Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures

Iradj Sobhani[a,b,1], Emma Bergsten[a,c], Séverine Couffin[a], Aurélien Amiot[a,b], Biba Nebbad[d], Caroline Barau[e], Nicola de'Angelis[f], Sylvie Rabot[g], Florence Canoui-Poitrine[h], Denis Mestivier[a,i], Thierry Pédron[c], Khashayarsha Khazaie[j,2], and Philippe J. Sansonetti[c,k,1,2]

[a]EA7375 (EC2M3 Research Team), Université Paris Est, Créteil 94000, France; [b]Service de Gastroenterologie, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; [c]Unité de Pathogénie Microbienne Moléculaire, INSERM U1202, Institut Pasteur, Paris 75015, France; [d]Service de Microbiologie, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; [e]Plateforme de Ressources Biologique, Hôpital Henri Mondor, Créteil 94000, France; [f]Service de Chirurgie Digestive, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; [g]Micalis Institute, Institut National de la Recherche Agronomique (INRA), AgroParisTech, Université Paris-Saclay, Jouy-en-Josas 78352, France; [h]Service de Santé Publique, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; [i]Bioinformatics Core Lab, INSERM U955, Institut Mondor de Recherche Biomédicale, Créteil 94000, France; [j]Department of Immunology, Mayo Clinic, Rochester, MN 55905; and [k]Chaire de Microbiologie et Maladies Infectieuses, Collège de France, Paris 75005, France

Contributed by Philippe J. Sansonetti, September 27, 2019 (sent for review July 17, 2019; reviewed by Guido Grandi and Sven Pettersson)

Sporadic colorectal cancer (CRC) is a result of complex interactions between the host and its environment. Environmental stressors act by causing host cell DNA alterations implicated in the onset of cancer. Here we investigate the stressor ability of CRC-associated gut dysbiosis as causal agent of host DNA alterations. The epigenetic nature of these alterations was investigated in humans and in mice. Germ-free mice receiving fecal samples from subjects with normal colonoscopy or from CRC patients were monitored for 7 or 14 wk. Aberrant crypt foci, luminal microbiota, and DNA alterations (colonic exome sequencing and methylation patterns) were monitored following human feces transfer. CRC-associated microbiota induced higher numbers of hypermethylated genes in murine colonic mucosa (vs. healthy controls' microbiota recipients). Several gene promoters including SFRP1,2,3, PENK, NPY, ALX4, SEPT9, and WIF1 promoters were found hypermethylated in CRC but not in normal tissues or effluents from fecal donors. In a pilot study (n = 266), the blood methylation levels of 3 genes (*Wif1*, *PENK*, and *NPY*) were shown closely associated with CRC dysbiosis. In a validation study (n = 1,000), the cumulative methylation index (CMI) of these genes was significantly higher in CRCs than in controls. Further, CMI appeared as an independent risk factor for CRC diagnosis as shown by multivariate analysis that included fecal immunochemical blood test. Consequently, fecal bacterial species in individuals with higher CMI in blood were identified by whole metagenomic analysis. Thus, CRC-related dysbiosis induces methylation of host genes, and corresponding CMIs together with associated bacteria are potential biomarkers for CRC.

colon | cancer | microbiota | gene methylation | biomarker

Colorectal cancer (CRC) is among the most common malignancies worldwide with a high mortality rate and is believed to result from interactions between the host and long-term environmental exposures. Environmental chemicals have been associated with a higher incidence of various cancers (1) and may act as either carcinogens or tumor-promoting agents (2) by causing an accumulation of DNA mutations (3) and epigenetic changes in DNA within host cells (4). At a cellular level, the *Wnt* pathway has been generally accepted as important contributor (5). Nevertheless, a comprehensive assessment of all factors involved remains an extremely challenging endeavor, primarily due to the great variety of environmental stressors and the long-term exposure period of many years. However, significant associations linking diet to features of the microbiota and to CRC were reported recently (6, 7). Therefore, analyzing the microbiota may provide new insights into the effects of environmental exposures on CRC.

The human gut microbiota contains trillions of microorganisms and is critical for overall health, playing a fundamental role in interactions with environmental drivers (i.e., nutrition and medicine) (8). We have shown that alterations in the colonic luminal and adherent microbiota composition are associated with CRC (9, 10). Interestingly, it has been reported that microbiota dysbiosis increases the risk for CRC even in individuals with genetic predisposition to CRC, including constitutional mutations of pivotal genes (i.e., APC) (11).

We previously demonstrated that hypermethylation of the *Wif1* promoter, the gene regulating the *Wnt* pathway, serves as surrogate diagnostic marker for early CRC (12).

The objective of this study was to assess the hypothesis that CRC-associated environmental factors may act by altering the composition of the gut microbiota and to investigate the underlying epigenetic pathways involved in CRC promotion. Thus, in a first step, we characterized the colonic luminal microbiota in humans and transferred it into germ-free mice in order to assess the

Significance

This study advances our appreciation and understanding of the role of colon dysbiosis in the pathogenesis of colorectal cancer. In a human pilot study of 266 individuals, greater epigenomic (methylation) DNA alterations correlated with CRC and microbiota composition. Beyond this correlative evidence, when germ-free mice received fresh feces from CRC patients and their healthy controls, the former animals developed colon epithelial renewal, more precancerous lesions, and increased tissue and blood DNA methylation in intestinal tissues. Confirmation was obtained in a larger cohort of 1,000 patients, indicating that CRC-associated dysbiosis may promote colon carcinogenesis via epigenome dysregulation. Gene methylation can therefore serve as a marker for CRC and likely for predicting efficacy of prebiotic supplementation in average-risk individuals.

Author contributions: I.S. designed research; I.S., E.B., S.C., A.A., B.N., C.B., N.d.A., S.R., F.C.-P., T.P., and P.J.S. performed research; I.S., E.B., A.A., N.d.A., F.C.-P., D.M., K.K., and P.J.S. analyzed data; and I.S., T.P., K.K., and P.J.S. wrote the paper.

Reviewers: G.G., University of Trento; and S.P., Lee Kong Chian School of Medicine.

Competing interest statement: I.S. shares rights in 3 patents: EP B31120, EP2635705, and EP 2955232 A1 20151216 based on methods for diagnosing adenomas and/or colorectal cancer.

Data deposition: Data related to this paper are available from the European Nucleotide Archive (ENA) database (http://www.ebi.ac.uk/ena) under the accession nos. ERX3622297–ERX3622402, ERR3628499–ERR3628604, ERS3936180–ERS3936285, and PRJEB35144.

[1]To whom correspondence may be addressed. Email: iradj.sobhani@aphp.fr or philippe.sansonetti@pasteur.fr.

[2]K.K. and P.J.S. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1912129116/-/DCSupplemental.

First published November 11, 2019.

potential effect of this microbiota as an environmental stressor for host DNA. Second, we investigated the links between dysbiosis and host gene methylated markers suggested by animal experiments in a pilot human study and validated in a large human cohort.

Results

Human Fecal Microbiota Transfer to Germ-Free Mice: Clinical and Routine Biochemical Parameters Among Groups. We performed human fecal microbiota transfer (FMT) to germ-free mice utilizing material from patient donors with CRC ($n = 9$) and healthy control donors ($n = 9$) as illustrated in *SI Appendix*, Fig. S1. In addition, mice received the carcinogen azoxymethane (AOM) or saline control for 10 wk in order to assess potential effects of chemically induced stress. As shown in *SI Appendix*, Tables S1 *A* and *B*, FMT had no effect on the general blood parameters of the recipient mice as no significant differences were observed among any of the groups (a total of 185 mice) at 7 and 14 wk. At 14 wk, mice that received both AOM and CRC microbiota (CRC-μ) showed a lower mean body weight compared to the normal microbiota (N-μ) controls. This was accounted to reduced food intake observed in the CRC-μ group when compared to the N-μ controls (1.5 ± 0.1 g/100 g food intake vs. 2.0 ± 0.1 g/100 g food intake, respectively; $P < 0.01$). Interestingly, this weight effect disappeared when mice were given PEG in their drinking water, which is known to prevent bacterial colonization of the gut (*SI Appendix*, Fig. S2). In line with this, the luminal bacterial load, adherence of luminal bacteria to the colonic mucosa, and short-chain fatty acid (SCFA) concentrations in the cecum were significantly lower in mice that received PEG as compared to the controls (*SI Appendix*, Fig. S2 *A* and *B*). These results suggest that human CRC FMT may induce systemic changes in mice mediated by the interaction of bacteria with the colonic mucosa.

Dysbiosis-Related Histological Alterations in Colonic Tissue After Human FMT in Mice. In order to delineate the predominant modifications occurring in murine colonic tissues following FMT, the number of aberrant crypt foci (ACF), i.e., clusters of enlarged thick crypts visualized by methylene blue at low magnification, were enumerated. Whereas histological features appeared normal in the group of mice receiving only N-μ, mild inflammation and ACF were seen at significant levels in the groups given CRC-μ, AOM, or both (Fig. 1 *A*–*D*, *H*, and *I*). After 7 or 14 wk, ACF numbers were highest in the group given CRC-μ + AOM, compared to all other groups, and were also significantly higher in the groups given CRC-μ alone or AOM alone, compared to the N-μ only group (Fig. 1 *H* and *I*). No ACFs were seen in mice given PEG, in which the microbiota was efficiently reduced in the mucus layer proximal to the mucosa. These latter samples were therefore not further examined. According to the colon length, numbers of ACFs became significantly higher in groups given CRC-μ versus N-μ, when only 20 mm of the left colonic mucosa was examined (13 ± 3 versus 5 ± 1; $P < 0.05$; Fig. 1*I*).

Human FMT to Germ-Free Mice: Effect on Epithelial Proliferation and Gene Expression. Epithelial cell proliferation in the colonic mucosa was visualized by KI67 staining (as illustrated in Fig. 1*G*). At 7 wk following FMT, a significant increase of proliferation was observed in the groups given AOM as compared to all other groups (Fig. 1*H*). However, at week 14, cell proliferation was significantly greater in the CRC-μ, as compared to N-μ recipients with AOM exerting an additional effect (Fig. 1 *H* and *I*), and ACFs increased with the rate of epithelial cell proliferation. These results are consistent with those reported after FMT from only 5 CRC patients as compared to 5 control individuals' microbiota (13). Further, to investigate the involvement of *Wnt* and *Notch* pathways that control intestinal cell fate, mRNA of key mediators was quantified. qPCR analysis of mRNA from the colonic mucosa of recipient mice revealed an increase of 1.7-fold, 1.9-fold, and 1.9-fold expression levels of the transcription factors *HES1*, *KLF4*, and *ELF3*, respectively, but not of *MATH1* in CRC-μ compared to N-μ recipients (*SI Appendix*, Table S2).

Human FMT to Germ-Free Mice: Mucosal Inflammation. Early after FMT (7 wk), a mild inflammatory response reflected by higher IL1β, IL6, and MIP2α levels and lower IL10, IL23, and INFγ levels was detected in CRC-μ recipient animals as compared to N-μ controls. AOM treatment tended to to amplify this stimulatory effect as assessed by mRNA cytokine quantification (Fig. 1*J*). In addition, semiquantitative evaluation of mucosal myeloid cells along the entire histologically examined intestine (10 fields per sample, 3 samples per mouse) showed a trend toward increased numbers in human CRC-μ as compared to N-μ recipients but without reaching statistical significance at any time point (Fig. 1*K*).

Microbiota Characterization Following Human FMT in Mice. To assess for the impact of the microbial communities on the observed differences in the mucosal phenotypes, murine fecal samples were analyzed by 16S rRNA gene sequencing. At baseline, the fecal microbiota of the CRC-μ ($n = 9$ donors) group contained a higher proportion of *Fusobacteria*, *Parvimonas*, *Butyrivibrio*, *Gemella*, and *Akkermansia* and lower proportions of *Ruminococcus*, *Bifidobacterium*, *Eubacteria*, and *Lachnospira*, as compared to stools from the N-μ ($n = 9$ donors) group (*SI Appendix*, Table S3 and Fig. S3).

During the follow-up, all groups ($n = 6$ per group and at each time point) showed a moderate decrease in *Coccoides*, *Clostridium leptum*, and *Bifidobacterium* on day 7 and stability in the community of bacteria afterward, as estimated by qPCR (*SI Appendix*, Table S4). Based on 16S rRNA gene sequencing of mouse stools ($n = 6$ in each group at each time point of the follow-up), 85% of OTUs remained unchanged over time. Overall, the bacterial species most associated with significant histopathological alterations, i.e., precancerous lesions (e.g., ACF), were *Firmicutes*, *Clostridia*, and *Clostridiales*. These histological changes were also associated with lower counts of genera with an antiinflammatory effect (i.e., *Faecalibaterium* and *Eubacterium*) and of butyrate-producing bacteria (*Firmicutes* species) (*SI Appendix*, Fig. S4). Interestingly, the abundance of *Coprococcus* was lower and of *Bacteroides* was higher in CRC-μ vs. N-μ. Covariations of these 2 genera with histological changes characterized by ACF numbers and with mRNA levels of the previously quantified transcription factors in mice colonic parameters were analyzed. Coinertia analysis revealed that these parameters were associated with variation of the fecal bacteria. The first component of the analysis showed a significant separation between N-μ and CRC-μ recipients. This first component was significantly associated linearly with *Bacteroides* and *Coprococcus* genera abundances in mice stools (*SI Appendix*, Fig. S5). These allowed a significant separation between N-μ and CRC-μ recipients regarding bacteria and mRNA levels of key transcription factors involved in *Wnt* and *Notch* pathways.

Human FMT to Germ-Free Mice Alters DNA.

DNA mutations. In order to analyze the potential of FMT to induce DNA modifications in the host, whole-genome sequencing of murine colonic mucosal tissues ($n = 12$, 3 in each group of recipients) was performed, covering 220,000 exons within 24,000 genes. Overall, the incidence of mutations at global exon/intron level was significantly higher in intestinal vs. spleen tissues. Results of the colonic mucosa showed a trend toward increased DNA alterations in exons or introns of CRC-μ compared to N-μ recipients, but this did not reach significance (Fig. 2*A*). In addition, mice given CRC-μ and AOM showed the highest levels of DNA alterations in exons or introns showing a trend to an additive effect of CRC-μ as compared to N-μ. Interestingly, Principal Component Analysis (PcA) scatter diagrams separated

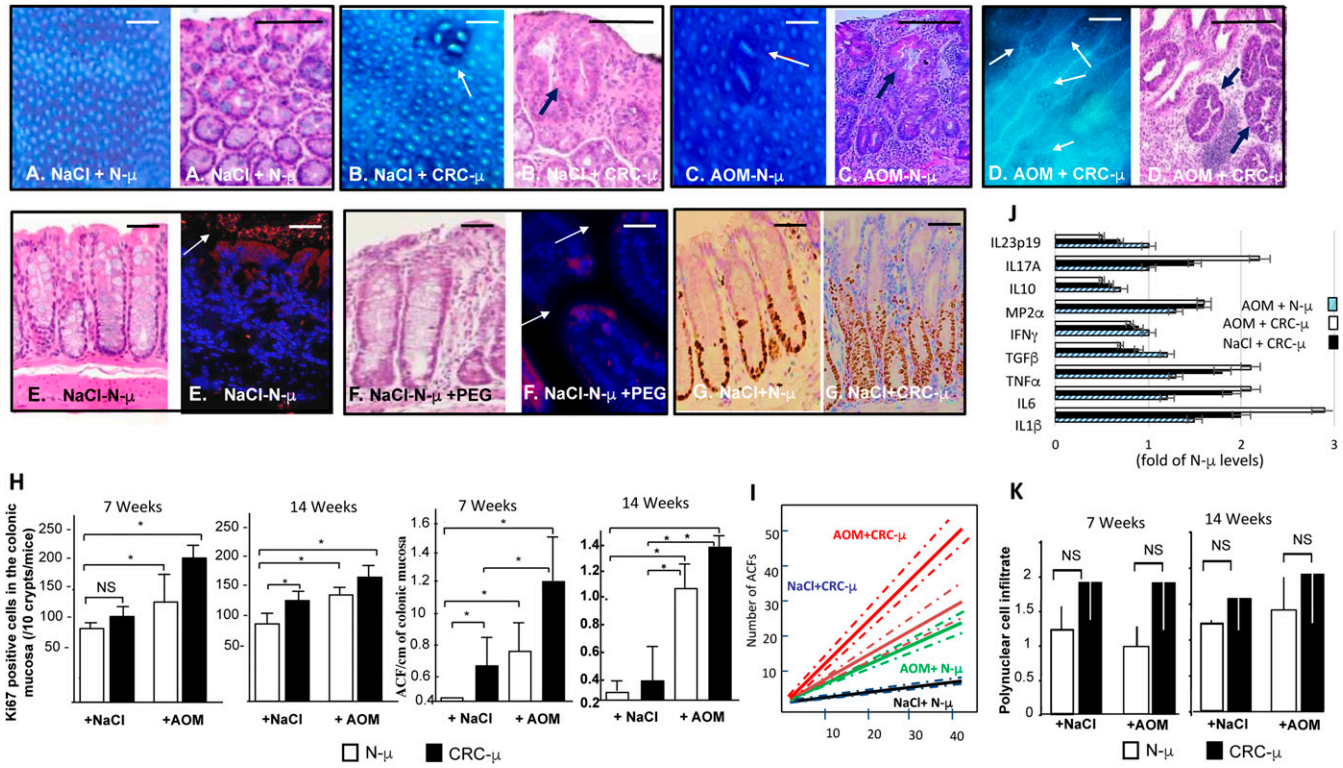

Fig. 1. Histological patterns of murine colonic mucosa following FMT. After the intestine was removed from cecum to anus, mucosa was carefully pinned flat, without folds, to examine the totality of the colonic mucosa which were stained with 0.2% methylene blue (left slides of coupled slides *A* to *D*) or HES (right slides of coupled slides *A* to *D*). Numbers of mice were as follows: $n = 53$ in 7-wk study (CRC-μ transfer, $n = 30$, and N-μ transfer, $n = 23$) and $n = 132$ in 14-wk study (CRC-μ transfer, $n = 66$, and N-μ transfer, $n = 66$). (Scale bar: 50 μm.) (*A*) After FMT from healthy human controls (N-μ), no ACF were visible in the colonic mucosa (here after 14 wk in *Left*), and pattern of crypts was normal by HES staining (*Right*). (*B*) Elevated numbers of ACF were observed after FMT from patients with CRC (CRC-μ) as compared to N-μ with multiple ACF (arrow in *Left*) as verified by HES (*Right*). (*C*) ACF counts were higher in the animals given AOM as compared to N-μ. Arrows indicate double ACF under blue coloration (*Left*) and illustration on HES slide of colonic mucosa. (*D*) The combination of CRC-μ and AOM increased the ACF count with dysplasia in rare cases (arrow). (*E*) No ACF nor inflammatory cell infiltrate was visible in mice given N-μ under NaCl (HES), although FISH staining showed density of bacteria trapped in the mucus layer (arrow, *Left*). (*F*) No ACF and no inflammatory cell infiltrate nor injury were noticed when mice received polyethylene glycol (PEG) in their drinking water as shown by H&S staining (HES) when FISH staining shows clear decrease in density of bacteria trapped in the proximal mucus layer (arrow). (*G*) Representative pictures of KI67 staining after human FMT from patients with CRC (CRC-μ) or from N-μ recipients. (*H*) Cell proliferation assessed by Ki67 staining and ACF quantification 7 and 14 wk after FMT in mice in the intestinal mucosa were higher in CRC-μ than in N-μ recipients. (*I*) The number of ACF counts were enhanced depending on the length of mucosa examined and the mice subgroups. (*J*) Comparative transcriptional levels of a set of inflammatory cytokines in the colonic mucosa assessed by murine cytokine qPCR quantification showing a trend to higher IL1, IL6, MIP2, and IL17 and lower IFNγ, IL10, and IL23 in mice given CRC-μ alone (fold vs. N-μ given mice) with AOM boosting this effect that reached significance for IL6, TNFα, and IL10 in mice given CRC-μ + AOM (vs. N-μ + AOM). (*K*) Inflammatory cell infiltrate in the colonic mucosa as assessed by semiquantification on HES stained slides (10 consecutive fields) under optic microscope magnification 20: a pathologist blinded to animal groups used a semiquantitative score to evaluate myeloid cell infiltrate in the colonic mucosa as 0, 1, and 2 indicating absence, scarce, and numerous inflammatory cells, respectively. The groups were compared by 1-way ANOVA followed by the Tukey–Kramer multiple comparisons post hoc test. No significant difference in between mice groups was observed. $^*P < 0.05$; NS, not significant.

animal subgroups according to the total DNA changes in single nucleotides in the colonic mucosa or spleen samples suggesting a link between gene mutation and type of microbiota (N or CRC) or treatment (saline or AOM) given to mice (Fig. 2 *B* and *C*).

In order to investigate the local oncogenic potential of CRC-μ, we went on to perform in-depth analysis of the following selected gene pathways (number of genes analyzed): Wnt and β–catenin (19 genes), Notch (4 genes), PPAR (3 genes), SMAD (2 genes), TGF-β (2 genes), ACRV (2 genes), DKK (4 genes), TCF (2 genes), MYC (1 gene), and SOCS (1 gene). Changes were most prominent in pivotal *Wnt* pathway genes (with indel and single nucleotide polymorphisms in 11 *Wnt* genes), with no significant differences between the CRC-μ and N-μ groups. When all DNA mutations were pooled, however, the number of DNA changes was highest in the CRC-μ + AOM recipients vs. the N-μ recipients (Fig. 2*D*). Again, DNA changes were significantly more numerous in colonic than in spleen tissues. Interestingly, in the spleen, DNA changes were similar among AOM-exposed animals that received CRC-μ or N-μ. Finally, an unsupervised analysis between mutated genes (including only selected genes as indicated above) and recipient status showed that changes in *Wnt*, *PPARγ*, and *Notch* pathway genes were associated with CRC-μ recipient, with an additive effect of AOM treatment (Fig. 2*C* and *SI Appendix*, Fig. S6 and Table S2).

DNA methylation/demethylation. We isolated colonic mucosal tissue from the colon ($n = 24$; 6 in each group) and evaluated epigenetic changes using the EPIC microarray to compare methylation rates in 63,987 probes reflecting 12,600 genes being expressed. For all probes pooled, both mean and median methylation rates decreased from the N-μ + NaCl group to the CRC-μ + NaCl and CRC-μ + AOM groups (Fig. 2*E*), with significantly lower methylation levels in the CRC-μ + AOM group than in the N-μ + NaCl group ($P < 0.01$). Overall, one third of the probes appeared unmethylated in all experimental conditions (<0.2 beta-value methylation mEPIC) in the colonic tissue after FMT (Fig. 2*F*). The number of methylated probes (>0.8) varied with N-μ, CRC-μ,

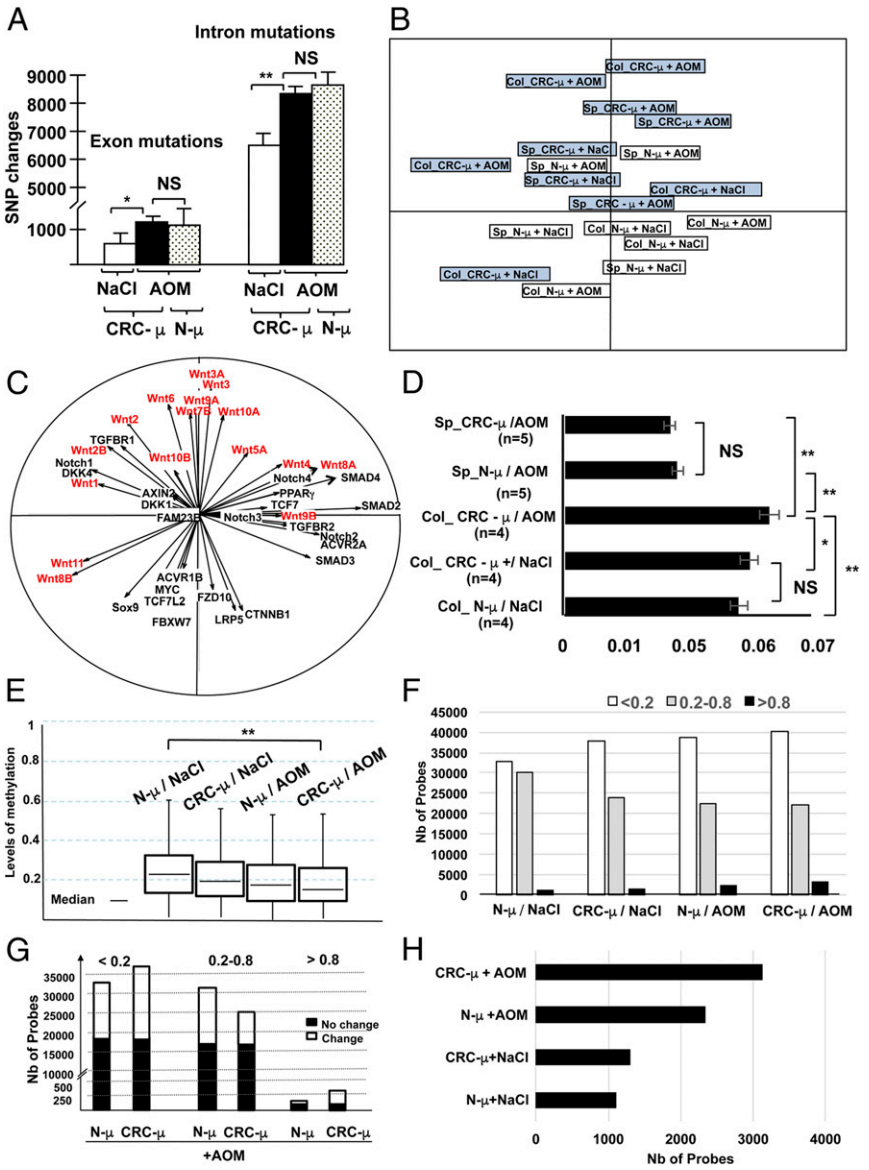

Fig. 2. DNA changes in mice after human FMT. Whole genome sequencing of total DNA extracted from colonic mucosa ($n = 12$) and spleen samples ($n = 6$) was performed. (*A*) All single-nucleotide polymorphisms (SNPs) in the colon and spleen samples are compared to the reference mouse genome (GRCm38), and mutation levels within gene segments are indicated. Mice given AOM showed the highest levels of DNA alterations in exons or introns with a trend of additive effect of CRC-μ as compared to N-μ. (*B*) Distribution of animal subgroups according to the total gene mutations. DNA changes in single nucleotides with PcA scatter diagrams for colonic or spleen samples in the groups of mice. The groups are identified by the type of human microbiota received (CRC-μ or N-μ for CRC patients' or controls' stool, respectively) and type of treatment (AOM or NaCl for azoxymethane or saline, respectively). (*C*) Correlation circle of targeted gene mutations in the colonic mucosa and spleen tissues according to PcA. Vector length reflects targeted gene mutation weight in the first 2 component analyses; targeted mutated genes are indicated (those of Wnt pathway in red color). (*D*) When mutations in all Wnt genes were pooled together, rates of mutations were significantly higher in mice given AOM with an additive effect of CRC-μ. The total number of mutations in *Wnt* pathway genes in both colonic mucosa and spleen (Sp) was the highest in the animals given the CRC-μ and AOM combination (see also *SI Appendix*, Fig. S6). The number of mutations was greater in colonic mucosa (but not in spleen tissues) with AOM combined with CRC microbiota compared to AOM combined with control microbiota. There was no significant effect in between colonic mucosa due to CRC-μ alone as compared to N-μ alone. Col, colonic mucosa; Sp, spleen. (*E*) DNA epigenetic changes were investigated by using mEPIC array (39). The methylation level of probes ($n = 63{,}987$) were estimated after bisulfite modification of DNAs. Changes based on the methylation of probes were investigated on DNAs from colon samples ($n = 16$; 4 mice from each experimental group): the level of methylated probes was quantified as reported (39) and ranged from 0 (not methylated) to 1 (fully methylated). DNAs were classified as unmethylated if the methylation value was <0.2 and as hypermethylated if the methylation value was >0.799. Overall, mean and median values of all probes pooled ($n = 63{,}987$) in each group of mice showed lowest values in the group of CRC-μ + AOM recipients and highest in control microbiota recipients. (*F*) Most of the probes were unmethylated in all animal groups; elevated numbers of both hypomethylated and hypermethylated probes were observed in the mice given CRC-μ + AOM. (*G*) Probes whose methylation level changed or remained unchanged under AOM in mice given N-μ or CRC-μ. Mice receiving CRC-μ had a greater number of genes with changed methylation levels. (*H*) Number of hypermethylated probes in each group showing highest level in mice given CRC-μ and AOM combination compared to all other groups. $^{**}P < 0.01$, $^{*}P < 0.05$; NS, not significant.

and AOM exposure, although the ratio of methylated/unmethylated probes was not significantly different (3.3% and 3.1%, respectively [$P = 0.25$, Fisher's test]) (Fig. 2*G*). The number of unmethylated probes was 11% higher in the CRC-μ + AOM group when compared to the N-μ + NaCl group ($P = 0.013$), and the number of fully methylated probes was higher in CRC-μ and AOM recipients

(Fig. 2*H*). In summary, human CRC-μ induced greater epigenetic alterations in murine colonic mucosal tissues when compared to N-μ. The ratios of methylation-shifted probes (methylated to unmethylated and vice versa) were higher in the CRC-μ group than in the N-μ group, with AOM. When we limited the analysis to in silico specific EPIC probes, two-thirds of the probes with hypermethylation levels (0.8 or more) in the CRC-μ + AOM group were unmethylated (<0.2) in the N-μ group, corresponding to 46 genes in the CRC-μ + AOM group being silenced vs. N-μ group. In contrast, various probes in *Wnt* and *Notch* gene families appeared unmethylated in the CRC-μ + AOM group but were methylated in the N-μ group. Overall, significant DNA epigenetic rather than mutation changes in several genes were associated with CRC-μ human FMT in the mouse colonic mucosa. Pooling of total genes with expected DNA alterations (i.e., hypermethylation, or mutation in exon) revealed the involvement of pathways implicated in cell growth, signal transduction, nucleic acid binding, protein synthesis, channel, and carrier protein (*SI Appendix*, Fig. S7).

Gene Methylation in Human Colonic Samples: From Bench to Bed. To investigate whether the gene methylation patterns observed in the mouse were also associated with CRC dysbiosis in humans, we first investigated methylated genes in CRC tissues and effluents such as blood and stool ($n = 9$) and in normal tissue effluent controls ($n = 9$), both obtained from the same individuals as for the FMT in mice (Fig. 3*A*). In order to develop a quick and easy methylation test, panels of genes were selected, based on the difference between normal and CRC in various effluents and tissues. Among the various genes classified as hypermethylated in the individual samples (tissue, stool, and blood) from the same individuals, only 8 (*Wif1*-regulating gene and *SEPT9*, *SFRP1,2,3*, *PENK*, *NPY*, and *ALX4* genes) of these were common to all samples and were therefore subjected to further characterization (Fig. 3*B* and *SI Appendix*, Figs. S8–S11). We selected *Wif1* to represent the *Wnt* pathway and *NPY* and *PENK*, 2 other tissue CRC-associated hypermethylated genes, for methylation testing, as at least either one was found hypermethylated in all tumor tissues. The sum of the methylation levels of the 3 selected genes was found to serve best when discriminating between effluents from normal and cancer individuals (details in *SI Appendix*). Finally, to conceive easy and reproducible biological testing for large cohort screening, a housekeeping gene plus selected genes (i.e., *ALB*, *WIF1*, *NPY*, and *PENK*) were combined in a multiplex procedure for qPCR measurement in blood. A cumulative methyl index (CMI) was determined in blood samples from participants with normal colonoscopy or with a cancer diagnosed (32 patients with CRC and 46 with extra colonic cancer; *SI Appendix*). To validate its performance, CMI assessment was performed in a pilot study including 266 individuals (*SI Appendix*, Table S5) from a well-characterized cohort designed as CCR1 (*SI Appendix*, Table S6) and in a final validation study including 999 individuals (Table 1) from a second cohort designed as CCR2 (*SI Appendix*, Table S7). The specificity and the sensitivity of CMI > 2 in blood was 95 and 59%, respectively, in the pilot study when those with positive fecal occult blood test (FOBT) in stool were 97 and 43%, respectively. In the validation study, half of individuals were enrolled through mass CRC-screening programs and had a new fecal immunochemical test (FIT) to directly measure human hemoglobin in stools, and all had a blood with a threshold of CMI > 2 for positivity.

Validation of a CMI. Demographic, clinical characteristics, and results of CMI were compared between CRC patients and those with normal colonoscopy in the validation study; in addition, FIT was performed in those asymptomatic individuals who underwent colonoscopy, and we investigated whether positive FIT and CMI > 2 could be independent parameters for diagnosis of CRC (Table 1 and *SI Appendix*, Table S7).

A CMI > 2 correlated significantly with CRC in 999 individuals in the validation study. Interestingly, a tendency of CMI increase with age was noted; however, this observation did not

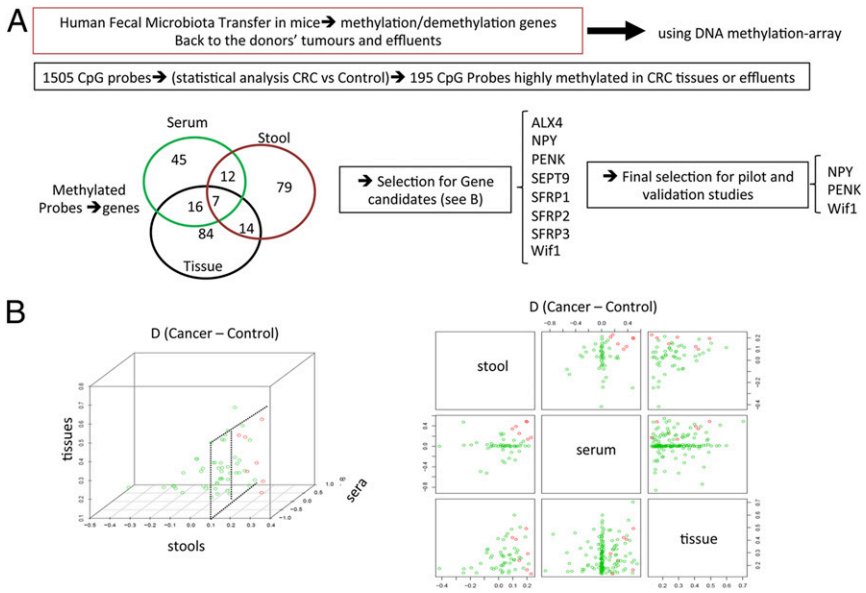

Fig. 3. Identification of hypermethylated genes related to fecal microbiota in human. Overview on the strategy from experimental approach for the validation of gene methylated targets in human based on microbiota donors (CRC patients or controls) in germ-free mice experiments. Methylated genes in CRC-associated tissues and fluids were identified based on their power for showing differences between normal colonoscopy individuals and CRC patients. (*A*) Human tissues and effluents were submitted to methylation gene array. Based on significant differences of methylation values in CpG probes between control ($n = 9$) and CRC patient ($n = 9$) donors, genes were selected according to the promoter segments hypermethylated in CRC patients. (*B*) Bidimensional (*Right*) and tridimensional (*Left*) distribution of genes regarding the difference in methylation values are indicated; in red color are indicated 7 selected more discriminant genes regarding CRC patients and controls. D, difference.

Table 1. Characteristics of participants in the validation CCR2 cohort

		Vatnimad (stage 0, I, and II)	Symptomatic (stage III and IV CRC)	
Age, y (n = 981)		63 [57 to 71]	60 [53 to 68]	<0.001
Gender (n = 981)	Female	77 (41.2)	392 (48.8)	0.06
Body mass index, kg/m^2		25.31 [22.86 to 28.09]	25.31 [22.39 to 28.7]	0.96
History of GI cancer (n = 869)	Yes	7 (4.4)	31 (4.4)	0.97
History of non-GI cancer (n = 868)	Yes	29 (18.4)	88 (12.4)	0.049
Family history of cancer (n = 866)	Yes	19 (11.7)	114 (16.4)	0.13
Diabetes (n = 980)	Yes	32 (17.1)	28 (16.3)	0.74
Blood methylation test				
Cumulative (*Wif1* + *PENK* + *NPY*) methylation index	>2	36 (20.3)	36 (4.6)	<0.001
Wif1 + *PENK* + *NPY*	Continuous	0.16 [0.01 to 1.16]	0.12 [0.01 to 0.40]	<0.001
Single gene				
Wif1 (n = 956)	>2	20 (11.4)	21 (2.7)	<0.001
	Continuous	0.06 [0 to 0.49]	0.07 [0 to 0.24]	0.002
PENK (n = 956)	>2	3 (1.7)	3 (0.4)	0.045
	Continuous	0 [0 to 0.04]	0 [0 to 0.02]	0.052
NPY (n = 956)	>2	11 (2.3)	12 (1.5)	<0.001
	Continuous	0 [0 to 0.03]	0 [0 to 0]	0.03
CMI[†] in multivariate analysis adjusted OR[‡] [95%CI]	>2	4.92 [2.79 to 0.68]		0.005
FIT limited to the Vatnimad subcohort (n = 468)	>150 ng[§]	45 (37.8)	20 (5.7)	<0.001
FIT limited to the Vatnimad subcohort (n = 468)	Continuous	61.66 [1 to 379]	1 [0 to 9.67	<0.001
FIT[†] in multivariate analysis adjusted OR[‡] [95%CI]		Reference		
Negative				
Positive				
Not done		8.69 [4.66 to 16.21]		<0.001
		0.50 [0.33 to 0.75]		0.24

The cohort included 981 individuals from 2 different cohorts, of whom 468 were asymptomatic individuals enrolled via a mass CRC-screening program and 513 were symptomatic patients. Invasive carcinoma or carcinoma in situ of any aspect in the rectum or colon = CRC; controls = normal colonoscopy.
[†]Student's t test and χ^2 test for continuous and binary variables, respectively.
[‡]Positivity cutoffs were 2.0 for the CMI and 150 ng/mL for the FIT; OR, odds ratio; 95%CI, 95% confidence interval.
[§]Adjusted for age, gender, and all variables listed in the table.
[¶]Concentration of hemoglobin for FIT is given per device.

reach significance. By keeping specificity of both tests superior to 95%, the negative predictive value (mean, extremes) of CMI (set at >2) test was better than FIT (set at >150 ng/mL): 84.09 [81.5 to 86.4] and 81.64 [77.5 to 85.3], respectively (*SI Appendix*, Table S8). A multivariate model including all individuals with and without neoplasia adjusted for age, gender, and FIT results showed CMI > 2 was independently associated with CRC (Table 1). Regarding CMI test and keeping with threshold for positivity set at ≥2, the sensitivity for the detection of CRC could reach 37% in the validation study.

After stratification on Tumor-Nodes-Metastasis classification from the American Committee on Cancer (*SI Appendix*, Table S9), the CMI value levels increased significantly and linearly with increasing tumor stages.

Dysbiosis and Associations with Methylation of Genes. The 16S rRNA gene sequencing on stool samples (n = 513) indicated a great β-diversity in patients with CRC and in those with a CMI > 2 (P < 0.05). Furthermore, CMI levels were significantly correlated with dysbiosis. In the validation cohort, we hierarchically clustered bacterial phylotypes on the genus level based on the similarity of their dynamics in patients with CRC and in those with CMI > 2 (Fig. 4).

Interestingly, principal coordinate analysis on the microbiota when incorporating the diagnosis and the level of blood methylation test (Fig. 4*A*), separated into 2 main clusters which differed significantly between participants with versus without a CMI > 2, regardless of whether they did or did not have CRC (Fig. 4*B*). A single cluster differed significantly in abundance between controls with and without a CMI > 2 (P < 0.05, Wilcoxon signed-rank test; Fig. 4*C*).

Although microbiota composition varied across cohorts, several genera such as *Parvimonas* and *Parasutterella* were more abundant in CRC patients in both cohorts, whereas *Eubacterium* was more abundant in controls (Table 2 and *SI Appendix*, Table S10). *Parvimonas* genus was also more abundant in individuals with blood CMI > 2 vs. ≤2 in the current as well as in our pilot study. The whole metagenomic analysis showed 20 bacteria species, including several *Parvimonas* species, differed in abundance in patients with CMI > 2 (n = 53) and in those with CMI ≤ 2 (n = 90) (Table 3).

Discussion

The data reported here constitute evidence that the relative abundance of some bacterial taxonomic groups within the microbiota in CRC is significantly associated with methylation/demethylation of host genes. In our study, we show clear phenotypical differences between mice receiving fecal transplants from CRC patients vs. healthy controls. The CRC-μ group presented with lower fecal SCFA concentrations and had significant colonic mucosal changes including higher ACF numbers and marked epigenetic alterations independent of AOM. However, 2 findings in the mouse study indicated an additive effect of AOM to the CRC microbiota: 1) the number of DNA alterations was greater in colonic than in spleen tissues under AOM (Fig. 2 *D* and *G*) and 2) the number of spleen tissue DNA alterations was similar in the CRC-μ and N-μ recipients. Interestingly, 16S rRNA gene sequencing of fecal microbiota from CRC patients and normal controls showed that CRC-associated dysbiosis was

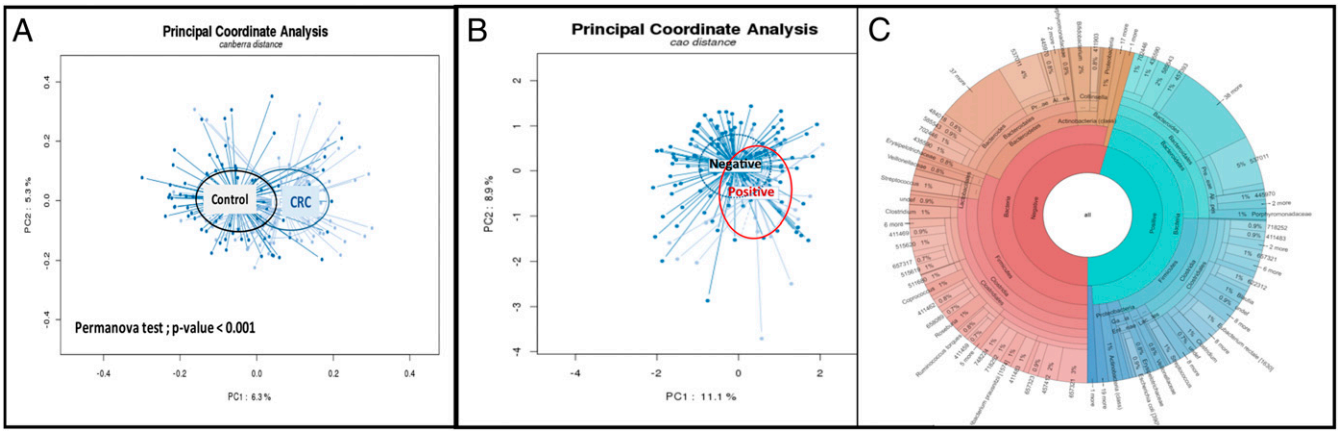

Fig. 4. Distribution of bacteria in fecal microbiota from patients with CRC and controls with normal colonoscopy findings. Overall, 348 individuals from CCR2 cohort (173 asymptomatic individuals enrolled via a mass CRC-screening program and 165 patients from Vatnimad and symptomatic subcohorts, respectively) enrolled. Invasive carcinoma, carcinoma in situ, or specific carcinoma either on flat mucosa or within a polyp in the rectum or colon were defined as CRC (n = 177); controls had no malignancy or significant polyp visible by full colonoscopy (n = 171). (*A*) Pattern of microbiota clustering according to the diagnosis as assessed by principal coordinate analysis. The genus-level analysis based on distance matrix variances showed significant differences between CRC patients and controls. Fecal DNA was subjected to metagenomic sequencing of the conserved V3 to V4 region of the 16S rRNA gene. The amplicons were purified, quantified, and pooled and then sequenced on an Illumina MiSeq platform. For the analysis of 16S rRNA gene sequences, raw MiSeq FASTQ files were demultiplexed, quality-filtered using Trimmomatic, and merged. Taxonomic assignations were performed using Qiime2 (default parameters) with the SILVA-123 database. The statistical analysis was done with MetagenomeSeq (36). (*B*) Pattern of microbiota clustering according to the blood methylation test as assessed by principal coordinate analysis. Analysis of variance using distance matrices on 789 OTUs (metagenomeSeq_1.16.0) from 362 individuals (175 with normal colonoscopy findings and 187 with advanced neoplasia) demonstrated a significant difference between the groups with positive and negative blood CMI values (>2 and ≤2, respectively). (*C*) Distribution of genera in fecal microbiota in the groups with positive and negative blood CMI values (>2 and ≤2, respectively) (Shaman c3bi platform; Institut Pasteur, http://shaman.c3bi.pasteur.fr/). A maximum likelihood phylogenetic tree was tested (*Individuals, Materials, and Methods*). Brown to red colors indicate negative CMI results (≤2), and blue colors indicate positive CMI results (>2). Note that diversity was less in the group with a positive CMI compared to the group with a negative CMI.

significantly associated with hypermethylation of several gene promoters, including *NPY* and *PENK* from the brain gut system and *Wif1* from the *Wnt* pathway. A CMI test constructed from these genes and performed in blood detected more (vs. controls) CRC patients even in asymptomatic individuals (n = 32) who were recruited through mass screening program and had early CRC at colonoscopy. However, the CMI showed higher sensitivity rate in the pilot study likely due to advanced CRCs being at stage III or IV in this series (*SI Appendix*, Table S9). Indeed, CRCs in the validation study were consistently at an early stage (0, I, or II). Nevertheless, more than 50% of CRC patients with symptoms and more than 35% of asymptomatic individuals presenting with an early CRC in the validation study showed a CMI ≥ 2 contrasting with only 4% in individuals with normal colonoscopy.

In keeping with earlier reports and with 2 very recent meta-analyses (10, 13–16), microbiota from CRC patients in the current study contained higher proportions of *Fusobacterium*, *Parvimonas*, *Butyrivibrio*, *Gemella*, *Fusobacteria*, and *Akkermansia* contrasting with lower proportions of *Ruminococcus*, *Bifidobacterium*, *Eubacteria*, and *Lachnospira*, compared to human control microbiota. Despite some early changes after microbiota transfer in germ-free mice, possibly due to the FMT itself as reported (17), most of the human microbiota components subsequently remained stable over time. The mouse fecal bacteria that showed the greatest decline over time (*SI Appendix*, Fig. S4) were *Clostridia* and *Clostridiales*, as analyzed at the class or the genus levels which might account for food uptake and weight differences among mouse groups. Numerous species belonging to this bacterial community have been associated with an increase in food intake and up-regulated production of proinflammatory molecules (18, 19). SCFAs, the main end-products of butyrate-producing bacteria (19), have been shown to inhibit intestinal inflammation and modulate immune responses (20), maintain barrier function (21), decrease precancerous lesions due to DNA damage (22–24), and regulate DNA methylation (25).

Of interest, the observed dysbiosis was correlated with the histological and DNA findings in the animals while the putative role of dysbiosis and inflammation in the initiation of cancer in the colonic mucosa remained unclear. We used AOM as a co-factor together with microbiota that could induce DNA alterations (22, 25). DNA alterations were more numerous in the colonic mucosa than in splenic tissue, effects which were abolished in the colon by PEG in CRC-μ + AOM recipients. Furthermore, transcripts of factors such as HES, KLF4, and ELF3 involved in the *Wnt* and *Notch* pathways were more abundant in CRC-μ than in N-μ mice recipients, depending on higher *Bacteroides* and lesser *Coprococcus* in mouse feces after human FMT (*SI Appendix*, Fig. S5). Thus, our findings suggest that histological alterations in the colonic mucosa might be due to an imbalance in microbiota composition with CRC-μ being associated with greater DNA damage and/or gene methylation/demethylation changes in the colonic mucosa. Although there was globally a greater number of unmethylated probes in the CRC-μ + AOM group than in the N-μ group, few genes (i.e., *Wif1*) might be methylated after FMT. This is suggested by the observed human CRC-μ induced DNA methylation/demethylation imbalance in mouse colonic mucosa with an additive effect of AOM. These findings are consistent with previous evidence of gene expression silencing by gene methylation in overall one third of human tumor tissues (5) and significant associations between abundance of *Parvimonas micra* and *Bacteroides fragilis* with highly methylated tumors (16). To evaluate this hypothesis, we confirmed that *Parvimonas micra* species was overabundant in the microbiota of those patients presenting a higher methylation gene index in the blood first by analyzing 16sRNA and then by using whole metagenome analysis for the confirmation at the species levels (Table 3). Thus, one would suggest that genera such as *Parvimonas* and *Proteobacteria* may use epigenetic pathways for adaptation to

Table 2. Differences in fecal microbiota in the validation CCR2 cohort

Vatnimad (stage 0, I, and II)			Symptomatic (stage III and IV CRC)		
Genus	Log_2 fold change	Adjusted *P* values	Genus	Log_2 fold change	Adjusted *P* values
Comparisons between controls and CRC patients according to tumor staging and cohorts					
Ruminococcus	−1.3	4.37E-02	—	—	—
Gemella	−1.2	4.61E-02	—	—	—
Parvimonas	−3.4	1.10E-09	*Parvimonas*	−2.0	1.16E-04
Parasutterella	−1.5	4.47E-02	*Parasutterella*	−1.4	1.72E-02
Mogibacterium	1.3	9.71E-03	*Mogibacterium*	1.2	5.41E-03
Butyrivibrio	1.0	2.18E-03	—	—	—
Lactonifactor	1.0	2.35E-02	*Megasphaera*	1.3	1.10E-02
Oscillospira	0.8	4.93E-02	*Olsenella*	1.6	1.64E-05
Howardella	1.1	4.16E-02	*Howardella*	1.1	1.10E-02
Abiotrophia	0.9	2.31E-02	*Abiotrophia*	0.9	1.03E-02
Eubacterium	2.5	1.49E-07	*Eubacterium*	2.3	4.64E-08
Acetitomaculum	0.8	4.16E-02	—	—	—
Ezakiella	1.1	2.76E-03	*Ezakiella*	0.9	1.03E-02
Comparison between individuals according to blood CMI values ≤ 2 (negative) versus >2 (positive)					
Coprococcus	1.1	4.717E-03	*Coprococcus*	1.26	4.0E-04
Gemella	−1.7	3.02E-02	*Dialister*	2.68	3.1E-04
Parvimonas	−2.3	4.10E-06	*Parvimonas*	−2.39	6.04E-06
Peptostreptococcus	−1.5	3.05E-5	*Peptostreptococcus*	−1.6	6.0E-04
Oxalobacter	−10.5	4.71E-04	*Fusobacterium*	−1.8	2.11E-03
Acidamicococcus	−1.9	3.05E-05	*Acidaminococcus*	−2.61	1.0E-04
			Mitsuokella	−2.2	5.24E-04
Howardella	1.9	3.06E-03	*Howardella*	1.1	1.10E-02
Eubacterium	2.5	1.49E-07	*Enterococcus*	2.7	5.9E-04
Acetitomaculum	0.8	4.16E-02	—	—	—
Shewanella	−11.2	1.13E-3	—	—	—
Phenylobacterium	11.4	3.71E-03	—	—	—

The 16sRNA metagenomics was assessed after stool DNA extraction and subjected to 16S rRNA sequencing on the Illumina Miseq platform. Only those bacteria detected in at least 20% of individuals are indicated. Log_2 is logarithmic value; *P* values are given after adjustment on age, gender, and BMI using Bonferroni-corrected Mann–Whitney *U* test. The cohort included 348 individuals of whom 173 were asymptomatic individuals enrolled via a mass CRC-screening program and 165 were symptomatic patients. Invasive carcinoma or carcinoma in situ of any aspect in the rectum or colon, CRC; controls, normal colonoscopy.

environmental factors (4, 26–30), with methylation/demethylation as a pivotal mechanism (31, 32).

In our cohorts, the *Prevotella* genus, a leading source of interindividual gut microbiota variation associated with long-term fiber intake (8), was more abundant in individuals with CMI ≤ 2. Of interest, *Parvimonas* and *Parasutterella* were the bacteria most closely associated with an animal-based diet and were related to a high fat intake (8), both of which probably result in greater bile acid release and higher enteric deoxycholic acid concentrations. These 2 bacteria coexclude antiinflammatory bacteria such as *Faecalibacterium* and *Eubacterium* as seen currently and reported by us and others (10, 12–15). Of interest was also *SFRP2* hypermethylation in tumor tissue and blood (*SI Appendix*, Fig. S10 *A* and *C*) which was associated with *Bilophila*, another proinflammatory genus (*SI Appendix*, Tables S3 and S10), and that coexcluded currently *Faecalibacterium* in the CRC microbiota. Increased abundance of the *Bilophila* genus was associated with *SFRP2* gene promoter demethylation after black raspberry supplementation (33) used as probiotics.

Our present results might be affected by various study limitations. First, in our prospective validation trial (NCT01270360), the blood CMI was determined at a single laboratory, and bias may therefore have occurred. The blood CMI was performed on serum and was negative in some CRC patients. Different results might have been obtained using plasma instead of serum. Thus, methylation was further investigated in random samples by performing a digital PCR test known to be more sensitive than routine qPCR as previously reported (34). Even with dPCR test, sensitivity was higher in the symptomatic CRC patients likely due to the more advanced CRC in symptomatic patients as compared to asymptomatic CRC individuals. This would suggest a blood CMI positivity test in CRC patients with symptoms was likely due to the presence of tumor cells in the bloodstream rather than to systemic plasticity of peripheral blood cells such as mononuclear cells. A second limitation of this study is that only 3 genes were incorporated in the blood CMI. Additional genes such as *SFRP2* might enhance sensitivity of the test and might yield different bacterial clustering patterns from those reported here. Furthermore, while viruses and CRC are still subject to discussion, we cannot rule out the hypothesis that in CRC patients, particular phages modulate the gut microbiota of patients, inducing a dysbiotic profile compared to healthy controls (35).

Nevertheless, our data present evidence for the association between a dysbiosis and CRC causing alterations in gene methylation. The cohort in which dysbiosis and gene methylation were investigated is the largest to date and strongly indicates a dysbiosis-induced imbalance in gene methylation and in bacterial species. We could identify bacteria species (Table 3) who were significantly associated with higher levels of methylation test (CMI > 2) by using whole metagenomic analysis. These measurements can now be proposed as markers for the effectiveness of and adherence to prebiotic and probiotic therapies.

Table 3. Species in the fecal microbiota in individuals with blood methylation test (negative versus positive)

Species	Mean at baseline	Fold change	Log_2 fold change	*P* value
Faecalibacterium prausnitzii [1574]	2,480.5	2.2e + 00	1.159	0.042
Ruminococcus sp. SR1/5 [1621]	2,022	2.5e + 00	1.354	0.004
Eubacterium hallii [1597]	1,942.6	2.2e + 00	1.14	0.045
Clostridium sp. L2-50 [1593]	1,306.6	3.8e + 00	1.945	0.004
Coprococcus comes [1616]	1,306.5	2.1e + 00	1.107	0.022
Dialister invisus [1259]	1,023	6.2e + 00	2.65	0.004
Coprococcus eutactus [1592]	941	2.9e + 00	1.569	0.042
Bacteroides eggerthii [1097]	938	2.9e − 01	−1.745	0.003
Ruminococcus obeum [1619]	714	2.3e + 00	1.246	0.002
Clostridium bolteae [1598]	478	3.1e − 01	−1.686	0.002
Bacteroides sp. D2 [1094]	367.5	3.8e − 01	−1.384	0.042
Enterococcus faecalis [1363]	240	6.6e + 00	2.723	0.042
Mitsuokella multacida [1653]	220	2.3e − 01	−2.12	0.003
Parvimonas micra [1505]	211	2.1e − 01	−2.22	0.003
Peptostreptococcus stomatis [1530]	110.5	2.0e − 01	−2.26	0.002
Veillonella atypica [1260]	84.9	1.9e − 01	−2.34	0.005
Streptococcus equinus [1381]	60.66	1.8e − 01	−2.42	0.042
Gemella morbillorum [1302]	53	1.3e − 01	−2.921	0.00004
Parvimonas sp. oral taxon 110 [1506]	35.6	2.7e − 01	−1.86	0.006
Parvimonas sp. oral taxon 393 [1507]	35	3.2e − 01	−1.622	0.026

Stool samples were submitted to whole metagenomic sequencing of fecal bacteria DNA (controls, $n = 61$; CRC patients, $n = 53$) according to CMI test, and results were generated using Shaman C3bi from Institut Pasteur de Paris (http://shaman.c3bi.pasteur.fr/) (for methods, see ref. 9). Log_2 is logarithmic value; *P* values are given after adjustment on age, gender, and BMI. The blood test defined as negative (CMI < 2) versus positive (CMI > 2) is considered to compare abundances of bacteria species in the stool milieu. Fold changes are indicated in log values with minus meaning the bacteria is higher in CRC patients' microbiota and plus meaning that the bacteria abundance was higher in controls.

MICROBIOLOGY

Individuals, Materials, and Methods

Recruitment of Participants and Collection of Samples. Patients referred to University hospitals for colonoscopy were enrolled in several prospective cohorts. Effluents and tissues from participants were used for experimental, proof-of-concept, and a pilot study that included 266 individuals from a cohort named CCR1 (*SI Appendix*, Table S6). A second cohort, named CCR2 (*SI Appendix*, Table S7), was constituted for validation; this was composed of 2 subcohorts: one including only symptomatic (named Valihybritest study) and the second including only asymptomatic individuals (named Vatnimad) recruited from mass screening programs. All individuals underwent colonoscopy due to symptoms or due to a positive fecal blood test (FOBT). Both subcohort studies were registered on ClinicalTrials.gov (NCT01270360), and 1,000 consecutive colonoscopies were to be enrolled, among them 500 individuals (Vatnimad) without any digestive symptoms. The study protocol was approved by the ethics committee of Comité de Protection des Personnes Paris Est-Henri Mondor (no. 10-006 in 2010). All participants signed an informed consent.

Exclusion criteria for these cohorts were a history of colorectal surgery due to CRC, familial adenomatous polyposis, Lynch syndrome, infection, and inflammatory bowel disease and exposure to antibiotics during the 3 wk preceding the colonoscopy. Asymptomatic individuals had an FOBT (10) or an FIT before colonoscopy as enrolled in the pilot or in the validation (Vatnimad) studies.

Studies of Microbiota from Human Participants.

The 16S rRNA gene and whole metagenomic sequencing on stools. Stool samples were collected in donors and patients during experimental and validations studies, respectively; samples of stool were collected and stored within 4 h for DNA extraction using the GNOME DNA Isolation Kit (MP Biomedicals) as previously described (9, 10, 12). After amplification by PCR of the V3 to V4 region of the 16S rRNA gene, sequencing was performed using a 250-bp paired-end sequencing protocol on the Illumina MiSeq platform. Raw FASTQ files were demultiplexed, quality-filtered using Trimmomatic (sliding windows of 2 with a quality score of 20), and merged using fastq-join from ea-utils (https://expressionanalysis.github.io/ea-utils). Taxonomic assignations were performed using Qiime2 (no quality filtering; default parameters) (36) with the SILVA-123 database; OTU were constructed using UCLUST (threshold of 97% of similarity), Chimera Slayer for chimera removing, and SILVA 16S rRNA database (version 123) for taxonomical assignation. The intergroup high similarity and intragroup low similarity of microbiota were assessed by β-diversity, PCoA (generated by Qiime using unweighted unifrac metrics). We subjected study populations to 2 principle coordinates analyses, independently of other datasets, and we investigated separation of CRC microbiota from control microbiota in donors in the experimental study as well as in controls versus CRC patients in the clinical trials. All microbiome statistical tests where produced using the MetagenomeSeq packages or the Shaman Webserver that used the DESeq2 packages for differential expression analyses; therefore, every statistical analysis for microbiome have been corrected according to gender, age, and BMI and adjusted for multiple testing.

For a deep identification of bacteria species in the current study, DNA samples were submitted to whole metagenomic analyses as previously described (10).

Studies of Methylation from Human Participants.

DNA isolation and bisulfite conversion. DNAs from colonic tissues (QIAamp DNA Mini Kit; Qiagen), blood, and stool samples (QiAamp DNA stool Mini Kit; Qiagen) were extracted using the ZR Serum DNA kit (Ozyme) according to the manufacturer's protocol. The DNA samples were then exposed to sodium bisulfite at 50 °C in the dark for 16 h (EZ DNA Methylation kit; Zymo Research) to convert unmethylated cytosine nucleotides into uracil nucleotides (subsequently converted to thymidine nucleotides during PCR cycling) without changing the methylated cytosines as detailed elsewhere (10, 12).

DNA methylation using Illumina Golden Gate methylation bead arrays. The GoldenGate Methylation Cancer Panel I (Illumina) was used to probe 500 ng of each bisulfite-converted DNA sample of human ($n = 18$, 9 CRC patients and 9 controls) including tissue ($n = 18$), stool ($n = 18$), and blood ($n = 18$); the stool samples were used for FMT to germ-free mice (*SI Appendix*, Table S11). Methylation levels ranged from 0 to 100.0% were used for the calculation of the ratio of the methylated signal intensity. The strategy for hierarchical clustering of gene candidates is further described in Fig. 3 and *SI Appendix*, Figs. S8–S11. Briefly, comparisons across tissue, stool, and blood samples identified genes with CpG loci methylation levels in the promotor above the expected number. The CMI was computed by addition of the methylation values of the 3 genes generated by the discovery study (characteristics of individuals in the pilot study in *SI Appendix*, Table S12). Primers targeting all genes (including albumin gene-ALB, devoid of CpG sites and used as a

housekeeping gene and for normalization of DNA amounts) are reported elsewhere (*SI Appendix*, Table S13).

Quantitative methylation-specific PCR amplification and verification of the specificity. Bisulfite-converted universal human methylated DNA standard (Zymo Research) served as a calibrator and positive control and DNA-free distilled water as a negative control. Quantitative, single-gene methylation-specific PCR (QS-MSP) and quantitative multiplex methylation-specific PCR (QM-MSP) were applied. The relative methylation level was determined using the $2^{-\Delta\Delta Ct}$ formula. Briefly, for each PCR run, a KAPA PROBE master mix (Kapa Biosystems) was prepared with predefined concentrations of genes candidates.

Statistical analyses of blood CMI and FIT data in the validation cohort. With the type I error set at 0.05 and assuming that CRC would be found in 8% of asymptomatic and 25% of symptomatic individuals, to detect at least 60% of CRCs and 20% of polyps with 90% specificity in average-risk (asymptomatic and aged ≥50 y) and high-risk (history of polyps or sibling with CRC) individuals with 80% power, 1,000 participants (with at least 400 in each subcohort) were needed. Characteristics of study populations were described using number (%) for qualitative variables and mean ± SD for quantitative variables.

The main endpoint was identifying those individuals with advanced neoplasia (invasive carcinoma or carcinoma in situ or specific carcinoma on any aspect in the rectum or colon, roughly called CRC). Demographic, clinical, blood methylation, and FIT data were compared between patients with and without CRC using Pearson's χ^2 test for binary variables and Student's *t* test for continuous variables. Blood methylation data (CMI and methylation of each of its 3 components, *Wif1*, *PENK*, and *NPY*) and FIT data were handled as both binary and continuous variables. The following values were predefined to classify tests as positive: >2 for the CMI and >150 ng per device for the FIT.

To determine whether CMI and FIT were associated with CRC, we built a multivariate logistic regression model as described (37) adjusted for age, gender, and BMI and the adjusted odds ratios (ORs) and hazard ratios with their 95% confidence intervals (95%CIs) were computed. All tests were 2-tailed, and *P* values < 0.05 were deemed significant. The statistical analysis software was Stata SE v15.0.

Studies from Experiments in Mice. Fresh stool samples were obtained from 9 females and 9 males in CCR1 cohort, for FMT to germ-free mice. They were 9 with CRC and 9 with normal colonoscopy (*SI Appendix*, Table S11).

FMT experiments. FMT was performed in male C3H/HeN germ-free 8-wk-old mice (design of the experimental study in *SI Appendix*, Fig. S1). Mice were maintained in gnotobiotic isolators, ad libitum for 1 wk of acclimation to the laboratory conditions followed by FMT (day 0) then by 7 or 14 wk of follow-up, after which the mice were killed. The fecal microbiota donors (*SI Appendix*, Table S11) were 9 consecutive individuals with normal and 9 consecutive patients with CRC at colonoscopy from cohorts. They were considered in the current study as normal microbiota (N-μ) and CRC microbiota (CRC-μ) donors, were informed about the experimental study, and accepted giving additional stools, if needed, for animal experimentation during the study period. Fresh stools were given by oral gavage, as follows: $n = 53$ in 7-wk study (CRC-μ transfer, $n = 30$, and N-μ transfer, $n = 23$) and $n = 132$ in 14-wk study (CRC-μ transfer, $n = 66$, and N-μ transfer, $n = 66$). In addition to the FMT, the mice were given i.p. injections of either the carcinogen AOM (Sigma; 8 mg/kg body weight once a week for 3 or 10 wk), here chosen as a potential environmental exposure, or saline. In the 14-wk study, 24 animals were given also polyethylene glycol (PEG3350) at a nonlaxative dosage (1%) in their drinking water to impede adherence of bacteria to the gut mucosa.

Associations between mice colonic tissue events, type of transferred microbiota (CRC-μ vs. N-μ), chemical product given (AOM vs. saline), and bacterial composition in stools were assessed to evaluate potential associations by using Spearman's tests. The experimental procedures were approved by the local ethics committee (committee reference no. 45; approval no. 12/076).

Cytokine and chemokine and cell signal measurements in mouse colonic mucosa in vivo. cDNA synthesis from total RNA extracted from colon mucosa scrapings of mice was quantified using SYBR Green PCR Master Mix cDNA with the Light Cycler 480 System (Roche Diagnostics) according to the manufacturer's instructions. The mRNA levels of each group ($n = 3$) were quantified, and transcripts involved in cell renewal and those of involved in inflammation (TNFα, MIP2, IL10, IL1β, IL6, IL17, and IL17R) were analyzed. Three housekeeping genes (GADPH, HRPT1, and TBP) were used. The comparative ΔΔCT method was used for relative mRNA quantification of target genes, normalized to GAPDH and a relevant control equal to 2-ΔΔCT. Relative levels of genes involved in cell renewal (*Kfl4*, *Elf3*, *MATH1*, and *Hes1*) or encoding proinflammatory cytokines and chemokines (TNFα, MIP2, IL10, IL1β, IL6, IL17, and IL17R) were analyzed using quantitative real-time PCR as previously described (38).

Data Availability. The shotgun metagenomic sequencing data and the 16S rRNA amplicon sequencing data are available from the European Nucleotide Archive (ENA) database (http://www.ebi.ac.uk/ena) under the accession number ERP005534. Data related to this paper are available from the European Nucleotide Archive (ENA) database (http://www.ebi.ac.uk/ena) under the accession nos. ERX3622297–ERX3622402, ERR3628499–ERR3628604, ERS3936180–ERS3936285, and PRJEB35144.

ACKNOWLEDGMENTS. We thank all patients for their participation and physicians who invited them to participate; they are the following gastroenterologists: Drs. E. Zrihen, O. Pecriaux, J. Samama, M. Petit, Ph. Cattan, M. Cavicchi, Ch. Locher, G. Gattineau, M. Parieto, M. Mozer, A. Rosenbaum, Ph. Capelle, D. Levoir, F. Maille, Ph. Lebourgeois, Ph. De Land, E. Chanteloup, M. Simon, F. Mal, and F. Iglicki. We thank also Drs. J. Tran VanHieu and M. L. Auriault for pathology analyses in human and animals; Prof. S. Loric for biochemistry analysis in mice; Prof. T. Simon, A. Touati, J. Tap, V. Jarrousse, A. Bado, and J. P. Fouret for animal experiments and help managing; A. Wolfe for revising the English; S. Peyvandi for assisting with the animal and molecular experiments; E. Guery and L. Segaux for their statistical contributions; C. Vialette for data managing; A. Caidia (Bioinformatics Core Laboratory) for 16S rRNA analysis; and Catherine Philippe for SCFA analyses in mice. We thank all technicians and scientific consultants from Faculté de Médecine site Pitié Salpêtrière (Assistance Publique-Hôpitaux de Paris [APHP]), Unité Mixte de Service (UMS) 29 Omic Platform p3s for technical help on EPIC methylation array study in mice, and all members of clinical monitoring group from the Unité de Recherche Clinique de l'Est Parisien (URC-Est) Hôpital Saint-Antoine, APHP. We thank all the technicians from the Anaxem germ-free animal facility of the Micalis Institute for breeding the germ-free mice and carrying out FMT and gnotobiotic mice care and monitoring. We thank Dr. Abdulmohammad Pezeshki and Ms. Shatha Awaad for assistance with the analysis of inflammation in mouse colons and Katja Brunner for editing the manuscript. Funding was provided by French Institut of Cancer and Ministry of Health (grant PHRC 2011-VatnimadAOM09268) and French Society of Gastroenterology (grant for fecal test screening), Ligue Nationale Contre le Cancer for fecal test screening and the Institut National du CAncer (INCA, Cancéropôle Ile de France, Grant for Microbiota and CRC), and National Institute of Health and Medical Research (INSERM) partially financed under the Institut Thématique Multi-Organisme program (ITMO). This work was sponsored by APHP, INSERM. ClinicalTrials.gov registration number is NCT 01270360.

1. K. Z. Guyton *et al.*; International Agency for Research on Cancer Monograph Working Group, IARC, Lyon, France, Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate. *Lancet Oncol.* **16**, 490–491 (2015).
2. G. López-Abente, J. García-Pérez, P. Fernández-Navarro, E. Boldo, R. Ramis, Colorectal cancer mortality and industrial pollution in Spain. *BMC Public Health* **12**, 589 (2012).
3. C. P. Wild, Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **14**, 1847–1850 (2005).
4. A. Ghantous, H. Hernandez-Vargas, G. Byrnes, T. Dwyer, Z. Herceg, Characterising the epigenome as a key component of the fetal exposome in evaluating in utero exposures and childhood cancer risk. *Mutagenesis* **30**, 733–742 (2015).
5. J. Guinney *et al.*, The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
6. S. J. O'Keefe *et al.*, Fat, fibre and cancer risk in African Americans and rural Africans. *Nat. Commun.* **6**, 6342 (2015).
7. E. D. Sonnenburg *et al.*, Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016).
8. L. A. David *et al.*, Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
9. I. Sobhani *et al.*, Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS One* **6**, e16393 (2011).
10. G. Zeller *et al.*, Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10**, 766 (2014).
11. C. M. Dejea *et al.*, Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592–597 (2018).
12. A. Amiot *et al.*; CRC group of Val De Marne, The detection of the methylated Wif-1 gene is more accurate than a fecal occult blood test for colorectal cancer screening. *PLoS One* **9**, e99233 (2014).
13. S. H. Wong *et al.*, Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. *Gastroenterology* **153**, 1621–1633.e6 (2017).
14. A. M. Thomas *et al.*, Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* **25**, 667–678 (2019).

15. J. Wirbel *et al*., Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
16. R. V. Purcell, M. Visnovska, P. J. Biggs, S. Schmeier, F. A. Frizelle, Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Sci. Rep.* **7**, 11590 (2017).
17. S. I. Halkjær *et al*., Faecal microbiota transplantation alters gut microbiota in patients with irritable bowel syndrome: Results from a randomised, double-blind placebo-controlled study. *Gut* **67**, 2107–2115 (2018).
18. V. C. Lombardi *et al*., Nutritional modulation of the intestinal microbiota; future opportunities for the prevention and treatment of neuroimmune and neuroinflammatory disease. *J. Nutr. Biochem.* **61**, 1–16 (2018).
19. T. O. Keku, S. Dulal, A. Deveaux, B. Jovov, X. Han, The gastrointestinal microbiota and colorectal cancer. *Am. J. Physiol. Gastrointest. Liver Physiol.* **308**, G351–G363 (2015).
20. R. F. Schwabe, C. Jobin, The microbiome and cancer. *Nat. Rev. Cancer* **13**, 800–812 (2013).
21. K. B. Islam *et al*., Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology* **141**, 1773–1781 (2011).
22. P. Rosignoli *et al*., Protective activity of butyrate on hydrogen peroxide-induced DNA damage in isolated human colonocytes and HT29 tumour cells. *Carcinogenesis* **22**, 1675–1680 (2001).
23. J. Winter *et al*., Inhibition by resistant starch of red meat-induced promutagenic adducts in mouse colon. *Cancer Prev. Res. (Phila.)* **4**, 1920–1928 (2011).
24. D. L. Worthley *et al*., DNA methylation in the rectal mucosa is associated with crypt proliferation and fecal short-chain fatty acids. *Dig. Dis. Sci.* **56**, 387–396 (2011).
25. J. Chen, X. F. Huang, The signal pathways in azoxymethane-induced colon cancer and preventive implications. *Cancer Biol. Ther.* **8**, 1313–1317 (2009).
26. S. Wu *et al*., A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat. Med.* **15**, 1016–1022 (2009).
27. I. P. Pogribny, I. Rusyn, Environmental toxicants, epigenetics, and cancer. *Adv. Exp. Med. Biol.* **754**, 215–232 (2013).
28. L. Hou *et al*., Ambient PM exposure and DNA methylation in tumor suppressor genes: A cross-sectional study. *Part. Fibre Toxicol.* **8**, 25 (2011).
29. M. J. Blow *et al*., The epigenomic landscape of prokaryotes. *PLoS Genet.* **12**, e1005854 (2016).
30. Y. D. Cho *et al*., Transcriptomics and methylomics in chronic periodontitis with tobacco use: A pilot study. *Clin. Epigenetics* **9**, 81 (2017).
31. R. Feil, M. F. Fraga, Epigenetics and the environment: Emerging patterns and implications. *Nat. Rev. Genet.* **13**, 97–109 (2012).
32. K. J. Howell *et al*., DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. *Gastroenterology* **154**, 585–598 (2018).
33. L. Chen *et al*., Chemoprevention of colorectal cancer by black raspberry anthocyanins involved the modulation of gut microbiota and SFRP2 demethylation. *Carcinogenesis* **39**, 471–481 (2018).
34. S. Garrigou *et al*., A study of hypermethylated circulating tumor DNA as a universal colorectal cancer biomarker. *Clin. Chem.* **62**, 1129–1139 (2016).
35. G. D. Hannigan, M. B. Duhaime, M. T. Ruffin, 4th, C. C. Koumpouras, P. D. Schloss, Diagnostic potential and interactive dynamics of the colorectal cancer virome. *MBio* **9**, e02248-18 (2018).
36. J. G. Caporaso *et al*., QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
37. Y. Huang, M. S. Pepe, Z. Feng, Logistic regression analysis with standardized markers. *Ann. Appl. Stat.*, **7** (2013).
38. M. Abolhassani *et al*., Leptin receptor-related immune response in colorectal tumors: The role of colonocytes and interleukin-8. *Cancer Res.* **68**, 9423–9432 (2008).
39. M. Needhamsen *et al*., Usability of human Infinium MethylationEPIC BeadChip for mouse DNA methylation studies. *BMC Bioinformatics* **18**, 486 (2017).