



21 <sup>6</sup>Department of Clinical Laboratory Medicine, the Cancer Hospital of Shantou University Medical  
22 College, Shantou, Guangdong 515041, China

23 <sup>7</sup>BioChain (Beijing) Science & Technology Inc., Beijing 102600, China.

24 <sup>8</sup>Department of Radiation Oncology, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi  
25 710032, China

26 <sup>9</sup>Department of Digestive Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi  
27 710032, China

28

29 # These authors contributed equally to this work.

30 \* Correspondence authors:

31 Yin Li, National Cancer Center, 17 Panjiayuan Nanli, Chaoyang District, Beijing, 100021, China.

32 Phone: 86.010.87788053; Email: [liyin@cicams.ac.cn](mailto:liyin@cicams.ac.cn);

33 Jie He, National Cancer Center, 17 Panjiayuan Nanli, Chaoyang District, Beijing, 100021, China. Phone:

34 86.010.87788713; Email: [hejie@cicams.ac.cn](mailto:hejie@cicams.ac.cn);

35 Xiaoliang Han, BioChain (Beijing) Science & Technology Inc., No.18 Hongda South Road, Daxing

36 District, Beijing, 100176, China. Phone: 86.18612824530; Email: [sean.han@biochainbj.com](mailto:sean.han@biochainbj.com);

37 Gang Ji, Xijing Hospital, Changle West Road, Xi'an, Shaanxi, 710032, China. Phone: 86.029.84771518;

38 Email: [Jigang@fmmu.edu.cn](mailto:Jigang@fmmu.edu.cn)

39

40

41 **Abstract**

42 *Background:* Current methods for detecting esophageal cancer (EC) are generally invasive or exhibit  
43 limited sensitivity and specificity, especially for the identification of early-stage tumors.

44 *Methods:* We identified potential methylated DNA markers (MDM) from multiple genomic regions in a  
45 discovery cohort and a diagnostic model was developed and verified in a model-verification cohort of  
46 297 participants. The accuracy of the MDM panel was validated in a multicenter, prospective cohort (n  
47 = 1429). The clinical performance of identified MDMs were compared with current tumor-associated  
48 protein markers.

49 *Results:* From 31 significant differentially methylated EC-associated regions identified in the marker  
50 discovery, we trained and validated a 3-MDM diagnostic model that could discriminate among EC  
51 patients and Non-EC volunteers in a multicenter clinical prospective cohort with a sensitivity of 85.5%  
52 and a specificity of 95.3%. This panel showed higher sensitivity in diagnosing early-stage tumors, with  
53 sensitivities of 56% for Stage 0 and 77% for Stage I, comparing with the performance of current  
54 biochemical markers. In population with high risk for EC, the sensitivity and specificity are 85.68% and  
55 93.61% respectively.

56 *Conclusion:* The assessment of tumor-associated methylation status in blood samples can facilitate non-  
57 invasive, and reliable diagnosis of early-stage EC, which warrants further development to expand  
58 screening and reduce mortality rates.

59 Trial registration number: ChiCTR2400083525

60

61 **Keywords:** Esophageal cancer; Early Diagnosis; DNA methylation; Plasma-based

62

63

## 64 **Introduction**

65 Esophageal squamous cell carcinoma (ESCC) remains the predominant histological subtype of  
66 esophageal cancer (EC) worldwide, representing nearly 90% of the 604,000 new cases reported in  
67 2020(1). As a result of the absence of common clinical manifestations and physical indicators during the  
68 initial phases of esophageal cancer, a large proportion of patients in China are diagnosed at intermediate  
69 to advanced stages, contributing to the unfavorable overall prognosis for individuals with this  
70 condition(2). However, regions in China where screening is routine demonstrate notably higher survival  
71 rates compared to those without screening (40.6% vs 32.8%)(3). In addition, substantial improvements  
72 in survival are observed when the disease is confined to superficial mucosal layers, with rates exceeding  
73 80% after endoscopic or surgical intervention(4), suggesting the critical value of early detection in  
74 clinical practice . However, endoscopy, the gold-standard technique for EC diagnosis, is not suitable for  
75 population-based screening due to its relatively high cost and invasiveness. In addition, although the  
76 preinvasive stage of esophageal squamous dysplasia is well-described and could serve as a reliable basis  
77 for development of less invasive, blood-based, early detection strategies, currently available biomarkers  
78 have shown generally insufficient accuracy and efficacy(5).

79 To overcome the deficiency, cancer-specific DNA methylation modifications have been proposed as  
80 potentially promising biomarkers for EC detection due to their prominent role in dysregulation of tumor

81 suppressor genes or oncogenes, and consequently, tumorigenesis(6). Aberrant methylated DNA markers  
82 (MDMs) have demonstrated to be an early event in oncogenesis and presented less heterogeneous than  
83 gene mutations, and could thus serve as ideal tools for early detection of different malignancy types,  
84 including EC. For example, MDMs identified from esophageal cytology specimens obtained via sponge  
85 sampling devices showed nearly perfect performance in detecting Barrett's esophagus, achieving 92%  
86 sensitivity and 94% specificity(7). This high accuracy suggests that MDMs could provide high diagnostic  
87 power for early detection of EC. Another study has shown that a panel of five MDMs (FER1L4, ZNF671,  
88 ST8SIA1, TBX15, ARHGEF4) identified from tissue samples could also to detect EC in plasma-based  
89 assays from limited clinical samples (8). It is necessary to prove that EC at early stage can be detected  
90 in plasma, such as at stage 0, and to validate the early diagnosis of EC in a clinical trial with sufficient  
91 clinical samples including EC at stage 0.

92 To improve early detection of EC, in this current study, we developed a method for detecting methylated  
93 DNA markers in multiple genomic regions in blood samples from EC at early stage. We applied this  
94 method in a training cohort, which resulted in a diagnostic model based on three markers for EC at early  
95 stage. Then we validated the model in a multi-center clinical cohort including diagnosed group and  
96 diagnosing group with high risk of EC to simulate applications of methylation testing in real clinical  
97 situation. Furthermore, we systematically validated the diagnostic performance of these EC-specific  
98 methylation markers for detecting early-stage EC by comparison with current protein-based markers.

99

## 100 **Results**

101 The study was conducted in three main phases: 1) Marker Discovery, a phase in which methylation  
102 markers were screened from tissue and plasma samples; 2) Model Verification, a phase in which the  
103 probes were optimized, and model was constructed; and 3) Clinical Validation, a phase wherein  
104 diagnostic performance for early EC was evaluated in a cohort of 1429 participants and control subjects  
105 **(Figure 1)**.

106

#### 107 Marker Discovery

108 In the biomarker discovery phase, we analyzed whole-genome bisulfite sequencing (WGBS) data from  
109 plasma samples of 56 esophageal cancer (EC) patients and 107 healthy controls. This dataset was  
110 augmented with the methylation data of 108 ESCC samples and 356 healthy individuals sourced from  
111 public databases to delineate differential methylation patterns. Our analysis revealed that overall  
112 methylation predominated in relatively rare genomic regions **(Supplementary Figure 1)**. We identified  
113 31 differentially methylated regions associated with esophageal cancer and some of the selected genes  
114 were shown in Integrative Genomics Viewer **(Figure 2A and Supplementary Figure 2)**. Using logistic  
115 regression modeling, we preliminarily identified 6 differentially methylated regions, which were  
116 annotated to the 6 genes, Epo, MT-1A, PDGFRA, HOXB13, TRIM15, and Septin9, as potential  
117 diagnostic markers for esophageal cancer. These six potential diagnostic markers were performed further  
118 validation through quantitative methylation-specific PCR (qMSP) in HeLa cell DNA, WBC (white blood  
119 cell) DNA and plasma samples collected from a subset of participants, including 20 EC patients, 12  
120 healthy individuals, and 10 patients with benign esophageal diseases. The original data of qMSP assays

121 based on these plasma samples were presented in **Supplementary Table 1**. The methylation level of the  
122 differentially methylated region on TRIM15 gene showed relative high background signal in WBC DNA  
123 and the amplification signal of the differentially methylated region on PDGFRA gene in HeLa cell DNA  
124 was below the limit of detection, which led us to focus on the remaining four potential markers (Epo,  
125 MT-1A, HOXB13, and Septin9) for further verification (**Supplementary Figure 3**). Notably, the Ct  
126 values in qMSP assays showed that the differentially methylated regions on Septin9, MT-1A, and Epo  
127 genes were significantly lower in esophageal cancer samples compared with controls (Wilcoxon test,  
128  $P < 0.05$ ), indicating these potential markers had a higher methylation status in cancer samples (**Figure**  
129 **2B**). Through this validation process, three highly unique differentially methylated regions emerged as  
130 candidate methylated DNA markers (MDMs).

#### 131 Model Verification

132 In the Model Verification phase, the three candidate MDMs were assessed by qMSP in plasma samples  
133 of the Model Verification cohort, including 87 patients with esophageal cancer, 5 patients with High-  
134 grade intraepithelial neoplasia, 16 patients with other types of cancers, and 189 healthy individuals.

135 The combination of candidate MDMs was employed by logistic regression and parallel techniques and  
136 evaluated by the diagnostic performance from a Model Verification cohort. The area under curve (AUC)  
137 values for the individual ROC curves of Septin9, Epo, and MT-1A were 0.857, 0.853, and 0.837,  
138 respectively. When combined, they reached AUC values of 0.947 (logistic regression) and 0.948 (parallel  
139 techniques), indicating improved diagnostic accuracy (**Figure 3A and Supplementary Table 2**). In the  
140 integration of three MDMs, the AUC values of logistic regression and parallel techniques were similar,

141 with slightly higher AUC for the parallel technique. Therefore, we opted for the parallel technique in our  
142 combined approach. Comparison of predicted (by qMSP and parallel techniques) versus observed  
143 classifications by confusion matrix showed that this panel of candidate MDMs provided 95.29%  
144 accuracy in discriminating between EC and healthy controls, suggesting relatively high consistency  
145 between the model and actual clinical diagnoses (**Figure 3B**, kappa=0.89).

146

## 147 Clinical Validation

### 148 1. Demographics in the Clinical Cohort

149 The Clinical Validation cohort consisted of 641 participants with EC, while the control group  
150 comprised 788 participants without EC, including healthy controls and participants with benign  
151 esophageal diseases or other types of cancer. Participants with EC were categorized according to the  
152 AJCC staging system, with 32 participants at stage 0, 106 participants at stage I, 111 participants at stage  
153 II, 204 participants at stage III, 117 participants at stage IV, and 71 participants with unknown staging  
154 information (**Table 1**).

155 For assessing the performance of diagnostic marker in confirmed participants with EC and illustrating  
156 its applicability for diagnosing EC in high-risk individuals, participants were assigned to the diagnosed  
157 group (including participants diagnosed with esophageal cancer and healthy controls before methylation  
158 testing, N=534) and the diagnosing group (including participants in diagnosing esophageal cancer,  
159 benign esophageal diseases, and healthy controls after methylation testing, N=697) as detailed in Table  
160 1. There are 12 participants at stage 0, 48 participants at stage I, 40 participants at stage II, 88 participants



161 at stage III, 32 participants at stage IV, and 37 participants with unknown staging information in the  
162 diagnosed group and 20 participants at stage 0, 58 participants at stage I, 71 participants at stage II, 116  
163 participants at stage III, 85 participants at stage IV, and 34 participants with unknown staging information  
164 in the diagnosing group. Besides, the Other Cancers Group comprises 198 cases (Age 60.39±10.2, 130  
165 Male and 68 Female) and the distributions of cancer types are as follows: lung cancer accounted for 14.14%  
166 with 28 cases, liver cancer represented 9.6% with 19 cases, colorectal cancer accounted for 25.76% with  
167 51 cases, breast cancer accounted for 15.15% with 30 cases, and gastric cancer was the most prevalent  
168 at 35.35% with 70 cases.

169

## 170 2. ROC analysis

171 In the clinical validation phase, the MDMs were again validated by qMSP in plasma samples from 609  
172 participants with EC, 32 participants with high-grade intraepithelial neoplasia (EC at Stage 0), 298  
173 participants with benign esophageal diseases, 198 participants with other types of cancer, and 292 healthy  
174 participants. The original data of qMSP assays were presented in **Supplementary Table 3**. We assessed  
175 the accuracy of the three MDMs by ROC analysis in the clinical cohort (n=641 EC samples; n=788 non-  
176 EC samples). Among the 1429 clinical samples, ROC curve analysis of qMSP Ct values in individual or  
177 multiplex detection assays of Septin9, Epo, and MT-1A yielded AUC values of 0.793, 0.758, and 0.795,  
178 respectively, and 0.904 for all three markers together (**Figure 4A**). These results suggested that multiplex  
179 detection using these candidates could provide higher diagnostic accuracy compared to detection of any  
180 individual MDM. Further comparison by confusion matrix of classifications predicted by qMSP assays

181 of the three-MDM panel with the observed clinical diagnoses showed an accuracy value of 90.17%  
182 (**Figure 4B**, kappa=0.80).

183

184

### 185 3. Sensitivity

186 To further assess whether the diagnostic efficacy of the three-MDM panel differed among stages of EC,  
187 we compared its sensitivity among participants in the clinical cohort stratified by disease stage. Multiplex  
188 qMSP analysis of Septin9, Epo, and MT-1A showed detection sensitivities of 85.49% (95%CI:  
189 82.55%~88.01%) for overall (n=641) patients and the detection performance showed positive correlation  
190 with the tumor progression. The sensitivity of stage 0 (n=32), stage I (n=106), stage II (n=111), stage III  
191 (n=204), stage IV (n=117) were 56.25% (95%CI: 39.06%~73.44%), 77.36% (95%CI: 69.39%~85.32%),  
192 86.69% (95%CI: 80.13%~92.85%), 89.70% (95%CI: 85.54% ~93.88%), 94.02% (95%CI:  
193 88.06%~97.56%) respectively (**Supplementary Figure 4**). The sensitivity in different age groups  
194 showed no significant difference in the range of 40 to over 80 (**Supplementary Table 4**). The sensitivity  
195 performance of MDMs in differentiation groups of different tumor were analysis, the medium-high  
196 differentiation and Medium differentiated groups showed highest performance of 92.31% (95%CI:  
197 66.69%~98.63%) (**Supplementary Table 5**).

### 198 4. Specificity

199 In assays testing whether our multiplex qMSP method could distinguish control subjects, including both  
200 healthy individuals and benign esophageal disease patients, the three-MDM panel achieved a specificity

201 95.25% (93.21%~96.82%). More specifically, healthy individuals could be identified with 97.26%  
202 specificity (95% CI, 94.67%-98.81%), while benign esophageal diseases were diagnosed with 93.29%  
203 specificity (95% CI, 89.82%-95.85%). We assembled a cohort of heterogeneous cancer types, including  
204 liver, colorectal, breast and lung cancers to examine the specificity of these three candidate markers for  
205 discriminating EC from other cancer types. Specificity decreased to 56.86% (95% CI, 43.27%-70.46%)  
206 among colorectal cancer patients. By contrast, the specificity in detecting lung cancer reached 100.00%  
207 (95% CI, 87.94%-100.00%), 100.00% for breast cancer (95% CI, 88.43%-100.00%), 78.95% for liver  
208 cancer (95% CI, 60.62%-97.28%), and a specificity of 70.00% for discriminating gastric cancer (95%  
209 CI, 59.26%-80.74%) (**Figure 4C and Supplementary Figure 5**).

## 210 5. PPV and NPV

211 The positive predictive value (PPV) in overall clinical cohort is 95.14% (548/576), while the negative  
212 predictive value (NPV) was 85.80% (562/655), while the Septin9 alone showed the highest PPV of  
213 96.72% and the best single marker NPV was detected by Epo with 65.15% (**Table 2**).

214

## 215 6. Comparison with conventional tumor markers

216 Further comparison of the Septin9, Epo, and MT-1A MDM panel with conventional tumor markers in  
217 diagnosing different tumor stages in the clinical cohort showed positive detection rates of 56.25%,  
218 77.36%, 86.49%, 89.71%, and 94.02% for cancer stages 0 to IV, respectively, notably higher than those  
219 of the conventional tumor markers, CEA, SCC, CA199 and NSE (**Figure 5A and Supplementary**  
220 **Figure 6**). In addition, we also calculated the Youden Index of each tumor markers, and the results

221 showed that the three-MDM panel detection method was optimal. The respective Youden Indexes of  
222 CEA, SCC, CA199, NSE and three-MDM panel were 0.11, 0.22, 0.03, -0.02, 0.76. Furthermore, the  
223 sensitivity of three-MDM panel for squamous cell carcinoma, adenocarcinoma and other rare cancers  
224 showed no significant difference, and can be applied to all types of esophageal cancers.

#### 225 7. Performance for high-risk populations in EC

226 To further evaluate whether the esophageal cancer methylation detection method in this study has the  
227 potential to be used for screening or as an adjunct diagnostic tool in populations with high-risk for  
228 esophageal cancer, we analysed performance of the three-marker panel in the diagnosed group  
229 (esophageal cancer diagnosed before methylation testing and healthy controls, N=534) and the  
230 diagnosing group (esophageal cancer diagnosed after methylation testing, benign esophageal diseases  
231 and healthy controls, N=697) as detailed in Table 1, the sensitivity, specificity, PPV and NPV in the  
232 diagnosed group were 85.21% (80.27%~89.31%) , 97.11%(94.39%~98.74%), 96.48% (93.17% ~  
233 98.47%) and 87.62% (83.41% ~ 91.09%) respectively, while the 85.68% (81.77%~89.02%), 93.61%  
234 (90.30%~96.05%), 94.27% (91.29% ~ 96.46%) and 84.20% (79.93% ~ 87.87%) were in the diagnosing  
235 group (**Table 3**). The diagnosing group includes populations with high-risk for esophageal cancer, while  
236 the diagnosed group includes participants with confirmed esophageal cancer as control for the diagnosing  
237 group. There is no statistically difference between the diagnosed group and the diagnosing group in terms  
238 of sensitivity, specificity, PPV and NPV, which indicates that the three-marker panel can be used for  
239 screening or diagnosing populations with high-risk for esophageal cancer.

#### 240 8. Treatment Monitoring

241 Additionally, examination of postoperative methylation levels in a subset of participants who underwent  
242 complete surgical resection revealed that 29 of 32 (90.6%) participants tested negative for methylation  
243 in the Septin9, Epo, and MT-1A promoter regions on the third day post-surgery. The methylation risk  
244 scores ( $45-\Delta\text{CT}$ ) of most participants included in the treatment monitoring decreased post-surgery  
245 compared to pre-surgery, showing statistical significance (Wilcoxon test,  $P < 0.001$ ) (**Figure 5B**).

246

## 247 **Discussion**

248 This study introduces a non-invasive approach for the detection of esophageal cancer (EC) using gene  
249 methylation profiles in plasma samples, offering marked advantages over the conventional invasive  
250 endoscopic and pathological examinations which are often painful and less accessible, thus impeding  
251 early diagnosis and treatment. By screening esophageal cancer methylation chip data in public databases,  
252 along with internal plasma WGBS data including EC at early stage, 31 differentially methylated regions  
253 were identified. Subsequent logistic regression analysis of 31 differentially methylated regions in  
254 esophageal and non-esophageal cancer samples pinpointed six differentially methylated regions with a  
255 strong association with EC. Subsequently, they were verified using qMSP technology in cancer cell lines  
256 and clinical plasma samples, leading to the selection of three genes—MT-1A, Epo, and Septin9—for the  
257 development of a methylation-based detection method for EC.

258 In our study, preclinical plasma sample verification was conducted, followed by a case-control and  
259 multicenter clinical study with sufficient participants to validate the effectiveness of this method. The  
260 results indicated high consistency with clinical gold standard, with superior sensitivity and specificity

261 compared to existing studies and commonly used tumor markers, particularly for early-stage esophageal  
262 cancer at stage 0 and I. And the approach by combining the detection of MT-1A, Epo, and Septin9 gene  
263 methylation for esophageal cancer diagnosis, which introduced by this study, is a strategy not previously  
264 documented. The combining method was more effective in the diagnosing esophageal cancer compared  
265 to single gene methylation detection.

266 From the perspective of clinical study design, this research utilized a multi-center trial approach, which  
267 enables the inclusion of a larger number of participants within the same timeframe compared to a single-  
268 center trial, thereby reducing the duration of the clinical trial. Multi-center trials involve collaboration  
269 among various regions, different trial institutions and numerous clinical researchers, leading to  
270 conclusions that are often broadly representative.

271 In this clinical validation stage, we employed a strategically designed two-group cohort to thoroughly  
272 evaluate our biomarkers' diagnostic performance. The diagnosed group included patients who had  
273 already been diagnosed with esophageal cancer at the time of methylation testing. By comparing the  
274 methylation detection results with confirmed diagnoses in this group, we could robustly assess the  
275 accuracy and reliability of the DNA methylation markers. The diagnosing group comprised high-risk  
276 individuals who had not yet been definitively diagnosed with esophageal cancer at the time of  
277 methylation testing. After performing the methylation testing on this group, the diagnosis was  
278 definitively made to simulate applications of methylation testing in real clinical situation. We later  
279 correlated the results of methylation testing with the definitive diagnostic outcomes and conducted  
280 integrated analysis. This setup allowed us to examine the practical application of the DNA methylation

281 testing in the assessment of EC high-risk individuals. Remarkably, the performance characteristics  
282 observed in the diagnosing group were consistent with those in the diagnosed group, demonstrating a  
283 promising tool for early detection in high-risk populations. Therefore, this two-group design provides  
284 compelling evidence that our biomarkers can be effectively utilized for both clinical diagnosis and early  
285 screening of esophageal cancer in high-risk individuals, highlighting the broad diagnostic potential.

286 Regarding the gene characters and functions, the metallothionein (MT) family, is a low molecular-weight  
287 protein family known for its strong affinity towards metal ions(9). This protein family consists for  
288 isomers and plays a crucial role in regulating the homeostasis and oxidation of transition metal ions with  
289 cells. Among its various functions are the maintenance of cellular balance, as well as involvement in  
290 processes such as cell proliferation, differentiation, and apoptosis. MT-1A is one of the four isoforms in  
291 MT family and aberrant MT expression has been observed in several human tumors, including  
292 esophageal cancer, gallbladder cancer, B-cell lymphoma, breast cancer, liver cancer, skin cancer,  
293 papillary thyroid cancer and prostate cancer(10). Studies have demonstrated that overexpression of MT  
294 can shield cancer cells from free DNA damage and lipid peroxidation induced by free radicals(11).

295 Recent investigations have highlighted that elevated expression of MT in squamous cell carcinoma,  
296 suggesting its potential utility as a diagnostic marker for esophageal squamous cell  
297 carcinoma(12).Erythropoietin (Epo) is a glycoprotein hormone(13). Chan K et al. discovered that Epo  
298 can rapidly induce the expression of the proto-oncogene c-myc, exert anti-apoptotic effects, and promote  
299 cell survival(14). Septin is a conserved family of skeleton protein genes with GTPase activity found in  
300 all eukaryotes except plants, playing a role in cell division. In humans, the family comprises 14 members

301 designated as SEPT 1 to 14. Research has indicated a direct association between Septin9 and tumor  
302 development, with varying expression and function across different tumor types(15). Particularly,  
303 Septin9 is highly expressed in gastrointestinal tumors(16, 17), serving as a reliable marker for their  
304 detection.

305 Numerous studies have explored gene methylation markers for esophageal cancer. For instance, Qin Y  
306 et al. used quantitative allele-specific real-time target and signal amplification technology to develop a  
307 diagnostic model based on five methylation genes, achieving a specificity of 91%, detecting 74% of 84  
308 esophageal cancers, with a sensitivity of 43% for 14 stage I cancer and no cancer at stage 0. Although  
309 they used a similar framework, our study employed Whole Genome Bisulfite Sequencing and included  
310 both tissue and plasma samples from a larger cohort in marker discovery stage. These methodological  
311 differences in data coverage, sample types, and sample sizes led to distinct gene signatures, emphasizing  
312 the robustness and specificity of our approach in identifying reliable biomarkers (8). Li D et al.  
313 established a diagnostic methylation classifier based on 12 CpG sites, effectively distinguish BE, EAC  
314 and ESCC from normal tissues (AUC = 0.992) (18). However, this study was solely based on  
315 bioinformatics analysis without validation using clinical plasma samples. Salta S et al. utilized  
316 quantitative methylation-specific PCR to assess the efficacy of detecting esophageal cancer tissue using  
317 two methylated gene combinations(19). Their study achieved the identification accuracy of 82.29% for  
318 adenocarcinoma and 81.73% for squamous cell carcinoma tissue, which was lower than that showed in  
319 our study. Qiao G et al. employed targeted methylation sequencing technology and a support vector  
320 machine algorithm to develop an early detection classifier for esophageal cancer based on 921



321 differentially methylated regions by sophisticated deep targeted sequencing, with a sensitivity of 74.7%  
322 and a specificity of 95.9% in 181 clinical samples(20). The sensitivity for detecting stage 0 to II  
323 esophageal cancer was lower than that observed in the current research.

324 Conventional tumor markers commonly used for adjunctive diagnostic, prognosis, and therapeutic  
325 monitoring purposes in esophageal cancer include cytokeratin-21-1-fragment (CYFRA21-1),  
326 carcinoembryonic antigen (CEA), squamous epithelial cell carcinoma Antigen (Squamous cell  
327 carcinoma antigen, SCC) and tissue polypeptide specific antigen (TPS), etc. While combined application  
328 of these tumor markers may enhance efficiency in the intermediate and advanced stages of esophageal  
329 cancer, the individual sensitivity of them for esophageal cancer at early stage is generally below 20%.

330 Our study not only confirmed the high diagnostic accuracy of the methylation-based approach but also  
331 demonstrates its superiority over conventional tumor markers like CYFRA21-1, CEA, SCC, and TPS.

332 Furthermore, there was a notable enhancement in specificity among individuals exhibiting symptoms of  
333 esophageal cancer but not gastrointestinal (GI) cancer.

334 In the study, there were 534 individuals in the diagnosed group and 697 individuals in the diagnosing  
335 group. The sensitivity results observed were consistent, suggesting that the screening method is suitable  
336 for identifying esophageal cancer in suspected and high-risk populations. The accuracy of esophageal  
337 cancer methylation detection method in Model Verification cohort (95.29%) was higher than that in  
338 Clinical Validation cohort (90.17%). It possibly due to differences in sample size, sample heterogeneity,  
339 and experimental errors. The findings from methylation testing for pre- and post-surgery patients  
340 indicated that 90.6% of patients exhibited a negative methylation status following the surgical procedure,

341 leading to a notable reduction in overall methylation levels. However, when we compared the  
342 methylation risk scores of EC patients before and after surgery, we observed that three patients did not  
343 experience a decline after treatment. For one patient, the pre-operative methylation levels of the  
344 biomarkers were already very low, resulting in a false-negative diagnosis before surgery. This might be  
345 attributed to individual differences, as some esophageal cancer (EC) patients do not exhibit abnormal  
346 methylation in peripheral blood cfDNA. Consequently, this patient's post-operative methylation risk  
347 score did not decrease. For the other two patients, the exact mechanisms underlying the lack of a decline  
348 in methylation risk scores remain unclear. In future studies, we intend to conduct a dedicated  
349 investigation to evaluate the performance of this esophageal cancer biomarker in treatment monitoring  
350 and better understand the clinical implications of changes in methylation risk scores during the treatment  
351 process.

## 352 Conclusions

353 In conclusion, there exists a robust association between the integration of MT-1A, Epo, and Septin9  
354 methylation testing and the development of esophageal cancer. And we offer a promising, highly accurate  
355 method not only for the early detection of EC and individuals with high-risk for EC, but also for the  
356 therapeutic monitoring.

357

## 358 **Methods**

359 Sex as a biological variable

360 Our study examined male and female human beings, but the sex was not considered as a biological

361 variable.

362 Study design and patient cohorts

363 Two cohorts of participants are enrolled in this study (the model-verification cohort containing 297  
364 participants and the clinical validation cohort containing 1429 participants). The participants are  
365 prospectively recruited, including participants with esophageal cancer (EC), participants with benign  
366 lesions, healthy controls, and participants with other cancers, from multiple centers in China.

367 The inclusion criteria are as follows:

368 (1) Esophageal cancer. participants at high risk of esophageal cancer (EC) who were advised by clinicians  
369 to undergo endoscopy. The included participants were those over 40 years old who satisfied any of the  
370 following criteria: participants with long-term residence in areas with high EC incidence or with family  
371 disease history. participants with symptoms of upper gastrointestinal discomfort. participants with  
372 presence of precancerous lesions of EC. And participants with a strong clinical suspicion of EC or high-  
373 grade intraepithelial neoplasia (HGIN) based on endoscopic, imaging, or pathological biopsy findings.  
374 And participants with benign digestive system diseases intending to perform endoscopy test or with prior  
375 endoscopic findings. All the included participants in this classification are with no history of esophageal  
376 cancer surgery and no prior treatment.

377 To validate the performance of the diagnostic marker in confirming esophageal cancer (EC) patients in  
378 this study, while also demonstrating its potential for diagnosing EC in high-risk populations. The  
379 included patients of the clinical validation cohort were divided into a diagnosed group and a diagnosing  
380 group. In the diagnosed group (esophageal cancer definitively diagnosed before methylation testing), the

381 participants were diagnosed definitively first, then provided blood for subsequent methylation testing. In  
382 the diagnosing group (esophageal cancer definitively diagnosed after methylation testing), the  
383 participants provide blood for methylation testing first, then start definitive diagnosing procedures to  
384 simulate applications of methylation testing in real clinical situation. EC was diagnosed based on  
385 characteristics observed during upper gastrointestinal endoscopy, computed tomography (CT), or  
386 magnetic resonance imaging (MRI) and confirmed through histopathology. Tumor staging was  
387 determined according to the American Joint Committee on Cancer/Union for International Cancer  
388 Control (AJCC/UICC) 8th edition system. During all the double-blinded experimental processes, the  
389 participants information was kept confidential from experimental operators and researchers to ensure the  
390 credibility and reliability of clinical trial outcomes.

391 (2) Benign esophageal conditions. Participants are clinically diagnosed with other esophageal diseases  
392 (such as reflux esophagitis, achalasia, esophageal hiatal hernia, diffuse esophageal spasm, and irregular  
393 esophageal spasm) based on laboratory tests (tumor markers or bronchoscopy or imaging, etc.), with no  
394 evidence of esophageal cancer, requiring further evaluation and management.

395 (3) Other types of cancer. Participants are clinically diagnosed with other cancers, such as gastric cancer,  
396 colorectal cancer, etc., who have not undergone treatment or surgery.

397 (4) Healthy controls. Participants with no history of malignant tumors, clinically confirmed to be free of  
398 esophageal cancer (EC), other digestive diseases and those with substantial medical conditions such as  
399 hepatitis, cirrhosis, and chronic obstructive pulmonary disease. All healthy volunteers will undergo a  
400 series of routine health assessments, including complete blood counts, urinalysis, blood biochemistry

401 tests, electrocardiograms, low-dose chest computed tomography (CT), and abdominal ultrasound  
402 examinations.

#### 403 Sample collection and preparation

404 The blood samples from esophageal cancer patients and other groups were collected from multiple  
405 hospitals in China. Blood samples collected from all participating hospitals were processed following the  
406 same protocol by trained technicians. A 5ml K<sub>2</sub>EDTA anticoagulant tube (BD Vacutainer®) was used to  
407 collect a 5ml peripheral blood sample to ensure the accuracy of the tests. Samples were processed and  
408 transported following the guidelines for nucleic acid extraction reagent (BioChain (Beijing) Science &  
409 Technology, Inc.) Plasma was separated from whole blood by centrifugation within 4 hours of blood  
410 sample collection and stored immediately at -80°C. Plasma was tested within 2 weeks of collection.

#### 411 DNA extraction and bisulfite conversion

412 DNA extraction and bisulfite conversion were carried out following the instructions provided in the  
413 manufacturer's manual for the nucleic acid extraction reagent.

#### 414 Quantitative methylation-specific PCR (qMSP)

415 When designing primers and probes for qMSP, primers are designed to include at least one CpG site in  
416 both forward and reverse primers, as well as in the probe binding sequence, ensuring that only methylated  
417 DNA templates are amplified. We extract cfDNA from cell-free plasma to prepare DNA templates. We  
418 simultaneously detect a reference gene, ACTB, when testing target genes. Through extensive clinical  
419 sample validation, we have established that cfDNA content in plasma should not fall below 0.9 ng/mL.  
420 Consequently, we set a reference ACTB Ct value threshold of  $\leq 34.8$ . If the reference gene meets this

421 criterion, the sample is deemed suitable for analysis, allowing us to determine the presence or absence  
422 of methylation in the marker gene. The sulfite-modified DNA served as the template for quantitative  
423 methylation-specific PCR (qMSP), following the detailed procedures outlined in the MT-1A, Epo, and  
424 Septin9 methylated gene detection kit, which employs the PCR fluorescent probe method. The  
425 amplification reactions were conducted in a total volume of 50  $\mu$ L, consisting of 25  $\mu$ L of reaction buffer  
426 and 25  $\mu$ L of sulfite-modified DNA template. The amplification process was conducted using either the  
427 Applied Biosystems 7500 Fast Real-Time PCR System or the SLAN-96S Fully Automatic Medical PCR  
428 Analysis System. Each experimental batch included patient DNA samples, positive controls and negative  
429 controls to maintain stringent quality control throughout the analysis.

#### 430 Marker Discovery

431 In this study, a total of 108 EC cancer tissue samples, 107 adjacent normal tissue samples, and 249 healthy  
432 human WBC (white blood cell) samples from public datasets were analyzed. To integrate data from two  
433 methylation detection chips, a mapping and matching process was conducted for the detection probes  
434 based on specific criteria. These criteria included ensuring that the probe design intervals overlapped in  
435 the genome coordinates or that the maximum distance between probes did not exceed 150 base pairs.  
436 Additionally, probes from the Human Methylation 450 chip were aligned with those from the GoldenGate  
437 chip. Cancer-specific hypermethylation markers were identified based on the following criteria: adjusted  
438 p-value ( $p. adj$ ) $<1e-2$ ,  $\Delta_{T2N}$  (difference in methylation levels between tumor and adjacent normal  
439 tissues)  $> 0.1$  and  $mean\_wbc$  (mean methylation level in WBCs)  $< 0.1$ .

440 The public data sets utilized in this research, including GSE51287, GSE26784, GSE40279, GSE52826

441 and GSE74693, were obtained from the NCBI GEO database (<https://www.ncbi.nlm.nih.gov/gds>). These  
442 datasets exclusively consisted of methylated Chip data based on two kinds of platforms, the Human  
443 Methylation 450 chip and the GoldenGate chip. Notably, GSE40279 specifically included samples from  
444 individuals aged between 30 and 60 years.

445 Furthermore, a subset of the inhouse plasma samples were utilized as a discovery-step validation set to  
446 confirm the markers previously selected based on the public datasets mentioned before. The inhouse  
447 samples were from 107 healthy individuals and 56 patients with esophageal cancer at early stage. These  
448 samples underwent WGBS assay. Given the presence of strong background signals in plasma detection  
449 outcomes, a one-hot approach was utilized to delineate the identification of cancer-specific  
450 hypermethylation signal patterns in plasma. It is a method for categorizing methylation signals based on  
451 a predefined threshold. For each methylation DNA region, if the methylation level in a sample exceeds  
452 the detection threshold, it is classified as "detected," with the signal value set to the methylation level;  
453 otherwise, it is classified as "not detected," with a signal value of zero. The threshold was set at the 95th  
454 percentile of plasma methylation levels in individuals without cancer to achieve 95% specificity in these  
455 control participants, minimizing false positives and allowing for the identification of suitable candidate  
456 markers. During this screening process, a total of 31 marker intervals were examined.

457 We utilized the methylation levels of esophageal cancer differential methylation regions and the sample  
458 type (esophageal cancer vs. non-esophageal cancer) to establish a logistic regression diagnostic model  
459 for esophageal cancer. The variables included in the model will be used as highly relevant candidate  
460 biomarkers for esophageal cancer for subsequent validation through quantitative methylation-specific

461 PCR (qMSP) analysis.

462 For the qMSP analysis, we used DNA from HeLa cell (75ng), which was verified by Sanger sequencing  
463 to be highly methylated for Epo, MT-1A, PDGFRA, HOXB13, TRIM15, and Septin9, as the positive  
464 control, and WBC DNA (35ng) as the negative control. Clinical blood samples from patients with  
465 esophageal cancer (n = 20), benign esophageal diseases (n = 10), and healthy individuals (n = 12) were  
466 also tested. This screening process effectively validated the methylation biomarkers related to esophageal  
467 cancer.

468 Model verification

469 To evaluate the diagnostic efficacy of the methylation-based markers for esophageal cancer (EC), a  
470 model-verification cohort comprising various participant groups was assembled, including patients  
471 diagnosed with EC, individuals with benign esophageal conditions, patients with other types of cancer,  
472 and healthy individuals. The model-verification cohort was consisted of 87 patients with EC, 5 patients  
473 with high-grade intraepithelial neoplasia, 16 patients with other cancers, 189 healthy individuals. The  
474 blood samples were collected from individuals of this cohort to perform qMSP. The experimental details  
475 were described above.

476 Clinical Validation

477 A multi-center, parallel comparison, blinded clinical trial design was utilized, with inclusion criteria  
478 consistent described above. The samples were used to assess the diagnostic efficacy of esophageal cancer  
479 methylation gene detection technology.

480 The clinical trial was approved prior to the commencement of the study. This Clinical Validation cohort



481 consisted of 609 patients with EC, 32 patients with high-grade intraepithelial neoplasia, 298 patients with  
482 benign esophageal disease, 198 patients with other cancers, 292 healthy individuals. The blood samples  
483 were collected from individuals of this cohort to perform qMSP. The experimental details were described  
484 above.

485 The pre-specified primary outcomes of this clinical trial were sensitivity and specificity, which are the  
486 key measures of diagnostic performance for the biomarkers being evaluated. The analyses presented in  
487 the manuscript focus on the pre-specified primary outcomes (sensitivity and specificity).

#### 488 Statistical analysis

489 A descriptive analysis of the demographic characteristics and initial data of the participants was  
490 conducted. Categorical variables were summarized using frequency and percentage composition, while  
491 quantitative variables were summarized using measures such as mean, standard deviation, median. The  
492 diagnostic efficacy was assessed through diagnostic test evaluation, including comparison with the gold  
493 standard, calculation of Kappa values and their corresponding 95% CI (21). Sensitivity was defined as  
494 the proportion of correctly identified positive esophageal cancer cases among all esophageal cancer cases,  
495 while specificity was defined as the proportion of correctly identified negative cases among all normal/  
496 esophageal benign disease and other cancer cases. Positive predictive value (PPV) and negative  
497 predictive value (NPV) were calculated to determine the probability of a positive or negative disease test  
498 result, respectively. Receiver operating characteristics (ROC) curves were generated using R software,  
499 and the area under the ROC curve (AUC) were analyzed. For sample sizes less than 5, the association  
500 between test positivity and demographic characteristics was assessed using either the chi-square test or

501 Fisher's exact test. Statistical significance was defined as  $P < 0.05$ .

502

503 Study approval

504 Prior to sample collection, all participants in this research provided written consent and were duly

505 informed about the utilization of sample and test outcomes. Approval for this study was obtained from

506 the Medical Ethics Committee of all the hospitals participated in the study. The ethical numbers are as

507 follows:

508 Cancer Hospital of the Chinese Academy of Medical Sciences, 21/223-2894

509 The First Affiliated Hospital of the Air Force Medical University, QX20211043-x-1

510 The Second Affiliated Hospital of the Air Force Medical University, 202110-09

511 Henan Provincial Cancer Hospital, 2021-240B-001

512 The Second Affiliated Hospital of Xi'an Jiaotong University, (2021) Ethical (056)

513 The Affiliated Cancer Hospital of Shantou University Medical College, 2021019

514

515 The ethical approvals were obtained from six ethic committees of hospitals in 2021. The recruitment

516 of participants and sample collection were conducted after obtaining ethical approvals. And the trial was

517 registered on the Chinese Clinical Trial Registry (ChiCTR) website in April 2024.

518

519 Data availability

520 The data values associated with the data points shown in graphs and values behind any reported means

521 were presented in the “Supporting data values” file. The qMSP data and the detailed Ct values in the  
522 three stages of present study are displayed in the “Supporting data values” file. The WGBS data of the  
523 inhouse cohort utilized in marker discovery step can be viewed in the Genome Sequence Archive  
524 (GSA) database under accession codes PRJCA035851  
525 (<https://ngdc.cnbc.ac.cn/bioproject/browse/PRJCA035851>).

526

#### 527 Author Contributions

528 Ruixiang Zhang, Yongzhan Nie, Xiaobing Chen, Tao Jiang, Jinhai Wang, Yuhui Peng and Guangpeng  
529 Zhou are co-first authors of the article. These co-first authors conducted the experiments and had full  
530 access to all data in the study and take responsibility for the integrity of the data and the accuracy of the  
531 data analysis. Gang Ji, Xiaoliang Han, Jie He and Yin Li were responsible for the study conception and  
532 design. Yong Li, Lina Zhao, Beibei Chen, Yunfeng Ni, Yan Cheng and Yiwei Xu were responsible for  
533 managing patients and data acquisition. Analysis and interpretation of the data were conducted by  
534 Zhenyu Zhu, Xianchun Gao, Zhen Wu, Tianbao Li, Jie Zhao, Cantong Liu, Gang Zhao, Jiakuan Chen,  
535 Jing Zhao. Drafting the manuscript was performed by Ruixiang Zhang, Yongzhan Nie, Xiaobing Chen,  
536 Tao Jiang, Jinhai Wang, Yuhui Peng, Guangpeng Zhou, Gang Ji, Xiaoliang Han, Jie He and Yin Li.  
537 Ruixiang Zhang, Zhenyu Zhu, Zhen Wu, Jie Zhao, Xiaoliang Han and Yin Li checked and confirmed the  
538 conclusion and the final manuscript.

539

#### 540 Footnotes

541 Author Details

542 <sup>1</sup>Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for  
543 Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College,  
544 Beijing 100021, China. <sup>2</sup>State Key Laboratory of Holistic Integrative Management of Gastrointestinal  
545 Cancers, National Clinical Research Center for Digestive Diseases, Xijing Hospital, Fourth Military  
546 Medical University, Xi'an, Shaanxi 710032, China. <sup>3</sup>Department of Oncology, the Affiliated Cancer  
547 Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, Henan 450008, China.  
548 <sup>4</sup>Department of Thoracic Surgery, Tangdu Hospital, The Air Force Military Medical University, Xi'an,  
549 Shaanxi 710038, China. <sup>5</sup>Department of Gastroenterology, the second affiliated hospital, Xi'an Jiaotong  
550 University, Xi'an, Shaanxi 710004, China. <sup>6</sup>Department of Clinical Laboratory Medicine, the Cancer  
551 Hospital of Shantou University Medical College, Shantou, Guangdong 515041, China. <sup>7</sup>BioChain  
552 (Beijing) Science & Technology Inc., Beijing 102600, China. <sup>8</sup>Department of Radiation Oncology, Xijing  
553 Hospital, Fourth Military Medical University, Xi'an, Shaanxi 710032, China. <sup>9</sup>Department of Digestive  
554 Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi 710032, China

555

556 Clinical Trial Information

557 Clinical trial number, ChiCTR2400083525

558

559 Medical Device Registration Information

560 NMPA Registration number 20243401368

561

562

563 Conflict of Interest

564 The authors declare the following potential conflict of interest with respect to the research, authorship,

565 and/or publication of this article: Dr. Guangpeng Zhou, Zhenyu Zhu, Zhen Wu, Dr. Tianbao Li, Dr. Jie

566 Zhao, Dr. Xiaoliang Han are current employees of BioChain (Beijing) Science & Technology, Inc. The

567 other authors have no potential conflict of interest to report.

568

569

570

571

## 572 **References**

573

574 1. Si M, Lang J. The roles of metallothioneins in carcinogenesis. *J Hematol Oncol*. 2018;11(1):107.

575 2. Zeng H, et al. Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based

576 cancer registries. *Lancet Glob Health*. 2018;6(5):e555-555e567.

577 3. An L, et al. The survival of esophageal cancer by subtype in China with comparison to the United States. *Int*

578 *J Cancer*. 2023;152(2):151-61.

579 4. Saowanee, et al. Survival of Patients With Superficial Esophageal Adenocarcinoma After Endoscopic

580 Treatment vs Surgery. *Clinical Gastroenterology and Hepatology*. 2013;11(11):1424-9.e2.

- 581 5. Chu LY, et al. Blood-based biomarkers for early detection of esophageal squamous cell carcinoma. *World J*  
582 *Gastroenterol.* 2020;26(15):1708-25.
- 583 6. Liu WJ, et al. Epigenetic modifications in esophageal cancer: An evolving biomarker. *Front Genet.*  
584 2022;13:1087479.
- 585 7. Shahsavari D, et al. Expanding beyond endoscopy: A review of non-invasive modalities in Barrett's  
586 esophagus screening and surveillance. *World J Gastroenterol.* 2022;28(32):4516-26.
- 587 8. Qin Y, et al. Discovery, Validation, and Application of Novel Methylated DNA Markers for Detection of  
588 Esophageal Cancer in Plasma. *Clin Cancer Res.* 2019;25(24):7396-404.
- 589 9. Raudenska M, et al. Metallothionein polymorphisms in pathological processes. *Metallomics.* 2014;6(1):55-  
590 68.
- 591 10. Mingfu, et al. Evaluation of MT Family Isoforms as Potential Biomarker for Predicting Progression and  
592 Prognosis in Gastric Cancer. *Biomed Res Int.* 2019;2019:2957821-2957821.
- 593 11. Ghoshal K, et al. Analysis of promoter methylation and its role in silencing metallothionein I gene expression  
594 in tumor cells. *Methods Enzymol.* 2002;353:476-86.
- 595 12. Merlos Rodrigo MA, et al. Metallothionein isoforms as double agents - Their roles in carcinogenesis, cancer  
596 progression and chemoresistance. *Drug Resist Updat.* 2020;52:100691.
- 597 13. Kimáková P, et al. Erythropoietin and Its Angiogenic Activity. *Int J Mol Sci.* 2017;18(7):1519.
- 598 14. Chan KK, et al. Erythropoietin drives breast cancer progression by activation of its receptor EPOR.  
599 *Oncotarget.* 2017;8(24):38251-63.
- 600 15. Lee HS, et al. Circulating methylated septin 9 nucleic Acid in the plasma of patients with gastrointestinal

601 cancer in the stomach and colon. *Transl Oncol.* 2013;6(3):290-6.

602 16. Sun J, et al. Screening and Prognostic Value of Methylated Septin9 and its Association With  
603 Clinicopathological and Molecular Characteristics in Colorectal Cancer. *Front Mol Biosci.* 2021;8:568818.

604 17. Nie Y, et al. Combining methylated SEPTIN9 and RNF180 plasma markers for diagnosis and early detection  
605 of gastric cancer. *Cancer Commun (Lond).* 2023;43(11):1275-9.

606 18. Li D, et al. Specific DNA methylation markers in the diagnosis and prognosis of esophageal cancer. *Aging*  
607 *(Albany NY).* 2019;11(23):11640-58.

608 19. Salta S, et al. A DNA methylation-based test for esophageal cancer detection. *Biomark Res.* 2020;8(1):68.

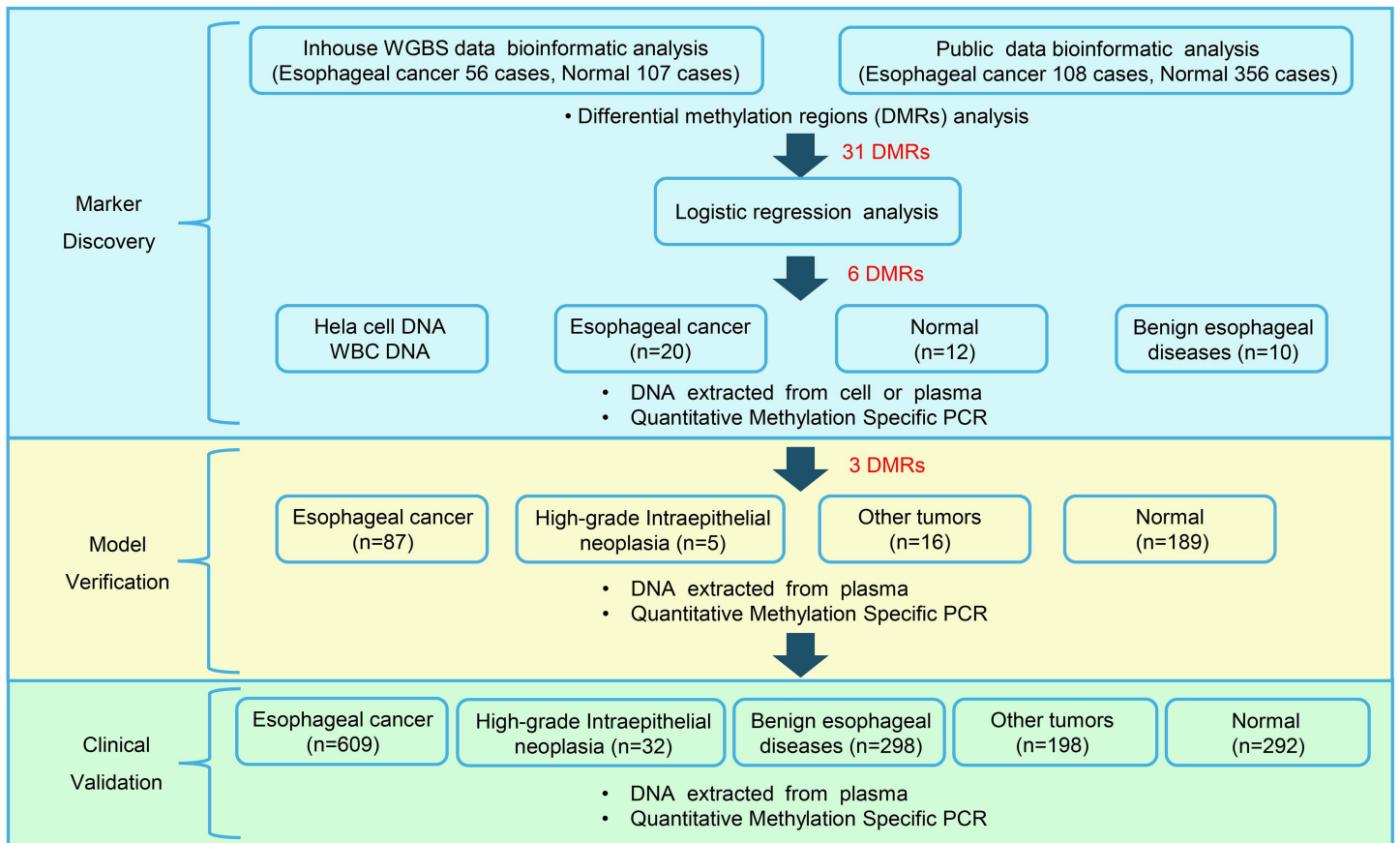
609 20. Qiao G, et al. Discovery and validation of methylation signatures in circulating cell-free DNA for early  
610 detection of esophageal cancer: a case-control study. *BMC Med.* 2021;19(1):243.

611 21. Roy A, et al. Treatment plan quality control using multivariate control charts. *Med Phys.* 2021;48(5):2118-  
612 26.

613

614

615



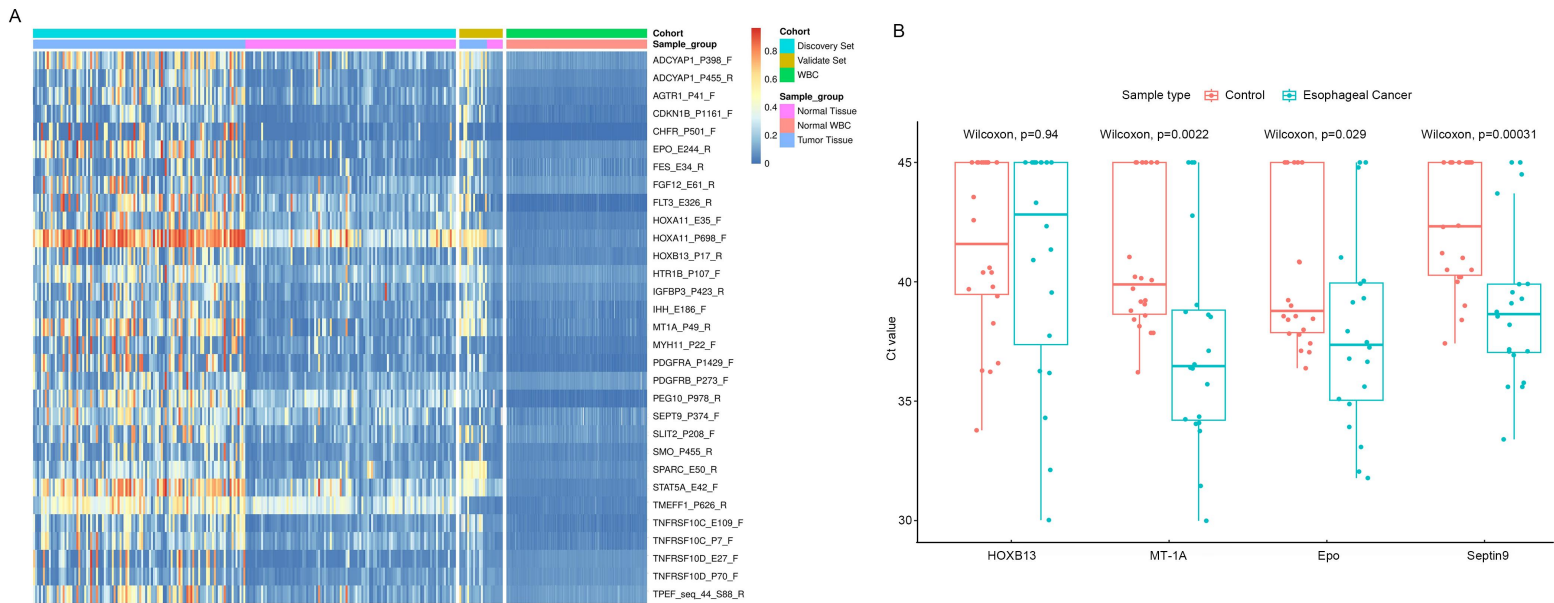
613

614 **Figure 1.** Workflow of the three stages study design, including Marker Discovery, Model Verification,

615 and Clinical Validation.

616





617

618 **Figure 2.** Differential Methylation of Candidate DNA Markers between esophageal cancer patients and

619 normal individuals. (A) Methylation levels of 31 differentially methylated regions between ESCC tumor

620 tissue (n = 108) and non-esophageal cancer cases (normal tissue and WBC) (n = 356) derived from public

621 datasets, illustrating distinct methylation profiles between cancerous and non-cancerous samples. (B)

622 Box plot presenting the qMSP Ct values for the selected 4 potential markers in blood samples of

623 esophageal cancer patients (n =20) and control cases (n = 22) in the Marker Discovery cohort. Control

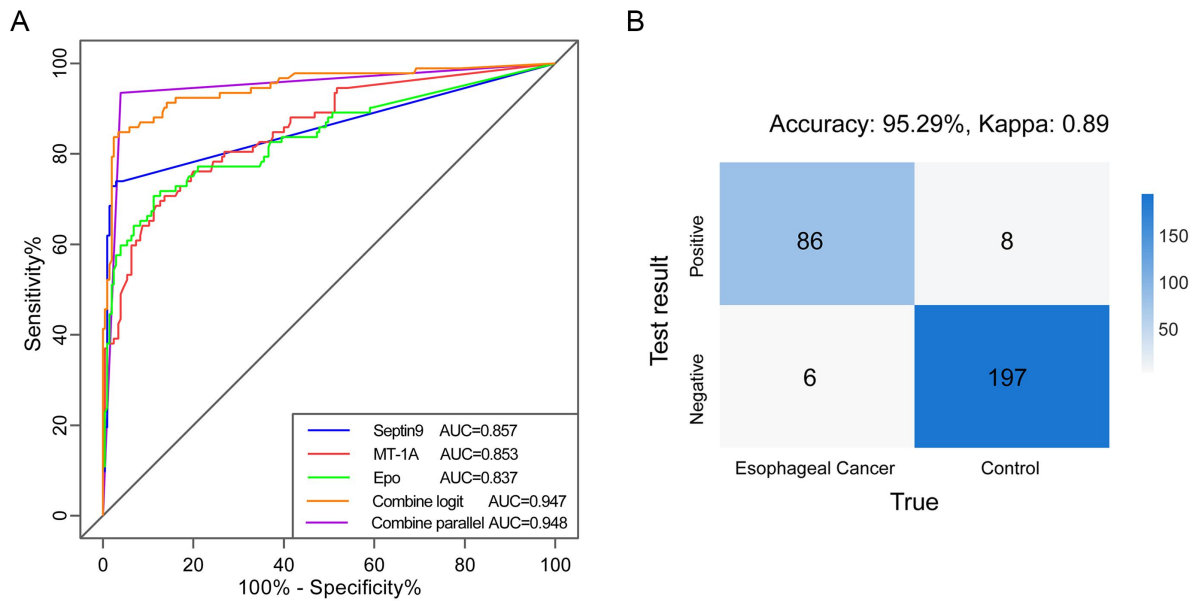
624 cases including benign esophageal diseases and normal individuals. The box-and-whisker plot

625 illustrates the interquartile range (IQR), with the line within the box denoting the median of the data and

626 the whiskers extending from the box to the minimum and maximum values within 1.5 times the IQR.

627 Each point represents one sample. The Wilcoxon test was used for pairwise comparison.

628



629 **Figure 3.** MDM model detection of methylation status by qMSP in the model verification cohort. (A)

630 Diagnostic efficacy of the three candidate MDMs and combined panel in samples from the model

631 verification cohort. The ROC curves indicated the performance for distinguishing esophageal cancer (n

632 = 92, including 87 patients with esophageal cancer and 5 patients with high-grade intraepithelial

633 neoplasia) from non-esophageal cancer (n = 205, including 16 patients with other cancers and 189 healthy

634 individuals). (B) Confusion matrix comparing true-observed classifications (reference detection methods)

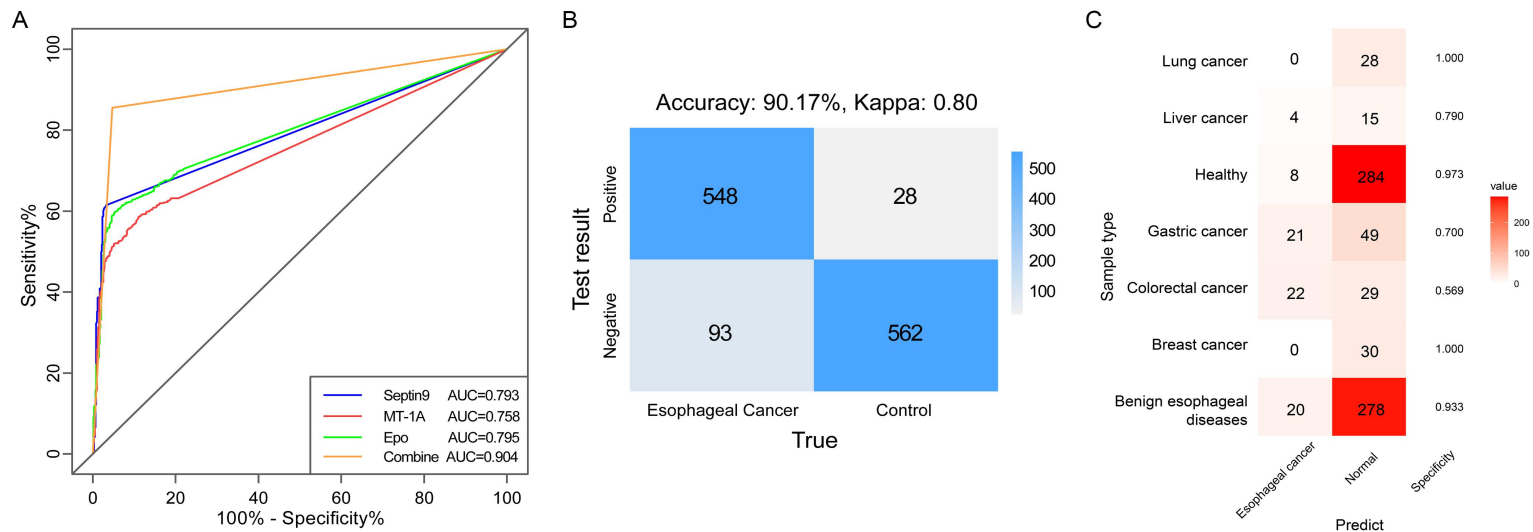
635 with three-MDM panel-predicted diagnoses in the model verification cohort. Esophageal cancer (n = 92)

636 according to true-observed classifications including 87 patients with esophageal cancer and 5 patients

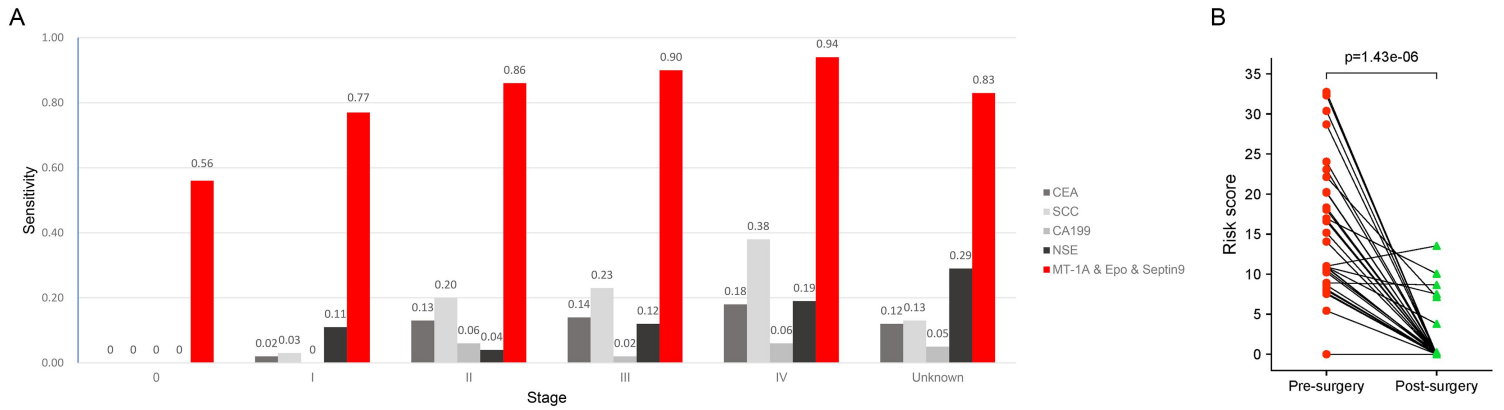
637 with high-grade intraepithelial neoplasia. Control group (n = 205) according to true-observed

638 classifications including 16 patients with other cancers and 189 healthy individuals.

639



640 **Figure 4.** Multiplex detection of Septin9, Epo, and MT-1A methylation status by qMSP in the clinical  
641 validation cohorts. (A) Diagnostic efficacy of the MDMs and combined panel in the clinical validation  
642 cohort. The ROC curves indicated the performance for distinguishing esophageal cancer (n = 641,  
643 including 609 patients with esophageal cancer and 32 patients with high-grade intraepithelial neoplasia)  
644 from non-esophageal cancer (n = 788, including 198 other cancer participants, 292 healthy participants,  
645 and 298 participants with benign esophageal diseases). (B) Confusion matrix comparing true-observed  
646 classifications (reference detection methods) with three-MDM panel-predicted diagnoses in the clinical  
647 validation cohort. Esophageal cancer (n = 641) according to true-observed classifications including 609  
648 patients with esophageal cancer and 32 patients with high-grade intraepithelial neoplasia. Control cases  
649 (n = 590) according to true-observed classifications including 298 patients with benign esophageal  
650 diseases and 292 healthy individuals. (C) Specificities of the three-MDM panel in each cancer type of  
651 198 other cancer participants, in 292 healthy participants, and in 298 participants with benign esophageal  
652 diseases. Different sample types are listed along the vertical axis and predictive results are shown in the  
653 heatmap, while the corresponding specificity for each sample type is on the right vertical axis.  
654



655 **Figure 5.** The three-MDM panel comparing with conventional markers and the application in treatment  
 656 monitoring. (A) Comparison of sensitivity between the three-MDM panel detection method and  
 657 conventional tumor protein markers in different cancer stages of the clinical validation cohort. The total  
 658 number of esophageal cancer samples is 609, with the sample sizes for each stage listed in Table 1. (B)  
 659 Pre-operative and Post-operative methylation levels in a subset of patients (n =32) who underwent  
 660 complete surgical resection. The Wilcoxon test was used for pairwise comparison.

661  
 662

664 **Table 1.** Patient demographics in the clinical cohort

	Overall			Diagnosed Group			Diagnosing Group			
	Cancer	Control	Total (n=1429)	Cancer	Control	Total (n=534)	Cancer	Benign	Control	Total
Case Number	(n=641)	(n=788)		(n=257)	(n=277)		(n=384)	(n=298)	(n=15)	(n=697)
Age (years)	64.5 ± 8.0	58.0 ± 11.0	61.0 ± 10.3	63.9±7.9	60.11±8.5	62.17±8.4	64.99±8.1	56.06±12.4	58.1±8.9	61.03±11.1
Sex (Male, n%)	516(80.50)	477(60.53)	993(69.49)	198(77.04)	133(48.01)	331(61.98)	318(82.81)	203(68.12)	11(73.33)	532(76.33)
AJCC stage, n (%)										
0	32(4.99)			12(4.67)			20(5.21)			
I	106(16.54)			48(18.68)			58(15.10)			
II	111(17.32)			40(15.56)			71(18.49)			
III	204(31.82)			88(34.24)			116(30.21)			
IV	117(18.25)			32(12.45)			85(22.14)			
Unknown	71(11.08)			37(14.40)			34(8.85)			

665

666

667 **Table 2.** The performance in clinical validation cohort

668

669

Group	Sensitivity(%)	Specificity(%)	PPV(%)	NPV(%)	Accuracy(%)
MT-1A	46.02(295/641)	97.12(573/590)	94.55(295/312)	62.35(573/919)	70.51(868/1231)
Epo	52.11(334/641)	97.29(574/590)	95.43(334/350)	65.15(574/881)	73.76(908/1231)
Septin9	40.49(257/641)	98.31(580/590)	96.72(295/305)	62.63(580/926)	71.08(875/1231)
MT-1A+Epo+Septin9	85.49(548/641)	95.25(562/590)	95.14(548/576)	85.80(562/655)	90.17(1110/1231)

674

675

676

677

678

679 **Table 3.** The detection performance of the three-MDM panel in diagnosed group and diagnosing group, respectively.

<b>Indicators</b>	<b>Diagnosed Group</b>	<b>Diagnosing Group</b>
Sensitivity (%) (95% CI)	85.21 (80.27~89.31)	85.68 (81.77~89.02)
Specificity (%) (95% CI)	97.11 (94.39~98.74)	93.61 (90.30~96.05)
Accuracy (%) (95% CI)	91.39 (88.68~93.62)	89.24 (86.70~91.44)
Positive Predictive Value (%) (95% CI)	96.48 (93.17~98.47)	94.27 (91.29~96.46)
Negative Predictive Value (%) (95% CI)	87.62 (83.41~91.09)	84.20 (79.93~87.87)

680

681

682