

1 **Early Diagnosis and Prognostic Prediction of Colorectal Cancer through Plasma**  
2 **Methylation Regions**

3 Lingqin Zhu<sup>1\*</sup>, Lang Yang<sup>2,3\*</sup>, Fangli Men<sup>4\*</sup>, Jianwei Yu<sup>5\*</sup>, Shuyang Sun<sup>1</sup>, Chenguang Li<sup>1</sup>,  
4 Xianzong Ma<sup>3</sup>, Junfeng Xu<sup>2,3</sup>, Yangjie Li<sup>3</sup>, Ju Tian<sup>6</sup>, Xin Wang<sup>3</sup>, Hui Xie<sup>3</sup>, Qian Kang<sup>1</sup>,  
5 Linghui Duan<sup>3</sup>, Xiang Yi<sup>7</sup>, Wei Guo<sup>8</sup>, Xueqing Gong<sup>8</sup>, Ni Guo<sup>4</sup>, Youyong Lu<sup>9#</sup>, Joseph  
6 Leung<sup>10#</sup>, Yuqi He<sup>1, 3#</sup>, Jianqiu Sheng<sup>2,3#</sup>

7 1. Department of Gastroenterology, Beijing Chest Hospital, Beijing Tuberculosis and  
8 Thoracic Tumor Research Institute, Laboratory for Clinical Medicine, Capital  
9 Medical University, No. 9 Beiguan Street, Tongzhou District, Beijing 101149, China

10 2. Senior Department of Gastroenterology, The First Medical Center of Chinese  
11 People's Liberation Army General Hospital, No. 28 of Fuxing Road, Haidian District,  
12 Beijing 100853, China;

13 3. Department of Gastroenterology, Seventh Medical Center of Chinese PLA General  
14 Hospital, No.5 Nanmencang, Dongcheng District, Beijing 100700, China;

15 4. Department of Gastroenterology, Dongying People's Hospital, Nan Yi Lu 317,  
16 Dongying District, Dongying 257000, China

17 5. Department of Gastroenterology, Longyan First Affiliated Hospital of Fujian  
18 Medical University, 105 Jiuyi Road (north), Longyan 364000, China

19 6. Department of Hepatobiliary and Pancreatic Surgery, The First Medical Center of  
20 Chinese PLA General Hospital, Institute of Hepatobiliary Surgery and Key  
21 Laboratory of Digital Hepetobiliary Surgery of Chinese PLA General Hospital, No.  
22 28 of Fuxing Road, Haidian District, Beijing 100853, China;

23 7. Mega Genomics Limited, 401 Health Work, North Garden Road, Haidian District,  
24 Beijing 100083, China;

25 8. Shanghai Yingce Biotechnology Co., Ltd, Room 202, 203, 2nd Floor, No. 4, Lane  
26 288, Mingnan Road, Songjiang District, Shanghai 201613, China;

27 9. Laboratory of Molecular Oncology, Key Laboratory of Carcinogenesis and  
28 Translational Research (Ministry of Education), Peking University Cancer  
29 Hospital/Institute, No. 52 Fucheng Road, Haidian District, Beijing 100142, China

30 10. Section of Gastroenterology, Sacramento VA Medical Center, Mather, California

31 \*These authors contributed equally to this work.

32 Word count of the manuscript text (3365)

33 #Correspondence to:

34 Email: [youyonglu@hsc.pku.edu.cn](mailto:youyonglu@hsc.pku.edu.cn)

35 Prof. Youyong Lu, MD, Ph.D: Laboratory of Molecular Oncology, Key Laboratory of  
36 Carcinogenesis and Translational Research (Ministry of Education), Peking University  
37 Cancer Hospital/Institute, No. 52 Fucheng Road, Haidian District, Beijing 100142, China;

38 Email: [ercpmaster@hotmail.com](mailto:ercpmaster@hotmail.com)

39 Prof. Joseph Leung, Section of Gastroenterology, Sacramento VA Medical Center, 10535  
40 Hospital Way, 111/G, Mather, CA 95655. Tel: (916)366-5339, Email: [jwleung@ucdavis.edu](mailto:jwleung@ucdavis.edu)

41 Email: [endohe@163.com](mailto:endohe@163.com)

42 Prof. Yuqi He, MD, Ph.D: Department of Gastroenterology, Beijing Chest Hospital, Capital  
43 Medical University, Beijing Tuberculosis and Thoracic Tumor Research Institute, No. 9

44 Beiguan Street, Tongzhou District, Beijing 101149, China; Department of Gastroenterology,  
45 the 7<sup>th</sup> Medical Center of Chinese PLA General Hospital, No.5 Nanmencang, Dongcheng  
46 District, Beijing 100700, China;  
47 Email: [jianqiu@263.net](mailto:jianqiu@263.net)  
48 Prof. Jianqiu Sheng, MD, Ph.D: Department of Gastroenterology, the 7<sup>th</sup> Medical Center of  
49 Chinese PLA General Hospital, No.5 Nanmencang, Dongcheng District, Beijing 100700,  
50 China. Phone: +86-10-66721299; Fax: +86-10-66721024;  
51  
52

53 **Abstract**

54 The methylation of plasma cell-free DNA (cfDNA) has emerged as a valuable diagnostic and  
55 prognostic biomarker in various cancers including colorectal cancer (CRC). Currently, there  
56 are no biomarkers that serve simultaneously for early diagnosis and prognostic prediction in  
57 CRC patients. Herein, we developed a plasma panel (27 DMRs, differential methylated  
58 regions) and validated its superior performance across CRC diagnosis and prognosis  
59 prediction in an independent cohort. We first conducted a preliminary screening of 119 CRC  
60 tissue samples to identify CRC-specific methylation features. Subsequently, a CRC-specific  
61 methylation panel was developed by further filtering 161 plasma samples. Then machine  
62 learning algorithms were applied to develop diagnosis and prognosis models using cfDNA  
63 samples from 51 CRC patients and 33 normal controls. The diagnosis model was tested in a  
64 cohort consisting of 30 CRC, 37 advanced adenoma (AA), and 14 healthy plasma samples,  
65 independently validated in a cohort consisting of 18 CRC, 91 NAA, 23 AA and 34 healthy  
66 plasma samples. In the tissue external validation cohort (GSE48684), the cfDNA methylation  
67 diagnosis model conducted with the panel, have the area under the curve (AUC) reached  
68 0.983, and for the plasma cfDNA model in the external validation cohort, the sensitivities for  
69 NAA, AA and CRC 0-II are 48.4%, 52.2% and 66.7% respectively, with a specificity of 88%.  
70 Additionally, the panel was applied to patient staging and metastasis, performing well in  
71 predicting CRC distant metastasis (AUC = 0.955) and prognosis (AUC = 0.867). Using  
72 normal samples as control, the changes in methylation score in both tissue and plasma were  
73 consistent across different lesions, although the degree of alterations varied with severity. The  
74 methylation scores vary between paired tissue and blood samples, suggesting distinct

75 mechanisms of migration from tumor tissue to blood for the 27 DMRs. Together, Our cfDNA  
76 methylation models based on 27 DMRs can identify different stages of CRC and predict  
77 metastasis and prognosis, ultimately enabling early intervention and risk stratification for  
78 CRC patients.

79

80 **Keywords: Colorectal cancer; cell-free DNA Methylation; Diagnosis prediction;**

81 **Prognosis prediction**

82

## 83 **Introduction**

84 Colorectal cancer (CRC) is the third most prevalent cancer globally and ranks as the fourth  
85 leading cause of cancer death, accounting for 9.4% of cancer-related mortalities <sup>(1)</sup>. When  
86 patients were diagnosed with CRC, over half of their family (62.9%) faced financial burdens  
87 <sup>(2,3)</sup>. Extensive research underscores that the survival rates of individuals diagnosed with  
88 advanced CRC (Stage III or IV) witness a significant decrement <sup>(4,5)</sup>. Early detection and  
89 removal of precancerous lesions remain the most effective strategy to prevent CRC-associated  
90 mortality <sup>(6)</sup>. While colonoscopy is regarded as the gold standard for CRC detection, its  
91 limitations including invasiveness, suboptimal patient compliance and risk of complications  
92 including intestinal perforation, warrant further consideration <sup>(7)</sup>. Guaiac-based fecal occult  
93 blood test (gFOBT), fecal immunochemical test (FIT), carcino-embryonic antigen (CEA), and  
94 multi-target stool DNA (mt-sDNA) testing, are constrained in clinical application due to the  
95 lower sensitivity<sup>(8,9)</sup>.

96 Recent research demonstrated a profound correlation between CRC and the development of  
97 genetic and epigenetic alterations. During early stage of CRC development, epigenetic  
98 modifications surpass the frequency of gene mutations, indicating their potential as diagnostic  
99 biomarkers in screening for colon polyps and cancer<sup>(10)</sup>. Circulating tumor DNA (ctDNA) in  
100 cell-free DNA (cfDNA) primarily come from apoptotic and necrotic tumor cells, carrying  
101 cancer-specific epigenetic alterations <sup>(11)</sup>. Notably, blood cfDNA methylation emerges as a  
102 promising cancer screening pathway because of its early appearance in tumorigenesis and  
103 abundant signal density <sup>(12)</sup>.

104 The value of cfDNA methylation in early diagnosis, detection of recurrence, molecular

105 subtyping and prognostic prediction of CRC has been proven<sup>(13)</sup>. However, there are still  
106 reported limitations. One study found a diagnostic panel for CRC demonstrated great  
107 performance, but it was derived from CRC tissue and normal blood leukocyte methylation  
108 data, potentially introducing bias due to the inconsistent sample types<sup>(14)</sup>. Another study  
109 addressed the potential of cfDNA methylation in patient risk stratification, but the  
110 requirement for blood sampling every three months led to poor patient compliance<sup>(15)</sup>.  
111 In this study, we aimed to develop plasma biomarkers derived from CRC tissues that exhibit  
112 superior performance in the diagnosis, metastasis and prognostic prediction for CRC. We  
113 meticulously screened CRC and normal tissues, and refining our selection within plasma  
114 samples. More importantly, the substantial proportion of paired tissue and plasma samples  
115 (from the same patient), effectively minimized potential data bias. The final resulting panel  
116 comprised 27 differential methylated regions (DMRs), with the area under the curve (AUC)  
117 reached 0.983 in the tissue validation cohort. Given that DNA methylation markers in plasma  
118 primarily stem from tumor tissues, these 27 DMRs exhibit significant potential in CRC  
119 plasma diagnostics. With a single blood test, we can ascertain the presence of CRC,  
120 distinguish the specific stages of the lesion (advanced adenoma, CRC 0-II, CRC III-IV), help  
121 identify distant metastasis, and achieve optimal risk stratification for CRC patients. This test  
122 is of clinical significance and holds promise in guiding the diagnosis and treatment for CRC.  
123

## 124 **Materials and Methods**

### 125 **Study design and samples**

126 Specific DNA methylation markers for CRC were identified through analysis of TCGA public  
127 database data, alongside collected tissue and plasma samples. Subsequently, diagnosis,  
128 metastasis, and prognosis models for CRC were established and validated. The 450k chip  
129 methylation data encompassing colorectal, esophageal, gastric, lung, liver, and breast cancers,  
130 and CRC transcriptome sequencing data were downloaded from The Cancer Genome Atlas  
131 (TCGA) database. The methylation data of tissue validation cohort was obtained from  
132 Fifteen pairs of tissue samples from the Seventh Medical Center of PLA General Hospital  
133 were sequenced with the Roche Nimble Gen Seq Cap Epi method. The 89 tissue samples and  
134 77 plasma samples used for marker screening were collected between June and December  
135 2020 at the Seventh Medical Center of PLA General Hospital and Dongying People's Hospital.  
136 Participants for model establishment and validation were enrolled from June 2020 to July  
137 2022 in the Seventh Medical Center of PLA General Hospital, Dongying People's Hospital.  
138 The participants for external validation cohort were enrolled from June 2021 to December  
139 2022 at the First Hospital of Longyan, Fujian Medical University. Inclusion and exclusion  
140 criteria are outlined in the Supplementary Materials. Detailed methods for DNA extraction,  
141 library construction, targeted bisulfite sequencing, and methylation data processing are  
142 provided in the Supplementary Materials.

### 143 **CRC-specific methylation markers selection**

144 Combining TCGA methylation data with our 15 pairs of tissue methylation sequencing results,  
145 probes were designed for further screening in the 89 samples (13 normal tissues and 38 paired

146 CRC tissues), 404 DMRs were identified. The detailed methods were listed in supplementary  
147 methods. Then, the methylation levels of 77 plasma samples from 13 Normal, 15 non-  
148 advanced adenoma (NAA), 12 advanced adenoma (AA) and 37 CRC participants, were  
149 analyzed. Subsequently, the Least Absolute Shrinkage and Selection Operator (LASSO)  
150 regression analysis was used to filter markers, and 27 regions that appeared in more than 50  
151 iterations were chosen as plasma diagnostic markers. The relationship between the  
152 methylation levels of the 27 DMRs and the transcription levels of their respective genes was  
153 compared by Spearman correlation coefficients.

#### 154 **Construction of the CRC diagnosis model**

155 The methylation data from the training cohort underwent machine learning, culminating in the  
156 establishment of a five-fold cross-validated model characterized by binary deviance  
157 minimization standards. The diagnosis model was constructed using logistic regression, and  
158 its performance was evaluated through Receiver Operating Characteristic (ROC) curve  
159 analysis. The optimal cutoff value was determined using the maximum Youden index. The  
160 methylation data of tissue validation cohort (GSE48684) was downloaded from GEO (Gene  
161 Expression Omnibus) dataset. Using the methylation scores derived from the diagnosis model,  
162 the CRC staging (AA, CRC 0-II, CRC III-IV) was predicted.

#### 163 **Construction of the CRC metastasis and prognosis models**

164 Utilizing the methylation scores derived from the diagnosis model, in conjunction with patient  
165 pathological parameters (metastasis), the plasma CRC metastasis model was constructed. The  
166 performance of this model underwent evaluation through ROC analysis, with the optimal  
167 cutoff value determined via the maximum Youden index. Furthermore, employing the same



168 methodology, the plasma prognosis model was established and validated, based on the  
169 methylation scores derived from the CRC diagnosis model and survival information from the  
170 patients.

#### 171 **Statistical analysis**

172 All statistical analyses were conducted using SPSS 26.0 or R 4.1.3. Two-sided tests were used  
173 for p-values, with differences deemed statistically significant at  $P < 0.05$ . Model performance  
174 was assessed through ROC analysis using the "roc" function from the R package pROC,  
175 generating AUC and a 95% confidence interval (CI). Kaplan-Meier (KM) curves were used  
176 for survival curve estimation, with comparisons made through the log-rank test and hazard  
177 ratios determined by Cox regression.

178

## 179 **Results**

### 180 **Patient characteristics**

181 To delineate DNA methylation biomarkers specific to CRC, 119 tissues and 77 plasma  
182 samples (originating from the same patients with tissues) were collected for methylation  
183 sequencing analysis, narrowing down to 142 DMRs. Then the 27 DMRs were selected with  
184 the sequencing data of 84 plasma samples. Following this, CRC diagnosis, metastasis, and  
185 prognosis models were constructed using plasma samples from 84 individuals (51 CRC, 33  
186 Normal). Validation was then carried out in an independent cohort (30 CRC, 37 AA, 14  
187 Normal). The CRC diagnosis model was further tested in an external validation cohort (18  
188 CRC 0-II, 91 NAA, 23 AA, 34 Normal). Detailed information about the study design is  
189 illustrated in Figure 1, and comprehensive patient characteristics are summarized in Table s2-  
190 4.

### 191 **Methylation markers selection**

192 Analyzing the 450k chip methylation data from TCGA, the differential methylated sites  
193 (DMCs) in CRC were identified (Figure s1A). Combining with additional CRC-related  
194 methylation sites reported in the literature, 1438 DMCs were selected for validation. The heat  
195 map depicted the methylation levels of 1438 DMCs, revealing distinct methylation patterns  
196 between Normal and Tumor. (Figure s1B). The top 1400 DMCs were carefully selected based  
197 on the results of DNA methylation sequencing of 15 paired tissue samples (Figure s2A-C).  
198 The distribution of the DMCs was predominantly observed in introns and promoters,  
199 signifying substantial implications for gene expression and cellular functions (Figure s2D).  
200 Merging the 1438 DMCs with the 1400 DMCs for probe design, 404 DMRs were discerned

201 across 89 tissues (Figure s3A). The different methylation patterns between Normal and Tumor  
202 were also quite evident (Figure s3B-C). These DMRs correspond to genes that play pivotal  
203 roles in biological processes such as cell-cell adhesion, digestive system development, and  
204 cAMP and cGMP pathways (Figure s3D). Subsequently, 142 regions remained after  
205 sequencing 77 plasma samples (Figure s3E-F). The final 27 DMRs were selected after  
206 analyzing methylation data of the training cohort samples. Detailed information of the 27  
207 DMRs is provided in Table s5. Correlation analysis revealed the methylation levels of the  
208 majority of the 27 DMRs are correlated with the transcription levels of their respective genes  
209 (Figure s4).

#### 210 **Development and validation of the CRC diagnosis model**

211 Using methylation data from 27 DMRs in tissues, we applied machine learning to construct a  
212 diagnosis model for CRC and the AUC reached a high value of 0.994 (Figure 2A). In the  
213 external tissue independent validation cohort (GSE48684), the AUC for CRC is 0.983 and for  
214 AA is 0.966. (Figure 2B-C). The sensitivities for AA and CRC were 95% and 94%,  
215 respectively (Figure 2D-E). Besides, the plasma-based diagnosis model for CRC was  
216 constructed with training cohort and validated in an external independent validation cohort,  
217 the 27 DMRs exhibited distinctive methylation patterns between CRC and normal individuals  
218 (Figure 3A-B). The methylation scores of the model were significantly elevated in the CRC  
219 group compared to the normal and AA groups ( $P < 0.001$ ), and they increased with advance  
220 tumor staging (Figure 3H). In the training cohort, the ROC curve analysis showed an AUC  
221 value of 0.928 [95% confidence interval (CI): 0.874–0.981] (Figure 3C). The sensitivities of  
222 the CRC diagnosis model in the training cohort for stages 0-II, III-IV, and all CRC stages

223 were 80%, 92%, and 86%, respectively, with a specificity of 85% (Figure 3I). ROC curve  
224 analysis in the validation cohort showed AUC values for AA, stages 0-II, III-IV, and all CRC  
225 stages as 0.714 (0.550–0.878), 0.890 (0.749–1.000), 0.967 (0.899–1.000), and 0.929 (0.827–  
226 1.000). (Figure 3D-G). The CRC diagnosis model demonstrated a specificity of 93% in the  
227 validation cohort, with sensitivities for AA, stages 0-II, III-IV, and all CRC stages of 43%,  
228 67%, 100%, and 83%, respectively (Figure 3I). We further tested the early diagnostic  
229 performance of the model in an external validation cohort, and the results showed that the  
230 model achieved a sensitivity of 52% for AA and 48% for NAA (Figure 4A-E),  
231 The plasma diagnosis model also serves to discern the specific staging of CRC patients. ROC  
232 curve analysis on the validation cohort revealed an AUC of 0.813 (0.656–0.970) for  
233 discriminating between CRC 0-II and CRC III-IV, and an AUC of 0.799 (0.693–0.905) for  
234 distinguishing between AA and CRC (Figure s5A-C). Kaplan-Meier survival curves revealed  
235 a significant reduction in overall survival (OS) for CRC III-IV patients compared to CRC 0-II  
236 patients identified by this diagnosis model (Figure s5D-E).

### 237 **Development and validation of the CRC metastasis and prognosis models**

238 Through an analysis of the relationship between the methylation scores in the diagnosis model  
239 of CRC patients' plasma samples and clinical-pathological parameters, we observed that  
240 methylation scores are associated with metastasis and staging ( $P < 0.001$ , Figure 5A-B), but  
241 not with age, gender, and lesion location ( $P > 0.05$ , Figure s6A-C). This suggests the possible  
242 role of 27 DMRs in metastasis and prognosis prediction. Therefore, we developed plasma-  
243 based CRC metastasis and prognosis models based on the 27 CRC-specific DMRs. In the  
244 training and validation cohorts, the metastasis model demonstrated high AUCs of 0.969 (95%

245 CI: 0.926-1) and 0.955 (95% CI: 0.878-1), respectively (Figure 5C). Methylation scores of  
246 M1-CRC individuals identified from the metastasis model were significantly higher than M0-  
247 CRC individuals (Figure s6D). Additionally, the M1-CRC patients exhibited shorter OS than  
248 M0-CRC patients. (Figure s6E-F). For the prognosis model, ROC curves demonstrated  
249 excellent performance in the training cohort (AUC = 0.883, 95% CI: 0.778-0.988) and  
250 validation cohort (AUC = 0.867, 95% CI = 0.728-1). (Figure 5D). Furthermore, Kaplan-Meier  
251 survival curves revealed a significant reduction in OS for the high-risk group of CRC patients  
252 identified by this prognosis model (Figure 5E). Based on the cutoff value determined by the  
253 algorithm, we divided CRC patients into high-risk and low-risk groups. Notably, almost all  
254 patients who developed distant metastases and those who had died belonged to the high-risk  
255 group (Figure 5F, H). We further characterized the survival status distribution and metastasis  
256 status between the two groups. As expected, the high-risk group had a higher proportion of  
257 deceased individuals, while non-metastatic patients were more prominent in the low-risk  
258 group (Figure 5G, I). These results indicate that the metastasis and prognosis models  
259 successfully identified patients who require further treatment. Multivariate regression analysis  
260 revealed a substantial correlation between the methylation score of the 27 DMRs and OS,  
261 indicating methylation score as an independent prognostic factor for CRC (Table s6). These  
262 findings underscore the considerable potential of the prognosis model based on the 27 DMRs  
263 specific to CRC in predicting the prognosis and conducting risk stratification for CRC  
264 patients.

### 265 **Changes in methylation scores of tissue and plasma samples across different populations**

266 Collectively, the CRC diagnosis model, established based on the methylation levels in these

267 27 DMRs within tissues, exhibits robust performance. However, the plasma-based diagnosis  
268 model exhibited reduced performance in AA and CRC 0-II compared to CRC III-IV.  
269 Therefore, we analyzed the methylation scores of the 27 DMRs generated from diagnosis  
270 models in paired tissue and plasma samples. Notably, in tissues, the methylation scores are  
271 ranked as AA > CRC > Normal (Figure 6A), suggesting significant alterations in the  
272 methylation of the 27 DMRs at the onset of precancerous lesions. In plasma, methylation  
273 scores gradually increase with CRC progression (CRC III-IV > CRC 0-II > AA > Normal,  
274 Figure 6B). Further analysis of individual DMR methylation levels in the blood revealed  
275 discernible distinctions in certain DMRs during both AA and CRC 0-II stages, while  
276 variances in another subset of DMRs became detectable exclusively during CRC III-IV stages  
277 (Figure 6C). We then conducted a gene analysis of DMRs identifiable in early-stage CRC  
278 blood on the Metascope website. Our findings revealed their association with processes such  
279 as "secretion by cell" and "cell-cell adhesion." (Figure s6G) Consequently, we hypothesize  
280 that although alterations of 26 DMRs were observed in AA tissues, only a selected few are  
281 actively released into the bloodstream through cell secretion by tumor cells. The majority of  
282 these alterations are likely to be identified in the blood only after the occurrence of apoptosis  
283 or necrosis in tumor cells. These results elucidate the inconsistency in the performance of  
284 tissue and plasma diagnosis models and the differences in diagnostic efficacy for different  
285 CRC stages using the same model.  
286

287 **Discussion**

288 Liquid biopsy of cfDNA has become an ideal clinical detection method because it is  
289 minimally invasive and easily sampled. However, due to the low concentration of ctDNA in  
290 bodily fluids and the heterogeneity of tumor cells, many DNA markers for CRC currently lack  
291 certainty whether they originate from CRC tissues. In this study, we integrated TCGA tissue  
292 data with self-collected tissue methylation data to identify CRC-related methylation sites.  
293 Probes were designed and applied in paired tissue and plasma samples to further screen the  
294 methylation sites. We successfully determined 27 DMRs originating from CRC tissues, which  
295 were subsequently used to construct diagnostic model. In the validation cohorts of tissue and  
296 plasma samples, the diagnosis models based on the 27 DMRs we established could effectively  
297 differentiate samples between CRC and normal participants, in addition, the plasma diagnosis  
298 model could distinguish the different stages of CRC. Furthermore, the transfer prognosis  
299 model, established based on diagnosis model scores, could effectively demonstrate whether  
300 the CRC had metastasized, separating CRC patients into high-risk and low-risk groups.  
301 Several blood-based methylation biomarker candidates have been proposed for early  
302 detection of CRC. For instance, the FDA-approved circulating methylated SEPT9 DNA  
303 (mSEPT9) demonstrates sensitivities of 11.2%, 35.0%, 63.0%, 46.0%, and 77.4% for AA and  
304 CRC stages I-IV, respectively, with a specificity of 91.5%<sup>(16)</sup>. In a recent study, a cfDNA  
305 methylation-based CRC screening model has a sensitivity of 86.4% and a specificity of 90.7%,  
306 utilizing 149 markers derived from blood samples<sup>(17)</sup>. In our study, 27 markers were selected  
307 through a layered screening process from tissue and plasma, undergoing marker selection,  
308 model development, and validation to ensure model robustness. Then the superior

309 performance of our plasma diagnosis model was validated, with a sensitivity of 83.3% and  
310 specificity of 92.9%. Notably, the sensitivity for AA and NAA reached 52.2% and 48.4%  
311 respectively, far surpassing the 11.2% sensitivity of mSEPT9 and the 33.3% reported in  
312 another blood screening study.<sup>(16, 17)</sup> In summary, our diagnostic model performs well in  
313 detecting precancerous lesions of CRC and has the potential to become a screening method  
314 for high-risk populations of CRC.

315 In plasma, we observed the methylation scores of 27 DMRs gradually increase with CRC  
316 progression. It is possibly due to less vascular infiltration in early CRC <sup>(18, 19)</sup>. On the other  
317 hand, the ctDNA detected in the blood of early-stage CRC is largely derived from tumor cells  
318 actively secreting into the bloodstream, making it challenging to be precisely captured with  
319 current detection technologies due to the limited quantity. Conversely, ctDNA in the blood of  
320 late-stage CRC primarily emanated from apoptotic and necrotic tumor cells, resulting in a  
321 larger quantity that facilitates easier detection. However, in reality, the methylation scores of  
322 the 27 DMRs have indeed exhibited noticeable changes in AA and early cancerous tissues.

323 Additionally, in the tissue diagnosis model using the 27 DMRs, the sensitivity was 94% and  
324 the specificity was 100%. Therefore, we have grounds to believe that the 27 DMRs detected  
325 in plasma originate from CRC tumor tissues. With the implementation of more sensitive  
326 detection methods, the performance of these 27 DMRs in early CRC plasma diagnosis is  
327 likely to be further enhanced.

328 Additionally, cfDNA contributes to risk stratification and early recurrence detection in  
329 CRC<sup>(20)</sup>. However, current methods rely on continuous blood cfDNA testing by patients <sup>(15, 21)</sup>.

330 This study discovered that the preoperative cfDNA methylation level of 27 DMRs is linked to



331 distant metastasis and is valuable for predicting the prognosis of CRC, consistent with certain  
332 prior research. <sup>(22, 23)</sup>. Our model predicted AUCs of 0.955 and 0.867 for distant metastasis and  
333 prognosis in CRC, respectively. This underscores the potential utility of the preoperative  
334 application of this cfDNA methylation model for risk stratification, serving as an effective  
335 tool to improve the perioperative management of CRC patients.

336 The treatment strategies for CRC differ significantly across various stages. Molecular  
337 stratification approaches for CRC patients are on the rise, and concurrently, the clinical  
338 application of biomarkers to determine treatment decisions is gradually gaining traction <sup>(24)</sup>.  
339 Studies have demonstrated the precise identification of T1 CRC patients at risk of lymph node  
340 metastasis using a specific set of miRNAs, thereby potentially mitigating unnecessary  
341 overtreatment <sup>(25, 26)</sup>. Our CRC diagnosis model, with an AUC of 0.813 for distinguishing  
342 CRC 0-II from CRC III-IV, could also serve as a potent, straightforward, and cost-effective  
343 preoperative screening/detection method, guiding patients to select more appropriate  
344 treatment plans.

345 DNA methylation regulates gene transcription, guiding the progression from normal mucosa  
346 to AA and ultimately CRC. This process involves silencing tumor suppressor genes and  
347 activating oncogene transcription <sup>(27)</sup>. Gene transcription levels correlate notably with  
348 methylation levels at specific sites, consistent with our research findings. Additionally, several  
349 genes housing the identified 27 DMRs have been partially explored in CRC. Some genes  
350 contribute to tumor growth. For instance, high methylation and diminished expression of  
351 transmembrane protein 240 (TMEM240) regulate CRC cell proliferation, predicting poor  
352 prognosis <sup>(28)</sup>. The transcription factor homeobox A3 (HOXA3) activates aerobic glycolysis,

353 promoting tumor growth <sup>(29)</sup>. KIFC3 controls mitotic spindle assembly initiation <sup>(30)</sup>. Moreover,  
354 several genes are implicated in tumor metastasis, such as NOVA alternative splicing regulator  
355 1 (NOVA1), which promotes CRC migration by activating the Notch pathway <sup>(31)</sup>. Protein  
356 tyrosine phosphatase receptor type T (PTPRT) contributes to early CRC dissemination <sup>(32)</sup>.  
357 SPARC-related modular calcium binding 2 (SMOC2) serves as the distinctive signature of  
358 cancer stem cells (CSCs) in CRC and promotes epithelial-to-mesenchymal transition (EMT)  
359 <sup>(33,34)</sup>. Increased methylation and expression of pancreatic and duodenal homeobox 1 (PDX1)  
360 facilitate CRC invasion and migration<sup>(35)</sup>. However, the mechanisms by which alterations in  
361 specific DNA methylation impact gene expression are intricate. Certain transcription factors  
362 selectively recognize sequences with methylated CpG (mCpG) and influence the expression  
363 of multiple genes<sup>(36)</sup>. Further exploration of the potential functional mechanisms of these  
364 DNA methylation markers may deepen our understanding of the molecular processes  
365 underlying CRC development, offering promising therapeutic targets.

366 This study has certain limitations. Firstly, although obtained from multiple institutions, the  
367 small sample size was insufficient. Therefore, the cfDNA methylation model should be  
368 further validated in large-sample trials in the future. Secondly, a prospective study is  
369 necessary to compare or combine the cfDNA methylation model with clinically commonly  
370 used markers such as CEA and CA19-9.

371 In conclusion, our cfDNA methylation model based on 27 DMRs can identify different stages  
372 of CRC, predict metastasis and prognosis, and ultimately achieve early intervention and risk  
373 stratification for CRC patients. The preoperative application of our DNA methylation  
374 biomarker as a robust, convenient, and cost-effective detection method can contribute to

375 making more informed clinical decisions and improving the perioperative management of  
376 CRC patients.

377

#### 378 **AUTHOR CONTRIBUTIONS**

379 Yuqi He and Jianqiu Sheng conceived and designed the study. Lang Yang, Fangli Men,  
380 Jianwei Yu, Xianzong Ma, Junfeng Xu, Yangjie Li, Ju Tian, Hui Xie, Qian Kang, Linghui  
381 Duan, Xiang Yi, Wei Guo, Ni Guo collected the samples and curated the data. Xueqing Gong,  
382 Lingqin Zhu, Lang Yang, Fangli Men, Jianwei Yu, Shuyang Sun, Chenguang Li, Xianzong  
383 Ma, Junfeng Xu, Yangjie Li, Xiang Yi, Wei Guo, Ni Guo, Youyong Lu, Joseph Leung, Yuqi  
384 He and Jianqiu Sheng analysed the data. Lingqin Zhu wrote the manuscript with the  
385 assistance of Youyong Lu, Joseph Leung, Yuqi He and Jianqiu Sheng.

#### 386 **ACKNOWLEDGEMENTS**

387 This research was funded by grants from the National Natural Science Foundation of China  
388 (Grant no. 82273245), the Capital's Funds for Health Improvement and Research (Grant no.  
389 2022-1-5082), and the Beijing Natural Science Foundation (Grant no. 7212107), Sponsored  
390 by Dongying City Natural Science Foundation (Grant no. 2023ZR026), Sponsored by  
391 Longyan City Science and Technology Plan Project (Grant no. 2022LYF17082).

#### 392 **CONFLICT OF INTEREST STATEMENT**

393 The authors declare no conflict of interest.

#### 394 **DATA AVAILABILITY STATEMENT**

395 All of the data supporting this work will be made available from the corresponding author  
396 upon reasonable request.

397 **ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

398 The study was conducted in accordance with the Declaration of Helsinki, and approved by the

399 Ethics Committee of the Seventh Medical Center of PLA General Hospital (Approval No.

400 2016-70, 2020-78), Dongying People's Hospital (Approval No. DYYW-2019-002-01), and

401 the First Hospital of Longyan, Fujian Medical University (Approval No. 2021-k0001).

402 Informed consents were obtained from all participants involved in the study.

403

404 REFERENCES

- 405 1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer  
406 Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185  
407 Countries. *CA Cancer J Clin.* 2021;71(3):209-49. Epub 2021/02/05. doi: 10.3322/caac.21660. PubMed  
408 PMID: 33538338.
- 409 2. Ghazal LV, Abrahamse P, Ward KC, Morris AM, Hawley ST, Veenstra CM. Financial Toxicity and Its  
410 Association With Health-Related Quality of Life Among Partners of Colorectal Cancer Survivors. *JAMA*  
411 network open. 2023;6(4):e235897. Epub 2023/04/07. doi: 10.1001/jamanetworkopen.2023.5897.  
412 PubMed PMID: 37022684; PubMed Central PMCID: PMCPCMC10080378 National Cancer Institute (NCI)  
413 during the conduct of the study. Mr Abrahamse reported receiving grants from the National Institutes  
414 of Health during the conduct of the study and outside the submitted work. No other disclosures were  
415 reported.
- 416 3. Herriges MJ, Shenhav-Goldberg R, Peck JI, Bhanvadia SK, Morgans A, Chino F, et al. Financial  
417 Toxicity and Its Association With Prostate and Colon Cancer Screening. *Journal of the National*  
418 *Comprehensive Cancer Network : JNCCN.* 2022;20(9):981-8. Epub 2022/09/09. doi:  
419 10.6004/jnccn.2022.7036. PubMed PMID: 36075394.
- 420 4. Olivier T, Prasad V. Molecular testing to deliver personalized chemotherapy recommendations:  
421 risking over and undertreatment. *BMC medicine.* 2022;20(1):392. Epub 2022/11/10. doi:  
422 10.1186/s12916-022-02589-6. PubMed PMID: 36348413; PubMed Central PMCID: PMCPCMC9644653.
- 423 5. Patel SG, Karlitz JJ, Yen T, Lieu CH, Boland CR. The rising tide of early-onset colorectal cancer: a  
424 comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early  
425 detection. *The lancet Gastroenterology & hepatology.* 2022;7(3):262-74. Epub 2022/01/30. doi:  
426 10.1016/s2468-1253(21)00426-x. PubMed PMID: 35090605.
- 427 6. He X, Hang D, Wu K, Nayor J, Drew DA, Giovannucci EL, et al. Long-term Risk of Colorectal Cancer  
428 After Removal of Conventional Adenomas and Serrated Polyps. *Gastroenterology.* 2020;158(4):852-  
429 61.e4. Epub 2019/07/16. doi: 10.1053/j.gastro.2019.06.039. PubMed PMID: 31302144; PubMed  
430 Central PMCID: PMCPCMC6954345.
- 431 7. Hayman CV, Vyas D. Screening colonoscopy: The present and the future. *World J Gastroenterol.*  
432 2021;27(3):233-9. Epub 2021/02/02. doi: 10.3748/wjg.v27.i3.233. PubMed PMID: 33519138; PubMed  
433 Central PMCID: PMCPCMC7814366.
- 434 8. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool  
435 DNA testing for colorectal-cancer screening. *The New England journal of medicine.*  
436 2014;370(14):1287-97. Epub 2014/03/22. doi: 10.1056/NEJMoa1311194. PubMed PMID: 24645800.
- 437 9. Young GP, Symonds EL, Allison JE, Cole SR, Fraser CG, Halloran SP, et al. Advances in Fecal Occult  
438 Blood Tests: the FIT revolution. *Digestive diseases and sciences.* 2015;60(3):609-22. Epub 2014/12/11.  
439 doi: 10.1007/s10620-014-3445-3. PubMed PMID: 25492500; PubMed Central PMCID:  
440 PMCPCMC4366567.
- 441 10. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers.  
442 *Gastroenterology.* 2015;149(5):1204-25.e12. Epub 2015/07/29. doi: 10.1053/j.gastro.2015.07.011.  
443 PubMed PMID: 26216839; PubMed Central PMCID: PMCPCMC4589488.
- 444 11. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients.  
445 *Nature reviews Cancer.* 2011;11(6):426-37. Epub 2011/05/13. doi: 10.1038/nrc3066. PubMed PMID:  
446 21562580.

- 447 12. Zeng C, Stroup EK, Zhang Z, Chiu BC, Zhang W. Towards precision medicine: advances in 5-  
448 hydroxymethylcytosine cancer biomarker discovery in liquid biopsy. *Cancer communications (London,*  
449 *England)*. 2019;39(1):12. Epub 2019/03/30. doi: 10.1186/s40880-019-0356-x. PubMed PMID:  
450 30922396; PubMed Central PMCID: PMC6440138.
- 451 13. Malla M, Loree JM, Kasi PM, Parikh AR. Using Circulating Tumor DNA in Colorectal Cancer:  
452 Current and Evolving Practices. *Journal of clinical oncology : official journal of the American Society of*  
453 *Clinical Oncology*. 2022;40(24):2846-57. Epub 2022/07/16. doi: 10.1200/jco.21.02615. PubMed PMID:  
454 35839443; PubMed Central PMCID: PMC69390824.
- 455 14. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, et al. Circulating tumor DNA methylation profiles  
456 enable early diagnosis, prognosis prediction, and screening for colorectal cancer. *Science translational*  
457 *medicine*. 2020;12(524). Epub 2020/01/03. doi: 10.1126/scitranslmed.aax7533. PubMed PMID:  
458 31894106.
- 459 15. Mo S, Ye L, Wang D, Han L, Zhou S, Wang H, et al. Early Detection of Molecular Residual Disease  
460 and Risk Stratification for Stage I to III Colorectal Cancer via Circulating Tumor DNA Methylation. *JAMA*  
461 *oncology*. 2023;9(6):770-8. Epub 2023/04/20. doi: 10.1001/jamaoncol.2023.0425. PubMed PMID:  
462 37079312; PubMed Central PMCID: PMC691019774  
463 Genomics (Shanghai) Ltd. Dr H. Wang reported being an employee of Singlera Genomics (Shanghai) Ltd and having a patent pending for  
464 202180000598.2. Dr Liu reported being an employee of Singlera Genomics (Shanghai) Ltd and having  
465 a patent pending for 202180000598.2. No other disclosures were reported.
- 466 16. Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective  
467 evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut*.  
468 2014;63(2):317-25. Epub 2013/02/15. doi: 10.1136/gutjnl-2012-304149. PubMed PMID: 23408352;  
469 PubMed Central PMCID: PMC3913123.
- 470 17. Zhao F, Bai P, Xu J, Li Z, Muhammad S, Li D, et al. Efficacy of cell-free DNA methylation-based  
471 blood test for colorectal cancer screening in high-risk population: a prospective cohort study.  
472 *Molecular cancer*. 2023;22(1):157. Epub 2023/09/29. doi: 10.1186/s12943-023-01866-z. PubMed  
473 PMID: 37770864; PubMed Central PMCID: PMC6910538018.
- 474 18. Cho SS, Park JW, Kang GH, Kim JH, Bae JM, Han SW, et al. Prognostic Impact of Extramural  
475 Lymphatic, Vascular, and Perineural Invasion in Stage II Colon Cancer: A Comparison With Intramural  
476 Invasion. *Diseases of the colon and rectum*. 2023;66(3):366-73. Epub 2022/03/26. doi:  
477 10.1097/dcr.0000000000002339. PubMed PMID: 35333785.
- 478 19. Hanrahan V, Currie MJ, Gunningham SP, Morrin HR, Scott PA, Robinson BA, et al. The angiogenic  
479 switch for vascular endothelial growth factor (VEGF)-A, VEGF-B, VEGF-C, and VEGF-D in the adenoma-  
480 carcinoma sequence during colorectal cancer progression. *The Journal of pathology*. 2003;200(2):183-  
481 94. Epub 2003/05/20. doi: 10.1002/path.1339. PubMed PMID: 12754739.
- 482 20. Zhou H, Zhu L, Song J, Wang G, Li P, Li W, et al. Liquid biopsy at the frontier of detection,  
483 prognosis and progression monitoring in colorectal cancer. *Molecular cancer*. 2022;21(1):86. Epub  
484 2022/03/27. doi: 10.1186/s12943-022-01556-2. PubMed PMID: 35337361; PubMed Central PMCID:  
485 PMC69108951719.
- 486 21. Reinert T, Henriksen TV, Christensen E, Sharma S, Salari R, Sethi H, et al. Analysis of Plasma Cell-  
487 Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. *JAMA oncology*.  
488 2019;5(8):1124-31. Epub 2019/05/10. doi: 10.1001/jamaoncol.2019.0528. PubMed PMID: 31070691;  
489 PubMed Central PMCID: PMC69106512280  
Balcioglu, Hafez, Goel, Rabinowitz, Billings, Swenerton,

490 Aleshin, Lin, and Zimmermann, Messrs Sethi, Srinivasan, Olson, and Dashner, and Ms Navarro,  
491 reported receiving support from Natera Inc outside the submitted work. Dr Billings reported receiving  
492 support from Trovagene, OmniSeq, MissionBio, and Metastat outside the submitted work. Dr  
493 Zimmermann has a pending patent for Provisional. Dr Lindbjerg Andersen reported receiving grants  
494 from Novo Nordisk Foundation, Danish Council for Strategic Research, Danish Council for Independent  
495 Research, and Danish Cancer Society during the conduct of the study. No other disclosures were  
496 reported.

497 22. Tie J, Wang Y, Cohen J, Li L, Hong W, Christie M, et al. Circulating tumor DNA dynamics and  
498 recurrence risk in patients undergoing curative intent resection of colorectal cancer liver metastases:  
499 A prospective cohort study. *PLoS medicine*. 2021;18(5):e1003620. Epub 2021/05/04. doi:  
500 10.1371/journal.pmed.1003620. PubMed PMID: 33939694; PubMed Central PMCID:  
501 PMCPMC8128260 following competing interests: BV, KWK, & NP are founders of, and hold equity in  
502 Thrive Earlier Detection and Personal Genome Diagnostics. KWK & NP are on the Board of Directors of,  
503 and consultants to, Thrive Earlier Detection. KWK & BV are consultants to Sysmex, Eisai, Personal  
504 Genome Diagnostics and CAGE Pharma and hold equity in CAGE Pharma. KWK, BV, and NP are  
505 consultants to and hold equity in NeoPhore. BV is a consultant to and holds equity in Catalio Capital  
506 Management. NP is an advisor to and holds equity in CAGE Pharma. The companies named above, as  
507 well as other companies, have licensed previously described technologies related to the work  
508 described in this paper from Johns Hopkins University. BV, KWK, and NP are inventors on some of  
509 these technologies. Licenses to these technologies are or will be associated with equity or royalty  
510 payments to the inventors as well as to Johns Hopkins University. CT and the University are entitled to  
511 royalty distributions related to technology licensed to Thrive Earlier Detection. CT is a consultant to  
512 Thrive. The terms of all these arrangements are being managed by Johns Hopkins University in  
513 accordance with its conflict of interest policies. All other authors declare no competing interests.

514 23. Kobayashi S, Nakamura Y, Taniguchi H, Odegaard JI, Nomura S, Kojima M, et al. Impact of  
515 Preoperative Circulating Tumor DNA Status on Survival Outcomes After Hepatectomy for Resectable  
516 Colorectal Liver Metastases. *Annals of surgical oncology*. 2021;28(8):4744-55. Epub 2021/01/05. doi:  
517 10.1245/s10434-020-09449-8. PubMed PMID: 33393041.

518 24. Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in  
519 complexity. *Nature reviews Clinical oncology*. 2020;17(1):11-32. Epub 2019/07/11. doi:  
520 10.1038/s41571-019-0241-1. PubMed PMID: 31289352; PubMed Central PMCID: PMCPMC7577509.

521 25. Miyazaki K, Wada Y, Okuno K, Murano T, Morine Y, Ikemoto T, et al. An exosome-based liquid  
522 biopsy signature for pre-operative identification of lymph node metastasis in patients with  
523 pathological high-risk T1 colorectal cancer. *Molecular cancer*. 2023;22(1):2. Epub 2023/01/08. doi:  
524 10.1186/s12943-022-01685-8. PubMed PMID: 36609320; PubMed Central PMCID: PMCPMC9817247.

525 26. Wada Y, Shimada M, Murano T, Takamaru H, Morine Y, Ikemoto T, et al. A Liquid Biopsy Assay for  
526 Noninvasive Identification of Lymph Node Metastases in T1 Colorectal Cancer. *Gastroenterology*.  
527 2021;161(1):151-62.e1. Epub 2021/04/06. doi: 10.1053/j.gastro.2021.03.062. PubMed PMID:  
528 33819484; PubMed Central PMCID: PMCPMC10360659.

529 27. Nishiyama A, Nakanishi M. Navigating the DNA methylation landscape of cancer. *Trends in*  
530 *genetics : TIG*. 2021;37(11):1012-27. Epub 2021/06/15. doi: 10.1016/j.tig.2021.05.002. PubMed PMID:  
531 34120771.

532 28. Chang SC, Liew PL, Ansar M, Lin SY, Wang SC, Hung CS, et al. Hypermethylation and decreased

533 expression of TMEM240 are potential early-onset biomarkers for colorectal cancer detection, poor  
534 prognosis, and early recurrence prediction. *Clinical epigenetics*. 2020;12(1):67. Epub 2020/05/14. doi:  
535 10.1186/s13148-020-00855-z. PubMed PMID: 32398064; PubMed Central PMCID: PMCPMC7218647.  
536 29. Yang R, Zhang G, Dong Z, Wang S, Li Y, Lian F, et al. Homeobox A3 and KDM6A cooperate in  
537 transcriptional control of aerobic glycolysis and glioblastoma progression. *Neuro-oncology*.  
538 2023;25(4):635-47. Epub 2022/10/11. doi: 10.1093/neuonc/noac231. PubMed PMID: 36215227;  
539 PubMed Central PMCID: PMCPMC10076951.  
540 30. Hata S, Pastor Peidro A, Panic M, Liu P, Atorino E, Funaya C, et al. The balance between KIF3 and  
541 EG5 tetrameric kinesins controls the onset of mitotic spindle assembly. *Nat Cell Biol*. 2019;21(9):1138-  
542 51. Epub 2019/09/05. doi: 10.1038/s41556-019-0382-6. PubMed PMID: 31481795.  
543 31. Zhang T, Chen S, Peng Y, Wang C, Cheng X, Zhao R, et al. NOVA1-Mediated SORBS2 Isoform  
544 Promotes Colorectal Cancer Migration by Activating the Notch Pathway. *Frontiers in cell and*  
545 *developmental biology*. 2021;9:673873. Epub 2021/10/26. doi: 10.3389/fcell.2021.673873. PubMed  
546 PMID: 34692669; PubMed Central PMCID: PMCPMC8531477.  
547 32. Hu Z, Ding J, Ma Z, Sun R, Seoane JA, Scott Shaffer J, et al. Quantitative evidence for early  
548 metastatic seeding in colorectal cancer. *Nature genetics*. 2019;51(7):1113-22. Epub 2019/06/19. doi:  
549 10.1038/s41588-019-0423-x. PubMed PMID: 31209394; PubMed Central PMCID: PMCPMC6982526.  
550 33. Wang H, Gong P, Chen T, Gao S, Wu Z, Wang X, et al. Colorectal Cancer Stem Cell States  
551 Uncovered by Simultaneous Single-Cell Analysis of Transcriptome and Telomeres. *Advanced science*  
552 (Weinheim, Baden-Wuerttemberg, Germany). 2021;8(8):2004320. Epub 2021/04/27. doi:  
553 10.1002/advs.202004320. PubMed PMID: 33898197; PubMed Central PMCID: PMCPMC8061397.  
554 34. Feng D, Gao P, Henley N, Dubuissez M, Chen N, Laurin LP, et al. SMOC2 promotes an epithelial-  
555 mesenchymal transition and a pro-metastatic phenotype in epithelial cells of renal cell carcinoma  
556 origin. *Cell death & disease*. 2022;13(7):639. Epub 2022/07/23. doi: 10.1038/s41419-022-05059-2.  
557 PubMed PMID: 35869056; PubMed Central PMCID: PMCPMC9307531.  
558 35. Lee Y, Dho SH, Lee J, Hwang JH, Kim M, Choi WY, et al. Hypermethylation of PDX1, EN2, and MSX1  
559 predicts the prognosis of colorectal cancer. *Experimental & molecular medicine*. 2022;54(2):156-68.  
560 Epub 2022/02/17. doi: 10.1038/s12276-022-00731-1. PubMed PMID: 35169223; PubMed Central  
561 PMCID: PMCPMC8894425 interests.  
562 36. Wan J, Su Y, Song Q, Tung B, Oyinlade O, Liu S, et al. Methylated cis-regulatory elements mediate  
563 KLF4-dependent gene transactivation and cell migration. *eLife*. 2017;6. Epub 2017/05/30. doi:  
564 10.7554/eLife.20068. PubMed PMID: 28553926; PubMed Central PMCID: PMCPMC5466421.  
565



566 **Figure Legends**

567 Figure 1. Flow diagram of the study.

568

569 Figure 2. Development and validation of tissue diagnosis model based on the 27 CRC-  
570 specific DMRs. (A) The use of ROC curve analysis to assess the performance of the tissue  
571 diagnostic model in differentiating CRC patients from normal individuals of the training  
572 cohort (51 tissues we collected). (B-C) The use of ROC curve analysis to assess the  
573 performance of the tissue diagnostic model in differentiating CRC patients from normal  
574 individuals(B) and distinguish AA from Normal subjects (C) in tissue validation cohort  
575 1(GSE 48684 dataset). (D-E) The sensitivity and specificity of the diagnosis model in the  
576 tissue training(D) and validation cohort 1(E).

577

578 Figure 3. Development and validation of a plasma diagnosis model based on the 27 CRC-  
579 specific DMRs. (A-B) Heatmap illustrating the DMRs between CRC and advanced adenoma  
580 and healthy controls in the training (A) and validation cohort (B). (C) The use of ROC curve  
581 analysis to assess the performance of the diagnosis model in differentiating CRC patients  
582 from normal individuals of the training cohort. (D) The use of ROC curve analysis to evaluate  
583 the performance of the diagnosis model in differentiating CRC patients from normal  
584 individuals of the validation cohort. (E-G) In the validation cohort, the performance of the  
585 diagnosis model in differentiating early-stage CRC (0-II) from normal individuals (E); in  
586 differentiating advanced-stage CRC (III-IV) from normal individuals (F); in differentiating  
587 AA from normal individuals (G). (H) Methylation scores of Normal 、 AA and CRC

588 individuals generated from the diagnosis model in the training and validation cohort (The  
589 dashed line represents the cutoff value). (I) The sensitivity and specificity of the diagnosis  
590 model stratified by stages in the training and validation cohort. Abbreviations: DMRs,  
591 differential methylated regions; ROC, receiver operating characteristic; AA, advanced  
592 adenoma; CRC, colorectal cancer.

593

594 Figure 4. External validation of plasma diagnosis model. (A-C) In the external validation  
595 cohort, the performance of the diagnosis model in differentiating early-stage CRC (0-II) from  
596 normal individuals (A); in differentiating AA from normal individuals (B); in differentiating  
597 NAA from normal individuals (C). (D) Methylation scores of Normal, NAA, AA and CRC  
598 (0-II) individuals generated from the diagnosis model in the external validation cohort (The  
599 dashed line represents the cutoff value). (E) The sensitivity and specificity of the diagnostic  
600 model for NAA, AA, and CRC in the external validation cohort.

601

602 Figure 5. Development and validation of a plasma-based metastasis model and a plasma-  
603 based prognosis model. (A-B) The methylation scores generated by the diagnosis model for  
604 CRC patients across various M stages (A) and different disease stages (B). (C) ROC curve  
605 analysis was performed to assess the performance of 27 DMRs methylation scores generated  
606 by the diagnosis model in distinguishing M1-CRC patients from M0-CRC patients in the  
607 training (AUC = 0.969) and validation (AUC = 0.955) cohorts. (D) ROC analysis evaluates  
608 the performance of 27 DMRs methylation scores generated by the diagnosis model in  
609 predicting the prognosis of CRC patients. (E) Kaplan-Meier survival curves comparing OS

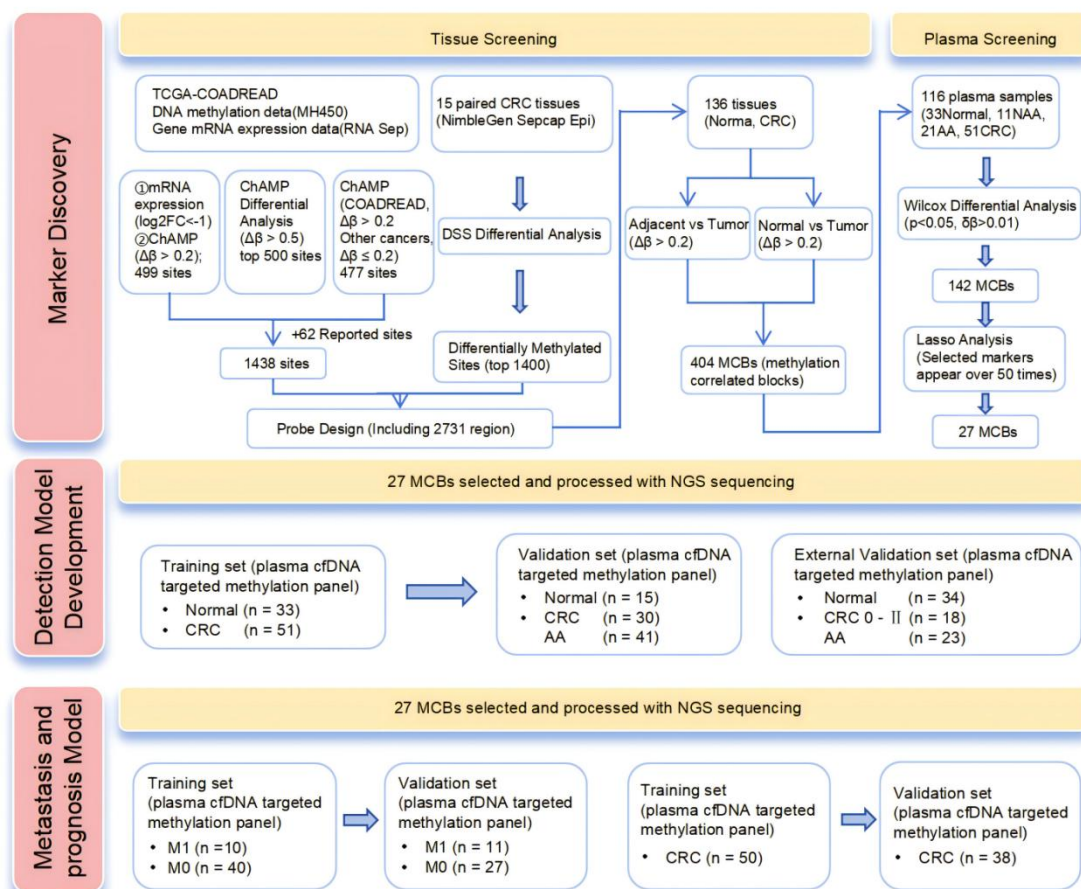
610 between the high-risk CRC group and low-risk CRC group of prognosis model in the training  
611 cohort and validation cohort. (F, H) Metastasis and prognosis prediction of the training and  
612 validation patients (n = 88). The black dotted line at the cutoff value divides the patients into  
613 high-risk and low-risk groups. Yellow circles and gray circles represent patients without  
614 distant metastasis and with distant metastasis, respectively (F). Yellow circles and gray circles  
615 represent patients survived and deceased, respectively (H). (G, I) The proportion of  
616 metastasis and deceased is higher in the high-risk group. The p-value was calculated using a  
617 two-sided Fisher's exact test. Abbreviations: DMRs, differential methylated regions; ROC,  
618 receiver operating characteristic; M1-CRC, colorectal cancer with distant metastasis; M0-  
619 CRC, colorectal cancer without distant metastasis; AUC, the area under the curve.

620

621 Figure 6. Methylation scores of paired tissue and plasma samples. (A) The methylation scores  
622 of tissues from Normal, AA, CRC 0-II, CRC III-IV individuals. (B) The methylation scores of  
623 blood samples from Normal, AA, CRC 0-II, CRC III-IV individuals. (C) The mechanism  
624 diagram interpreting the methylation scores changes of tissue and plasma across different  
625 populations.

626

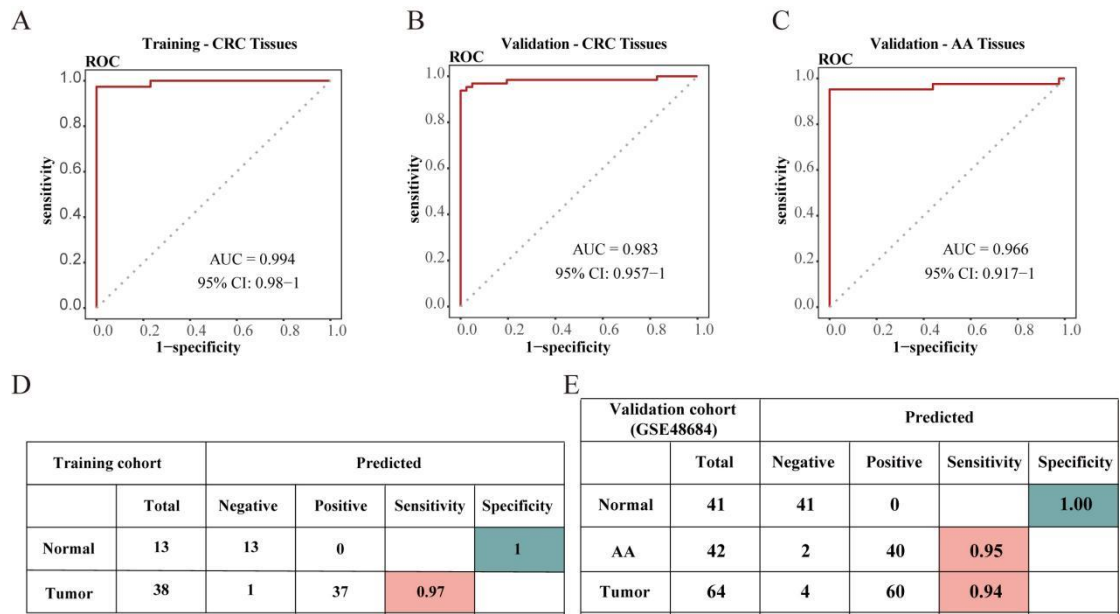
627 Figure 1



628

629

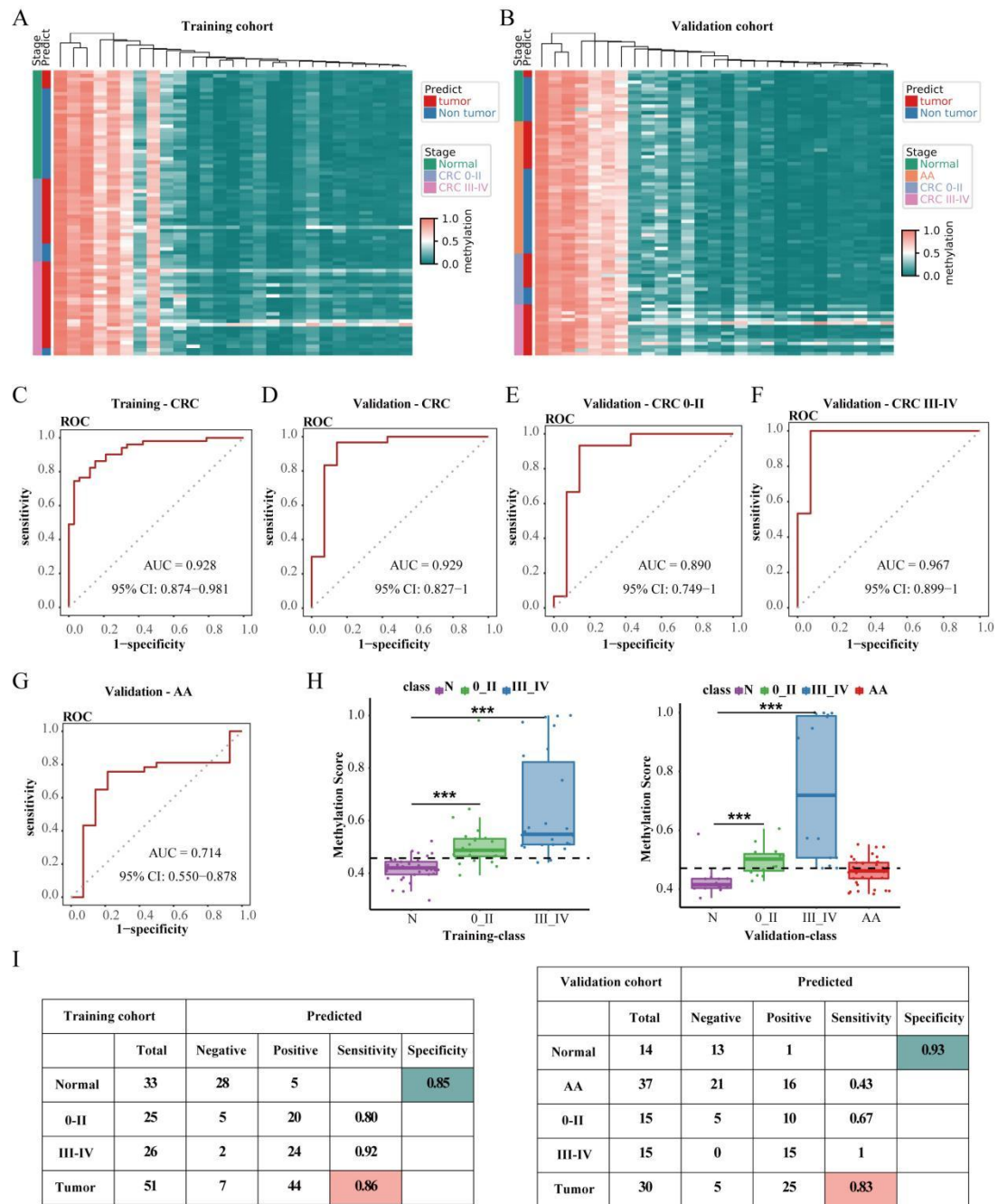
630 Figure 2



631

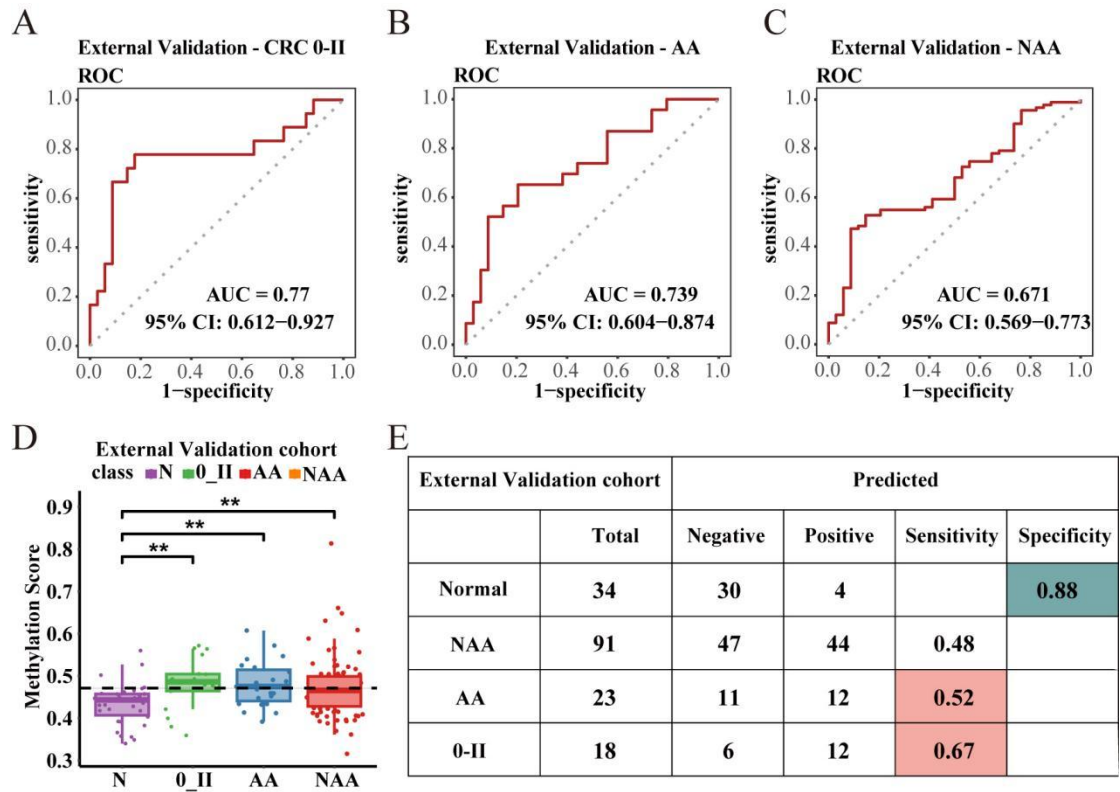
632

633 Figure 3



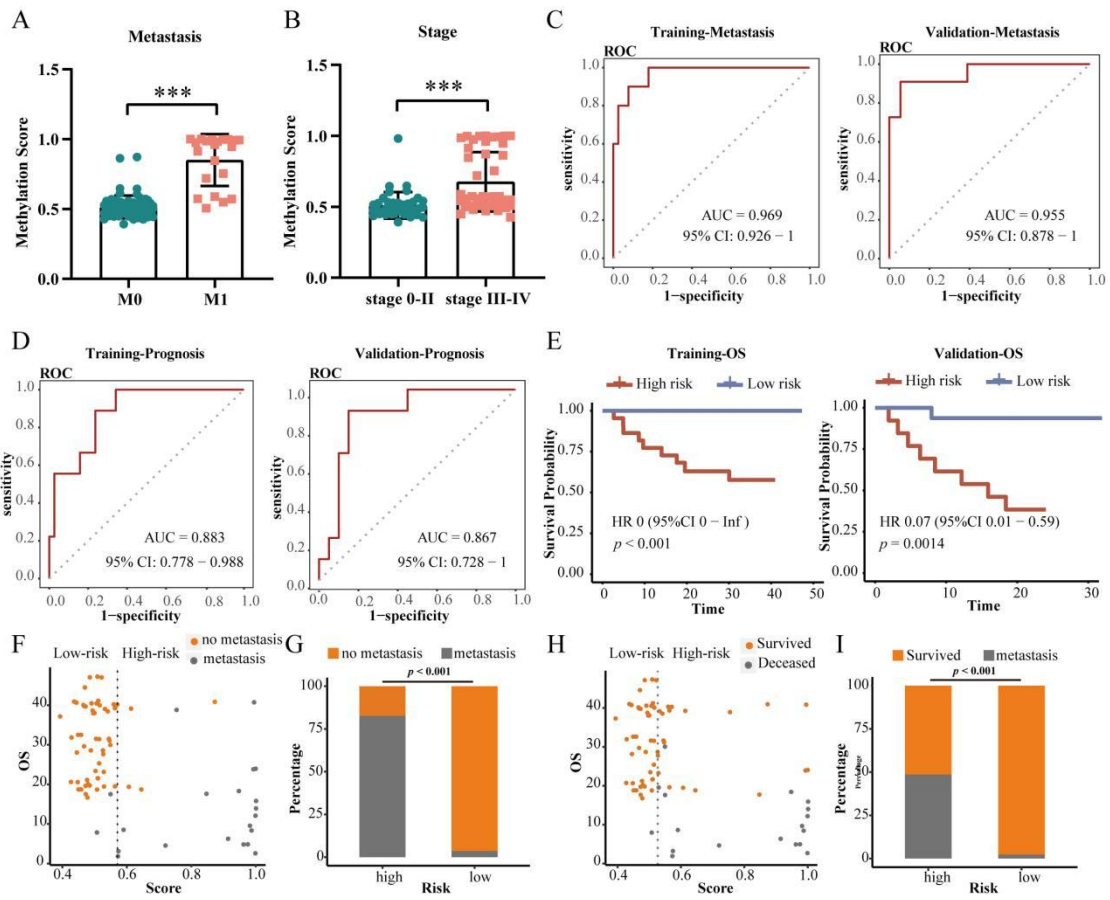
634  
635

636 Figure 4



637  
638

639 Figure 5

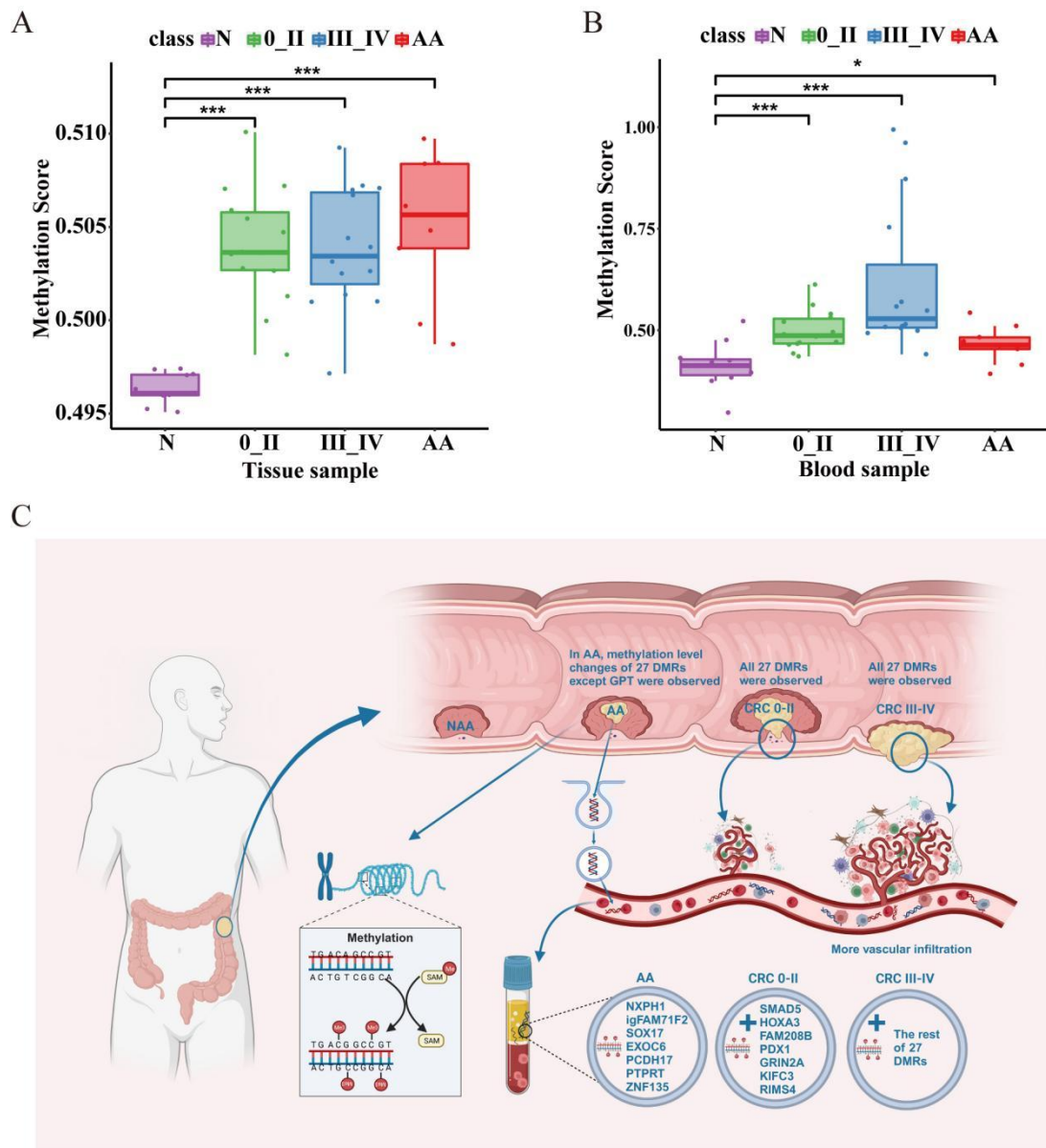


640

641



642 Figure 6



643