# Genomewide methylation profiling identifies a novel gene signature for patients with synchronous colorectal cancer

Yasuyuki Okada[1,2#], Fuduan Peng[3#], José Perea[4], Luis Corchete[5], Luis Bujanda[6], Wei Li[3#*], and Ajay Goel[1,7#*]

[1] Department of Molecular Diagnostics and Experimental Therapeutics, Beckman Research Institute of City of Hope, Monrovia, CA, USA.

[2] Department of Gastroenterology and Oncology, Tokushima University Graduate School, Tokushima, Japan

[3] Department of Biological Chemistry, School of Medicine, University of California Irvine, Irvine, CA, USA.

[4] Surgery Department, Fundación Jiménez Díaz University Hospital, Madrid, Spain.

[5] Hematology Department, University Hospital of Salamanca, Salamanca, Spain. Institute of Biomedical Research of Salamanca (IBSAL). Cancer Research Center (CiC-IBMCC, CSIC/USAL). Center for Biomedical Research in Network of Cancer (CIBERONC).

[6] Gastroenterology Department, Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Universidad del País Vasco (UPV/EHU), San Sebastián, Spain.

[7] City of Hope Comprehensive Cancer Center, Duarte, CA, USA.

[#] These authors equally contributed to this work.

Running title: Methylation signature for synchronous CRC

*Corresponding Authors: Ajay Goel, Professor and Chair, Department of Molecular Diagnostics and Experimental Therapeutics, Beckman Research Institute of City of Hope Comprehensive Center; 1218 S. Fifth Avenue, Suite 2226, Biomedical Research Center, Monrovia, CA 91016; Phone: 626-218-3452; Email: ajgoel@coh.org; and Wei Li, wei.li@uci.edu.

Keywords: synchronous colorectal cancer, gene methylation, predictive biomarker

Abbreviations: AUC, area under the curve; CI, confidence interval; CIMP, CpG island methylation phenotype; CIN, chromosomal instability; CRC, colorectal cancer; CT, computed tomography; DMP,

differentially methylated CpG site; DMR, differentially methylated region; FFPE, formalin-fixed paraffin-embedded; MCRC, metachronous colorectal cancer; MSI, microsatellite instability; OR, odds ratio; OS, overall survival; PMR, percentage of methylated reference; RFS, relapse-free survival; ROC, receiver operating characteristic; SoCRC, solitary colorectal cancer; SyCRC, synchronous colorectal cancer

Word count: Abstract: 200 words; Manuscript(text): 3931 words.
Total number of figures and tables: 4 figures, 2 Tables

## ABSTRACT

Background: There are no robust tools for the diagnosis of synchronous colorectal cancer (SyCRC). Herein, we developed the first methylation signature to identify and characterize patients with SyCRC.

Methods: For biomarker discovery, we analyzed the genome-wide methylation profiles of 16 SyCRC and 18 solitary colorectal cancer (SoCRC) specimens. We thereafter established a methylation signature risk-scoring model to identify SyCRC in an independent cohort of 38 SyCRC and 42 SoCRC patients. In addition, we evaluated the prognostic value of the identified methylation profile.

Results: We identified six differentially methylated CpG probes/sites that distinguished SyCRC from SoCRC. In the validation cohort, we developed a methylation panel that identified patients with SyCRC from not only larger tumor (AUC=0.91) but also the paired remaining tumor (AUC=0.93). Moreover, high risk scores of our panel were associated with the development of metachronous CRC among patients with SyCRC (AUC=0.87) and emerged as an independent predictor for relapse-free survival (hazard ratio=2.72; 95% CI=1.12–6.61). Furthermore, the risk stratification model which combined with clinical risk factors was a diagnostic predictor of recurrence (AUC=0.90).

Conclusions: Our novel six-gene methylation panel robustly identifies patients with SyCRC, which has the clinical potential to improve the diagnosis and management of patients with CRC.

## INTRODUCTION

Patients with colorectal cancer (CRC) may present with solitary cancer (SoCRC) or multiple primary CRCs, involving two or more neoplasms. Synchronous CRC (SyCRC) is diagnosed when two or more tumors are detected in a single patient at the same time or within 6 months of the initial diagnosis [1]. In contrast, metachronous CRC (MCRC) is diagnosed when the new primary tumor is detected at least 6 months after the resection of the primary lesion and in present in a different part of the large intestine, hence to rule out cancer recurrence from the initial diagnosis [2]. Multiple primary CRCs are thought to have characteristics different from those of SoCRCs due to various environmental and hereditary factors [3-5]. For example, compared to SoCRC, a SyCRC—which accounts for about 1.2-8.1% of all CRCs – is more frequently found in men, at a proximal location, and are generally of a mucinous subtype [6]. The precise and accurate diagnosis of SyCRC is important because patients with such cancers may require extensive resection around the cancer or may even be considered for more extensive segmental resection [7,8]. If overlooked, a synchronous tumor might progress to a more advanced stage and could metastasize. Furthermore, complete pre-operative colonoscopy is often unachievable for patients with distal colonic obstruction or stenosis; hence, raising the possibility of missing such lesions.

Although computed tomography (CT) colonography has improved the detection of synchronous lesions, its diagnostic accuracy still remains largely uncertain [9,10]. NCCN guidelines recommend colonoscopy in 3-6 months when the patients with CRC could not be achieved total colonoscopy before surgery due to obstructing lesion [11]. Unfortunately, up to 50% of patients experience the postoperative complications [12]. Therefore, these patients with lower health related quality of life are not often able to receive the colonoscopy within the recommended interval. Thus, it is critical to develop more robust strategies to identify patients with or likely to develop SyCRC before treatment. Known risk factors for SyCRC include familial adenomatous polyposis (FAP), Lynch syndrome, inflammatory bowel diseases (IBD), and serrated polyps/hyperplastic polyposis [6,13]; however, these features are present in only 10% of patients with SyCRC [14].

Previous studies have identified DNA methylation biomarkers of several cancers based on differentially methylated CpG sites/probes (DMPs) or genes [15-20]. DNA methylation alterations are

remarkably stable, cancer-specific and often occur early during carcinogenesis, representing a promising tool for minimally and noninvasive cancer detection [21]. Considering that aberrant DNA methylation is the most common epigenetic variation in sporadic CRCs [22], we sought to develop a DNA methylation-based signature that can facilitate detection of SyCRC on its own, or in conjunction with currently used diagnostic screening approaches.

Recent studies have identified various genetic and epigenetic features of SyCRC [3,23-25]. For example, long interspersed nucleotide element-1 (LINE1) are frequently methylated in SyCRC [3]. Moreover, approximately 60% of patients with SyCRC exhibit chromosomal instability (CIN)[25], and the presence of these lesions is also highly correlated with the microsatellite instability pathway, with high-frequency microsatellite instability (MSI-high) occurring in about 30% of SyCRC, compared to only 10-12% of SoCRC [3,26,27]. Similarly, in contrast to SoCRC, patients with SyCRC more frequently exhibit the CpG island methylation phenotype (CIMP), which arises through increased accumulation of aberrantly methylated CpG sites within gene promoters of various tumor suppressor genes[3,28]. In spite of this, none of the previous studies have performed a thorough interrogation of DNA methylation profiles in SyCRC, which could offer additional clues for the underlying disease biology and may yield clinically useful biomarkers for disease detection.

To address this important unmet need and gap in knowledge, herein, we performed a systematic and comprehensive genomewide analysis of SoCRC and SyCRC specimens to discover DNA methylation biomarkers for the identification of SyCRC. By undertaking an extensive analysis of methylation sequencing data and using rigorous bioinformatic and statistical approaches, we established six-gene methylation signature that robustly identified patients with SyCRC. These results were subsequently validated in an independent clinical cohort of patients with SyCRC. Equally importantly, we also compared the methylation signatures of *paired-SyCRCs* with SoCRC, which also showed significant difference of this methylation panel. We subsequently evaluated the prognostic potential of our methylation panel for its ability to identify patients that are likely to develop MCRC. Our identified methylation panel could predict the patients which were developed MCRC and the patient with recurrence. Furthermore, our final risk stratification model which combined the methylation panel with clinical risk factors dichotomized high- and low-risk patient with recurrence. In summary, thorough genome-wide DNA methylation profiling analysis,

we successfully established a novel methylation signature for the identification of SyCRC, which has the potential to more accurately identify and risk-stratify patients with SyCRC in the clinic.

## MATERIALS AND METHODS

### Study design and patient cohorts

We analyzed a total of 114 formalin-fixed paraffin-embedded (FFPE) CRC specimens (54 SyCRC and 60 SoCRC) from patients enrolled at the Hospital Universitario 12 de Octubre, Spain, between 2006 and 2018 and at the Hospital Universitario Donostia, Spain, between 2010 and 2017. A SyCRC was defined by the presence of two or more histologically distinct colorectal tumors identified in the same patient at the same time or within six months of the first diagnosis[1,29]. A metachronous CRC (MCRC) was defined as distinctly separated from the previous line of anastomosis and diagnosed at a minimum interval of 6 months after the initial CRC [2]. For patients with SyCRC, we primarily analyzed the higher-stage tumor or larger tumor if the synchronous tumors were of the same stage. Patients with hereditary CRC syndrome, including FAP, Lynch syndrome, patients with history of previous CRC, and those with IBD were excluded in this study. None of the patients received preoperative cancer treatment. All patients were followed until death or March 2020. Relapse-free survival (RFS) times were calculated from the date of surgery to the date of death from any cause or recurrence or last follow-up date.

Our study workflow is summarized in **Supplementary Figure S1** and the clinicopathological characteristics of the clinical cohorts are shown in **Table 1**. In the biomarker discovery phase, 16 SyCRC and 18 SoCRC specimens were profiled for genomewide DNA methylation sequencing to identify candidate biomarkers of SyCRC. Thereafter, 38 SyCRC and 42 SoCRC specimens were analyzed to validate the candidate genes and establish a methylation signature-based risk score.

### DNA extraction and bisulfite conversion

For these experiments, FFPE surgical and endoscopic biopsy slides (10 mm-thick) were hematoxylin and eosin-stained, and DNA was isolated from microdissected, cancer cell-rich areas,

using an AllPrep DNA/RNA FFPE Kit (Qiagen, Hilden, Germany). After quantification of the extracted DNA using a NanoDrop system (Thermo Fisher Scientific, Massachusetts, USA), 500 ng of genomic DNA was bisulfite-converted using an EZ DNA Methylation-Gold Kit (Zymo, Irvine, CA, USA)[30,31]. All procedures were conducted according to the manufacturers' instructions.

### Mutational analyses for *KRAS* and *BRAF* genes

Mutational analysis for *KRAS* and *BRAF* genes was performed using the Ion Torrent PGM platform with a commercial panel. The protocols for the NGS library preparation, emulsion PCR, sequencing analysis, bioinformatics processing and data analysis were performed as previously reported[29].

### MSI and CIMP characterization of the tumors

MSI analysis was performed as previously described[29]. In brief, we used the Bethesda panel to assess the MSI status and considered two or more altered markers as a positive result. MSI tumors were first analyzed for the BRAF V600E mutation and hypermethylation of the MLH1 gene promoter to confirm their sporadic nature and, when negative, they were subsequently screened for germline mutations in the DNA mismatch repair (MMR) genes MLH1, MSH2, MSH6 and PMS2. For the evaluation of CIMP, we investigated the methylation status of the promoter regions of *CACNA1G*, *CDKN2A*, *CRABP1*, *IGF2*, *MLH1*, *NEUROG1*, *RUNX3* and *SOCS1*. Each patient was categorized as CIMP-High, CIMP-Low or CIMP-0 depending on whether their simultaneous tumors showed ≥5/8, 2/8 to 4/8, or 0/8 to 1/8 methylated promoters, respectively.

### Genome-wide DNA methylation analysis

To comprehensively discover biomarkers of SyCRC, we first performed genomewide DNA methylation analysis using an Infinium MethylationEPIC array (GenomesSan B.V., Leiden, Netherlands), which covers more than 850,000 CpG sites [30,32]. Raw fluorescence intensities were loaded into BeadStudio software to generate β values (i.e., the methylation score of each CpG site), ranging from 0 (non-methylated) to 1 (fully methylated). Prior to the identification of DMPs, data preprocessing included data filtering, correction, and normalization. DMPs were detected

based on a β value difference > 0.15 between SyCRC and SoCRC, with a Benjamini-Hochberg adjusted P value < 0.05. Differentially methylated regions (DMRs) were defined as 100-bp genomic windows containing more than two adjacent DMPs[16].

### MethyLight – quantitative polymerase chain reaction (qPCR) assays

We performed MethyLight qPCR assays using a QuantStudio 7 Flex RT-PCR System (Applied Biosystems, Foster City, CA) with a SensiFAST™ Probe Lo-ROX Kit (Bioline, London, UK), as described previously[33]. The primers and probes (listed in **Supplementary Table S1**) were designed using Beacon Designer™ version 8.21 (Premier Biosoft International, Palo Alto, CA, USA). B-actin was used as an internal reference and fully methylated, bisulfite-converted human DNA (Qiagen Hilden, Germany) was used as a positive control to calculate the percentage of methylated reference (PMR) values of the samples (i.e., the degree of methylation of each sample relative to the fully methylated control)[34].

### Statistical analysis

Statistical analyses were performed using MedCalc Statistical Software version 16.2.0 (MedCalc Software, Ostend, Belgium), GraphPad Prism version 8.0 (GraphPad Software, San Diego, CA), and R version 3.5.0 (R Development Core Team; https://cran.r-project.org/). The BMIQ method was used for original β value normalization. The 'limma' package was used to detect DMPs, and the Bumphunter method was used to detect DMRs. The associations between categorical variables were assessed using $\chi 2$, Fisher's exact, and Mann-Whitney U test. Paired t-tests and Mann-Whitney U tests were used to compare methylation signature risk scores between tumors. Correlations between two continuous values were analyzed by Pearson's correlation. Kaplan-Meier analysis and log-rank tests were used to estimate and compare overall survival (OS) and RFS between groups. Risk stratification model was dichotomized into low and high values based on receiver operating characteristic (ROC) curves with Youden's index correction. A univariate and multivariate logistic regression model was used to develop the gene methylation panel. Regarding prognosis prediction, univariate Cox proportional hazard regression model were employed to evaluate the gene methylation panel and several clinical factors. All *P*-values were two-sided, and

*P*-values < 0.05 were considered statistically significant.

## RESULTS

### Genome-wide methylation profiling identifies a panel of six DMPs that discriminate patients with SyCRC from those with SoCRC

Relevant methylation changes are regional during cancer progression; thus, the general pattern demonstrated by several adjacent CpGs represents a more robust biologic effect than any single CpG alone[31,35,36]. Therefore, to obtain SyCRC specific biomarkers, we initially identified 1,184 DMPs which were associated with 175 DMRs after performing DMR filtering based on the restrictive criteria (**Supplementary Fig. S1**). Next, we used the LASSO-based regression algorithms to establish a methylation-based signature that discriminated patients with SyCRC from those with SoCRC. This analysis further reduced the list of candidate DMPs to 12, among which half were significantly hypermethylated and the other half were hypomethylated in SyCRCs (**Fig. 1A**).

Subsequently, we visualized the distribution of all CRC samples based on these DMPs using a two-dimensional scatter plot produced by multidimensional scaling, which revealed that the two clusters corresponding to SyCRC and SoCRC were distinct and clearly discriminated by these differentially methylated loci (**Fig. 1B**). From this initial set of 12 DMPs, we excluded those that were highly correlated with each other and did not add any further value to the discriminatory model, which led us to finally establish a panel of six DMPs: cg20275528, cg03578926, cg22084339, cg27332938, cg10461088, and cg11255039, which corresponded to *SEPT9*, *SHANK2*, *PRKAR1B*, *ZNF511*, *ARFGAP2*, and *KIF22* genes, respectively (**Fig. 1C**). Next, we constructed a logistic regression model with these six DMPs to calculate the risk scores for patients with SyCRC in the discovery cohort. Our model demonstrated excellent predictive performance (area under the curve [AUC]=1.00; 95% confidence interval [CI]=1.00–1.00; **Fig. 1D**), highlighting the significance of the epigenetic biomarkers we discovered and their ability to discriminate patients with SyCRC from those with SoCRC which provides a rationale for these alterations for their biological and clinical significance for interrogating the differences between these two subtypes of CRC.

## Successful clinical validation of the six-gene methylation panel for its ability to identify SyCRC in an independent clinical cohort

In order to validate the performance of the six-gene methylation panel and assess its potential as a clinically translatable prognostic assay, we first performed MethyLight-based qPCR assays to quantitatively measure the methylation status of each gene in an independent clinical cohort of patients (n=80; 38 SyCRC and 42 SoCRC). Using the logistic regression analysis, we developed this risk-assessment scoring model based on the coefficients derived from individual markers with the following model parameters: risk score = (0.076389* methylation level of *SEPT9*) + (-0.054988* methylation level of *SHANK2*) + (0.00072293* methylation level of *PRKAR1B*) + (0.2938* methylation level of *ZNF511*) + (0.086657* methylation level of *ARFGAP2*) + (-0.03117* methylation level of *KIF22*) – 5.71048. When we evaluated the cumulative risk score based upon this scoring model, we observed that patients with SyCRC had significantly higher risk scores than those with SoCRC ($P < 0.001$, Mann-Whitney test; **Fig. 2A**). Moreover, the six-gene methylation model robustly identified patients with SyCRC (AUC=0.91; 95% CI=0.82–0.96; **Fig. 2B**); with 93.9% of patients with SyCRC had a positive score. In contrast, as much as 85.1% of patients with SoCRC had a negative risk score (**Fig. 2C**), indicating the specificity of this six gene methylation panel for patients with SyCRC. Next, we compared patients with SoCRC and SyCRC without stage 0 due to exclude the possibility of influence the difference in the baseline characteristics. We observed that the diagnostic performance of our methylation panel under the exclusion of patients with stage 0 was quite comparable to that observed in the total cohort (AUC = 0.94, 95% CI = 0.85-0.92; **Supplementary Fig. S2**). Taken together, these data illustrate that we successfully established a novel methylation panel for the robust identification of patients with SyCRC.

## The gene methylation panel has a diagnostic role in the identification of paired tumors

In our analysis to this point, we compared specimens from patients with SoCRC to the larger or higher-stage tumor in SyCRC pairs. However, there is a growing body of evidence that suggests significant genetic differences between SyCRC tumor pairs [37,38]. Accordingly, we next compared the methylation signatures of paired synchronous tumors (n=38 pairs; **Supplementary Table S2**).

In twenty-four patients (63.2%), SyCRC pair tumors were located within the same segment of colon; 16 in the distal colon (from splenic flexure to rectum), and 8 in the proximal colon (from cecum to transverse colon). In contrast, fourteen patients (36.8%) had the SyCRC tumors in different segments of colon. Of a total of 76 SyCRC, 32 tumors exhibited *KRAS* mutations (42.1%), 3 had *BRAF* mutations (3.9%), 3 were MSI-H (3.9%), and 40 were CIMP-high (52.6%). With regards to the concordance between paired tumors, 8 patients harbored *KRAS* mutations (21.2%), 1 had a mutation in the *BRAF* gene (2.6%), and 13 exhibited CIMP-high (34.2%). None of the patients possessed MSI-H in both tumors. Interestingly, no significant differences were observed in their risk scores, and our model yielded an AUC value of 0.51 (95% CI=0.39–0.62) for differentiating the larger and higher-stage synchronous tumors from their pairs (**Fig. 3A** and **Supplementary Fig. S3**). Moreover, the risk scores of the paired tumors were positively correlated (r=0.52, *P*<0.001; **Fig. 3B**). Next, we were curious to assess the performance of our methylation risk scores in distinguishing patients with SoCRC from the remaining cohort of SyCRC cases (smaller or lower stage tumors). It was quite encouraging to observe that as with the initial set of SyCRC tumors, we observed a significant association between high-risk scores and patients with SyCRC (AUC=0.93; 95% CI=0.84-0.98; **Fig. 3C** and **3D**). Taken together, these data indicate that the methylation risk score of either tumor can be used to identify patients with SyCRC.

## Development of a risk stratification model incorporating the methylation panel and clinical risk factors to identify the high-risk patients with SyCRC.

In our total cohort, among the 38 patients with SyCRC, seven developed metachronous lesions. We hypothesized that our methylation panel might be able to predict development of MCRC. It was reassuring to witness that indeed we observed higher methylation risk scores that significantly associated with the development of MCRC (AUC=0.87; 95% CI=0.72–0.96; **Fig.4A**). Moreover, we compared the overall risk score distributions among SyCRC patients and observed that those who developed MCRC had significantly higher methylation scores than those who did not (*P*<0.001; **Fig.4B**).

Because the development of MCRC is often associated with poor prognosis, we examined the prognostic potential of our methylation risk score. In our cohort, 4 of 38 patients were

diagnosed with disease recurrence. One patient who got carcinomatosis was treated with palliative chemotherapy. Locoregional recurrence was observed in two patients, and they did not receive any treatment. Another patient presented with metachronous liver metastasis and received adjuvant therapy after surgery. Interestingly, we observed that our score showed a robust identification of recurrence in patient with SyCRC (AUC=0.76, 95% CI=0.56-0.96; **Fig. 4C**). In addition, we conducted univariate analysis using Cox proportional hazard regression to estimate the prognostic ability of the methylation panel and other clinicopathologic factors, and the methylation panel was the only factor associated with significantly worse RFS (hazard ratio=2.72; 95% CI=1.12–6.61; **Table 2**). However, considering that some of risk factors currently used for predicting some prognostic potential in patients with CRC in the clinic, we examined whether a risk-stratification model that includes our methylation risk score and any other clinical risk factors might serve as further accurate recurrence marker. When we established the risk stratification model by combining out methylation risk score with CEA status, tumor size, and the presence of lymph node metastasis, the risk model further augmented the diagnostic accuracy of the methylation panel and other risk factors (AUC=0.90, 95% CI=0.80-1.00; **Fig. 4C**). Next, Kaplan-Meier analysis for OS and RFS was performed in order to evaluate the risk stratification model. The median follow-up time was 133.42 months (95% CI=120.73–146.12) in our clinical cohort. Importantly, 11 high-risk SyCRC patients exhibited significantly poorer RFS than the 26 low-risk SyCRC patients ($P$<0.01; **Fig.4D**), whereas there were no significant differences in OS between the two groups (**Supplementary Fig. S4**). Collectively, these results indicate that our methylation signature has not only the diagnostic ability to identify patients with SyCRC, but significant prognostic potential, as well.

## DISCUSSION

The clinicopathological features of SyCRC are poorly understood, and no tools are currently available for the accurate diagnosis of these lesions. In this study, we used a systematic genomewide methylation sequencing approach to identify DMPs that are significantly associated with SyCRC and subsequently developed and validated a six-gene methylation signature that

328 robustly identified patients with such lesions in an independent clinical cohort. Moreover, we
329 demonstrated that the methylation gene-based risk-scoring model yielded high risk scores in
330 patients with paired synchronous tumors, and these scores were significantly correlated.
331 Furthermore, our methylation signature was robust in identifying patients with poor RFS and
332 could predict the subgroup of patients that developed MCRC. Taken together, these results
333 highlight the potential clinical significance of our novel methylation as the first and one-of-a-kind
334 molecular signature for a more accurate identification and characterization of patients with SyCRC.
335 Our highly specific methylation signature includes six genes that have not been previously
336 associated with SyCRC. *SEPT9* is known to be highly methylated in the tumor tissue and plasma
337 of patients with CRC, and hypermethylated *SEPT9* is associated with CRC tumorigenesis [39-41]. In
338 fact, a blood test to detect circulating methylated *SEPT9* has been approved by the U.S. Food and
339 Drug Administration (FDA) for CRC screening. Whereas the other five genes have never been
340 reported to associate with CRC, *SHANK2* is a member of the Shank family of synaptic proteins,
341 which is often amplified in human cancers and potently promotes tumor formation[42].
342 Furthermore, *KIF22* is involved in the regulation of the cell cycle via MEK/ERK/P21 and promotes
343 the occurrence and development of pancreatic cancer[43]. The function of the remaining three
344 genes in cancer remains unclear. However, these genes corresponding to the DMPs in SyCRC are
345 the first to be revealed due to our genomewide methylation profiling efforts in SyCRC and may
346 also be related to the development and progression of the disease. Additional functional
347 investigations into these six genes may provide new insights and novel approaches to managing
348 and treating patients with SyCRC.
349 Although our initial studies compared SoCRC with larger and higher-stage SyCRC tumors,
350 we observed similar results when we compared SoCRC to smaller and lower-stager SyCRC tumors.
351 Moreover, although previous studies have shown a high degree of heterogeneity between paired
352 SyCRC tumors based on whole-genome sequencing [37,38], we found that the methylation signature
353 risk scores of paired SyCRC tumors were highly correlated, hence superior, and might indicate a
354 field cancerization defect [44]. Several studies have previously demonstrated similarities between
355 the epigenetic signatures of paired tumors from patients with SyCRC; however, to the best of our
356 knowledge, none has conducted an analysis as comprehensive as the one presented in our

present study [3,29,45]. These results indicate that the analysis of any primary CRC tumor may be sufficient for detecting SyCRC.

The prognosis of patients affected by SyCRC is controversial [3,5,14]. Whereas one study observed poorer survival among patients with SyCRC than patients with SoCRC [3], others have demonstrated that there are no significant differences in their OS[5,14]. In the present study, we did not observe any significant differences in survival between patients with SyCRC vs. SoCRC (data not shown), as our cohort may have included stage 0 cancer in SyCRC group. However, our methylation risk score was able to predict poor prognosis among patients with SyCRC by detecting not only those who developed MCRC but also recurrence. Epigenetic changes of DNA methylation have a critical role in cancer progression and metastasis [22]. Some of our six gene methylation may contribute to cancer progression. Therefore, risk stratification model demonstrated further improved diagnostic accuracy, indicating that our novel methylation signature has prognostic, as well as diagnostic, potential.

Although our results are promising, we would like to acknowledge a few potential limitations to the present study. First, we were unable to validate our biomarkers in an independent prospective cohort of patients. Furthermore, our study had a retrospective design; therefore, our results could have inadvertently affected due to a potential selection bias between SyCRC and SoCRC groups. Therefore, to further confirm the accuracy of our stratification model, prospective studies with larger patient cohorts are required before translating our biomarkers to the clinical setting. Second, we used surgical specimens rather than tissue biopsy specimens; therefore, we could not evaluate the ability of our methylation signature to identify the patients with SyCRC before treatment. Nevertheless, despite these limitations, this study remains valuable and demonstrates the significant potential of our methylation signature for clinically identifying patients with SyCRC. Moreover, the diagnosis of SyCRC sometimes conditions the need for more extensive surgeries, or studies to rule out their hereditary nature. On the other hand, the capacity for identify cases that will also develop metachronous neoplasms would also be highlighted somewhat more, since even those cases would condition extensive surgery in the diagnosis of SyCRC, and thus be able to prevent the development of the consequent metachronous disease.

In conclusion, we used genome-wide methylation profiling to identify DMPs that distinguished patients with SyCRC from those with SoCRC, followed by robust analyses to develop a novel methylation signature to identify SyCRC. We successfully validated our signature in an independent clinical cohort and demonstrated its potential diagnostic and prognostic clinical significance, which could have major implications for the management of CRC in the clinic.

## Additional Information


### Acknowledgements

We thank Tatsuhiko Kakisaka, Priyanka Sharma, Satoshi Nishiwada, Yuma Wada, In Seob Lee, Geeta Sharma, Divya Sahu, Xiao Zhang, Yinghui Zhao, and Yuetong Chen for helping to conduct the experiments and data analysis. We also would like to extend our thanks to Dr. Kerin Higa for her significant editing and useful suggestions for improving the quality of our manuscript.


### Authors' contributions

**YO:** study concept and design; analysis and interpretation of data and statistical analysis; drafting of the manuscript. **FP:** analysis and interpretation of data, statistical analysis; drafting of the manuscript. **JP:** specimen provider; acquisition of clinical data; drafting of the manuscript. **LC:** specimen provider; acquisition of clinical data; drafting of the manuscript. **LB:** specimen provider; acquisition of clinical data; drafting of the manuscript. **WL** analysis and interpretation of data, statistical analysis; drafting of the manuscript. **AG:** study concept and design, analysis and interpretation of data, statistical analysis, drafting of the manuscript.

All authors read and approved the final manuscript.

### Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. A written informed consent was obtained from all patients, and the study was approved by the institutional review boards of the participating institutions.

### Data availability

The datasets used for the current study are available from the corresponding author on reasonable request.

### Competing Interests

None of the authors has any potential conflicts to disclose.


### Funding information

This work was supported by CA72851, CA181572, CA184792, CA202797, and CA227602 grants from the National Cancer Institute, National Institutes of Health.

## REFERENCES


1 Lee, B. C. *et al.* Clinicopathological features and surgical options for synchronous colorectal cancer. *Medicine (Baltimore)* **96**, e6224 (2017).

2 Cunliffe, W. J., Hasleton, P. S., Tweedle, D. E. & Schofield, P. F. Incidence of synchronous and metachronous colorectal carcinoma. *Br J Surg* **71**, 941-943 (1984).

3 Nosho, K. *et al.* A prospective cohort study shows unique epigenetic, genetic, and prognostic features of synchronous colorectal cancers. *Gastroenterology* **137**, 1609-1620.e1601-1603 (2009).

4 Pajares, J. A. & Perea, J. Multiple primary colorectal cancer: Individual or familial predisposition? *World J Gastrointest Oncol* **7**, 434-444 (2015).

5 Yang, J., Peng, J. Y. & Chen, W. Synchronous colorectal cancers: a review of clinical features, diagnosis, treatment, and prognosis. *Dig Surg* **28**, 379-385 (2011).

6 Lam, A. K., Chan, S. S. & Leung, M. Synchronous colorectal cancer: clinical, pathological and molecular implications. *World J Gastroenterol* **20**, 6815-6820 (2014).

7 Wang, H. Z., Huang, X. F., Wang, Y., Ji, J. F. & Gu, J. Clinical features, diagnosis, treatment and prognosis of multiple primary colorectal carcinoma. *World J Gastroenterol* **10**, 2136-2139 (2004).

8 Passman, M. A., Pommier, R. F. & Vetto, J. T. Synchronous colon primaries have the same prognosis as solitary colon cancers. *Dis Colon Rectum* **39**, 329-334 (1996).

9 Latournerie, M. *et al.* Epidemiology and prognosis of synchronous colorectal cancers. *Br J Surg* **95**, 1528-1533 (2008).

10 Johnson, C. D. *et al.* Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med* **359**, 1207-1217 (2008).

11 Provenzale, D. *et al.* NCCN Guidelines Insights: Colorectal Cancer Screening, Version 1.2018. *J Natl Compr Canc Netw* **16**, 939-949 (2018).

12 van Rooijen, S. *et al.* Multimodal prehabilitation in colorectal cancer patients to improve functional capacity and reduce postoperative complications: the first international randomized controlled trial for multimodal prehabilitation. *BMC Cancer* **19**, 98 (2019).

13 Chin, C. C., Kuo, Y. H. & Chiang, J. M. Synchronous colorectal carcinoma: predisposing factors and characteristics. *Colorectal Dis* **21**, 432-440 (2019).

14 Lam, A. K. *et al.* Clinicopathological significance of synchronous carcinoma in colorectal cancer. *Am J Surg* **202**, 39-44 (2011).

15 Qiu, J. *et al.* CpG Methylation Signature Predicts Recurrence in Early-Stage Hepatocellular Carcinoma: Results From a Multicenter Study. *J Clin Oncol* **35**, 734-742 (2017).

16 Dong, S. *et al.* Histone-Related Genes Are Hypermethylated in Lung Cancer and Hypermethylated HIST1H4F Could Serve as a Pan-Cancer Biomarker. *Cancer Res* **79**, 6101-6112 (2019).

17 Diaz-Lagares, A. *et al.* A Novel Epigenetic Signature for Early Diagnosis in Lung Cancer. *Clin Cancer Res* **22**, 3361-3371 (2016).

18 Fan, J. *et al.* Genome-wide DNA methylation profiles of low- and high-grade adenoma reveals potential biomarkers for early detection of colorectal carcinoma. *Clin Epigenetics* **12**, 56 (2020).

19 Luo, Y. *et al.* Differences in DNA methylation signatures reveal multiple pathways of

progression from adenoma to colorectal cancer. *Gastroenterology* **147**, 418-429.e418 (2014).

20 Gallardo-Gómez, M. *et al.* A new approach to epigenome-wide discovery of non-invasive methylation biomarkers for colorectal cancer screening in circulating cell-free DNA using pooled samples. *Clin Epigenetics* **10**, 53 (2018).

21 Heyn, H. & Esteller, M. DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* **13**, 679-692 (2012).

22 Jung, G., Hernández-Illán, E., Moreira, L., Balaguer, F. & Goel, A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat Rev Gastroenterol Hepatol* **17**, 111-130 (2020).

23 de Macedo, M. P. *et al.* RAS mutations vary between lesions in synchronous primary colorectal cancer: testing only one lesion is not sufficient to guide anti-EGFR treatment decisions. *Oncoscience* **2**, 125-130 (2015).

24 Lawes, D. A., Pearson, T., Sengupta, S. & Boulos, P. B. The role of MLH1, MSH2 and MSH6 in the development of multiple colorectal cancers. *Br J Cancer* **93**, 472-477 (2005).

25 Leggett, B. A. & Worthley, D. L. Synchronous colorectal cancer: not just bad luck? *Gastroenterology* **137**, 1559-1562 (2009).

26 Dykes, S. L., Qui, H., Rothenberger, D. A. & García-Aguilar, J. Evidence of a preferred molecular pathway in patients with synchronous colorectal cancer. *Cancer* **98**, 48-54 (2003).

27 Boland, C. R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087.e2073 (2010).

28 Lao, V. V. & Grady, W. M. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* **8**, 686-700 (2011).

29 Perea, J. *et al.* Redefining synchronous colorectal cancers based on tumor clonality. *Int J Cancer* **144**, 1596-1608 (2019).

30 Kandimalla, R. *et al.* EpiPanGI Dx: A Cell-free DNA Methylation Fingerprint for the Early Detection of Gastrointestinal Cancers. *Clin Cancer Res* **27**, 6135-6144 (2021).

31 Roy, R. *et al.* A comprehensive methylation signature identifies lymph node metastasis in esophageal squamous cell carcinoma. *Int J Cancer* **144**, 1160-1169 (2019).

32 Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8**, 389-399 (2016).

33 Eads, C. A. *et al.* MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* **28**, E32 (2000).

34 Friedrich, M. G. *et al.* Detection of methylated apoptosis-associated genes in urine sediments of bladder cancer patients. *Clin Cancer Res* **10**, 7457-7465 (2004).

35 Hotta, K. *et al.* Identification of differentially methylated region (DMR) networks associated with progression of nonalcoholic fatty liver disease. *Sci Rep* **8**, 13567 (2018).

36 Lehmann-Werman, R. *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A* **113**, E1826-1834 (2016).

37 Cereda, M. *et al.* Patients with genetically heterogeneous synchronous colorectal cancer carry rare damaging germline mutations in immune-related genes. *Nat Commun* **7**, 12072 (2016).

38 Wang, X. *et al.* The molecular landscape of synchronous colorectal cancer reveals genetic heterogeneity. *Carcinogenesis* **39**, 708-718 (2018).
39 Model, F. *et al.* Identification and validation of colorectal neoplasia-specific methylation markers for accurate classification of disease. *Mol Cancer Res* **5**, 153-163 (2007).
40 Church, T. R. *et al.* Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* **63**, 317-325 (2014).
41 Lofton-Day, C. *et al.* DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin Chem* **54**, 414-423 (2008).
42 Xu, L. *et al.* SHANK2 is a frequently amplified oncogene with evolutionarily conserved roles in regulating Hippo signaling. *Protein Cell* **12**, 174-193 (2021).
43 Zhang, R. *et al.* KIF22 Promotes Development of Pancreatic Cancer by Regulating the MEK/ERK/P21 Signaling Axis. *Biomed Res Int* **2022**, 6000925 (2022).
44 Braakhuis, B. J., Tabor, M. P., Kummer, J. A., Leemans, C. R. & Brakenhoff, R. H. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* **63**, 1727-1730 (2003).
45 Tapial, S. *et al.* Cimp-Positive Status is More Representative in Multiple Colorectal Cancers than in Unique Primary Colorectal Cancers. *Sci Rep* **9**, 10516 (2019).

## FIGURE LEGENDS

### Figure 1. Identification of a methylation panel for the identification of patients with SyCRC.

(A) Heatmap representing the methylation of 12 significant DMPs in patients with SyCRC (n=16) and SoCRC (n=18), identified based on methylation array analysis. (B) Multidimensional scaling (MDS) plot of SyCRC and SoCRC cases, based on the methylation levels of the 12 DMPs. (C) Heatmap of the Spearman's correlation coefficients between the methylation levels of the 12 DMPs. DMPs that were highly correlated with others were excluded, leaving six DMPs in the candidate methylation signature. (D) ROC curve analysis of the six-gene methylation panel for identifying patients with SyCRC in the discovery cohort.

### Figure 2. Establishment of a six-gene methylation signature using MethyLight qPCR assays in an independent clinical cohort.

(A) Violin plots representing the risk scores of patients with SyCRC and SoCRC in the validation cohort. (B) ROC curve analysis of the risk score for identifying patients with SyCRC. (C) Distribution of risk scores in the validation cohort.

### Figure 3. Comparison of methylation signatures between pairs of synchronous tumors.

(A) Comparison of the methylation panel risk scores of pairs of synchronous tumors. (B) Correlations between the methylation panel risk scores of paired synchronous tumors. (C) ROC curve analysis of the risk score for identifying patients with SyCRC. (D) Distribution of risk scores in the clinical validation cohort, based on SoCRC and lower-stage or smaller SyCRC tumors. Tumor 1: the higher-stage tumor or larger tumor if the paired tumors were of the same stage. Tumor 2: the lower-stage tumor or smaller tumor if the paired tumors were of the same stage. N.S., not significant.

### Figure 4. Prognostic potential of the methylation signature for the patients with SyCRC.

(A) ROC curve analysis of the risk score for predicting which patients with SyCRC developed MCRC. (B) Violin plots representing the risk scores of patients with SyCRC in the validation cohort who did and did not develop MCRC. (C) ROC curve analysis of risk stratification model, the methylation

554 panel, tumor size, CEA status and lymph node metastasis for the prediction of recurrence in the
555 patients with SyCRC. (D) Kaplan-Meier plots of RFS in high- and low-risk SyCRC patients in the
556 validation cohort, stratified based on their risk stratification models.
557

**Table 1.** Clinicopathological characteristics of patients in the clinical discovery and validation cohorts.

| Characteristics | Discovery cohort | | | Validation cohort | | |
|---|---|---|---|---|---|---|
| | SoCRC (n=18) | SyCRC (n=16) | *P* value | SoCRC (n=42) | SyCRC (n=38) | *P* value |
| Gender (percentage) | | | 0.08* | | | 0.2 |
| Male | 4 (22.2) | 9 (56.3) | | 24 (57.1) | 27 (71.1) | |
| Female | 14 (77.8) | 7 (43.7) | | 18 (42.9) | 11 (28.9) | |
| Age (years) | | | 0.16** | | | **0.02**** |
| Median (range) | 70 (31–85) | 75 (60–88) | | 73 (47–96) | 69 (45–91) | |
| Location | | | 0.83 | | | 0.42 |
| Left | 10 | 10 | | 28 | 22 | |
| Right | 7 | 6 | | 14 | 16 | |
| Not available | 1 | 0 | | 0 | 0 | |
| Histology | | | | | | 0.09* |
| Differentiated | 16 | 12 | | 38 | 19 | |
| Undifferentiated | 1 | 0 | | 4 | 7 | |
| Not available | 1 | 4 | | 0 | 12 | |
| Tumor size (mm) | | | 0.19** | | | **0.01**** |
| Median (range) | 40 (20–100) | 30 (20–150) | | 40 (20–100) | 30 (4–110) | |
| Lymph node metastasis | | | 0.17* | | | 0.15 |
| Negative | 9 | 12 | | 28 | 30 | |
| Positive | 9 | 4 | | 14 | 7 | |
| Not available | 0 | 0 | | 0 | 1 | |
| AJCC stage (ver. 8) | | | 0.13 | | | **0.002** |
| 0 | 0 | 0 | | 0 | 12 | |
| I | 1 | 5 | | 11 | 7 | |
| II | 9 | 7 | | 17 | 10 | |
| III | 8 | 4 | | 13 | 7 | |
| IV | 0 | 0 | | 1 | 2 | |

*AJCC, American Joint Committee on Cancer; * Fisher's exact test; ** Mann-Whitney U test*

558

559

**Table 2.** Univariate Cox proportional analysis of RFS in patients with SyCRC in the clinical validation cohort.

| | Univariate analysis | | |
|---|---|---|---|
| **Characteristics** | **HR** | **95% CI** | ***P* value** |
| Age (≥70 vs. <70 years) | 1.18 | 0.16–8.50 | 0.87 |
| Gender (Female vs. Male) | 6.86 | 0.71–65.94 | 0.10 |
| CEA (≥5.0 vs. <5.0 ng/mL) | 1.93 | 0.27–13.73 | 0.51 |
| Tumor size (≥30 mm vs. <30 mm) | 2.90 | 0.30–27.89 | 0.36 |
| Lymph node metastasis | 1.83 | 0.19–17.70 | 0.60 |
| Six-gene methylation signature risk score | 2.72 | 1.12–6.61 | **0.03** |

*HR*, hazard ratio; *CI*, confidence interval

560

## Supplemental Material

### Supplementary Figure S1. Overview of the study design.

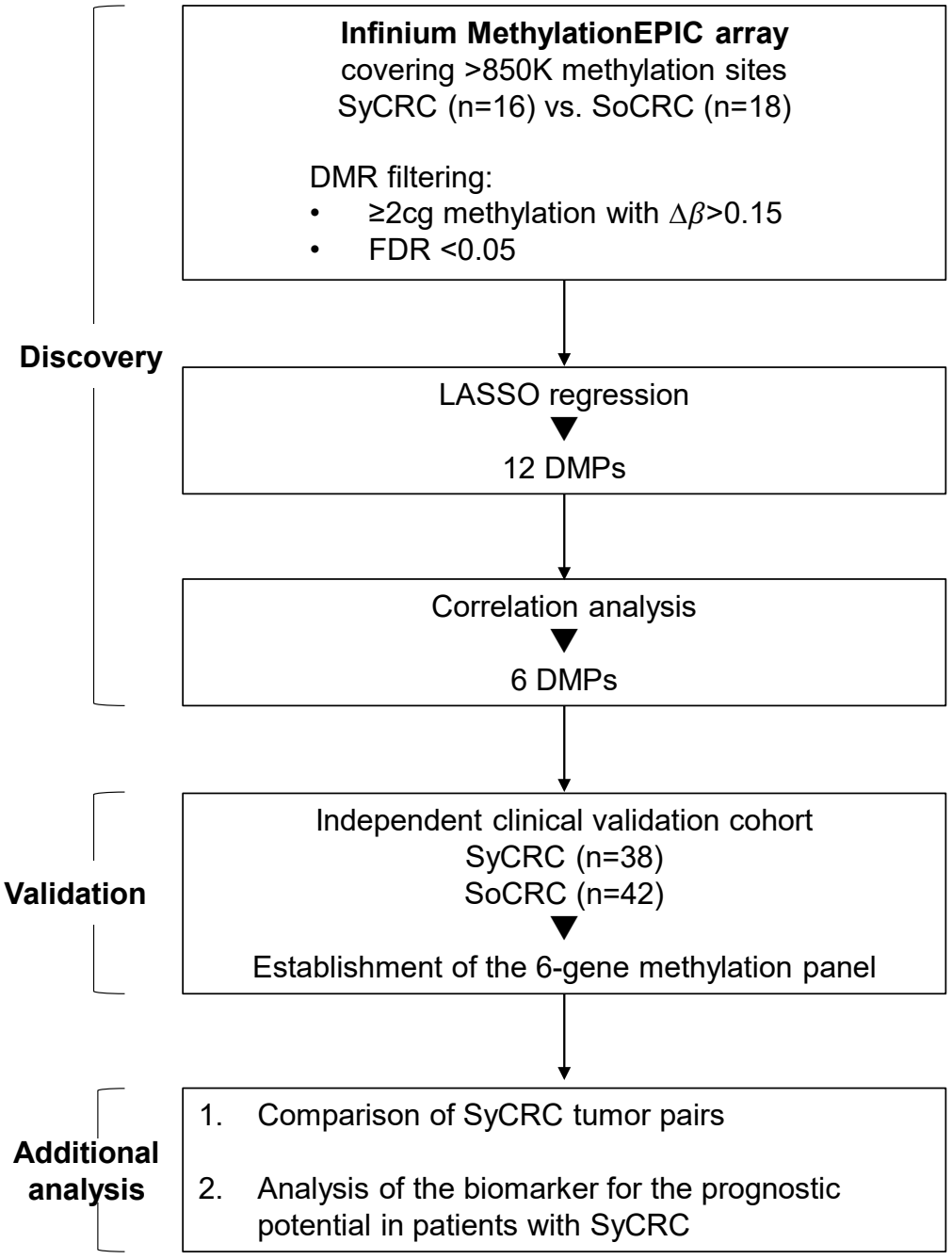
$\beta$ values (0 to 1) represent the methylation score. DMR, differentially methylated region; DMPs, differentially methylated CpG sites; FDR, false discovery rate; SyCRC, synchronous colorectal cancer; SoCRC, solitary colorectal cancer.

### Supplementary Figure S2. ROC curve analysis of the risk score for identifying patients with SyCRC (excluding patients with stage 0).

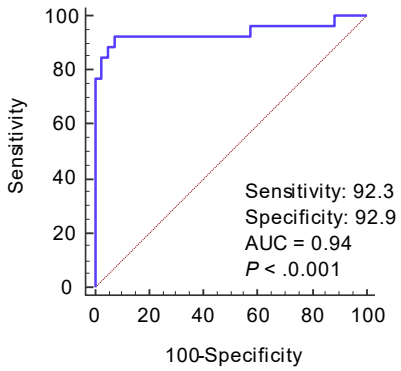
### Supplementary Figure S3. ROC curve analysis of the risk score for differentiating larger and higher-stage synchronous tumors from their pairs.

### Supplementary Figure S4. Kaplan-Meier plots of OS in high- and low-risk SyCRC patients in the validation cohort, stratified based on their risk stratification models.

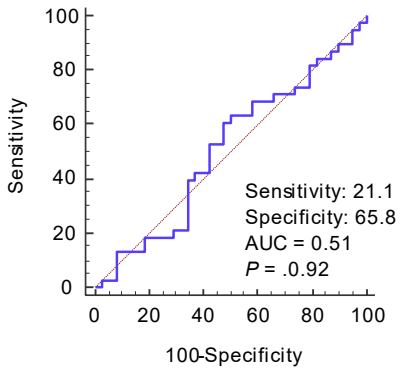
Supplementary Figure S1



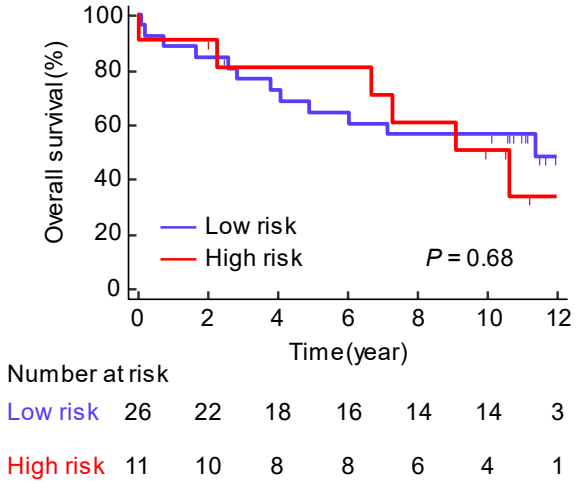
Supplementary Figure S2



Supplementary Figure S3



Supplementary Figure S4




Number at risk

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Low risk | 26 | 22 | 18 | 16 | 14 | 14 | 3 |
| High risk | 11 | 10 | 8 | 8 | 6 | 4 | 1 |

**Supplementary Table S1 Methylight primer pairs and probes for 6 genes**

| No. | Gene | Forward primer | Reverse primer | Probe |
|---|---|---|---|---|
| 1 | *SEPT9* | TTTATTTAGTCGGAGGTGAGGAA | CTTTAACTCTCCCCGACGAC | CCCGCTTAAACCCGACAACGAAATAAA |
| 2 | *SHANK2* | GCGGGATGACGTTTAGGTAG | CCGACGATATACGACAAACAAA | CCACAATCATCTAACGAACCCACAATACG |
| 3 | *PRKAR1B* | GTGGGTTTTAGGTCGGTTTT | TCCCGTAATCCTCGAAAACTA | AAACTATCTACCGTCCTACAAATCCTTCC |
| 4 | *ZNF511* | GGAGTAAATATTTTCGTGTAGCG | CCAAATAACCGACTACTACCAAA | CCCTAAACGACTACGAACACCACTACC |
| 5 | *ARFGAP2* | CGGAGAGTTTATTTGATGAAGT | GTACGATATTTTCTTATACATTAACTAT | CCGAAATACACCGCTCCCTAAACG |
| 6 | *KIFF22* | GGTATTCGTTTTGTTTAGGTCG | AAACGACGCGAAATAACGAC | ACTTCAACGACGACGATCTCAAATACTT |
| 7 | β-actin | TGGTGATGGAGGAGGTTTAGTAAGT | AACCAATAAAACCTACTCCTCCCTTAA | ACCACCACCCAACACACAATAACAAACACA |

**Supplementary Table 2. Synchronous colorectal cancer pairs with clinical data**

| Case | Age | Sex | Lesion | Location | size (mm) | Differentiation | *KRAS* | *BRAF* | MSI | CIMP |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 68 | Female | 1 | Distal | 40 | differentiation | *KRAS* G12D | wild | MSS | Low |
| | | | 2 | Distal | 10 | N/A | *KRAS* G12D | wild | MSS | Low |
| 2 | 68 | Female | 1 | Distal | 27 | undifferentiation | wild | V600E | MSS | Low |
| | | | 2 | Proximal | 17 | undifferentiation | wild | V600E | MSS | High |
| 3 | 80 | Male | 1 | Proximal | 47 | differentiation | *KRAS* G12D | wild | MSS | Low |
| | | | 2 | Proximal | 30 | differentiation | *KRAS* G12D | wild | MSS | 0 |
| 4 | 63 | Female | 1 | Proximal | 40 | differentiation | wild | wild | MSS | Low |
| | | | 2 | Proximal | 12 | N/A | *KRAS* G13D | wild | MSS | Low |
| 5 | 69 | Male | 1 | Proximal | 110 | differentiation | *KRAS* G12V | wild | MSS | 0 |
| | | | 2 | Distal | 10 | N/A | *KRAS* G12V | wild | MSS | Low |
| 6 | 64 | Female | 1 | Distal | 24 | differentiation | *KRAS* G13D | wild | MSS | Low |
| | | | 2 | Distal | 15 | differentiation | *KRAS* G13D | wild | MSS | Low |
| 7 | 91 | Male | 1 | Distal | 70 | differentiation | *KRAS* G12A | wild | MSS | Low |
| | | | 2 | Proximal | 55 | differentiation | wild | V600E | MSS | High |
| 8 | 75 | Female | 1 | Proximal | 10 | N/A | wild | wild | MSS | 0 |
| | | | 2 | Distal | 6 | N/A | wild | wild | MSS | Low |
| 9 | 54 | Male | 1 | Distal | 20 | N/A | wild | wild | MSS | 0 |
| | | | 2 | Distal | 15 | N/A | wild | wild | MSS | 0 |
| 10 | 70 | Female | 1 | Proximal | 4 | N/A | wild | wild | MSS | Low |
| | | | 2 | Proxima | 4 | N/A | wild | wild | MSS | Low |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | l | | | | | | |
| 11 | 75 | Female | 1 | Proxima l | 60 | differentiation | wild | wild | MSS | Low |
| | | | 2 | Proxima l | 50 | differentiation | wild | wild | MSI-H | High |
| 12 | 60 | Male | 1 | Proxima l | 20 | N/A | wild | wild | MSS | 0 |
| | | | 2 | Distal | 15 | N/A | wild | wild | MSI-H | Low |
| 13 | 59 | Male | 1 | Distal | 20 | differentiation | *KRAS* G13D | wild | MSS | High |
| | | | 2 | Distal | 15 | N/A | wild | wild | MSS | 0 |
| 14 | 73 | Male | 1 | Distal | 95 | differentiation | wild | wild | MSS | High |
| | | | 2 | Distal | 5 | N/A | wild | wild | MSS | Low |
| 15 | 59 | Female | 1 | Distal | 20 | differentiation | *KRAS* G12V | wild | MSS | 0 |
| | | | 2 | Distal | 5 | N/A | wild | wild | MSS | High |
| 16 | 71 | Male | 1 | Distal | 5 | undifferentiation | wild | wild | MSS | High |
| | | | 2 | Distal | 6 | N/A | *KRAS* G12D | wild | MSS | High |
| 17 | 58 | Male | 1 | Proxima l | 15 | N/A | *KRAS* G13D | wild | MSS | High |
| | | | 2 | Proxima l | 4 | N/A | wild | wild | MSS | High |
| 18 | 67 | Male | 1 | Proxima l | 25 | differentiation | *KRAS* G12D | wild | MSS | High |
| | | | 2 | Distal | 50 | differentiation | *KRAS* G12D | wild | MSS | Low |
| 19 | 71 | Female | 1 | Proxima l | 20 | N/A | *KRAS* G12D | wild | MSS | 0 |
| | | | 2 | Distal | 15 | N/A | *KRAS* G12D | wild | MSS | Low |
| 20 | 58 | Female | 1 | Proxima l | 70 | undifferentiation | wild | wild | MSS | High |
| | | | 2 | Proxima l | 45 | undifferentiation | wild | wild | MSI-H | Low |
| 21 | 69 | Male | 1 | Proxima l | 20 | differentiation | wild | wild | MSS | High |
| | | | 2 | Proxima l | 20 | differentiation | wild | wild | MSS | Low |

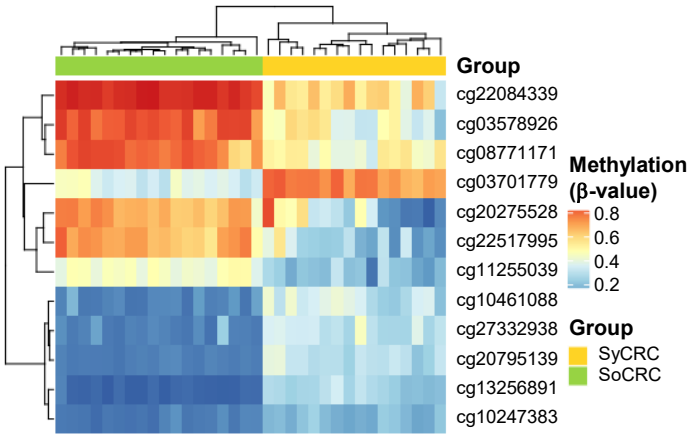
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 73 | Male | 1 | Distal | 35 | undifferentiation | wild | wild | MSS | High |
| | | | 2 | Proximal | 10 | differentiation | *KRAS* G12D | wild | MSS | High |
| 23 | 45 | Male | 1 | Proximal | 55 | differentiation | wild | wild | MSS | High |
| | | | 2 | Distal | 7 | N/A | wild | wild | MSS | High |
| 24 | 84 | Male | 1 | Distal | 25 | N/A | *KRAS* G12V | wild | MSS | 0 |
| | | | 2 | Distal | 15 | N/A | *KRAS* G12C | wild | MSS | High |
| 25 | 82 | Male | 1 | Proximal | 15 | differentiation | *KRAS* G12V | wild | MSS | High |
| | | | 2 | Proximal | 20 | N/A | wild | wild | MSS | High |
| 26 | 71 | Male | 1 | Proximal | 30 | differentiation | wild | wild | MSS | High |
| | | | 2 | Distal | 21 | N/A | *KRAS* A146T | wild | MSS | 0 |
| 27 | 61 | Male | 1 | Proximal | 35 | N/A | wild | wild | MSS | Low |
| | | | 2 | Distal | 20 | N/A | *KRAS* G12V | wild | MSS | Low |
| 28 | 72 | Male | 1 | Distal | 30 | N/A | *KRAS* G12D | wild | MSS | Low |
| | | | 2 | Distal | 20 | N/A | *KRAS* G13D | wild | MSS | High |
| 29 | 62 | Male | 1 | Distal | 20 | undifferentiation | wild | wild | MSS | High |
| | | | 2 | Proximal | 10 | N/A | wild | wild | MSS | Low |
| 30 | 64 | Male | 1 | Distal | 60 | undifferentiation | wild | wild | MSS | High |
| | | | 2 | Distal | 3 | N/A | wild | wild | MSS | High |
| 31 | 77 | Male | 1 | Distal | 45 | differentiation | wild | wild | MSS | High |
| | | | 2 | Distal | 25 | N/A | wild | wild | MSS | High |
| 32 | 76 | Female | 1 | Distal | 50 | N/A | *KRAS* G12D | wild | MSS | High |
| | | | 2 | Proximal | 50 | N/A | *KRAS* G12D | wild | MSS | High |
| 33 | 83 | Male | 1 | Distal | 50 | differentiation | wild | wild | MSS | High |
| | | | 2 | Distal | 15 | differentiation | wild | wild | MSS | High |
| 34 | 50 | Male | 1 | Distal | 40 | differentiation | wild | wild | MSS | High |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2 | Distal | 10 | N/A | wild | wild | MSS | High |
| 35 | 77 | Male | 1 | Distal | 15 | N/A | wild | wild | MSS | High |
| | | | 2 | Distal | 2 | N/A | *KRAS* K117N | wild | MSS | Low |
| 36 | 70 | Male | 1 | Distal | 40 | undifferentiation | *KRAS* K117N | wild | MSS | High |
| | | | 2 | Distal | 20 | differentiation | *KRAS* K117N | wild | MSS | High |
| 37 | 55 | Male | 1 | Distal | 9 | N/A | wild | wild | MSS | High |
| | | | 2 | Distal | 8 | N/A | wild | wild | MSS | High |
| 38 | 55 | Male | 1 | Distal | 50 | differentiation | *KRAS* G12D | wild | MSS | High |
| | | | 2 | Proximal | 20 | differentiation | wild | wild | MSS | High |

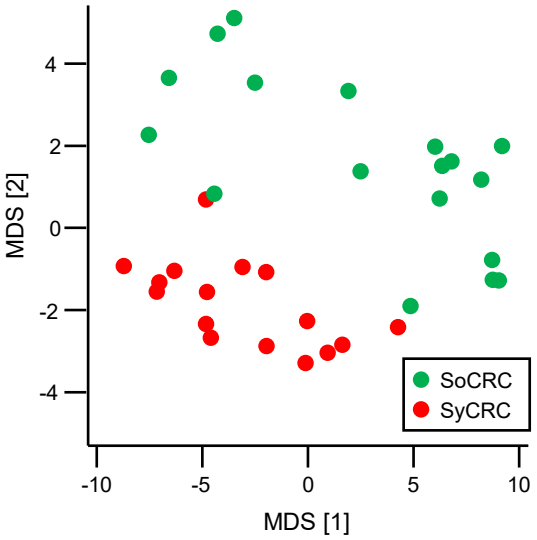
*MSI, microsatellite instability; MSS, microsatellite stable; MSI-H microsatellite instability-high; N/A, not available*

Figure 1

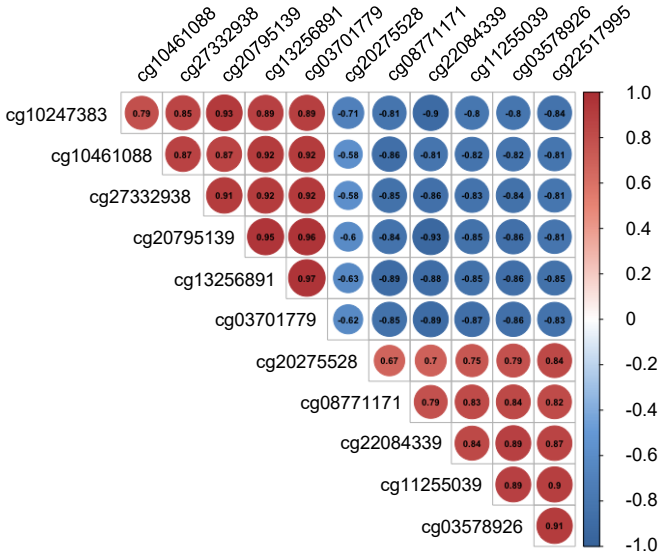
A




B




C




cg20275528 ***(SEPT9)***
cg03578926 ***(SHANK2)***
cg22084339 ***(PRKAR1B)***
cg27332938 ***(ZNF511)***
cg10461088 ***(ARFGAP2)***
cg11255039 ***(KIF22)***

D

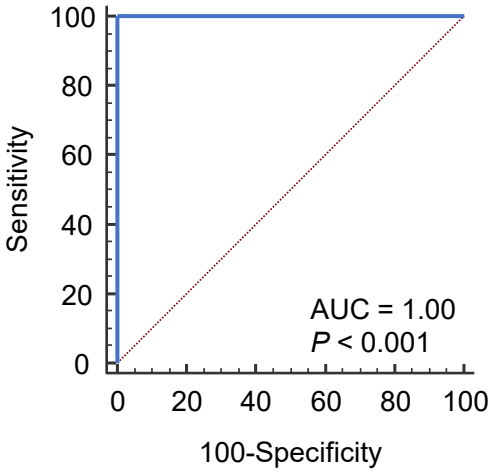


Figure 2

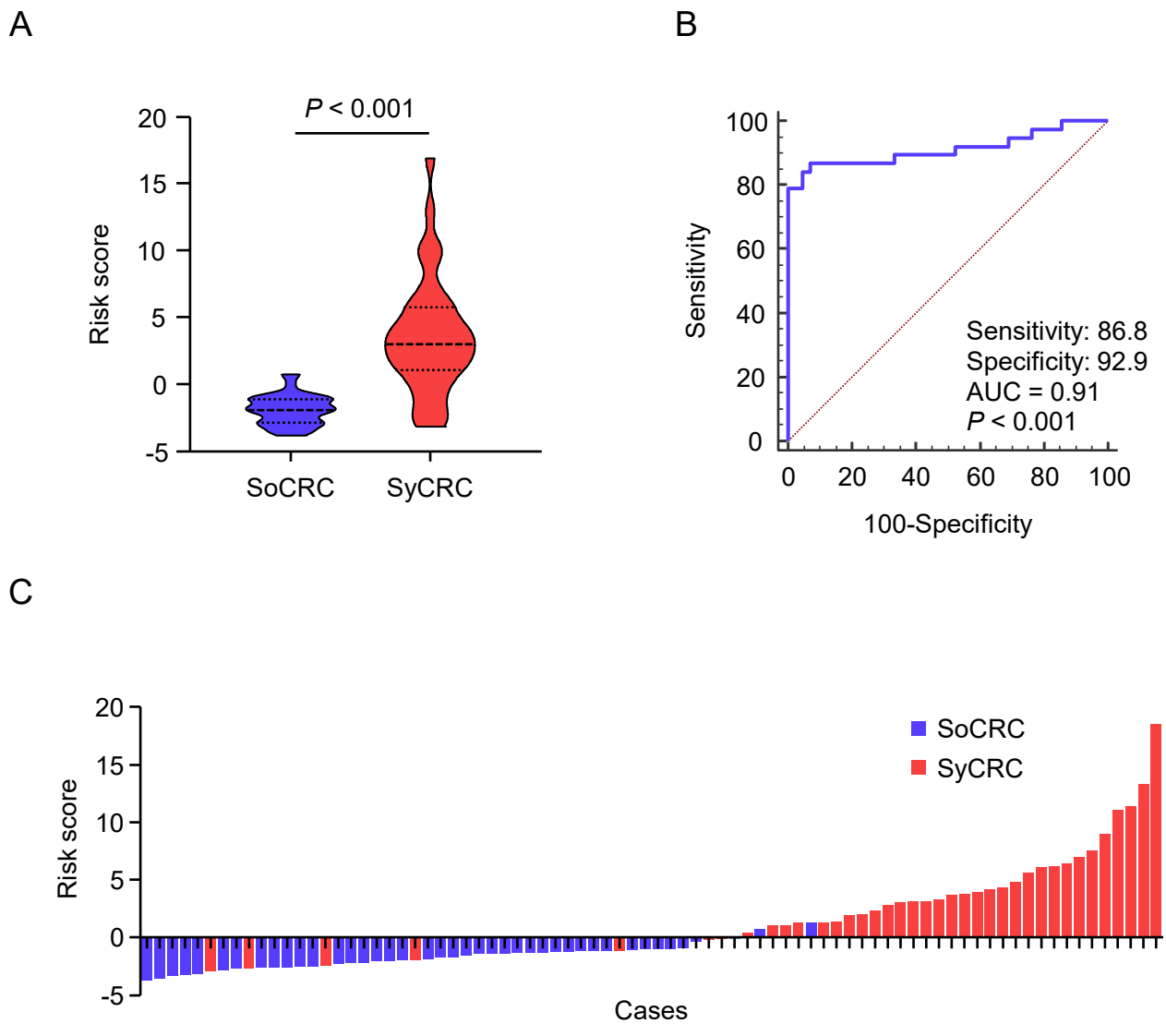
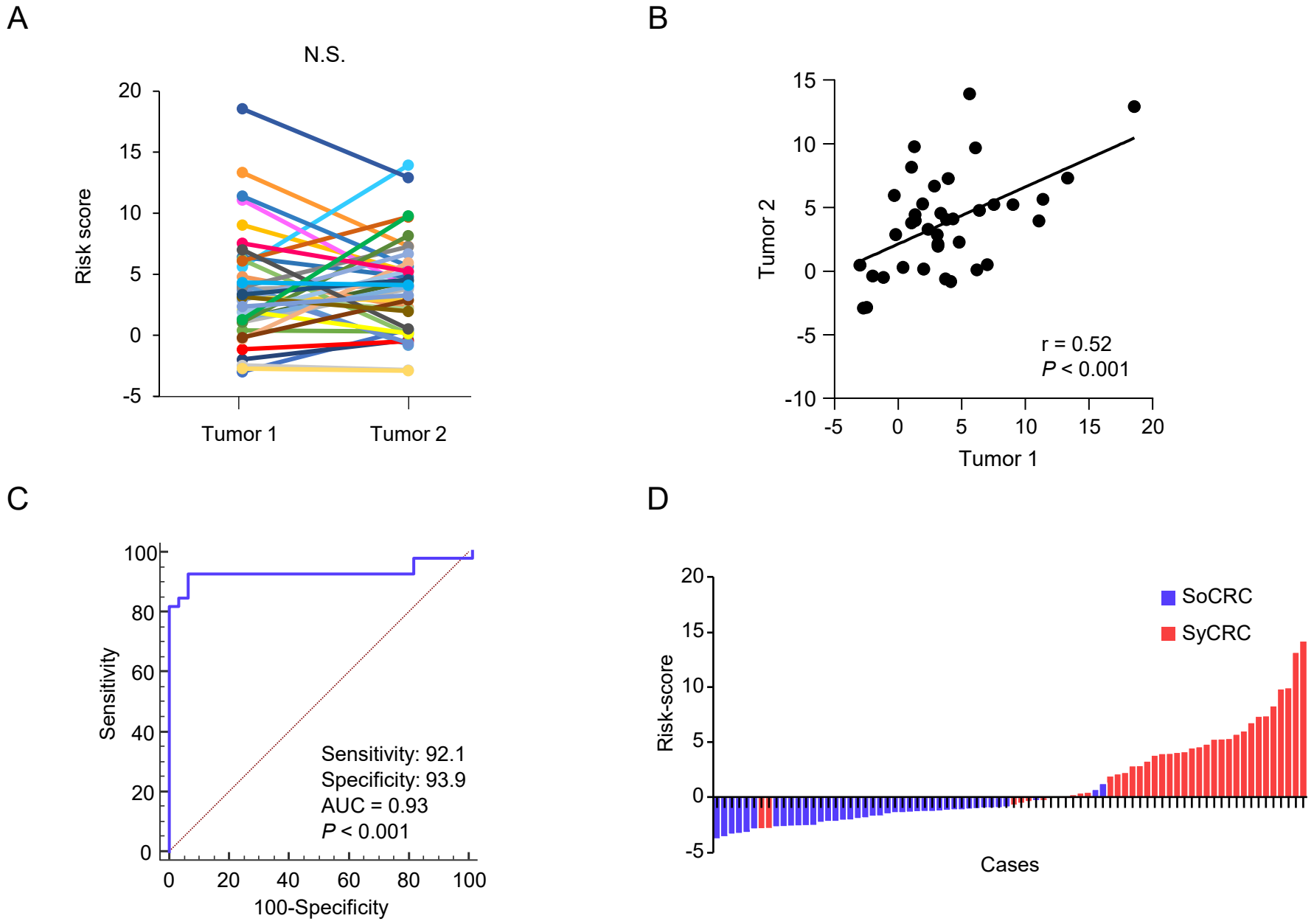
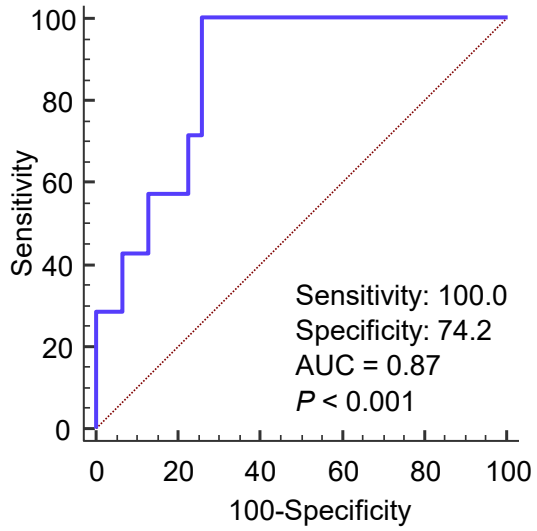

A
P < 0.001
Risk score
20
15
10
5
0
-5
SoCRC
SyCRC
B
Sensitivity
100
80
60
40
20
0
0
20
40
60
80
100
100-Specificity
Sensitivity: 86.8
Specificity: 92.9
AUC = 0.91
P < 0.001
C
Risk score
20
15
10
5
0
-5
SoCRC
SyCRC
Cases

Figure 3

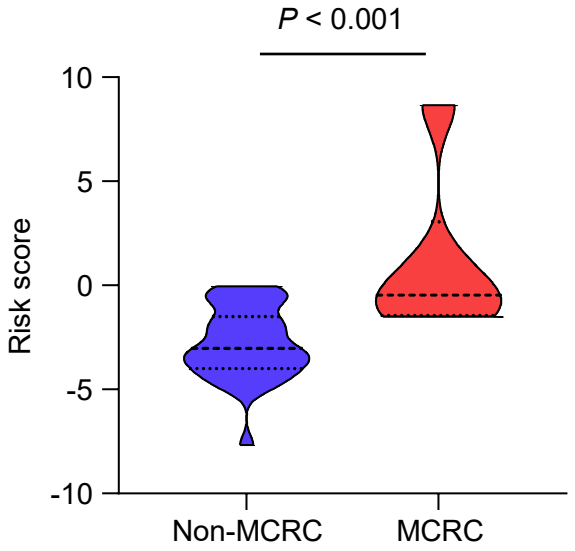

A
N.S.
Risk score
20
15
10
5
0
-5
Tumor 1
Tumor 2
B
Tumor 2
15
10
5
0
-5
-10
-5
0
5
10
15
20
Tumor 1
r = 0.52
P < 0.001
C
Sensitivity
100
80
60
40
20
0
0
20
40
60
80
100
100-Specificity
Sensitivity: 92.1
Specificity: 93.9
AUC = 0.93
P < 0.001
D
Risk-score
20
15
10
5
0
-5
Cases
SoCRC
SyCRC

# Figure 4

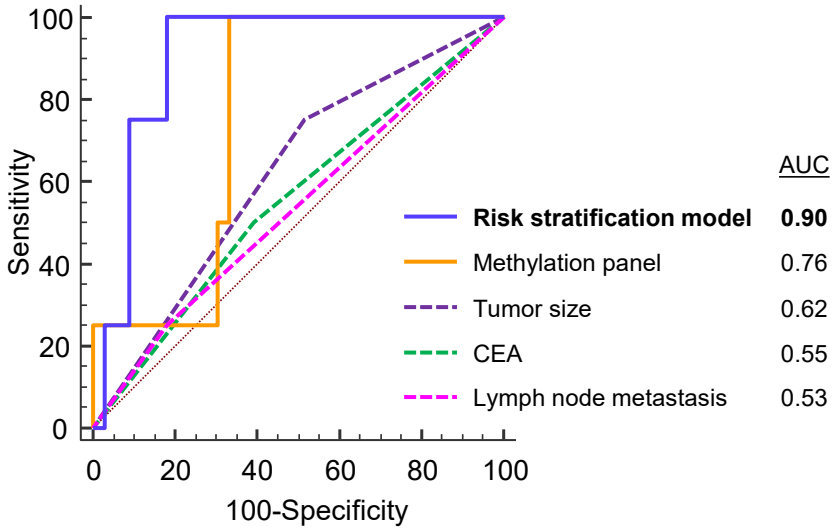
A



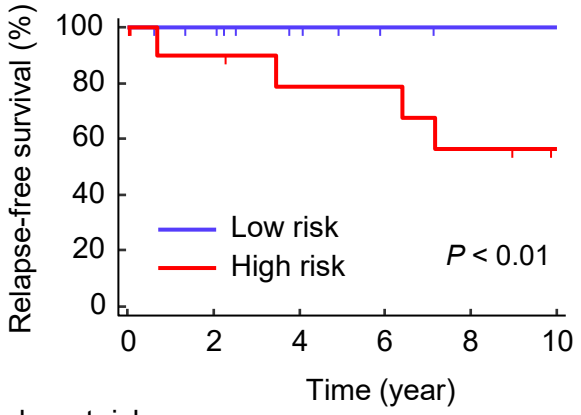

B




C




D




Number at risk

| | | | | | | |
|---|---|---|---|---|---|---|
| Low risk | 26 | 22 | 18 | 15 | 14 | 14 |
| High risk | 11 | 9 | 7 | 7 | 5 | 3 |