Check for updates

# ARTICLE

Clinical Studies

# A DNA methylation signature for the prediction of tumour recurrence in stage II colorectal cancer

Min Li [1,2,10], Congcong Zhu [3,4,10], Ying Xue [1,2,10], Changhong Miao [1,5], Ruiping He [1,2], Wei Li [6], Baolong Zhang [6], Wenqiang Yu [6], Xingxu Huang [7], Minzhi Lv [2,8 ✉], Ye Xu [3,4 ✉] and Qihong Huang [1,2,9 ✉]

**BACKGROUND:** A major challenge in stage II colorectal carcinoma is to identify patients with increased risk of recurrence. Biomarkers that distinguish patients with poor prognosis from patients without recurrence are currently lacking. This study aims to develop a robust DNA methylation classifier that allows the prediction of recurrence and chemotherapy benefit in patients with stage II colorectal cancer. We performed a genome-wide DNA methylation capture sequencing in 243 stage II colorectal carcinoma samples and identified a relapse-specific DNA methylation signature consisting of eight CpG sites.
**METHODS:** Two hundred and forty-three patients with stage II CRC were enrolled in this study. In order to select differential methylation sites among recurrence and non-recurrence stage II CRC samples, DNA methylation profiles of 62 tumour samples including 31 recurrence and 31 nonrecurrence samples were analysed using the Agilent SureSelectXT Human Methyl-Seq, a comprehensive target enrichment system to analyse CpG methylation. Pyrosequencing was applied to quantify the methylation level of candidate DNA methylation sites in 243 patients. Least absolute shrinkage and selection operator (LASSO) method was employed to build the disease recurrence prediction classifier.
**RESULTS:** We identified a relapse-related DNA methylation signature consisting of eight CpG sites in stage II CRC by DNA methylation capture sequencing. The classifier showed significantly higher prognostic accuracy than any clinicopathological risk factors. The Kaplan–Meier survival curve showed an association of high-risk score with poor prognosis. In multivariate analysis, the signature was the most significant prognosis factor, with an HR of 2.80 (95% CI, 1.71–4.58, $P < 0.001$). The signature could identify patients who are suitable candidates for adjuvant chemotherapy.
**CONCLUSIONS:** An eight-CpG DNA methylation signature is a reliable prognostic and predictive tool for disease recurrence in patients with stage II CRC.

## BACKGROUND

Colorectal cancer (CRC) is the third most common cancers and the second leading cause of cancer-related deaths worldwide [1]. Surgery and chemotherapy are the most common therapeutics for CRC. Surgery is used for stage I CRC with 5-year survival rate approximately 90% [2]. Postoperative adjuvant chemotherapy is a standard treatment for patients with stage III CRC, which improves survival and reduces risk in disease recurrence [3–5]. However, although a large number of clinical trials have been carried out, whether adjuvant chemotherapy is beneficial to stage II CRC is still controversial [6–8]. About one-quarter of CRC patients are stage II CRC, and approximately 15–25% of these patients suffer from disease recurrence [9, 10]. Current guidelines recommend adjuvant

chemotherapy for stage II CRC with clinicopathological risk factors including T4 stage, high tumour grade, <12 lymph nodes harvested, positive or unknown margins, intestinal obstruction or perforation, lymphovascular and perineural invasion [8]. However, those clinicopathological risk factors can not clearly classify which patients have a high or low risk of recurrence, and can not precisely predict which patients benefit from adjuvant chemotherapy. One reason that adjuvant chemotherapy failed to show a survival benefit for patients at stage II CRC could be the lack of biologic factors predicting their recurrence, making it more difficult to draw meaningful clinical conclusions. Hence, the role of biomarkers is becoming increasingly important for the future personalised treatment options. MSI status has proven accurate and effective in

[1]Cancer Center, Zhongshan Hospital, Fudan University, 180 Fenglin Road, 200032 Shanghai, P. R. China. [2]Institute of Clinical Sciences, Zhongshan Hospital, Fudan University, 180 Fenglin Road, 200032 Shanghai, P. R. China. [3]Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, 270 Dong'An Road, 200032 Shanghai, P. R. China. [4]Department of Oncology, Shanghai Medical College, Fudan University, 270 Dong'An Road, 200032 Shanghai, P. R. China. [5]Department of Anesthesiology, Zhongshan Hospital, Fudan University, 180 Fenglin Road, 200032 Shanghai, P. R. China. [6]Laboratory of RNA Epigenetics, Institute of Biomedical Sciences, Fudan University, 130 Dong'An Road, 200032 Shanghai, P. R. China. [7]School of Life Science and Technology, ShanghaiTech University, 201210 Shanghai, P. R. China. [8]Department of Biostatistics, Clinical Research Unit, Zhongshan Hospital, Fudan University, 180 Fenglin Road, 200032 Shanghai, P. R. China. [9]Shanghai Respiratory Research Institute, 180 Fenglin Road, 200032 Shanghai, P. R. China. [10]These authors contributed equally: Min Li, Congcong Zhu, Ying Xue. ✉email: lv.minzhi@zs-hospital.sh.cn; yexu@shmu.edu.cn; qihong_huang@fudan.edu.cn
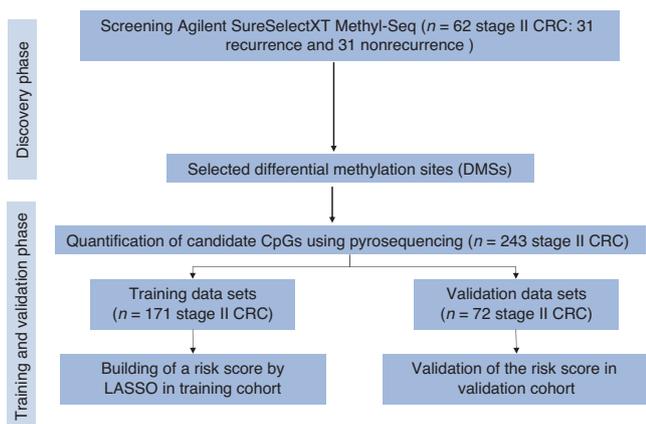
**Fig. 1 Study flowchart.** Differential methylation sites were selected in the discovery phase and validated in the training and validation phase.

identifying approximately 10–15% of patients who do not benefit from 5-FU adjuvant chemotherapy [11, 12]. In addition, some molecular biomarkers have been reported, for example, BRAF mutation, CDX2 gene expression, tumour-infiltrating T cell counts, non-coding RNAs, Oncotype DX colon and so on, but there is still a lack of sufficient evidence resulting in been not applied in clinic [9, 13–22]. Recently a randomised controlled clinical study DYNAMIC was reported to show that circulating tumour DNA (ctDNA) guides adjuvant therapy in stage II colon cancer. Compared to standard-management group, ctDNA-guided approach reduced the use of adjuvant therapy without compromising recurrence risk [23]. Despite of these efforts, there is still an urgent need for effective biomarkers with prognostic and predictive value to classify the stage II CRC patients.

Tumourigenesis of CRC is a multistep process involving the progression from adenomas to adenocarcinomas, and this complex process harbours the accumulation of genetic and epigenetic alterations [24–27]. Studies have shown that epigenetic changes including DNA methylation, histone modifications, small non-coding RNAs, chromatin remodelling, etc. play important roles in the development and progression of diseases [28, 29]. These epigenetic modifications are heritable and reversible, involving in conferring cellular plasticity by establishing specific cellular states and responding to changes in microenvironment. Aberrant epigenetic modifications have been shown to play important roles in colorectal cancer [30, 31].

Among the epigenetic modifications, DNA methylation has received a lot of attention because of its stability and heritability [32]. In eukaryotes, DNA methylation occurs at C5 in CpG dinucleotides termed as CpG sites, and CpG islands are regions with a high frequency of CpG sites. CpG island methylator phenotype (CIMP) has been described in CRC, which is characterised by a high frequency of aberrant CpG island methylation [33]. CIMP-positive CRC are associated with the clinical features of female, older age, right side colon, MSI-H status, and BRAF mutations [34, 35]. Aberrant DNA methylation results in dysregulation of various genes and occurs in all stages of cancer, including initiation, growth, and metastasis [36–38]. DNA methylation inhibitors have been used clinically as antitumour drugs, suggesting the importance of DNA methylation in cancers [39]. Furthermore, changes in DNA methylation in cancer have been regarded as promising targets for the development of powerful diagnostic, prognostic, and predictive biomarkers. So far, 14 DNA methylation-based biomarkers have been translated into a commercially available clinical test [40].

The lack of reliable biomarkers for the identification of patients with stage II CRC who are at high risk for relapse has made it

difficult to make clinical decision in treatment. There is an urgent need to identify prognostic and predictive biomarkers of stage II CRC. To investigate potential DNA methylation sites associated with stage II CRC recurrence, we used high-throughput sequencing to screen 2,521,730 CpG methylation sites in recurrence and nonrecurrence stage II CRC tissue samples in this study. We developed a classifier based on the characteristics of eight DNA methylation sites to predict the recurrence of stage II CRC and potentially guide adjuvant therapy in the current study.

## METHODS
### Patients
This study was approved by the Institutional Review Board of Fudan University Shanghai Cancer Center and was performed in accordance with the Declaration of Helsinki. This study recruited 243 patients of stage II CRC with clinicopathological characteristics and follow-up information from January 2010 to December 2015 at Fudan University Shanghai Cancer Center. Patients with neoadjuvant therapy were excluded in this study. Written informed consent for the sample collection was obtained. Tumour tissue samples were collected at surgery and conserved in RNAlater. All patients were treated and followed up according to the Chinese Society of Clinical Oncology guideline.

### Procedures
In the discovery phase, 62 stage II CRC patients, including 31 recurrence and 31 nonrecurrence, who did not receive adjuvant chemotherapy were selected to evaluate the genome-wide DNA methylation profile by Agilent SureSelectXT Human Methyl-Seq, a target enrichment system which delivers more information than methylation microarrays and reduces costs compared with whole genome bisulfite sequencing. Differential methylation sites (DMSs) between recurrence and nonrecurrence samples were identified and the candidate CpG sites were selection based on (1) methylation-level differences; (2) statistical significance; (3) AUC values; and (4) more than one differential CpG site within 100 base pairs. We then quantified the methylation level of candidate CpG sites in 243 stage II CRC patients by pyrosequencing to build and validate a predictive signature of recurrence. Least absolute shrinkage and selection operator (LASSO) method was applied to build the disease recurrence prediction classifier (Fig. 1).

### Methyl-seq and data analysis
Sixty-two of the 243 samples were selected for Methyl-seq. Genomic DNA from tissue samples were extracted by Qiagen genomic DNA extraction kit (QIAGEN, catalogue no. 69506), and all the samples passed the quality control tests. Agilent SureSelectXT Human Methyl-Seq, a comprehensive target enrichment system to analysis CpG methylation, were used to quantify CpG methylation levels according to instruction. For each CpG site, reads represented either methylation and unmethylation were obtained by Bismark, a software package to map and determine the methylation state of sequencing reads [41]. The reference genome is hg19 human genome. The methylation level of each CpG site is represented as a $\beta$ value, calculated as the ratio of the number of reads supporting methylation to the sum of the number of reads supporting both methylation and unmethylation. The $\beta$ values range from 0 (unmethylated) to 1 (fully methylated). $\Delta\beta$ was calculated using the following formula: $\Delta\beta =$ (mean $\beta$ nonrecurrence samples) − (mean $\beta$ recurrence samples).

### Bisulfite-PCR pyrosequencing
Pyrosequencing performed on the PyroMark Q96 instrument (QIAGEN, Hilden, Germany) to analyse candidate CpGs in 243 samples. In all, 100 ng genomic DNA were used for bisulfite treatment (ZYMO Research, catalogue no. D5006), and the product was used as PCR template subsequently. A detailed pyrosequencing protocol has been described previously [42]. The genomic sequences and primers designed for CpG sites are listed in Supplementary Table S6.

### Statistical analysis
Continuous variables were presented as mean (standard deviation) or median (interquartile range), and categorical variables were described as frequencies and percentages. Student's t test or the Mann–Whitney U-test
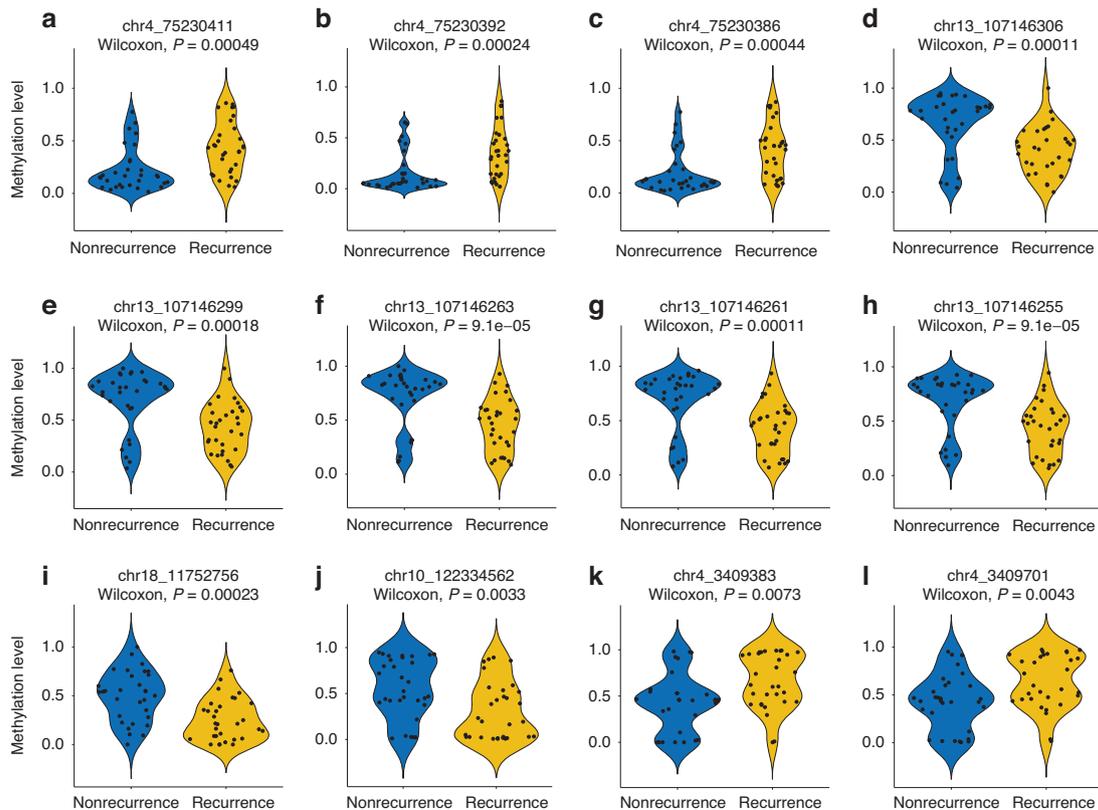
**Fig. 2 CpG methylation sites are significantly associated with recurrence in stage II CRC. a–l** $\beta$ values of candidate sites among nonrecurrence ($n = 31$) and recurrence ($n = 31$) stage II CRC samples.

were used to identify significant differences between groups. The LASSO model was used to select prognostic markers of all the relapse-associated CpGs in the training data sets, and constructed a classifier for predicting the recurrence of patients with stage II CRC. We conducted the receiver operating characteristic (ROC) analysis and calculated the area under the curve to measure prognostic or predictive accuracy.

Disease recurrence was defined as whether the patient developed recurrence before the follow-up deadline. Disease-free survival (DFS) was defined as the time from the data of resection to the data of confirmed tumour relapse or death. OS was calculated from the data of surgery to death. We used the Kaplan–Meier method to analyses the correlation between the classifier and DFS or OS, and the log-rank test to compare survival curves. The univariable and multivariable Cox regression model were used to assess the association between the classifier and survival. Common clinical features were adjusted in the multivariable model, and Cox regression coefficients were used to generate nomograms. Calibration plots were generated to explore the performance characteristics of the nomograms. Calibration is useful for assessing whether actual outcomes approximate predicted outcomes for every nomogram. The x-axis represents the prediction calculated with use of the nomogram, and the y-axis represents the actual freedom from cancer recurrence for our patients. The 45-degree line represents the performance of an ideal nomogram, in which predicted outcome perfectly corresponds with actual outcome. In a well-calibrated model, points are close to the 45-degree line.

All statistical significance was set at P value of <0.05. Statistical analyses were performed using the R software version 3.6.0 (https://www.r-project.org/).

## RESULTS
### Patient clinical characteristics
A total of 243 patients diagnosed with clinical stage II CRC were enrolled in this study, and patients treated with neoadjuvant therapy were excluded. The clinical characteristics of the patients were listed in Supplementary Table S1. The median

age was 62 years (range 25–87) and 38.7% of the patients were female.

### Selection of recurrence-related candidate CpG sites
We first evaluated a genome-wide DNA methylation profile of a discovery cohort of 62 patients with stage II colorectal tumours, including 31 recurrence and 31 nonrecurrence by the Agilent SureSelectXT Human Methyl-Seq. We analysed a total of 2,521,730 CpG methylation sites, and identified 294 CpG sites as the most differential methylation sites (DMSs) between recurrence and nonrecurrence tissues (Supplementary Table S2). Most of the differently methylation sites (220 of 294 sites, 75%) were hypomethylated in recurrence sample. Then, the following 12 CpG sites were selected as top candidates CpGs based on the selection criteria mentioned above: chr13_107146306, chr13_107146299, chr13_107146263, chr13_107146261, chr13_107146255, chr4_75230386, chr4_75230392, chr4_75230411, chr18_11752756, chr10_122334562, chr4_3409383, chr4_3409701 (Fig. 2). Using unsupervised hierarchical clustering, these 12 DMSs successfully separated the 62 patients into 2 discrete clusters (Supplementary Fig. S1).

### Building and validation of a predictive signature of recurrence in stage II CRC
In order to obtain a more reliable signature of recurrence in stage II CRC, 243 stage II CRC patients were included to build and validate the predictive model. Samples were randomly divided into a training cohort ($n = 171$, 70%) and a validation cohort ($n = 72$, 30%). We quantified the methylation level of candidate DNA methylation sites in 243 patients by pyrosequencing. When pyrosequencing was used to detect site chr18_11752756, chr10_122334562, chr4_3409383, and chr4_3409701, we also obtained the methylation information of two CpG sites adjacent
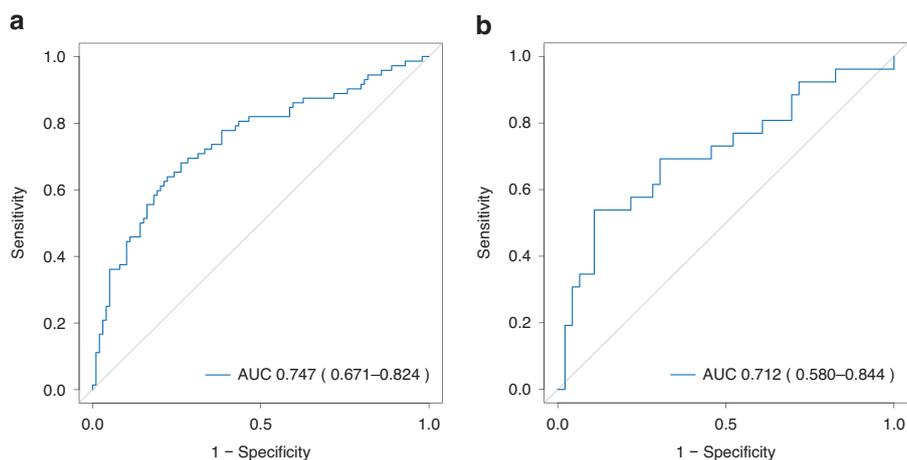
**Fig. 3 DNA methylation signature is a biomarker for recurrence in stage II CRC.** ROC curves in the training (**a**) and validation cohort (**b**).

to each of these sites in the genome: chr18_11752740, chr18_11752738, chr10_122334583, chr10_122334569, chr4_3409388, chr4_3409399, chr4_3409739 and chr4_3409727. Hence, a total of 20 CpG sites comprising of 12 DMSs and 8 adjacent CpG sites were used to generate a methylation signature of recurrence in stage II CRC.

We first determined the association of 20 individual CpG sites with prognosis in the training samples (Supplementary Table S3 and Supplementary Fig. S2). We then determined the combination of these sites in the prediction of disease progression. Eight CpG sites were selected by LASSO methods to build a disease recurrence prediction classifier (Supplementary Fig. S3A, B). The risk score combing the eight CpG sites (chr4_75230411, chr13_107146255, chr13_107146299, chr18_11752740, chr10_122334569, chr10_122334562, chr4_3409383, chr4_3409727) was calculated as follows:

Risk score $= 1/1 + \exp$ [$-(1.815 \times$ methylation level of chr4_75230411 $+ 6.533 \times$ methylation level of chr13_107146299-7.637 $\times$ methylation level of chr13_10714625-10.165 $\times$ methylation level of chr18_11752740 $+ 2.384 \times$ methylation level of chr10_122334569-1.673 $\times$ methylation level of chr10_122334562 $+ 1.198 \times$ methylation level of chr4_340938 $+ 3.474 \times$ methylation level of chr4_3409727 $- 1.657$)].

In this formula, methylation levels range from 0 to 1.

With disease recurrence as an event, ROC analysis was performed in the training data sets and validation data sets. The AUC of the training cohort and validation cohort was 0.75 (95% CI: 0.67–0.82) and 0.71 (95% CI: 0.58–0.84) respectively, indicating that the signature potentially represents a robust prognostic biomarker for stage II RCR (Fig. 3a, b).

**The prognostic value of the DNA methylation classifier**
To study the prognostic value of the classifier, the risk score for each patient was calculated. The distributions of risk score and recurrence status of stage II patients in training cohort and validation cohort were analysed. As expected, patients with recurrence had a higher risk score and patients with lower risk score were less likely to relapse (Fig. 4a, b).

All patients in the training cohort were further divided into either a high-risk group or low-risk group based on the median risk score as the cutoff. The Kaplan–Meier survival curve showed that patients in high-risk group had a poorer DFS than the low-risk group (Logrank test: $P$ value <0.0001) (Fig. 4c). Similar results showed that high-risk patients had poorer DFS than low-risk patients in the validation cohort (Logrank test: $P$ value = 0.026) (Fig. 4d).

**DNA methylation classifier is an independent prognostic factor for stage II CRC**
The univariate Cox proportional hazard regression model indicates that the risk score was highly associated with prognosis of DFS (HR = 2.96, 95% CI: 1.84–4.75, $P$ value < 0.001) in training cohort (Supplementary Table S4). In validation cohort, risk score (HR = 2.27, 95% CI: 1.08–4.75, $P$ value = 0.017) was also a predictive factor of DFS (Supplementary Table S5).

To further evaluate the role of the DNA methylation classifier in the prediction of prognosis, we included the CpG methylation risk score and common clinicopathological features including age, gender, TMN stage, adjuvant chemotherapy, tumour grade, lymphovascular invasion, perineural invasion, number of lymph nodes harvested, positive or unknown margins and intestinal obstruction or perforation, in the multivariate Cox regression model. In the training cohort, the model showed that risk score (HR = 2.80, 95% CI: 1.71–4.58, $P$ value < 0.001) was an independent prognostic factors for stage II CRC (Supplementary Table S4). In validation cohort, multivariate Cox analysis showed the risk score was an independent prognostic factor associated with DFS (HR = 2.82, 95% CI: 1.20–6.61, $P$ value = 0.017) (Supplementary Table S5). The methylation classifier was still a statistically significant prognostic factor after clinicopathological risk factors are stratified (Fig. 5 and Supplementary Fig. S4).

The combination of the methylation classifier with clinicopathological risk factors performed slightly better than the methylation classifier alone in predicting disease recurrence, and the AUC was 0.734 (Fig. 6a). The classifier also showed significantly higher accuracy in the prediction of DFS (Fig. 6b, c) and overall survival (Supplementary Fig. S5A, B) than any other clinicopathological risk factors. These results indicated the methylation classifier or the combination of DNA methylation classifier with clinicopathological risk factors can serve as a prognostic tool in predicting disease recurrence, DFS and OS in stage II CRC.

**Association of the risk score with the benefit of adjuvant chemotherapy**
To explore the potential utility of our risk score in treatment decision, we first determined the benefit of adjuvant therapy using current clinical criteria. Adjuvant chemotherapy did not improve DFS in all 243 patients (HR 1.24, 95% CI: 0.83–1.86; $P = 0.3$; Supplementary Fig. S6A). Compared with the non-adjuvant chemotherapy group, the DFS analysis of adjuvant chemotherapy group showed that there was no significant difference between the survival of patients with classifier-defined high-risk combined with T4 stage and other groups after adjuvant chemotherapy (Supplementary Fig. S6B, C), which
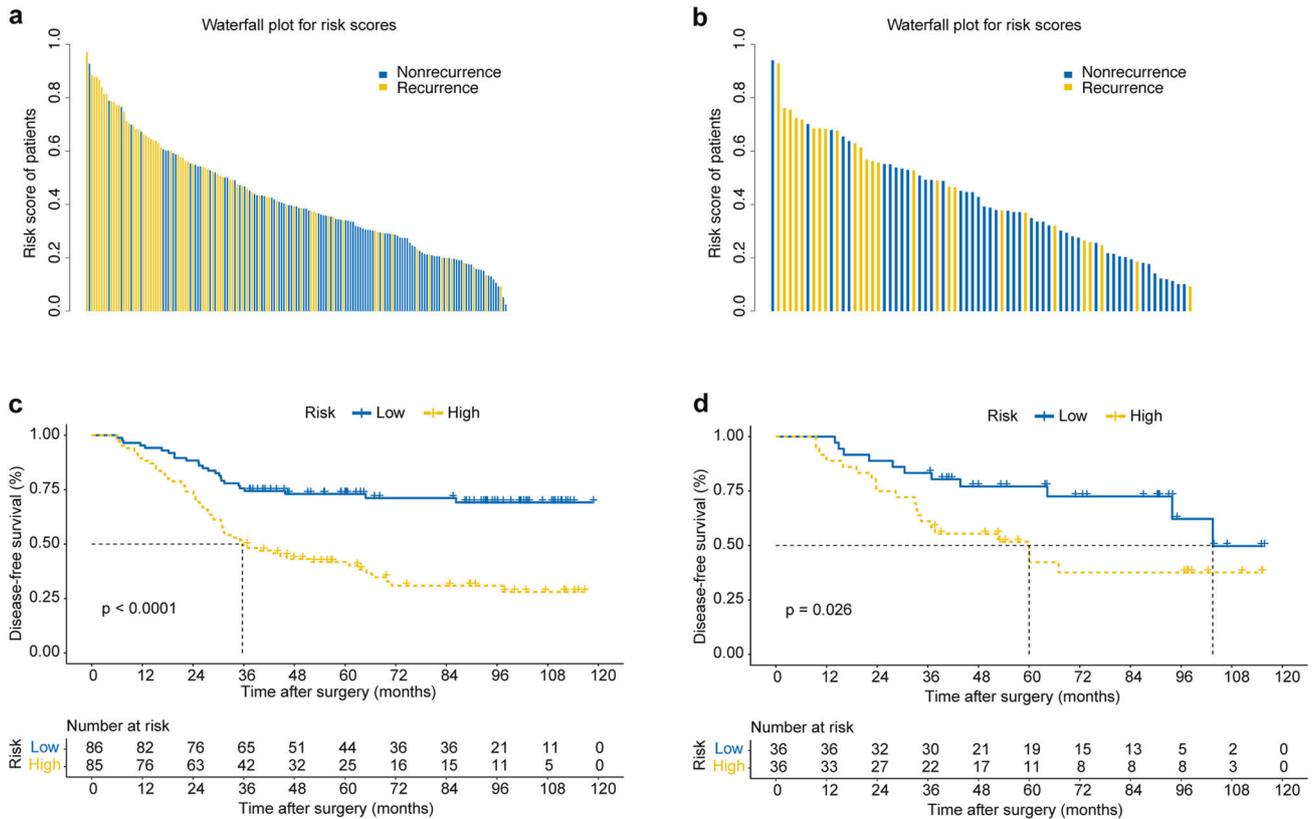
**Fig. 4  Association between the DNA methylation classifier and prognosis.** Waterfall plot of risk score using DNA methylation classifier in training (**a**) and validation (**b**) cohort. DFS Kaplan–Meier survival in the training (**c**) and validation cohort (**d**).

indicated that the patients with classifier-defined high-risk combined T4 stage had a favourable response to adjuvant chemotherapy.

## Construction of nomogram based on methylation classifier

To visualise the prediction model, we established a nomogram based on the eight-methylation site classifier that integrated the classifier and clinicopathological risk factors, including TNM stage, tumour grade, number of lymph nodes examined, positive or unknown margins, intestinal obstruction or perforation, lympho-vascular and perineural invasion (Supplementary Fig. S7A). Among them, DNA methylation risk score had the greatest impact on prognosis, followed by positive margins and tumour grade. Calibration curve showed that the nomogram did well compare with an ideal model represented by the 45-degree line (Supplementary Fig. S7B). The predictive accuracy of the nomogram was calculated by ROC analysis, the AUC of nomogram was 0.734 (Supplementary Fig. S7C). In addition, we also established nomograms with and without adjuvant chemotherapy to predict, the AUC of without and with adjuvant therapy nomograms were 0.868 and 0.7, respectively (Supplementary Fig. S8).

## DISCUSSION

Identification of effective prognostic and predictive biomarkers is critical for risk stratification and guiding treatment of stage II CRC. In this study, we presented the discovery and validation of a relapse-related DNA methylation classifier which predicts tumour recurrence in stage II CRC. In predicting disease recurrence, the AUC of our DNA methylation classifier in training and validation cohort was 0.75 and 0.71 respectively. Our DNA methylation classifier can distinguish high- and low-risk group and the risk

score calculated from the classifier serves as a prognostic indicator for stage II CRC patients. High-risk group defined by the signature had a poorer DFS than the low-risk group regardless of clinicopathological factors. Furthermore, our DNA methylation signature showed significant high prognostic accuracy than any clinicopathological risk factor. Multivariate Cox analysis showed the risk score was an independent prognostic factor associated with DFS. In addition, our classifier can potentially identify patients who are suitable candidates for adjuvant chemotherapy.

The benefits of adjuvant chemotherapy are small in the patients with stage II CRC, thus it is unnecessary to apply chemotherapy in all stage II CRC patients. Over-treatment occurs in 75% of stage II CRC patients who will not relapse [43]. Our data also indicated that adjuvant chemotherapy did not improve DFS in stage II CRC patients using current clinical stratification. According to the updated guideline, adjuvant chemotherapy should not routinely be offered to patients who are at low risk for recurrence [44]. Identification of patients who can benefit from adjuvant therapy is critical in the management of stage II CRC patients. Our analysis showed that the classifier-defined high-risk combined with T4 stage group can significantly benefit from adjuvant chemotherapy. Studies with larger sample size are needed to further validate the utility of the classifier.

Considering DNA methylation is more stable and heritable than RNAs, several studies have reported DNA methylation play important roles in predicting prognosis in lung cancer, prostate cancer, acute myelocytic leukaemia, breast cancer, ovarian cancer and hepatocellular carcinoma [45–50], but it remains unclear whether DNA methylation can act as prognostic factors for stage II CRC. In our research, the high-throughput technology Agilent SureSelectXT Human Methyl-Seq was applied to the analysis CpG methylation. Compared with the methylation beadchips, it determined the methylation level of more than 2.5 million CpG
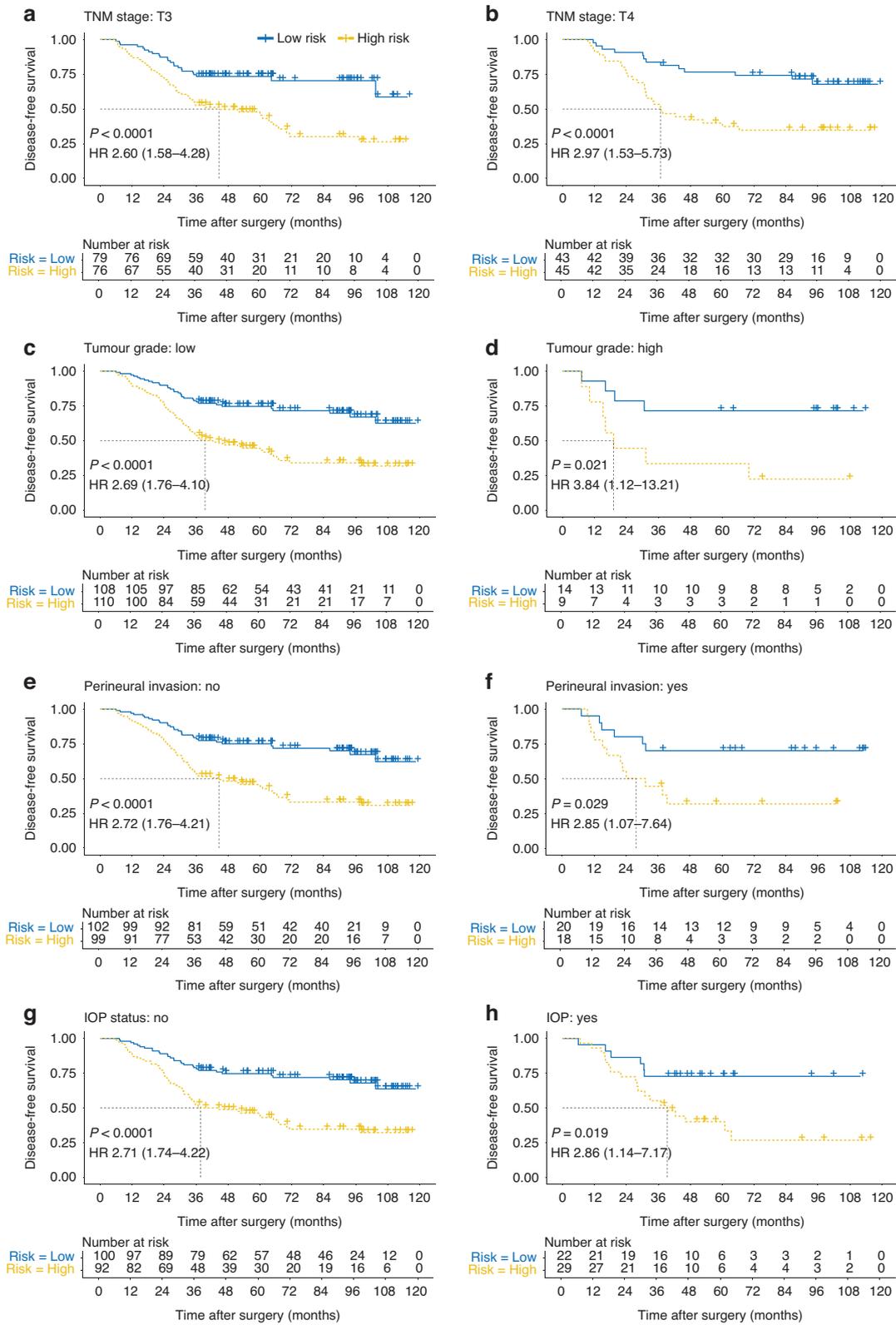
**Fig. 5 DNA methylation signature is an independent prognostic factor in stage II CRC.** Kaplan–Meier survival analysis of 243 patients with stage II CRC using the DNA methylation classifier stratified by clinicopathological risk factors. **a**, **b** TNM stage. **c**, **d** Tumour grade. **e**, **f** Perineural invasion. **g**, **h** Intestinal obstruction of perforation (IOP) status.
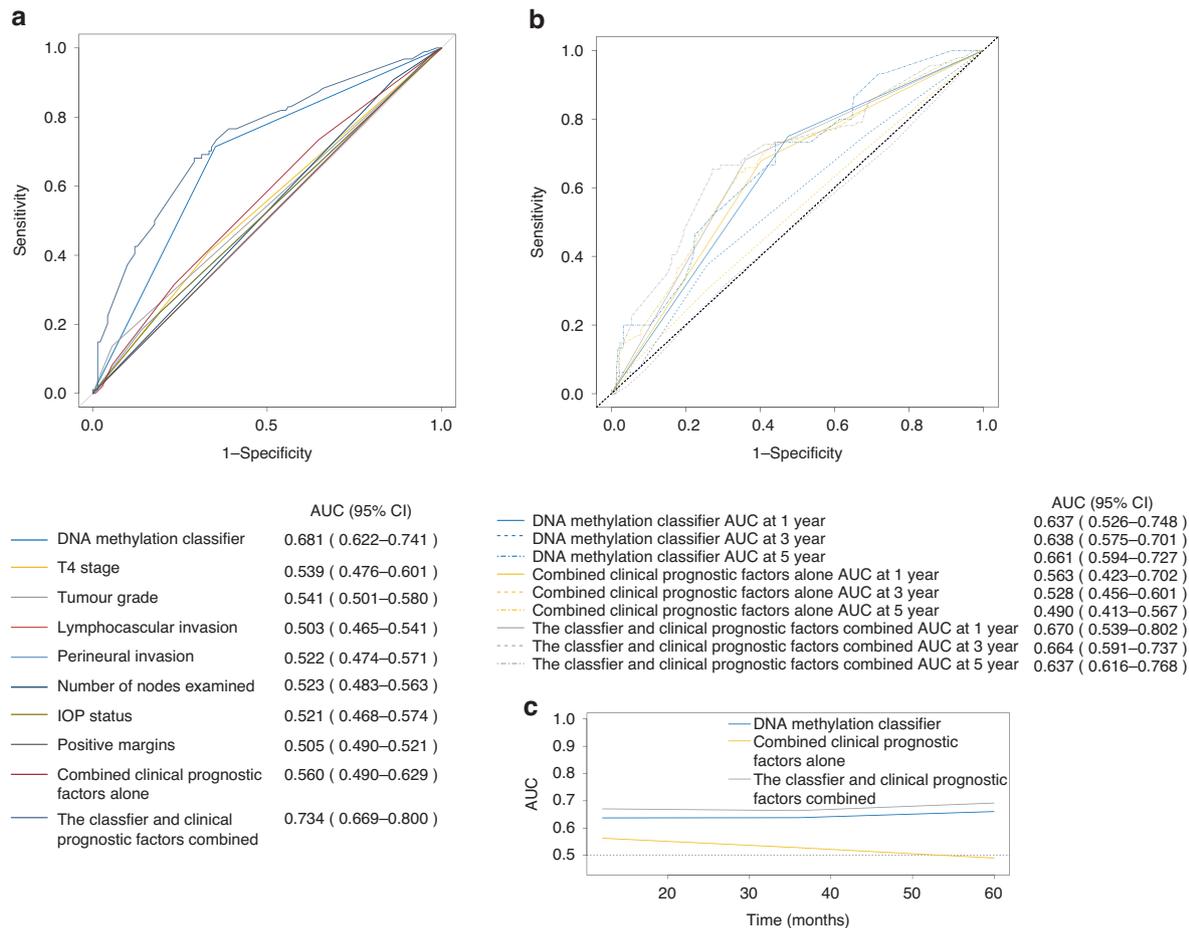
**Fig. 6 DNA methylation classifier is more accurate in the prediction of disease recurrence and disease-free survival than other clinicopathological risk factors. a** Disease recurrence ROC curves of prognosis analysis using the DNA methylation classifier (high risk vs low risk), TNM stage (T4 vs T3), tumour grade (high vs low), IOP (yes vs no), number of lymph node examined (12 or more vs fewer than 12), lymphocasular invasion (yes vs no), perineurial invasion (yes vs no), positive margins (yes vs no), combined clinicopathological risk factors alone, and the combination of classifier with clinicopathological risk factors. **b**, **c** Disease-free survival time-dependent ROC curves comparisons of DNA methylation classifier, clinicopathological risk factors, and the combination of classifier with clinicopathological risk factors.

sites, significantly more than the 850K sites in the methylation beadchips. The eight CpG sites in our classifier were not included in Illumina Infinium 450K Human Methylation Beadchips or Illumina Infinium MethylationEPIC (850K) Beadchips, which were widely used in methylation studies including TCGA project. The identification of these sites that were not discovered previously demonstrated that the methylation information beyond 450K or 850K sites may yield more comprehensive prognostic biomarkers in stage II CRC.

The eight CpG sites in our classifier were related to the methylation in the following genes: *EREG*, *EFNB2*, *GNAL*, *PPAPDC1A* and *RGS12*. Studies of their biological functions showed they are involved in tumour development and therapeutic response. EREG is one of the EGFR ligands. High-expression of *EREG* was correlated with prolonged OS and PFS in patients with CRC [51–54]. *EFNB2* encodes a member of the epherin family and variants in *EFNB2* are associated with overall survival in colorectal cancer patients [55]. Previous studies have shown *EFNB2* and its receptor *EPHB4* are transcriptional upregulated in colon cancer and promote cancer cell proliferation, migration and invasion [56]. *EFNB2* overexpression was correlated with poor prognosis in patients who received chemotherapy or radiotherapy [57]. *GNAL* encodes the α subunit of a heterotrimeric GTP-binding protein. It is significantly correlated with grade and prognosis in glioma [58]. *PPAPDC1A*

encodes a phospholipid phosphatase that converts phosphatidic acids to diacylglycerols. High expression of PPAPDC1A is significant associated with poor OS and DFS in lung cancer [59]. RGS12 expression is lower in various cancer types. Studies have shown its roles in cell proliferation, migration and invasion [60–62]. Genes in our methylation signature have all been found to be involved in tumourigenesis, prognosis and therapy response. However, their functions in CRC recurrence have yet to be elucidated.

Unlike several studies on the prognosis of stage II CRC that included both stage II and III CRC cases [18, 63, 64], our study focused only on stage II CRC. It is unclear whether the recurrence factors in stage II and stage III CRC are the same. The inclusion of only stage II CRC patients may identify stage II specific biomarkers. The Kaplan–Meier analysis of DFS using DNA methylation classifier stratified by the number of lymph nodes examined showed that our classifier has no prognostic value in <12 lymph nodes examined population, who may be unidentified stage III patients due to poor lymph nodes harvest. These results indicated that our classifier is potentially a stage II-specific prognostic biomarker. Whether our classifier can serve as a biomarker in adjuvant therapy guidance need to be further explored. Our results indicated that patients in the classifier-defined high risk combined with T4 stage group may benefit from adjuvant therapy. These

results need to be further validated in larger multi-centre cohorts. In addition, the number of stage II CRC cases included in our study is not large, and the evidence from prospective randomised clinical trials is a necessary step to verify that our classifier is a clinically viable biomarker.

## CONCLUSIONS
In summary, the DNA methylation-based risk score model constructed in this study provided more accurate information on the risk of recurrence compared with the use of conventional clinicopathologic criteria alone, performed effectively in predicting the recurrence risk of stage II CRC patients, possessed good power to discriminate high-risk patients from low-risk patients, and could potentially identify patients who benefit from adjuvant chemotherapy.

## DATA AVAILABILITY
The data that support the findings of this study are available upon request from the corresponding authors.

## REFERENCES
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality world-wide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin. 2021;71:7–33.
3. André T, Boni C, Mounedji-Boudiaf L, Navarro M, Tabernero J, Hickish T, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. N Engl J Med. 2004;350:2343–51.
4. Gramont A. Adjuvant therapy of stage II and III colon cancer. Semin Oncol. 2005;32:11–4.
5. Kelly H, Goldberg RM. Systemic therapy for metastatic colorectal cancer: current options, current evidence. J Clin Oncol. 2005;23:4553–60.
6. Kannarkatt J, Joseph J, Kurniali PC, Al-Janadi A, Hrinczenko B. Adjuvant che-motherapy for stage II colon cancer: a clinical dilemma. J Oncol Pract. 2017;13:233–41.
7. André T, de Gramont A, Vernerey D, Chibaudel B, Bonnetain F, Tijeras-Raballand A, et al. Adjuvant fluorouracil, leucovorin, and oxaliplatin in stage II to III colon cancer: updated 10-year survival and outcomes according to BRAF mutation and mismatch repair status of the MOSAIC study. J Clin Oncol. 2015;33:4176–87.
8. O'Connor ES, Greenblatt DY, LoConte NK, Gangnon RE, Liou JI, Heise CP, et al. Adjuvant chemotherapy for stage II colon cancer with poor prognostic features. J Clin Oncol. 2011;29:3381–8.
9. Zhang JX, Song W, Chen ZH, Wei JH, Liao YJ, Lei J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. Lancet Oncol. 2013;14:1295–306.
10. Feng Q, Chang W, Mao Y, He G, Zheng P, Tang W, et al. Tumor-associated macrophages as prognostic and predictive biomarkers for postoperative adju-vant chemotherapy in patients with stage II colon cancer. Clin Cancer Res. 2019;25:3896–907.
11. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N Engl J Med. 2003;349:247–57.
12. Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, et al. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. J Clin Oncol. 2010;28:3219–26.
13. Dalerba P, Sahoo D, Paik S, Guo X, Yothers G, Song N, et al. CDX2 as a prognostic biomarker in stage II and stage III colon cancer. N Engl J Med. 2016;374:211–22.
14. Parent P, Cohen R, Rassy E, Svrcek M, Taieb J, André T, et al. A comprehensive overview of promising biomarkers in stage II colorectal cancer. Cancer Treat Rev. 2020;88:102059.
15. Gangadhar T, Schilsky RL. Molecular markers to individualize adjuvant therapy for colon cancer. Nat Rev Clin Oncol. 2010;7:318–25.
16. Zhu L, Dong C, Cao Y, Fang X, Zhong C, Li D, et al. Prognostic role of BRAF mutation in stage II/III colorectal cancer receiving curative resection and adjuvant chemotherapy: a meta-analysis based on randomized clinical trials. PLoS ONE. 2016;11:e0154795.
17. Fariña-Sarasqueta A, van Lijnschoten G, Moerland E, Creemers GJ, Lemmens V, Rutten HJT, et al. The BRAF V600E mutation is an independent prognostic factor for survival in stage II and stage III colon cancer patients. Ann Oncol. 2010;21:2396–402.
18. Kim SK, Kim SY, Kim CW, Roh SA, Ha YJ, Lee JL, et al. A prognostic index based on an eleven gene signature to predict systemic recurrences in colorectal cancer. Exp Mol Med. 2019;51:1–12.
19. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet. 2018;391:2128–39.
20. Lanzi A, Pagès F, Lagorce-Pagès C, Galon J. The consensus immunoscore: toward a new classification of colorectal cancer. Oncoimmunology. 2020;9:1789032.
21. Yamanaka T, Oki E, Yamazaki K, Yamaguchi K, Muro K, Uetake H, et al. 12-Gene recurrence score assay stratifies the recurrence risk in stage II/III colon cancer with surgery alone: the SUNRISE Study. J Clin Oncol. 2016;34:2906–13.
22. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med. 2016;8:346ra92.
23. Tie J, Cohen JD, Lahouel K, Lo SN, Wang Y, Kosmider S, et al. Circulating tumor DNA analysis guiding adjuvant therapy in stage II colon cancer. N Engl J Med. 2022;386:2261–72.
24. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61:759–67.
25. Fearon ER. Genetic alterations underlying colorectal tumorigenesis. Cancer Surv. 1992;12:119–36.
26. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology. 2008;135:1079–99.
27. Hanley MP, Hahn MA, Li AX, Wu X, Lin J, Wang J, et al. Genome-wide DNA methylation profiling reveals cancer-associated changes within early colonic neoplasia. Oncogene. 2017;36:5035–44.
28. Esteller M. Epigenetics in cancer. N Engl J Med. 2008;358:1148–59.
29. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150:12–27.
30. Okugawa Y, Grady WM, Goel A. Epigenetic alterations in colorectal cancer: emerging biomarkers. Gastroenterology. 2015;149:1204.e12–25.e12.
31. Oster B, Thorsen K, Lamy P, Wojdacz TK, Hansen LL, Birkenkamp-Demtröder K, et al. Identification and validation of highly frequent CpG island hypermethyla-tion in colorectal adenomas and carcinomas. Int J Cancer. 2011;129:2855–66.
32. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nat Rev Genet. 2012;13:679–92.
33. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci USA. 1999;96:8681–6.
34. Issa JP. CpG island methylator phenotype in cancer. Nat Rev Cancer. 2004;4:988–93.
35. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet. 2006;38:787–93.
36. Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, et al. Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat Genet. 2006;38:149–53.
37. Kong X, Chen J, Xie W, Brown SM, Cai Y, Wu K, et al. Defining UHRF1 domains that support maintenance of human colon cancer DNA methylation and oncogenic properties. Cancer Cell. 2019;35:633.e7–48.e7.
38. Gkountela S, Castro-Giner F, Szczerba BM, Vetter M, Landin J, Scherrer R, et al. Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding. Cell. 2019;176:98.e14–112.e14.
39. Kaminskas E, Farrell A, Abraham S, Baird A, Hsieh LS, Lee SL, et al. Approval summary: azacitidine for treatment of myelodysplastic syndrome subtypes. Clin Cancer Res. 2005;11:3604–8.
40. Koch A, Joosten SC, Feng Z, de Ruijter TC, Draht MX, Melotte V, et al. Analysis of DNA methylation in cancer: location revisited. Nat Rev Clin Oncol. 2018;15:459–66.
41. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.
42. Dong S, Li W, Wang L, Hu J, Song Y, Zhang B, et al. Histone-related genes are hypermethylated in lung cancer and hypermethylated HIST1H4F could serve as a pan-cancer biomarker. Cancer Res. 2019;79:6101–12.
43. Benson AB 3rd, Schrag D, Somerfield MR, Cohen AM, Figueredo AT, Flynn PJ, et al. American Society of Clinical Oncology recommendations on adjuvant che-motherapy for stage II colon cancer. J Clin Oncol. 2004;22:3408–19.
44. Baxter NN, Kennedy EB, Bergsland E, Berlin J, George TJ, Gill S, et al. Adjuvant therapy for stage II colon cancer: ASCO guideline update. J Clin Oncol. 2022;40:892–910.
45. Ding W, Chen G, Shi T. Integrative analysis identifies potential DNA methylation biomarkers for pan-cancer diagnosis and prognosis. Epigenetics. 2019;14:67–80.

46. Zhang H, Song G, Song G, Li R, Gao M, Ye L, et al. Identification of DNA methylation prognostic signature of acute myelocytic leukemia. PLoS ONE. 2018;13:e0199689.

47. Guo W, Zhu L, Yu M, Zhu R, Chen Q, Wang Q. A five-DNA methylation signature act as a novel prognostic biomarker in patients with ovarian serous cystadenocarcinoma. Clin Epigenetics. 2018;10:142.

48. Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Méndez-González J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology. 2015;61:1945–56.

49. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. J Clin Oncol. 2013;31:4140–7.

50. Haldrup C, Mundbjerg K, Vestergaard EM, Lamy P, Wild P, Schulz WA, et al. DNA methylation signatures for prediction of biochemical recurrence after radical prostatectomy of clinically localized prostate cancer. J Clin Oncol. 2013;31:3250–8.

51. Jonker DJ, Karapetis CS, Harbison C, O'Callaghan CJ, Tu D, Simes RJ, et al. Epiregulin gene expression as a biomarker of benefit from cetuximab in the treatment of advanced colorectal cancer. Br J Cancer. 2014;110:648–55.

52. Lin CY, Hsieh PL, Chou CL, Yang CC, Lee SW, Tian YF, et al. High EREG expression is predictive of better outcomes in rectal cancer patients receiving neoadjuvant concurrent chemoradiotherapy. Oncology. 2020;98:549–57.

53. Li XD, Miao SY, Wang GL, Yang L, Shu YQ, Yin YM. Amphiregulin and epiregulin expression in colorectal carcinoma and the correlation with clinicopathological characteristics. Onkologie. 2010;33:353–8.

54. Stahler A, Heinemann V, Giessen-Jung C, Crispin A, Schalhorn A, Stintzing S, et al. Influence of mRNA expression of epiregulin and amphiregulin on outcome of patients with metastatic colorectal cancer treated with 5-FU/LV plus irinotecan or irinotecan plus oxaliplatin as first-line treatment (FIRE 1-trial). Int J Cancer. 2016;138:739–46.

55. Scherer D, Deutelmoser H, Balavarca Y, Toth R, Habermann N, Buck K, et al. Polymorphisms in the angiogenesis-related genes EFNB2, MMP2 and JAG1 are associated with survival of colorectal cancer patients. Int J Mol Sci. 2020;21:5395.

56. Zhu F, Dai SN, Xu DL, Hou CQ, Liu TT, Chen QY, et al. EFNB2 facilitates cell proliferation, migration, and invasion in pancreatic ductal adenocarcinoma via the p53/p21 pathway and EMT. Biomed Pharmacother. 2020;125:109972.

57. Oweida A, Bhatia S, Hirsch K, Calame D, Griego A, Keysar S, et al. Ephrin-B2 overexpression predicts for poor prognosis and response to therapy in solid tumors. Mol Carcinog. 2017;56:1189–96.

58. Zhang D, Zhao J, Han C, Liu X, Liu J, Yang H. Identification of hub genes related to prognosis in glioma. Biosci Rep. 2020;40:5.

59. Zhang X, Zhang L, Lin B, Chai X, Li R, Liao Y, et al. Phospholipid phosphatase 4 promotes proliferation and tumorigenesis, and activates Ca(2+)-permeable cationic channel in lung carcinoma cells. Mol Cancer. 2017;16:147.

60. Wang Y, Wang J, Zhang L, Karatas OF, Shao L, Zhang Y, et al. RGS12 is a novel tumor-suppressor gene in african american prostate cancer that represses AKT and MNX1 expression. Cancer Res. 2017;77:4247–57.

61. Li Y, Liu M, Yang S, Fuller AM, Karin Eisinger-Mathason TS, Yang S. RGS12 is a novel tumor suppressor in osteosarcoma that inhibits YAP-TEAD1-Ezrin signaling. Oncogene. 2021;40:2553–66.

62. Fu C, Yuan G, Yang ST, Zhang D, Yang S. RGS12 represses oral cancer via the phosphorylation and SUMOylation of PTEN. J Dent Res. 2021;100:522–31.

63. Ahluwalia P, Mondal AK, Bloomer C, Fulzele S, Jones K, Ananth S, et al. Identification and clinical validation of a novel 4 gene-signature with prognostic utility in colorectal cancer. Int J Mol Sci. 2019;20:3818.

64. Kandimalla R, Gao F, Matsuyama T, Ishikawa T, Uetake H, Takahashi N, et al. Genome-wide discovery and identification of a novel miRNA signature for recurrence prediction in stage II and III colorectal cancer. Clin Cancer Res. 2018;24:3867–77.

## AUTHOR CONTRIBUTIONS

M Li: data curation, formal analysis, methodology, validation, writing; CZ: data curation, formal analysis; YX: data curation, formal analysis; CM: funding acquisition, resources; RH: data curation; WL: data curation; BZ: data curation; WY: funding acquisition, resources; XH: funding acquisition, resources; M Lv: formal analysis, validation, writing—original draft, writing—review and editing; YX: conceptualisation, formal analysis, funding acquisition, resources, supervision, writing—review and editing; QH: conceptualisation, formal analysis, funding acquisition, resources, supervision, writing—original draft, writing—review and editing.

## FUNDING

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Institutional Review Board of the Fudan University Shanghai Cancer Center. All included patients gave their written consent to participate in this study.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41416-023-02155-8.

**Correspondence** and requests for materials should be addressed to Minzhi Lv, Ye Xu or Qihong Huang.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.