Ji et al. Clinical Epigenetics (2026) 18:95
https://doi.org/10.1186/s13148-026-02097-x

Clinical Epigenetics

RESEARCH Open Access

Circulating tumor DNA methylation-based method for noninvasive detection and stage stratification of colorectal tumor

Hongli Ji[1†], Mimi Xu[2†], Wei Jiang[1†], Yaowen Hu[1†], Botao Yan[1], Qiaolin Chen[1], Jixiang Zheng[1], Wei Wu[1], Xiarong Hu[1,3], Zhenbang Chen[1], Zelong Han[4], Hanlin Lu[5] and Jun Yan[1,6*]

Abstract

Background Early detection and optimal treatment could improve the outcomes of patients with colorectal cancer (CRC). No adequate method has been developed to meet these two requirements. Here, we aimed to identify differential circulating tumor DNA (ctDNA) methylation biomarkers associated with CRC, and then establish models for detection, stage stratification and clinical decision-making.

Results A total of 636 participants were included in this prospective study. To identify differential ctDNA methylation biomarkers, we first performed a genome-wide analysis between tumor and adjacent normal tissues using the α-value, which is sensitivity to ctDNA methylation signals. After filtering with PBMC samples, 4965 biomarkers were identified. A panel of 21 biomarkers was selected after shrinkage. A ctDNA methylation-based CRC diagnostic model (cMCD) was constructed. The cMCD yielded a sensitivity of 87.82% (83.39%–91.41%) for the combined detection of advanced adenomas and CRC, and a specificity of 91.88% (88.54%–94.49%). The overall accuracy was 89.85% (87.50%–92.30%), which was much greater than that of tumor markers [CEA: 67.30% (63.50%–70.93%); CA19-9: 59.91% (55.98%–63.74%); CA72-4: 55.66% (51.70%–59.57%); all $P < 0.001$]. As the risk score and sensitivity of the cMCD tended to increase with tumor progression, we constructed another ctDNA methylation-based model to stratify stage (cMCSS) and further guide treatment selection, specifically for discriminating the Early-stage patients eligible for curative endoscopic resection and avoiding overtreatment. The cMCSS could discriminate 75.86% (68.81%–82.02%) of Early-stage patients (advanced adenomas and T1N0M0 CRCs), for whom endoscopic resection could achieve curative intent, and 89.45% (84.59%–93.19%) of Advanced-stage patients. The accuracy [83.42% (79.36%–86.96%)] was significantly greater than that of tumor markers [CEA: 60.20% (55.17%–65.08%); CA19-9: 51.77% (46.71%–56.83%); CA724: 46.17% (41.16%–51.25%); all $P < 0.001$].

Conclusions The approach based on ctDNA methylation is a noninvasive and robust method for both early detection and tumor stratification of CRC and could benefit patients in clinical decision-making.

†Hongli Ji, Mimi Xu, Wei Jiang and Yaowen Hu contributed equally to this work.

*Correspondence:
Jun Yan
yanjunfudan@163.com

Full list of author information is available at the end of the article

Keywords Colorectal cancer, ctDNA methylation, Biomarker, Early detection, Stage stratification

Background

Colorectal cancer (CRC) is among the most common malignancies worldwide, with almost 2 million new cases occurring each year [1, 2]. Owing to changes in food and lifestyle, the incidence and mortality rates of CRC are increasing in developing countries, such as China [2, 3]. Early detection and optimal treatment can not only improve the outcomes and quality of life of patients, but also mitigate the disease burden [4, 5]. Colonoscopy combined with tissue biopsy is considered the gold standard for diagnosing CRC [6, 7]. Successful treatment begins with accurate discrimination during colonoscopy [8]. However, as an invasive method, colonoscopy is expensive and requires bowel preparation [9]. These disadvantages limit the compliance within the population [10]. Fewer than 20% of the eligible individuals in China have undergone colonoscopy [11]. The evidence indicates that more affordable and less invasive methods may be more popular and that a blood-based method is preferred [2, 12]. However, no adequate blood-based method is available. Traditional tumor markers, including serum carcinoembryonic antigen (CEA), carbohydrate antigen 19 – 9 (CA19-9) and CA72-4, have insufficient sensitivity [13–15]. Other methods, including fecal immunochemical testing and computed tomography, are inadequate for poor diagnostic efficacy, especially for advanced adenomas (AAs) and early-stage CRC [16–18]. In addition, discriminating stage during endoscopy is crucial but challenging. According to the National Comprehensive Cancer Network, endoscopic resection (ER) is recommended for treating early-stage patients (those with suspected T1 stage), yet 70–80% of patients who undergo ER require additional radial surgery for high-risk pathological features of lymph node metastasis (LNM) [19, 20]. The high-risk features include deep submucosal invasion (≥ 1000 μm), lymphatic or vascular invasion, high-grade tumor budding, and poorly differentiated histology [8, 21]. However, this post-treatment risk stratification results in more than 80% of high-risk patients without LNM being overtreated [22, 23]. The selection of ER or radial surgery remains confusing. No adequate pre-treatment method is available to discriminate stage [23, 24]. Therefore, it is imperative to develop an accurate blood-based method not only for both early detection and stage stratification of colorectal tumors, but also as a noninvasive tool to guide optimal treatment selection (e.g., suitability for ER). Circulating tumor DNA (ctDNA), as a type of cell-free DNA (cfDNA), is the DNA fragment derived from dead tumor cells that contains cancer-specific genetic and epigenetic aberrations [25, 26]. DNA methylation is a well-known epigenetic alteration and has been extensively studied. Global hypomethylation and hypermethylation of cytosine–phosphoric–guanine (CpG) islands (CGIs) in tumor suppressor genes results in carcinogenesis and tumor progression [27, 28]. Aberrant methylation can also be detected early in ctDNA [29, 30]. Several studies have identified CRC-specific ctDNA methylation biomarkers, such as *septin-9* (*SEPT9*), a biomarker that has been approved by the Food and Drug Administration [31, 32]. However, the *SEPT9* was not recommended as the first-line method for diagnosis and screening for the inadequate sensitivity [33]. Aberrations in other biomarkers, such as *SFRP2*, *SDC2*, and *ADHFE1*, have also been reported [34–36]. The verification of these markers was based mainly on sequencing methods involving a few biomarkers rather than genome-wide methylation profiling. Verified biomarkers may not reflect the entire DNA methylation profile and the above methods are mainly aimed at early diagnosis but not at stage discrimination. Therefore, potential ctDNA methylation signatures for stage discrimination remain to be further investigated. In addition, conventional methods for methylation marker discovery rely on population-average methylation values, the β-value. However, such β-value based approaches are not sensitive enough to detect tumor signals especially when the tumor fraction is ultralow in the cfDNA samples, whereas the α-value, defined as the percent of methylated CpGs out of all CpG sites in a sequencing read, shows superior sensitivity [37]. In this study, genome-wide methylation profiling of all samples was performed with Methylome-Seq and the α-value was used to define the differential methylation region (DMR), which can capture and characterize DNA methylation biomarkers with high sensitivity. The purpose of this study was to establish ctDNA methylation-based models for CRC diagnosis, stage stratification and treatment selection.

Methods

Study design

This prospective study was approved by the Institutional Ethical Review Board at Nanfang Hospital of Southern Medical University. This study is registered at clinicaltrials.gov with the identifier NCT05558436. Participants were recruited consecutively at Nanfang Hospital from June 2022 to September 2023. Informed consent was obtained. Patients with CRC or intestinal polyps were diagnosed by histological examination. AA was defined as the adenoma with a diameter of ≥ 10 mm or with villous components or with high-grade dysplasia. Participants with AA or CRC were included in the tumor group. Individuals with no evidence of disease (NED) and

patients with benign polyps, non-advanced adenomas or benign anorectal diseases were included in the control group. Participants with negative colonoscopies and no previous diagnosis of significant disease were considered to have NED. Within the tumor group, patients with AA and pT1N0M0 CRC were eligible for curative ER. These patients constituted the Early-stage group. The patients with more advanced-stage disease constituted the Advanced-stage group, for whom ER alone could not achieve curative intent.

The participants in the tumor group met the following criteria: (1) aged between 18 and 80 years; (2) willing to provide blood and/or tissue samples; and (3) tumor confirmed by pathological examination. The exclusion criteria were (1) receiving tumor treatment prior to blood sample collection, including surgical resection, neoadjuvant chemoradiotherapy or targeted therapy; (2) having other malignancies, inflammatory bowel disease or Lynch syndrome; (3) having an acute or chronic infection; (4) having indications for emergency surgery, including bleeding, obstruction or perforation; or (5) having liver and/or kidney failure. Participants with a previous history or presence of cancer or critical illness were excluded from the control group.

The sample size was calculated according to the detailed information in the Supplementary Materials. All clinicopathological data were collected. The results of CEA, CA19-9 and CA72-4 were divided into normal and abnormal.

Sample collection

The tumor tissues and corresponding adjacent normal tissues were obtained during colonoscopy or after radical resection. Whole blood samples (15–20 mL) were collected from each participant using Cell-Free DNA blood collection tubes (Streck, USA). The plasma and peripheral blood mononuclear cells (PBMCs) were separated within 2 h. All samples were stored at −80 °C until use.

DNA isolation and methylation profiling

Genomic DNA was isolated from tissue samples via the Qiagen DNeasy Blood & Tissue Kit (BioMag Scientific Inc., China).

ctDNA and PBMC DNA were harvested with a Circulating DNA Extraction Kit (Shenzhen Apostle-Sustech Ltd., China).

The isolated DNA was purified, eluted and then analyzed via the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). The sequencing libraries were constructed with the NEBNext® Enzymatic Methyl-seq Conversion Kit. Following hybridization and purification, the DNA was quantified again. Only DNA with a concentration of ≥ 5 ng/L and an abundance of ≥ 150 ng was amplified via PCR. Methylome-seq libraries were generated by sequencing 150 bp paired-end reads on the Illumina NovaSeq 6000 platform (Illumina, USA). Among them, 636 samples that passed the quality control criteria (coverage higher than 15x and bisulfite conversion rate higher than 99.0%) were used in this study.

Detection of read-based methylation markers

After preprocessing of the methylome-seq data (for detailed methods, see the Supplementary Materials), we explored the differential biomarkers for further study. Previous studies proposed the concept of the α-value, defined as the percentage of methylation CpGs out of all CpG sites in a sequencing read [37, 38]. In this study, we defined a two-step framework for using the α-value to robustly identify markers in cfDNA samples controlled by CRC-specific methylated markers in tissue and PBMC samples (Figure S1). First, we used Metilene (version 2.8) to detect cancer-specific DMRs by comparing colorectal tumor tissue and adjacent normal tissue with the 2D KS-test. The *P* values were corrected for multiple hypothesis testing using the Benjamini–Hochberg procedure to obtain adjusted *P* values. The tissue-based DMRs that met the following criteria were selected: (1) containing ≥ 5 CpGs, and the distance between two adjacent cytosines was within 100 bp; (2) with a depth of ≥ 10X; and (3) the α-values of the region were significantly different (the absolute value ≥ 0.15 and adjusted *P* value ≤ 0.01). The background noise within the potential DMRs was further filtered using PBMC samples. DMRs with α-values ≤ 0.05 in the PBMC samples were retained as candidate markers. Second, the α-value for each candidate marker in all cfDNA samples was determined, which is the normalized fraction of methylated reads out of all mapped reads in this region. The mapped reads in cfDNA samples were extracted from alignment BAM files, and methylated read was defined as the mapped read containing at least two adjacent methylated CpG sites.

Construction of the model for diagnosis

On the basis of the above analysis, the cfDNA methylation dataset was randomly allocated to the training and validation datasets. The least absolute shrinkage and selection operator (LASSO) and recursive feature elimination (RFE) based on 10-fold cross-validation were used for biomarker shrinkage in the training cohort. The intersection of the two methods was identified as the final DMRs. A ctDNA methylation-based prediction model for CRC diagnosis (cMCD) was constructed using eXtreme Gradient Boosting (XGBoost). The best cutoff values were determined by Youden's index. Models combining all tumor markers were also constructed with XGBoost. Hyperparameter fine-tuning for the models was implemented on the basis of 10-fold cross-validation to achieve optimal performance. The predictability of the

models, including the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV), was evaluated and compared to that of traditional tumor markers. The calibration curve was used to compare the predicted probabilities with the real probabilities. Decision curve analysis (DCA) was used to assess the clinical application value.

Construction of the model for stage stratification and guiding treatment selection

ER is less invasive than radial surgery. However, the accuracy of estimating whether ER could achieve curative resection for the lesion is not satisfactory in the clinic. To guide treatment selection, we constructed another model for stage stratification. The participants in the tumor group were divided into the Early-stage group and the Advanced-stage group. Owing to the imbalanced sample size between the two groups, the adaptive synthetic oversampling (ADASYN) algorithm was applied to ensure the comparability of the training data. A ctDNA methylation-based prediction model for stage stratification (cMCSS) was subsequently constructed with the final DMRs. Hyperparameter fine-tuning for the models was implemented. The performance of the novel model was evaluated.

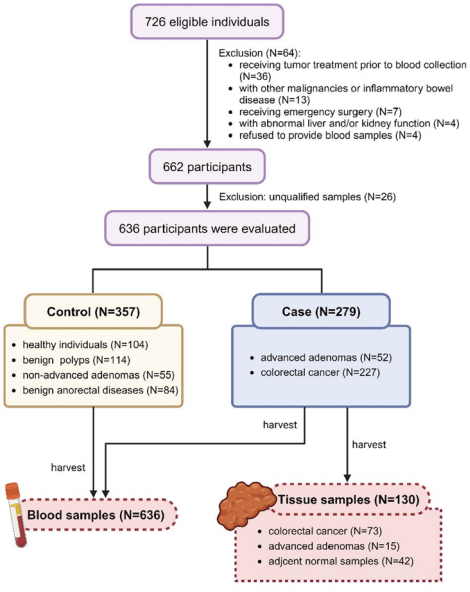

Fig. 1 Participants and sample enrollment of the study. This picture was created with BioRender.com. *AA* advanced adenoma. *CRC* colorectal cancer

Model interpretability

To interpret how the selected DMRs influence the model, the artificial intelligence SHapley Additive exPlanations (SHAP) was used [39, 40]. SHAP can help address the challenge of black-box predictions and gain insights into the decision-making process of the model. On the basis of SHAP, the importance of individual predictors for early detection and stage stratification could be determined.

Statistical analysis

Continuous variables are presented as median (interquartile range [IQR]) and were compared via the Mann–Whitney *U* test. Categorical variables are presented as number (percentage) and were compared via the chi-square test or Fisher's exact test. The AUROCs were compared via a permutation-based Wilcoxon signed-rank test. A two-tailed *P* value less than 0.05 was considered to indicate statistical significance. R (v4.2.2) and MedCalc (v20.0) were used for statistical analysis.

Results

Participants and sample characteristics

A total of 636 plasma samples from 227 CRC patients, 52 AA patients, and 357 control individuals were profiled (Fig. 1). The number meets the sample size standard. The control group consisted of 104 NEDs, 114 patients with polyps, 55 patients with non-advanced adenomas and 84 patients with benign anorectal diseases. The plasma samples were randomly split into training and validation datasets at a ratio of 7:3, and the clinicopathological characteristics were comparable between the training and validation datasets (Table 1).

Genome-wide DNA methylation analysis reveals the differential biomarkers for colorectal tumors

To elucidate the colorectal tumor-specific DNA methylation patterns, we first performed genome-wide DNA methylation profiling in 88 tumor (73 CRC and 15 AA) and 42 adjacent normal tissue samples. The average methylation levels of different genomic elements in different samples are shown in Fig. 2A. A total of 306,438 markers were identified. The methylation regions with the absolute $\Delta\alpha \geq 0.15$ and adjusted *P* value ≤ 0.01 were recognized as DMRs. A total of 27,433 DMRs were significantly hypermethylated, while 278,822 DMRs were significantly hypomethylated in tumors (Fig. 2B). DMRs with α-values ≤ 0.05 in the PBMC samples were retained as candidates for the next step. After the methylation level for each candidate in all the cfDNA samples was calculated and verified, 4965 DMRs were identified. These DMRs were associated with multiple types of biological processes (Figure S2). On the basis of the gene location annotation, the proportion of DMRs in gene bodies was the greatest, whereas fewer DMRs were located in

Table 1 Characteristics of the training and validation datasets

Characteristic	Training dataset (*n* = 446)	Validation dataset (*n* = 190)	*P*
Control group, *n*	**251**	**106**	
Age[years], Median (IQR)	44.0 (33.0–54.0)	46.0 (33.0–53.0)	0.732
Gender, No. (%)			0.282
Male	120 (47.8%)	58 (54.7%)	
Female	131 (52.2%)	48 (45.3%)	
BMI [kg/m^2], Median (IQR)	23.2 (21.0–25.5.0.5)	23.25 (20.8–25.2)	0.789
Type			0.770
NED	74 (29.5%)	30 (28.3%)	
Polyp	76 (30.3%)	38 (35.8%)	
Adenoma	40 (15.9%)	15 (14.2%)	
Benign Anorectal Diseases	61 (24.3%)	23 (21.7%)	
Tumor group, n	**195**	**84**	
Age [years], Median (IQR)	60.0 (52.0–68.0)	58.0 (50.5–66.5)	0.303
Gender, No. (%)			0.211
Male	108 (55.4%)	54 (64.3%)	
Female	87 (44.6%)	30 (35.7%)	
BMI [kg/m^2], Median (IQR)	22.7 (20.0–25.0)	23.1 (20.8–25.7)	0.422
Advanced Adenoma, n	**34**	**18**	
Max size [cm], No. (%)			> 0.999
< 4	25 (73.5%)	14 (77.8%)	
≥ 4	9 (26.5%)	4 (22.2%)	
Location, No. (%)			0.945
Right	9 (26.5%)	4 (22.2%)	
Left	16 (47.1%)	9 (50.0%)	
Rectum	9 (26.5%)	5 (27.8%)	
Colorectal cancer, n	**161**	**66**	
Max size [cm], No. (%)			0.401
< 4	70 (43.5%)	24 (36.4%)	
≥ 4	91 (56.5%)	42 (63.6%)	
Location, No. (%)			0.054
Right	43 (26.7%)	21 (31.8%)	
Left	52 (32.3%)	29 (43.9%)	
Rectum	66 (41%)	16 (24.2%)	
Stage, No. (%)			0.313
I-II	72 (44.7%)	24 (36.4%)	
III-IV	89 (55.3%)	42 (63.6%)	
Differentiation, No. (%)			> 0.999
Well to moderate	142 (88.2%)	58 (87.9%)	
Poor	19 (11.8%)	8 (12.1%)	
MSI state			0.548
pMMR	138 (85.7%)	53 (80.3%)	
dMMR	10 (6.2%)	6 (9.1%)	
NA	13 (8.1%)	7 (10.6%)	

IQR interquartile range, *BMI* body mass index, *NED* no evidence of disease, *MSI* microsatellite instability, *pMMR* mismatch repair proficient, *dMMR* mismatch repair deficient, *NA* not available

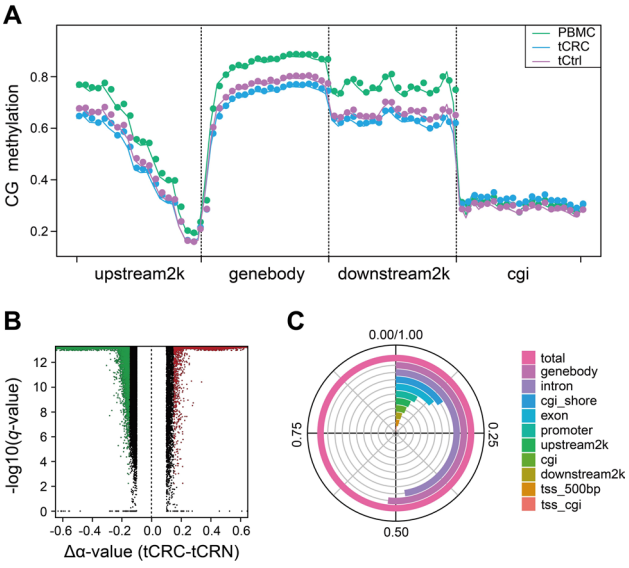

Fig. 2 The landscape of CRC-associated cfDNA methylation signatures by Methylome-Seq. **A** DNA methylation levels across genomic functional elements among three groups: PBMC, tCRC (DNA from CRC tissues), and tCtrl (DNA from normal control tissues) groups. Each genomic element, including upstream2k (2 kb region upstream of the transcription start site), the gene body, downstream2k (2 kb region downstream of the transcription termination site), and cgi (CpG island), is divided into 20 windows. For each window, methylation levels were calculated as the average within each group. **B** Volcano plot representing differential methylation regions (DMRs) between CRC and normal tissues. Hypermethylated DMRs are shown in red, hypomethylated DMRs are shown in green, and non-significant methylated DMRs are shown in gray. **C** Classification of DMRs according to their location relative to their genomic distribution. *CRC* colorectal cancer. *TSS* transcription start site

the regulatory regions (Fig. 2C). For biomarker shrinkage, LASSO and RFE were used in the training dataset (Fig. 3A and B). Eventually, a panel of 21 DMRs was selected for further studies (Fig. 3C and D, Table S1).

Construction of the cMCD for the combined diagnosis of AA and all-stage CRC

The AUROCs of the individual DMRs for detecting AA and CRC were all around 0.7, and the performance was insufficient for use (Table S2). Thus, the diagnostic model, cMCD, was constructed based on the 21 DMRs. A model combining all tumor markers was also constructed. The performance of each method is presented in Table S3. The AUROCs of the cMCD were 0.97 (95% CI 0.95–0.98) and 0.92 (95% CI 0.88–0.96) in the training and validation datasets, respectively, which were significantly greater than those of the tumor markers (all $P < 0.001$; Fig. 4A and B). The AUROCs were 0.92 (95% CI 0.87–0.96) and 0.96 (95% CI 0.94–0.98) for AA and CRC, respectively. With the best cutoff value of 0.4412 determined by Youden's index, the cMCD yielded a sensitivity of 87.82% (95% CI 83.39%–91.41%), a specificity of 91.88% (95% CI 88.54%–94.49%), a PPV of 89.42% (95%

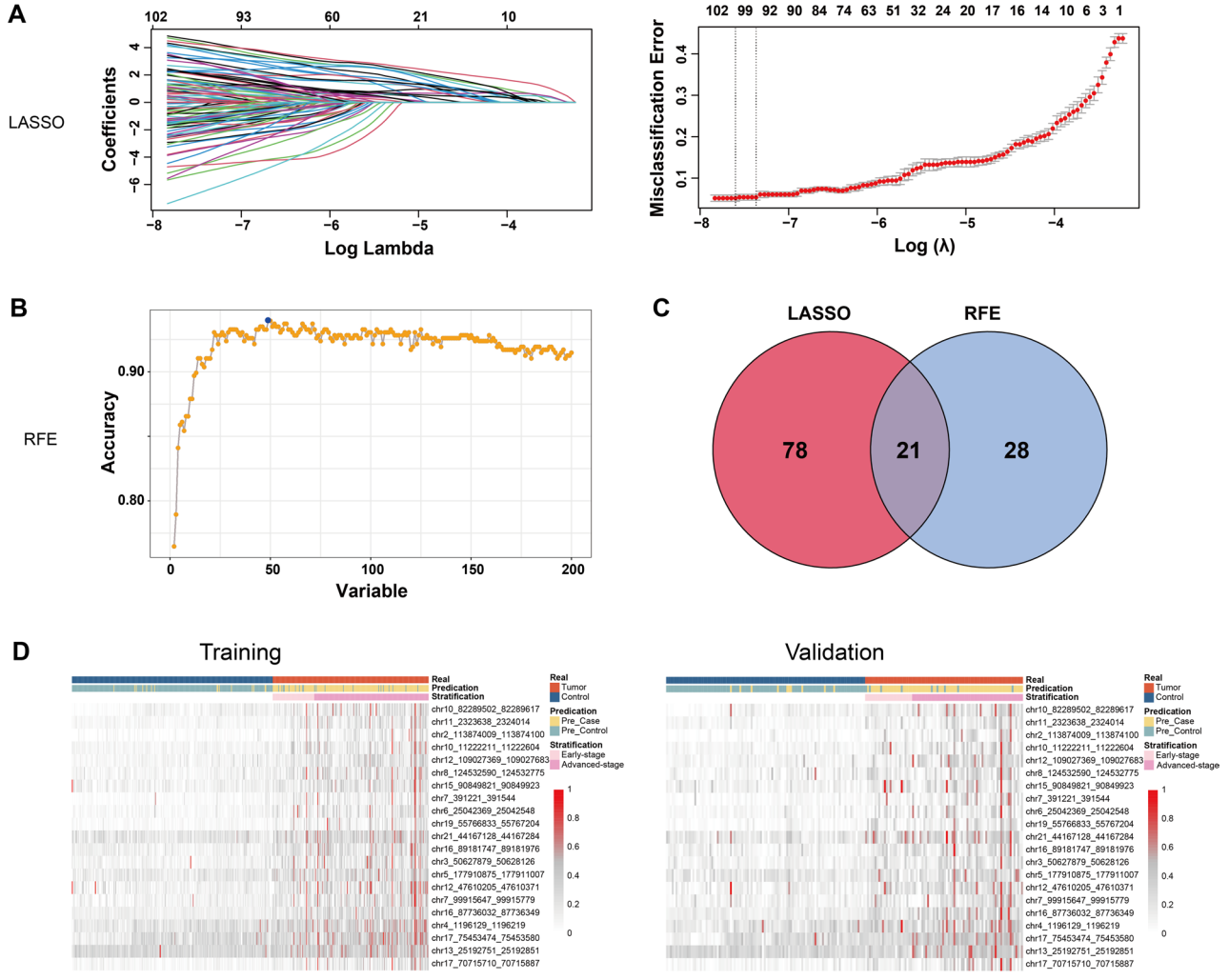

Fig. 3 Selection of differentially methylated regions. **A** Selection of the optimal penalty parameter (λ) via 10-fold cross-validation. The misclassification error is plotted versus log (λ). The optimal λ value of 0.00049156 with log (λ) = −7.617927 is selected. The dotted vertical line is the optimal λ value of −7.617927 with 99 nonzero coefficients. **B** Feature selection with RFE based on 10-fold cross-validation, the optimal accuracy was achieved when 49 biomarkers were included. **C** The intersection of LASSO and RFE consisted of 21 DMRs. **D** Heatmaps of the selected DMRs in the training and validation datasets. Clinical features and predicted outcomes are indicated by the annotation bars above the heatmaps. *LASSO* least absolute shrinkage and selection operator. *RFE* recursive feature elimination

CI 85.60%–92.31%), an NPV of 90.61% (95% CI 87.55%–92.98%), and an accuracy of 89.85% (95% CI 87.50%–92.30%; Table S3). The feature importance and impact were calculated with SHAP. According to the SHAP values, chr16_87736032_87736349, chr7_391221_391544 and chr12_47610205_47610371 were the three most important predictors for the cMCD (Figure S3). The sensitivity, accuracy and NPV of the cMCD were superior to those of all tumor markers (all $P < 0.001$; Fig. 4C). The calibration curves revealed good agreement between the cMCD and the real outcomes, while DCA revealed the clinical benefits (Figure S4). These results indicated that both positive and negative classifications of the cMCD were reliable.

Influence of confounding factors

To determine whether age, gender, tumor location or other confounding factors influence the performance of the cMCD, subgroup analyses were performed. The sensitivity of the cMCD was greater for late-onset patients (late-onset vs. early-onset, 91.67% vs. 70.59%, $P < 0.001$) and patients with larger tumor (≥ 4 cm vs. <4 cm, 93.15% vs. 81.95%, $P = 0.004$; Figure S5). Intriguingly, we found that both the risk score and the sensitivity of the cMCD increased with tumor progression. The sensitivities were 78.85% (95% CI 65.30%–88.94%) for AA, 84.38% (95% CI 75.54%–90.98%) for stage I–II CRC and 93.89% (95% CI 88.32%–97.33%) for stage III–IV CRC, which were significantly greater than those of CEA (13.46%, 23.96% and 42.45%), CA19-9 (3.85%, 2.08% and 24.43%), CA72-4 (15.38%, 11.46% and 11.45%) and the combination

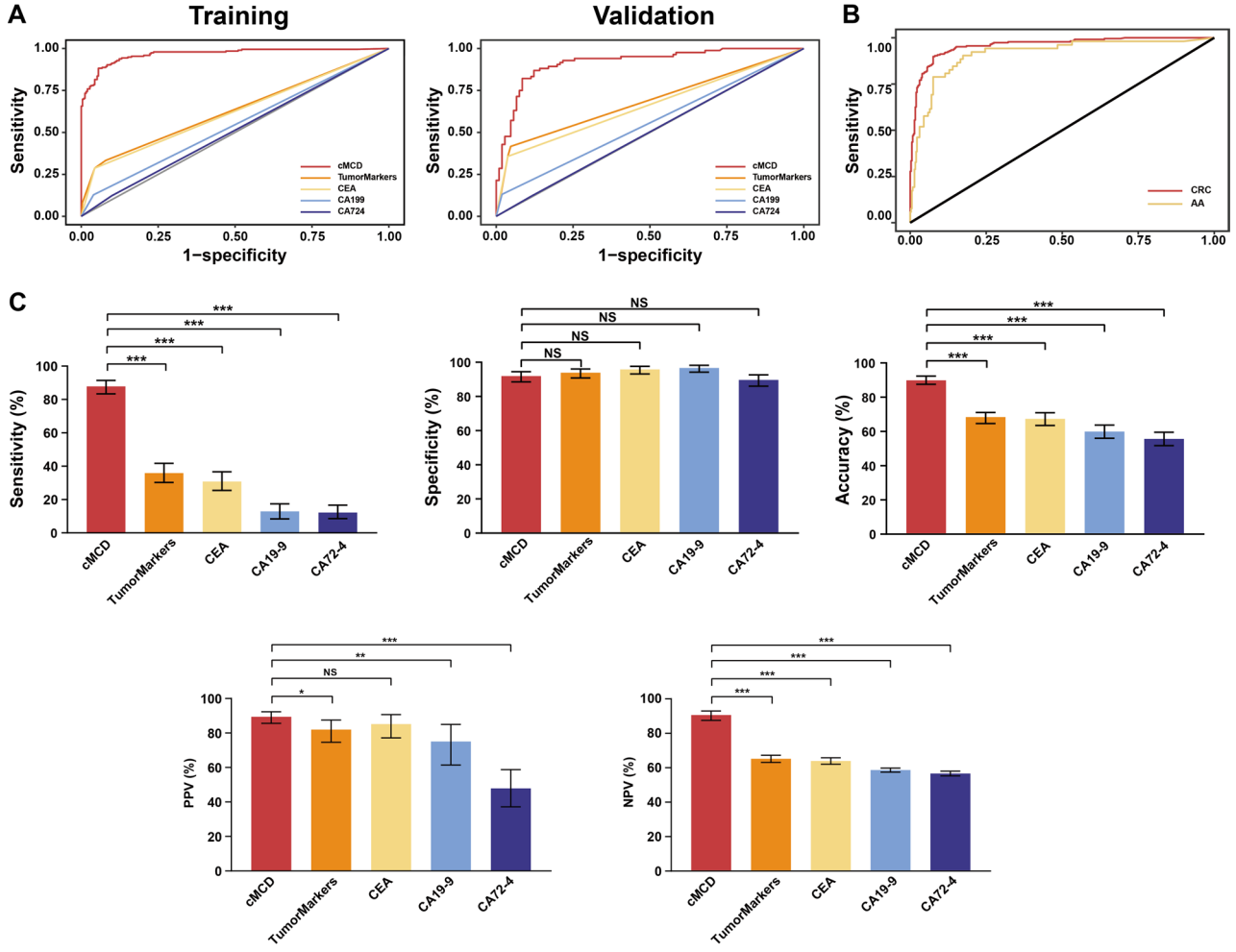

Fig. 4 The performance of the cMCD and comparison with tumor markers. **A** The ROCs of the cMCD and tumor markers for distinguishing colorectal tumor in the training and validation datasets. **B** The ROCs of the cMCD for distinguishing patients with AA and all-stage CRC from the controls. **C** Comparison of sensitivity, specificity, accuracy, PPV and NPV between different biomarkers: cMCD (red), CEA (yellow), CA19-9 (blue), CA72-4 (purple) and the combination (orange). NS, not significant; *$p<0.05$; **$p<0.01$; ***$p<0.001$. *CEA* carcinoembryonic antigen. *CA19-9* carbohydrate antigen 19–9. *CA72-4* carbohydrate antigen 72–4

(15.39%, 26.04% and 51.15%) (all $P<0.001$; Figure S5). In addition to the reliable predictive efficacy in all stages, the risk score and sensitivity of the cMCD tended to increase with tumor progression according to T and N stage (Fig. 5). Therefore, it is reasonable to assume that the ctDNA methylation-based approach could also be applied in stage stratification and treatment selection.

Construction of the cMCSS for stage stratification and guiding treatment selection

To verify the above hypothesis, we performed further studies on the same 21 DMRs. Patients in the tumor group were divided into the Early–stage group (for whom ER could achieve curative resection) and the Advanced–stage group. To balance the data, ADASYN was performed in the training dataset. The ratio of Early–stage to Advanced–stage increased from 42:153 to 155:153 (Fig. 6A).

The AUROCs of the individual DMRs for stratifying stage and guide treatment selection for CRC ranged from 0.44 to 0.77, which was insufficient (Table S4). Afterward, the combined stage stratification model, cMCSS, was constructed. The AUROCs were 0.89 (95% CI 0.85–0.93) and 0.89 (95% CI 0.81–0.97) in the training and validation datasets, respectively, which were significantly greater than those of tumor markers (all $P<0.001$; Fig. 6B). The risk score of the cMCSS was significantly higher in Advanced-stage than that in Early-stage ($P<0.001$; Fig. 6C). With the best cutoff value of 0.4063, the cMCSS could discriminate 75.86% (95% CI 68.81%–82.02%) of the Early-stage, and 89.45% (95% CI: 84.59%–93.19%) of the Advanced–stage. The accuracy was 83.42% (95% CI 79.36%–86.96%; Table S5). On the basis of SHAP, chr17_75453474_75453580, chr7_391221_391544 and chr13_25192751_25192851 were the three most important predictors for cMCSS (Figure S6). Although the

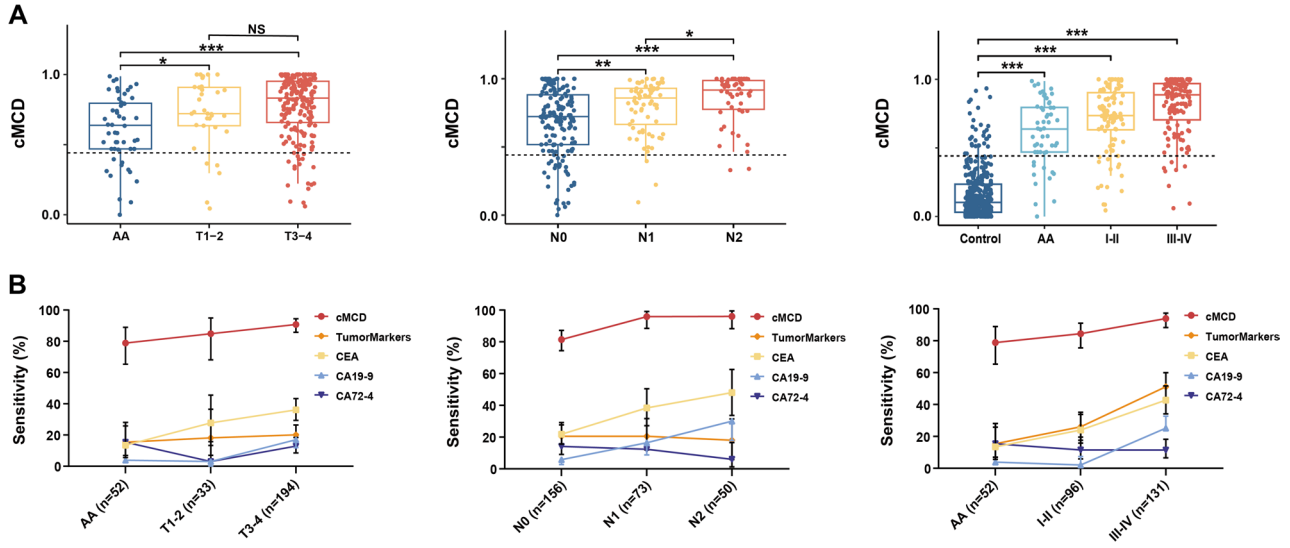

Fig. 5 The risk score and sensitivity of the cMCD at different T, N and AJCC stages. Both the risk score (**A**) and sensitivity (**B**) of the cMCD tended to increase with the progression of colorectal tumor. NS, not significant; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. *CEA* carcinoembryonic antigen. *CA19-9* carbohydrate antigen 19–9. *CA72-4* carbohydrate antigen 72–4

sensitivity to detect the Early-stage was lower, the PPV of cMCSS (85.16% [95% CI 79.44%–89.50%]) was significantly greater than tumor markers (all $P < 0.001$; Fig. 6D). The specificity and accuracy of the cMCSS were superior to those of tumor markers (Fig. 6D; Table S5). The calibration curves revealed the high reliability of the cMCSS for stage stratification, and DCA highlighted the potential clinical benefits (Figure S7). In summary, the cMCSS yielded adequate performance in stage stratification and could help to determine whether ER could achieve curative intent.

Discussion

Early detection and optimal treatment selection are crucial for the management of CRC patients in the clinic, especially early-stage patients [41, 42]. An adequate non-invasive blood-based method for detecting colorectal tumors, stratifying stage and guiding treatment selection would be of great clinical value. In this study, we identified 21 DMRs and established two novel methods, of which the first was the cMCD for detecting AA and all-stage CRC, and the second was the cMCSS for stage stratification. The sensitivity of the cMCD was 78.85% for AA, 84.38% for stage I-II CRC and 93.89% for stage III-IV CRC, and all of these values were significantly greater than those of tumor markers. Robust diagnostic efficacy revealed that patients could be early detected and receive prompt treatment to halt the advancement of the disease with the application of the cMCD. The risk score and sensitivity of the cMCD increased with tumor stage, which is consistent with the findings of other studies [9, 43]. To further stratify stage, the cMCSS was constructed with the same 21 DMRs, which could discriminate 75.86% of the Early-stage and 89.45% of the Advanced-stage. The accuracy (83.42%) was much greater than that of all tumor markers. Thus, cMCSS could assist in endoscopic optical diagnosis and guide optimal treatment. For the Early-stage patients, endoscopic resection alone can achieve curative intent, which could avoid overtreatment and reduce postoperative complications [44]. For the Advanced-stage patients requiring surgical intervention, this approach also helps prevent undertreatment, avoid secondary surgeries, and minimize trauma. In addition, we provided the selected DMRs with explanations on how they participated in the early detection and stage stratification via a unified method of SHAP.

ctDNA methylation is a promising approach because aberrant signatures may be present early in tumorigenesis and be associated with tumor progression [45]. In addition, tumor cells with the same origin tend to present similar DNA methylation signatures [27, 46]. The amount of ctDNA is low, especially in the early stage. To take full advantage of ctDNA fragments, Methylome-Seq was employed, and more ctDNA fragments could be retained [47]. The α-value-based approach was also used in this study for methylation biomarker discovery, which can reveal the joint methylation states of multiple adjacent CpG sites. Previous studies have shown that this approach carries more sensitive tumor signals than the β-value does, even with low tumor fractions and sequencing coverage [37]. In addition, the α-value could provide a sensitivity estimate of the fraction across all markers instead of a set of markers. A previous study identified liver cancer-specific methylation biomarkers for detecting hepatocellular carcinoma with the α-value [38]. Another study developed a general framework with

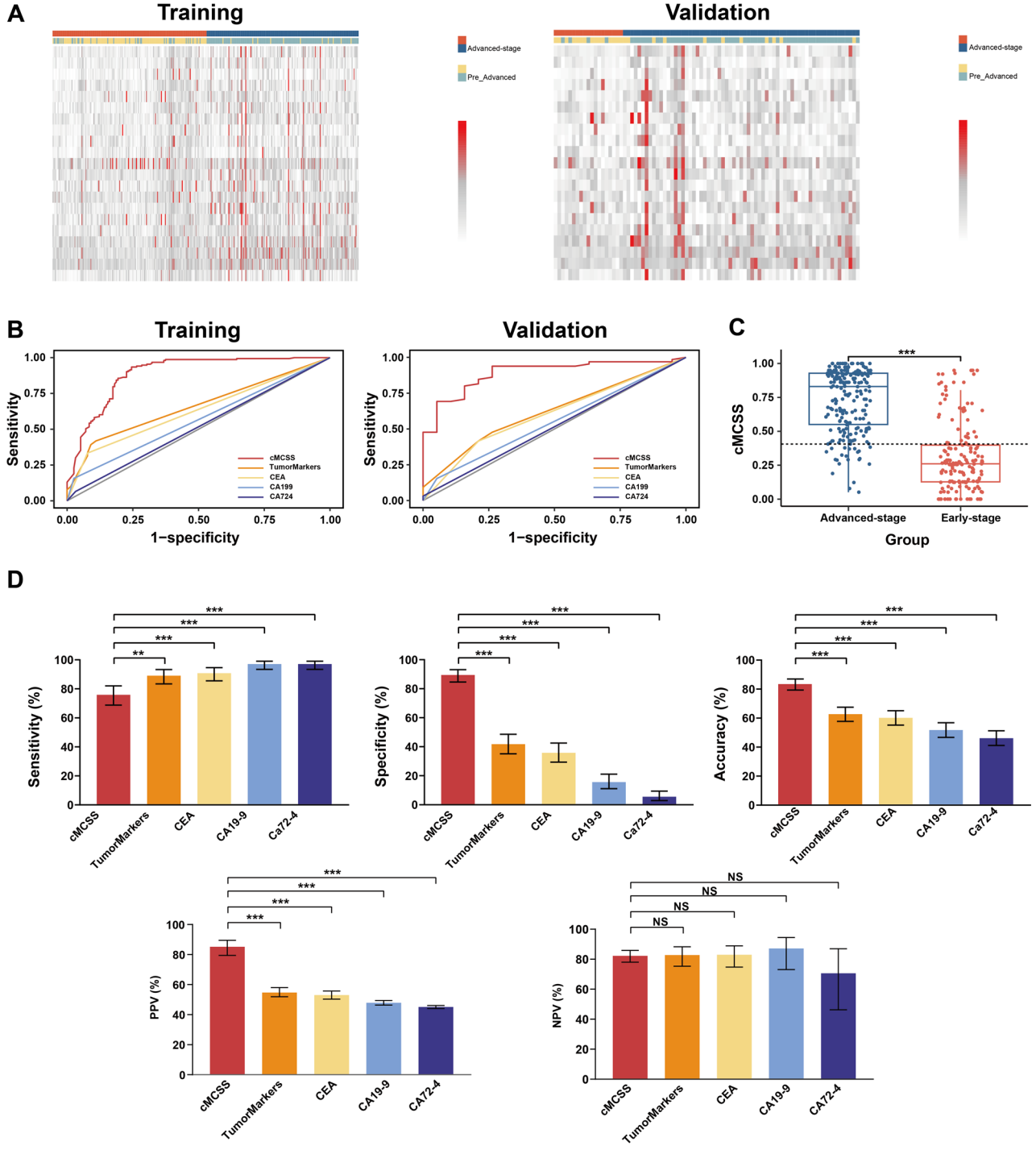

Fig. 6 The performance of the cMCSS and comparison with tumor markers. **A** Heatmap of DMRs in the training (after oversampling) and validation datasets. Clinical features and predicted outcomes are indicated by the annotation bars above the heatmaps. **B** The ROCs of the cMCSS and tumor markers for stratifying the stages in the training and validation datasets. **C** Distribution of the risk score of the cMCSS in the two groups. The risk score in the Advanced-stage was significantly greater than that in the Early-stage. **D** Comparison of sensitivity, specificity, accuracy, PPV and NPV between the cMCSS and tumor markers: cMCSS (red), CEA (yellow), CA19-9 (blue), CA72-4 (purple) and the combination (orange). NS, not significant; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. *CEA* carcinoembryonic antigen. *CA19-9* carbohydrate antigen 19–9. *CA72-4* carbohydrate antigen 72–4

read-level α-values to identify biomarkers from impure tissue samples or cfDNA plasma samples [47]. Both of these approaches achieved satisfactory performance. In the discovery phase, thousands of colorectal tumor-specific DMRs were identified at the tissue-level. Normal cfDNA is derived mainly from white blood cells; thus, the background noise from PBMCs was also filtered out in this study. These innovations allowed our model to fully cover the methylation signatures and maximize the precision.

To explore the potential associations between the biomarkers and CRC development and evolution of time, we reviewed the literature. Among the selected biomarkers, chr17_75453474_75453580 constitutes part of *SPET9*, which has been widely studied. *SEPT9* contributes to the development and progression of colorectal cancer through signaling pathways such as the JNK and HIF-1 pathways [48, 49]. In addition to *SEPT9*, the associations between some of the selected biomarkers have also been studied. chr4_1196129_1196219 constitutes part of *SPON2*. *SPON2* was reported to induce M2-polarized tumor-associated macrophage infiltration through the upregulation of cytokines in tumor cells and further promote tumor growth and metastasis [50]. chr12_47610205_47610371 constitutes part of *miR4496*. As a tumor suppressor, upregulated *miR4496* could bind to the 3' UTR of β-catenin mRNA and decrease the expression of β-catenin, resulting in impaired self-renewal capacity of colon cancer-initiating cells and reduced tumor formation in mouse models [51]. chr12_109027369_109027683 constitutes part of *PCED1B*. In *vitro* and in *vivo* experiments demonstrated that *PCED1B* acts as a molecular sponge for *miR633*, negatively regulating its expression. Downstream *HOXA9* was subsequently evaluated, which could promote cancer cell proliferation, invasion, and inhibits apoptosis [52]. chr21_44167128_44167284 is a component of *PDE9A*. Bioinformatics analysis revealed that *PDE9A* has significant prognostic value in colorectal cancer [53]. chr16_89181747_89181976 constitutes part of *ACSF3*. *ACSF3* encodes acyl-CoA synthetase family member 3, which is important for mitochondrial fatty acid biosynthesis and energy metabolism. Mendelian randomization analysis revealed that *ACSF3* was linked to colorectal cancer at the mQTL-eQTL level [54].

As a novel ctDNA methylation-based approach, EpiproColon 2.0 has been widely tested for tumor screening. However, it could detect only 17.1%–27.6% of AA and 26.5%–52.6% of stage I-II CRC [55, 56]. These values are much lower than those obtained in our study. Another biomarker, cg10673833, was reported to be able to detect 30% of AA, which is also inferior to the cMCD [57]. In a previous study, a panel containing 191 methylation markers detected 79.0% AA and 86.6% CRC [43]. Another panel containing 149 methylation markers was able to detect 85% of stage I-II CRCs, with an AUROC of 0.956 [13]. The performance of the cMCD is comparable to that of panels with 149 or 191 biomarkers, but the cMCD contains a panel of only 21 differential biomarkers, which makes this novel test much more cost-effective and easier to implement in a clinical laboratory setting. In this research, the novel approach involving the use of 21 differential biomarkers as a single-test could not only assist the secondary prevention of CRC, but also stratify stage and guide optimal treatment selection, fulfilling a dual clinical need simultaneously.

Compared with radial surgery, ER is preferred because it is less invasive and is associated with lower morbidity rates. Prospective studies have demonstrated that 80% of high-risk T1 patients underwent radial surgery were overtreated [19, 22, 23]. Few studies have focused on developing accurate noninvasive methods to inform optimal treatment selection. In this study, we first constructed the cMCSS to stratify tumor stage, which yielded an AUROC of 0.89. The results showed that the cMCSS could discriminate 75.86% of the Early-stage, for whom ER could achieve curative intent, and 89.45% of the Advanced-stage. The overall accuracy was 83.42%. The cMCSS achieved lower sensitivity but significantly greater PPV than tumor markers did. The main reason for these results is that tumor markers usually show negative results in the clinic, which revealed that tumor markers were not suitable for stratifying these Early-stage patients. These results demonstrated the high confidence of the cMCSS in mitigating undertreatment and overtreatment. Currently, endoscopic optical diagnosis is mainly used to evaluate whether the lesion could be curatively resected with endoscopic resection. However, providing a timely diagnosis through subjective evaluation by the endoscopists is still challenging and susceptible to the expertise and interobserver variability [58, 59]. To develop an objective and accurate method, Takamatsu et al. [60] proposed a machine-learning-based model using HE stained histologic whole-slide images to predict LNM in T1-stage CRC. This model achieved AUROCs of 0.971 and 0.760 in the training and validation datasets, respectively. However, predicting with HE images means that coloscopy must be performed at least 2 times for patients. Considering the disadvantages of coloscopy, the compliance is further limited within the population [10]. Several studies have reported the development of objective and noninvasive methods. Miyazaki et al. [61] reported that cf-miRNAs and exosomal miRNAs could yield AUROCs of 0.82 and 0.86 for detecting LNM in T1-stage patients. The combination model of the two panels yielded an AUROC of 0.91.Ying et al. [62] reported that cancer-related inflammatory biomarkers could be used to detect stage I-II CRC, achieving an AUROC of

0.83. The aim of these studies was to identify LNM in patients with stage I and/or II colorectal cancer. By discriminating these patients, researchers could evaluate the necessity of radial resection and spare patients with a low-risk of the presence of LNM from overtreatment. However, only patients with relatively early-stage disease were included in these studies, which does not fully constitute the real portion of CRC. To address the problem, we developed a ctDNA methylation-based method, the cMCSS, with a consecutive cohort of patients with AA and all-stage CRC. Patients with AA and pT1N0M0 CRC constituted the Early-stage group, who were eligible for curative endoscopic resection. Patients with more advanced-stage disease constituted the Advanced-stage group. With the application of the cMCSS, patients could be evaluated and the treatment strategy could be guided objectively at the first interval rather than by endoscopy. In addition, cMCD plus cMCSS, which were constructed using the same 21 biomarkers, for both early detection and guide treatment selection would be more popular for both patients and clinicians.

Despite the satisfactory outcomes obtained, this study has several limitations. First, this study is based on a single center cohort. Second, ADASYN was used to balance the data in this study. The application of ADASYN may make the classifier less effective because it fails to address the issue of class overlap [63]. Third, direct comparisons between the novel methods and other methylation-based technologies in the same cohort were not performed. The difference in the performance remains unclear. In the future, novel methods should be validated in multicenter, prospective trials to confirm the accuracy and generalizability, and direct comparisons with other methylation-based technologies should be performed.

Conclusion

In conclusion, the approach based on genome-wide ctDNA methylation profiling is a noninvasive and robust method for both early detection and stage stratification of CRC, which could diagnose and guide treatment selection simultaneously. We believe that this approach has potential for personalized diagnosis and treatment of CRC.

Abbreviations

CRC	colorectal cancer
CEA	carcinoembryonic antigen.
CA19-9	carbohydrate antigen 19-9
CA72-4	carbohydrate antigen 72-4
AA	advanced adenoma
ER	endoscopic resection
LNM	lymph node metastasis
ctDNA	circulating tumor DNA
cfDNA	cell-free DNA
CpG	cytosine–phosphoric–guanine
SEPT9	septin-9
NED	no evidence of disease
PBMC	peripheral blood mononuclear cells
DMR	differential methylation region
LASSO	least absolute shrinkage and selection operator
RFE	recursive feature elimination
SHAP	SHapley Additive exPlanation
XGBoost	eXtreme Gradient Boosting
AUROC	the area under the receiver operating characteristic curve
PPV	positive predictive value
NPV	negative predictive value
DCA	decision curve analysis
ADASYN	adaptive synthetic oversampling
IQR	interquartile range

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13148-026-02097-x.

Supplementary Material 1.

Acknowledgements

Not applicable.

Author contributions

H.J., M.X., W.J., and Y.H. contributed equally to this work. J.Y. and H.J. designed and supervised the research; M.X., Y.H., B.Y., Q.C., J.Z., W.W., Y.W., and Z.H. enrolled patients and collected the clinicopathological data; H.J., M.X. and Q.C. performed DNA isolation and methylation profiling; X.H., Z.C and H.L. analyzed the data; H.J., W.J., and Y.H. performed result interpretation; H.J., Y.H. and J.Y. completed the clinical trial registration and ethical procedures. H.J. and J.Y. wrote the paper with contributions from all other authors.

Funding

This work was supported by grants from the National Natural Science Foundation of China (82273360); the Science and Technology Planning Project of Guangzhou City (202206010085); the Fujian Province University-Industry Cooperation Project (2024Y4018); the Joint Funds for the Innovation of Science and Technology, Fujian Province (2024Y9282); the Postdoctoral Fellowship Program of CPSF (GZC20231069); the Clinical Research Program of Nanfang Hospital of Southern Medical University (2022CR003); the President Foundation of Nanfang Hospital, Southern Medical University (2023A053 and 2023B016); the Research Project of Fujian Medical University Union Hospital (2024XH030).

Data availability

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This prospective study was approved by the Institutional Ethical Review Board at Nanfang Hospital of Southern Medical University. This study is registered at https://register.clinicaltrials.gov with the identifier NCT05558436. Informed consent was obtained.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Guangdong Provincial Key Laboratory of Precision Medicine for Gastrointestinal Tumor, Department of General Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, P.R. China
[2]Department of Plastic and Cosmetic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, P.R. China

[3]Department of General Surgery, The Tenth Affiliated Hospital of Southern Medical University (Dongguan people's hospital), Southern Medical University, Guangzhou, P.R. China
[4]Department of Gastroenterology, Guangdong Provincial Key Laboratory of Gastroenterology, Nanfang Hospital, Southern Medical University, Guangzhou, P.R. China
[5]Acegen (Shenzhen) Ltd, Shenzhen, Guangdong, P.R. China
[6]Department of Colorectal Surgery, Fujian Medical University Union Hospital, Fuzhou, P.R. China

Received: 30 March 2025 / Accepted: 22 February 2026
Published online: 08 March 2026

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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):209–49.
3. Xia C, Dong X, Li H, Cao M, Sun D, He S, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. Chin Med J. 2022;135(5):584–90.
4. Joranger P, Nesbakken A, Sorbye H, Hoff G, Oshaug A, Aas E. Survival and costs of colorectal cancer treatment and effects of changing treatment strategies: a model approach. Eur J health economics: HEPAC : health Econ Prev care. 2020;21(3):321–34.
5. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2018;68(6):394–424.
6. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal Cancer Screening: Recommendations for Physicians and Patients From the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology. 2017;153(1):307–23.
7. Shaukat A, Kahi CJ, Burke CA, Rabeneck L, Sauer BG, Rex DK. ACG clinical guidelines: colorectal cancer screening 2021. Am J Gastroenterol. 2021;116(3):458–79.
8. Wada Y, Shimada M, Murano T, Takamaru H, Morine Y, Ikemoto T, et al. A liquid biopsy assay for noninvasive identification of lymph node metastases in T1 colorectal cancer. Gastroenterology. 2021;161(1):151-62.e1.
9. Wu X, Zhang Y, Hu T, He X, Zou Y, Deng Q, et al. A novel cell-free DNA methylation-based model improves the early detection of colorectal cancer. Mol Oncol. 2021;15(10):2702–14.
10. Gimeno García AZ. Factors influencing colorectal cancer screening participation. Gastroenterol Res Pract. 2012;2012:483417.
11. Chen H, Li N, Ren J, Feng X, Lyu Z, Wei L, et al. Participation and yield of a population-based colorectal cancer screening programme in China. Gut. 2019;68(8):1450–7.
12. Navarro M, Nicolas A, Ferrandez A, Lanas A. Colorectal cancer population screening programs worldwide in 2016: an update. World J Gastroenterol. 2017;23(20):3632–42.
13. Zhao F, Bai P, Xu J, Li Z, Muhammad S, Li D, et al. Efficacy of cell-free DNA methylation-based blood test for colorectal cancer screening in high-risk population: a prospective cohort study. Mol Cancer. 2023;22(1):157.
14. Palmqvist R, Engarås B, Lindmark G, Hallmans G, Tavelin B, Nilsson O, et al. Prediagnostic levels of carcinoembryonic antigen and CA 242 in colorectal cancer: a matched case-control study. Dis Colon Rectum. 2003;46(11):1538–44.
15. Chen CC, Yang SH, Lin JK, Lin TC, Chen WS, Jiang JK, et al. Is it reasonable to add preoperative serum level of CEA and CA19-9 to staging for colorectal cancer? J Surg Res. 2005;124(2):169–74.
16. Malla M, Loree JM, Parikh AR, Kasi PM. Using circulating tumor DNA in colorectal cancer: current and evolving practices. J Clin Oncol. 2022;40(24):2846–57.
17. Buskermolen M, Cenin DR, Helsingen LM, Guyatt G, Vandvik PO, Haug U, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a microsimulation modelling study. BMJ Clin Res Ed. 2019;367:l5383.
18. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med. 2012;366(8):697–706.
19. Vleugels JLA, Koens L, Dijkgraaf MGW, Houwen B, Hazewinkel Y, Fockens P, et al. Suboptimal endoscopic cancer recognition in colorectal lesions in a national bowel screening programme. Gut. 2020;69(6):977–80.
20. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2021;19(3):329–59.
21. Kandimalla R, Ozawa T, Gao F, Wang X, Goel A. Gene expression signature in surgical tissues and endoscopic biopsies identifies high-risk T1 colorectal cancers. Gastroenterology. 2019;156(8):2338-e413.
22. Ikematsu H, Yoda Y, Matsuda T, Yamaguchi Y, Hotta K, Kobayashi N, et al. Long-term outcomes after resection for submucosal invasive colorectal cancers. Gastroenterology. 2013;144(3):551–9. quiz e14.
23. Pimentel-Nunes P, Libânio D, Bastiaansen BAJ, Bhandari P, Bisschops R, Bourke MJ, et al. <article-title update="added">Endoscopic submucosal dissection for superficial gastrointestinal lesions: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2022. Endoscopy. 2022;54(6):591–622.
24. Muto T, Oya M. Recent advances in diagnosis and treatment of colorectal T1 carcinoma. Dis Colon Rectum. 2003;46(10 Suppl):S89–93.
25. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11(6):426–37.
26. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: current technology and clinical applications. J Hematol Oncol. 2022;15(1):131.
27. Luo H, Wei W, Ye Z, Zheng J, Xu RH. Liquid biopsy of methylation biomarkers in cell-free DNA. Trends Mol Med. 2021;27(5):482–500.
28. Esteller M. Epigenetics in cancer. N Engl J Med. 2008;358(11):1148–59.
29. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14(9):985–90.
30. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 2001;61(4):1659–65.
31. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut. 2014;63(2):317–25.
32. deVos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J, et al. Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. Clin Chem. 2009;55(7):1337–46.
33. Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, Davis EM, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2021;325(19):1965–77.
34. Pedersen SK, Symonds EL, Baker RT, Murray DH, McEvoy A, Van Doorn SC, et al. Evaluation of an assay for methylated BCAT1 and IKZF1 in plasma for detection of colorectal neoplasia. BMC Cancer. 2015;15:654.
35. Barták BK, Kalmár A, Péterfia B, Patai ÁV, Galamb O, Valcz G, et al. Colorectal adenoma and cancer detection based on altered methylation pattern of SFRP1, SFRP2, SDC2, and PRIMA1 in plasma samples. Epigenetics. 2017;12(9):751–63.
36. Huang H, Cao W, Long Z, Kuang L, Li X, Feng Y, et al. DNA methylation-based patterns for early diagnostic prediction and prognostic evaluation in colorectal cancer patients with high tumor mutation burden. Front Oncol. 2022;12:1030335.
37. Li W, Li Q, Kang S, Same M, Zhou Y, Sun C, et al. CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. Nucleic Acids Res. 2018;46(15):e89.
38. Li J, Wei L, Zhang X, Zhang W, Wang H, Zhong B, et al. Dismir: deep learning-based noninvasive cancer detection by integrating DNA sequence and methylation information of individual cell-free DNA reads. Brief Bioinform. 2021. https://doi.org/10.1093/bib/bbab250.
39. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56–67.
40. Han Z, Zhang Z, Yang X, Li Z, Sang S, Islam MT, et al. Development and interpretation of a pathomics-driven ensemble model for predicting the response to immunotherapy in gastric cancer. J Immunother Cancer. 2024. https://doi.org/10.1136/jitc-2024-008927.
41. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA Cancer J Clin. 2023;73(3):233–54.
42. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin. 2019;69(5):363–85.

43. Mo S, Dai W, Wang H, Lan X, Ma C, Su Z, et al. Early detection and prognosis prediction for colorectal cancer by circulating tumour DNA methylation haplotypes: a multicentre cohort study. EClinicalMedicine. 2023;55:101717.
44. Kim JB, Lee HS, Lee HJ, Kim J, Yang DH, Yu CS, et al. Long-Term Outcomes of Endoscopic Versus Surgical Resection of Superficial Submucosal Colorectal Cancer. Dig Dis Sci. 2015;60(9):2785–92.
45. Kanwal R, Gupta K, Gupta S. Cancer epigenetics: an introduction. Methods in molecular biology (Clifton, NJ). 2015;1238:3–25.
46. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer. 2011;11(10):726–34.
47. Stackpole ML, Zeng W, Li S, Liu CC, Zhou Y, He S, et al. Cost-effective methylome sequencing of cell-free DNA for accurately detecting and locating cancer. Nat Commun. 2022;13(1):5566.
48. Gonzalez ME, Makarova O, Peterson EA, Privette LM, Petty EM. Up-regulation of SEPT9_v1 stabilizes c-Jun-N-terminal kinase and contributes to its pro-proliferative activity in mammary epithelial cells. Cell Signal. 2009;21(4):477–87.
49. Jędrzejczak P, Saramowicz K, Kuś J, Barczuk J, Rozpędek-Kamińska W, Siwecka N, et al. SEPT9_i1 and Septin Dynamics in Oncogenesis and Cancer Treatment. Biomolecules. 2024. https://doi.org/10.3390/biom14091194.
50. Huang C, Ou R, Chen X, Zhang Y, Li J, Liang Y, et al. Tumor cell-derived SPON2 promotes M2-polarized tumor-associated macrophage infiltration and cancer progression by activating PYK2 in CRC. Journal of experimental & clinical cancer research : CR. 2021;40(1):304.
51. Kang DW, Choi CY, Cho YH, Tian H, Di Paolo G, Choi KY, et al. Targeting phospholipase D1 attenuates intestinal tumorigenesis by controlling β-catenin signaling in cancer-initiating cells. J Exp Med. 2015;212(8):1219–37.
52. Liu J, Qian J, Mo Q, Tang L, Xu Q. Long non-coding RNA PCED1B-AS1 promotes the proliferation of colorectal adenocarcinoma through regulating the miR-633/*HOXA9* axis. Bioengineered. 2022;13(3):5407–20.
53. Susmi TF, Rahman A, Khan MMR, Yasmin F, Islam MS, Nasif O, et al. Prognostic and clinicopathological insights of phosphodiesterase 9A gene as novel biomarker in human colorectal cancer. BMC Cancer. 2021;21(1):577.
54. Zhang Z, Lu S, Peng L, Ge F, Zhang B, Zhang Y, et al. Causal association between mitochondrial genes and colorectal cancer: a multi-omics Mendelian randomization study. Discov Oncol. 2025;16(1):1864.
55. Ma ZY, Law WL, Ng EKO, Chan CSY, Lau KS, Cheng YY, et al. Methylated septin 9 and carcinoembryonic antigen for serological diagnosis and monitoring of patients with colorectal cancer after surgery. Sci Rep. 2019;9(1):10326.
56. Sun J, Fei F, Zhang M, Li Y, Zhang X, Zhu S, et al. The role of (m)SEPT9 in screening, diagnosis, and recurrence monitoring of colorectal cancer. BMC Cancer. 2019;19(1):450.
57. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med. 2020. https://doi.org/10.1126/scitranslmed.aax7533.
58. Moss A, Bourke MJ, Williams SJ, Hourigan LF, Brown G, Tam W, et al. Endoscopic mucosal resection outcomes and prediction of submucosal cancer from advanced colonic mucosal neoplasia. Gastroenterology. 2011;140(7):1909–18.
59. Yao L, Lu Z, Yang G, Zhou W, Xu Y, Guo M, et al. Development and validation of an artificial intelligence-based system for predicting colorectal cancer invasion depth using multi-modal data. Dig Endosc. 2023;35(5):625–35.
60. Takamatsu M, Yamamoto N, Kawachi H, Nakano K, Saito S, Fukunaga Y, et al. Prediction of lymph node metastasis in early colorectal cancer based on histologic images by artificial intelligence. Sci Rep. 2022;12(1):2963.
61. Miyazaki K, Wada Y, Okuno K, Murano T, Morine Y, Ikemoto T, et al. An exosome-based liquid biopsy signature for pre-operative identification of lymph node metastasis in patients with pathological high-risk T1 colorectal cancer. Mol Cancer. 2023;22(1):2.
62. Ying HQ, Chen W, Xiong CF, Wang Y, Li XJ, Cheng XX. Quantification of fibrinogen-to-pre-albumin ratio provides an integrating parameter for differential diagnosis and risk stratification of early-stage colorectal cancer. Cancer Cell Int. 2022;22(1):137.
63. Vuttipittayamongkol P, Elyan E. Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. Inf Sci. 2020;509:47–70.

Publisher's Note