Citation: Clinical and Translational Gastroenterology (2016) 7, e200; doi:10.1038/ctg.2016.53
 2155-384X/16

www.nature.com/ctg

Fusobacterium nucleatum in Colorectal Carcinoma Tissue According to Tumor Location

Kosuke Mima, MD, PhD[1,15], Yin Cao, ScD[2,3,4,15], Andrew T. Chan, MD, MPH[2,3,5,15], Zhi Rong Qian, MD, PhD[1,15], Jonathan A. Nowak, MD, PhD[6], Yohei Masugi, MD, PhD[1], Yan Shi, MD, PhD[1], Mingyang Song, MD, ScD[2,3,4], Annacarolina da Silva, MD[1], Mancang Gu, PhD[1], Wanwan Li, MD[1], Tsuyoshi Hamada, MD, PhD[1], Keisuke Kosumi, MD, PhD[1], Akiko Hanyuda, MD, MPH[1], Li Liu, PhD[1], Aleksandar D. Kostic, PhD[7,8,9], Marios Giannakis, MD, PhD[1,8,10], Susan Bullman, PhD[1,8], Caitlin A. Brennan, PhD[11], Danny A. Milner, MD, MSc[6,11], Hideo Baba, MD, PhD[12], Levi A. Garraway, MD, PhD[1,8,10], Jeffrey A. Meyerhardt, MD, MPH[1], Wendy S. Garrett, MD, PhD[1,8,11], Curtis Huttenhower, PhD[7,8,13], Matthew Meyerson, MD, PhD[1,8], Edward L. Giovannucci, MD, ScD[4,5,14,16], Charles S. Fuchs, MD, MPH[1,5,16], Reiko Nishihara, PhD[1,4,7,14,16] and Shuji Ogino, MD, PhD, MS[1,6,14,16]

OBJECTIVES: Evidence suggests a possible role of *Fusobacterium nucleatum* in colorectal carcinogenesis, especially in right-sided proximal colorectum. Considering a change in bowel contents and microbiome from proximal to distal colorectal segments, we hypothesized that the proportion of colorectal carcinoma enriched with *F. nucleatum* might gradually increase along the bowel subsites from rectum to cecum.

METHODS: A retrospective, cross-sectional analysis was conducted on 1,102 colon and rectal carcinomas in molecular pathological epidemiology databases of the Nurses' Health Study and the Health Professionals Follow-up Study. We measured the amount of *F. nucleatum* DNA in colorectal tumor tissue using a quantitative PCR assay and equally dichotomized *F. nucleatum*-positive cases (high vs. low). We used multivariable logistic regression analysis to examine the relationship of a bowel subsite variable (rectum, rectosigmoid junction, sigmoid colon, descending colon, splenic flexure, transverse colon, hepatic flexure, ascending colon, and cecum) with the amount of *F. nucleatum*.

RESULTS: The proportion of *F. nucleatum*-high colorectal cancers gradually increased from rectal cancers (2.5%; 4/157) to cecal cancers (11%; 19/178), with a statistically significant linear trend along all subsites ($P < 0.0001$) and little evidence of non-linearity. The proportion of *F. nucleatum*-low cancers was higher in rectal, ascending colon, and cecal cancers than in cancers of middle segments.

CONCLUSIONS: The proportion of *F. nucleatum*-high colorectal cancers gradually increases from rectum to cecum. Our data support the colorectal continuum model that reflects pathogenic influences of the gut microbiota on neoplastic and immune cells and challenges the prevailing two-colon (proximal vs. distal) dichotomy paradigm.

Clinical and Translational Gastroenterology (2016) **7,** e200; doi:10.1038/ctg.2016.53; published online 3 November 2016

Subject Category: Colon/Small Bowel

INTRODUCTION

An increasing body of evidence suggests possible roles of microorganisms in colorectal carcinogenesis.[1–6] Among various microbial species, *Fusobacterium nucleatum* appears to inhibit antitumor immune response and potentiate colonic neoplasia development in animal models.[7–10] In addition, an enrichment of *F. nucleatum* can be observed in a subset of human colorectal neoplasms, and a high amount of *F. nucleatum* in carcinoma tissue has been associated with proximal tumor location, high-level microsatellite instability (MSI-high), and lower density of T cells in tumor tissue.[11–18]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA; [2]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; [3]Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, USA; [4]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; [5]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [6]Division of MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [7]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; [8]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA; [9]Center for Computational and Integrative Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; [10]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [11]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; [12]Department of Gastroenterological Surgery, Graduate School of Medical Science, Kumamoto University, Kumamoto, Japan; [13]Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA and [14]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

Correspondence: Reiko Nishihara, PhD, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Room M422, Boston, Massachusetts 02215, USA or Shuji Ogino, MD, PhD, MS, Division of MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA.

E-mail: rnishiha@hsph.harvard.edu or shuji_ogino@dfci.harvard.edu

[15]These authors contributed equally as co-first authors.

[16]These authors contributed equally as co-last authors.

Received 27 March 2016; accepted 13 September 2016

As a long organ, the colorectum is typically divided into proximal colon (cecum to transverse colon), distal colon (splenic flexure to sigmoid colon), and rectum in clinical, pathological, and epidemiological studies.[19–21] However, multiple studies have demonstrated that proportions of colorectal cancers with specific molecular features such as MSI-high, CpG island methylator phenotype (CIMP)-high, and *BRAF* and *PIK3CA* mutations gradually increase along the bowel subsites from rectum to ascending colon.[22–25] These findings are consistent with the fact that microbiota, bacterial metabolites, and other contents of the large intestine continually (rather than abruptly) change from the proximal to distal segments[26–29] and support the colorectal continuum model rather than the dichotomy or trichotomy model.[19–21] Hence, we hypothesized that the proportion of colorectal cancer enriched with *F. nucleatum* might gradually change along the bowel subsites from cecum to rectum.

To test this hypothesis, we utilized a database of colorectal carcinoma cases in two US nationwide prospective cohort studies, the Nurses' Health Study and the Health Professionals Follow-up Study and examined the amount of *F. nucleatum* in colorectal cancer tissue according to the bowel subsites.

METHODS

Study population. We utilized the database of colorectal carcinoma cases in two US nationwide prospective cohort studies, the Nurses' Health Study (121,701 women enrolled in 1976) and the Health Professionals Follow-up Study (51,529 men enrolled in 1986)[30,31] and conducted a retrospective, cross-sectional analysis to assess the association of the amount of *F. nucleatum* in colorectal cancer tissue with tumor location. Every 2 years, we sent participants follow-up questionnaires to collect information on lifestyle factors and asked whether they had received diagnoses of major disease, including cancers. Study physicians reviewed medical records for incident colorectal cancer cases and recorded cancer stage (Tumor, Node, Metastasis) and tumor location (cecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, rectosigmoid junction, and rectum).[22]

We collected formalin-fixed paraffin-embedded (FFPE) tissue blocks from hospitals where participants with colorectal carcinoma had undergone tumor resection. A single pathologist (S.O.), who was unaware of other data, conducted a centralized review of hematoxylin and eosin–stained tissue sections of all colorectal carcinoma cases and recorded pathological features. Tumor differentiation was classified into well to moderate or poor ($>50\%$ vs. $\leq 50\%$ glandular area). Written informed consent was obtained from all study participants. The institutional review boards at the Harvard T.H. Chan School of Public Health and the Brigham and Women's Hospital (Boston, MA) approved the cohort studies.

Quantitative PCR for *F. nucleatum*. We dissected colorectal cancer tissues from whole-tissue sections of FFPE tissue blocks, and DNA was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen, Valencia, CA). After the quantitative PCR assay for *F. nucleatum* was developed and validated as previously described,[16] we measured the amount of tissue *F. nucleatum* DNA in 1,102 colorectal carcinoma cases, while blinded to data on tumor location and other clinical, pathological, and tumor molecular features. Custom TaqMan primer/probe sets (Applied Biosystems, San Diego, CA) for the *nusG* gene of *F. nucleatum* and for the reference human gene *SLCO2A1* were used as previously described.[16] Each reaction contained 80 ng of genomic DNA and was assayed in 20 µl reactions containing $1\times$ final concentration TaqMan Environmental Master Mix 2.0 (Applied Biosystems) and each TaqMan Gene Expression Assay (Applied Biosystems) in a 96-well optical PCR plate. Amplification and detection of DNA was performed with the StepOnePlus Real-Time PCR Systems (Applied Biosystems) using the following reaction conditions: 10 min at 95 °C and 45 cycles of 15 s at 95 °C and 1 min at 60 °C.

Our validation study has previously shown that, in colorectal carcinoma cases with detectable *F. nucleatum* DNA, the cycle threshold (Ct) values in the quantitative PCR for *F. nucleatum* and *SLCO2A1* decreased linearly with the log-transformed amount of input DNA from the same specimen ($r^2 > 0.99$), and that the interassay coefficient of variation of Ct values from the same specimen in five different batches was $\leq 1\%$ for all targets.[16] Each specimen was analyzed in duplicate for each target in a single batch, and we used the mean of the two Ct values for each target. Spearman's rank-correlation coefficients between the two Ct values (in duplicated runs) in each of cases with detectable target amplification in the quantitative PCR assays for *F. nucleatum* and *SLCO2A1* were 0.95 and 0.92, respectively.[16] The amount of tissue *F. nucleatum* DNA in each specimen was calculated as a relative unitless value normalized with *SLCO2A1* using the $2^{-\Delta Ct}$ method (where $\Delta Ct =$ "the mean Ct value of *F. nucleatum*" – "the mean Ct value of *SLCO2A1*").[16]

Cases with detectable *F. nucleatum* DNA were categorized as low or high based on the median cutpoint while cases without detectable *F. nucleatum* DNA were categorized as negative, to keep consistent classification system with our previous study.[32]

Analyses of MSI, DNA methylation, and *KRAS*, *BRAF*, and *PIK3CA* mutations. Using DNA extracted from FFPE colorectal carcinoma tissue, MSI status was analyzed with the use of 10 microsatellite markers (D2S123, D5S346, D17S250, BAT25, BAT26, BAT40, D18S55, D18S56, D18S67, and D18S487) as previously described.[33] We defined MSI-high as the presence of instability in $\geq 30\%$ of the markers, and MSI-low/microsatellite stable (MSS) as instability in $<30\%$ of the markers. Methylation analyses of long interspersed nucleotide element-1 (LINE-1)[34] and eight promoter CpG islands specific for CIMP (*CACNA1G*, *CDKN2A*, *CRABP1*, *IGF2*, *MLH1*, *NEUROG1*, *RUNX3*, and *SOCS1*) were performed as previously described.[35,36] PCR reaction and pyrosequencing were performed for *KRAS* (codons 12, 13, 61, and 146),[37,38] *BRAF* (codon 600),[33] and *PIK3CA* (exons 9 and 20).[39,40]

Statistical analysis. All statistical analyses were conducted using SAS (version 9.3, SAS Institute, Cary, NC) and all

P values were two-sided. For our primary hypothesis testing, we examined the relationship of the tumor location variable (the nine subsites) with the amount of *F. nucleatum* DNA in colorectal cancer tissue (as an outcome variable). There was an initial flexibility of the outcome variable, which could be raw continuous, log-transformed (continuous), ordinal three-tiered (high vs. low vs. negative), binary (high vs. low/negative), or binary (high/low vs. negative). Considering these five possibilities, we used adjusted two-sided α level of 0.01 ($=0.05/5$). Neither the amount of *F. nucleatum* DNA nor its log-transformed value fitted a normal distribution with the use of the Kolmogorov–Smirnov test for normality ($P<0.01$). Thus we used multivariable logistic regression analysis to examine the relationship of the nine subsites (as a predictor variable; tested for a linear trend with one degree of freedom) with categorical outcome variables of *F. nucleatum*. For the ordinal outcome variable, the proportionality of odds assumption was not satisfied in ordinal logistic regression models ($P=0.019$). According to the distribution of colorectal cancer cases by the subsites and the amount of *F. nucleatum* (Figure 1), we used a binary outcome variable (*F. nucleatum*-high vs. -low/negative) in the logistic regression model. For the subsite variable, we assigned population average distance from anal verge to each bowel subsite (either the midpoint or junction/flexure), which was calculated based on published data using computed tomographic colonography,[22,41] as follows: rectum (the midpoint), 9.8 cm; rectosigmoid junction, 20 cm; sigmoid colon (the midpoint), 44 cm; descending colon (the midpoint), 85 cm; splenic flexure, 102 cm; transverse colon (the midpoint), 131 cm; hepatic flexure, 160 cm; ascending colon (the midpoint), 171 cm; and cecum (the midpoint), 186 cm. A significant *P* value by the Wald's test on the bowel subsite variable indicated a linear relationship of the bowel subsite with tissue *F. nucleatum*, but a curvilinear relationship might exist. Thus we assessed the non-linearity by a likelihood ratio test comparing the model with squared and/or cubic subsite variables with the model without squared or cubic subsite variable; a significant likelihood ratio test result would indicate the presence of non-linearity (curvilinearity).

The multivariable logistic regression models were adjusted for clinical features, including age (continuous), sex, year of diagnosis (continuous), and family history of colorectal carcinoma in any first-degree relative (present vs. absent). Studies have shown an enrichment of *F. nucleatum* in colorectal adenomas (before progression to carcinomas),[9,13,15] suggesting that *F. nucleatum* may be involved in early colorectal carcinogenesis. Hence, pathological and tumor molecular features of colorectal carcinoma may be present downstream in the causal sequence of events after the *Fusobacterium* variable, and adjusting for the tumor pathological and molecular features might cause biased results. Thus we did not include the pathological and tumor molecular variables in the multivariable logistic regression models. For cases with missing information on family history of colorectal carcinoma in a first-degree relative (1.3%), we included those cases in a majority category of a given covariate to minimize the number of variables in multivariable logistic regression models. We confirmed that excluding the cases with missing information on family history of colorectal carcinoma in a first-degree relative did not substantially alter the results (data not shown).

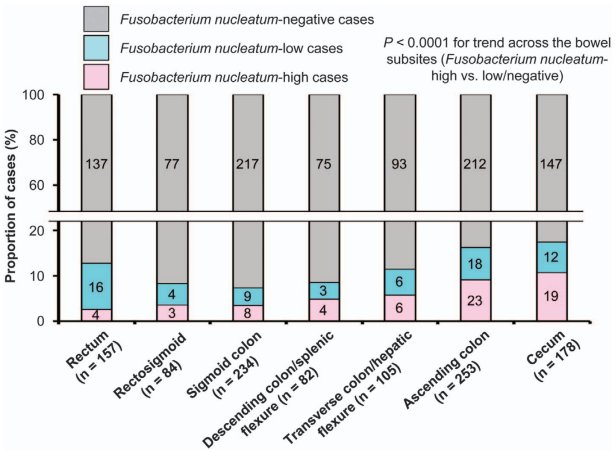

Figure 1 Proportions of *Fusobacterium nucleatum*-negative, *F. nucleatum*-low, and *F. nucleatum*-high colorectal carcinoma cases along the bowel subsites. *P*-value was calculated by the linear trend test across the bowel subsite variable (population average distance from anal verge to each subsite (cm)) as a continuous variable in the univariable logistic regression model to predict the amount of tissue *F. nucleatum* (as a binary outcome variable (high vs. low/negative)).

To assess the associations between the amount of tissue *F. nucleatum* and other categorical variables, chi-square test was performed. To compare mean age and mean LINE-1 methylation levels, an analysis of variance was performed. These comparisons represented secondary analyses, and we used adjusted α level of 0.003 ($=0.05/14$) by simple Bonferroni correction for multiple hypothesis testing.

RESULTS

***F. nucleatum* in colorectal carcinoma tissue.** We measured the amount of tissue *F. nucleatum* DNA in 1,102 colorectal carcinoma cases within the two prospective cohort studies using the quantitative PCR assay that was previously validated.[16] *F. nucleatum* DNA was detected (positive) in colorectal carcinoma tissue in 138 (13%) of the 1,102 cases and undetectable (negative) in the remaining 964 cases (87%). We equally dichotomized the 138 cases with detectable *F. nucleatum* DNA levels into two groups to keep consistency with our previous study.[32] Clinical, pathological, and tumor molecular features according to the amount (high vs. low. vs. negative) of tissue *F. nucleatum* are summarized in Table 1. High-level *F. nucleatum* in colorectal cancer tissue was associated with proximal tumor location, poor tumor differentiation, MSI-high, *MLH1* hypermethylation, CIMP-high, and *BRAF* mutation ($P\leq0.0004$ with the adjusted α level of 0.003 for multiple hypothesis testing).

The relationship of the bowel subsites with the amount of *F. nucleatum* in colorectal cancer tissue. The amount of *F. nucleatum* in colorectal cancer tissue according to the bowel subsites from rectum to cecum is shown in Figure 1. The proportion of *F. nucleatum*-high cancers gradually

Table 1 Characteristics of colorectal cancer cases according to the amount of tissue *Fusobacterium nucleatum*

Characteristics[a]	All patients ($n=1,102$)	The amount of tissue *Fusobacterium nucleatum*			*P* value[b]
		Negative ($n=964$)	Low ($n=69$)	High ($n=69$)	
Mean age ± s.d. (years)	69.5 ± 8.9	69.4 ± 8.9	70.9 ± 9.0	69.2 ± 8.5	0.39
Sex					0.35
Men	466 (42%)	415 (43%)	27 (39%)	24 (35%)	
Women	636 (58%)	549 (57%)	42 (61%)	45 (65%)	
Year of diagnosis					0.026
Prior to 1995	352 (32%)	323 (34%)	12 (17%)	17 (25%)	
1996–2000	301 (27%)	262 (27%)	19 (28%)	20 (29%)	
2001–2008	449 (41%)	379 (39%)	38 (55%)	32 (46%)	
Family history of colorectal carcinoma in a first-degree relative					0.26
Absent	877 (81%)	762 (80%)	60 (88%)	55 (81%)	
Present	211 (19%)	190 (20%)	8 (12%)	13 (19%)	
Tumor location					0.0004
Proximal colon	536 (49%)	452 (47%)	36 (53%)	48 (72%)	
Distal colon	316 (29%)	292 (31%)	12 (18%)	12 (18%)	
Rectum	241 (22%)	214 (22%)	20 (29%)	7 (10%)	
Nine bowel subsites[c]					<0.0001[d]
Cecum	178	147 (83%)	12 (6.7%)	19 (11%)	
Ascending colon	253	212 (84%)	18 (7.1%)	23 (9.1%)	
Hepatic flexure	32	28 (88%)	2 (6.3%)	2 (6.3%)	
Transverse colon	73	65 (89%)	4 (5.5%)	4 (5.5%)	
Splenic flexure	29	27 (93%)	1 (3.5%)	1 (3.5%)	
Descending colon	53	48 (91%)	2 (3.8%)	3 (5.7%)	
Sigmoid colon	234	217 (93%)	9 (3.8%)	8 (3.4%)	
Rectosigmoid junction	84	77 (92%)	4 (4.8%)	3 (3.6%)	
Rectum	157	137 (87%)	16 (10%)	4 (2.5%)	
Disease stage					0.006
I	247 (25%)	230 (26%)	10 (16%)	7 (11%)	
II	331 (33%)	279 (32%)	23 (37%)	29 (45%)	
III	286 (29%)	246 (28%)	25 (40%)	15 (24%)	
IV	135 (13%)	117 (14%)	5 (7.9%)	13 (20%)	
Tumor differentiation					<0.0001
Well to moderate	994 (90%)	887 (92%)	57 (84%)	50 (72%)	
Poor	106 (9.6%)	76 (7.9%)	11 (16%)	19 (28%)	
MSI status					<0.0001
MSI-low/MSS	885 (84%)	805 (87%)	44 (67%)	36 (54%)	
MSI-high	171 (16%)	118 (13%)	22 (33%)	31 (46%)	
MLH1 hypermethylation					<0.0001
Absent	869 (86%)	782 (89%)	50 (79%)	37 (58%)	
Present	140 (14%)	100 (11%)	13 (21%)	27 (42%)	
CIMP status					<0.0001
Low/negative	823 (82%)	737 (84%)	50 (79%)	36 (56%)	
High	186 (18%)	145 (16%)	13 (21%)	28 (44%)	
BRAF mutation					<0.0001
Wild type	892 (84%)	795 (85%)	52 (79%)	45 (66%)	
Mutant	172 (16%)	135 (15%)	14 (21%)	23 (34%)	
KRAS mutation					0.51
Wild type	569 (57%)	501 (57%)	30 (51%)	38 (61%)	
Mutant	435 (43%)	382 (43%)	29 (49%)	24 (39%)	
PIK3CA mutation					0.88
Wild type	841 (84%)	738 (84%)	49 (82%)	54 (83%)	
Mutant	162 (16%)	140 (16%)	11 (18%)	11 (17%)	
Mean LINE-1 methylation level, % ± s.d.	63.5 ± 10.2	63.3 ± 10.2	65.0 ± 10.6	65.4 ± 8.9	0.14

CIMP, CpG island methylator phenotype; LINE-1, long interspersed nucleotide element-1; MSI, microsatellite instability; MSS, microsatellite stable.
[a]Percentage (%) indicates the proportion of cases with a specific clinical, pathological, or tumor molecular feature according to the amount of tissue *Fusobacterium nucleatum*. There were cases which had missing values for any of the characteristics except for age, sex, and year of diagnosis.
[b]To assess associations between the ordinal categories (negative, low, and high) of the amount of tissue *F. nucleatum* and categorical variables, the chi-square test was performed. To compare mean age and mean LINE-1 methylation levels, an analysis of variance was performed. We adjusted two-sided α level to 0.003 (= 0.05/14) by simple Bonferroni correction for multiple hypothesis testing.
[c]Percentage indicates the proportion of *F. nucleatum*-negative, *F. nucleatum*-low, or *F. nucleatum*-high cases among all tumors in a given bowel subsite.
[d]*P* value was calculated by the linear trend test across the bowel subsite variable (population average distance from anal verge to each subsite (cm)) as a continuous variable in the univariable logistic regression model to predict the amount of tissue *F. nucleatum* (as a binary outcome variable (high vs. low/negative)).

Table 2 Assessment of the linearity and non-linearity on the relationship of the bowel subsites with the amount of *Fusobacterium nucleatum* in colorectal cancer tissue by multivariable logistic regression analyses

Bowel subsite variable (distance from anal verge to each subsite (cm))	Squared subsite variable		Cubic subsite variable		Likelihood ratio test	
***P* value (Wald's test)[a]**	**Included**	***P* value (Wald's test)**	**Included**	***P* value (Wald's test)**	**Degree of freedom**	***P* value[b]**
Model for the amount of tissue *Fusobacterium nucleatum* (as an outcome variable (high vs. low/negative))						
All cases						
<0.0001	No	—	No	—	—	Referent
0.87	Yes	0.61	No	—	1	0.61
0.65	Yes	0.74	Yes	0.67	2	0.80
Cases from sigmoid colon to cecum (excluding rectal and rectosigmoid cancers)						
0.002	No	—	No	—	—	Referent
0.82	Yes	0.51	No	—	1	0.50
0.84	Yes	0.85	Yes	0.80	2	0.77

Multivariable logistic regression model included age, sex, year of diagnosis, family history of colorectal cancer in parent or sibling, and the bowel subsite variable with or without the squared and cubic subsite variable, as indicated in the Table. We adjusted two-sided α level to 0.01 ($=0.05/5$) for multiple hypothesis testing.
[a]*P* value was calculated by the Wald's test on the bowel subsite variable (population average distance from anal verge to each subsite (cm)) as a continuous variable in the multivariable logistic regression model to predict the amount of tissue *Fusobacterium nucleatum* (as a binary outcome variable (high vs. low/negative)).
[b]A significant *P* value by the likelihood ratio test indicates a non-linear (curvilinear) relationship, and a combination of insignificant *P* values by the likelihood ratio test and a significant *P* value by the Wald test on the bowel subsite variable in the model without the squared or cubic subsite variable indicates a linear relationship.

increased from rectal cancers (2.5% = 4/157) to cecal cancers (11% = 19/178). In contrast, the proportion of *F. nucleatum*-low cancers was higher in rectal cancers (10% = 16/157), ascending colon cancers (7.1% = 18/253), and cecal cancers (6.7% = 12/178) than in cancers of middle segments.

We assessed the relationship of the bowel subsite (as a predictor variable) with the amount of *F. nucleatum* in colorectal cancer tissue (as a binary outcome variable (high vs. low/negative)) by multivariable logistic regression analysis that adjusted for potential confounders (Table 2). The bowel subsite variable was significantly associated with high-level *F. nucleatum* in colorectal cancer tissue ($P_{trend}<0.0001$ for trend across the bowel subsites from rectum to cecum, with the adjusted α level of 0.01). We demonstrated little evidence of non-linearity ($P\geq 0.61$) using likelihood ratio test, which compared the model with squared and/or cubic subsite variables to the model without squared or cubic subsite variable.

To exclude a potential influence of preoperative chemotherapy and/or radiation therapy for rectal cancers as a secondary analysis, we excluded cancers in the rectum and rectosigmoid and performed a linearity test. The bowel subsite variable (from sigmoid colon to cecum) was significantly associated with high-level tissue *F. nucleatum* ($P_{trend}=0.002$ for trend across the bowel subsites, with the adjusted α level of 0.01), and there was no evidence for non-linearity ($P\geq 0.50$; Table 2).

DISCUSSION

We found that the proportion of *F. nucleatum*-high colorectal cancers increased linearly along the large intestine from rectum to cecum. Although differences in clinical, pathological, and epidemiological features between proximal and distal colon cancers and between colon and rectal cancers have been known for decades,[19–21] emerging evidence indicates a gradual change in the proportions of key molecular features of colorectal cancer along the bowel subsites.[22–25] Our data challenge the common dichotomy model and support the colorectal continuum model that recently emerged.[42,43]

Analyses of molecular pathology are increasing importance in cancer research.[44–51] Accumulating evidence suggests that colorectal tumors arise with sets of genomic and epigenomic alterations through interactions between neoplastic cells, immune cells, and microbiota that vary along the proximal to distal axis of colorectum.[42] In the current study, high-level tissue *F. nucleatum* DNA was associated with the molecular features of colorectal cancer, including MSI-high, CIMP-high, and *BRAF* mutation. Some studies reported that butyrate might suppress colonic inflammation and tumor development.[52,53] In contrast, emerging evidence suggests that butyrate may promote the growth of colonic tumors that exhibit DNA mismatch repair deficiency.[5,54] *F. nucleatum* is known to be one of the microbes that produce butyrate.[1,5] Thus *F. nucleatum* might promote the development and progression of colorectal tumors through the production of butyrate, although additional studies are needed to elucidate the exact mechanisms underlying the association between *F. nucleatum* and colorectal carcinogenesis.

Cecal carcinomas represent an interesting subgroup of colorectal carcinomas characterized by high prevalence of *KRAS* mutations.[22,55] In addition, along with rectum, cecum shows the highest incidence of carcinoma occurrence per surface area of mucosa.[56] Our current study has shown that *F. nucleatum*-enriched carcinomas are most prevalent in cecum compared with other subsites. Future studies should investigate the role of microbiota in cecal carcinogenesis.

Epidemiological evidence indicates a recent increase in the proportion of proximal colon cancers in Western countries[57,58] and the association between postcolonoscopy cancer and proximal tumor location.[59,60] Because any experimental system cannot perfectly recapitulate the complex nature of human tumor or microorganisms, analyses of human cancer tissue in a large population are useful in elucidating the relationship between microorganisms and cancer. However, no previous study has examined the amount of specific microbial species in human colorectal cancers according to

detailed subsites (using an enough sample size), as we did in this study. Our population-based human data would guide future mechanistic investigations. Considering that diet, lifestyle, pharmacological factors (including antibiotics), and probiotics and prebiotics can influence the composition of intestinal microbiota,[61–63] future investigations may be warranted to examine potential influences of those modifiable factors on the intestinal microflora and tumorigenic processes.

Strengths of this study include the use of our molecular pathological epidemiology[64,65] database (of 1,102 colorectal carcinoma cases in the two US nationwide, prospective cohort studies), which integrates epidemiological exposures, clinical characteristics, and tissue *F. nucleatum* in colorectal carcinoma. The sample size and the comprehensiveness of the colorectal cancer database enabled us to examine the amounts of *F. nucleatum* in colorectal cancer tissue in each of the bowel subsites and test the linearity of the relationship of the bowel subsites with the amount of *F. nucleatum*, while adjusting for clinical features. Importantly, our data set of colorectal cancer cases represented a population-based sample derived from a large number of hospitals in diverse settings across the United States that increases the generalizability of our findings.

We recognize limitations of our study. First, routine histopathology processing might have influenced the performance of the quantitative PCR assay to detect microorganisms in FFPE tissue specimens. Although measurement errors in FFPE tissue specimens would have likely driven our results toward the null hypothesis, we cannot exclude unmeasured confounding factors. However, our validation study has demonstrated a high linearity ($r^2 > 0.99$) and a high reproducibility (interassay coefficient of variation $\leq 1\%$) of the quantitative PCR assay for *F. nucleatum* with the use of FFPE tissue specimens.[16] In addition, our data on the relationships of *F. nucleatum* with clinicopathological characteristics and tumor molecular features, including MSI and CIMP status, are consistent with the study using a quantitative PCR assay for frozen tissue specimens.[14] Second, rectal cancers are commonly treated by preoperative chemotherapy and/or radiation, which might have changed the gut and tumor microbiota. Therefore, we excluded preoperatively treated rectal cancers in which adequate pretreatment biopsy specimens were unavailable. In addition, as a secondary analysis, we excluded rectal and rectosigmoid cancers and observed similar findings on the association of bowel subsites (from the sigmoid colon to cecum) with the amount of tissue *F. nucleatum*. In the current study, rectal cancers showed a high proportion of *F. nucleatum*-low cases (10%) compared with rectosigmoid (4.8%) and sigmoid colon cancers (3.8%). These findings need to be validated by additional studies. Third, we did not examine other microbes (including *Escherichia coli* and *Bacteroides fragilis*[66,67]) in colorectal cancer tissue or data on stool microbiota. Certainly, future comprehensive metagenomic analyses on tissue and stool microbiota may provide further insights on roles of gut microorganisms in the development and progression of colorectal tumors. Fourth, as an observational study, we could not conclude on a potential causal effect of *F. nucleatum*. Nonetheless, given complex roles of interactions between microbial and host factors in human carcinogenesis, we believe that our novel data on *F. nucleatum* (which appears to have a role in carcinogenesis in experimental studies[8–10]) in $>1{,}000$ colorectal cancer cases along the bowel subsites represent valuable information.

In conclusion, utilizing the database of the 1,102 colorectal carcinoma cases in the US nationwide prospective cohort studies, we have found that the proportion of colorectal cancer enriched with *F. nucleatum* increases linearly along the bowel subsites from rectum to cecum. Our human population-based data suggest a continuum model of pathogenic influences of *F. nucleatum* on colorectal carcinogenesis, which may be targeted for colorectal cancer prevention and treatment in the future.

CONFLICT OF INTEREST

Guarantor of the article: Shuji Ogino, MD, PhD, MS.
Specific author contributions: All authors contributed to review and revision. Kosuke Mima, Caitlin A. Brennan, Danny A. Milner, Levi A. Garraway, Jeffrey A. Meyerhardt, Wendy S. Garrett, Curtis Huttenhower, Matthew Meyerson, Edward L. Giovannucci, Andrew T. Chan, Charles S. Fuchs, and Shuji Ogino developed the main concept and designed the study. Andrew T. Chan, Charles S. Fuchs, and Shuji Ogino wrote grant applications. Kosuke Mima, Yin Cao, Reiko Nishihara, Zhi Rong Qian, Jonathan A. Nowak, Yohei Masugi, Yan Shi, Annacarolina da Silva, Mancang Gu, Wanwan Li, Tsuyoshi Hamada, Keisuke Kosumi, Akiko Hanyuda, Li Liu, Mingyang Song, Jeffrey A. Meyerhardt, Edward L. Giovannucci, Andrew T. Chan, Charles S. Fuchs, and Shuji Ogino were responsible for collection of tumor tissue and acquisition of epidemiological, clinical and tumor tissue data, including histopathological and immunohistochemical characteristics. Kosuke Mima, Aleksandar D. Kostic, Susan Bullman, Caitlin A. Brennan, Wendy S. Garrett, Curtis Huttenhower, Matthew Meyerson, Charles S. Fuchs, and Shuji Ogino performed data analysis and interpretation. Kosuke Mima, Yin Cao, Reiko Nishihara, and Shuji Ogino drafted the manuscript. Yin Cao, Andrew T. Chan, Mingyang Song, Marios Giannakis, Caitlin A. Brennan, Hideo Baba, Wendy S. Garrett, Matthew Meyerson, Jeffrey A. Meyerhardt, Edward L. Giovannucci, Charles S. Fuchs, Reiko Nishihara, and Shuji Ogino contributed to editing and critical revision for important intellectual contents. All authors approved the final draft submitted.
Financial support: This work was supported by US National Institutes of Health (NIH) grants (P01 CA87969 to M.J. Stampfer; UM1 CA186107 to M.J. Stampfer; P01 CA55075 to W.C. Willett; UM1 CA167552 to W.C. Willett; P50 CA127003 to C.S.F.; R01 CA137178 to A.T.C.; R01 CA151993 to S.O.; R35 CA197735 to S.O.; and K07 CA190673 to R.N.); and by grants from the Project P Fund, The Friends of the Dana-Farber Cancer Institute, Bennett Family Fund, and the Entertainment Industry Foundation through National Colorectal Cancer Research Alliance. K.M. is supported by a grant from Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers from 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 NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Potential competing interests: Chan previously served as a consultant for Bayer Healthcare, Millennium Pharmaceuticals, Pozen, and Pfizer. This study was not funded by Bayer Healthcare, Millennium Pharmaceuticals, Pozen, or Pfizer. This is redundant, and we will keep the last sentence. Meyerson has applied for a patent on Fusobacterium in colorectal cancer diagnosis and had ownership interest in and was a consultant and advisory board member for Foundation Medicine. He also receives research support from Bayer. The other authors declare no competing financial interest.

Acknowledgments. We 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: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY. We assume full responsibility for analyses and interpretation of these data. We have used HUGO (Human Genome Organisation)-approved official symbols for genes and gene products, including BRAF, CACNA1G, CDKN2A, CRABP1, IGF2, KRAS, MLH1, NEUROG1, PIK3CA, RUNX3, SLCO2A1, and SOCS1, all of which are described at www.genenames.org.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ Colorectal cancer is typically classified into rectal, distal colon, and proximal colon cancers.
- ✓ Emerging evidence indicates a gradual change in molecular features of colorectal cancer along bowel subsites.
- ✓ Contents and microbiota of the large intestine may change gradually from the proximal to distal segments.
- ✓ *Fusobacterium nucleatum* has been detected predominantly in proximal colon cancer and may potentiate colonic neoplasia development.

WHAT IS NEW HERE

- ✓ The proportion of colorectal cancers with a high amount of *F. nucleatum* gradually increases from rectum to cecum.
- ✓ This trend along colorectal subsites is statistically significantly linear.
- ✓ Our data support the colorectal continuum model reflecting carcinogenic influences of *F. nucleatum*.

1. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* 2014; **12**: 661–672.
2. Keku TO, Dulal S, Deveaux A *et al.* The gastrointestinal microbiota and colorectal cancer. *Am J Physiol Gastrointest Liver Physiol* 2015; **308**: G351–G363.
3. Ericsson AC, Akter S, Hanson MM *et al.* Differential susceptibility to colorectal cancer due to naturally occurring gut microbiota. *Oncotarget* 2015; **6**: 33689–33704.
4. Dejea CM, Wick EC, Hechenbleikner EM *et al.* Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc Natl Acad Sci USA* 2014; **111**: 18321–18326.
5. Belcheva A, Irrazabal T, Robertson SJ *et al.* Gut microbial metabolism drives transformation of MSH2-deficient colon epithelial cells. *Cell* 2014; **158**: 288–299.
6. Dulal S, Keku TO. Gut microbiome and colorectal adenomas. *Cancer J* 2014; **20**: 225–231.
7. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. *Cell Host Microbe* 2014; **15**: 317–328.
8. Gur C, Ibrahim Y, Isaacson B *et al.* Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* 2015; **42**: 344–355.
9. Kostic AD, Chun E, Robertson L *et al. Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 2013; **14**: 207–215.
10. Rubinstein MR, Wang X, Liu W *et al. Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe* 2013; **14**: 195–206.
11. Kostic AD, Gevers D, Pedamallu CS *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* 2012; **22**: 292–298.
12. Castellarin M, Warren RL, Freeman JD *et al. Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res* 2012; **22**: 299–306.
13. McCoy AN, Araujo-Perez F, Azcarate-Peril A *et al.* Fusobacterium is associated with colorectal adenomas. *PLoS One* 2013; **8**: e53653.
14. Tahara T, Yamamoto E, Suzuki H *et al.* Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res* 2014; **74**: 1311–1318.
15. Ito M, Kanno S, Nosho K *et al.* Association of *Fusobacterium nucleatum* with clinical and molecular features in colorectal serrated pathway. *Int J Cancer* 2015; **137**: 1258–1268.
16. Mima K, Sukawa Y, Nishihara R *et al. Fusobacterium nucleatum* and T cells in colorectal carcinoma. *JAMA Oncol* 2015; **1**: 653–661.
17. Nosho K, Sukawa Y, Adachi Y *et al.* Association of *Fusobacterium nucleatum* with immunity and molecular alterations in colorectal cancer. *World J Gastroenterol* 2016; **22**: 557–566.
18. Li YY, Ge QX, Cao J *et al.* Association of *Fusobacterium nucleatum* infection with colorectal cancer in Chinese patients. *World J Gastroenterol* 2016; **22**: 3227–3233.
19. Iacopetta B. Are there two sides to colorectal cancer? *Int J Cancer* 2002; **101**: 403–408.
20. Gervaz P, Bucher P, Morel P. Two colons-two cancers: paradigm shift and clinical implications. *J Surg Oncol* 2004; **88**: 261–266.
21. Carethers JM. One colon lumen but two organs. *Gastroenterology* 2011; **141**: 411–412.
22. Yamauchi M, Morikawa T, Kuchiba A *et al.* Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut* 2012; **61**: 847–854.
23. Phipps AI, Buchanan DD, Makar KW *et al.* BRAF mutation status and survival after colorectal cancer diagnosis according to patient and tumor characteristics. *Cancer Epidemiol Biomarkers Prev* 2012; **21**: 1792–1798.
24. Rosty C, Young JP, Walsh MD *et al.* PIK3CA activating mutation in colorectal carcinoma: associations with molecular features and survival. *PLoS One* 2013; **8**: e65479.
25. Phipps AI, Limburg PJ, Baron JA *et al.* Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology* 2015; **148**: 77–87 e72.
26. Cummings JH, Pomare EW, Branch WJ *et al.* Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* 1987; **28**: 1221–1227.
27. Hayashi H, Takahashi R, Nishi T *et al.* Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol* 2005; **54**: 1093–1101.
28. Li X, LeBlanc J, Truong A *et al.* A metaproteomic approach to study human-microbial ecosystems at the mucosal luminal interface. *PLoS One* 2011; **6**: e26542.
29. Mowat AM, Agace WW. Regional specialization within the intestinal immune system. *Nat Rev Immunol* 2014; **14**: 667–685.
30. Liao X, Lochhead P, Nishihara R *et al.* Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. *N Engl J Med* 2012; **367**: 1596–1606.
31. Nishihara R, Wu K, Lochhead P *et al.* Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med* 2013; **369**: 1095–1105.
32. Mima K, Nishihara R, Qian ZR *et al. Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut* 2016 in press.
33. Ogino S, Nosho K, Kirkner GJ *et al.* CpG island methylator phenotype, micro-satellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 2009; **58**: 90–96.
34. Irahara N, Nosho 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**: 177–183.
35. Ogino S, Kawasaki T, Brahmandam M *et al.* Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. *J Mol Diagn* 2006; **8**: 209–217.
36. 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**: 305–314.
37. Ogino S, Kawasaki T, Brahmandam M *et al.* Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. *J Mol Diagn* 2005; **7**: 413–421.
38. Imamura Y, Lochhead P, Yamauchi M *et al.* Analyses of clinicopathological, molecular, and prognostic associations of KRAS codon 61 and codon 146 mutations in colorectal cancer: cohort study and literature review. *Mol Cancer* 2014; **13**: 135.
39. Nosho K, Kawasaki T, Ohnishi M *et al.* PIK3CA mutation in colorectal cancer: relationship with genetic and epigenetic alterations. *Neoplasia* 2008; **10**: 534–541.
40. Liao X, Morikawa T, Lochhead P *et al.* Prognostic role of PIK3CA mutation in colorectal cancer: cohort study and literature review. *Clin Cancer Res* 2012; **18**: 2257–2268.
41. 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**: 674–678.
42. Yamauchi M, Lochhead P, Morikawa T *et al.* Colorectal cancer: a tale of two sides or a continuum? *Gut* 2012; **61**: 794–797.
43. Jess P, Hansen IO, Gamborg M *et al.* A nationwide Danish cohort study challenging the categorisation into right-sided and left-sided colon cancer. *BMJ Open* 2013; **3**: e002608.

44. Kudryavtseva AV, Lipatova AV, Zaretsky AR *et al.* Important molecular genetic markers of colorectal cancer. *Oncotarget* 2016 in press.
45. Colussi D, Brandi G, Bazzoli F *et al.* Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *Int J Mol Sci* 2013; **14**: 16365–16385.
46. Lin JH, Giovannucci E. Environmental exposure and tumor heterogeneity in colorectal cancer risk and outcomes. *Curr Colorectal Cancer Rep* 2014; **10**: 94–104.
47. Kocarnik JM, Shiovitz S, Phipps AI. Molecular phenotypes of colorectal cancer and potential clinical applications. *Gastroenterol Rep (Oxf)* 2015; **3**: 269–276.
48. Kuipers EJ, Grady WM, Lieberman D *et al.* Colorectal cancer. *Nat Rev Dis Primers* 2015; **1**: 15065.
49. Zong L, Abe M, Ji J *et al.* Tracking the correlation between CpG island methylator phenotype and other molecular features and clinicopathological features in human colorectal cancers: a systematic review and meta-analysis. *Clin Transl Gastroenterol* 2016; **7**: e151.
50. Pellatt DF, Stevens JR, Wolff RK *et al.* Expression profiles of miRNA subsets distinguish human colorectal carcinoma and normal colonic mucosa. *Clin Transl Gastroenterol* 2016; **7**: e152.
51. Kedrin D, Gala MK. Genetics of the serrated pathway to colorectal cancer. *Clin Transl Gastroenterol* 2015; **6**: e84.
52. Lupton JR. Microbial degradation products influence colon cancer risk: the butyrate controversy. *J Nutr* 2004; **134**: 479–482.
53. Singh N, Gurav A, Sivaprakasam S *et al.* Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* 2014; **40**: 128–139.
54. Donohoe DR, Collins LB, Wali A *et al.* The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol Cell* 2012; **48**: 612–626.
55. Rosty C, Young JP, Walsh MD *et al.* Colorectal carcinomas with KRAS mutation are associated with distinctive morphological and molecular features. *Mod Pathol* 2013; **26**: 825–834.
56. Phipps AI, Chan AT, Ogino S. Anatomic subsite of primary colorectal cancer and subsequent risk and distribution of second cancers. *Cancer* 2013; **119**: 3140–3147.
57. Wong RJ. Marked variations in proximal colon cancer survival by race/ethnicity within the United States. *J Clin Gastroenterol* 2010; **44**: 625–630.
58. Cheng L, Eng C, Nieman LZ *et al.* Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. *Am J Clin Oncol* 2011; **34**: 573–580.
59. Samadder NJ, Curtin K, Tuohy TM *et al.* Characteristics of missed or interval colorectal cancer and patient survival: a population-based study. *Gastroenterology* 2014; **146**: 950–960.
60. Robertson DJ, Lieberman DA, Winawer SJ *et al.* Colorectal cancers soon after colonoscopy: a pooled multicohort analysis. *Gut* 2014; **63**: 949–956.
61. Biedermann L, Zeitz J, Mwinyi J *et al.* Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. *PLoS One* 2013; **8**: e59260.
62. O'Keefe SJ, Li JV, Lahti L *et al.* Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun* 2015; **6**: 6342.
63. Zitvogel L, Galluzzi L, Viaud S *et al.* Cancer and the gut microbiota: an unexpected link. *Sci Transl Med* 2015; **7**: 271ps271.
64. 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.
65. Ogino S, Nishihara R, VanderWeele TJ *et al.* Review article: The role of molecular pathological epidemiology in the study of neoplastic and non-neoplastic diseases in the era of precision medicine. *Epidemiology* 2016; **27**: 602–611.
66. Arthur JC, Perez-Chanona E, Muhlbauer M *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 2012; **338**: 120–123.
67. Wu S, Rhee KJ, Albesiano E *et al.* A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17T cell responses. *Nat Med* 2009; **15**: 1016–1022.

Supplementary Information accompanies this paper on the Clinical and Translational Gastroenterology website (http://www.nature.com/ctg)