Am J Cancer Res 2024;14(2):744-761
www.ajcr.us /ISSN:2156-6976/ajcr0152310

Original Article

Cell-free DNA methylation profiles enable early detection of colorectal and gastric cancer

Xiaotian Lei[1*], Dongxun Zhou[2*], Ying Wen[3*], Weihong Sha[4*], Juan Ma[4,5*], Xixiang Tu[3], Kewei Zhai[6], Caixia Li[7], Hong Wang[3], Jinsheng Tao[3], Zhiwei Chen[3,8], Weimei Ruan[3], Jian-Bing Fan[3,8,9], Bin Wang[10], Chunhui Cui[1]

[1]Department of Surgery, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China; [2]Department of Endoscopy and Gastroenterology, Eastern Hepatobiliary Hospital, Naval Medical University, 225 Changhai Road, Shanghai, China; [3]AnchorDx Medical Co., Ltd., Guangzhou, Guangdong, China; [4]Guangdong Provincial People's Hospital, Guangzhou, Guangdong, China; [5]Diagnosis and Treatment Center of High Altitude Digestive Disease, Xining Second People's Hospital, Xining, Qinghai, China; [6]The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, Henan, China; [7]Jiyuan Second People's Hospital, Jiyuan, Henan, China; [8]AnchorDx, Inc., Fremont, CA, The United States; [9]Southern Medical University, Guangzhou, Guangdong, China; [10]Department of Oncology, Changhai Hospital, Naval Medical University, 168 Changhai Road, Shanghai, China. [*]Equal contributors.

Received July 19, 2023; Accepted December 7, 2023; Epub February 15, 2024; Published February 28, 2024

Abstract: Colorectal cancer (CRC) and gastric cancer (GC) rank the top five common and lethal cancers worldwide. Early detection can significantly reduce the mortality of CRC and GC. However, current clinical screening methods including invasive endoscopic techniques and noninvasive fecal occult blood test screening tests/fecal immunochemical test have shown low sensitivity or unsatisfactory patient's compliance. Aberrant DNA methylation occurs frequently in tumorigenesis and cell-free DNA (cfDNA) methylation has shown the potential in multi-cancer detection. Herein, we aimed to explore the value of cfDNA methylation in the gastrointestinal cancer detection and develop a noninvasive method for CRC and GC detection. We applied targeted methylation sequencing on a total of 407 plasma samples from patients diagnosed with CRC, GC, and noncancerous gastrointestinal benign diseases (Non-Ca). By analyzing the methylation profiles of 34 CRC, 62 GC and 107 Non-Ca plasma samples in the training set (n=203), we identified 40,110 gastrointestinal cancer-specific markers and 63 tissue of origin (TOO) prediction markers. A new integrated model composed of gastrointestinal cancer detection and TOO prediction for three types of classification of CRC, GC and Non-Ca patients was further developed through logistic regression algorithm and validated in an independent validation set (n=103). The model achieved overall sensitivities of 83% and 81.3% at specificities of 81.5% and 80% for identifying gastrointestinal cancers in the test set and validation set, respectively. The detection sensitivities for GC and CRC were respectively 81.4% and 83.3% in the cohort of the test and validation sets. Among these true positive cancer samples, further TOO prediction showed accuracies of 95.8% and 95.8% for GC patients and accuracies of 86.7% and 93.3% for CRC patients, in test set and validation set, respectively. Collectively, we have identified novel cfDNA methylation biomarkers for CRC and GC detection and shown the promising potential of cfDNA as a noninvasive gastrointestinal cancer detection tool.

Keywords: Colorectal cancer, gastric cancer, cfDNA methylation, cancer early detection, liquid biopsy

Introduction

Colorectal cancer (CRC) and gastric cancer (GC) are the two most frequent gastrointestinal malignancies and both rank the top five cancers worldwide by incidence and mortality, causing an increasing health burden globally. According to the GLOBOCAN 2020 estimates [1], CRC is the third most frequently diagnosed cancer and the second-leading cause of cancer-related death while GC accounts for 5.6% of all new cancer cases and 7.7% of mortality worldwide. CRC and GC are both asymptomatic at early stage and are easily missed-diagnosis if cost-effective screening approaches are not applied. As a result, these two cancer types are mostly diagnosed at advanced stage leading to a high mortality in countries of large population and of poor compliant screening policies such as in China [2, 3]. However, both colorectal and

https://doi.org/10.62347/TPTQ3682

gastric cancer are preventable if detected early. The five-year survival rate is estimated at 10% for patient with advanced GC while it attains 90% when GC is detected at an early stage [4, 5]. Likewise, the five-year survival rate is approximately 90% in patient with early-stage CRC whereas it decreases to less than 20% if the tumor is detected late and is spread to other organs [6]. However, there are very limited effective early detection methods in the current clinic. The invasive endoscopic techniques have been used as the gold standard in diagnosis of gastrointestinal cancers but limited by poor patient compliance, high cost and invasive procedure [7-10], while the noninvasive screening methods including fecal occult blood test (FOBT), fecal immunochemical test (FIT) and serum-based ABC method lack satisfactory sensitivities and specificities limiting their clinical utility [2, 10]. Thus, it is imperative in the clinic to develop a noninvasive, affordable method with high sensitivity and specificity for early detection of CRC and GC.

Circulating tumor DNA (ctDNA), a fraction of circulating cell-free DNA (cfDNA), is generated by shedding of apoptotic and necrotic tumor cells or by active release of tumor cells into bloodstream [11, 12]. CtDNA carries genetic and epigenomic signatures of tumor cells of origin, making it a promising biomarker in developing noninvasive liquid biopsy for cancer diagnosis [13, 14]. The wide diversity of genetic mutations and their distribution across large genomic regions make it challenging to develop genetic mutation-based cancer diagnostic tests [10]. In contrary, aberrant DNA methylation as the most well-studied epigenetic alteration in humans happens early in tumorigenesis and abundantly exists in the entire cancer process [15]. Notably, the methylation patterns are consistent with the origin of specific cancers, enabling feasible trace of tissue of origin (TOO) for source-unknown cancers [16]. These advantages and the stability of ctDNA make cfDNA methylation a robust biomarker of liquid biopsy and stand out in the detection of cancer.

CfDNA methylation has been explored in early detection and diagnosis of multiple cancers including CRC and GC [10]. Methylated SEPT9 has been demonstrated as an effective biomarker for plasma-based CRC detection and approved by the US Food and Drug Administration as the first blood-based IVD assay for CRC detection, of which the clinical significance in CRC diagnosis has been validated by a number of independent studies [3]. Apart from SEPT9, a number of other methylation markers such as APC, ALX4, BCAT1, CDH1, C9orf50, HLTF, HPP1, IKZF1, ITGA4, NGFR, PCDH10, RASSF1A, SDC2, SPG20, TMEFF2 and VIM have been reported for blood-based CRC detection [3, 9]. Additionally, previous studies have explored individual cfDNA methylation markers such as *Reprimo*, RNF180, DAPK1, GSTP1, SFRP2, SLC19A3 and ZICK1 for blood-based GC detection although most of them have not yet been validated by clinic [3]. Moreover, Tang *et al* recently reported that a panel of 153 cfDNA methylation markers is effective to detect GC in blood, with a sensitivity of 64% at a specificity of 93% [17]. Although these studies have identified some specific DNA methylation biomarkers of CRC or GC, the potential value of cfDNA methylation markers for early detection and diagnosis of gastrointestinal cancers and especially for identification of tissue of origin remains to be further investigated.

In this study, we focused on early detection of the two most common gastrointestinal cancers with high incidence and mortality, as early detection can result in a great prognosis and survival. However, current efficient screening tools were lacking. We aimed to explore the potential value of cfDNA methylation in gastrointestinal cancer detection and develop an effective blood-based noninvasive diagnostic method to assist diagnosis of gastrointestinal cancers and further discrimination of CRC and GC in the clinic where patients might show digestive discomfort. Herein, we analyzed the cfDNA methylation profiles of plasma samples from patients diagnosed with CRC, GC, or non-cancerous gastrointestinal benign diseases (Non-Ca) by high-throughput targeted DNA methylation sequencing. We further established and validated a diagnostic model for gastrointestinal cancers detection and a TOO prediction model for discriminating between CRC and GC. Our results identified novel cfDNA methylation biomarkers for CRC and GC detection and showed cfDNA methylation profile as a promising and powerful tool for noninvasive gastrointestinal cancer detection.

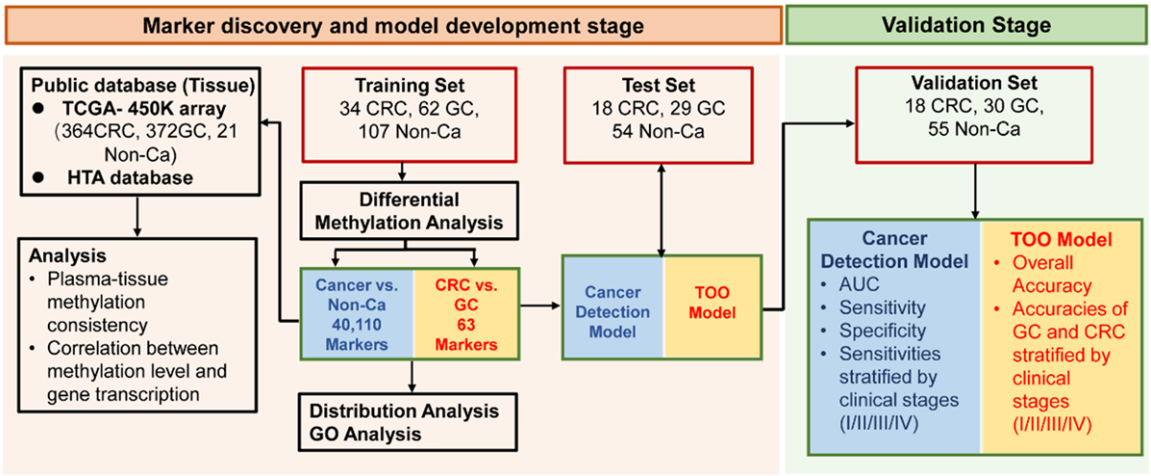


Figure 1. Schematic workflow of the study design. Abbreviations: CRC, colorectal cancer; GC, gastric cancer; Non-Ca, noncancerous gastrointestinal benign disease; TOO, tissue of origin; GO, Gene Ontology; TCGA, The Cancer Genome Atlas; HPA, The Human Protein Atlas; AUC, area under ROC curve.

Material and methods

Participants and study design

The workflow of the study is shown in **Figure 1**. Plasma samples were prospectively collected from 407 participants who showed discomfort in digestive tract in the clinic at multiple hospital centers, including Zhujiang Hospital, Guangdong Provincial People's Hospital and Xining Second People's Hospital. Out of these participants, 70 were later pathologically confirmed to have CRC, 121 were confirmed to have GC, and 216 participants served as Non-Ca control. All 216 Non-Ca patients had undergone either gastroscopy or colonoscopy. Among them, 57 (26.4%) patients had both gastroscopy and colonoscopy, while 86 (39.8%) patients had only gastroscopy, and 73 (33.8%) patients had only colonoscopy, respectively. These Non-Ca controls included patients diagnosed with chronic/astrophic gastritis, intestinal metaplasia, intestinal dysplasia and colorectal adenoma and polyps (**Table 1**). The patients with CRC, GC, or Non-Ca were randomly divided into training set, test set and a validation set at a ratio of 2:1:1 by a bioinformatician who was blind to the methylation sequencing results, in which the training set contained 34 CRC, 62 GC, and 107 Non-Ca and the test set contained 18 CRC, 29 GC and 54 Non-Ca while there were 18 CRC, 30 GC and 55 Non-Ca in the validation set. The clinical features of participants are provided in **Table 1**.

Plasma sample collection and cfDNA isolation

A total of 10 ml of blood was drawn from each participant and promptly stored in Streck cell-free DNA BCT tube (Streck, catalog 218962), of which the unique preservative and specialized chemistry prevents the cell lysis and release of genomic DNA and stabilizes cell-free DNA for up to 14 days at 6°C to 37°C enabling the sample integrity and high-quality isolation of cfDNA, as described by manufactory. Then the blood samples were stored and transported within 4 days at room temperature as following the instructions of the collection tube. Once arrived at laboratory, the plasma was separated immediately from the whole blood according to the standard protocol described previously [18] and stored at -80°C before further usages. cfDNA was extracted using the MagMAX Cell-Free DNA Isolation Kit (Thermo Fisher Scientific, catalog A29319) according to the manufactory's manual. After DNA extraction, Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, catalog Q32854) was used to measure the concentration of the extracted cfDNA and the Agilent High Sensitivity DNA Kit (Agilent, catalog 5067-4626) was used for examination of cfDNA quality and checking if any gDNA contamination using Agilent 2100 bioanalyzer. CfDNA with a

Table 1. Participant demographics and baseline characteristics

Characteristics	Marker discovery and model development cohort		Validation cohort
	Training set, n=203	Test set, n=101	n=103
Non-Ca	n=107 (%)	n=54 (%)	n=55 (%)
Age			
Mean (range)	53.1 (27-73)	52.2 (31-78)	53.5 (41-78)
Gender			
Male	35 (32.71%)	15 (27.78%)	18 (32.73%)
Female	72 (67.29%)	39 (72.22%)	36 (65.45%)
NA			1 (1.82%)
Premalignant lesions			
Atrophic gastritis	30 (28.0%)	15 (27.8%)	15 (27.3%)
Intestinal metaplasia	15 (14.0%)	7 (13.0%)	8 (14.5%)
Intraepithelial neoplasia/dysplasia	2 (1.9%)	1 (1.19%)	/
Colorectal adenoma	18 (16.8%)	9 (16.7%)	8 (14.5%)
CRC	n=34 (%)	n=18 (%)	n=18 (%)
Age			
Mean (range)	53.2 (38-77)	52.5 (25-76)	55.7 (32-89)
Gender			
Male	12 (35.29%)	11 (61.11%)	5 (27.78%)
Female	22 (64.71%)	7 (38.89%)	13 (72.22%)
Clinical stage			
I	5 (14.71%)	2 (11.11%)	3 (16.67%)
II	11 (32.35%)	6 (33.33%)	8 (44.44%)
III	12 (35.29%)	6 (33.33%)	4 (22.22%)
IV	6 (17.65%)	4 (22.22%)	3 (16.67%)
GC	n=62 (%)	n=29 (%)	n=30 (%)
Age			
Mean (range)	54.5 (27-78)	55.4 (25-78)	55.3 (29-73)
Gender			
Male	28 (45.28%)	14 (48.28%)	12 (40.00%)
Female	34 (54.84%)	15 (51.72%)	18 (60.00%)
Clinical stage			
I	11 (17.74%)	7 (24.14%)	4 (13.33%)
II	9 (14.52%)	5 (17.24%)	6 (20.00%)
III	30 (48.39%)	9 (31.03%)	17 (56.67%)
IV	10 (16.13%)	7 (24.14%)	2 (6.67%)
Undefined	2 (3.23%)	1 (3.45%)	1 (3.33%)

Non-Ca, noncancerous gastrointestinal benign diseases; CRC, colorectal cancer; GC, gastric cancer. Percentages were calculated as the proportion of the subgroup patients/individuals to the number of Non-Ca, CRC or GC participants in the training, test, or validation set.

yield greater than 3 ng and without excessive genomic DNA contamination was subjected to library construction.

DNA methylation library preparation and targeted sequencing

10 ng of cfDNA from each of the samples was used for DNA methylation library preparation and targeted sequencing as described previously [19]. Briefly, 10 ng extracted cfDNA was bisulfite-treated and purified using the Zymo Lightning Conversion Reagent (Zymo Research, catalog D5031) according to the manufacturer's protocol. The bisulfite-converted DNA were then used to construct the pre-libraries using the AnchorDx EpiVisio Methylation Library Prep Kit (AnchorDx, catalog A0UX00019) and subse-

quently amplified with the AnchorDx EpiVisio Indexing PCR Kit (AnchorDx, catalog A2DX-00025) according to the manufacture's manuals. The amplified prehybridization libraries were subsequently purified using the IPB1 Magnetic Beads and the concentration was measured using the Qubit dsDNA HS Assay Kit. These amplified libraries containing more than 400 ng DNA were considered qualified for the following target enrichment. Next, target enrichment was performed using the AnchorDx EpiVisio Target Enrichment Kit (AnchorDx, catalog A0UX00031) and with a proprietary custom-made pan-cancer methylation panel (AnchorDx, catalog B0UX00040) consisted of 24654 pre-selected regions enriched for cancer-specific methylation. The enriched libraries were further amplified with P5 and P7 indexing primers for Illumina using KAPA HiFi HotStart Ready Mix (KAPA Biosystems, catalog KK2602), and PCR products were subsequently purified with Agencourt AMPure XP Magnetic Beads (Beckman Coulter, catalog A63882). The resulting libraries were sequenced on the NovaSeq 6000 System (Illumina Inc.).

Sequencing data analysis

FASTQ files were generated from raw BCL data using bcl2fastq version 2.19.1. The sequencing data was processed as described previously [19], sequencing adapters and 3' low quality bases were trimmed from raw sequencing reads using Trim Galore version 0.4.1 (https://github.com/FelixKrueger/TrimGalore) and then the trimmed sequences were aligned to bisulfite treatment converted (non CpG cytosine converted to thymidine, guanidine converted to adenine) human genome (hg19) using Bismark version 0.15.0 (Bowtie2 as the default aligner behind Bismark). Aligned reads were evaluated by Picard version 2.5.0 and the PCR duplicates labelled by Picard were removed from further downstream analysis. Methylation metrics including genomic location, mapped reads of methylated and unmethylated bases of all CpG sites were extracted by MethylDackel version 0.3.0. For each sample and each CpG site, we calculated a β-value, defined as the ratio of the aligned methylated bases to the total aligned bases, which was used for subsequent predictive modeling.

Identification of differential methylation signatures

Differential methylation analysis was performed on the training set of the model development cohort using R package DSS, version 2.14.0 [19]. To identified gastrointestinal cancer-specific methylated biomarkers, methylation profiles of patients diagnosed with CRC or GC were compared with that of Non-Ca patients. A differential methylation locus (DML) was defined as false discovery rate (FDR) < 0.01, mean β-value difference ≥ 0.02 or fold change ≥ 1.5. Hypermethylation was defined as a higher mean β-value in CRC and GC group than in Non-Ca group while markers showed lower mean β-value in CRC and GC group was recognized as hypomethylated. To identify TOO prediction methylation markers of GC and CRC, the methylation profiles of the CRC group were compared to GC group, and CpG sites with FDR < 0.05, mean β-value difference ≥ 0.02 or fold change ≥ 1.5 were defined as a DML for TOO prediction.

Gene Ontology (GO) enrichment analysis of differential methylation markers

The gene annotation of differentially methylated CpG sites were performed by using ANNOVAR. GO enrichment analyses were applied to determine the biological roles of the differentially methylated genes by using R package clusterProfiler version 3.10.1. The GO terms were considered significantly enriched with BH-adjusted *p*-values < 0.05.

Model development for gastrointestinal cancer detection and tissue of origin prediction

The model included a two-step binary classifications of gastrointestinal cancer detection and TOO prediction, in which a sample was firstly classified as gastrointestinal cancer or non-cancer and if determined as gastrointestinal cancer, the sample was further classified as CRC or GC. All DMLs identified as gastrointestinal cancer-specific markers were used to develop the gastrointestinal cancer detection model while the identified TOO prediction methylation signatures were used for TOO prediction model development. Logistic regression was fitted with L2 regularization parameter using sklearn package to build the cancer detection model

and the TOO prediction mode in the training set, resulting in a cancer risk score and GC prediction score, respectively. L2 regularization was achieved by adding a penalty term to the loss function to control the complexity of the model and prevent overfitting. The formula for calculating cancer risk score or GC prediction score were similar and equaled to 1/(1 + exp^(-x)), where $x = w_1*m_1 + w_2*m_2 + ... + w_i*m_i +$ intercept, where i represents the number of the gastrointestinal cancer-specific marker or TOO marker and w_i is the coefficient of the $marker_i$ obtained during the logistic regression modeling, m_i is the β-value of the $marker_i$. The coefficients and intercepts of the gastrointestinal cancer-specific markers and TOO markers were listed in the supplementary file, respectively. The cutoff of the gastrointestinal cancer detection model was set at the point by fixing the specificity at 80% in the test set. The true positive cancer samples identified by the gastrointestinal cancer detection model were further used for TOO model development and the cutoff value was determined by Youden's index. The final models built from the training set were used for testing and validation.

Statistical analysis

Logistic regression was fitted to build the model for gastrointestinal cancer detection and TOO prediction, resulting in a cancer risk score and GC prediction score, respectively. The receiver operating characteristic (ROC) was generated based the cancer risk score and GC prediction score. Scaled β-value, calculated as the z-score of β-value, was used to indicate methylation level of each sample. The scaled β-values of a marker from samples of different clinical categories were presented as violin plot with median and the differences between two groups were analyzed with Mann-Whitney-Wilcoxon's test. The cancer risk score and GC prediction score distribution of different clinical categories was presented as box plots with median and the interquartile range marks. Differences between 2 groups were analyzed with the unpaired Student's t test and one-way ANOVA followed by Dunnett's multiple comparisons tests when more than 2 groups were compared. The sensitivity, specificity of gastrointestinal cancer detection model and accuracy of TOO model were presented as univariate values with 95% confident intervals (CIs). All statistical analysis and data visualizations were carried out in R software (version 3.5.1) with R packages and GraphPad Prism9.

Results

CfDNA methylation profiling of gastrointestinal cancer signatures

To identify gastrointestinal cancer-specific methylation markers, we conducted a comparison of the methylation profiles between cancerous (CRC and GC) plasma samples and Non-Ca plasma samples in the training set. A total of 40,110 differential methylation CpG sites were found (Supplementary Figure 1; FDR < 0.01, β-value difference ≥ 0.02 or fold change ≥ 1.5). Among these sites, 95.7% (38,384 CpG sites) were identified as cancer-specific hypermethylation markers, while 4.3% (1,726 CpG sites) were identified as cancer-specific hypomethylation markers for the detection of gastrointestinal cancers (Supplementary Figure 1). The methylation levels of the top 2000 markers (most significant β-value difference) were representatively depicted in the heatmap (**Figure 2A**), exhibiting different methylation patterns in plasma samples between gastrointestinal cancers and Non-Ca group. The DMLs were mainly located in the island (60.3%) and open sea (24.1%), and distributed in the gene body and promotor regions in terms of gene distribution, implying their function in gene regulatory (**Figure 2B**). Among all the identified markers, most of them were newly identified to be specific to gastrointestinal cancers, for instance, EVI2A, MRTFA, ARHGAP3, RPS6KA1 and CPNE3. In addition, the study also consistently identified well-known markers that have been reported to be associated with CRC and GC, including markers of SEPT9, ATXN1, PCDH10, MYO1G, NGFR, IKZF1 and ITGA4 for CRC detection [3, 9, 20] and DAPK1, SFRP2, DOCK10, BMP3, NDRG4, SFMBT2, ELMO1, ZNF569, SP9, EMX1, PPP2R5C, RHGEF4 and ZEB2 for GC detection [3, 17, 21].

We further compared these gastrointestinal cancer-specific methylation signatures to the publicly available database to confirm the correlations between methylation levels in plasma and tissue samples, as well as the associations between methylation and RNA expression. Based on the 450K methylation array data of The Cancer Genome Atlas (TCGA), which includ-

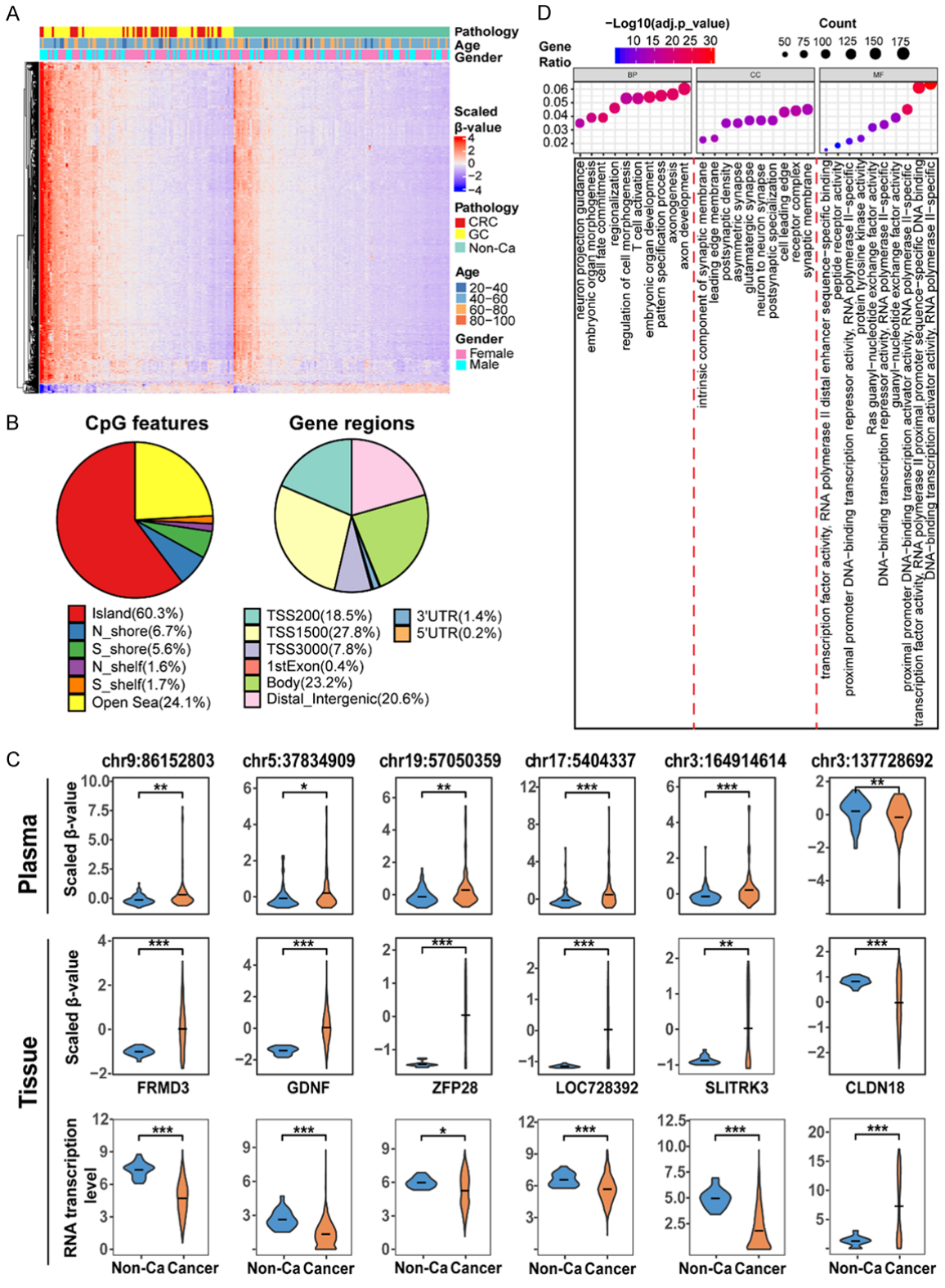


Figure 2. Discovery of cell-free DNA methylation markers for gastrointestinal cancer detection. A. Heatmap of the top 2000 methylation markers differentially methylated between plasma of gastrointestinal cancer (gastric cancer and colorectal cancer) (n=96) and noncancerous gastrointestinal benign diseases (n=107). Unsupervised hierarchical clustering was performed on markers (rows) while samples (columns) were sorted based on their pathology group. Each column represents an individual participant, and each row is a CpG marker. Methylation level in each partici-

pant is denoted by scaled β-value calculated as the z-score of β-value. B. Distribution of identified gastrointestinal cancer-specific differential methylation loci (DMLs) in the genome. C. Methylation levels of representative DMLs in plasma (n=203) and tissue (n=757), and the corresponding gene transcription level in tissue. 450K methylation array data of The Cancer Genome Atlas was used as the data source of methylation level and RNA transcription level in tissue and for the analysis. Scaled-β-value denotes the methylation level. RNAseq data was downloaded from UCSC Xena and RNA transcription level was estimated as in log2(x+1) transformed RSEM normalized count. Statistical significance was assessed using Mann-Whitney-Wilcoxon's test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. D. Gene Ontology enrichment analysis. The top 10 enriched signaling pathways were reprehensively shown.

ed 364 CRC, 372 GC and 21 cancer-free tissue samples, we found that a portion of the identified DMLs, including chr9:86152803 (FRMD3), chr5:37834909 (GDNF), chr19:57050359 (ZFP28), chr17:5404337 (LOC728392), chr3:16-4914614 (SLITRK3) and chr3:137728692 (CLDN18) showed consistent methylation difference between solid cancers and noncancerous tissues (**Figure 2C**). Notably, these corresponding genes (FRMD3, GDNF, ZFP28, LOC728392 and SLITRK3) regulated by the identified hypermethylation markers showed significant transcriptional repression in cancer group compared to noncancerous group, while CLDN18 regulated by the hypomethylated cancer-specific markers exhibited higher transcriptional level in the cancer group (**Figure 2C**), which further confirmed the role of these DMLs in mediating gene expression. In addition, we performed a GO enrichment analysis and found that the GO categories including tissue proliferation and differentiation such as embryonic organ morphogenesis and development, cell fate commitment, regulation of cell morphogenesis, and regionalization, were significantly enriched (**Figure 2D**). Additionally, transcription activator activities such as RNA polymerase II-specific DNA-binding transcription activator activity, and RNA polymerase II proximal promotor sequence-specific DNA binding transcription factor activity were also enriched (**Figure 2D**). These results suggested that the epigenetic signaling pathways regulating cell differentiation/reprogramming might play an important role in gastrointestinal tumorigenesis.

CfDNA methylation signatures for tissue of origin prediction of gastrointestinal cancers

In addition to identifying gastrointestinal cancer-specific methylation signatures, we have also identified the markers for TOO prediction. Through the comparison of methylation profiles between CRC plasma samples and GC plasma samples in the training set, a total of 63 DMLs were found (Supplementary Figure 2; FDR < 0.05, β-value difference ≥ 0.02 or fold change ≥ 1.5). The hierarchical clustering showed a clear differential methylation pattern between the GC and CRC patients (**Figure 3A**). Among these markers, 12.7% (8) of the markers were hypermethylated and 87.3% (55) of the markers were hypomethylated for GC identification (Supplementary Figure 2). Analysis of the genomic region distribution revealed that these DMLs were mainly located in the island and open sea, and a large portion were concentrated on gene body and protein-coding area (**Figure 3B**). Furthermore, we compared these methylation signatures to publicly available tissue database for to confirm the consistency of tissue-plasma methylation pattern and the heatmap showed the representative consistent methylation profiles of some markers at tissue and plasma levels (**Figure 3C**). In addition to the consistency observed between tissue and plasma methylation levels, in terms of the potential regulatory functions of these methylation markers in gene expression, some of the markers, for example, CD164L2 (chr1:27709735), KCP (chr7:128550946) and especially CPLX3 (chr15:75118958) which have been shown to be hypermethylated in CRC patients showed lower transcriptional level in colon and rectum than stomach or other tissues of origin, while CNST (chr1:246733010) that was hypomethylated in CRC patients exhibited opposite transcriptional profiles as compare the stomach or other tissues of origin (**Figure 3D**, **3E**). Their expression in solid tumors have also been analyzed (Supplementary Figure 3). These results further confirmed the tissue-specificity of these markers and implied their potential as TOO markers in discrimination of GC and CRC.

Development and validation of a gastrointestinal cancer detection model

Based on the all identified gastrointestinal cancer-specific markers, a binary diagnostic model for gastrointestinal cancer (CRC and GC) detection in plasma samples was developed using

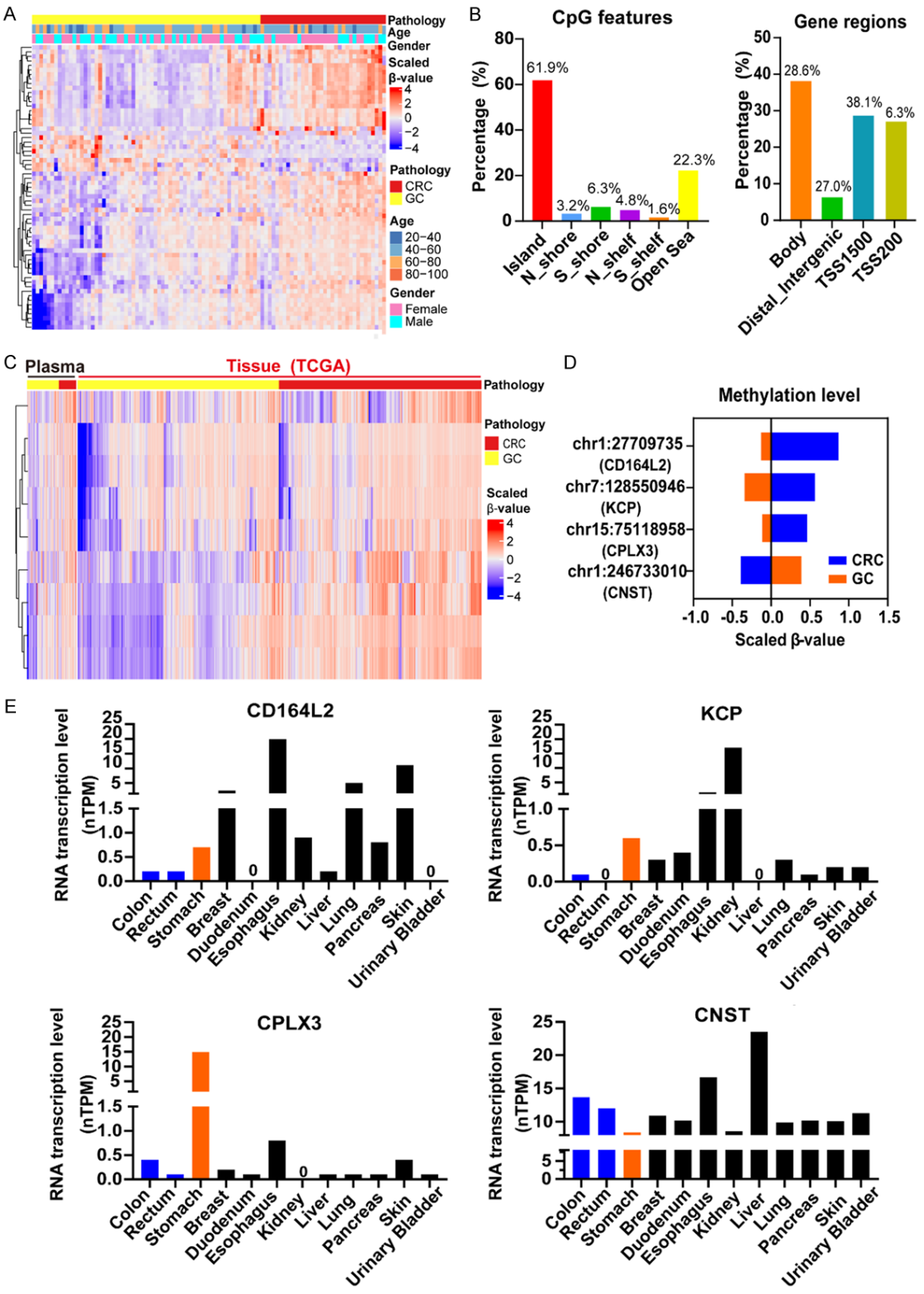


Figure 3. Methylation markers for tissue of origin (TOO) prediction of colorectal and gastric cancer. (A, C) Heatmap of methylation markers differentially methylated between colorectal cancer (CRC, n=34) and gastric cancer (GC, n=62) plasma (A), and their corresponding methylation level in tissue samples (n=736) (C). Each row represents a CpG marker and each column represents an individual participant. Unsupervised hierarchical clustering was performed

on markers (rows) while samples (columns) were sorted based on their pathology group. Methylation level in each participant is denoted by scaled β-value calculated as the z-score of β-value. The 450K methylation array data of The Cancer Genome Atlas (TCGA) was used as the data source of tissue methylation profiles (C). (B) Distribution of identified TOO prediction markers in the genome. (D) Differential methylation levels of representative TOO markers in CRC plasma (n=34) and GC plasma (n=62). (E) RNA expression level of the representative TOO marker in different tissues. RNAseq data of the corresponding genes in human tissues was retrieved from Human Protein Atlas database, and the nTPM (normalized transcripts per million) of genes was used to denote the transcription level.

logistic regression algorithm in the training set. The cut-off was locked down in the test set using this model and the diagnostic performance was also evaluated. In the test set, this gastrointestinal cancer detection model achieved an AUC (area under ROC curve) of 88.3% (95% CIs: 81.5-95%) (**Figure 4A**). According to the predicted cancer risk scores generated by the model, patients with CRC or GC showed significantly higher cancer risk scores compared to Non-Ca controls (**Figure 4C**). In summary, the gastrointestinal cancer detection model exhibited sensitivity of 83% (95% CIs: 72.3-93.6%) and specificity of 81.5% (95% CIs: 69.2-89.6%) in the test set (**Figure 4E**).

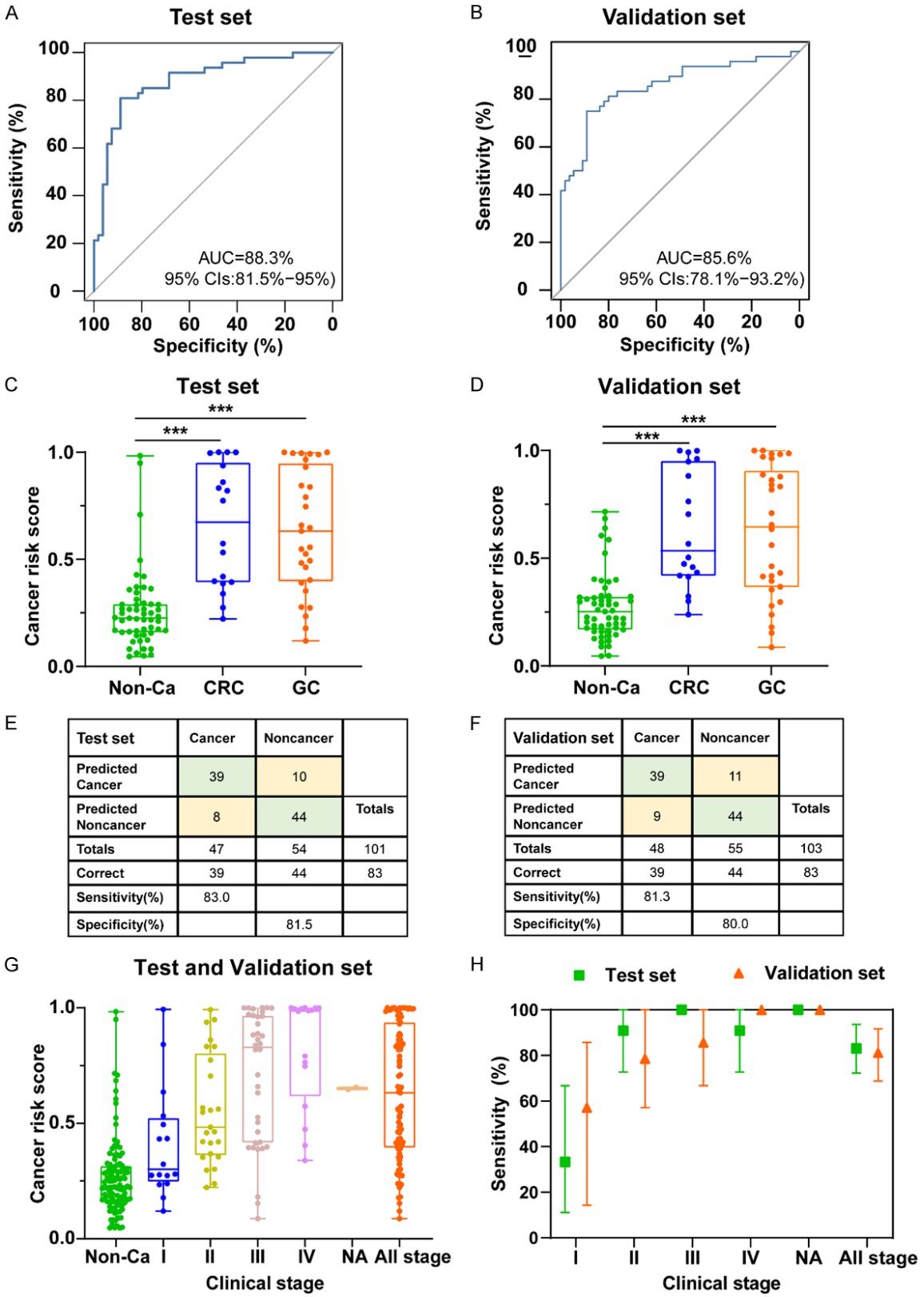
We further used an independent validation set comprised of 103 plasma samples (18 CRC, 30 GC and 55 Non-Ca) to further evaluate the performance of the gastrointestinal cancer detection model (**Figure 1**). It achieved consistent performance as compared to the test set with an AUC of 85.6% (95% CIs: 78.1-93.2%), sensitivity of 81.3% (95% CIs: 68.8-91.7%) and specificity of 80% (95% CIs: 67.6-88.4%) (**Figure 4B**, **4F**). The cancer risk score distributions were also consistent with the results in the test set, which showed significant higher scores in patients diagnosed with CRC or GC than in Non-Ca group (**Figure 4D**).

In addition, we also assessed the detection sensitivities of the model to subgroups of cancer patients with different clinical stages. The sensitivities were 33.3% (95% CIs: 11.1-66.7%), 90.9% (95% CIs: 72.7-100%), 100% (95% CIs: 100-100%) and 90.9% (95% CIs: 72.7-100%) for stage I, II, III and IV, respectively, in the test set (**Figure 4H**). Consistently, the sensitivities were 57.1% (95% CIs: 14.3-85.7%), 78.6% (95% CIs: 57.1-100%), 85.7% (95% CIs: 66.7-100%) and 100% (95% CIs: 100-100%), respectively, in the validation set (**Figure 4H**). Statistical analysis showed no significant differences in detective sensitivities among different tumor stages in the test or validation set, except for the sensitivity in detecting stage I tumor in the test set which showed a difference when compared to tumors in other stages (Supplementary Figure 4). In addition, the cancer risk scores were increased with cancer clinical stages in the cohort of the test and validation sets (**Figure 4G**). These results implied that the gastrointestinal cancer detection model can be used as a potential tool for early cancer detection.

Development and validation of tissue of origin prediction model

For patients identified by the gastrointestinal cancer (CRC and GC) detection model, we further developed a TOO prediction model using logistic regression algorithm in the training set. The cut-off was locked down in the test set using this model and the TOO prediction performance was evaluated in the test set and validation set. The model showed significantly higher GC prediction scores in GC patients compared to CRC patients in both test set and validation set (**Figure 5A**, **5B**). Furthermore, the GC prediction scores in the cohort of test and validation sets showed consistently distinguishing differences between patients with CRC and GC across different clinical stages, specifically the scores of CRC patients were significantly lower than those of GC patients in both early and advanced stages (**Figure 5C**), further suggesting the high accuracy of TOO prediction.

Furthermore, the model showed high accuracies in classifying all stages GC patients with average accuracies up to 95.8% (79.8-99.3%) in the test set and 95.8% (79.8-99.3%) in the validation set, respectively (**Figure 5D**). The average accuracies for predicting patients with CRC were 86.7% (62.1-96.3%) in the test set and 93.3% (70.2-98.8%) in the validation set (**Figure 5D**), respectively. The TOO prediction for both gastrointestinal cancers exhibited overall performance with average accuracies of 92.3% (79.7-97.3%) in the test set and 94.9% (83.1-98.6%) in the validation set, respectively (**Figure 5D**). In subgroup analysis of different clinical stages, accuracies of 85.7% and above were achieved for stage I-IV GC patients (**Figure 5D**). Additionally, the model showed accu-



Test set	Cancer	Noncancer	
Predicted Cancer	39	10	
Predicted Noncancer	8	44	Totals
Totals	47	54	101
Correct	39	44	83
Sensitivity(%)	83.0		
Specificity(%)		81.5	

Validation set	Cancer	Noncancer	
Predicted Cancer	39	11	
Predicted Noncancer	9	44	Totals
Totals	48	55	103
Correct	39	44	83
Sensitivity(%)	81.3		
Specificity(%)		80.0	

Figure 4. Development and validation of gastrointestinal cancer detection model. (A, B) Receiver operating characteristic (ROC) curves and the associated areas under curves (AUCs) of the gastrointestinal cancer detection model in the test (A) and validation (B) sets. (C, D) Cancer risk scores of patients with noncancerous gastrointestinal benign

diseases (Non-Ca) and patients with colorectal cancer and gastric cancer in the test set (C, n=101) and validation set (D, n=103). The data are shown as median scores with the corresponding the interquartile range. Statistical significance was assessed using 1-way ANOVA followed by Dunnett's tests. ***$P < 0.001$. (E, F) Confusion matrices of the gastrointestinal cancer detection model in the test set (E) and validation set (F). (G) The cancer risk scores of Non-Ca patients (n=109) and of patients with gastrointestinal cancers of stage I (n=16), stage II (n=25), stage III (n=36), stage IV (n=16), undefined stage (n=2) and all stages (n=95) in the cohort of the test and validation sets (n=204). (H) The detection sensitivities or for different stages of gastrointestinal cancers in the test set (n=101) and validation set (n=103).

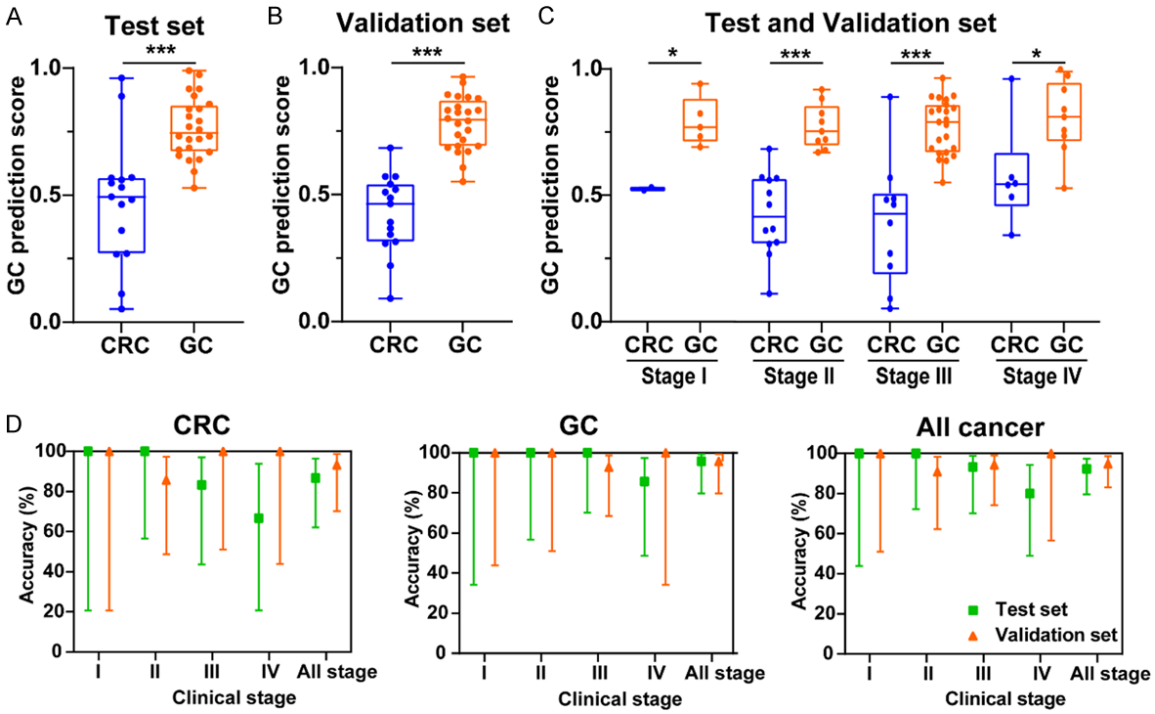


Figure 5. Development and validation of tissue of origin (TOO) prediction model for colorectal and gastric cancer. (A, B) Gastric cancer (GC) prediction scores of colorectal cancer (CRC) and GC patients in the test set (A, n=39) and validation set (B, n=39). The data are shown as median scores with the corresponding interquartile range. Statistical significance was assessed using unpaired t tests. ***$P < 0.001$. (C) GC prediction scores of patients with CRC or GC at indicated stages in the cohort of the test and validation sets (n=78); two GC patients with undefined clinic stages were excluded in the analysis. (D) Accuracies of TOO prediction in patients with CRC and GC, and the overall accuracies of gastrointestinal cancers regarding to different stages in the test set (n=39) and validation set (n=39).

racies of 80% and above for predicting stage I/II/III CRC patients (**Figure 5D**). Although there was a slight decreased in accuracies for stage IV CRC patients in the test set, no statistically significant differences were found when compared to earlier stages (Supplementary Figure 5). These results demonstrated that the TOO prediction model was an accurate classification of CRC and GC.

An integrated model for three-types classification of CRC, GC and non-gastrointestinal patients

We further combined the gastrointestinal detection model and TOO prediction model to create an integrated model for the classification of three types: CRC, GC and Non-Ca patients. We also assessed the overall performance of the classification in the cohort of the test and validation sets. The integrated model showed specificities of 80.7% (88/109) for identifying noncancerous individuals (**Figure 6A**). Out of 36 CRC patients, 30 (83.3%) were identified as having either CRC or GC. Among the 30 true positive cancer patients, 27 (90%) were accurately classified as having CRC (**Figure 6A**). Additionally, out of the 59 GC patients, 48 (81.4%) were correctly classified as having cancer, with 46 (95.8%) of them further predicted as GC (**Figure 6A**).

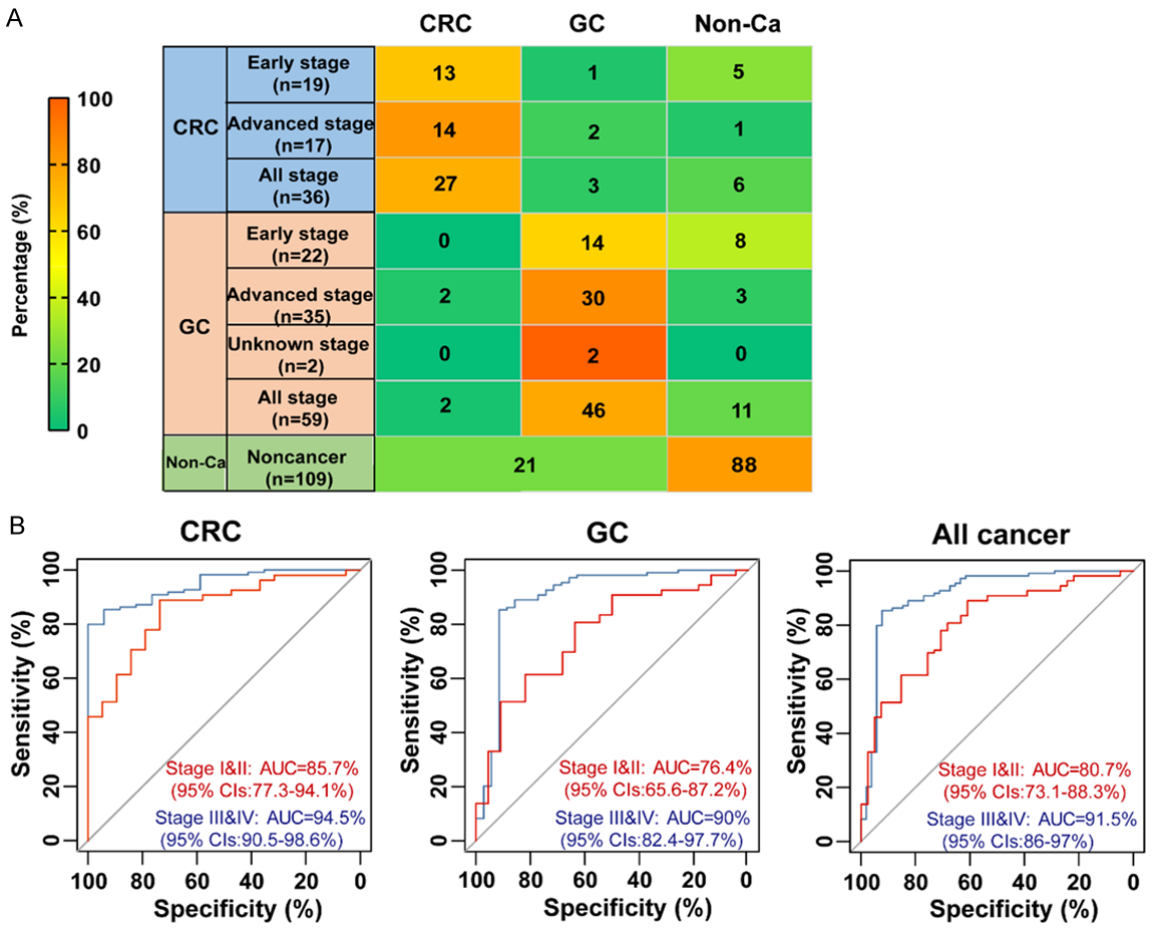


Figure 6. Performance of the integrated model in classification of colorectal cancer (CRC), gastric cancer (GC) and noncancerous gastrointestinal disease (Non-Ca). A. Summary of the prediction results in three types classification of CRC (n=36), GC (n=59) and Non-Ca (n=109) patients in the cohort of the test and validation sets (n=204). Numbers in the box with black line denotes the real cases of CRC, GC and Non-Ca patients in the denoted clinical stages. Numbers in CRC, GC and Non-Ca columns denotes the predictive cases predicted by the model as CRC, GC or Non-Ca respectively in the indicated clinical subgroups and the colors from green to red represent the percentages of the predicted CRC, GC and Non-Ca cases in the indicated clinical stages. B. The receiver operating characteristic (ROC) curves of the model in early-stage and advanced-stage CRC (n=19/17 for early/advanced stage), GC (n=22/35 for early/advanced stage), all cancer (n=41/52 for early/advanced stage), and Non-Ca patients (n=109) in the cohort of the test and validation set (n=204); Two GC patients with undefined clinical stage were excluded in the analysis.

We also evaluated the performance of the integrated model to detect CRC and GC of early (stage I&II) and advanced (stage III&IV) stages in the combined cohort of the test and validation sets. The integrated model showed comparable sensitivities of 73.7% (14/19) and 63.6% (14/22) in diagnosing early-stage CRC and GC patients as gastrointestinal cancer, respectively (**Figure 6A**). However, the sensitivities of the model distinguishing gastrointestinal cancers from benign disease increased to 94.1% (16/17) in advanced-stage CRC patients and 91.4% (32/35) in advanced-stage GC patients, respectively (**Figure 6A**). Moreover, the integrated model achieved AUCs of 85.7% (95% CI: 77.3-94.1%), 76.4% (95% CI: 65.6-87.2%) in cancer detection for early-stage CRC and GC patients and an overall AUC of 80.7% (95% CI: 73.1-88.3%) for all patients at early stages (**Figure 6B**). The model exhibited improved performance in distinguishing gastrointestinal cancer from noncancerous benign diseases as the cancer progressed to more advanced stages. It reached AUCs of 90% or higher in advanced-stage CRC, GC and all cancer patients and this improvement could be attributed to the increased release of ctDNA released as the cancer spreading in the advanced stages

(**Figure 6B**). Overall, our binary model displayed favorable predictive power in simultaneously identifying CRC and GC in high-risk population.

Discussion

Early detection of colorectal and gastric cancers can significantly reduce cancer mortality and disease management costs. However, currently, no effective and convenient diagnostic approach is for the early detection of gastrointestinal cancers in clinics. Although endoscopic techniques have been widely regarded as the golden standard for diagnosing CRC and GC, their poor patient compliance, high cost and invasive procedure limit their widespread use in clinics. Moreover, interobserver variability may introduce bias affect the diagnosis efficacy, particularly in detecting early lesions. Hence, the development of a noninvasive, affordable, and accurate method for the early detection of CRC and GC is imperative in the clinic. In this study, we aimed to explore the potential value of cfDNA methylation in gastrointestinal cancer detection and develop an effective blood-based noninvasive diagnostic method to assist the diagnosis of gastrointestinal cancers and further discrimination of CRC and GC, in the clinic where patients might show discomfort in digestive tract. Firstly, we identified 40,110 gastrointestinal cancer-specific DMLs and 63 tissue-specific DMLs by analyzing the methylation profiles of plasma samples from CRC, GC and Non-Ca patients. Furthermore, based on these methylation signatures, we developed an effective blood-based noninvasive diagnostic model comprised of gastrointestinal cancer detection and TOO prediction, capable of detecting CRC and GC simultaneously. Our study established a convenient, noninvasive, and accurate method to simultaneously detect CRC and GC and demonstrated that cfDNA methylation pattern as a promising tool in noninvasive gastrointestinal cancer detection.

Through targeted methylation sequencing using a custom-made pan-cancer panel, we identified informative plasma gastrointestinal cancer-specific markers. It is worth noting that some of the markers we identified for cancer detection have previously been reported in blood-based gastrointestinal cancer detection. For instance, the gastrointestinal cancer-specific markers we identified, such as SEPT9, PCDH10, NGFR, IKZF1, ITGA4, and MYO1G, have been well reported as methylation markers for CRC [3, 9, 20]. Moreover, some of these marker have been reported for GC detection. For example, our study identified 9 out of the 16 cfDNA methylation markers for GC detection that previously identified by Anderson *et al* based on tissue DNA methylation profiles [21], suggesting the consistency between the tissue-based markers and the plasma gastrointestinal cancer-specific markers we identified. Collectively, the consistency of our results with previous studies in gastrointestinal cancer marker discovery provided evidence for the reliability of methylation signature identification. Most importantly, besides the CRC or GC methylation markers already reported, we also identified novel signature methylation markers including EVI2A, MRTFA, ARHGAP3, RPS6KA1 and CPNE3 for the detection of gastrointestinal cancer, and markers such as DIP2C, CNST, KITLG, KCP, KIF26A and ANKRD18B for TOO prediction of CRC and GC.

GO enrichment analysis revealed that most of the cancer-specific markers we identified were involved in the regulation of transcription, as well as cell and tissue proliferation and differentiation, indicating the potential relevance of epigenetic signaling pathways with tumorigenesis and suggesting the biological feasibility of our noninvasive plasma assay for detecting cancers including CRC and GC. To be specific, the genes corresponding to some of our identified methylation markers have been reported to be relevant to tumorigenesis. For example, ATXN1, a chromatin-binding factor, has been shown to be an important regulator in tumorigenesis and cancer metastasis, of which the down-regulation of expression promotes tumor cell migration and invasion and the protein family to which also regulates extracellular matrix remodeling [22-24]. Notably, ATXN1 was found to be a component of Notch signaling pathway which is essential for the development and homeostasis of gastrointestinal epithelial cells [25, 26]. Likewise, MYO1G as an essential regulator of membrane intension in T cells has been found involved in the lymphoblastic leukemia, and the class I myosins have been shown to be important for development and metastasis of diverse cancer types, including colorectal cancer [27, 28]. Additionally, GNDF was reported to increase cell motility in colon cancer through

VEGF-VEGFR1 interaction, as well as the GNDF receptor complex component RET has also been implicated in regulating the activity of colorectal cancer cells [29-32]. However, most of these gastrointestinal cancer-specific markers are not yet uncovered their clear relationships with cancer development and tumorigenesis. Further investigations into the potential functional mechanisms of these markers might deepen our understanding of the occurrence and development of gastrointestinal cancers, and also possibly provide therapeutic targets.

Currently, besides the invasive and uncomfortable endoscopic tests used for gastrointestinal cancer diagnosis, two commonly used noninvasive screening tests for colorectal cancer worldwide are FOBT and FIT. However, these tests showed limitations of low specificities and sensitivities due to the multiple sources of fecal blood which include not only cancerous lesions but also non-cancerous sources such as polys [33]. Besides that, the inconvenience and discomfort of dealing feces decrease the patient compliance. Compared to fecal-based detection test, liquid biopsy blood-based tests exhibit prominent advantages in terms of accessibility and patients' acceptance. Several serum protein markers such as carcinoembryonic antigen (CEA), carbohydrate antigen (CA19-9) and CA72-4 have been identified and used in the clinic for the diagnosis and recurrence monitoring of gastrointestinal cancers [4, 33]. But due to their limited sensitivity and specificity, these markers are not recommended for early detection of gastrointestinal cancers and have limited use in discrimination of CRC and GC [33, 34]. In our study, we developed a blood-based binary diagnostic model that achieved sensitivities of 83% (95% CIs: 72.3-93.6%) and 81.3% (95% CIs: 68.8-91.7%) at specificities of 81.5% (95% CIs: 69.2-89.6%) and 80% (95% CIs: 67.6-88.4%) for detecting gastrointestinal cancers in the test set and independent validation set, yielding AUCs of 88.3% (95% CIs: 81.5-95%) and 85.6% (95% CIs: 78.1-93.2%), respectively. The performance was superior than the currently available non-invasive tests, including FOBT (sensitivity of 61.4%, specificity of 70.3%) [35], FIT (sensitivity of 27.6%, specificity of 94.1%) [36], CEA (sensitivity of 35.0%, specificity of 62.6%) [35], CA19-9 (sensitivity of 30%, specificity of 87%) [34] and CA72-4 (sensitivity of 40%, specificity of 95%) [34]. This model also showed high average accuracies of 92.3% (79.7-97.3%) and 94.9% (83.1-98.6%) for TOO prediction of CRC and GC in the test and validation set, respectively. The performance was comparable to currently reported multi-cancer TOO prediction [37]. In addition to the advantage of detecting CRC and GC simultaneously, our model displayed better performance in CRC detection than the currently available SEPT9-based methylation blood test which has a sensitivity of 68% at the specificity of 80% [38]. Moreover, the model could efficiently distinguish GC from CRC with an accuracy of 85.7% or higher for stage I-IV GC rather than just classify the tissue of origin for advanced GC, suggesting the robustness of the model in TOO prediction and its potential clinical application in diagnosing gastrointestinal cancers.

However, it was worth noting that the current model was tested in a clinical setting with patients showing discomfort in digestive tract and the test's potential application was primary as an assistant diagnostic tool. For the test to further be applied as a cost-effective screening tool, a future study further involving large scale validation based on high-risk population is required.

Conclusion

Collectively, we conducted comprehensive cfDNA methylation analysis and identified gastrointestinal cancer-specific methylation signatures and TOO prediction markers for CRC and GC, demonstrating that cfDNA methylation signature is a promising tool for the early detection of gastrointestinal cancers and noninvasive cancer screening.

Acknowledgements

This study is supported by the Scheme of Guangzhou Economic and Technological Development District for Leading Talents in Innovation and Entrepreneurship (grant 2017-L152), Scheme of Guangzhou for Leading Talents in Innovation and Entrepreneurship (grant 2016007), Scheme of Guangzhou for Leading Team in Innovation (grant 20190-9010010), 2020 Guangzhou Development Zone International Science and Technology Cooperation Project (grant 2020GH15), Science and Technology Planning Project of Guangzhou (grant 202206080013), 2021

Guangzhou Development Zone International Science and Technology Cooperation Project (grant 2021GH17), 2022 National Key Research and Development Program for Young Scientists (grant 2022YFC2408300), Science and Technology Program of Guangzhou (grant 202102080007), Science and Technology Program of Xining (grant 2022-M-19) and Science and Technology Program of Qinghai (grant 2023-ZJ-936M).

Patient consents were obtained from all participants.

Disclosure of conflict of interest

Ying Wen, Xixiang Tu, Hong Wang, Jinsheng Tao, Weimei Ruan, Zhiwei Chen and Jian-Bing Fan are/were employee of AnchorDx. Medical. Co., Ltd. or AnchorDx, Inc.

Address correspondence to: Chunhui Cui, Department of Surgery, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, Guangdong, China. E-mail: drcuich@163.com; Bin Wang, Department of Oncology, Changhai Hospital, Naval Medical University, 168 Changhai Road, Shanghai 200433, China. E-mail: wangbinch@smmu.edu.cn; Jian-Bing Fan and Weimei Ruan, AnchorDx Medical Co., Ltd., Guangzhou 510300, Guangdong, China. E-mail: jianbing_fan@anchordx.com (JBF); weimei_ruan@anchordx.com (WMR)

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Cai Q, Zhu C, Yuan Y, Feng Q, Feng Y, Hao Y, Li J, Zhang K, Ye G, Ye L, Lv N, Zhang S, Liu C, Li M, Liu Q, Li R, Pan J, Yang X, Zhu X, Li Y, Lao B, Ling A, Chen H, Li X, Xu P, Zhou J, Liu B, Du Z, Du Y and Li Z; Gastrointestinal Early Cancer Prevention & Treatment Alliance of China (GECA). Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. Gut 2019; 68: 1576-1587.

[3] Wong CC, Li W, Chan B and Yu J. Epigenomic biomarkers for prognostication and diagnosis of gastrointestinal cancers. Semin Cancer Biol 2019; 55: 90-105.

[4] Necula L, Matei L, Dragu D, Neagu AI, Mambet C, Nedeianu S, Bleotu C, Diaconu CC and Chivu-Economescu M. Recent advances in gastric cancer early diagnosis. World J Gastroenterol 2019; 25: 2029-2044.

[5] Kakeji Y, Ishikawa T, Suzuki S, Akazawa K, Irino T, Miyashiro I, Ono H, Suzuki H, Tanabe S, Kadowaki S, Muro K, Fukagawa T, Nunobe S, Wada T, Katai H and Kodera Y; Registration Committee of the Japanese Gastric Cancer Association. A retrospective 5-year survival analysis of surgically resected gastric cancer cases from the Japanese Gastric Cancer Association nationwide registry (2001-2013). Gastric Cancer 2022; 25: 1082-1093.

[6] Kusumoto H, Tashiro K, Shimaoka S, Tsukasa K, Baba Y, Furukawa S, Furukawa J, Niihara T, Hirotsu T and Uozumi T. Efficiency of gastrointestinal cancer detection by nematode-NOSE (N-NOSE). In Vivo 2020; 34: 73-80.

[7] Roy S, Kanda M, Nomura S, Zhu Z, Toiyama Y, Taketomi A, Goldenring J, Baba H, Kodera Y and Goel A. Diagnostic efficacy of circular RNAs as noninvasive, liquid biopsy biomarkers for early detection of gastric cancer. Mol Cancer 2022; 21: 42.

[8] Karimi P, Islami F, Anandasabapathy S, Freedman ND and Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer Epidemiol Biomarkers Prev 2014; 23: 700-13.

[9] Li J, Zhou X, Liu X, Ren J, Wang J, Wang W, Zheng Y, Shi X, Sun T, Li Z, Kang A, Tang F, Wen L and Fu W. Detection of colorectal cancer in circulating cell-free DNA by methylated CpG tandem amplification and sequencing. Clin Chem 2019; 65: 916-926.

[10] Constâncio V, Nunes SP, Henrique R and Jerónimo C. DNA methylation-based testing in liquid biopsies as detection and prognostic biomarkers for the four major cancer types. Cells 2020; 9: 624.

[11] De Rubis G, Krishnan SR and Bebawy M. Circulating tumor DNA - Current state of play and future perspectives. Pharmacol Res 2018; 136: 35-44.

[12] Liang L, Zhang Y, Li C, Liao Y, Wang G, Xu J, Li Y, Yuan G, Sun Y, Zhang R, Li X, Nian W, Zhao J, Zhang Y, Zhu X, Wen X, Cai S, Li N and Wu L. Plasma cfDNA methylation markers for the detection and prognosis of ovarian cancer. EBioMedicine 2022; 83: 104222.

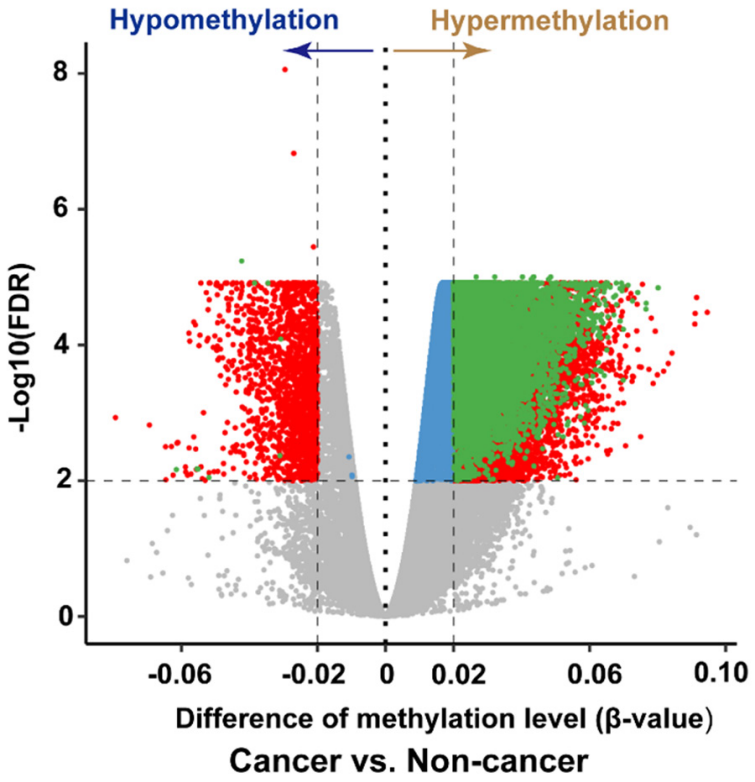
[13] Van der Pol Y and Mouliere F. Toward the early detection of cancer by decoding the epigenetic and environmental fingerprints of cell-free DNA. Cancer Cell 2019; 36: 350-368.

[14] Chen W, Yan H, Li X, Ge K and Wu J. Circulating tumor DNA detection and its application status in gastric cancer: a narrative review. Transl Cancer Res 2021; 10: 529-536.

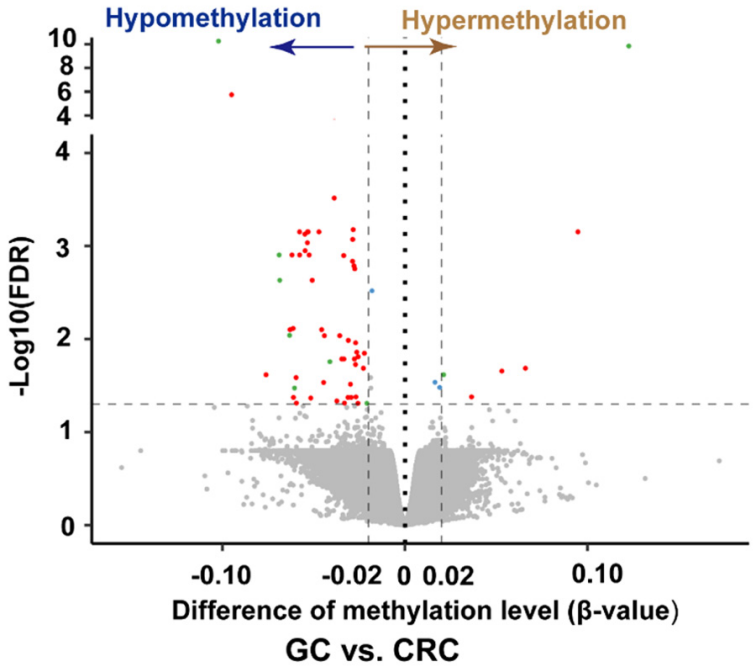
[15] Huang J and Wang L. Cell-free DNA methylation profiling analysis-technologies and bioinformatics. Cancers (Basel) 2019; 11: 1741.
[16] Luo H, Wei W, Ye Z, Zheng J and Xu RH. Liquid biopsy of methylation biomarkers in cell-free DNA. Trends Mol Med 2021; 27: 482-500.
[17] Ren J, Lu P, Zhou X, Liao Y, Liu X, Li J, Wang W, Wang J, Wen L, Fu W and Tang F. Genome-scale methylation analysis of circulating cell-free DNA in gastric cancer patients. Clin Chem 2022; 68: 354-364.
[18] Zhang X, Zhao D, Yin Y, Yang T, You Z, Li D, Chen Y, Jiang Y, Xu S, Geng J, Zhao Y, Wang J, Li H, Tao J, Lei S, Jiang Z, Chen Z, Yu S, Fan JB and Pang D. Circulating cell-free DNA-based methylation patterns for breast cancer diagnosis. NPJ Breast Cancer 2021; 7: 106.
[19] Liang W, Zhao Y, Huang W, Gao Y, Xu W, Tao J, Yang M, Li L, Ping W, Shen H, Fu X, Chen Z, Laird PW, Cai X, Fan JB and He J. Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). Theranostics 2019; 9: 2056-2070.
[20] Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, Wang W, Sheng H, Pu H, Mo H, Zuo Z, Liu Z, Li C, Xie C, Zeng Z, Li W, Hao X, Liu Y, Cao S, Liu W, Gibson S, Zhang K, Xu G and Xu RH. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med 2020; 12: eaax7533.
[21] Anderson BW, Suh YS, Choi B, Lee HJ, Yab TC, Taylor WR, Dukek BA, Berger CK, Cao X, Foote PH, Devens ME, Boardman LA, Kisiel JB, Mahoney DW, Slettedahl SW, Allawi HT, Lidgard GP, Smyrk TC, Yang HK and Ahlquist DA. Detection of gastric cancer with novel methylated DNA markers: discovery, tissue validation, and pilot testing in plasma. Clin Cancer Res 2018; 24: 5724-5734.
[22] Wong D, Lounsbury K, Lum A, Song J, Chan S, LeBlanc V, Chittaranjan S, Marra M and Yip S. Transcriptomic analysis of CIC and ATXN1L reveal a functional relationship exploited by cancer. Oncogene 2019; 38: 273-290.
[23] Kang AR, An HT, Ko J and Kang S. Ataxin-1 regulates epithelial-mesenchymal transition of cervical cancer cells. Oncotarget 2017; 8: 18248-18259.
[24] Lee Y, Fryer JD, Kang H, Crespo-Barreto J, Bowman AB, Gao Y, Kahle JJ, Hong JS, Kheradmand F, Orr HT, Finegold MJ and Zoghbi HY. ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization. Dev Cell 2011; 21: 746-757.
[25] Tong X, Gui H, Jin F, Heck BW, Lin P, Ma J, Fondell JD and Tsai CC. Ataxin-1 and brother of ataxin-1 are components of the Notch signalling pathway. EMBO Rep 2011; 12: 428-35.
[26] Zhou B, Lin W, Long Y, Yang Y, Zhang H, Wu K and Chu Q. Notch signaling pathway: architecture, disease, and therapeutics. Signal Transduct Target Ther 2022; 7: 95.
[27] Diaz-Valencia JD, Estrada-Abreo LA, Rodríguez-Cruz L, Salgado-Aguayo AR and Patiño-López G. Class I myosins, molecular motors involved in cell migration and cancer. Cell Adh Migr 2022; 16: 1-12.
[28] Estrada-Abreo LA, Rodríguez-Cruz L, Garfias-Gómez Y, Araujo-Cardenas JE, Antonio-Andrés G, Salgado-Aguayo AR, Orozco-Ruiz D, Torres-Nava JR, Díaz-Valencia JD, Huerta-Yépez S and Patiño-López G. High expression of myosin 1g in pediatric acute lymphoblastic leukemia. Oncotarget 2021; 12: 1937-1945.
[29] Huang SM, Chen TS, Chiu CM, Chang LK, Liao KF, Tan HM, Yeh WL, Chang GR, Wang MY and Lu DY. GDNF increases cell motility in human colon cancer through VEGF-VEGFR1 interaction. Endocr Relat Cancer 2013; 21: 73-84.
[30] Fielder GC, Yang TW, Razdan M, Li Y, Lu J, Perry JK, Lobie PE and Liu DX. The GDNF family: a role in cancer? Neoplasia 2018; 20: 99-117.
[31] Furuta A, Funahashi H, Sawai H, Sato M, Okada Y, Takeyama H and Manabe T. The relationship between GDNF and integrins in human colorectal cancer cell activity. Hepatogastroenterology 2007; 54: 1398-402.
[32] Kohno T, Tabata J and Nakaoku T. RETOma: a cancer subtype with a shared driver oncogene. Carcinogenesis 2020; 41: 123-129.
[33] Jelski W and Mroczko B. Biochemical markers of colorectal cancer - Present and future. Cancer Manag Res 2020; 12: 4789-4797.
[34] Kotzev AI and Draganov PV. Carbohydrate antigen 19-9, carcinoembryonic antigen, and carbohydrate antigen 72-4 in gastric cancer: is the old band still playing? Gastrointest Tumors 2018; 5: 1-13.
[35] Xie L, Jiang X, Li Q, Sun Z, Quan W, Duan Y, Li D and Chen T. Diagnostic value of methylated septin9 for colorectal cancer detection. Front Oncol 2018; 8: 247.
[36] Cross AJ, Wooldrage K, Robbins EC, Kralj-Hans I, MacRae E, Piggott C, Stenson I, Prendergast A, Patel B, Pack K, Howe R, Swart N, Snowball J, Duffy SW, Morris S, von Wagner C, Halloran SP and Atkin WS. Faecal immunochemical tests (FIT) versus colonoscopy for surveillance after screening and polypectomy: a diagnostic accuracy and cost-effectiveness study. Gut 2019; 68: 1642-1652.
[37] Liu L, Toung JM, Jassowicz AF, Vijayaraghavan R, Kang H, Zhang R, Kruglyak KM, Huang HJ, Hinoue T, Shen H, Salathia NS, Hong DS, Naing

A, Subbiah V, Piha-Paul SA, Bibikova M, Granger G, Barnes B, Shen R, Gutekunst K, Fu S, Tsimberidou AM, Lu C, Eng C, Moulder SL, Kopetz ES, Amaria RN, Meric-Bernstam F, Laird PW, Fan JB and Janku F. Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification. Ann Oncol 2018; 29: 1445-1453.

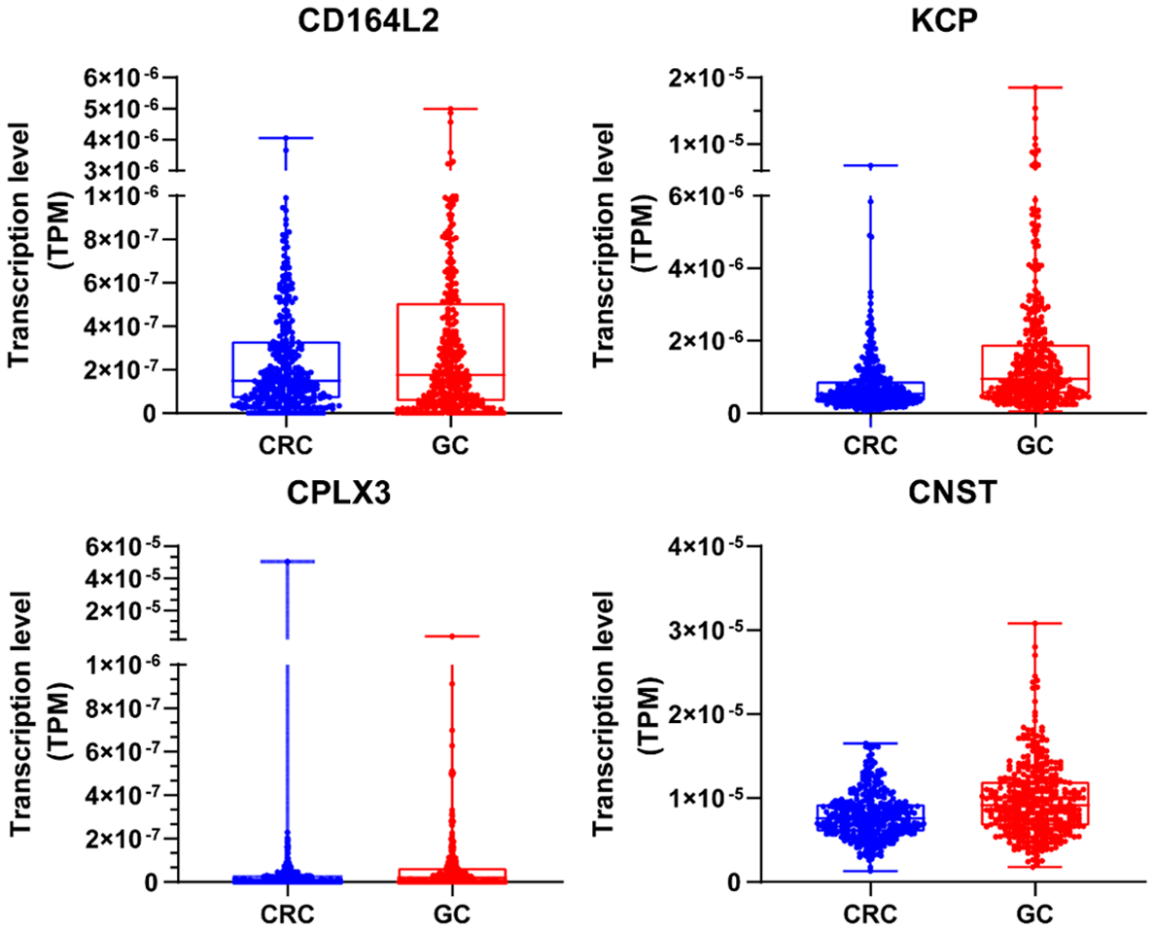
[38] Potter NT, Hurban P, White MN, Whitlock KD, Lofton-Day CE, Tetzner R, Koenig T, Quigley NB and Weiss G. Validation of a real-time PCR-based qualitative assay for the detection of methylated SEPT9 DNA in human plasma. Clin Chem 2014; 60: 1183-91.



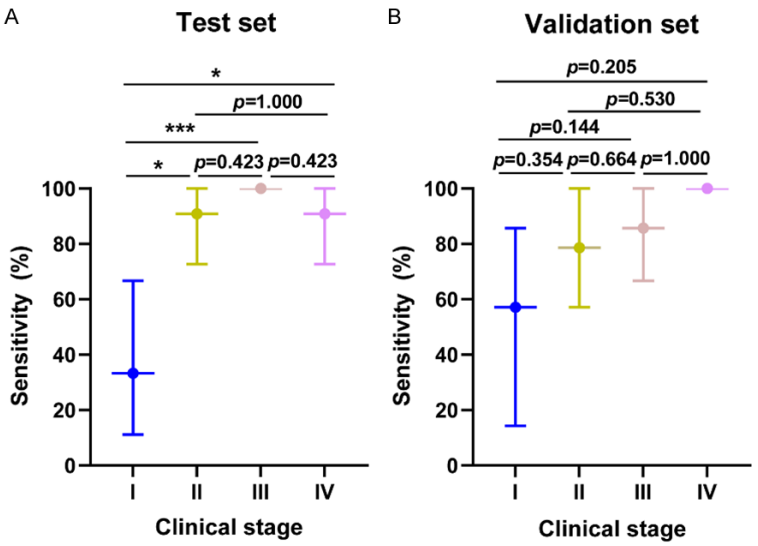
Supplementary Figure 1. Volcano plots illustrating the gastrointestinal cancer-specific hyper- and hypo-methylation CpG sites. The green plots denote differential methylation markers with mean β-value difference ≥ 0.02 and fold change ≥ 1.5; the red plots denote differential methylation markers with mean β-value difference ≥ 0.02 but fold change < 1.5; the blue plots denote markers with fold change ≥ 1.5 but β-value difference < 0.02. The markers that are not differentially methylated are denoted as gray dots. Markers with mean β-value difference ≥ 0 was defined as hypermethylation markers while markers with mean β-value difference < 0 was defined as hypomethylation markers.



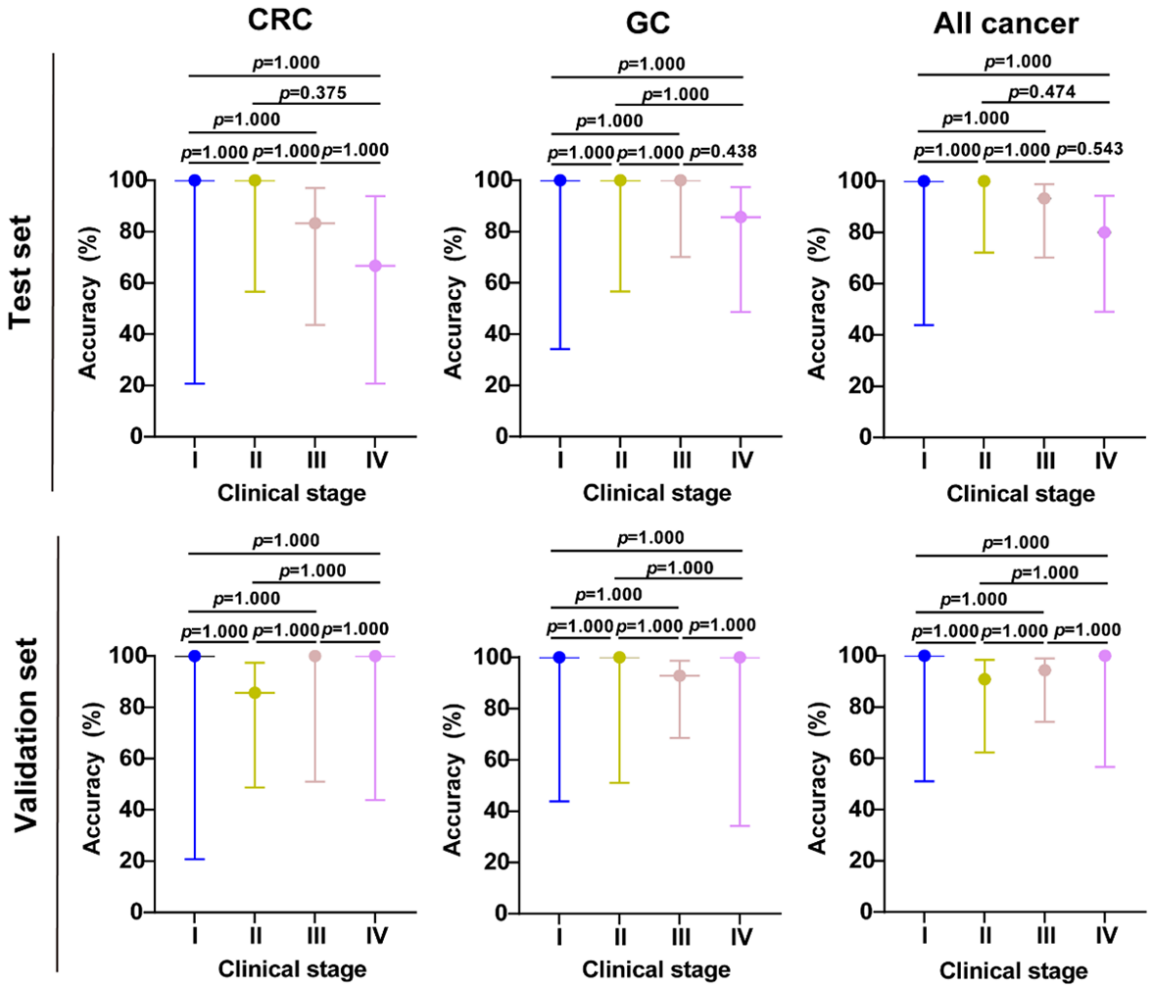
Supplementary Figure 2. Volcano plots illustrating the methylation markers of tissue of origin prediction of GC and CRC. The green plots denote differential methylation markers with mean β-value difference ≥ 0.02 and fold change ≥ 1.5; the red plots denote differential methylation markers with mean β-value difference ≥ 0.02 but fold change < 1.5; the blue plots denote differential methylation markers with fold change ≥ 1.5 but β-value difference < 0.02. The markers that are not differentially methylated are denoted as gray dots. Markers of which mean β-value difference ≥ 0 was defined as hypermethylation markers in GC while that with mean β-value difference < 0 was defined as hypomethylation markers.



Supplementary Figure 3. RNA expression level of the indicated TOO markers in CRC and GC tumor. RNAseq data of the corresponding genes in colorectal tumor and gastric tumor was downloaded from UCSC Xena database, and TPM (transcripts per million) of genes was used to denote the transcription level.



Supplementary Figure 4. Significant difference analysis on sensitivities of the gastrointestinal cancer detection among different clinical stages by Fisher's Exact Test.



Supplementary Figure 5. Significant difference analysis on accuracies of the tissue of origin prediction among different clinical stages by Fisher's Exact Test.