



## ORIGINAL ARTICLE

# Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum

Mai Yamauchi,<sup>1</sup> Teppei Morikawa,<sup>1</sup> Aya Kuchiba,<sup>1</sup> Yu Imamura,<sup>1</sup> Zhi Rong Qian,<sup>1</sup> Reiko Nishihara,<sup>1</sup> Xiaoyun Liao,<sup>1</sup> Levi Waldron,<sup>2,3</sup> Yujin Hoshida,<sup>4</sup> Curtis Huttenhower,<sup>2</sup> Andrew T Chan,<sup>5,6</sup> Edward Giovannucci,<sup>6,7,8</sup> Charles Fuchs,<sup>1,6</sup> Shuji Ogino<sup>1,9</sup>

► Additional tables are published online only. To view these files please visit the journal online (<http://gut.bmj.com/content/61/6.toc>).

For numbered affiliations see end of article.

## Correspondence to

Dr Shuji Ogino, Epidemiology, Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, 450 Brookline Avenue, Room JF-215C, Boston, MA 02215, USA; [shuji\\_ogino@dfci.harvard.edu](mailto:shuji_ogino@dfci.harvard.edu)

Revised 14 October 2011  
Accepted 22 October 2011  
Published Online First  
17 March 2012

## ABSTRACT

**Objective** Colorectal cancer is typically classified into proximal colon, distal colon and rectal cancer. Tumour genetic and epigenetic features differ by tumour location. Considering a possible role of bowel contents (including microbiome) in carcinogenesis, this study hypothesised that tumour molecular features might gradually change along bowel subsites, rather than change abruptly at splenic flexure.

**Design** Utilising 1443 colorectal cancers in two US nationwide prospective cohort studies, the frequencies of molecular features (CpG island methylator phenotype (CIMP), microsatellite instability (MSI), LINE-1 methylation and *BRAF*, *KRAS* and *PIK3CA* mutations) were examined along bowel subsites (rectum, rectosigmoid junction, sigmoid, descending colon, splenic flexure, transverse colon, hepatic flexure, ascending colon and caecum). The linearity and non-linearity of molecular relations along subsites were statistically tested by multivariate logistic or linear regression analysis.

**Results** The frequencies of CIMP-high, MSI-high and *BRAF* mutations gradually increased from the rectum (<2.3%) to ascending colon (36–40%), followed by falls in the caecum (12–22%). By linearity tests, these molecular relations were significantly linear from rectum to ascending colon ( $p < 0.0001$ ), and there was little evidence of non-linearity ( $p > 0.09$ ). Caecal cancers exhibited the highest frequency of *KRAS* mutations (52% vs 27–35% in other sites;  $p < 0.0001$ ).

**Conclusions** The frequencies of CIMP-high, MSI-high and *BRAF* mutations in cancer increased gradually along colorectum subsites from the rectum to ascending colon. These novel data challenge the common conception of discrete molecular features of proximal versus distal colorectal cancers, and have a substantial impact on clinical, translational and epidemiology research, which has typically been performed with the dichotomous classification of proximal versus distal tumours.

Over the past decades, clinical, pathological or epidemiological investigations into the large bowel have semi-automatically divided the colorectum into three compartments, namely, the rectum, distal colon and proximal colon.<sup>1–3</sup> In 1990, Bufill<sup>1</sup> proposed the existence of two distinct genetic

## Significance of this study

### What is already known about this subject?

- Colorectal cancer is typically classified into rectal, distal colon and proximal colon cancers.
- Proximal colon cancers and distal cancers differ in clinical, pathological and molecular features.
- Although it remains uncertain whether colorectal cancer molecular features change abruptly at splenic flexure, some investigators believe that there are distinct molecular features of proximal tumours and distal tumours.

### What are the new findings?

- The frequencies of CIMP-high, MSI-high and *BRAF* mutations in colorectal cancer increase gradually (statistically linearly) along the bowel from the rectum to ascending colon, rather than change abruptly at splenic flexure.
- Caecal cancers represent a unique subtype characterised by a high frequency of *KRAS* mutations, and caecal cancers do not follow the linearity trend in terms of the frequencies of CIMP-high, MSI-high and *BRAF* mutations.
- Mean tumour LINE-1 methylation levels show non-linear changes along the bowel subsites, and do not show an abrupt change at splenic flexure.

### How might it impact on clinical practice in the foreseeable future?

- Over the past decades, most clinical, translational and epidemiological studies have gathered and published colorectal tumour location data as proximal colon versus distal colon (vs rectum). Future studies on colorectal neoplastic diseases should include information on detailed bowel subsites (beyond proximal colon, distal colon and rectum), which will further improve our understanding of the mechanisms of colorectal carcinogenesis.

categories of colorectal cancers according to tumour location in the proximal or distal segment of the large bowel, divided at splenic flexure. This concept

of distinct molecular features of proximal cancer versus distal cancer has repeatedly been discussed.<sup>2-3</sup>

Colorectal cancers encompass a heterogeneous group of diseases with complex genetic and epigenetic alterations.<sup>4</sup> Molecular classification is thus increasingly important for clinical decision making.<sup>5</sup> Microsatellite instability (MSI) represents a distinct form of genomic instability.<sup>5-6</sup> The CpG island methylator phenotype (CIMP) is a distinct form of epigenomic instability,<sup>7-17</sup> which causes most sporadic MSI-high colorectal cancers through epigenetic inactivation of *MLH1*.<sup>18-21</sup> Independent of MSI, CIMP-high colorectal cancer is associated with proximal tumour location, old age of onset, female sex and the *BRAF* mutation.<sup>18-19</sup>

Accumulating evidence suggests that proximal colon cancer and distal colon cancer differ in various molecular features including CIMP and MSI.<sup>22-25</sup> However, it remains uncertain whether the tumour molecular features change abruptly at splenic flexure. Considering a possible role of bowel contents (including microbiome) in colorectal carcinogenesis,<sup>26</sup> we hypothesised that tumour molecular characteristics might change gradually along the large bowel. This hypothesis is not inconsistent with the differences between proximal versus distal cancers, as long as tumour molecular features change along the large bowel.

To test the hypothesis, we conducted this study utilising a database of 1443 colorectal cancers in two prospective cohort studies. We examined the frequencies of relevant molecular features along the bowel subsites (rectum, rectosigmoid, sigmoid colon, descending colon, splenic flexure, transverse colon, hepatic flexure, ascending colon and caecum), and statistically assessed the linearity and non-linearity of molecular relations

along the bowel subsites. Our novel findings of gradual increases of CIMP-high, MSI-high and *BRAF* mutation from the rectum to ascending colon challenge the common conception of discrete dichotomy of tumour molecular features in proximal colon versus distal colorectum.

## MATERIALS AND METHODS

### Study group

We utilised the database of two independent, prospective cohort studies; the Nurses' Health Study (N=121 701 women followed since 1976) and the Health Professionals Follow-up Study (N=51 529 men followed since 1986).<sup>27-28</sup> Every 2 years, participants have been sent follow-up questionnaires to update information on potential risk factors and to identify newly diagnosed cancers in themselves and their first-degree relatives. In addition, we searched the National Death Index for those who died of colorectal cancer. Our study physicians reviewed medical records and obtained information on disease stage and tumour location (rectum, rectosigmoid, sigmoid colon, descending colon, splenic flexure, transverse colon, hepatic flexure, ascending colon and caecum). We collected paraffin-embedded tissue blocks from hospitals where patients underwent tumour resections. We collected diagnostic biopsy specimens for rectal cancer patients who received preoperative treatment, in order to avoid artifacts or bias introduced by treatment. Based on the availability of adequate tissue specimens, 1443 colorectal cancer cases (diagnosed up to 2006) were included (tables 1 and 2). Among our cohort studies, there was

**Table 1** Clinical characteristics of colorectal cancer according to tumour location bowel subsite

Clinical feature	Total N	Caecum	Ascending colon	Hepatic flexure	Transverse colon	Splenic flexure	Descending colon	Sigmoid colon	Rectosigmoid	Rectum	p Value
All cases	1443	243	295	46	91	33	83	314	106	232	
Sex											<0.0001
Male (HPFS)	649 (45%)	129 (53%)	95 (32%)	28 (61%)	34 (37%)	18 (55%)	34 (41%)	149 (47%)	51 (48%)	111 (48%)	
Female (NHS)	794 (55%)	114 (47%)	200 (68%)	18 (39%)	57 (63%)	15 (46%)	49 (59%)	165 (53%)	55 (52%)	121 (52%)	
Mean age ±SD	68.4±8.9	69.2±8.6	70.9±8.4	71.3±8.8	67.0±8.4	67.2±9.6	65.8±8.8	67.4±8.3	67.4±9.1	67.4±9.6	<0.0001
Body mass index											0.35
<30 kg/m <sup>2</sup>	1161 (81%)	209 (86%)	241 (83%)	35 (76%)	72 (79%)	24 (75%)	65 (78%)	246 (79%)	83 (81%)	186 (81%)	
≥30 kg/m <sup>2</sup>	269 (19%)	33 (14%)	48 (17%)	11 (24%)	19 (21%)	8 (25%)	18 (22%)	67 (21%)	20 (19%)	45 (19%)	
Year of diagnosis											0.0024
Before 1995	459 (32%)	65 (27%)	75 (25%)	10 (22%)	32 (35%)	13 (39%)	29 (35%)	124 (39%)	28 (26%)	83 (36%)	
1995–2006	984 (68%)	178 (73%)	220 (75%)	36 (78%)	59 (65%)	20 (61%)	54 (65%)	190 (61%)	78 (74%)	149 (64%)	
Family history of colorectal cancer in any first-degree relative											0.15
Absent	1162 (81%)	190 (79%)	225 (78%)	38 (83%)	73 (80%)	23 (72%)	69 (83%)	254 (81%)	90 (87%)	200 (87%)	
Present	270 (19%)	52 (21%)	65 (22%)	8 (17%)	18 (20%)	9 (28%)	14 (17%)	59 (19%)	14 (13%)	31 (13%)	
Prediagnosis physical activity (MET-hours/week)											0.058
<18	803 (59%)	129 (55%)	184 (65%)	20 (43%)	44 (53%)	18 (60%)	50 (63%)	182 (63%)	56 (55%)	120 (57%)	
≥18	550 (41%)	104 (45%)	97 (35%)	26 (57%)	39 (47%)	12 (40%)	29 (37%)	108 (37%)	45 (45%)	90 (43%)	
Prediagnosis smoking status											0.16
Never	571 (41%)	108 (45%)	108 (38%)	19 (41%)	36 (40%)	12 (38%)	24 (29%)	141 (45%)	38 (38%)	85 (38%)	
Former/current smokers	839 (60%)	131 (55%)	178 (62%)	27 (59%)	54 (60%)	20 (63%)	59 (71%)	169 (55%)	62 (62%)	139 (62%)	
Prediagnosis alcohol consumption											0.50
No	468 (33%)	84 (35%)	104 (36%)	12 (26%)	31 (34%)	14 (44%)	18 (22%)	98 (31%)	32 (31%)	75 (33%)	
<15 g/day	661 (46%)	113 (47%)	132 (46%)	21 (46%)	44 (48%)	11 (34%)	45 (54%)	146 (47%)	52 (50%)	95 (43%)	
≥15 g/day	298 (21%)	45 (19%)	52 (18%)	13 (28%)	16 (18%)	7 (22%)	20 (24%)	69 (22%)	20 (19%)	56 (25%)	

(%) Indicates the proportion of cases with a specific clinical feature among cancers located in each subsite (caecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, rectosigmoid or rectum). p Values were calculated by the  $\chi^2$  test, except for age (one-way analysis of variance test).

HPFS, Health Professionals Follow-up Study; MET, metabolic equivalent task; NHS, Nurses' Health Study.

**Table 2** Pathological characteristics of colorectal cancer according to tumour location bowel subsite

Pathological feature	Total N	Caecum	Ascending colon	Hepatic flexure	Transverse colon	Splenic flexure	Descending colon	Sigmoid colon	Rectosigmoid	Rectum	p Value
Disease stage											<0.0001
I	339 (26%)	53 (24%)	59 (22%)	12 (29%)	16 (18%)	7 (22%)	12 (16%)	84 (31%)	33 (34%)	63 (33%)	
II	403 (31%)	75 (34%)	108 (40%)	15 (36%)	35 (40%)	9 (28%)	32 (43%)	67 (25%)	24 (25%)	38 (20%)	
III	357 (28%)	57 (26%)	61 (22%)	9 (21%)	21 (24%)	9 (28%)	25 (34%)	77 (28%)	30 (31%)	68 (36%)	
IV	187 (15%)	36 (16%)	44 (16%)	6 (14%)	16 (18%)	7 (22%)	5 (6.8%)	44 (16%)	9 (9.4%)	20 (11%)	
Tumour differentiation											<0.0001
Well to moderate	1286 (90%)	214 (88%)	234 (80%)	37 (80%)	77 (86%)	28 (85%)	81 (98%)	298 (96%)	103 (98%)	214 (96%)	
Poor	140 (9.8%)	28 (12%)	58 (20%)	9 (20%)	13 (14%)	5 (15%)	2 (2.4%)	13 (4.2%)	2 (1.9%)	10 (4.5%)	
Mucinous component											<0.0001
0%	836 (62%)	116 (49%)	138 (48%)	18 (45%)	41 (48%)	17 (52%)	57 (71%)	217 (75%)	71 (75%)	161 (79%)	
1–49%	340 (25%)	82 (35%)	93 (33%)	10 (25%)	27 (31%)	12 (36%)	16 (20%)	52 (18%)	19 (20%)	29 (14%)	
≥50%	175 (13%)	39 (16%)	55 (19%)	12 (30%)	18 (21%)	4 (12%)	7 (8.8%)	22 (7.6%)	5 (5.3%)	13 (6.4%)	
Signet ring cell component											<0.0001
0%	1201 (89%)	207 (88%)	224 (79%)	33 (83%)	70 (82%)	32 (97%)	75 (94%)	278 (96%)	94 (99%)	189 (93%)	
1–49%	123 (9.1%)	27 (11%)	50 (18%)	7 (18%)	12 (14%)	1 (3.0%)	5 (6.3%)	10 (3.4%)	1 (1.1%)	10 (4.9%)	
≥50%	23 (1.7%)	2 (0.9%)	11 (3.9%)	0	3 (3.5%)	0	0	3 (1.0%)	0	4 (2.0%)	

(%) Indicates the proportion of cases with a specific pathological feature among cancers located in each subsite (caecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, rectosigmoid or rectum). p Values were calculated by the  $\chi^2$  test.

no significant difference in demographic features between cases with tissue available and those without available tissue.<sup>27</sup> This current study represents a new analysis of tumour molecular features along the detailed bowel subsites on the existing colorectal cancer database that has previously been characterised for CIMP, MSI, LINE-1 methylation and *BRAF* and *KRAS* mutations.<sup>29–30</sup> Informed consent was obtained from all study subjects. This study was approved by the Harvard School of Public Health and Brigham and Women's Hospital institutional review boards.

### Assessment of physical activity

Leisure-time physical activity has been assessed every 2 years. Subjects reported duration of participation (ranging from 0 to 11 h or more per week) on walking (along with usual pace), jogging, running, bicycling, swimming laps, racket sports, other aerobic exercises, lower intensity exercise (yoga, toning, stretching), or other vigorous activities. Each activity on the questionnaire was assigned a metabolic equivalent task (MET) score. One MET is the energy expenditure for sitting quietly. MET scores are defined as the ratio of the metabolic rate associated with specific activities divided by the resting metabolic rate. The values from the individual activities were summed for a total MET-hours per week score.

### Assessment of cigarette smoking and alcohol consumption

Cigarette smoking has been assessed every 2 years in both cohorts. Alcohol consumption was the sum of the values for three types of beverages: beer, wine and spirits. We assumed an ethanol content of 13.1 g for a 12 oz (38 dl) can or bottle of beer, 11.0 g for a 4 oz (12 dl) glass of wine and 14.0 g for a standard portion of spirits.

### Histopathological evaluations

Tissue sections from all colorectal cancer cases were reviewed by a pathologist (SO) unaware of other data. Tumour differentiation was categorised as well–moderate versus poor (>50% vs ≤50% glandular area). Extent of mucin and signet ring cells were recorded.

### Sequencing of *BRAF*, *KRAS* and *PIK3CA*, and MSI analysis

DNA was extracted from tumour and PCR, and pyrosequencing targeted for *BRAF* (codon 600),<sup>31</sup> *KRAS* (codons 12 and 13)<sup>32</sup> and *PIK3CA* (exons 9 and 20) were performed as previously described.<sup>33</sup> MSI analysis was performed, using 10 microsatellite markers (BAT25, BAT26, BAT40, D2S123, D5S346, D17S250, D18S55, D18S56, D18S67 and D18S487).<sup>30</sup> MSI-high was defined as the presence of instability in 30% or more of the markers. MSI-low (1–29% unstable markers) tumours were grouped into microsatellite stable (MSS) tumours (no unstable markers) because those showed similar features.

### Methylation analyses for CpG islands and LINE-1

Using real-time PCR (MethylLight)<sup>34</sup> on bisulfite-treated DNA,<sup>35</sup> we quantified DNA methylation in eight CIMP-specific promoters (*CACNA1G*, *CDKN2A (p16)*, *CRABP1*, *IGF2*, *MLH1*, *NEUROG1*, *RUNX3* and *SOCS1*).<sup>9 18 36</sup> CIMP-high was defined as the presence of six/eight or more methylated promoters, and CIMP-low/zero as zero/eight to five/eight methylated promoters, according to the previously established criteria.<sup>18 36</sup> In order to quantify methylation levels accurately in LINE-1 repetitive elements we utilised pyrosequencing, as previously described.<sup>37 38</sup>

### Analysis of gene expression

RNA was extracted and gene expression profiling was performed according to the complementary DNA-mediated annealing, selection, extension and ligation assay (Illumina, San Diego, California, USA), as previously described.<sup>39</sup>

### Statistical analysis

For all statistical analyses, we used SAS software (V9.1.3). All p values were two-sided. For categorical data, the  $\chi^2$  test was performed. One-way analysis of variance was used to compare mean age or mean LINE-1 methylation level across bowel subsites. The Kruskal–Wallis test was used to compare the *ABCB1* expression levels across bowel subsites.

To test the linearity and non-linearity of the relationship of tumour location–molecular feature along bowel subsites,

multivariate logistic regression analysis (or linear regression analysis for LINE-1 methylation level) was performed. First, a numerical subsite location variable that represented an average distance (cm) from anal verge to each subsite was made, utilising recent CT colonography data.<sup>40</sup> In the logistic or linear regression model with a tumour molecular variable as an outcome variable, a significant p value by the Wald test on the bowel subsite variable indicated a linear relationship of the molecular variable along the bowel subsites, but a non-linear relationship might be present. To test non-linearity of the relationship along the bowel subsites, we used a likelihood ratio test (LRT) comparing the model with squared and/or cubic subsite variables with the model without squared or cubic subsite variables. With a significant p value by the Wald test (mentioned above), a non-significant LRT p value would support a linear relationship excluding non-linearity, while a significant LRT p value would indicate the presence of non-linearity. All logistic and linear regression models were adjusted for age (continuous), sex, year of diagnosis (continuous), family history of colorectal cancer in any first-degree relative (present vs absent), body mass index (<30 vs ≥30 kg/m<sup>2</sup>), physical activity (<18 vs ≥18 MET-hours/week), smoking (never vs former/current smokers) and alcohol consumption (no vs <15 vs ≥15 g/day). For cases with missing information in any of the covariates (family history of colorectal cancer (0.9%), body mass index (0.8%), physical activity (5.1%), smoking (1.1%)), we included those cases in a majority category of a given covariate to avoid overfitting. We confirmed that excluding cases with missing information in any of the covariates did not substantially alter the results (data not shown).

## RESULTS

### Colorectal cancer molecular features along bowel subsites

To assess the frequencies of various tumour molecular features along the bowel subsites (rectum, rectosigmoid, sigmoid colon, descending colon, splenic flexure, transverse colon, hepatic flexure, ascending colon and caecum), we examined the database of 1443 colorectal cancer cases (excluding appendiceal cancers) in the two prospective cohort studies. Table 3 and supplementary tables 1 and 2 (available online only) show the frequencies of various clinical, pathological or molecular features along the bowel subsites in our subject population. The frequencies of CIMP-high, MSI-high and *BRAF* mutation gradually increased from the rectum (<2.3%) to the ascending colon (36–40%; figure 1), supporting our hypothesis that these tumour molecular features might change gradually along the large bowel. There was no abrupt change at splenic flexure. Caecal cancers showed lower frequencies of CIMP-high, MSI-high and *BRAF* mutation (12–22%) than ascending colon cancers. Notably, caecal cancers showed a higher frequency of *KRAS* mutations (52%) than any other sites (27–35%;  $p<0.0001$ ).

Although there was no striking pattern of *PIK3CA* mutation frequency along bowel subsites, it was generally low in the rectum and rectosigmoid (10–11%) and higher proximally ( $p=0.0016$ ).

With regard to the tumour LINE-1 methylation level (mean±SD), it gradually decreased from the rectum ( $63.2\pm 8.5$ ) to descending colon ( $60.2\pm 11.7$ ), and then increased from descending colon to ascending colon ( $64.7\pm 9.4$ ;  $p=0.0003$ ). Again, there was no abrupt change at splenic flexure.

There was no significant relationship between bowel subsites and *ABCB1* expression level ( $p=0.19$ ).

Considering the importance of molecular classification based on combined CIMP and MSI status,<sup>41</sup> we also examined the

frequency of each CIMP/MSI subtype along bowel subsites (figure 2). The frequency of CIMP-high MSI-high tumours increased gradually along the bowel subsites from the rectum to ascending colon, while the frequency of CIMP-low/zero MSS tumours decreased from the rectum to ascending colon. There was no abrupt change at splenic flexure.

### Assessment of linearity of tumour location–molecular relationship

We assessed the linearity of tumour location–molecular relationship along the bowel subsites by a multivariate logistic regression model (or linear regression model for LINE-1 methylation) (table 4).

In our multivariate analysis strategy, we could assess whether data presented in table 3 and figure 1 were independent of other variables. We used bowel subsite as a predictor (independent) variable, and a molecular feature as an outcome (dependent) variable. When we assessed the relationship between subsite (rectum to ascending colon) and CIMP, bowel subsite was significantly linearly associated with CIMP-high ( $p<0.0001$ ). To assess non-linearity, we performed LRT comparing a model with squared and/or cubic subsite variable(s) with a model without squared or cubic variables. As a result, LRT yielded  $p>0.09$ , excluding a non-linear relationship and supporting a linear relationship of bowel subsites with CIMP-high.

When we assessed the relationship between subsite (rectum to ascending colon) and MSI (or *BRAF* mutation) by logistic regression models (table 4), the results were similar to those on the relationship between bowel subsite and CIMP. Tumour location bowel subsite (from rectum to ascending colon) was significantly linearly associated with MSI-high or *BRAF* mutation ( $p<0.0001$ ). In addition, bowel subsite was also linearly associated with *PIK3CA* mutation ( $p=0.0034$ ). When assessing non-linearity, LRT comparing a model with squared and/or cubic subsite variable(s) with a model without squared or cubic variables yielded non-significant p values ( $p>0.19$ ), excluding non-linearity and supporting a linear relationship between subsite and MSI (or *BRAF* mutation or *PIK3CA* mutation).

To exclude a potential influence of differential selection bias as a result of preoperative treatment for rectal cancers, we excluded cancers in the rectum and rectosigmoid and performed a linearity test. Tumour location bowel subsite (sigmoid colon to ascending colon) was significantly linearly associated with CIMP-high, MSI-high, or *BRAF* mutation ( $p<0.0001$ ), but not with *PIK3CA* mutation ( $p=0.13$ ), and there was no evidence for non-linearity (LRT  $p>0.05$ ).

## DISCUSSION

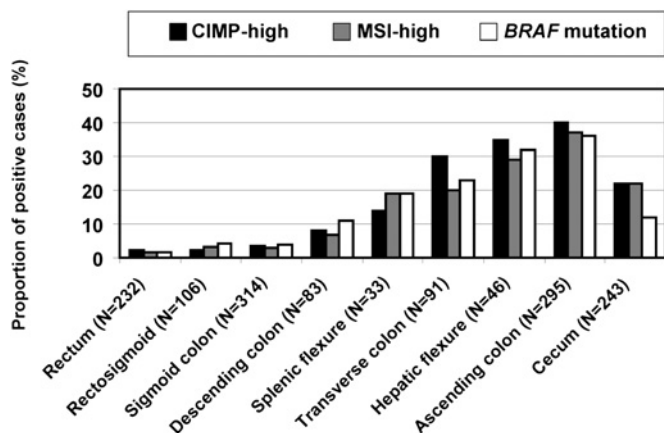
We performed this study to test the hypothesis that molecular features of colorectal cancer change gradually along bowel subsites, rather than change abruptly at splenic flexure. Accumulating evidence suggests that proximal colon cancers differ in clinical, pathological and molecular features from distal cancers.<sup>22–25</sup> However, it is still uncertain whether those features change abruptly at splenic flexure. Utilising the tumour database in the two prospective cohort studies, our current study is unique in examining tumour molecular features along the detailed bowel subsites (rectum, rectosigmoid, sigmoid colon, descending colon, splenic flexure, transverse colon, hepatic flexure, ascending colon and caecum). Notably, we found that the frequencies of CIMP-high, MSI-high and *BRAF* mutation increased (statistically) linearly along the bowel from the rectum to ascending colon. These data support our hypothesis of

**Table 3** Molecular characteristics of colorectal cancer according to tumour location bowel subsite

Molecular feature	Total N	Caecum	Ascending colon	Hepatic flexure	Transverse colon	Splenic flexure	Descending colon	Sigmoid colon	Rectosigmoid	Rectum	p Value
CIMP status											
CIMP-low/zero	1010 (83%)	160 (78%)	152 (60%)	24 (65%)	54 (70%)	24 (86%)	68 (92%)	266 (96%)	85 (98%)	177 (98%)	<0.0001
CIMP-high	208 (17%)	44 (22%)	102 (40%)	13 (35%)	23 (30%)	4 (14%)	6 (8.1%)	10 (3.6%)	2 (2.3%)	4 (2.2%)	
MSI status											
MSS	1061 (84%)	163 (78%)	165 (63%)	27 (71%)	64 (80%)	25 (81%)	70 (93%)	275 (97%)	87 (97%)	183 (98%)	<0.0001
MSI-high	196 (16%)	46 (22%)	98 (37%)	11 (29%)	16 (20%)	6 (19%)	5 (6.7%)	8 (2.8%)	3 (3.3%)	3 (1.6%)	
BRAF mutation											
(-)	1093 (86%)	186 (88%)	172 (64%)	25 (68%)	62 (78%)	25 (81%)	68 (89%)	276 (96%)	90 (96%)	189 (98%)	<0.0001
(+)	183 (14%)	26 (12%)	95 (36%)	12 (32%)	18 (23%)	6 (19%)	8 (11%)	11 (3.8%)	4 (4.3%)	3 (1.6%)	
KRAS mutation											
(-)	819 (64%)	101 (48%)	182 (68%)	26 (70%)	59 (73%)	21 (66%)	50 (66%)	186 (65%)	63 (68%)	131 (68%)	<0.0001
(+)	458 (36%)	111 (52%)	85 (32%)	11 (30%)	22 (27%)	11 (34%)	26 (34%)	101 (35%)	30 (32%)	61 (32%)	
PIK3CA mutation											
(-)	962 (82%)	146 (75%)	191 (78%)	31 (84%)	62 (85%)	18 (64%)	57 (80%)	220 (83%)	77 (90%)	160 (89%)	0.0016
(+)	217 (18%)	49 (25%)	54 (22%)	6 (16%)	11 (15%)	10 (36%)	14 (20%)	44 (17%)	9 (10%)	20 (11%)	
LINE-1 methylation level (mean±SD)	62.7±9.5	63.7±9.0	64.7±9.4	62.4±8.7	62.5±10.0	61.9±10.2	60.2±11.7	61.3±9.1	61.0±10.1	63.2±8.5	0.0003
ABCB7 expression level (log <sub>2</sub> intensity)	6.47 (6.35–6.81)	6.46 (6.35–6.71)	6.43 (6.35–6.73)	6.43 (6.35–6.68)	6.44 (6.35–6.67)	6.48 (6.38–6.91)	6.60 (6.36–6.84)	6.51 (6.37–6.91)	6.53 (6.35–6.82)	6.51 (6.39–6.80)	0.19

(%) Indicates the proportion of cases with a specific molecular feature among cancers located in each subsite (caecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, rectosigmoid or rectum). p Values were calculated by the  $\chi^2$  test, except for LINE-1 (one-way analysis of variance test) and ABCB7 (Kruskal–Wallis test).

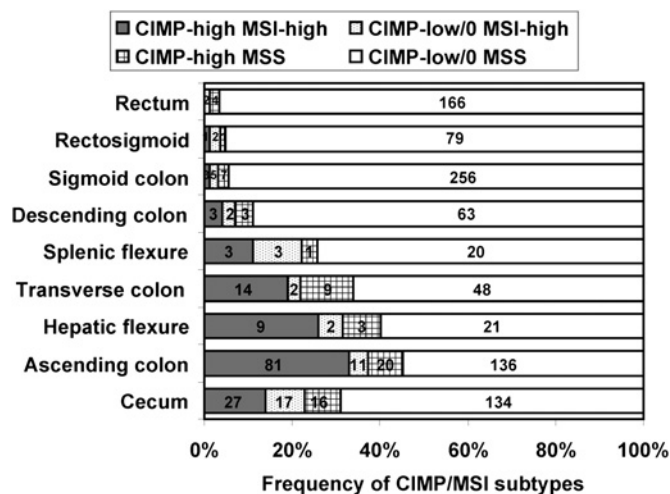
CIMP, CpG island methylator phenotype; MSI, microsatellite instability; MSS, microsatellite stable.



**Figure 1** Frequencies of CIMP-high, MSI-high, and *BRAF* mutation in colorectal cancer along bowel subsites. The frequencies of these molecular features increase gradually from the rectum to the ascending colon. Formal multivariate statistical analyses for linearity and non-linearity were performed as described in the Materials and methods section and the results are shown in table 4. CIMP, CpG island methylator phenotype; MSI, microsatellite instability.

gradual changes in tumour molecular features along the bowel subsites, rather than abrupt changes at splenic flexure. Importantly, our hypothesis and data are not inconsistent with repeated observations of differences in molecular features (such as CIMP and MSI) between proximal colon cancer and distal colorectal cancer,<sup>22–25</sup> so long as molecular features change along the bowel subsites.

Examining molecular changes in colorectal neoplasias is increasingly important for a better understanding of the carcinogenic process.<sup>42–44</sup> In past decades, colorectal cancers were typically divided into three compartments: rectum, distal colon (sigmoid to splenic flexure) and proximal colon (transverse colon



**Figure 2** Frequencies of CIMP/MSI subtypes of colorectal cancer along bowel subsites. The frequency of CIMP-high MSI-high tumours increased gradually from the rectum to the ascending colon, while that of CIMP-low/zero MSS tumours decreased gradually from the rectum to the ascending colon. CIMP, CpG island methylator phenotype; MSI, microsatellite instability; MSS, microsatellite stable.

to caecum) in most clinical, pathological and epidemiological publications.<sup>1–3</sup> As a result, our epidemiological, clinical and molecular pathological knowledge on colorectal cancer in detailed bowel subsites is currently deficient. Therefore, our data demonstrating gradual changes in tumour molecular features along the bowel may have considerable implications in clinical, epidemiological and pathological research. We would propose that future studies on colorectal neoplasia should include information on detailed bowel subsites (beyond the proximal colon, distal colon and rectum), which will further improve our understanding of the mechanisms of colorectal carcinogenesis.

**Table 4** Assessment of linearity and non-linearity on subsite–molecular relationship in colorectal cancer by multivariate logistic or linear regression model

Multivariate regression model	Outcome variable (molecular feature)	Bowel subsite variable (from rectum to ascending colon)		Squared subsite variable		Cubic subsite variable		LRT	
		p Value (Wald test)	Included	p Value (Wald test)	Included	p Value (Wald test)	Degrees of freedom	p Value*	
Logistic	CIMP	<0.0001	No	—	No	—	—	—	Referent
		0.0017	Yes	0.17	No	—	1	0.17	
		0.68	Yes	0.15	Yes	0.096	2	0.098	
Logistic	MSI	<0.0001	No	—	No	—	—	—	Referent
		0.020	Yes	0.56	No	—	1	0.56	
		0.93	Yes	0.48	Yes	0.42	2	0.61	
Logistic	<i>BRAF</i> mutation	<0.0001	No	—	No	—	—	—	Referent
		0.0041	Yes	0.26	No	—	1	0.26	
		0.56	Yes	0.76	Yes	0.63	2	0.47	
Logistic	<i>KRAS</i> mutation	0.66	No	—	No	—	—	—	Referent
		0.48	Yes	0.42	No	—	1	0.42	
		0.16	Yes	0.19	Yes	0.23	2	0.35	
Logistic	<i>PIK3CA</i> mutation	0.0034	No	—	No	—	—	—	Referent
		0.070	Yes	0.20	No	—	1	0.20	
		0.096	Yes	0.23	Yes	0.30	2	0.26	
Linear	LINE-1 methylation level	0.020	No	—	No	—	—	—	Referent
		0.0070	Yes	0.0006	No	—	1	0.0036	
		0.036	Yes	0.0022	Yes	<0.0001	2	<0.0001	

A multivariate regression model included age, sex, year of diagnosis, family history of colorectal cancer, body mass index, physical activity, smoking, alcohol consumption and the bowel subsite variable with or without the squared and cubic subsite variables, as indicated in the table.

\*A significant p value by the LRT indicates a non-linearity, and a combination of insignificant p values by LRT and a significant p value by the Wald test on the subsite variable in the model without the squared or cubic location variable indicates a linear relationship.

CIMP, CpG island methylator phenotype; LRT, likelihood ratio test; MSI, microsatellite instability.

Colorectal epithelial cells are constantly in contact with bowel contents, which may play a critical role in cellular transformation and tumour development and progression. Bowel contents (food debris, microbiome and bacterial fermentation products) and their interactions with host cells (epithelial and immune cells) may directly cause cellular molecular changes, or alternatively, may influence tumour progression differentially according to molecular features in preneoplastic or premalignant cells.<sup>45–46</sup> In fact, bowel contents gradually change along the bowel subsites, and this fact may explain why tumour molecular features change gradually along the bowel subsites. In support of this hypothesis, studies on synchronous primary colorectal cancers have shown that CIMP-high (or MSI-high or *BRAF*-mutated) proximal cancer may co-exist with CIMP-negative (or MSS or *BRAF* wild type) distal cancer,<sup>47–50</sup> and another study has shown a gradual gradient of CpG island methylation along normal bowel mucosa.<sup>51</sup> Together with these data, our current study supports the role of bowel contents in predisposing colon epithelial cells to certain molecular insults. However, further investigations such as identifying the components of bowel contents or factors participating in the host–bacterial interactions are needed to understand how colorectal cancer develops.

The ATP-binding cassette (ABC) transporters constitute a large family of active transporter molecules, and play a role in the process of absorption. Because of the diverse substrates that can be transported, ABC proteins are found to be expressed in a number of specialised cell types.<sup>52</sup> ABCB1 has been known to play a critical role in host–bacterial interactions in the gastrointestinal tract,<sup>53</sup> and has been implicated in colorectal cancer development and progression.<sup>54</sup> Potocnik *et al*<sup>55</sup> have shown that *ABCB1* gene polymorphisms may be associated with MSI-high cancer. Although our current study did not show that bowel subsite was significantly associated with *ABCB1* expression in colorectal cancer, ABC transporters may play roles in modifying the risks of colorectal epithelial cells for neoplastic transformation/progression differentially according to the cellular molecular status.

Interestingly, our data indicate that caecal cancers have unique molecular features different from cancers in other subsites. The frequency of the *KRAS* mutation was highest in caecal cancers among all subsites. In addition, for the relations of bowel subsites with the frequencies of CIMP-high, MSI-high and *BRAF* mutation, caecal cancers did not follow the trend of an increase from the rectum to ascending colon. Kucherlapati *et al*<sup>56</sup> have shown that loss of *Rb1* in the gastrointestinal tract of *Apc*<sup>1638N</sup> mice promotes caecal tumour formation. Loss of RB1 (retinoblastoma protein) has been found specifically in caecal cancers.<sup>57</sup> Taken together, caecal cancer may arise through somewhat unique carcinogenic mechanisms different from cancers in other subsites.

There are advantages in utilising the database of the two US nationwide prospective cohort studies to study molecular features of colorectal cancer along bowel subsites. Our large database readily enabled us to examine the frequencies of various molecular features in cancers in each bowel subsite with adequate statistical power, and test linearity of the molecular relations along the bowel subsites while adjusting for patient and clinical characteristics. In addition, cohort participants who developed cancer resided throughout the USA, and thus were more representative colorectal cancer cases in the general US population than highly selected patients in one to a few academic hospitals. These facts increase the generalisability of our study findings.

One limitation of our study is that a vast majority (94%) of our cohort participants were non-Hispanic Caucasians. It thus remains to be seen whether our findings can be applicable to other racial or ethnic groups. As another limitation, rectal cancer is commonly treated by preoperative radiation, which could cause bias or artifacts. Therefore, we collected pretreatment biopsy materials to overcome this issue. In addition, as a secondary analysis, we excluded rectal and rectosigmoid cancers, and we obtained similar findings of statistically linear increases in the frequencies of CIMP-high, MSI-high and *BRAF* mutation from the sigmoid colon to ascending colon.

In summary, our data suggest that the frequencies of CIMP-high, MSI-high and *BRAF* mutation in colorectal cancer do not change abruptly at splenic flexure. Instead, the frequencies of CIMP-high, MSI-high and *BRAF* mutation appear to increase gradually (statistically linearly) along the bowel from the rectum to ascending colon. In addition, caecal cancers represent a unique subtype characterised by a high frequency of *KRAS* mutations, and caecal cancers do not follow the linearity trend in terms of CIMP, MSI and *BRAF* mutation. Our novel data indicate that future studies on colorectal cancers or neoplasias should include information on detailed bowel subsites (beyond the proximal colon, distal colon and rectum), which will further improve our understanding of the mechanisms of colorectal carcinogenesis.

#### Author affiliations

<sup>1</sup>Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA

<sup>2</sup>Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA

<sup>3</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA

<sup>4</sup>Cancer Program, Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA

<sup>5</sup>Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>6</sup>Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

<sup>7</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA

<sup>8</sup>Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA

<sup>9</sup>Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

**Acknowledgements** The authors would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Nebraska, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Virginia, Washington, Wyoming.

**Contributors** SO conceived the study. All authors contributed to data acquisition, analyses and interpretations. MY, TM, AK and SO drafted the manuscript. All authors critically revised the manuscript. All authors approved the final version. ATC, CSF and SO provided funding support. MY, TM and AK contributed equally to this work.

**Funding** This work was supported by US National Institutes of Health grants (P01CA87969 (to SE Hankinson), P01CA55075 (to WC Willett), P50CA127003 (to CSF), R01CA151993 (to SO) and R01CA137178 (to ATC)); the Bennett Family Fund for Targeted Therapies Research; and the Entertainment Industry Foundation through National Colorectal Cancer Research Alliance. TM was supported by a fellowship grant from the Japan Society for the Promotion of Science. The content is solely the responsibility of the authors and does not necessarily represent the official views of NCI or NIH. Funding agencies did not have any role in the design of the study, the collection, analysis, or interpretation of the data, the decision to submit the manuscript for publication or the writing of the manuscript.

**Competing interests** None.

**Ethics approval** This study was approved by the Harvard School of Public Health and Brigham and Women's Hospital institutional review boards.

**Patient consent** Obtained.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

- Buflin JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med* 1990;**113**:779–88.
- Iacopetta B. Are there two sides to colorectal cancer? *Int J Cancer* 2002;**101**:403–8.
- Gervaz P, Bucher P, Morel P. Two colons — two cancers: paradigm shift and clinical implications. *J Surg Oncol* 2004;**88**:261–6.
- Ogino S, Chan AT, Fuchs CS, et al. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut* 2011;**60**:397–411.
- Pritchard CC, Grady WM. Colorectal cancer molecular biology moves into clinical practice. *Gut* 2011;**60**:116–29.
- Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology* 2010;**138**:2088–100.
- Toyota M, Ahuja N, Ohe-Toyota M, et al. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* 1999;**96**:8681–6.
- Ogino S, Cantor M, Kawasaki T, et al. CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut* 2006;**55**:1000–6.
- Weisenberger DJ, Siegmund KD, Campan M, 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.
- Yagi K, Akagi K, Hayashi H, et al. Three DNA methylation epigenotypes in human colorectal cancer. *Clin Cancer Res* 2010;**16**:21–33.
- Hughes LA, Simons CC, van den Brandt PA, et al. Body size, physical activity and risk of colorectal cancer with or without the CpG island methylator phenotype (CIMP). *PLoS One* 2011;**6**:e18571.
- Zlobec I, Bihl M, Foerster A, et al. Comprehensive analysis of CpG island methylator phenotype (CIMP)-high, -low, and -negative colorectal cancers based on protein marker expression and molecular features. *J Pathol* 2011;**225**:336–43.
- Dahlin AM, Palmqvist R, Henriksson ML, et al. The role of the CpG island methylator phenotype in colorectal cancer prognosis depends on microsatellite instability screening status. *Clin Cancer Res* 2010;**16**:1845–55.
- Tanaka N, Huttenhower C, Noshu K, et al. Novel application of structural equation modeling to correlation structure analysis of CpG island methylation in colorectal cancer. *Am J Pathol* 2010;**177**:2731–40.
- Hinoue T, Weisenberger DJ, Lange CP, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res* 2012;**22**:271–82.
- Teodoridis JM, Hardie C, Brown R. CpG island methylator phenotype (CIMP) in cancer: causes and implications. *Cancer Lett* 2008;**268**:177–86.
- Curtin K, Slattery ML, Samowitz WS. CpG island methylation in colorectal cancer: past, present and future. *Pathol Res Int* 2011;**2011**:902674.
- Noshu K, Irahara N, Shima K, et al. Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample. *PLoS One* 2008;**3**:e3698.
- Samowitz WS, Albertsen H, Herrick J, et al. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* 2005;**129**:837–45.
- Barault L, Charon-Barra C, Jooste V, et al. Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res* 2008;**68**:8541–6.
- Wong JJ, Hawkins NJ, Ward RL, et al. Methylation of the 3p22 region encompassing MLH1 is representative of the CpG island methylator phenotype in colorectal cancer. *Mod Pathol* 2011;**24**:396–411.
- van Engeland M, Derks S, Smits KM, et al. Colorectal cancer epigenetics: complex simplicity. *J Clin Oncol* 2011;**29**:1382–91.
- Ang PW, Loh M, Liem N, et al. Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer* 2010;**10**:227.
- Slattery ML, Curtin K, Wolff RK, et al. A comparison of colon and rectal somatic DNA alterations. *Dis Colon Rectum* 2009;**52**:1304–11.
- Minoo P, Zlobec I, Peterson M, et al. Characterization of rectal, proximal and distal colon cancers based on clinicopathological, molecular and protein profiles. *Int J Oncol* 2010;**37**:707–18.
- Neish AS. Microbes in gastrointestinal health and disease. *Gastroenterology* 2009;**136**:65–80.
- Chan AT, Ogino S, Fuchs CS. Aspirin and the risk of colorectal cancer in relation to the expression of COX-2. *N Engl J Med* 2007;**356**:2131–42.
- Morikawa T, Kuchiba A, Yamauchi M, et al. Association of CTNNB1 (beta-catenin) alterations, body mass index, and physical activity with survival in patients with colorectal cancer. *JAMA* 2011;**305**:1685–94.
- Ogino S, Noshu K, Kirkner GJ, et al. A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer. *J Natl Cancer Inst* 2008;**100**:1734–8.
- Ogino S, Noshu K, Kirkner GJ, et al. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 2009;**58**:90–6.
- Ogino S, Kawasaki T, Kirkner GJ, et al. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. *J Mol Diagn* 2006;**8**:582–8.
- Ogino S, Kawasaki T, Brahmandam M, et al. Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. *J Mol Diagn* 2005;**7**:413–21.
- Noshu K, Kawasaki T, Ohnishi M, et al. PIK3CA mutation in colorectal cancer: relationship with genetic and epigenetic alterations. *Neoplasia* 2008;**10**:534–41.
- Eads CA, Danenberg KD, Kawakami K, et al. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* 2000;**28**:E32.
- Ogino S, Kawasaki T, Brahmandam M, et al. Precision and performance characteristics of bisulfite conversion and real-time PCR (MethylLight) for quantitative DNA methylation analysis. *J Mol Diagn* 2006;**8**:209–17.
- Ogino S, Kawasaki T, Kirkner GJ, et al. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn* 2007;**9**:305–14.
- Ogino S, Kawasaki T, Noshu K, et al. LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. *Int J Cancer* 2008;**122**:2767–73.
- Irahara N, Noshu K, Baba Y, et al. Precision of pyrosequencing assay to measure LINE-1 methylation in colon cancer, normal colonic mucosa, and peripheral blood cells. *J Mol Diagn* 2010;**12**:177–83.
- Hoshida Y, Villanueva A, Kobayashi M, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* 2008;**359**:1995–2004.
- Khashab MA, Pickhardt PJ, Kim DH, et al. Colorectal anatomy in adults at computed tomography colonography: normal distribution and the effect of age, sex, and body mass index. *Endoscopy* 2009;**41**:674–8.
- Ogino S, Goel A. Molecular classification and correlates in colorectal cancer. *J Mol Diagn* 2008;**10**:13–27.
- Ibrahim AE, Arends MJ, Silva AL, et al. Sequential DNA methylation changes are associated with DNMT3B overexpression in colorectal neoplastic progression. *Gut* 2011;**60**:499–508.
- Schernhammer ES, Giovannucci E, Kawasaki T, et al. Dietary folate, alcohol, and B vitamins in relation to LINE-1 hypomethylation in colon cancer. *Gut* 2010;**59**:794–9.
- Svrcek M, Buhard O, Colas C, et al. Methylation tolerance due to an O6-methylguanine DNA methyltransferase (MGMT) field defect in the colonic mucosa: an initiating step in the development of mismatch repair-deficient colorectal cancers. *Gut* 2010;**59**:1516–26.
- Saleh M, Trinchieri G. Innate immune mechanisms of colitis and colitis-associated colorectal cancer. *Nat Rev Immunol* 2011;**11**:9–20.
- Ogino S, Galon J, Fuchs CS, et al. Cancer immunology — analysis of host and tumor factors for personalized medicine. *Nat Rev Clin Oncol* 2011;**8**:711–19.
- Konishi K, Shen L, Jelinek J, et al. Concordant DNA methylation in synchronous colorectal carcinomas. *Cancer Prev Res (Phila)* 2009;**2**:814–22.
- Noshu K, Kure S, Irahara N, et al. A prospective cohort study shows unique epigenetic, genetic, and prognostic features of synchronous colorectal cancers. *Gastroenterology* 2009;**137**:1609–20 e1–3.
- Gonzalo V, Lozano JJ, Munoz J, et al. Aberrant gene promoter methylation associated with sporadic multiple colorectal cancer. *PLoS One* 2010;**5**:e8777.
- Norrie MW, Hawkins NJ, Todd AV, et al. The role of hMLH1 methylation in the development of synchronous sporadic colorectal carcinomas. *Dis Colon Rectum* 2002;**45**:674–80.
- Worthley DL, Whitehall VL, Buttenshaw RL, et al. DNA methylation within the normal colorectal mucosa is associated with pathway-specific predisposition to cancer. *Oncogene* 2010;**29**:1653–62.
- Berggren S, Gall C, Wollnitz N, et al. Gene and protein expression of P-glycoprotein, MRP1, MRP2, and CYP3A4 in the small and large human intestine. *Mol Pharm* 2007;**4**:252–7.
- Ho GT, Moodie FM, Satsangi J. Multidrug resistance 1 gene (P-glycoprotein 170): an important determinant in gastrointestinal disease? *Gut* 2003;**52**:759–66.
- Mochida Y, Taguchi K, Taniguchi S, et al. The role of P-glycoprotein in intestinal tumorigenesis: disruption of mdr1a suppresses polyp formation in Apc(Min/+) mice. *Carcinogenesis* 2003;**24**:1219–24.
- Potocnik U, Glavac D, Dean M. Common germline MDR1/ABCB1 functional polymorphisms and haplotypes modify susceptibility to colorectal cancers with high microsatellite instability. *Cancer Genet Cytogenet* 2008;**183**:28–34.
- Kuchelapati MH, Yang K, Fan K, et al. Loss of Rb1 in the gastrointestinal tract of Apc1638N mice promotes tumors of the cecum and proximal colon. *Proc Natl Acad Sci U S A* 2008;**105**:15493–8.
- Cui X, Shirai Y, Wakai T, et al. Aberrant expression of pRb and p16(INK4), alone or in combination, indicates poor outcome after resection in patients with colorectal carcinoma. *Hum Pathol* 2004;**35**:1189–95.



Supplementary Table 1. Occurrence of *BRAF* mutation among colorectal cancers according to MSI status

	Total N	MSS	MSI-high	P value (chi-square test)
<i>BRAF</i> mutation				<0.0001
(-)	1081 (86%)	988 (92%)	93 (48%)	
(+)	183 (15%)	81 (7.6%)	102 (52%)	

MSI, microsatellite instability; MSS, microsatellite stable.

Supplementary Table 2. Clinical, pathologic or molecular characteristics of colorectal cancer according to CIMP/MSI subtypes

Clinical, pathologic or molecular feature	Total N	CIMP-high MSI-high	CIMP-low/0 MSI-high	CIMP-high MSS	CIMP-low/0 MSS	P value
All cases	1188	141	46	65	936	
Sex						0.0001
Male (HPFS)	520 (44%)	37 (26%)	22 (48%)	28 (43%)	433 (46%)	
Female (NHS)	668 (56%)	104 (74%)	24 (52%)	37 (57%)	503 (54%)	
Mean age ± SD	68.4 ± 8.9	71.5 ± 6.5	66.2 ± 8.6	71.0 ± 8.6	67.7 ± 8.8	<0.0001
Body mass index						0.75
<30 kg/m <sup>2</sup>	963 (82%)	112 (80%)	36 (80%)	56 (86%)	759 (82%)	
≥30 kg/m <sup>2</sup>	216 (18%)	28 (20%)	9 (20%)	9 (14%)	170 (18%)	
Year of diagnosis						0.0026
Prior to 1995	390 (33%)	31 (22%)	12 (26%)	15 (23%)	332 (35%)	
1995 to 2006	798 (67%)	110 (78%)	34 (74%)	50 (77%)	604 (65%)	
No. of first-degree relatives with colorectal cancer						0.0008
0	954 (81%)	109 (78%)	32 (71%)	49 (75%)	764 (82%)	
1	173 (15%)	22 (16%)	5 (11%)	13 (20%)	133 (14%)	
≥2	54 (4.6%)	9 (6.4%)	8 (18%)	3 (4.6%)	34 (3.7%)	
Pre-diagnosis physical activity (MET-hours/week)						0.45
<18	680 (60%)	91 (66%)	28 (67%)	39 (61%)	522 (60%)	
≥18	440 (39%)	47 (34%)	14 (33%)	25 (39%)	354 (40%)	
Pre-diagnosis smoking status						0.16
Never	451 (39%)	43 (31%)	19 (44%)	29 (45%)	360 (39%)	
Former/current smokers	712 (61%)	96 (69%)	24 (56%)	36 (55%)	556 (61%)	

(Supplementary Table 2, continued)

(Supplementary Table 2, continuing from the previous page)

Clinical, pathologic or molecular feature	Total N	CIMP-high MSI-high	CIMP-low/0 MSI-high	CIMP-high MSS	CIMP-low/0 MSS	P value
Pre-diagnosis alcohol consumption						0.0032
No	389 (33%)	65 (46%)	20 (44%)	18 (28%)	286 (31%)	
<15 g/day	549 (47%)	50 (36%)	21 (47%)	35 (54%)	443 (48%)	
≥15 g/day	238 (20%)	25 (18%)	4 (8.9%)	12 (18%)	197 (21%)	
Disease stage						<0.0001
I	272 (25%)	21 (15%)	15 (37%)	13 (22%)	223 (27%)	
II	336 (31%)	85 (63%)	16 (39%)	11 (18%)	224 (27%)	
III	313 (29%)	23 (17%)	7 (17%)	18 (30%)	265 (32%)	
IV	157 (15%)	7 (5.1%)	3 (7.3%)	18 (30%)	129 (15%)	
Tumor differentiation						<0.0001
Well to moderate	1064 (90%)	87 (62%)	41 (89%)	58 (89%)	878 (95%)	
Poor	117 (9.9%)	54 (38%)	5 (11%)	7 (11%)	51 (5.5%)	
Mucinous component						<0.0001
0%	702 (61%)	44 (32%)	19 (41%)	29 (45%)	610 (68%)	
1-49%	294 (26%)	49 (35%)	18 (39%)	26 (40%)	201 (22%)	
≥50%	155 (13%)	46 (33%)	9 (20%)	10 (15%)	90 (10%)	
Signet ring cell component						<0.0001
0%	1026 (89%)	98 (71%)	39 (85%)	50 (78%)	839 (93%)	
1-49%	104 (9.1%)	34 (25%)	7 (15%)	14 (22%)	49 (5.4%)	
≥50%	19 (1.7%)	6 (4.4%)	0	0	13 (1.4%)	
<i>BRAF</i> mutation						<0.0001
(-)	1005 (85%)	44 (31%)	44 (98%)	34 (52%)	883 (95%)	
(+)	176 (15%)	97 (69%)	1 (2.2%)	31 (48%)	47 (5.1%)	

(Supplementary Table 2, continued)

(Supplementary Table 2, continuing from the previous page)

Clinical, pathologic or molecular feature	Total N	CIMP-high MSI-high	CIMP-low/0 MSI-high	CIMP-high MSS	CIMP-low/0 MSS	P value
<i>KRAS</i> mutation						<0.0001
(-)	754 (64%)	134 (95%)	27 (61%)	42 (65%)	551 (59%)	
(+)	429 (36%)	7 (5.0%)	17 (39%)	23 (35%)	382 (41%)	
<i>PIK3CA</i> mutation						0.24
(-)	893 (81%)	100 (77%)	33 (83%)	44 (75%)	716 (82%)	
(+)	204 (19%)	30 (23%)	7 (18%)	15 (25%)	152 (18%)	
LINE-1 methylation level (mean $\pm$ SD)						<0.0001
	62.7 $\pm$ 9.5	67.5 $\pm$ 8.5	66.8 $\pm$ 10.1	67.2 $\pm$ 8.5	61.4 $\pm$ 9.4	
<i>ABCBI</i> expression level (Log <sub>2</sub> intensity)						0.0011
Median	6.47	6.41	6.40	6.46	6.50	
(interquartile range)	(6.35-6.81)	(6.33-6.60)	(6.35-6.54)	(6.36-6.64)	(6.36-6.89)	

(%) indicates the proportion of cases with a specific clinical, pathologic or molecular feature among cancers located in each CIMP/MSI subtypes (CIMP-high MSI-high, CIMP-low/0 MSI-high, CIMP-high MSS, CIMP-low/0 MSS). P values were calculated by chi-square test, except for age, LINE-1 (one-way ANOVA test), and ABCB1 (Kruskal-Wallis test). CIMP, CpG island methylator phenotype; HPFS, Health Professionals Follow-up Study; MET, metabolic equivalent task; MSI, microsatellite instability; MSS, microsatellite stable; NHS, Nurses' Health Study; SD, standard deviation.