CANCER

Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer

Huiyan Luo[1]*, Qi Zhao[1]*, Wei Wei[1]*, Lianghong Zheng[2]*, Shaohua Yi[3]*, Gen Li[4], Wenqiu Wang[5], Hui Sheng[1], Hengying Pu[1], Haiyu Mo[1], Zhixiang Zuo[1], Zexian Liu[1], Chaofeng Li[1], Chuanbo Xie[1], Zhaolei Zeng[1], Weimin Li[6], Xiaoke Hao[7], Yuying Liu[1], Sumei Cao[1], Wanli Liu[1], Sarah Gibson[4], Kang Zhang[6,8], Guoliang Xu[1], Rui-hua Xu[1†]

Circulating tumor DNA (ctDNA) has emerged as a useful diagnostic and prognostic biomarker in many cancers. Here, we conducted a study to investigate the potential use of ctDNA methylation markers for the diagnosis and prognostication of colorectal cancer (CRC) and used a prospective cohort to validate their effectiveness in screening patients at high risk of CRC. We first identified CRC-specific methylation signatures by comparing CRC tissues to normal blood leukocytes. Then, we applied a machine learning algorithm to develop a predictive diagnostic and a prognostic model using cell-free DNA (cfDNA) samples from a cohort of 801 patients with CRC and 1021 normal controls. The obtained diagnostic prediction model discriminated patients with CRC from normal controls with high accuracy (area under curve = 0.96). The prognostic prediction model also effectively predicted the prognosis and survival of patients with CRC ($P < 0.001$). In addition, we generated a ctDNA-based molecular classification of CRC using an unsupervised clustering method and obtained two subgroups of patients with CRC with significantly different overall survival ($P = 0.011$ in validation cohort). Last, we found that a single ctDNA methylation marker, cg10673833, could yield high sensitivity (89.7%) and specificity (86.8%) for detection of CRC and precancerous lesions in a high-risk population of 1493 participants in a prospective cohort study. Together, our findings showed the value of ctDNA methylation markers in the diagnosis, surveillance, and prognosis of CRC.

INTRODUCTION

Colorectal cancer (CRC) is the third most common fatal cancer worldwide (*1*). As with other cancers, patients with CRC diagnosed at earlier or intermediate stages have better prognoses than those at advanced stages (*2*, *3*). Thus, early detection is helpful for improving the survival of these patients, but there has been limited clinical success in developing effective, noninvasive diagnostic approaches. Serum carcinoembryonic antigen (CEA) quantification, a noninvasive biomarker, has shown good specificity for identifying occult CRC, but its use has been limited due to its low sensitivity, 40 to 60% (*4*, *5*). Colonoscopy, usually considered to be the best test for early visual detection and screening of CRC, is uncomfortable, invasive, time-consuming, and expensive and may lead to complications. All of these considerations may negatively affect patients' compliance with recommended screening, indicating the need for the development of specific, sensitive, and noninvasive biomarkers for the early detection of CRC.

Circulating tumor DNA (ctDNA) is tumor-derived fragmented DNA in the cell-free fraction of the blood, mainly derived from dead tumor cells through necrosis and apoptosis (*6*, *7*). Given its origin, ctDNA carries cancer-specific genetic and epigenetic aberrations, which can be used as a surrogate source of tumor DNA in cancer diagnosis and prognosis. Several studies have assessed the usefulness of quantitative and qualitative tumor-specific alterations of cell-free DNA (cfDNA) as diagnostic, prognostic, and monitoring markers in patients with cancer (*8*, *9*), introducing the concept of liquid biopsy (*10*–*12*). Recently, Cohen *et al.* reported a blood test for early detection of eight common cancer types through assessment of the expression of circulating proteins and mutations in cfDNA (*13*).

DNA methylation is a major epigenetic modification that is involved in differentiation and development, aging, tumorigenesis, and other diseases. Aberrant methylation is a central feature of carcinogenesis and usually causes defective gene expression (*14*). Increased methylation of tumor suppressor genes is an early event in many tumors, and it may also be one of the first detectable neoplastic changes associated with tumorigenesis (*15*–*17*). ctDNA methylation profiling provides several advantages over somatic mutation analysis for cancer detection, including higher clinical sensitivity and dynamic range, multiple detectable methylation target regions, and multiple altered CpG sites within each targeted genomic region (*18*, *19*). The alterations in CpG methylation are relatively constant in each type of cancer, whereas there are usually no predominant somatic mutations. Despite the relatively high frequency of mutations, specific aberrant patterns are still highly heterogeneous in individual patients, making somatic mutations less than ideal markers for early detection of cancers (*20*). A number of studies have identified some specific DNA methylation sites, such as *SEPT9*, as biomarkers of CRC (*21*–*23*). However, the potential value of ctDNA bearing cancer-specific methylation biomarkers for screening and early detection of CRC remained to be further investigated.

It is challenging to obtain reliable and quantitative methylation measurement values with low amounts of cfDNA and considering

[1]Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, P. R. China. [2]Guangzhou Youze Biological Pharmaceutical Technology Company Ltd., Guangzhou 510005, P.R. China. [3]Huazhong University of Science and Technology Tongji Medical College, Wuhan 430030, P. R. China. [4]Guangzhou Women and Children's Medical Center, Guangzhou 510623, P. R. China. [5]Shanghai General Hospital, Shanghai 200080, P. R. China. [6]Molecular Medicine Center, West China Hospital, Sichuan University, Chengdu 610041, P. R. China. [7]Department of Clinical Laboratory Medicine, Xijing Hospital, Fourth Military Medical University, Xi'an 710032, P. R. China. [8]Faculty of Medicine, Macau University of Science and Technology, Macau 999078, P. R. China.
*These authors contributed equally to this work.
†Corresponding author. Email: xurh@sysucc.org.cn

that methyltransferase or demethylase might simultaneously modify adjacent CpG sites within the same DNA strand. Similar to the concept of haplotype blocks of adjacent single-nucleotide polymorphisms in DNA, these adjacent stretches of CpG methylation increase the accuracy of determining allele methylation status (*24*–*26*). We have termed these stretches of DNA "methylation correlated blocks (MCBs)" (*24*, *27*).

Our previous study showed that the cfDNA methylation profile can be used in the diagnosis, surveillance, and prognostication of hepatocellular carcinoma (HCC) (*24*). Here, we focused on evaluating the potential usefulness of cfDNA methylation markers for CRC surveillance and the efficacy of CpG markers in screening for CRC. Multiple statistical methods were applied to construct diagnostic and prognostic prediction models with selected methylation markers (Fig. 1). We further obtained a ctDNA-based molecular subtype of CRC using an unsupervised clustering method. In addition, we determined the sensitivity and specificity of this methylation-based screening approach in a large prospectively enrolled Chinese cohort with high risk of CRC.

RESULTS

Patient and sample characteristics

The demographics and clinical characteristics of the participants are summarized in tables S1 and S2. The methylation profiles were collected for 459 CRC tumor samples from The Cancer Genome Atlas (TCGA) and 754 normal samples from a dataset used in a previous methylation study on aging (GSE40279) (*28*) to identify CRC-specific methylation markers. Compared with the TCGA dataset, the healthy controls in the GSE dataset were relatively younger (mean age, 63 versus 68 years) and had a lower proportion of males (46.8% versus 52.9%).

To study the cfDNA in CRC, the plasma samples of 801 CRC Chinese patients and a contemporary population comprising 1021 healthy controls were analyzed (Fig. 1A). The patients with CRC enrolled from our institute were older (58 versus 47 years) and had a higher proportion of males (61.9% versus 46.0%) than those from the control group.

A total of 16,890 participants were enrolled in the prospective screening cohort study, and 1493 participants with a high risk of CRC underwent colonoscopy and cfDNA methylation tests (Fig. 1B). Clinical characteristics of all participants from this cohort are listed in table S2. A total of 29 participants were found to have CRC on colonoscopy (prevalence, 1.9%). A total of 78 participants had advanced precancerous lesions (prevalence, 5.2%).

cfDNA-based diagnostic prediction model for CRC

We analyzed the entire methylation dataset of 544 markers using the least absolute shrinkage and selection operator (LASSO) and random forest algorithms to reduce the number of markers. Samples were randomly assigned to training and validation sets with a 2:1 ratio (Fig. 2A). We obtained nine overlapping markers from the two algorithms and constructed a diagnostic score (cd-score) that was obtained according to the coefficients from multinomial logistic regression (fig. S1 and table S3). Using this score, we observed a high consistency between predicted results and pathological diagnosis results in both the training and validation datasets (Fig. 2, B to E). CEA has been explored for CRC diagnosis since decades ago, but its clinical usage has been hindered by its low sensitivity and specificity (*5*), and invasive approaches such as colonoscopy have been instead investigated for patients with the highest suspicion of CRC. In

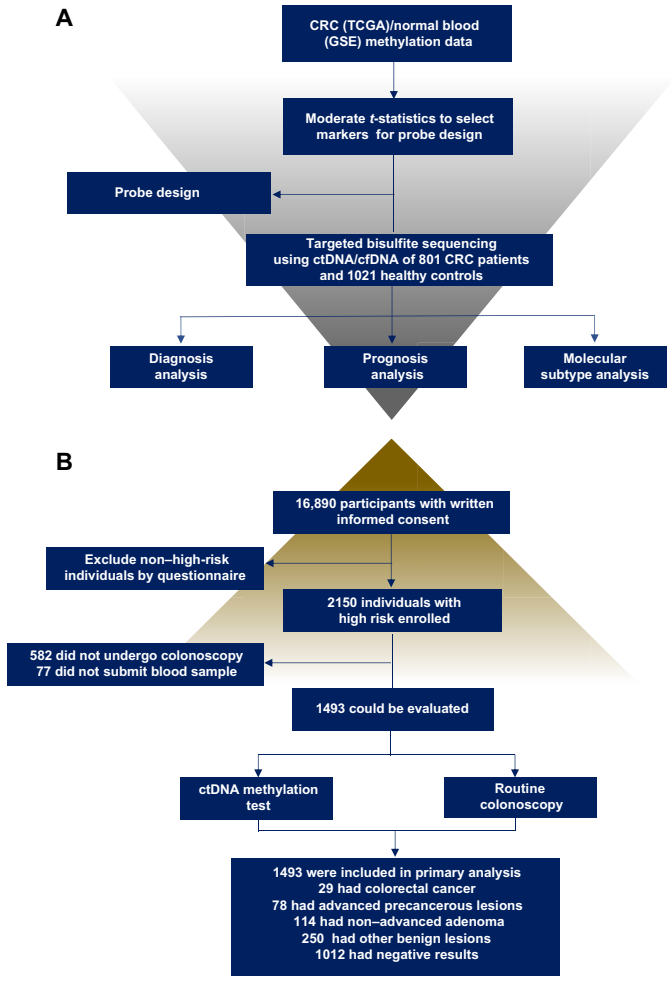
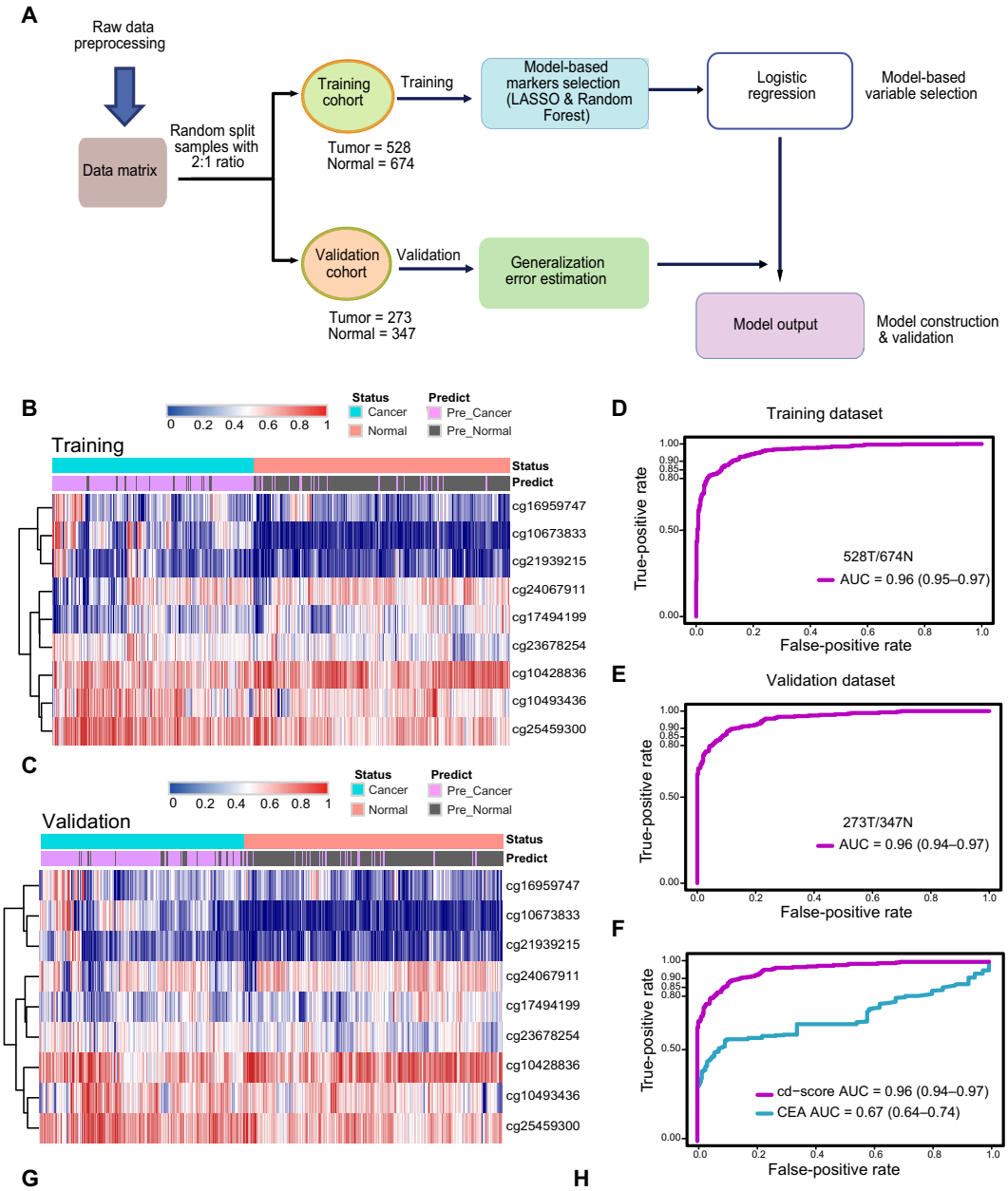


Fig. 1. Workflow of model generation and subject enrolment. (**A**) Workflow for building the diagnostic and prognostic models with cfDNA methylation markers. (**B**) Enrollment and outcomes of the prospective screening cohort study.

contrast, the cd-score demonstrated superior sensitivity and specificity to CEA for CRC diagnosis [area under curve (AUC), 0.96 versus 0.67; Fig. 2F]. In addition, using a best cutoff value, as determined via the Youden index method (*29*), the cd-score demonstrated sensitivity and specificity of 87.5 and 89.9%, respectively, for discriminating CRC from normal controls in the training dataset, and 87.9% and 89.6%, respectively, in the validation dataset (Fig. 2, G and H).

We then examined the usefulness of the cd-score in assessing the staging of CRC, the presence of residual tumors after treatment, and the response to treatment (such as surgery or chemotherapy). The cd-scores of patients with detectable residual tumor after treatment were significantly higher than those without detectable tumor ($P < 0.001$; fig. S2A). Similarly, there was good correlation between the cd-scores and tumor stage. Patients with early-stage (I and II) disease had substantially lower cd-scores compared with those with advanced-stage (III and IV) disease ($P < 0.001$; fig. S2B). The cd-scores of patients with primary tumor on the right side were also higher than those with primary tumors on the left side (fig. S2C). Furthermore, the cd-scores were significantly higher in patients before treatment compared with those who underwent surgical resection ($P < 0.001$; fig. S2D). The scores also increased after relapse (fig. S2D).



G

Training cohort	CRC	Normal	
CRC	462	68	
Normal	66	606	Totals
Totals	528	674	1202
Correct	462	606	1068
Sensitivity (%)	87.5		
Specificity (%)		89.9	88.9

H

Validation cohort	CRC	Normal	
CRC	240	36	
Normal	33	311	Totals
Totals	273	347	620
Correct	240	311	551
Sensitivity (%)	87.9		
Specificity (%)		89.6	88.9

Fig. 2. cfDNA methylation analysis for CRC diagnosis. (**A**) Workflow for building the diagnostic model. (**B** and **C**) Unsupervised hierarchical clustering of methylation markers differentially methylated between CRC and normal subject DNA in the training (B) and validation (C) testing cohorts. Each row represents an individual patient, and each column is a CpG marker. (**D** and **E**) Receiver operating characteristic (ROC) curves and the associated areas under curves (AUCs) of the diagnostic prediction model (cd-score) using cfDNA methylation analysis in the training (D) and validation (E) testing cohorts. (**F**) ROC curves and corresponding AUCs of cd-score and CEA for CRC diagnosis in the validation dataset. (**G** and **H**) Confusion matrices built from the diagnostic model prediction in the training (G) and validation (H) testing cohorts.

Both the cd-score and CEA values were correlated with the tumor stage (fig. S2, B and E). The cd-score was more strongly affected by surgical resection than CEA (fig. S2, D and F).

We also calculated the diagnosis efficiency of each of the nine markers in the diagnostic model (fig. S3), and cg10673833 showed the best diagnostic performance, yielding AUCs of 0.904 and 0.91 in the training and validation datasets, respectively (fig. S3). We, thus, investigated the usefulness of cg10673833 as a marker of treatment response. We monitored longitudinal dynamic changes of the methylation values of this marker in patients with CRC with a series of specimens before and after treatment. The results showed that the dynamic changes in cg10673833 methylation were consistent with treatment outcomes, and these changes were more pronounced than those of CEA (fig. S4). In patients with serial samples, those with positive treatment (surgery or chemotherapy) responses had a concomitant decrease in cg10673833 methylation compared with untreated patients, and a further reduction in cg10673833 methylation was observed in patients after surgery. In contrast, patients with progressive or recurrent disease showed increased methylation (fig. S5).

cfDNA-based prognostic prediction model for CRC

On the basis of the cfDNA methylation analysis, we introduced a combined prognosis score (cp-score) for the prognostication of CRC, in combination with clinical and demographic characteristics including age, gender, primary tumor site, and AJCC (American Joint Committee on Cancer) stage. We analyzed the same training dataset as in the diagnostic section, which contained 528 observations with 157 events, and a validation dataset containing 273 observations with 77 events. The median follow-up time was 26.6 months (range, 1 to 42 months). We conducted a variable selection on the training set and built the composite score on the validation set (Fig. 1). The UniCox and LASSO-Cox methods were implemented to reduce the dimensionality, and a Cox-model was constructed to predict prognosis with a five-marker panel (Fig. 3A and table S4). Kaplan-Meier curves were generated using the dichotomized composite score, which separated the patients into high-risk and low-risk groups relative to the median. The median survival time in the low-risk group was significantly better than that in the high-risk group ($P < 0.001$ in the training cohort and $P = 0.0012$ in the validation cohort) (Fig. 3, B and C).

We used time-dependent ROC (*30*) to characterize the discrimination potential of the composite score, AJCC stage, CEA concentration, primary tumor location, and the combination of all the existing biomarkers. Multivariate Cox regression analysis indicated that the cp-score highly correlated with risk of death and was an independent factor of survival in both the training and validation sets (table S5). As expected, the TNM (tumor, node, metastasis) stage (as defined by the AJCC guidelines), CEA status, and primary tumor location were also prognostic factors for survival of patients with CRC (table S5). Time-dependent ROC analysis showed that the combination of cp-score and clinical characteristics improved our ability to predict prognosis [training cohort: AUC, 0.82; and 95% confidence intervals (CI), 0.77 to 0.87; validation cohort: AUC, 0.87; 95% CI, 0.82 to 0.93] (Fig. 3, D and E), when compared with cp-score or clinical characteristics, such as TNM stages, primary tumor site, and CEA status.

We developed a nomogram with a point scale of the four variables (cp-score, CEA concentration, TNM stage, and primary tumor location, identified as independent predictive factors in multivariate Cox regression analysis; table S5) to predict the overall survival of patients with CRC (fig. S6). Figure S6B shows the calibration graph for the nomogram, in which the probability of 3-year overall survival as predicted by the nomogram is plotted against the corresponding observed survival rates obtained by the Kaplan-Meier method. The c-index of this model was 0.78 in the validation cohort, indicating good discrimination.

cfDNA-based subtyping of CRC

To generate cfDNA methylation–based subtypes of CRC, we used an unsupervised clustering method modified from a recent study (*31*, *32*). The method applied an iteration strategy that could derive the optimal signature and clusters from a consensus similarity matrix generated by consensus clustering (Fig. 4A). Using the same training dataset as in the prognostic model, we obtained two clusters of CRC samples with 45 markers that were differentially methylated between the clusters (Fig. 4, B and C, and table S6). We also observed distinctly different methylation profiles of the 45 markers between the two clusters (Fig. 4D) in the validation dataset. Among these markers, three were also in the diagnostic markers list, and one was in both the diagnostic and prognostic marker lists (fig. S7A).

To explore the clinical relevance of the two subtypes, we systematically tested the associations between the subtype and clinical factors including TNM stage, tumor site, mismatch repair status, microsatellite stability status, tumor burden, sex, mutation status of a limited gene panel (including *KRAS*, *NRAS*, *BRAF*, *PIK3CA*, and *PTEN*), and survival outcomes. Cluster 1 tumors were frequently observed in females with left-sided lesions (table S7) and usually diagnosed in earlier stages (I and II) (Fig. 4E, lower panel, both $P < 0.05$, χ^2 test). Cluster 2, in both the training and validation datasets, showed a significantly poorer survival rate than that of cluster 1 (Fig. 4E, upper panel, both $P < 0.01$, log-rank test). Further analysis showed that cp-scores in cluster 2 were significantly higher than those in cluster 1 in both datasets (fig. S7, B and C, both $P < 0.001$, Wilcoxon test, and table S8). These differences in prognosis with unsupervised ctDNA methylation signatures confirmed the clinical relevance of the intrinsic biological processes implicated in each cluster group.

Methylation marker cg10673833 for screening and early diagnosis of CRC in high-risk populations

It is essential to make screening methods for cancer as simple as possible. In our analysis, methylation status of CpG site cg10673833 demonstrated great efficiency in diagnostic performance in CRC. Thus, we prospectively investigated the potential of cg10673833 as a methylation marker for the detection of CRC and precancerous lesions in high-risk populations in plasma samples. From January 2015 to December 2017, we enrolled 16,890 participants for this prospective cohort study. All the participants were first invited to take a cancer risk assessment by an established Clinical Cancer Risk Score System (*33*). A total of 1493 participants between the ages of 45 and 75 years, who were considered to be at high risk for CRC, were scheduled to undergo screening colonoscopy and were recruited into the study to undergo methylation profiling at the time of the screening procedure (Fig. 1B). Table 1 and table S9 show the colonoscopy screening results and cg10673833 methylation test. The cg10673833 methylation test identified 19 of 21 participants with CRC, and 7 of 8 participants with CRC in situ (diagnosis as high-grade dysplasia), with a sensitivity of 89.7% (95% CI, 0.727 to 0.978), a specificity of 86.8% (95% CI, 0.849 to 0.884), and AUC of 0.90 (95% CI, 0.885 to 0.942). The positive predictive value and negative predictive value were 0.118 (95% CI, 0.101 to 0.138) and 0.998 (95% CI, 0.993 to 0.999), respectively (table S10). For advanced precancerous lesions, the sensitivity

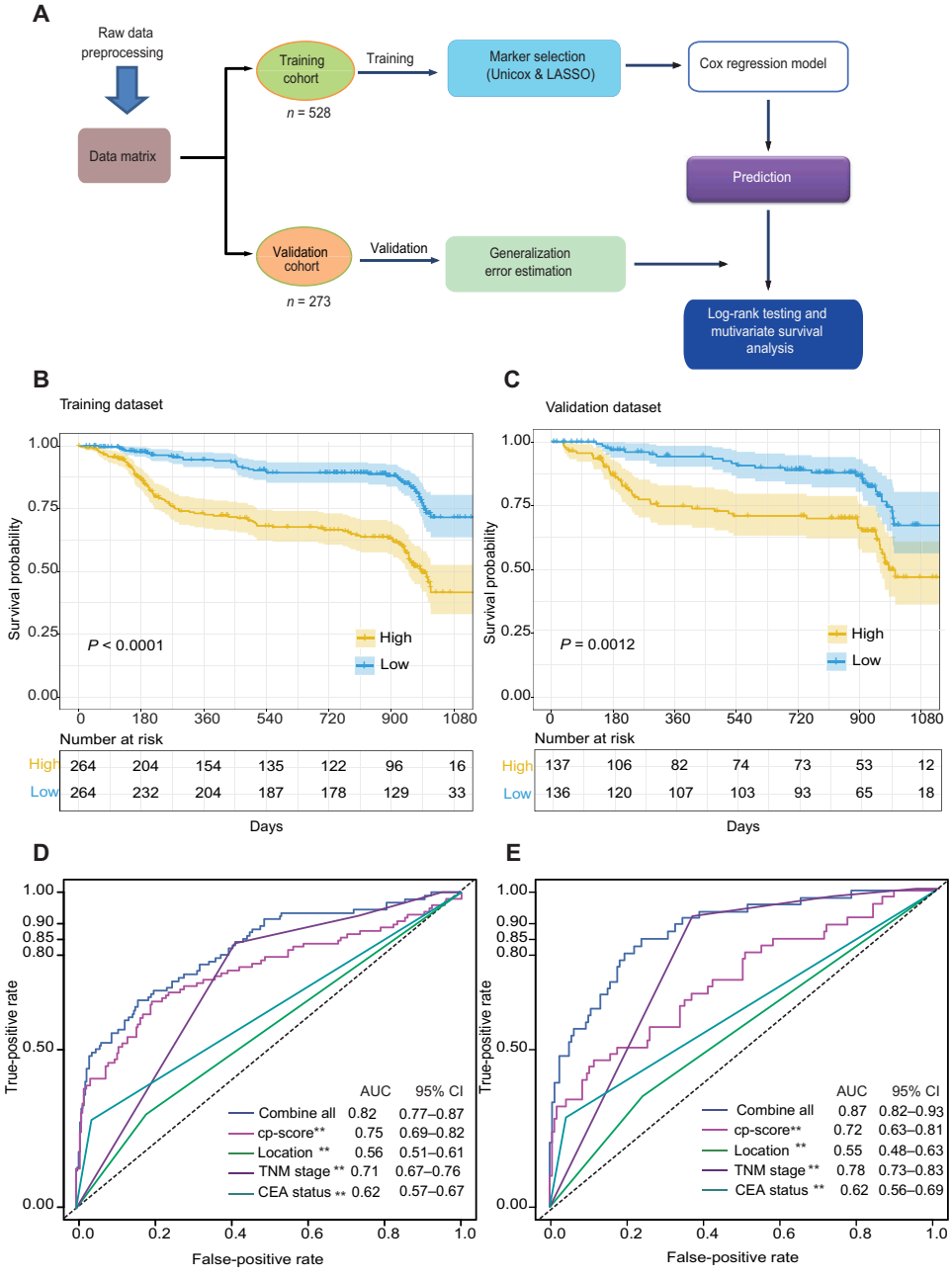


Fig. 3. Prognostic prediction of CRC survival based on cfDNA methylation profiling. (**A**) Workflow for building the prognostic models. (**B**) Overall survival curves of patients with CRC with low or high risk of death according to the combined prognosis score (cp-score) in the training testing cohort. (**C**) Overall survival curves of patients with CRC with low or high risk of death according to the combined prognosis score (cp-score) in the validation testing cohort. (D to E) ROC and corresponding AUCs for 6-month survival predicted by cp-score, primary tumor location, TNM stage, CEA status, and all combined in the training (D) and validation (E) testing cohorts. $^{**}P < 0.001$.

was 33.3% (95% CI, 0.231 to 0.449), much higher than the positivity rate for subjects without cancer or advanced precancerous lesions (12.1%; 95% CI, 0.101 to 0.139).

DISCUSSION

The majority of CRC cases can be successfully treated if detected early (*2*). Colonoscopy is widely recognized as an effective screening tool for CRC, but its cost and invasive nature limit its use. Moreover, colonoscopy requires bowel cleansing, is often painful, and may at times be biased by inter-observer variability, especially for early lesions, lessening screening efficacy. An accurate, noninvasive diagnostic test for both CRC and advanced precancerous lesions is highly desirable, and to this end, the emergence of liquid biopsy technology has shown to be a promising approach. The methylation of oncogenes and tumor suppressor genes may be present at early stages of malignant transformation, suggesting that methylation patterns could provide reliable discriminatory markers for the detection and diagnosis of malignancy. In our previous study, we showed that DNA methylation signatures were useful for differentiating between tumoral and non-tumoral tissue in four common cancers, namely, breast, colon, liver, and lung cancer (*27*). Despite substantial variability in the somatic mutations of individual tumors, with some notable exceptions (*20*), methylation patterns turned out to be remarkably consistent (*14*). In a separate study, we have also demonstrated the usefulness of cfDNA methylation markers in diagnosis, prognostication, and surveillance of HCC (*24*).

In this study, we developed a diagnostic model (cd-score) using nine selected cfDNA methylation markers and found that this model could accurately discriminate patients with CRC from normal individuals. The sensitivity and specificity of this model for CRC diagnosis were superior to those of CEA, the only globally used blood test for this disease. The cd-score correlated with the staging of CRC, the presence of residual tumor after treatment, and the response to treatment. These results suggest that this model may also be useful for the detection of residual tumor, evaluation of treatment efficacy, and surveillance of recurrence. We presume that methylation markers in the diagnostic model and the downstream genes might play accumulative roles in carcinogenesis and development of CRC. Further, the elucidation of the underlying mechanism might also provide potential targets for therapeutic interventions and prevention of CRC.

A prognostic prediction model (cp-score) was then constructed using another five-marker panel. In the model, the cp-score could effectively distinguish patients with CRC with different prognoses and was validated as an independent prognostic risk factor in a multivariable analysis. When compared to other prognostic risk

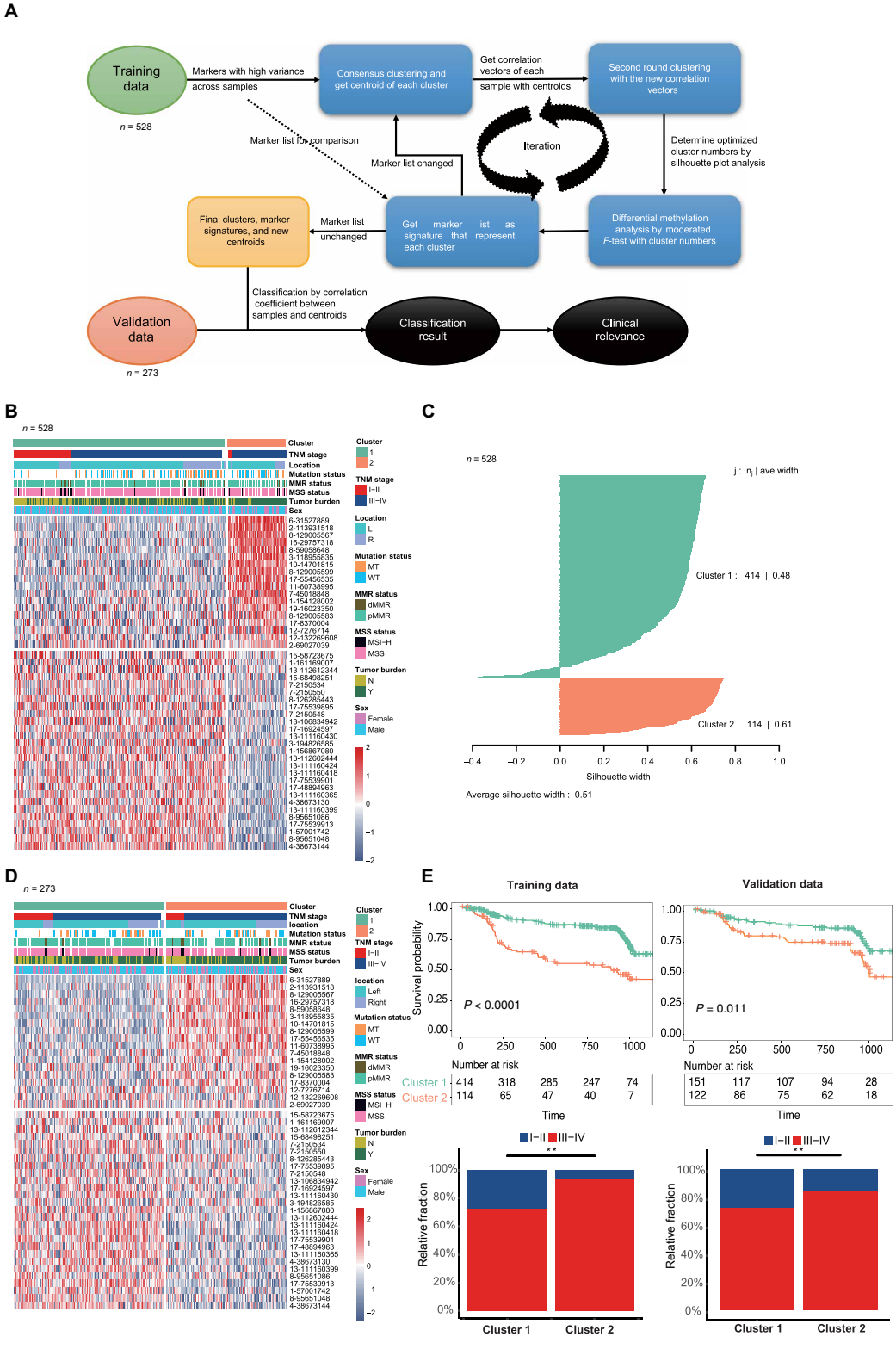


Fig. 4. cfDNA methylation subtyping analysis in 801 patients with CRC. (**A**) Schematic diagram of the core algorithm used in the sample clustering. (**B**) Iteratively unsupervised clustering of cfDNA methylation markers identified two subtypes/clusters in training data. Clinical and molecular features are indicated by the annotation bars above the heatmap. Patients without such information were colored in white. Mutation status was defined by the mutation detected in one or more of the following genes: *BRAF*, *KRAS*, *NRAS*, and *PIK3CA*. (**C**) Silhouette analysis of the clusters in the last iteration. (**D**) Predicted subtypes/clusters of validation using the 45 markers. (**E**) Upper panel: Overall survival for each of the cfDNA methylation patterns in each subtype (log-rank test, $P < 0.05$). Lower panel: Proportion of patients with stage III to IV CRC in two clusters (χ^2 test, $^{**}P < 0.01$; left, training cohort; right, validation cohort).

Table 1. Sensitivity and specificity of the cfDNA methylation test for colonoscopy findings.

Colonoscopy finding	Patients (*n* = 1493)	ctDNA methylation test (*n* = 1493)			
		Positive results	Negative results	Sensitivity	Specificity
CRC	**No.**	**No.**	**No.**	**% (95% CI)**	**% (95% CI)**
Stage I–III CRC	21	19	2	90.5 (72.7–97.8)	
High-grade dysplasia	8	7	1	87.5 (47.4–99.7)	
All CRC				89.7 (72.7–97.8)	
Advanced precancerous lesion*	78	26	52	33.3 (23.1–44.9)	66.7 (55.1–76.9)
Nonadvanced adenoma	114	25	89	21.9 (14.7–30.7)	78.1 (69.4–85.3)
Other benign lesions	250	20	230	8.0 (4.9–12.1)	92.0 (87.9–95.1)
Negative on colonoscopy	1012	123	889		87.9 (85.7–89.8)
All nonadvanced adenoma and non-neoplastic findings and negative results on colonoscopy	1386	168	1218		87.9 (86.0–89.6)
All noncancer	1464	194	1270		86.8 (84.9–88.4)

*Advanced precancerous lesion includes the following: villous adenoma, adenoma >1.0 cm in size, and sessile serrated polyps >1.0 cm in size.

factors (CEA status, TNM stage, and primary tumor location), the discrimination potential of the cp-score was found to be superior. The combination of cp-score with clinical characteristics improved prognostic estimation, which helped identify patients who would need more aggressive treatment and surveillance. In addition, a nomogram consisting of the cp-score, CEA status, TNM stage, and primary tumor location to predict patient survival was constructed for clinical decision-making, and its performance was validated in both training and validation cohorts. The nomogram showed favorable predictive capability and can, therefore, be considered as a potential tool for CRC prognostication. However, our study was limited by a relatively short clinical follow-up (median follow-up time, 26.6 months), and further investigations with longer clinical surveillance to adequately assess the reliability of this score in clinical decision-making for patients are still needed.

Gene expression–based subtyping is widely accepted as a relevant source of disease stratification (*34*), but similar cfDNA-based subtyping is still lacking. With an iteration strategy, we could divide patients into two molecular subgroups based on 45 cfDNA methylation markers. The patients in these two subtypes have different staging distributions and prognoses. Further study of correlation between this cfDNA-based methylation subtyping and clinical factors [including TNM stage, tumor site, microsatellite stable (MSS) status, tumor burden, sex, mutation status of limited gene panels, and survival outcomes] might deepen the understanding surrounding the evolution of CRC while providing a more personalized treatment strategy. We found an association between cfDNA-based subtype group and clinical variables; however, the subtyping part of the study was a retrospective analysis, and most of our patients did not have enough gene mutation and transcriptomic information to type based on current consensus molecular subtypes. It is therefore difficult to compare our cfDNA-based subtypes and the current consensus molecular subtypes based on available data.

Through sequencing of bisulfate-converted DNA (bis-DNA), we identified many CpG markers that are differentially methylated in cancer versus normal plasma. Among the downstream genes of these CpG markers, some have known functions [such as *ATXN1* as a chromatin-binding factor that represses Notch signaling (*35*), *BMPR1A* as a receptor of transforming growth factor–β (TGF-β) pathway associated with juvenile polyposis syndrome (*36*), *MYO1G* as a master regulator of membrane tension in T cells (*37*)], but the majority of them do not have clear relationships with carcinogenesis and development of cancers, including CRC. On the other hand, the mechanism by which the methylation of specific CpG markers affects the expression of downstream genes is very complex. Some transcription factors can even preferentially recognize methylated CpG and activate more than 100 genes (*38*). Further investigation of the underlying functional mechanism of these CpG markers might deepen our understanding of the origin of CRC and provide potential therapeutic targets.

Several blood-based methylation marker candidates have been proposed for the early detection of CRC, such as *TMEFF2*, *NGFR*, and *SEPT9*, with AUC values for discrimination between CRC and healthy controls of 0.72, 0.70, and 0.80, respectively (*39*). Church *et al.* conducted a large, prospective trial to assess the accuracy of circulating methylated *SEPT9* DNA for detecting CRC in 7941 patients using a commercially available assay. The results showed low sensitivity and specificity of 48.2% and 91.5%, respectively (*22*). In our study, we evaluated the efficacy of colonoscopy and cg10673833 methylation testing for CRC screening in a prospective study. The results showed that cg10673833 methylation testing identified 26 of 29 participants with CRC, with a sensitivity of 89.7% (95% CI, 0.727 to 0.978), a specificity of 86.8% (95% CI, 0.849 to 0.884), and AUC of 0.9 (95% CI, 0.885 to 0.942). For advanced precancerous lesions, the sensitivity was 33.3%, much higher than that of *SEPT9*, another blood methylation marker used in CRC screening, whose sensitivity was only 11.2%

(*22*). The cg10673833 marker performed less well for precancerous lesions, as would be expected; however, in this circumstance, it might be beneficial to use more than just one marker. Our results indicated that cg10673833 was superior to other currently reported cfDNA methylation markers for CRC screening. It is estimated that less than 20% of the eligible population in China has been screened by colonoscopy, mostly due to its inconvenience (*33*). The cfDNA methylation marker described here provides a noninvasive, effective screening tool with likely good compliance for early detection of CRC. The noninvasive screening strategy investigated here may enhance screening adherence and increase participation rates.

However, there are some limitations that need to be emphasized. We identified a CRC special marker panel by comparing CRC tissue DNA methylation data from TCGA and normal blood leukocyte methylation data from an aging study, but inconsistencies in sample types might increase data deviation in marker screening. Second, this was not a randomized controlled study, which might have introduced some amount of selection bias.

Collectively, our findings demonstrated the usefulness of cfDNA methylation markers for diagnosis, prognostication, and surveillance of CRC, with the potential to be used for early detection of asymptomatic patients with CRC. The results of this study offer support for setting up large-scale randomized clinical trials to validate its clinical applicability.

MATERIALS AND METHODS

Study design

This study aimed to identify cfDNA methylation–based biomarkers for the early detection of CRC (*2*). First, we identified differential methylation markers from public CRC and healthy blood DNA methylation datasets. We then tested the markers in a retrospective cfDNA cohort consisting of blood samples from patients with CRC and healthy people. Patients who presented with CRC from stages I to IV were selected and enrolled in this study. The cfDNA methylation data from this cohort were randomly divided into training and validation datasets with a 2:1 ratio to build both diagnostic and prognostic models. To note, variables were selected with a machine learning algorithm before the model construction. The models were evaluated with cross-validation and ROC methods and then compared with the CEA concentrations in the cohort. With the cfDNA methylation profile of those patients, we also built two cfDNA methylation CRC clusters via a modified unsupervised clustering method. We further designed a prospective study to investigate the potential value of cg10673833, which demonstrated high efficiency as a cfDNA methylation marker for malignant lesions and advanced adenomas of the colon and rectum in the high-risk screening population, using colonoscopy as the reference method.

Screening study sample size calculation

On the basis of published reports, the area under the ROC of *SEPT9* or other clinical characteristics for screening CRC was estimated to be 70 to 75% (*22*, *39*). We hypothesized that the diagnostic accuracy rate could be increased to 85% with the introduction of the cd-score. With a two-sided significance level at 0.05, dropout rate of 20%, and nuisance parameter at 0.30, 1284 participants were needed in the prospective screening cohort to ensure a power of 90% to detect the assumed improvement in area under the ROC (according to the PASS 15.0 software of equivalence tests for the difference between two correlated proportions). In this study, we enrolled 1493 high-risk participants into the prospective CRC high-risk cohort, meeting the sample size standard. Considering that the missing data only accounted for a very small proportion (3.6%) of the study participants, all those with missing data were excluded.

Patients and sample collection

Tissue DNA methylation data were obtained from the TCGA (TCGA, TCGA-COAD, and TCGA-READ). Complete clinical, molecular, and histopathological datasets are available at the TCGA website: https://tcga-data.nci.nih.gov/tcga/. Whole-blood DNA methylation profiles from healthy donors were generated in an aging study (GSE40279) (*28*). Of note, the dataset from the TCGA and aging study used the same platform (Illumina 450K) for profiling methylation status. The cfDNA cohort consisted of 801 patients with CRC and 1021 healthy controls. Patients' characteristics and tumor features are summarized in table S1. This cohort was collected from the Sun Yat-sen University Cancer Center in Guangzhou, Xijing Hospital in Xi'an, and the West China Hospital in Chengdu, China. The prospective CRC screening cohort was composed of a total of 16,890 subjects, aged between 45 and 75 years and without CRC-related symptoms, who participated in this study between January 2015 and December 2017.

Prospective CRC screening cohort study design

The CRC screening cohort originated from a subset of the Cancer Screening Program in Urban China or individuals undergoing a screening test for CRC. All the participants were first invited to take a cancer risk assessment by an established Clinical Cancer Risk Score System (see data file S1), containing information on demographic characteristics, smoking history, family history of cancer, height, weight, body mass index (BMI), medical history, health behaviors, and health status. Participants enrolled into the study if they met the following criteria: (i) men and women, (ii) age 45 to 75 years, (iii) never diagnosed with cancer, (iv) consent to receive and complete investigation questionnaire, (v) able and willing to undergo a screening colonoscopy within 90 days of enrollment, and (vi) able and willing to provide plasma samples. Subjects who had a personal history of colorectal neoplasia, digestive cancer, or inflammatory bowel disease; had undergone colonoscopy within the previous 10 years or a barium enema, computed tomographic colonography, or sigmoidoscopy within the previous 5 years; had undergone colorectal resection for any reason other than sigmoid diverticula; or had overt rectal bleeding within the previous 30 days were excluded.

The detailed procedure regarding how the Clinical Cancer Risk Score System defined high-risk CRC can be found elsewhere (*33*). Briefly, the system defined high-risk CRC based on the revised Harvard Risk Index (*33*), which considered the following risk factors: BMI, dietary intake of whole grains, fresh vegetables, processed meat, high-fat diet intake, history of gallstones and chronic colitis, family history of CRC in first-degree relatives, results of previous fecal occult blood test, and history of colonic polyps. Each risk factor was assigned a score by the expert panel (*33*) based on the magnitude of its association with CRC. The cumulative risk scores were calculated and then divided by the average risk score in the general population to get the final individual relative risks. Individuals with relative risks over 1.50 were defined as high-risk for CRC.

Those participants who were deemed to be at high risk for CRC were recommended to undergo colonoscopy and cfDNA methylation tests. Figure 1B shows the analytic sample selection process for the prospective CRC screening cohort. Among the 16,890 subjects,

1493 subjects were identified as high-risk for CRC and were enrolled into the CRC high-risk cohort. All 1493 participants were scheduled for colonoscopy and cfDNA methylation tests within 2 months after risk assessment. Any abnormal findings discovered by colonoscopy were sent to pathologists to confirm whether they were CRC. This project was approved by the institutional review board from ethics committee of Sun Yat-sen University Cancer Center (accession nos. YB2014-11-10 and B2017-019-01).

Identification of methylation markers discriminating between CRC and normal blood

To identify putative markers, we first compared the methylation data derived from CRC tissue DNA from the TCGA and healthy blood from a previous study (*28*), including 459 CRC tissue samples and blood from 754 healthy controls. We used a moderated *t* test with empirical Bayes for shrinking the variance (*40*) and selected the top 1000 significant markers with an adjusted *P* value <0.05 (fig. S8). We then designed the molecular-inversion (padlock) probes corresponding to these 1000 markers for capture-sequencing cfDNA from plasma, and selected 544 markers with a good experimental amplification profile for further analysis. We applied the concept of genetic linkage disequilibrium (LD block) (*26*) to study the degree of comethylation among different DNA strands, with the underlying assumption that DNA sites in close proximity were more likely to be comethylated than distant sites (*24*).

Isolation and methylation profiling of cfDNA

For each sample, cfDNA was extracted from 1.5 ml of plasma using an EliteHealth cfDNA extraction Kit (EliteHealth) according to the manufacturer's recommendations. At least 10 ng of DNA was subjected to bisulphite conversion using an EZ DNA Methylation-Lightning Kit (Zymo Research). DNA methylation rate at each MCB was determined using deep sequencing of bis-DNA captured with molecular inversion probes. To measure the methylation status of a single marker (cg10673833), we adopted a droplet digital polymerase chain reaction (PCR) paradigm featuring a Bio-Rad QX-200 Droplet Reader and an Automated Droplet Generator (AutoDG) on 10 ng of bisulphite-converted DNA. The detailed procedure of bisulphite conversion efficiency assessment, probe design, sequencing analysis, droplet digital PCR, and data processing can be found in the Supplementary Materials and Methods.

Statistical analysis

A logistic regression model was fitted to build the cd-score, and a Cox regression model was fitted to build the cp-score. The ROC curve was adopted to assess the performance of the cd-score–based classifier. The cd-score distribution between clinical categories was examined using the Wilcoxon test because the cd-score was shown to be non-normally distributed using the Shapiro-Wilk test. Kaplan-Meier curves and log-rank tests were used for survival analysis with the dichotomized cp-score, which provided a high-risk and low-risk group assignment relative to the median. We used time-dependent ROC to compare the discrimination performance of the cp-score, AJCC stage, CEA status, primary tumor location, and the combination of all factors. Multivariate Cox regression analysis was performed to assess the effect of potential risk factors upon the survival time. All hypothesis testing in the prognostic analysis section was done in a two-sided manner, with *P* value <0.05 considered to be statistically significant. We used the percentile method to calculate 95% CIs. All analysis was conducted in R software, version 3.4.3 (see Supplementary Materials and Methods).

SUPPLEMENTARY MATERIALS

stm.sciencemag.org/cgi/content/full/12/524/eaax7533/DC1

Materials and Methods

Fig. S1. List of methylation correlated blocks used for cd-score generation.

Fig. S2. ctDNA methylation analysis for predicting tumor burden, staging, and treatment response using a cd-score in patients with CRC.

Fig. S3. The diagnosis efficiency of each marker among the nine markers in the diagnostic model.

Fig. S4. Patient treatment response monitoring with methylation rate of cg10673833.

Fig. S5. Methylation values correlated with treatment outcomes in patients with CRC with serial plasma samples.

Fig. S6. Nomogram for predicting overall survival of patients with CRC.

Fig. S7. Comparison of subtype markers, diagnosis markers, and prognosis markers.

Fig. S8. Unsupervised hierarchical clustering of the top 1000 methylation markers differentially methylated between CRC tumor DNA and normal blood.

Table S1. Clinical characteristics of the entire study cohort.

Table S2. Clinical characteristics of the screening study cohort.

Table S3. Characteristics of the nine methylation markers and their coefficients in diagnosis.

Table S4. Characteristics of the five methylation markers and their coefficients in prognosis.

Table S5. Multivariable Cox regression analysis with covariates including cp-score, gender, age, tumor location, TNM stage, and CEA for overall survival.

Table S6. Characteristics of the 45 methylation markers in ctDNA methylation–based subtyping of CRC.

Table S7. Clinicopathological and molecular associations of subtype groups.

Table S8. Association between ctDNA methylation–based CRC subtypes and CRC prognosis in both the training and validation sets (the same cohort as the prognosis model analysis).

Table S9. Methylation value of cg10673833 in different categories reported by colonoscopy.

Table S10. Positive and negative predictive values of ctDNA methylation test.

Data file S1. Questionnaire for screening patients at high risk of CRC.

References (*41*–*57*)

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30 (2018).
2. R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. S. Meester, A. Barzi, A. Jemal, Colorectal cancer statistics, 2017. *CA Cancer J. Clin.* **67**, 177–193 (2017).
3. R.-M. Feng, Y.-N. Zong, S.-M. Cao, R.-H. Xu, Current cancer situation in China: Good or bad news from the 2018 Global Cancer Statistics? *Cancer Commun.* **39**, 22 (2019).
4. R. Palmqvist, B. Engarås, G. Lindmark, G. Hallmans, B. Tavelin, O. Nilsson, S. Hammarstrom, L. Hafström, Prediagnostic levels of carcinoembryonic antigen and CA 242 in colorectal cancer: A matched case-control study. *Dis. Colon rectum* **46**, 1538–1544 (2003).
5. J. S. Macdonald, Carcinoembryonic antigen screening: Pros and cons. *Semin. Oncol.* **26**, 556–560 (1999).
6. H. Schwarzenbach, D. S. Hoon, K. Pantel, Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437 (2011).
7. C. Zeng, E. K. Stroup, Z. Zhang, B. C. Chiu, W. Zhang, Towards precision medicine: Advances in 5-hydroxymethylcytosine cancer biomarker discovery in liquid biopsy. *Cancer Commun.* **39**, 12 (2019).
8. E. Heitzer, P. Ulz, J. B. Geigl, Circulating tumor DNA as a liquid biopsy for cancer. *Clin. Chem.* **61**, 112–123 (2015).
9. G. Francis, S. Stein, Circulating cell-free tumour DNA in the management of cancer. *Int. J. Mol. Sci.* **16**, 14122–14142 (2015).
10. L. A. Diaz Jr., A. Bardelli, Liquid biopsies: Genotyping circulating tumor DNA. *J. Clin. Oncol.* **32**, 579–586 (2014).
11. P. Mokarram, M. Rismanchi, M. Alizadeh Naeeni, S. Mirab Samiee, M. Paryan, A. Alipour, Z. Honardar, S. Kavousipour, F. Naghibalhossaini, Z. Mostafavi-Pour, A. Monabati, S. V. Hosseni, S. A. Shamsdin, Microsatellite instability typing in serum and tissue of patients with colorectal cancer: Comparing real time PCR with hybridization probe and high-performance liquid chromatography. *Mol. Biol. Rep.* **41**, 2835–2844 (2014).
12. K. Warton, G. Samimi, Methylation of cell-free circulating DNA in the diagnosis of cancer. *Front. Mol. Biosci.* **2**, 13 (2015).
13. J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T. L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H. L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. Diaz Jr., C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).

14. M. Esteller, Epigenetics in cancer. *N. Engl. J. Med.* **358**, 1148–1159 (2008).
15. S. B. Baylin, P. A. Jones, A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).
16. S. B. Baylin, P. A. Jones, Epigenetic determinants of cancer. *Cold Spring Harb. Perspect. Biol.* **8**, a019505 (2016).
17. R. A. Irizarry, C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. Potash, S. Sabunciyan, A. P. Feinberg, The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41**, 178–186 (2009).
18. P. W. Laird, The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* **3**, 253–266 (2003).
19. H. Heyn, M. Esteller, DNA methylation profiling in the clinic: Applications and challenges. *Nat. Rev. Genet.* **13**, 679–692 (2012).
20. C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. M. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, L. Ding, Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
21. M. Toyota, C. Ho, N. Ahuja, K.-W. Jair, Q. Li, M. Ohe-Toyota, S. B. Baylin, J.-P. Issa, Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res.* **59**, 2307–2312 (1999).
22. T. R. Church, M. Wandell, C. Lofton-Day, S. J. Mongin, M. Burger, S. R. Payne, E. Castaños-Vélez, B. A. Blumenstein, T. Rösch, N. Osborn, D. Snover, R. W. Day, D. F. Ransohoff, PRESEPT Clinical Study Steering Committee, Investigators and Study Team, Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* **63**, 317–325 (2014).
23. W. D. Chen, Z. J. Han, J. Skoletsky, J. Olson, J. Sah, L. Myeroff, P. Platzer, S. Lu, D. Dawson, J. Willis, T. P. Pretlow, J. Lutterbaugh, L. Kasturi, J. K. Willson, J. S. Rao, A. Shuber, S. D. Markowitz, Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J. Natl. Cancer Inst.* **97**, 1124–1132 (2005).
24. R. H. Xu, W. Wei, M. Krawczyk, W. Wang, H. Luo, K. Flagg, S. Yi, W. Shi, Q. Quan, K. Li, L. Zheng, H. Zhang, B. A. Caughey, Q. Zhao, J. Hou, R. Zhang, Y. Xu, H. Cai, G. Li, R. Hou, Z. Zhong, D. Lin, X. Fu, J. Zhu, Y. Duan, M. Yu, B. Ying, W. Zhang, J. Wang, E. Zhang, C. Zhang, O. Li, R. Guo, H. Carter, J. K. Zhu, X. Hao, K. Zhang, Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* **16**, 1155–1161 (2017).
25. J. D. Wall, J. K. Pritchard, Haplotype blocks and linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **4**, 587–597 (2003).
26. D. E. Reich, M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, T. Lavery, R. Kouyoumjian, S. F. Farhadian, R. Ward, E. S. Lander, Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
27. X. Hao, H. Luo, M. Krawczyk, W. Wei, W. Wang, J. Wang, K. Flagg, J. Hou, H. Zhang, S. Yi, M. Jafari, D. Lin, C. Chung, B. A. Caughey, G. Li, D. Dhar, W. Shi, L. Zheng, R. Hou, J. Zhu, L. Zhao, X. Fu, E. Zhang, C. Zhang, J. K. Zhu, M. Karin, R. H. Xu, K. Zhang, DNA methylation markers for diagnosis and prognosis of common cancers. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7414–7419 (2017).
28. G. Hannum, J. Guinney, L. Zhao, L. Zhang, G. Hughes, S. Sadda, B. Klotzle, M. Bibikova, J. B. Fan, Y. Gao, R. Deconde, M. Chen, I. Rajapakse, S. Friend, T. Ideker, K. Zhang, Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
29. W. J. Youden, Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
30. P. J. Heagerty, T. Lumley, M. S. Pepe, Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337–344 (2000).
31. M. K. Keck, Z. Zuo, A. Khattri, T. P. Stricker, C. D. Brown, M. Imanguli, D. Rieke, K. Endhardt, P. Fang, J. Brägelmann, R. DeBoer, M. El-Dinali, S. Aktolga, Z. Lei, P. Tan, S. G. Rozen, R. Salgia, R. R. Weichselbaum, M. W. Lingen, M. D. Story, K. K. Ang, E. E. Cohen, K. P. White, E. E. Vokes, T. Y. Seiwert, Integrative analysis of head and neck cancer identifies two biologically distinct HPV and three non-HPV subtypes. *Clin. Cancer Res.* **21**, 870–881 (2015).
32. Q. Zhao, Y. Sun, Z. Liu, H. Zhang, X. Li, K. Zhu, Z.-X. Liu, J. Ren, Z. Zuo, CrossICC: Iterative consensus clustering of cross-platform gene expression data without adjusting batch effect. *Brief. Bioinform.* , (2019).
33. H. Chen, N. Li, J. Ren, X. Feng, Z. Lyu, L. Wei, X. Li, L. Guo, Z. Zheng, S. Zou, Y. Zhang, J. Li, K. Zhang, W. Chen, M. Dai, J. He; group of Cancer Screening Program in Urban China (CanSPUC), Participation and yield of a population-based colorectal cancer screening programme in China. *Gut* **68**, 1450–1457 (2019).
34. J. Guinney, R. Dienstmann, X. Wang, A. de Reynies, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B. M. Bot, J. S. Morris, I. M. Simon, S. Gerster, E. Fessler, E. M. F. De Sousa, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G. C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J. W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S. H. Friend, P. Laurent-Puig, J. P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, S. Tejpar, The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
35. X. Tong, H. Gui, F. Jin, B. W. Heck, P. Lin, J. Ma, J. D. Fondell, C.-C. Tsai, Ataxin-1 and Brother of ataxin-1 are components of the Notch signalling pathway. *EMBO Rep.* **12**, 428–435 (2011).
36. T. Hirschhorn, M. Levi-Hofman, O. Danziger, N. I. Smorodinsky, M. Ehrlich, Differential molecular regulation of processing and membrane expression of Type-I BMP receptors: Implications for signaling. *Cell. Mol. Life Sci.* **74**, 2645–2662 (2017).
37. A. Gérard, G. Patino-Lopez, P. Beemiller, R. Nambiar, K. Ben-Aissa, Y. Liu, F. J. Totah, M. J. Tyska, S. Shaw, M. F. Krummel, Detection of rare antigen-presenting cells through T cell-intrinsic meandering motility, mediated by Myo1g. *Cell* **158**, 492–505 (2014).
38. J. Wan, Y. Su, Q. Song, B. Tung, O. Oyinlade, S. Liu, M. Ying, G.-l. Ming, H. Song, J. Qian, H. Zhu, S. Xia, Methylated *cis*-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. *eLife* **6**, e20068 (2017).
39. C. Lofton-Day, F. Model, T. Devos, R. Tetzner, J. Distler, M. Schuster, X. Song, R. Lesche, V. Liebenberg, M. Ebert, B. Molnar, R. Grutzmann, C. Pilarsky, A. Sledziewski, DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin. Chem.* **54**, 414–423 (2008).
40. G. K. Smyth, in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (Springer, 2005), pp. 397–420.
41. S. B. Edge, C. C. Compton, The American Joint Committee on Cancer: The 7th edition of the AJCC *Cancer Staging Manual* and the future of TNM. *Ann. Surg. Oncol.* **17**, 1471–1474 (2010).
42. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B. Methodol.* **58**, 267–288 (1996).
43. R. Diaz-Uriarte, GeneSrF and varSelRF: A web-based tool and R package for gene selection and classification using random forest. *BMC Bioinformatics* **8**, 328 (2007).
44. E. F. Harrell, K. L. Lee, D. B. Mark, Tutorial in biostatistics multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* **15**, 361–387 (1996).
45. E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, J. Verweij, New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
46. L. Wasserman, K. Roeder, High dimensional variable selection. *Ann. Stat.* **37**, 2178–2201 (2009).
47. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
48. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
49. P. Blanche, J. F. Dartigues, H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).
50. M. S. Schröder, A. C. Culhane, J. Quackenbush, B. Haibe-Kains, *survcomp*: An R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* **27**, 3206–3208 (2011).
51. M. D. Wilkerson, D. N. Hayes, ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
52. G. J. Porreca, K. Zhang, J. B. Li, B. Xie, D. Austin, S. L. Vassallo, E. M. LeProust, B. J. Peck, C. J. Emig, F. Dahl, Y. Gao, G. M. Church, J. Shendure, Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931–936 (2007).
53. D. Diep, N. Plongthongkum, A. Gore, H.-L. Fung, R. Shoemaker, K. Zhang, Library-free methylation sequencing with bisulfite padlock probes. *Nat. Methods* **9**, 270–272 (2012).
54. J. Deng, R. Shoemaker, B. Xie, A. Gore, E. M. LeProust, J. Antosiewicz-Bourget, D. Egli, N. Maherali, I.-H. Park, J. Yu, G. Q. Daley, K. Eggan, K. Hochedlinger, J. Thomson, W. Wang, Y. Gao, K. Zhang, Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **27**, 353–360 (2009).
55. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
56. P. A. Jones, S. B. Baylin, The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).
57. M. Yu, K. T. Carter, K. W. Makar, K. Vickers, C. M. Ulrich, R. E. Schoen, D. Brenner, S. D. Markowitz, W. M. Grady, MethyLight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. *Epigenetics* **10**, 803–809 (2015).

Acknowledgments: The results presented here are, in part, based on data generated by the TCGA Research Network (http://cancergenome.nih.gov/). We thank the staff of the Ruihua Xu and Kang Zhang laboratories for technical assistance and helpful discussions. We apologize for any omission of reference citations owing to space limitations. **Funding:** This research was supported by the National Key R&D Program of China (2018YFC1313300 to R.-h.X.), the Key Project of Natural Science Foundation of China (no. 81930065 to R.-h.X.), the National Natural Science Foundation of China (no. 81871985 to W. Wei), the Natural Science Foundation of

Guangdong Province (2017A030313485 and 2014A030312015 to R.-h.X.), the Science and Technology Program of Guangdong (2015B020232008 to H.L.), the Science and Technology Program of Guangzhou (201508020250 to H.L., and 201604020003 and 2019B020227002 to R.-h.X.), and the Fundamental Research Funds for the Central Universities (17ykpy82 to H.L.).
Author contributions: H.L. contributed to the study design, experiment performance, data collection and interpretation, statistical analysis, and drafting of the manuscript. Q.Z. contributed to the data collection, statistical analysis, data interpretation, and drafting of the manuscript. W. Wei contributed to the study design, experiment performance, data collection and interpretation, statistical analysis, and drafting of the manuscript. L.Z. contributed to the data collection, statistical analysis, data interpretation, and drafting of the manuscript. S.Y. and G.L. contributed to the experiment performance, data collection and interpretation, and drafting of the manuscript. W. Wang, H.S., H.P., H.M., and Z. Zuo contributed to the sample collection and experiment performance. Z.L. and C.L. contributed to the data collection and interpretation and statistical analysis. C.X., Z. Zeng, W.L., X.H., Y.L., S.C., G.X., and W.L. contributed to the sample collection, data collection, and prospective screening study. S.G. contributed to the data collection and interpretation and the critical review of the manuscript. K.Z. contributed to the study design, data collection and interpretation, statistical analysis, and critical review of the manuscript. R.-h.X. contributed to the study design, data collection and interpretation, statistical analysis, prospective screening study, and drafting of the manuscript. R.-h.X. is the principal investigator of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. **Competing interests:** The authors (H.L., L.Z., R.-h.X., K.Z., Q.Z., and W. Wei) have filed two patents related to our findings in this paper: application no. 20180094325 (title: Solid Tumour Methylation Markers and Uses Thereof) and application no. 201810820758.8 (title: Gene Methylation Panel for Predicting the Efficacy and Prognosis of Colorectal Cancer). All other authors declare that they have no competing interests. **Data and materials availability:** All data associated with this study are present in the paper or Supplementary Materials. The raw sequencing data reported in this paper have been deposited in the Sequence Read Archive (SRA) database, under accession number PRJNA574555, which can be accessed at http://www.ncbi.nlm.nih.gov/bioproject/574555. The clinical and beta value matrix data in this paper have been uploaded and locked onto the Research Data Deposit with RDD number RDDB2019000675.

Submitted 20 April 2019
Accepted 4 October 2019
Published 1 January 2020
10.1126/scitranslmed.aax7533

Citation: H. Luo, Q. Zhao, W. Wei, L. Zheng, S. Yi, G. Li, W. Wang, H. Sheng, H. Pu, H. Mo, Z. Zuo, Z. Liu, C. Li, C. Xie, Z. Zeng, W. Li, X. Hao, Y. Liu, S. Cao, W. Liu, S. Gibson, K. Zhang, G. Xu, R.-h. Xu, Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. *Sci. Transl. Med.* **12**, eaax7533 (2020).

Science Translational Medicine

Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer

Huiyan Luo, Qi Zhao, Wei Wei, Lianghong Zheng, Shaohua Yi, Gen Li, Wenqiu Wang, Hui Sheng, Hengying Pu, Haiyu Mo, Zhixiang Zuo, Zexian Liu, Chaofeng Li, Chuanbo Xie, Zhaolei Zeng, Weimin Li, Xiaoke Hao, Yuying Liu, Sumei Cao, Wanli Liu, Sarah Gibson, Kang Zhang, Guoliang Xu and Rui-hua Xu

Sci Transl Med **12**, eaax7533.
DOI: 10.1126/scitranslmed.aax7533

Methylation marks the spot

The detection of circulating tumor DNA in the blood is a noninvasive method that may help detect cancer at early stages if one knows the correct markers for evaluation. Luo *et al.* analyzed methylation patterns in blood samples from multiple large cohorts of patients, including a prospective screening cohort of people at high risk of colorectal cancer. The authors identified and validated a methylation-based diagnostic score to help distinguish patients with colorectal cancer from healthy controls, as well as a prognostic score that correlated with patients' survival. One methylation marker in particular appeared to have high sensitivity and specificity for identifying patients with cancer.