stage II and III CRC and samples were easily available. Smaller sample sizes were found in stages IV due to the difficulty in obtaining clinical samples. The outcomes of the predictions will be more affected by an imbalance or insufficiency of sample size in each stage of the tumor. In addition, the number and range of gene loci studied may also have an impact on staging predictions. Follow-up studies will include additional gene loci to more accurately predict colorectal cancer staging.

According to this study, these loci and the modal may contribute to a clinically CRC diagnosis. The quantity and range of samples that are obtained, however, still need to be greatly increased. In addition to samples from various tumor stages, fecal and tissue samples from adenomas, precancerous lesions, and high-risk populations should also be obtained, if available. After collecting more samples, we will conduct a study on early screening for colorectal cancer, because the current research is not entirely applicable to screen. Fecal is used in this study as a non-invasive genetic testing method that can compensate for other diagnostic tools. It might be more helpful in the diagnosis of colorectal cancer if they are combined with other diagnostic tools.

Conclusions

We successfully developed a CRC methylation diagnostic model containing 29 mutation sites in three genes. The detection efficacy of the model has been validated in an independent cohort, emphasizing its potential for effective detection of CRC.

Abbreviations

CRC	Colorectal cancer
NGS	Next-generation sequencing
IFOBT	Immunochemical fecal occult blood test
QC	Quality control
CpG	Cytosine-phosphate-guanine
PCR	Polymerase chain reaction
MDS	Multidimensional scaling
PPV	Positive predictive value
NPV	Negative predictive value
ROC	Receiver operating characteristic
AUC	Area under the curve
AUROC	Area under the receiver operating characteristic curve
SD	Standard deviation

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-12990-4.

Supplementary Material 1: Supplement 1. Target gene description. Supplement 2. List of primers used for bisulfite and multiplex MethyLight polymerase chain reactions. Supplement 3. The 29 methylated sites and their corresponding genes selected in this study. Supplement 4. Table of information on the model coefficients. Supplement 5. The accuracy stats for the model in training and test set. Supplement 6. Accuracy of the models for staging colorectal cancer. Supplement 7. Confusion matrix for predicting tumor staging with model validation.

Supplementary Material 2: Methylation rates of positive samples.

Acknowledgements

We are grateful for the sincere cooperation of the patients enrolled in the study for voluntarily providing samples and clinical information. We also thank the R language researchers who developed the packages which were open-source packages and available free of charge.

Authors' contributions

Yue Liu and Hongzhi Li conducted the experiments and contributed to the conceptualization and design of the project, as well as manuscript writing. Lizhi Xu, Cundi Zhong and Hongbo Ming conducted the data analysis and contributed to conceptualization and manuscript writing. Lizhen Li contributed to the project design and manuscript writing. Qi Liu and Jinyin Zhao contributed to the data analysis. All of the authors have read and approved the manuscript.

Funding

This study was supported by Dalian Gentalker Biotech Co., Ltd., and the following projects: "Study on the application of DNA methylation detection in the early diagnosis and screening of colorectal cancer", Department of Laboratory, The Second Affiliated Hospital of Dalian Medical University.

Availability of data and materials

We did not generate new datasets for this article. Please contact the author if needed.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

All procedures conducted in this study adhered to the ethical standards outlined by the institutional research committee and were in accordance with the principles of the 1964 Helsinki Declaration and its subsequent amendments. Approval for the studies was granted by the institutional review board of the Second Affiliated Hospital of Dalian Medical University in Liaoning Province, China (LZ001). Written informed consent was obtained from each study subject.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Dalian Gentalker Biotech Co., Ltd., 9-2, Jinqi Road, Jinpu New District , Dalian, Liaoning 116635, China. [2]Department of Laboratory, The Second Affiliated Hospital of Dalian Medical University, 216 Zhongshan Street, Ganjingzi District, Dalian, Liaoning 116031, China.

Received: 3 July 2024 Accepted: 24 September 2024
Published online: 30 September 2024

References

1. Araghi M, Soerjomataram I, Jenkins M, Brierley J, Morris E, Bray F, Arnold MJIjoc: Global trends in colorectal cancer mortality: projections to the year 2035. 2019;144(12):2992–3000. https://doi.org/10.1002/ijc.32055.
2. Holtedahl K, Borgquist L, Donker GA, Buntinx F, Weller D, Campbell C, Månsson J, Hammersley V. Braaten T. Parajuli RJBfp: Symptoms and signs of colorectal cancer, with differences between proximal and distal colon cancer: a prospective cohort study of diagnostic accuracy in primary care. 2021;22:1–13. https://doi.org/10.1186/s12875-021-01452-6.
3. Bishehsari F, Mahdavinia M, Vacca M. Malekzadeh R. Mariani-Costantini RJWjogW: Epidemiological transition of colorectal cancer in developing countries: environmental factors, molecular pathways, and opportunities for prevention. 2014;20(20):6055. https://doi.org/10.3748/wjg.v20.i20.6055.
4. Kanth P, Inadomi JMJB: Screening and prevention of colorectal cancer. 2021;374. https://doi.org/10.1136/bmj.n1855.
5. Van Engeland M, Derks S, Smits KM, Meijer GA. Herman JGJJoco: Colorectal cancer epigenetics: complex simplicity. 2011;29(10):1382–91. https://doi.org/10.1200/JCO.2010.28.2319.
6. Lao VV, Grady WMJNrG, hepatology: Epigenetics and colorectal cancer. 2011;8(12):686–700. https://doi.org/10.1038/nrgastro.2011.173.
7. Laugsand EA. Brenne SS. Skorpen FJIJoCD: DNA methylation markers detected in blood, stool, urine, and tissue in colorectal cancer: a systematic review of paired samples. 2021;36:239–51. https://doi.org/10.1007/s00384-020-03757-x.
8. Tinmouth J, Lansdorp-Vogelaar I, Allison JEJG. Faecal immunochemical tests versus guaiac faecal occult blood tests: what clinicians and colorectal cancer screening programme organisers need to know. 2015;64(8):1327–37. https://doi.org/10.1136/gutjnl-2014-308074.
9. Bruns-Toepler M, Hardt PJAR: Evaluation of a new technique for iFOBT utilising a new sample collection device with increased buffer stability. 2017;37(7):3735–3740. https://doi.org/10.21873/anticanres.11747.
10. Högberg C, Samuelsson E, Lilja M, Fhärm EJBFP: Could it be colorectal cancer? General practitioners' use of the faecal occult blood test and decision making–a qualitative study. 2015;16:1–8. https://doi.org/10.1186/s12875-015-0371-1.
11. Ladabaum U, Dominitz JA, Kahi C, Schoen REJG. Strategies for colorectal cancer screening. 2020;158(2):418–32. https://doi.org/10.1053/j.gastro.2019.06.043.
12. Steffenssen MW, Al-Najami I, Baatrup GJAO. Patient-reported minor adverse events after colonoscopy: a systematic review. 2019;58(sup1):S22–8. https://doi.org/10.1080/0284186X.2019.1574979.
13. Shaukat A, Levin TRJNrG, hepatology: Current and future colorectal cancer screening strategies. 2022;19(8):521–531. https://doi.org/10.1038/s41575-022-00612-y.
14. Zhao G, Liu X, Liu Y, Li H, Ma Y, Li S, Zhu Y, Miao J, Xiong S, Fei S. Aberrant DNA methylation of SEPT9 and SDC2 in stool specimens as an integrated biomarker for colorectal cancer early detection. Front Genet. 2020;11:643. https://doi.org/10.3389/fgene.2020.00643.
15. Anghel SA, Ioniță-Mîndrican C-B, Luca I, Pop ALJC. Promising epigenetic biomarkers for the early detection of colorectal cancer: a systematic review. 2021;13(19):4965. https://doi.org/10.3390/cancers13194965.
16. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA. Berger BMJNEJoM: Multitarget stool DNA testing for colorectal-cancer screening. 2014;370(14):1287–97. https://doi.org/10.1056/nejmoa1405215.
17. Sun M, Liu J, Hu H, Guo P, Shan Z, Yang H, Wang J. Xiao W. Zhou XJJocr, oncology c: A novel panel of stool-based DNA biomarkers for early screening of colorectal neoplasms in a Chinese population. 2019;145:2423–32. https://doi.org/10.1007/s00432-019-02992-2.
18. Vacante M, Borzì AM, Basile F, Biondi AJWjocc: Biomarkers in colorectal cancer: Current clinical utility and future perspectives. 2018;6(15):869. https://doi.org/10.12998/wjcc.v6.i15.869.
19. Singh MP, Rai S, Suyal S, Singh SK, Singh NK, Agarwal A. Srivastava SJERoMD: Genetic and epigenetic markers in colorectal cancer screening: recent advances. 2017;17(7):665–85. https://doi.org/10.1080/14737159.2017.1337511.
20. Rasmussen SL, Krarup HB, Sunesen KG, Pedersen IS, Madsen PH, Thorlacius-Ussing OJCD. Hypermethylated DNA as a biomarker for colorectal cancer: a systematic review. 2016;18(6):549–61. https://doi.org/10.1111/codi.13336.
21. Devos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J, Steiger KV, Grutzmann R, Pilarsky C, Habermann JKJCc: Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. 2009;55(7):1337–1346. https://doi.org/10.1373/clinchem.2008.115808.
22. Carmona FJ, Azuara D, Berenguer-Llergo A, Fernández AF, Biondo S, de Oca J, Rodriguez-Moranta F, Salazar R, Villanueva A. Fraga MFJCpr: DNA methylation biomarkers for noninvasive diagnosis of colorectal cancer. 2013;6(7):656–65. https://doi.org/10.1158/1940-6207.CAPR-12-0501.
23. Fatemi N, Tierling S, Es HA, Varkiani M, Mojarad EN, Aghdaei HA, Walter J. Totonchi MJIjoc: DNA methylation biomarkers in colorectal cancer: Clinical applications for precision medicine. 2022;151(12):2068–81. https://doi.org/10.1002/ijc.34186.
24. Liang X, Liu Y, Sun J, Feng YJTC-GJoCO: Detection of serum aberrant CDKN2/P16 DNA in colorectal cancer. 2005;4:361–364. https://doi.org/10.1007/s10330-004-0268-2.
25. Melotte V, Lentjes MH, Van den Bosch SM, Hellebrekers DM, De Hoon JP, Wouters KA, Daenen KL, Partouns-Hendriks IE, Stessels F, Louwagie JJJJotNCI: N-Myc downstream-regulated gene 4 (NDRG4): a candidate tumor suppressor gene and potential biomarker for colorectal cancer. 2009;101(13):916–927. https://doi.org/10.1093/jnci/djp131.
26. Li L-C. Dahiya RJB: MethPrimer: designing primers for methylation PCRs. 2002;18(11):1427–31. https://doi.org/10.1093/bioinformatics/18.11.1427.
27. Wu X, Zhang Y, Hu T, He X, Zou Y, Deng Q, Ke J, Lian L, He X. Zhao DJMo: A novel cell-free DNA methylation-based model improves the early detection of colorectal cancer. 2021;15(10):2702–14. https://doi.org/10.1002/1878-0261.12942.
28. Dor Y, Cedar HJTL. Principles of DNA methylation and their implications for biology and medicine. 2018;392(10149):777–86. https://doi.org/10.1016/s0140-6736(18)31268-6.
29. Hajebi Khaniki S, Shokoohi F, Esmaily H. Kerachian MAJSR: Analyzing aberrant DNA methylation in colorectal cancer uncovered intangible heterogeneity of gene effects in the survival time of patients. 2023;13(1):22104. https://doi.org/10.1038/s41598-023-47377-1.
30. Zhang W, Yang C, Wang S, Xiang Z, Dou R, Lin Z, Zheng J, Xiong BJCm, research: SDC2 and TFPI2 methylation in stool samples as an integrated biomarker for early detection of colorectal cancer. 2021:3601–3617. https://doi.org/10.2147/cmar.s300861.
31. Itzkowitz SH, Jandorf L, Brand R, Rabeneck L, Schroy PC III, Sontag S, Johnson D, Skoletsky J, Durkee K, Markowitz SJCG, et al. Improved fecal DNA test for colorectal cancer screening. 2007;5(1):111–7. https://doi.org/10.1016/j.cgh.2006.10.006.
32. Fang Y, Peng J, Li Z, Jiang R, Lin Y, Shi Y, Sun J, Zhuo D, Ou Q, Chen JJM: Identification of multi-omic biomarkers from Fecal DNA for improved Detection of Colorectal Cancer and precancerous lesions. 2022:2022.2011.2008.22282099. https://doi.org/10.1101/2022.11.08.22282099.
33. Wasserkort R, Kalmar A, Valcz G, Spisak S, Krispin M, Toth K, Tulassay Z, Sledziewski AZ, Molnar BJBc: Aberrant septin 9 DNA methylation in colorectal cancer is restricted to a single CpG island. 2013;13:1–11. https://doi.org/10.1186/1471-2407-13-398.
34. Oh TJ, Oh HI, Seo YY, Jeong D, Kim C, Kang HW, Han YD, Chung HC, Kim NK, An SJCe: Feasibility of quantifying SDC2 methylation in stool DNA for early detection of colorectal cancer. 2017;9:1–11. https://doi.org/10.1186/s13148-017-0426-3.
35. Raut JR, Guan Z, Schrotz-King P, Brenner HJCE. Fecal DNA methylation markers for detecting stages of colorectal cancer and its precursors: a systematic review. 2020;12:1–19. https://doi.org/10.1186/s13148-020-00904-7.
36. Xiao W, Zhao H, Dong W, Li Q, Zhu J, Li G, Zhang S, Ye MJOl: Quantitative detection of methylated NDRG4 gene as a candidate biomarker for diagnosis of colorectal cancer. 2015;9(3):1383–1387. https://doi.org/10.3892/ol.2014.2815.
37. Ferrari A, Neefs I, Hoeck S, Peeters M, Van Hal GJC. Towards novel non-invasive colorectal cancer screening methods: a comprehensive review. 2021;13(8):1820. https://doi.org/10.3390/cancers13081820.
38. Ahlquist DAJNpo: Universal cancer screening: revolutionary, rational, and realizable. 2018;2(1):23. https://doi.org/10.1038/s41698-018-0066-x.

39. Chen X, Gole J, Gore A, He Q, Lu M, Min J, Yuan Z, Yang X, Jiang Y. Zhang TJNc: Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. 2020;11(1):1–10. https://doi.org/10.1038/s41467-020-17316-z.
40. Feng CH, Disis ML, Cheng C, Zhang LJLI. Multimetric feature selection for analyzing multicategory outcomes of colorectal cancer: random forest and multinomial logistic regression models. 2022;102(3):236–44. https://doi.org/10.1038/s41374-021-00662-x.

Publisher's Note