npj Precision Oncology

https://doi.org/10.1038/s41698-026-01525-8

Article in Press

Cell-free DNA methylation biomarkers for the early detection and tumor burden monitoring of gastric cancer

Received: 5 September 2025

Accepted: 21 May 2026

Cite this article as: Lin, Q., Liu, Y., Lu, H. *et al.* Cell-free DNA methylation biomarkers for the early detection and tumor burden monitoring of gastric cancer. *npj Precis. Onc.* (2026). https://doi.org/10.1038/s41698-026-01525-8

Qi Lin, Yifan Liu, Hanlin Lu, Zhixiong Wang, Wei Tang, Yulong He, Juyong Zhang & Guanghua Li

Cell-free DNA methylation biomarkers for the early detection and tumor burden monitoring of gastric cancer

Qi Lin[1,2]#, Yifan Liu[2]#, Hanlin Lu[3,4]#, Zhixiong Wang[2], Wei Tang[2], Yulong He[1]*, Juyong Zhang[3]*, Guanghua Li[2]*

1 Guangdong Provincial Key Laboratory of Digestive Cancer Research, Digestive Diseases Center, The Seventh Affiliated Hospital of Sun Yat-Sen University, 518107, Shenzhen, Guangdong, China.

2 Department of Gastrointestinal Surgery, First Affiliated Hospital of Sun Yat-sen University, 510080, Guangzhou, Guangdong, China.

3 Shenzhen Acegen Technology Co. Ltd, 518107, Shenzhen, Guangdong, China.

4 Agricultural Genomics Institute, the Chinese Academy of Agricultural Sciences, 518107, Shenzhen, Guangdong, China.

#Contributed equally

*Corresponding author:

Guanghua Li, Email: ligh26@mail.sysu.edu.cn

Juyong Zhang, Email: juyong_zhang@sz-acegen.com

Yulong He, Email: heyulong@sysush.com

Abstract

Development of sensitive biomarkers is required to achieve early detection and tumor burden monitoring in gastric cancer (GC). We performed genome-wide methylation sequencing on 78 tissue and 241 plasma samples from 171 GC patients and 114 healthy controls from two independent clinical centers. Differentially methylated regions (DMRs) were screened using paired GC and normal tissues, and refined through cfDNA profiles with LASSO regression to construct a cfDNA-based biomarker, the GCML-Score. The GCML-Score, consisting of 13 DMRs, demonstrated excellent diagnostic performance (AUC: 0.95/0.99/0.95 overall and 0.96/0.99/0.82 in early GC for training/internal validation/ external validation cohorts). In 12 patients receiving neoadjuvant chemotherapy, dynamic changes in GCML-Score were consistent with radiological tumor burden, highlighting its monitoring potential. The GCML-Score, derived from genome-wide cfDNA methylation profiling, provides a robust tool for early GC detection and real-time tumor burden monitoring, facilitating improved prognosis and personalized therapeutic strategies.

Key Words: cell-free DNA methylation; liquid biopsy; gastric cancer; tumor burden; early diagnosis

Introduction

Gastric cancer (GC) ranks as the fifth most prevalent malignant tumor and is the fourth leading cause of cancer-related mortality globally[1,2]. Because characteristic symptoms are not commonly observed in gastric cancer, most patients are diagnosed at an advanced local stage. The diagnosis of gastric cancer mainly depends on endoscopy. However, since endoscopy is an invasive procedure, it is unlikely to serve as the primary tool for gastric cancer screening, considering factors such as cost and patient compliance[3]. For patients with T2N0 or more advanced disease, neoadjuvant chemotherapy (NAC) facilitates downstaging of the primary tumor, reducing micro-metastatic lesions, and achieving optimal systemic therapy outcomes more effectively[4]. Due to the heterogeneity of tumors, not all gastric cancer patients benefit from NAC[5,6]. Current clinical assessment of NAC response primarily relies on radiographic imaging. However, restricted scheduling frequency on account of cumulative radiation exposure or substantial costs often causes critical delays in therapeutic evaluation, potentially compromising timely treatment adjustment. Therefore, there is an urgent clinical need for an effective, non-invasive tool to aid in the early diagnosis of gastric cancer and the monitoring of tumor burden during NAC.

Both normal and tumor tissues release circulating free DNA (cfDNA) into the circulation. CfDNA sequencing can detect subtle genomic changes in tumor tissues, enabling non-invasive tumor detection. This approach provides valuable clinical information such as tumor burden, response to treatment, and recurrence[7-9]. Furthermore, due to its short lifespan in circulation, cfDNA is an ideal biomarker for monitoring dynamic changes in tumors[10]. DNA methylation is an epigenetic regulatory mechanism that mainly occurs at cytosine residues followed by a guanine nucleotide (CpG sites). It promotes tumorigenesis by activating oncogenes or silencing tumor suppressor genes[11,12]. Compared to genomic mutations, epigenetic alteration occurs earlier in the tumorigenesis process, which is an ideal biomarker for early cancer detection[13]. Beyond genetic alterations, cfDNA methylation provides an early and tissue-informative signal of tumor presence. Integrating methylation profiling with conventional biomarkers and imaging has been shown to significantly improve diagnostic sensitivity and clinical applicability, particularly in low tumor burden settings, supporting its translational potential in gastrointestinal cancer management[14-17]. At the same time, abnormal methylation modification accumulates throughout tumor progression. Therefore, the status of abnormal methylation can, to some extent, reflect the disease progression and tumor burden[18]. Previous studies have shown that gastric cancer is associated with specific cfDNA abnormal methylation patterns[19-21]. Therefore, we hypothesized that quantifying the degree of characteristic abnormal methylation in cfDNA could reflect the disease burden in gastric cancer patients.

In this study, we performed whole-genome methylation sequencing of paired tumor tissues and plasma cfDNA from gastric cancer patients, identified and quantified cancer-specific cfDNA methylation

patterns, and developed a non-invasive clinical diagnostic marker capable of tracking tumor burden during neoadjuvant chemotherapy.

Results

Screening cfDNA methylation biomarkers for GC

Reduced Representation Bisulfite Sequencing (RRBS) was performed on matched tissue (39 GC tissue samples and 39 corresponding adjacent non-cancerous mucosal tissue samples) and 50 PBMC samples from healthy controls. Methylome-seq was performed on plasma-derived cfDNA (57 pre-treatment/during-treatment plasma samples from Chemotherapy cohort, 120 samples from Surgery Cohort and 114 plasma samples from healthy control individuals). The detailed grouping and workflow are shown in Figure 1a. Table 1 provides the clinical features of included patients.

In order to analyze the methylation modification differences between gastric cancer tissue and normal mucosal tissue, we compared the methylation signatures of GC tissues and matched adjacent mucosa. A total of 22051 differentially methylated regions (DMRs) were identified in tissue samples (Figure 1b). Of these, 16,136 (73.1%) DMRs were significantly hypermethylated in gastric cancer tissues, while 5,915 (26.9%) DMRs were significantly hypomethylated. The genome-wide distribution significance of DMRs are shown in Figure 1d. More than half of the DMRs were located in CGI open sea regions (15816, 71.7%) or intergenic regions (13761, 62.4%) (Figure 2a). Studies have shown that DNA methylation regulation is not limited to CpG islands of promoters but also occurs across the entire genome, including genic and intergenic regions. The extensive intergenic regions contain distal regulatory elements, such as enhancers and non-coding RNA genes, and DNA methylation in these distal regulatory regions is also associated with transcriptional regulation[22,23] . We next performed enrichment analysis on the genes corresponding to the DMRs. GO pathway enrichment indicated that these genes were associated with pathways related to digestive organ development, cell differentiation, cell junctions and so on. KEGG pathway analysis revealed significant enrichment in the cAMP signaling pathway, calcium signaling pathway, and others (Figure 1c).

To avoid interference from the genetic material carried by peripheral blood mononuclear cells (PBMCs) in subsequent cfDNA analysis, we screened out the DMRs with a methylation level lower than 5% in the PBMCs samples. As a result, 2079 of 22051 DMRs were selected as candidate markers (Figure 1b).

Cohort	Tissue-sample Cohort	Surgery Cohort			Healthy Controls			Chemotherapy Cohort
		Training	Internal Validating	External Validating	Training	Internal Validating	External Validating	
Total	39 (100)	47 (100)	33 (100)	40 (100)	63 (100)	41 (100)	10 (100)	12 (100)
Gender								
Male (%)	25 (64.10)	27 (57.45)	27 (81.82)	18 (45.00)	15 (23.81)	11 (26.83)	2 (20.00)	9 (75.00)
Female (%)	14 (35.89)	20 (42.55)	6 (18.18)	22 (55.00)	48 (76.19)	30 (73.17)	8 (80.00)	3 (25.00)
Age,year								
< 65 (%)	31 (79.48)	37 (78.72)	19 (57.58)	34 (85.00)	61 (96.83)	41 (100)	9 (90.00)	10 (83.33)
≥65 (%)	8 (20.51)	10 (21.28)	14 (42.42)	6 (15.00)	2 (3.17)	0 (0)	1 (10.00)	2 (16.67)
T stage (Pathological)								**T stage (Clinical)**
T1 (%)	6 (15.38)	11 (23.40)	5 (15.15)	6 (15.00)	--	--	--	0 (0)
T2 (%)	5 (12.82)	3 (6.38)	6 (18.18)	8 (20.00)	--	--	--	0 (0)
T3 (%)	4 (10.25)	9 (19.15)	1 (3.03)	18 (45.00)	--	--	--	2 (16.67)
T4 (%)	24 (61.53)	24 (51.06)	21 (63.64)	8 (20.00)	--	--	--	10 (83.33)
N stage (Pathological)								**N stage (Clinical)**
N0 (%)	9 (23.07)	11 (23.40)	8 (24.24)	25 (62.50)	--	--	--	0 (0)
N1 (%)	9 (23.07)	11 (23.40)	3 (9.09)	6 (15.00)	--	--	--	4 (33.33)
N2 (%)	6 (15.38)	8 (17.02)	15 (45.45)	8 (20.00)	--	--	--	3 (25.00)
N3 (%)	15 (38.46)	17 (36.17)	7 (21.21)	1 (2.50)	--	--	--	5 (41.67)
M stage (Pathological)								**M stage (Clinical)**
M0 (%)	35 (89.74)	36 (76.60)	23 (69.70)	37 (92.50)	--	--	--	7 (58.33)
M1 (%)	4 (10.25)	11 (23.40)	10 (30.30)	3 (7.50)	--	--	--	5 (41.67)
AJCC stage (Pathological)								**AJCC stage (Clinical)**
Ⅰ (%)	7 (17.94)	11 (23.40)	6 (18.18)	14 (35.00)	--	--	--	0 (0)
Ⅱ (%)	4 (10.25)	6 (12.77)	6 (18.18)	11 (27.50)	--	--	--	0 (0)
Ⅲ (%)	24 (61.53)	19 (40.43)	11 (33.33)	12 (30.00)	--	--	--	7 (58.33)
Ⅳ (%)	4 (10.25)	11 (23.40)	10 (30.30)	3 (7.50)	--	--	--	5 (41.67)
Differentiation								
highly/moderately (%)	5 (12.82)	7 (14.89)	7 (21.21)	6 (15.00)	--	--	--	0 (0)
poorly (%)	33 (84.61)	35 (74.47)	26 (78.79)	32 (80.00)	--	--	--	0 (0)
signet-ring cell (%)	1 (2.56)	5 (10.64)	0 (0.00)	2 (5.00)	--	--	--	12 (100.00)
Position								
Upper (%)	9 (23.07)	12 (25.53)	14 (42.42)	14 (35.00)	--	--	--	5 (41.67)
Middle (%)	9 (23.07)	11 (23.40)	7 (21.21)	12 (30.00)	--	--	--	4 (33.33)
Lower (%)	20 (51.28)	22 (46.81)	12 (36.36)	13 (32.50)	--	--	--	2 (16.67)
Total Stomach (%)	1 (2.56)	2 (4.26)	0 (0.00)	1 (2.50)	--	--	--	1 (8.33)

Table 1. Clinical characteristics of the study population

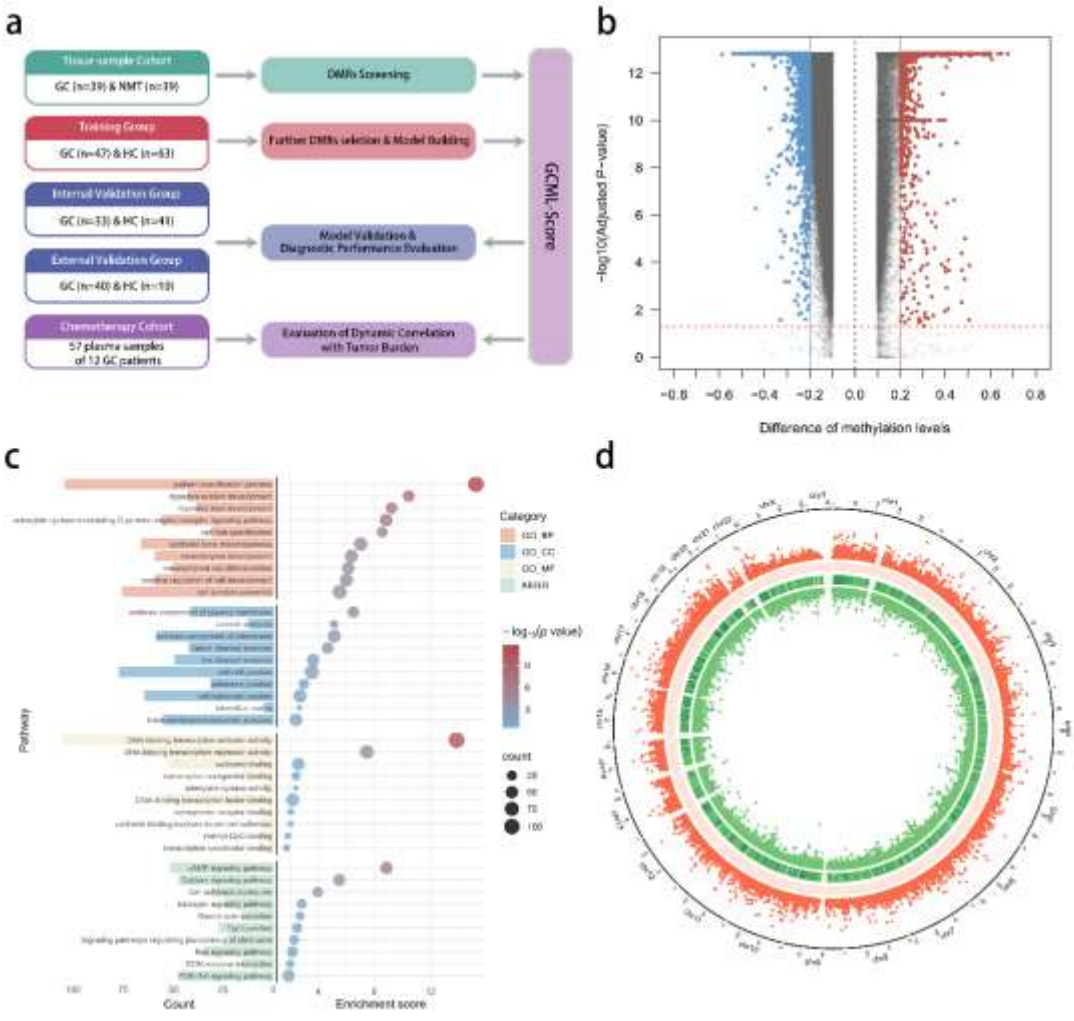

Figure 1

a The cohort design and workflow of the research. **b** Volcano plot showing the relationship between methylation levels of DMRs and −log10 P values in gastric cancer tissues versus adjacent noncancerous tissues. Red and blue dots represent significantly hypermethylated and significantly hypomethylated DMRs, respectively. **c** The bar chart (left) and the bubble plot (right) showing GO/KEGG pathways enrichment analysis of DMRs-associated genes. The bar chart displays gene counts per biological pathway, with categorical color schemes denoting shared functional modules. Bubble color responds to the enrichment significance and bubble size is related to the ratio of the number of genes mapped to a certain pathway in the bubble plot, x-axis position responds to the enrichment score. **d** Circos plot of the genome-wide distribution significance of DMRs. From the outside to the inside, the figure displays: a scatter plot showing the distribution of hyper-DMRs, oriented outward, where points farther from the center indicate higher significance; a heatmap of GC content, with darker colors representing higher GC content; a heatmap of gene density, with darker colors representing higher gene density; and a scatter plot showing the distribution of hypo-DMRs, oriented inward, where points closer to the center indicate higher significance.

A GC diagnostic model based on cfDNA methylation markers

We subsequently investigated whether a diagnostic model can be constructed based on the methylation level of cfDNA. The plasma samples were randomly divided in a 6:4 ratio from 104 healthy controls and surgery GC patients of FAH-SYSU into training and internal validation cohorts. The LASSO regression model was employed to screen for significant DMRs. The training process utilized an iterative 10-fold cross-validation approach. When λ reached 0.1198 (log(λ) = -2.122), the model achieved optimal performance with the fewest variables (Figure 2c-d). Considering in cfDNA-based methylation analyses, hypomethylated signals are more susceptible to background noise derived from normal cfDNA and tend to show reduced robustness and reproducibility. Therefore, we deliberately focused on hypermethylated DMRs during LASSO-based feature selection. Variables demonstrating coefficients greater than 0 under this model configuration were selected, ultimately identifying 13 gastric cancer-related DMRs (GC-DMRs) (Figure 3a). All GC-DMRs exhibited significant hypermethylation in gastric cancer patients' cfDNA compared to healthy controls ($p<0.001$, Figure 3b). Then we retained the original LASSO tuning strategy and conducted bootstrap resampling (B = 200) for model retraining, during which the variables selected into the model were recorded for each iteration. We then summarized the selection frequencies of 2,079 DMRs. The results showed that all 13 DMRs originally included in the model ranked among the top-selected features, with selection frequencies exceeding 50%, indicating that the previous variable selection procedure was relatively robust and reliable (Figure 3c).

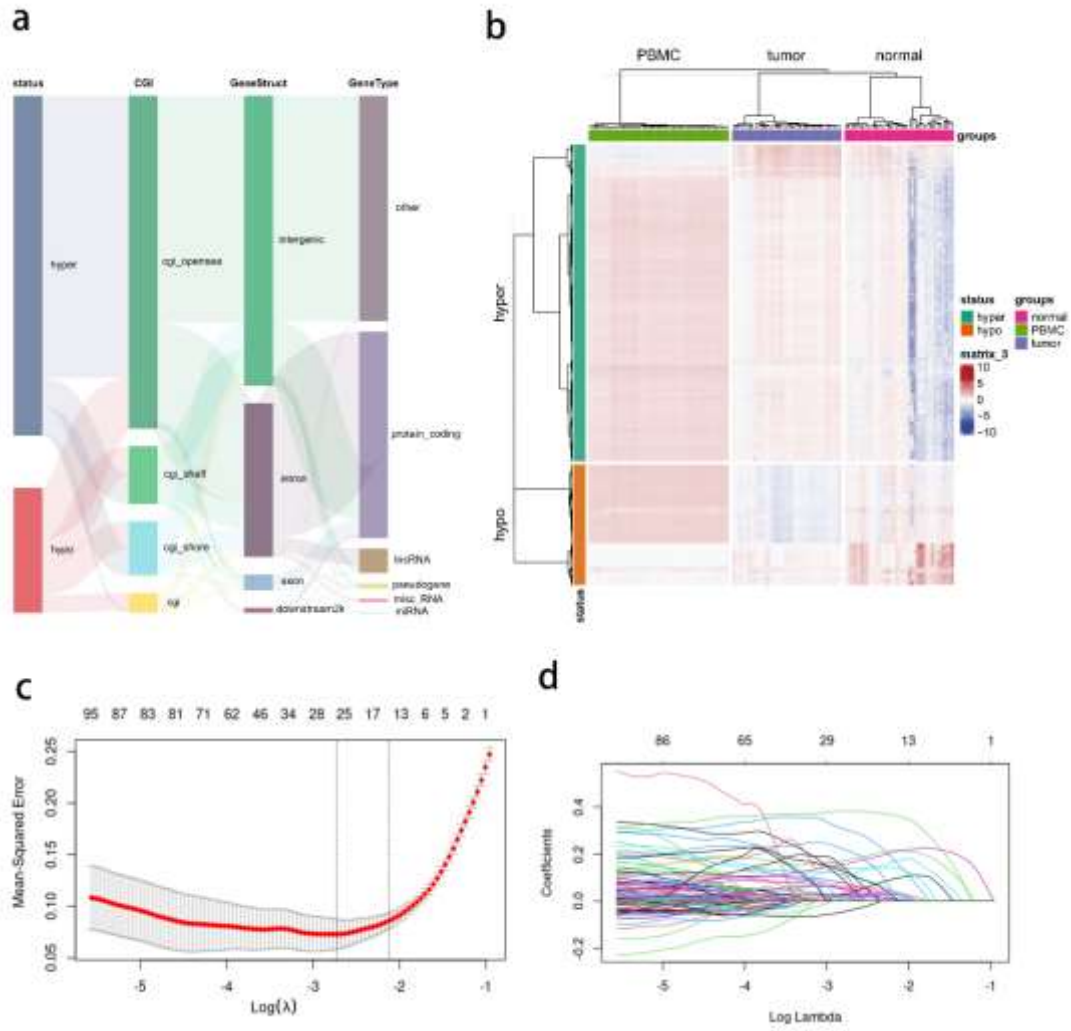

Figure 2

a The Sankey diagram showing the relationships between DMRs and CpG island structures, gene structures, and the corresponding gene functions. **b** Unsupervised hierarchical clustering of 22051 DMRs across PBMCs, gastric tumor tissues, and adjacent non-tumorous tissues. **c** 10-fold cross-validation was employed for parameter selection in the least absolute shrinkage and selection operator (LASSO) model. LASSO regression was applied to high-dimensional predictor variables. This method utilizes L1 penalty to shrink specific regression coefficients exactly to zero. The binomial deviance curve was plotted as a function of log(λ), where λ denotes the regularization parameter. **d** LASSO coefficient profiles.

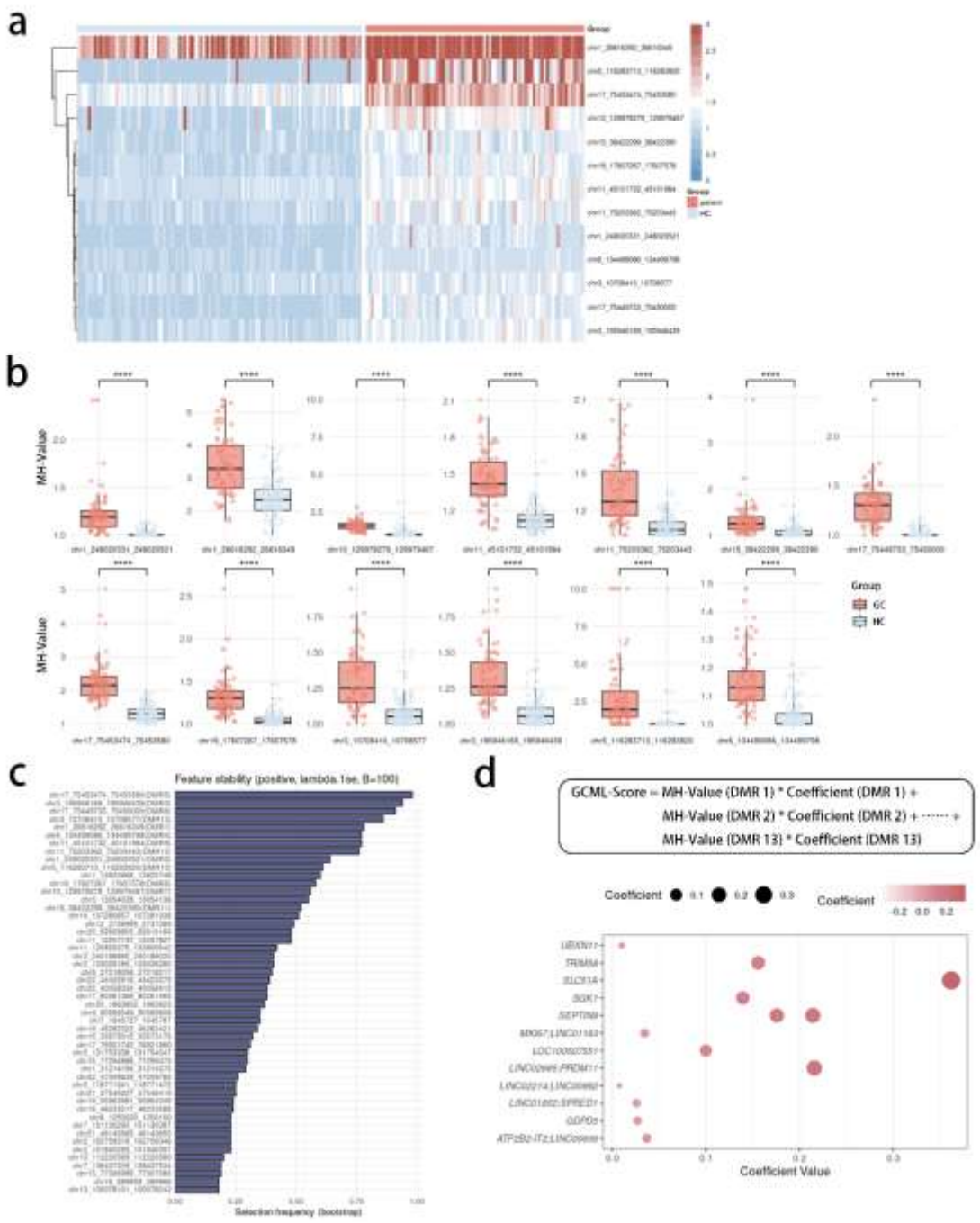

Figure 3

a An unsupervised hierarchical clustering plot of MH-values of GC-DMRs demonstrated that these regions, exhibiting significantly differential methylation levels in cfDNA between gastric cancer (GC) patients and healthy controls (HC) individuals, optimally discriminated HC individuals from GC patients. **b** Box plots demonstrated significant hypermethylation of GC-DMRs in plasma cfDNA from gastric cancer patients. **c** Stability of DMR selection assessed by bootstrap-LASSO resampling (B=200). **d** Formula for calculating the GCML-Score (top); Bubble plot (bottom) displaying genes and coefficients corresponding to GC-DMRs.

Among the screened DMRs, five were located in intergenic regions, one in a non-coding region, and seven in gene coding regions. Notably, 2 DMRs (chr17_75449733_75450000, chr17_75453474_75453580) were positioned within the SEPTIN9 gene. As a protein-coding gene associated with the cell cycle,

hypermethylation of this gene has been linked to the initiation and progression of gastrointestinal tumors in multiple studies[24-27]. Our study independently validated the clinical significance of SEPTIN9 methylation as a biomarker for gastric cancer screening. Subsequently, we aimed to quantify methylation levels of the selected genes, construct a cfDNA methylation-derived score, and evaluate its diagnostic potential for gastric cancer. Weighting based on LASSO-derived coefficients was applied to the MH-values of GC-DMRs, with the summation of these weighted values designated as the Gastric Cancer Methylation Level score (GCML-Score) (Figure 3d, Table 2).

The ideal threshold values (1.70) for diagnosis were pinpointed in the training cohort by maximizing the difference between the true positive and false positive rates (TPR-FPR) (Figure 4a). Notably, The GCML-Score demonstrated robust diagnostic performance across both cohorts, achieving AUC values of 0.95 in the training cohort and 0.99 in the internal validation cohort, with corresponding accuracy of 86.4% and 93.2%. Furthermore, the diagnostic performance of the GCML-Score for early-stage gastric cancer was assessed, revealing AUCs of 0.96 (95% CI: 0.91–1.00) in the training cohort and 0.99 (95% CI: 0.97–1.00) in the validation cohort (Figure 4b-d). To evaluate the generalizability of the GCML-Score, we additionally collected plasma samples from 40 previously untreated gastric cancer patients and 10 healthy controls at SAH-SYSU to construct an external validation cohort, and cfDNA methylation profiling was performed using the same analytical approach. In this independent external validation cohort, the GCML-Score achieved an overall diagnostic AUC of 0.95 (95% CI: 0.90–1.00), while the AUC for early-stage gastric cancer reached 0.82 (95% CI: 0.55–1.00) (Figure 4e-f).

Given the relatively small sample size of the cohorts, to reduce potential bias, we performed bootstrap resampling (5-fold, B = 500) on the training cohort, internal validation cohort, and external validation cohort to further assess the robustness of the GCML-Score. Across multiple resampling iterations, the GCML-Score consistently maintained high AUC values, and the calibration curves further supported its diagnostic reliability. The Brier scores for the above cohorts were 0.086, 0.095, and 0.072, respectively (Figure 5a). Furthermore, ROC curves were constructed based on each bootstrap resampling iteration, and the median AUC values in all three cohorts exceeded 0.95 (Figure 5b). The above results demonstrate that GCML-Score is an effective blood biomarker for early diagnosing gastric cancer.

DMR_ID	DMR	Chromosome	Start	End	Gene	Region	Coefficient
DMR1	chr1_26616292_26616349	chr1	26616292	26616349	UBXN11	intronic	0.010342357
DMR2	chr1_248020331_248020521	chr1	248020331	248020521	TRIM58	UTR5	0.155892985
DMR3	chr3_195946169_195946439	chr3	195946169	195946439	SLC51A	intronic	0.362055173

DMR4	chr6_134499086_134499798	chr6	134499086	134499798	SGK1	intronic	0.13948758
DMR5	chr17_75453474_75453580	chr17	75453474	75453580	SEPTIN9	intronic	0.214324466
DMR6	chr17_75449733_75450000	chr17	75449733	75450000	SEPTIN9	intronic	0.175936703
DMR7	chr10_129979279_129979467	chr10	129979279	129979467	MKI67;LINC01163	intergenic	0.034726013
DMR8	chr19_17607267_17607578	chr19	17607267	17607578	LOC100507551	intronic	0.100077171
DMR9	chr11_45101732_45101984	chr11	45101732	45101984	LINC02685;PRDM11	intergenic	0.215988872
DMR10	chr5_116283713_116283920	chr5	116283713	116283920	LINC02214;LINC00992	intergenic	0.007695582
DMR11	chr15_38422299_38422390	chr15	38422299	38422390	LINC01852;SPRED1	intergenic	0.026036065
DMR12	chr11_75203362_75203443	chr11	75203362	75203443	GDPD5	intronic	0.027115591
DMR13	chr3_10708410_10708577	chr3	10708410	10708577	ATP2B2-IT2;LINC00606	intergenic	0.037018087

Table 2. Genes and coefficients corresponding to GC-related DMRs

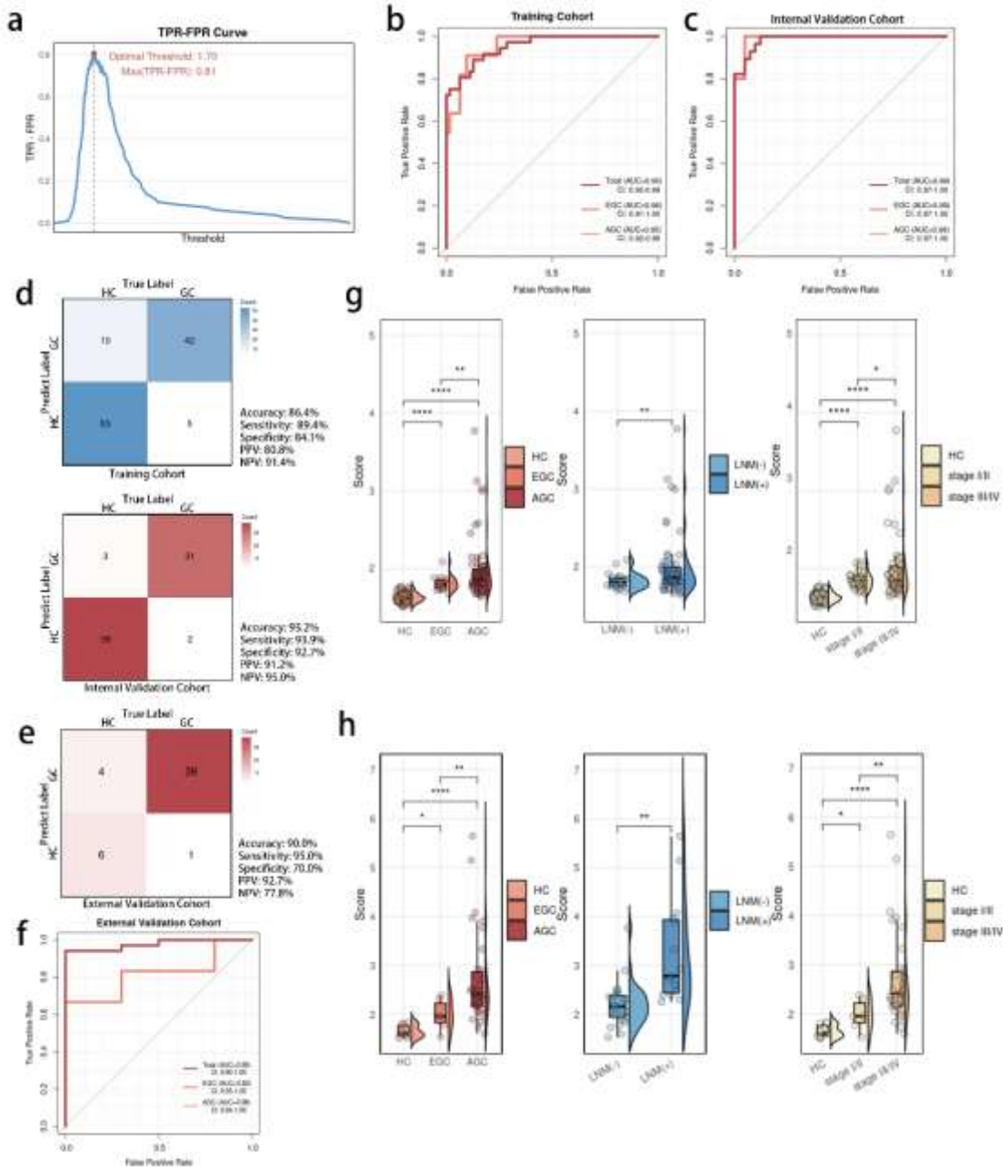

Figure 4

a True Positive Rate - False Positive Rate curve in training cohort. **b-c** Receiver Operative Characteristic curves of GCML-Score for early gastric cancer (EGC), advanced gastric cancer (AGC) and total gastric cancer patients in the training cohort(b) and the internal validation cohort (c). **d** Confusion matrix for the GCML-Score in the training cohort and the internal validation cohort. **e** Confusion matrix for the GCML-Score in the external validation cohort. **f** Receiver Operative Characteristic curves of GCML-Score for early gastric cancer (EGC), advanced gastric cancer (AGC) and total gastric cancer patients in the external validation cohort. **g** Comparison of GCML-Score across clinical cohort in the training and internal validation cohort. Left: healthy controls (HC), early gastric cancer (EGC), and advanced gastric cancer (AGC); Middle: lymph node metastasis negative [LNM (−)] vs. positive [LNM (+)]; Right: HC, stage I–II, and stage III–IV patients. **h** Comparison of GCML-Score across clinical cohorts in the

external validation cohort. Left: healthy controls (HC), early gastric cancer (EGC), and advanced gastric cancer (AGC); Middle: lymph node metastasis negative [LNM (−)] vs. positive [LNM (+)]; Right: HC, stage I–II, and stage III–IV patients. ns, not significant, *P ⩽ 0.05, **P ⩽ 0.01, ***P ⩽ 0.001, ****P ⩽ 0.0001

To assess potential bias introduced by data partitioning, we performed 30 rounds of random resampling of all healthy controls and surgical cohort samples using different split ratios, and calculated the AUC and accuracy for each new grouping. As the proportion of positive samples gradually decreased through dilution, both the AUC and accuracy exhibited certain degrees of fluctuation and decline. Nevertheless, even when the ratio of gastric cancer patients to healthy controls reached 1:9, the accuracy remained consistently above 0.7, and the AUC remained above 0.8 in the majority of resampling iterations (Figure 5c). To more closely approximate real-world screening conditions and evaluate the diagnostic performance of the GCML-Score, we performed random upsampling of healthy control samples at GC-to-HC ratios ranging from 1:10 to $1:10^6$, with 10 resampling iterations for each ratio. We then calculated the AUC and accuracy of the GCML-Score at each ratio. As the proportion of gastric cancer cases was progressively diluted, the AUC remained consistently above 0.98, while the accuracy was stably maintained above 0.7 (Figure 5d), indicating strong robustness under low-prevalence conditions. Importantly, these findings suggest that the GCML-Score maintains robust and stable diagnostic performance in real-world population-based screening settings, where disease prevalence is typically low.

Next, we investigated whether the GCML-Score were associated with the tumor burden of GC patients. Significantly elevated GCML-Scores was observed in advanced gastric cancer (AGC) patients compared to early-stage cases ($p<0.01$). According to AJCC staging criteria, stage III/IV GC patients demonstrated markedly higher scores than stage I/II patients ($p<0.05$). Moreover, the GCML-Score showed strong association with lymph node metastasis (LNM), with LNM-positive patients exhibiting significantly increased scores relative to LNM-negative counterparts ($p<0.01$, Figure 4g-h), indicating the scores were associated with the known reported tumor burden factors.

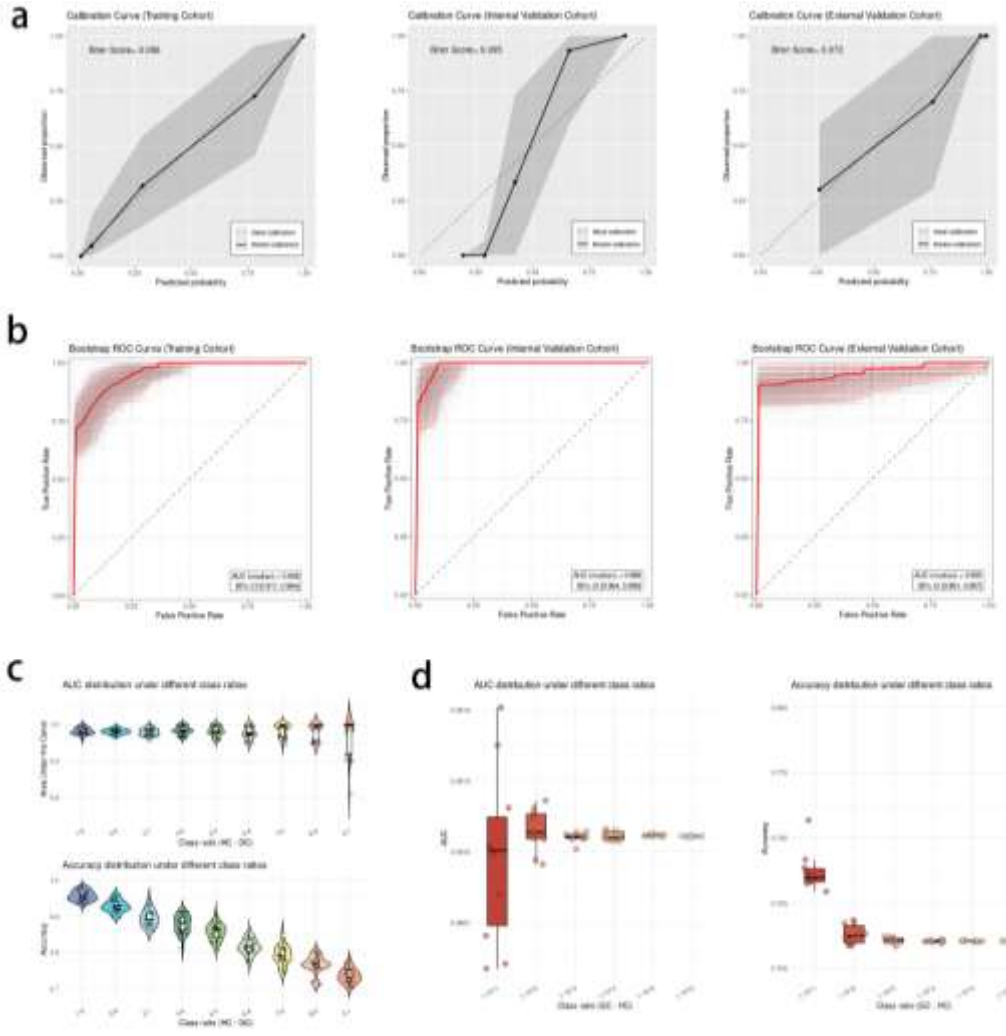

Figure 5

a Calibration curves of GCML-Score after bootstrap resampling (B = 500, g=5). Left: Training cohort; Middle: internal validation cohort; Right: external validation cohort. **b** Receiver Operative Characteristic curves of GCML-Score after bootstrap resampling (B = 500, g=5). Left: Training cohort; Middle: internal validation cohort; Right: external validation cohort. **c** Area Under the Curve (top) and accuracy (bottom) of GCML-Score after 30 rounds of random resampling of all healthy controls and surgical cohort samples using different split ratios (HC:GC from 1:9 to 9:1). **d** Area Under the Curve (left) and accuracy (right) of GCML-Score after random upsampling of healthy control samples at GC:HC ratios ranging from 1:10 to 1:10^6

Tumor burden assessment of GC patients undergoing preoperative chemotherapy by GCML-Score monitoring

This study included a total of 12 gastric cancer patients who received neoadjuvant chemotherapy. Each of these patients underwent blood sample collection before treatment and at least once during the

treatment process. Among them, 1 patient (P005) underwent surgical treatment during the follow-up period.

The surgical treatment, radiological tumor burden changes, and GCML-Score were marked on each patient's follow-up timeline. We found that the trend of GCML-Score changes was generally consistent with the trend of tumor burden changes in the patients. 7/12 of patients experienced a decrease in GCML-Score after the first cycle of chemotherapy. Consistent correlations were observed between longitudinal changes in GCML-Score and radiographic tumor burden dynamics across the 12-patient cohort (Figure 6a, Table 3). Then we calculated the average daily change rate (ADCR) of the GCML-Score between each two consecutive sampling time points. Based on radiological assessment results, each monitoring interval was categorized into PR, SD, or PD phases. Statistical analysis showed that the ADCR during SD phases intervals was close to zero, whereas intervals corresponding to PR and PD phases were predominantly associated with negative and positive ADCR, respectively. Using the Jonckheere–Terpstra trend test, we observed a significant increasing trend in the ADCR across intervals of PR, SD, and PD phases ($J = 594$, $p = 4.48 \times 10^{-10}$) (Figure 6b). In addition, we performed a lead-time analysis in this cohort. Patient disease courses were stratified into different phases based on CT assessment results, and for each phase, the percentage change of each GCML-Score relative to the phase-specific baseline was calculated. Statistical analysis showed that, across all PR or PD phases, the median lead time for detection intervals with changes exceeding 10% was 36.5 days, and the median lead time for changes exceeding 20% was 43 days, both of which were substantially earlier than the recommended imaging follow-up interval of 8 weeks (Figure 6c).

Patient ID	Date	GCML-Score	CT assessment /Surgery	Patient ID	Date	GCML-Score	CT assessment /Surgery
P001	2021/10/01		BL		2022/7/11	3.256	
	2021/11/12	3.529399			2022/7/29		PD
	2021/12/01		PR	**P007**	2022/2/25	3.134166	BL
	2021/12/02	1.787171			2022/3/21	1.506697	
	2021/12/22	1.655418			2022/4/21	1.506697	
	2022/01/05		SD		2022/5/10		PR
	2022/01/06	1.75251			2022/5/13	1.562664	
	2022/02/11	1.562994			2022/6/22	3.48907	
	2022/03/03	1.506697			2022/8/9	5.691498	PD
	2022/03/17		PR	**P008**	2022/2/7	2.779587	BL
P002	2021/11/28	2.052097	BL		2022/3/2	1.599778	
	2021/12/24	2.350555			2022/3/24		PR
	2022/1/18	2.669984			2022/4/21	3.137912	
	2022/2/11	3.485909	PD		2022/5/10		PD
P003	2021/12/21	2.405423	BL	**P009**	2022/2/28	4.192528	
	2022/1/14	1.506697			2022/3/2		BL
	2022/3/7		PR		2022/4/12	2.019794	
	2022/3/8	1.552872			2022/5/11	1.874739	
	2022/4/6	1.579625			2022/5/27		PR
	2022/4/7		PR		2022/6/23	1.506697	
	2022/4/17		OR		2022/7/20	1.506697	PR
	2022/4/18	2.399228			2022/8/2		OR
	2022/5/13		NET		2022/8/3	1.819231	
P004	2021/12/21	1.553514	BL		2022/8/4		NET
	2022/1/15	1.697238		**P010**	2022/5/20		BL
	2022/4/28	1.709801	SD		2022/5/25	3.480467	
P005	2021/12/25	1.682605	BL		2022/7/5	1.702159	
	2022/1/21	2.063227			2022/7/25	1.506697	
	2022/2/21	3.923062			2022/7/26		PR
	2022/5/26	5.222771	PD		2022/8/2		OR
	2022/5/27		OR		2022/8/3	2.024253	
	2022/7/27	1.690149			2022/8/23		NET
	2022/8/18	1.907939	PD	**P011**	2022/7/21	1.839859	BL
	2022/9/10	3.402661			2022/9/27	1.750622	
P006	2022/1/19	2.262923			2022/10/19		SD
	2022/2/1		BL		2022/10/29	1.857021	
	2022/2/25	1.875528			2022/11/20	1.940123	SD
	2022/3/24	1.506697		**P012**	2022/9/5	2.507363	BL
	2022/4/12	1.506697			2022/10/24	1.584585	

2022/4/19		PR	2022/11/16	1.565231	
2022/6/20	3.119304		2022/12/9		PR

Table 3. Longitudinal changes in GCML-Score and corresponding radiographic tumor burden in the chemotherapy cohort

Patient P007 was a 42-year-old female who underwent radiological re-evaluation 74 days after receiving neoadjuvant chemotherapy, which indicated a partial response. Correspondingly, the GCML-Score decreased from 3.13 before treatment to 1.51. Subsequently, a second radiological re-evaluation conducted 91 days later showed a marked progression because the tumor developed resistance. Similarly, during this period, the GCML-Score gradually increased (from 1.56 to 5.69) (Figure 7). Meanwhile, patient P005 (63-year-old male with hepatic metastatic gastric cardia adenocarcinoma) demonstrated progressive GCML-Score elevation during systemic chemotherapy. Crucially, a 2-fold score increase over baseline (from 1.68 to 3.92) occurred 94 days prior to radiographic confirmation of primary tumor progression and L3 vertebral metastasis. Following palliative L3 spondylectomy, postoperative day 60 cfDNA methylation analysis revealed significant score reduction (from 5.22 to 1.69), attributable to metastatic burden resection. However, subsequent chemotherapy was accompanied by continued primary tumor progression and recurrent progressive GCML-Score elevation, indicating chemoresistance development. The above example indicates that the GCML-Score exhibited a consistent trend with tumor burden in patients undergoing chemotherapy.

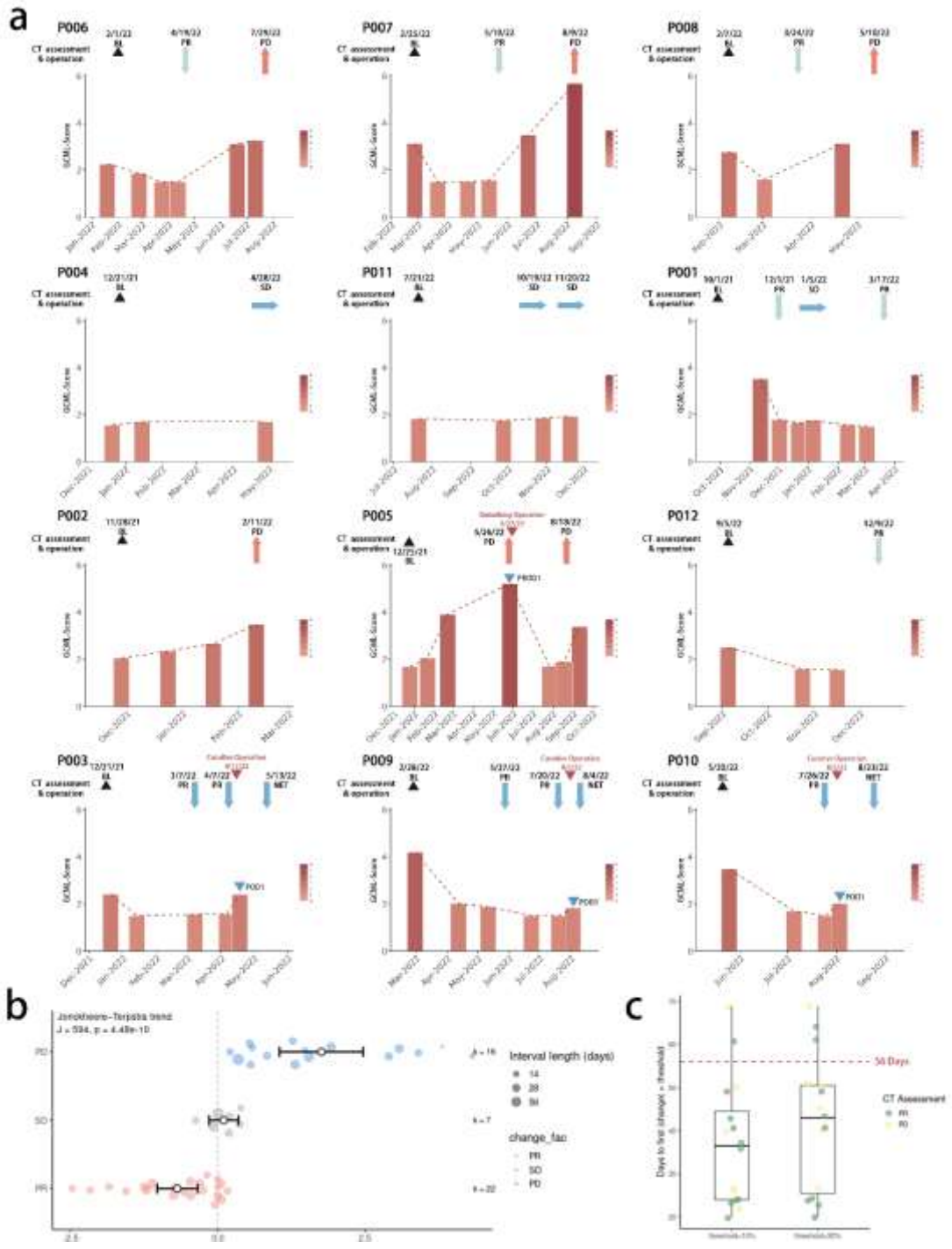

Figure 6

a Dynamic changes in GCML-Score, CT-based tumor burden assessment, and surgical events across NAC treatment course in 12 patients of Chemotherapy Cohort. BL: baseline, PR: partial response, PD: progressive disease, SD: stable disease, NET: no evidence of tumor. PROD1: preoperative day 1, POD1: postoperative day 1. **b** Average Daily Change Rate (ADCR) of GCML-Score across treatment response phases (PD, SD, and PR) and the corresponding Jonckheere–Terpstra trend test. **c** Time to achieve GCML-Score changes >10% (left) and >20% (right) across non-SD treatment response phases.

Among the preoperative chemotherapy cohort, 3 patients (P003, P009, P010) underwent curative operation (R0 resection) eventually. All 3 patients exhibited obvious reductions in GCML-Scores at preoperative reassessment compared to baseline levels, consistent with corresponding CT evaluations. Patient P009, a 36-year-old female, achieved partial response after five cycles of neoadjuvant

chemotherapy. Concurrently, her GCML-Score decreased from 4.19 at baseline to 1.51 prior to surgical resection. But interestingly, retesting of cfDNA on postoperative day 1 (POD1) revealed a marginal score elevation relative to preoperative values (from 1.51 to 1.82) (Figure 8). This consistent fluctuation pattern was reproducibly observed in two additional patients who underwent radical operation. We hypothesize that this increase may reflect intraoperative release of tumor-derived materials into circulation due to surgical manipulation.

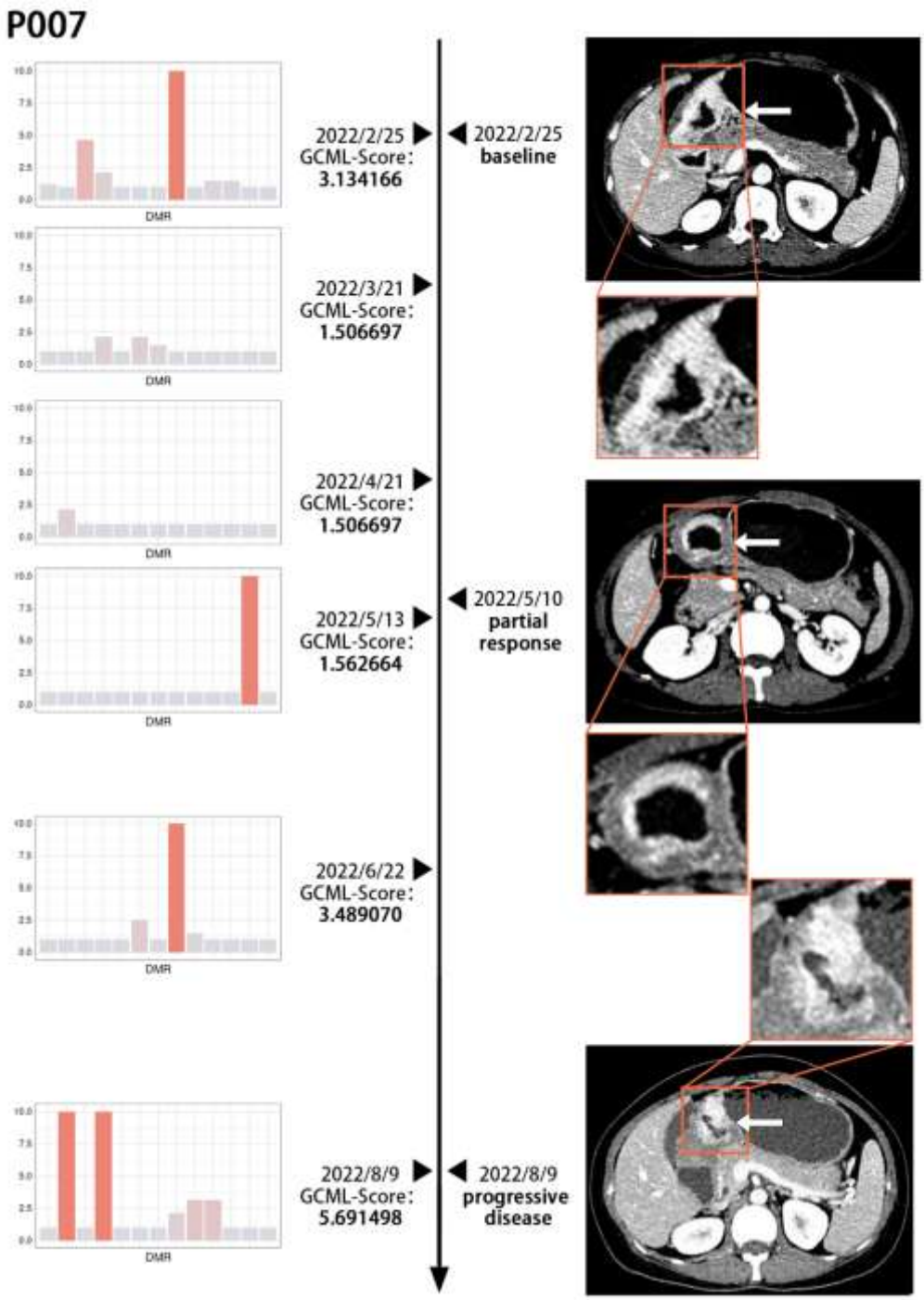

Figure 7

Timeline of patient P007 with corresponding GCML-Scores and CT images at multiple time points.

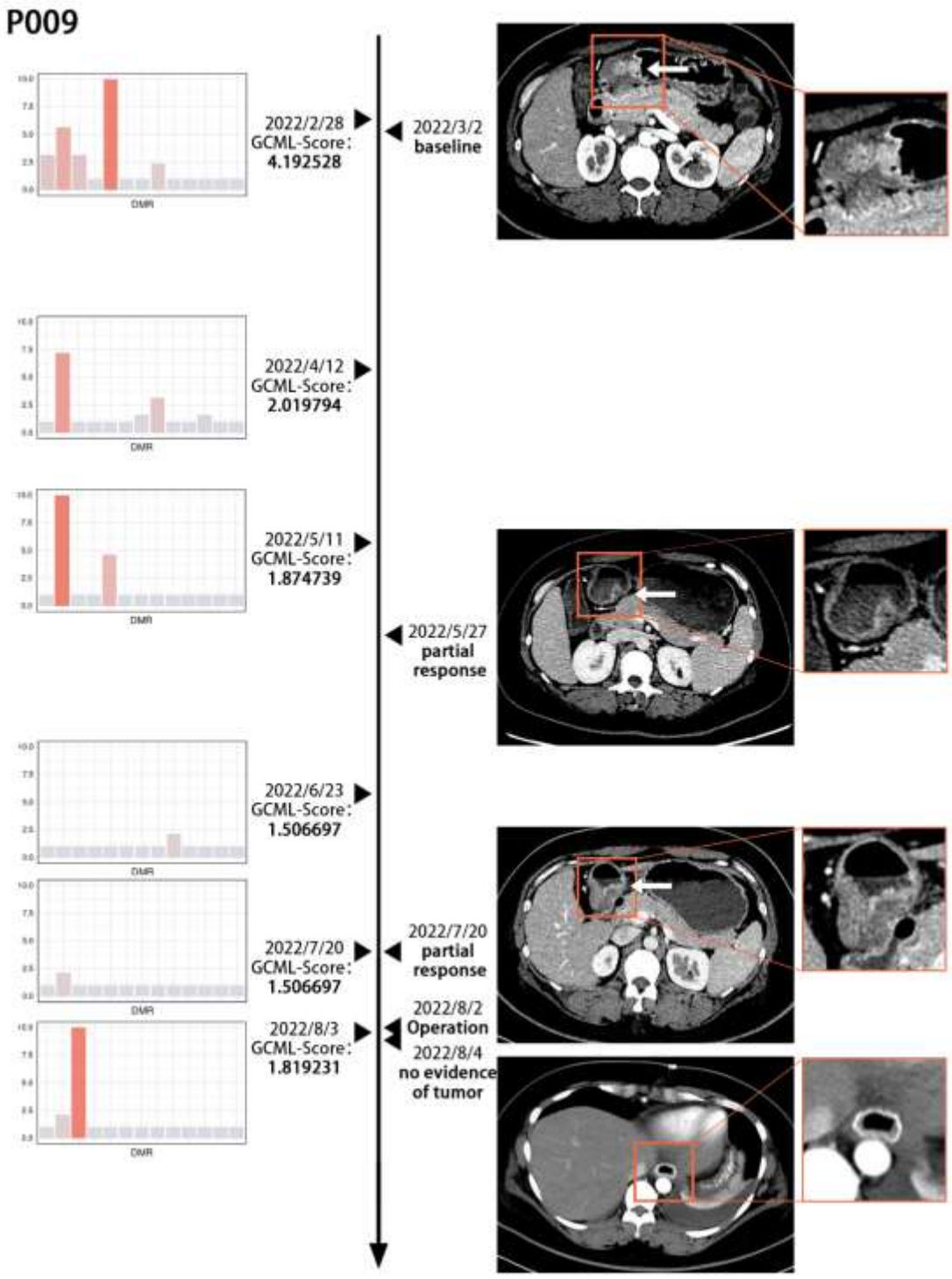

Figure 8

Timeline of patient P009 with corresponding GCML-Scores and CT images at multiple time points.

Discussion

This study created a cfDNA-based methylation biomarker for gastric cancer through whole-genome methylation sequencing, Methylome-seq, GCML-Score, composed of the weighted methylation value of 13 DMRs. A dual-track DMR selection strategy was implemented using paired tissue and plasma samples. Tissue-derived methylation signatures (gastric cancer vs. normal mucosa) captured universal epigenetic alterations associated with tumorigenesis. Concurrently, cfDNA-based signatures (cancer patients vs. healthy controls) effectively filtered out leukocyte background noise to amplify tumor-specific circulating signals. This integrated approach yielded blood-optimized DMR markers with enhanced detection capability. The GCML-Score demonstrated robust diagnostic performance for gastric cancer, particularly excelling in early-stage detection (training AUC: 0.96; internal validation AUC: 0.99; external validation AUC: 0.82). This enhanced sensitivity may be attributable to our cohort's elevated proportion of early-stage cases (15.38% in Tissue-sample Cohort; 20% in surgery Cohort), suggesting that the GC-DMRs capture epigenetic signatures enriched in early tumorigenesis. Moreover, patients with advanced-stage cancer or lymph node metastasis exhibited significantly elevated GCML-Scores, indicating its capacity to reflect tumor burden.

Clinically, the diagnosis and monitoring of gastric cancer mainly rely on endoscopy and imaging techniques. In recent years, liquid biopsy has provided a less invasive and more convenient solution to above issues[28]. Epigenetic modifications of DNA are early events in tumorigenesis, and the detection of abnormal methylation patterns in tumor-derived cfDNA as tumor biomarkers has been widely studied[29,30]. In 2021, Ren et al. identified 153 gastric cancer-associated cfDNA methylation sites through whole-genome methylation sequencing and developed a cfDNA methylation-based diagnostic tool for gastric cancer with clinical application value. However, despite the overall good diagnostic performance, the model had low diagnostic sensitivity for stage I patients (44%), making it insufficient for early screening tasks[19]. In contrast, the GCML-Score has a higher detection rate for early gastric cancer.

Due to the high heterogeneity of gastric cancer, the choice of preoperative chemotherapy regimen and the duration of treatment have been subjects of ongoing debate[31-33]. In current clinical practice, the efficacy of preoperative chemotherapy is primarily assessed by comparing tumor characteristics on imaging studies obtained before and after treatment[34]. However, CT is not suitable as a sensitive and frequently monitored indicator, leading to a delay in assessing the treatment response, which may result in missed opportunities for adjusting the treatment plan for some patients. In recent years, some studies have focused on applying ctDNA/cfDNA in the context of tumor preoperative chemotherapy, including breast cancer, prostate cancer and so on[7,35,36]. However, few studies have tried to use abnormal methylation

to assess tumor burden during the preoperative chemotherapy process of gastric cancer. In this study, through a longitudinal monitoring of 12 neoadjuvant chemotherapy recipients, GCML-Score trajectories consistently paralleled radiographically assessed tumor burden changes. GCML-Score fluctuations preceded radiographic alterations, suggesting its potential utility for early chemotherapy response assessment. GCML-Score provides a new approach for tumor burden monitoring in gastric cancer neoadjuvant chemotherapy.

SEPTIN9, a cell cycle-associated protein, suppresses tumorigenesis through modulation of hypoxia-inducible factor (HIF) pathways, Rho GTPase signaling, and mitotic regulation and so on[37-39]. Extensive clinical evidence has established blood-based SEPTIN9 methylation analysis as a high-performance diagnostic biomarker for colorectal cancer, particularly for early detection applications[40-42]. In 2023, Nie et al. pioneered a dual-marker approach detecting concurrent SEPTIN9 and RNF180 methylation in peripheral blood. Their multicenter prospective cohort study demonstrated clinically significant performance for gastric cancer diagnosis and early detection, achieving 50.9% sensitivity for Stage I tumors—surpassing conventional serum biomarkers[26]. In our study, 2 DMRs (chr17:75449733-75450000; chr17:75453474-75453580) among the 13 GC-DMRs mapped to the SEPTIN9 gene. Both demonstrated substantial weighting in the GCML-Score algorithm (0.175 and 0.214 respectively), establishing SEPTIN9 methylation as an integral contributor to the model's predictive capacity, which was consistent with previous research. Data from TCGA-STAD showed that the expression level of SEPTIN9 in gastric adenocarcinoma tissues was significantly higher than that in normal gastric mucosa, suggesting that DNA methylation of SEPTIN9 is likely involved in regulating its expression and may represent an important event in the development of gastric adenocarcinoma (Supplementary Figure 1a). The apparent discrepancy between SEPTIN9 hypermethylation in cfDNA and its increased expression in tumor tissue highlights the context-dependent nature of DNA methylation. Importantly, the DMR identified in this study is not located within the SEPTIN9 promoter region, and gene-body or distal methylation has been reported to show variable or even positive associations with transcriptional activity[43]. In addition, cfDNA methylation captures tumor-derived epigenetic signals rather than functional gene expression states[44], supporting the interpretation of SEPTIN9 methylation as a biomarker of tumor burden rather than a direct epigenetic driver[45].

Notably, the DMR within SLC51A (chr3_195946169_195946439) demonstrated the highest weighting contribution to the GCML-Score (0.281). This gene encodes the alpha subunit of the organic solute transporter (OSTα), primarily responsible for transmembrane transport of bile acids and steroids—a function extensively documented in cholestatic disorders. While SLC51A remains unstudied in gastric cancer, Li et al.[46] established that bile acids promote gastric intestinal metaplasia, which is a precancerous lesion. What's more, TCGA-STAD data reveal significant SLC51A downregulation in gastric adenocarcinoma (Supplementary Figure 1b). We postulate that SLC51A hypermethylation may dysregulate

intracellular bile acid levels, thereby facilitating gastric carcinogenesis, but the specific regulatory mechanisms remain to be further investigated. Although certain DMRs were mapped to genes previously implicated in cancer-related pathways, functional validation would be required to determine whether these methylation changes represent epigenetic driver events or passenger alterations associated with tumor burden.

Several limitations merit consideration in this study: First, the single-center retrospective cohort design necessitates external validation through prospective multicenter studies to confirm generalizability across diverse populations. Second, although the GC-DMRs panel demonstrates clinical utility, mechanistic insights into specific methylation loci require functional validation to elucidate their biological roles in gastric carcinogenesis.

In conclusion, our study establishes that cfDNA methylation analysis serves as a powerful biomarker for tumor burden assessment, enabling not only early detection of gastric cancer with high diagnostic accuracy but also early prediction of therapeutic response during neoadjuvant chemotherapy. This approach may facilitate timely and evidence-based treatment modifications to optimize long-term survival outcomes.

Methods

Patients and cohorts

369 samples from 285 patients and healthy control were included in this study in total. This study included a total of 131 gastric cancer patients from the First Affiliated Hospital of Sun Yat-sen University (FAH-SYSU), 40 gastric cancer patients from the Seventh Affiliated Hospital of Sun Yat-sen University (SAH-SYSU) and 114 healthy control individuals. Among the patients, 39 provided both gastric cancer tissues and matched adjacent mucosal tissues (Tissue-sample Cohort); 80 patients provided preoperative plasma samples and subsequently underwent surgical treatment (Surgery Cohort); an additional 12 patients received perioperative chemotherapy and provided plasma samples before and during treatment (Chemotherapy Cohort). The healthy control cohort was recruited from individuals undergoing routine health examinations, and participants with comorbidities, inflammatory conditions, or infections were excluded prior to enrollment. Among them, PBMCs from 50 healthy controls were subjected to methylation sequencing to eliminate potential cfDNA noise derived from blood cells. No individual contributed samples to more than one cohort. The training, internal/external validation cohorts and chemotherapy consisted of independent participants. The inclusion and exclusion criteria for each cohort are shown in Supplementary Table 1, and the CONSORT flow diagram illustrating the study enrollment process has been provided in Supplementary Figure 3.

For the patients in the chemotherapy cohort, we prospectively collected blood samples for cfDNA analysis at baseline, before each chemotherapy cycle and before operation. These patients had the option to decline blood collection, so blood samples were not obtained from all time points for most patients.

This study was conducted in accordance with the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University (Ethics Review [2022] No. 078) and the Ethics Committee of the Seventh Affiliated Hospital of Sun Yat-sen University (Approval No. KY-2022-039-03). Written informed consent was obtained from all participants prior to sample collection.

Blood Collection and Plasma Separation

Peripheral blood was collected according to standard venipuncture procedures using specialized Streck blood collection tubes (10 mL capacity). Two tubes were drawn from each participant, with 8–10 mL of blood per tube. After collection, the tubes were gently inverted four times and stored at room temperature. Blood samples were maintained at room temperature (15–35°C) for no longer than 3 days before plasma separation.

Plasma separation was performed as follows:

First centrifugation: Samples were centrifuged at 3,000 × g for 10 minutes at 16°C or room temperature using a low-speed centrifuge with swing buckets. The 10 mL collection tubes were directly placed into the centrifuge. The supernatant plasma was carefully transferred into a new 10 mL EP tube. Plasma derived from the two tubes of the same participant was combined into a single 10 mL EP tube. Care was taken to avoid disturbing the buffy coat layer containing leukocytes and erythrocytes.

Second centrifugation: The collected plasma was centrifuged again at 3,000 × g for 10 minutes at 16°C or room temperature. The supernatant was slowly aspirated and aliquoted into new 5 mL EP tubes, with 4 mL plasma per tube. Care was taken to avoid aspirating any pellet at the bottom of the tube, and the pipette tip was positioned opposite the sediment.

Blood cell preservation: After plasma processing, 1 mL of the buffy coat layer from the first centrifugation was transferred into a new 2.0 mL EP tube and labeled with the corresponding sample identification number.

DNA Isolation and Library Construction

At least 300ng genomic DNA was isolated from fresh-frozen tissue samples using the Qiagen DNeasy Blood & Tissue Kit (BioMag Scientific Inc., China).

cfDNA and PBMCs DNA were drawn with Circulating DNA Extraction kit (Shenzhen Apostle-Sustech Ltd., China) from plasma and PBMC according to manufacturer's instructions.

The isolated DNA was purified, eluted and then quantified by using the QubitTM dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). The sequencing libraries were constructed using the NEBNext® Enzymatic Methyl-seq Conversion kit. The EM-converted libraries were hybridized with a customized capture probe panel to enrich for DNA regions. Following the hybridization and purification, the quality of DNA was assessed using QubitTM again. Only the DNA whose concentration was ⩾5ng/mL and abundance was ⩾15ng was then PCR amplified. Finally, high-quality libraries were sequenced on the NovaSeq 6000 platform (Illumina, USA).

Reduced Representation Bisulfite Sequencing (RRBS) library preparation and sequencing

1.Genomic DNA Extraction and Quality Control

Genomic DNA from blood and animal tissues was extracted using the Magen-HiPure Universal DNA Kit (Magen, Guangzhou, China) or Xingchun Tissue & Whole Blood DNA-2230R.

All extraction experiments were strictly performed according to the kit instructions. The extracted DNA was quantified accurately using Qubit (Invitrogen, Carlsbad, California, USA). DNA with a concentration ⩾25 ng/μl and a total amount ⩾500ng was suitable for downstream library construction experiments.

2. Library Construction and Quality Control:

Library preparation was performed using the Acegen Rapid RRBS-seq Kit AG0311 (Acegen, Shenzhen, China) according to the library construction SOP. After adding lambda DNA to the gDNA, first, an enzymatic digestion reaction was carried out using NEB's MspI (for the plant version, MspI and DpnII are used); then, end repair, adapter ligation, purification, and fragment selection were performed; next, bisulfite conversion and purification were performed using the Zymo Lightning Conversion Reagent Kit (Zymo Research, Irvine, CA, USA); finally, index PCR amplification was carried out. Library concentration was measured using Qubit (Invitrogen, Carlsbad, California, USA) reagents, and fragment size was assessed using the Qsep400 instrument system. Qualified libraries were sequenced on the Illumina NovaSeq X Plus platform using the PE150 sequencing mode.

Probe-capture methylome-seq library preparation and sequencing

1.Genomic DNA Extraction and Quality Control

Genomic DNA from blood and animal tissues was extracted using the Magen-HiPure Universal DNA Kit (Magen, Guangzhou, China) or Xingchun Tissue & Whole Blood DNA-2230R. cfDNA was extracted from plasma, cerebrospinal fluid, and other sources using the Apostle MiniMaxTM High Efficiency cfDNA Isolation Kit (Apostle, USA). For FFPE samples, DNA extraction was carried out using the QIAamp® DNA FFPE Advanced Handbook (QIAGEN, Germany). All extraction experiments were strictly performed

according to the kit instructions. The extracted DNA was quantified accurately using Qubit (Invitrogen, Carlsbad, California, USA). DNA with a concentration ⩾25 ng/μl and a total amount ⩾500 ng was suitable for downstream library construction experiments.

2. Library Construction and Quality Control:

Library preparation was carried out using the Acegen Bisulfite-seq Kit BS0311 (Acegen, Shenzhen, China) following the library SOP manual. gDNA, along with quality control materials, was fragmented (if cfDNA, no fragmentation is required; the already fragmented quality control material is directly added) and subjected to bisulfite conversion using the Zymo Lightning Conversion Reagent kit (Zymo Research, Irvine, CA, USA) for single-strand denaturation. 3' adapter ligation and extension, 5' sequencing adapter ligation, product purification, and fragment selection were performed. The ligated products were PCR-amplified using index primers matched to the kit and then captured by the target panel using the Acegen Methylation Rapid Library Prep Kit (Agitech, Beijing, China). Finally, the library was purified and size-selected to obtain the final product. Library concentration was measured using Qubit (Invitrogen, Carlsbad, California, USA) reagents, and fragment size was analyzed using the Qsep400 system. Qualified libraries were sequenced on the Illumina Novaseq X Plus platform, using the PE150 sequencing mode.

Methylome-seq data preprocessing

Methylome-seq libraries were generated for cfDNA samples and were sequenced with 150 bp paired-end reads on Illumina NovaSeq 6000 platform (Illumina, USA). Samples passing the quality control criteria (coverage higher than 15x and bisulfite conversion rate higher than 99.0%) were used in this study. We performed 4 steps to preprocess the methylome-seq data.

In step 1, the adapter sequence was removed and low-quality base was trimmed. Trimmomatic (Version 0.36) was used to trim the default Illumina adapters from the sequencing reads and perform a sliding window trimming, cutting once the average quality within the window falls below a threshold (using the options as --phred33 ILLUMINACLIP:TruSeq3-PE.fa:2:30:7:1:true MINLEN:36 TRAILING:3 SLIDINGWINDOW:4:15). The software fastp (v0.21.0) was used to visualize quality control results and make a summary HTML reports.

In step 2, we performed paired-end sequencing reads merging, alignment and methylation calling. We first used pear (https://github.com/tseemann/PEAR) to merge read1 and read2 to form one fragment with default parameters. We then used bsmap (v2.73) to align the merged reads to the reference genome hg19 (GRCh37). And then samtools (version 1.19) was used to sort mapped reads with mapping locations. Bsmap's methylation extractor 'metratio.py' was then used to call methylation of each cytosine in the mapped reads.

In step 3, the chromosome-wise sequence alignment statistics and whole-methylome methylation

statistics were summarized from the individual read information obtained in Step 2. Summary statistics of the methylome-seq data including quality control, mapping rate and bisulfite conversion rate are available in Supplementary Table 2.

In step 4, tissue and PBMCs samples were sequenced and processed in the same manner as methylome-seq data.

Detection of read-based methylation markers

Conventional methods for methylation marker discovery rely on population-average methylation values, the β-value, defined as the number of methylated alleles out of all alleles mapped to a CpG site in a DNA sample. Such β-value based approaches, however, are not sensitive to tumor signals if the tumor fraction is low in the cfDNA samples. Li, W. et al[47] previously proposed the concept of α-value, defined as the percent of methylated CpGs out of all CpG sites in a sequencing read. Here we defined a two-step framework for using read based "methylated haplotype value" (or "MH-Value") to robustly identify markers in cfDNA samples controlled by GC-specific methylated markers in tissue samples and PBMCs samples.

Firstly, we used metilene (version 2.8) to detected cancer-specific DMRs with CpGs β-value comparing tumor tissue and adjacent normal tissue. We only selected tissue-based DMRs that met the following criteria: (1) contained >5 CpGs, and the distance between two adjacent cytosines was within 100 bp, (2) had a depth of ⩾10X, and (3) whose β-values significantly differed between case and control tissue samples (absolute β-value ⩾0.2 and q value ⩽0.01). The background noise within the potential DMRs was further filtered using 50 PBMC samples. DMR with β-values ⩽0.05 in PBMC samples was retained as a candidate marker for next step.

Secondly, we calculated MH-Value for each candidate markers in all cfDNA samples, which is the normalized fraction of methylated reads out of all mapped reads at this region. Mapped reads in cfDNA samples were extracted from alignment BAM files, and methylated read was defined as the mapped read contained two adjacent methylated CpG sites at least.

To quantitatively assess the methylation levels across different simulated concentration gradients, we simulated five different methylation concentration gradients (0%, 0.1%, 1%, 10%, and 100%) and calculated the methylated haplotype value under each gradient according to our defined approach. The specific procedures were as follows:

(1) After read alignment, tools such as wgbs_tools were used to extract methylation information for each aligned read from the BAM files, generating PAT-format files. (2) From the PAT files, the total number of aligned reads and the number of methylated reads were quantified. (3) The methylated haplotype value was then calculated as: methylated haplotype value = number of methylated reads / total

number of reads.

As shown in Supplementary Figure 2, a strict linear relationship between methylation level and dilution ratio was preserved down to 0.1%, demonstrating accurate low-concentration detection by methylome-seq. These findings indicate that the MH-Value remains capable of capturing stable cfDNA methylation features under conditions of low cfDNA concentration or low tumor burden.

CT tumor burden assessment

CT imaging for Chemotherapy Cohort patients was independently reviewed by at least 2 experienced radiologists at attending physician level or above. Therapeutic response was evaluated according to RECIST (Response Evaluation Criteria in Solid Tumors) v1.1 guidelines, with patients categorized into complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD)[48].

Statistical analysis

Least absolute shrinkage and selection operator (LASSO) logistic regression was performed using the glmnet package (R Foundation for Statistical Computing, Vienna, Austria) for feature selection and predictor construction. This method applies L1 regularization to shrink coefficients toward zero, with penalty intensity controlled by parameter λ. Optimal λ determination employed 10-fold cross-validation, selecting the value minimizing binomial deviance. Continuous variables were expressed as median (IQR) and compared via Mann-Whitney U test, while categorical variables presented as count (%) utilized Chi-square or Fisher's exact tests. Functional annotation of differentially methylated genes was conducted via Gene Ontology (GO) enrichment analysis using DAVID (v6.8; https://david.ncifcrf.gov/). Heatmap visualization implemented tidyheatmaps (v0.8.3), and ROC analysis utilized pROC package to generate AUC with 95% CI. Optimal cutoff determination applied Youden index. All analyses used R 4.4.1.

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to patient privacy restrictions but are available from the corresponding author upon reasonable request.

Code Availability

The code used for data analysis in this study is available from the corresponding author upon reasonable request.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 82220108013), the National Natural Science Foundation of China (Grant No. 81801246), the Guangdong Provincial Science and Technology Plan (2017A020215094), and The Kelin New Star Program of Sun Yat-

sen University (No. R08011). The authors would like to thank Shenzhen Acegen Technology Co. Ltd, for guidance in data analysis and statistics and for technical guidance in RRBS library preparation. The authors thank all the patients and volunteers who participated in this study. We also acknowledge the support from the clinical staff involved in sample collection and data management.

Author Contributions

The contribution of all authors as follows:

Q.L.: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing-Original Draft Writing;

Y.L.: Conceptualization, Data Curation, Formal Analysis, Writing-Original Draft Writing

H.L.: Data Curation, Formal Analysis, Methodology, Software, Visualization

Z.W.: Investigation, Methodology, Supervision

W.T.: Methodology, Supervision

Y.H.: Resources, Supervision, Writing-Review &Editing;

J.Z.: Resources, Supervision, Conceptualization;

G.L.: Funding Acquisition, Resources, Supervision, Conceptualization, Writing-Review & Editing

Competing Interests

The authors declare no competing financial or non-financial interests.

Reference

1 Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA Cancer J Clin* **71**, 7-33, doi:10.3322/caac.21654 (2021).

2 Culp, M. B., Soerjomataram, I., Efstathiou, J. A., Bray, F. & Jemal, A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *Eur Urol* **77**, 38-52, doi:10.1016/j.eururo.2019.08.005 (2020).

3 Pasechnikov, V., Chukov, S., Fedorov, E., Kikuste, I. & Leja, M. Gastric cancer: prevention, screening and early diagnosis. *World J Gastroenterol* **20**, 13842-13862, doi:10.3748/wjg.v20.i38.13842 (2014).

4 Guan, W. L., He, Y. & Xu, R. H. Gastric cancer treatment: recent progress and future perspectives. *J Hematol Oncol* **16**, 57, doi:10.1186/s13045-023-01451-3 (2023).

5 Deng, J., Zhang, W., Xu, M. & Zhou, J. Imaging advances in efficacy assessment of gastric cancer neoadjuvant chemotherapy. *Abdom Radiol (NY)* **48**, 3661-3676, doi:10.1007/s00261-023-04046-1 (2023).

6 Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C. & Lordick, F. Gastric cancer. *Lancet* **396**, 635-648, doi:10.1016/s0140-6736(20)31288-5 (2020).

7 Stover, D. G. *et al.* Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. *J Clin Oncol* **36**, 543-553, doi:10.1200/jco.2017.76.0033 (2018).

8 Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224, doi:10.1126/scitranslmed.3007094 (2014).

9 Corcoran, R. B. & Chabner, B. A. Application of Cell-free DNA Analysis to Cancer Treatment. *N Engl J Med* **379**, 1754-1765, doi:10.1056/NEJMra1706174 (2018).

10 Jiang, P. & Lo, Y. M. D. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. *Trends Genet* **32**, 360-371, doi:10.1016/j.tig.2016.03.009 (2016).

11 Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and medicine. *Lancet* **392**, 777-786, doi:10.1016/s0140-6736(18)31268-6 (2018).

12 Skvortsova, K., Stirzaker, C. & Taberlay, P. The DNA methylation landscape in cancer. *Essays Biochem* **63**, 797-811, doi:10.1042/ebc20190037 (2019).

13 Matsusaka, K., Funata, S., Fukayama, M. & Kaneda, A. DNA methylation in gastric cancer, related to Helicobacter pylori and Epstein-Barr virus. *World J Gastroenterol* **20**, 3916-3926,

doi:10.3748/wjg.v20.i14.3916 (2014).
14 Leal, A., Sidransky, D. & Brait, M. Tissue and Cell-Free DNA-Based Epigenomic Approaches for Cancer Detection. *Clin Chem* **66**, 105-116, doi:10.1373/clinchem.2019.303594 (2020).
15 Liu, J. *et al.* Genome-wide cell-free DNA methylation analyses improve accuracy of non-invasive diagnostic imaging for early-stage breast cancer. *Mol Cancer* **20**, 36, doi:10.1186/s12943-021-01330-w (2021).
16 Zhao, G. *et al.* Circulating cell-free DNA methylation-based multi-omics analysis allows early diagnosis of pancreatic ductal adenocarcinoma. *Mol Oncol* **18**, 2801-2813, doi:10.1002/1878-0261.13643 (2024).
17 Su, C. *et al.* Application of cfDNA methylation in cancer prognostic assessment: Progress and challenges (Review). *Int J Oncol* **67**, doi:10.3892/ijo.2025.5804 (2025).
18 Katsurano, M. *et al.* Early-stage formation of an epigenetic field defect in a mouse colitis model, and non-essential roles of T- and B-cells in DNA methylation induction. *Oncogene* **31**, 342-351, doi:10.1038/onc.2011.241 (2012).
19 Ren, J. *et al.* Genome-Scale Methylation Analysis of Circulating Cell-Free DNA in Gastric Cancer Patients. *Clin Chem* **68**, 354-364, doi:10.1093/clinchem/hvab204 (2022).
20 Han, Q. *et al.* Association between the methylation of the STAT1 and SOCS3 in peripheral blood and gastric cancer. *J Gastroenterol Hepatol* **35**, 1347-1354, doi:10.1111/jgh.15021 (2020).
21 Fattahi, S. *et al.* Infection-associated epigenetic alterations in gastric cancer: New insight in cancer therapy. *J Cell Physiol* **233**, 9261-9270, doi:10.1002/jcp.27030 (2018).
22 Stirzaker, C., Taberlay, P. C., Statham, A. L. & Clark, S. J. Mining cancer methylomes: prospects and challenges. *Trends Genet* **30**, 75-84, doi:10.1016/j.tig.2013.11.004 (2014).
23 Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251-263, doi:10.1016/j.cell.2012.06.024 (2012).
24 Wasserkort, R. *et al.* Aberrant septin 9 DNA methylation in colorectal cancer is restricted to a single CpG island. *BMC Cancer* **13**, 398, doi:10.1186/1471-2407-13-398 (2013).
25 Sun, J., Zheng, M. Y., Li, Y. W. & Zhang, S. W. Structure and function of Septin 9 and its role in human malignant tumors. *World J Gastrointest Oncol* **12**, 619-631, doi:10.4251/wjgo.v12.i6.619 (2020).
26 Nie, Y. *et al.* Combining methylated SEPTIN9 and RNF180 plasma markers for diagnosis and early detection of gastric cancer. *Cancer Commun (Lond)* **43**, 1275-1279, doi:10.1002/cac2.12478 (2023).
27 Zhao, L., Li, M., Zhang, S. & Liu, Y. Plasma-Methylated SEPT9 for the Noninvasive Diagnosis of Gastric Cancer. *J Clin Med* **11**, doi:10.3390/jcm11216399 (2022).
28 Alix-Panabières, C. & Pantel, K. Liquid Biopsy: From Discovery to Clinical Application. *Cancer Discov* **11**, 858-873, doi:10.1158/2159-8290.Cd-20-1311 (2021).
29 Nikanjam, M., Kato, S. & Kurzrock, R. Liquid biopsy: current technology and clinical applications. *J Hematol Oncol* **15**, 131, doi:10.1186/s13045-022-01351-y (2022).
30 Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926-930, doi:10.1126/science.aar3247 (2018).
31 Mukherjee, S. *et al.* NEOSCOPE: a randomised Phase II study of induction chemotherapy followed by either oxaliplatin/capecitabine or paclitaxel/carboplatin based chemoradiation as pre-operative

regimen for resectable oesophageal adenocarcinoma. *BMC Cancer* **15**, 48, doi:10.1186/s12885-015-1062-y (2015).

32 Al-Batran, S. E. *et al.* Histopathological regression after neoadjuvant docetaxel, oxaliplatin, fluorouracil, and leucovorin versus epirubicin, cisplatin, and fluorouracil or capecitabine in patients with resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4-AIO): results from the phase 2 part of a multicentre, open-label, randomised phase 2/3 trial. *Lancet Oncol* **17**, 1697-1708, doi:10.1016/s1470-2045(16)30531-9 (2016).

33 Elimova, E. *et al.* It Is Time to Stop Using Epirubicin to Treat Any Patient With Gastroesophageal Adenocarcinoma. *J Clin Oncol* **35**, 475-477, doi:10.1200/jco.2016.69.7276 (2017).

34 Badgwell, B., Das, P. & Ajani, J. Treatment of localized gastric and gastroesophageal adenocarcinoma: the role of accurate staging and preoperative therapy. *J Hematol Oncol* **10**, 149, doi:10.1186/s13045-017-0517-9 (2017).

35 Gordevičius, J. *et al.* Cell-Free DNA Modification Dynamics in Abiraterone Acetate-Treated Prostate Cancer Patients. *Clin Cancer Res* **24**, 3317-3324, doi:10.1158/1078-0432.Ccr-18-0101 (2018).

36 Moss, J. *et al.* Circulating breast-derived DNA allows universal detection and monitoring of localized breast cancer. *Ann Oncol* **31**, 395-403, doi:10.1016/j.annonc.2019.11.014 (2020).

37 Amir, S., Wang, R., Matzkin, H., Simons, J. W. & Mabjeesh, N. J. MSF-A interacts with hypoxia-inducible factor-1alpha and augments hypoxia-inducible factor transcriptional activation to affect tumorigenicity and angiogenesis. *Cancer Res* **66**, 856-866, doi:10.1158/0008-5472.Can-05-2738 (2006).

38 Yeh, Y. T. *et al.* Matrix stiffness regulates endothelial cell proliferation through septin 9. *PLoS One* **7**, e46889, doi:10.1371/journal.pone.0046889 (2012).

39 Cortez, B. A., Rezende-Teixeira, P., Redick, S., Doxsey, S. & Machado-Santelli, G. M. Multipolar mitosis and aneuploidy after chrysotile treatment: a consequence of abscission failure and cytokinesis regression. *Oncotarget* **7**, 8979-8992, doi:10.18632/oncotarget.6924 (2016).

40 Tóth, K. *et al.* Detection of methylated SEPT9 in plasma is a reliable screening method for both left- and right-sided colon cancers. *PLoS One* **7**, e46000, doi:10.1371/journal.pone.0046000 (2012).

41 Nian, J. *et al.* Diagnostic Accuracy of Methylated SEPT9 for Blood-based Colorectal Cancer Detection: A Systematic Review and Meta-Analysis. *Clin Transl Gastroenterol* **8**, e216, doi:10.1038/ctg.2016.66 (2017).

42 Sun, J. *et al.* The role of (m)SEPT9 in screening, diagnosis, and recurrence monitoring of colorectal cancer. *BMC Cancer* **19**, 450, doi:10.1186/s12885-019-5663-8 (2019).

43 Yang, X. *et al.* Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26**, 577-590, doi:10.1016/j.ccr.2014.07.028 (2014).

44 Galardi, F. *et al.* Cell-Free DNA-Methylation-Based Methods and Applications in Oncology. *Biomolecules* **10**, doi:10.3390/biom10121677 (2020).

45 Li, Y., Song, L., Gong, Y. & He, B. Detection of colorectal cancer by DNA methylation biomarker SEPT9: past, present and future. *Biomark Med* **8**, 755-769, doi:10.2217/bmm.14.8 (2014).

46 Li, T. *et al.* MicroRNA-92a-1-5p increases CDX2 by targeting FOXD1 in bile acids-induced gastric intestinal metaplasia. *Gut* **68**, 1751-1763, doi:10.1136/gutjnl-2017-315318 (2019).

47 Li, W. *et al.* CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. *Nucleic Acids Res* **46**, e89, doi:10.1093/nar/gky423 (2018).

48 Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-247, doi:10.1016/j.ejca.2008.10.026 (2009).