Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures

Iradj Sobhani\textsuperscript{a,1}, Emma Bergsten\textsuperscript{a,3}, Séverine Couffin\textsuperscript{a}, Aurélien Amiot\textsuperscript{a,b}, Biba Nebbad\textsuperscript{a}, Caroline Barau\textsuperscript{a}, Nicola de’Angelis\textsuperscript{a}, Sylvie Rabot\textsuperscript{a}, Florence Canou-Poitrins\textsuperscript{b}, Denis Mestivier\textsuperscript{a,1}, Thierry Pédron\textsuperscript{b}, Khashayarsha Khazaie\textsuperscript{a,2}, and Philippe J. Sansonetti\textsuperscript{c,1,2}

\textsuperscript{a}EA7375 (ECM3 Research Team), Université Paris Est, Créteil 94000, France; \textsuperscript{b}Service de Gastroenterologie, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; \textsuperscript{c}Unité de Pathogénie Microbienne Moléculaire, INSERM U1202, Institut Pasteur, Paris 75015, France; \textsuperscript{d}Service de Microbiologie, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; \textsuperscript{e}Plateforme de Ressources Biologique, Hôpital Henri Mondor, Créteil 94000, France; \textsuperscript{f}Service de Chirurgie Digestive, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; \textsuperscript{g}Micalis Institute, Institut National de la Recherche Agronomique (INRA), AgroParisTech, Université Paris-Saclay, Jouy-en-Josas 78352, France; \textsuperscript{h}Service de Santé Publique, Hôpital Henri Mondor, Assistance Publique-Hôpitaux de Paris, Créteil 94000, France; \textsuperscript{i}Bioinformatics Core Lab, INSERM U955, Institut Mondor de Recherche Biomédicale, Créteil 94000, France; \textsuperscript{j}Department of Immunology, Mayo Clinic, Rochester, MN 55905; and \textsuperscript{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, 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 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 immunocompositional 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.

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 value of CMIs as CRC markers.

\textbf{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.


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


Published under the PNAS license.

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. ERX322297–ERX322402, ERX328499–ERX328604, ERX336180–ERX336285, and PRJEB35144.

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

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.

www.pnas.org/cgi/doi/10.1073/pnas.1912129116
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 S1A 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. The number of aberrant crypt foci (ACF), i.e., clusters of enlarged crypts visualized by methylene blue at low magnification, were enumerated. Whereas histological features appeared normal in the groups given CRC alone or CRC alone + AOM, or both (Fig. 1 A–D, H, and J), 7 after 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 J).

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. 1G). At 7 wk following FMT, a significant increase of proliferation was observed in the groups given AOM as compared to all other groups (Fig. 1H). 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. 1J). 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. 1K).

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 Clostridium, 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 antimicrobial effect (i.e., Faecalibacterium 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 two 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 the microbiota. The principal Component Analysis (PCA) scatter diagrams separated...
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. 2D). 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. 2C 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. 2E), 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. 2F). The number of methylated probes (>0.8) varied with N-μ, CRC-μ,
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. 2G). 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. 2H). 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. 3A). In order to develop a quick and easy methylation test, panels of genes were selected, based on the difference between normal CRC-μ + AOM group being silenced vs. N-μ group. 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. 3B 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
reach significance. By keeping specificity of both tests superior to 95%, the negative predictive value (mean, extremes) of CMI (set at ≥ 2) 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 methylation test (Fig. 4A), 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. 4B). A single cluster differed significantly in abundance between controls with and without a CMI > 2 (P < 0.05, Wilcoxon signed-rank test; Fig. 4C).

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. 2D 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...
Materials and Methods

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 (173 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.

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 Vietnam 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 (173 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.
Of interest was also \( Eubacterium \) as seen currently and reported by us and others performing a digital PCR test known to be more sensitive than methylation was further investigated in random samples by 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 hypothesis that in CRC patients, particular phages modulate the gut microbiota of patients, inducing a dysbiotic profile compared to healthy controls (35).

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

<table>
<thead>
<tr>
<th>Genus</th>
<th>Log2 fold change</th>
<th>Adjusted ( P ) values</th>
<th>Genus</th>
<th>Log2 fold change</th>
<th>Adjusted ( P ) values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruminococcus</td>
<td>(-1.3)</td>
<td>4.37E-02</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gemella</td>
<td>(-1.2)</td>
<td>4.61E-02</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Parvimonas</td>
<td>(-3.4)</td>
<td>1.10E-09</td>
<td>Parvimonas</td>
<td>(-2.0)</td>
<td>1.16E-04</td>
</tr>
<tr>
<td>Parasutterella</td>
<td>(-1.5)</td>
<td>4.47E-02</td>
<td>Parasutterella</td>
<td>(-1.4)</td>
<td>1.72E-02</td>
</tr>
<tr>
<td>Mogibacterium</td>
<td>(1.3)</td>
<td>9.71E-03</td>
<td>Mogibacterium</td>
<td>(1.2)</td>
<td>5.41E-03</td>
</tr>
<tr>
<td>B. Virribrio</td>
<td>(1.0)</td>
<td>2.18E-03</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lactonifactor</td>
<td>(1.0)</td>
<td>2.35E-02</td>
<td>Megasphaera</td>
<td>(1.3)</td>
<td>1.10E-02</td>
</tr>
<tr>
<td>Oscillospora</td>
<td>(0.8)</td>
<td>4.93E-02</td>
<td>Olesenella</td>
<td>(1.6)</td>
<td>1.64E-05</td>
</tr>
<tr>
<td>Howardella</td>
<td>(1.1)</td>
<td>4.16E-02</td>
<td>Howardella</td>
<td>(1.1)</td>
<td>1.10E-02</td>
</tr>
<tr>
<td>Abiotrophia</td>
<td>(0.9)</td>
<td>2.31E-02</td>
<td>Abiotrophia</td>
<td>(0.9)</td>
<td>1.03E-02</td>
</tr>
<tr>
<td>Eubacterium</td>
<td>(2.5)</td>
<td>1.49E-07</td>
<td>Eubacterium</td>
<td>(2.3)</td>
<td>4.64E-08</td>
</tr>
<tr>
<td>Acetitomaculum</td>
<td>(0.8)</td>
<td>4.16E-02</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ezakiella</td>
<td>(1.1)</td>
<td>2.76E-03</td>
<td>Ezakiella</td>
<td>(0.9)</td>
<td>1.03E-02</td>
</tr>
</tbody>
</table>

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. Log2 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 \( \leq 2 \). 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 coexcluded 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, 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).
proof-of-concept, and a pilot study that included 266 individuals from University hospitals for colonoscopy were enrolled in several prospective studies. DNA extraction using the GNOME DNA Isolation Kit (MP Biomedicals) as described (10) or an FIT before colonoscopy as enrolled in the pilot or in the validation group. Additionally, 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). Log2 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.

<table>
<thead>
<tr>
<th>Species</th>
<th>Mean at baseline</th>
<th>Fold change</th>
<th>Log2 fold change</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faecalibacterium prausnitzii [1574]</td>
<td>2,480.5</td>
<td>2.2e + 00</td>
<td>1,159</td>
<td>0.042</td>
</tr>
<tr>
<td>Ruminococcus sp. 5R1/5 [1621]</td>
<td>2,325</td>
<td>2.5e + 00</td>
<td>1.839</td>
<td>0.004</td>
</tr>
<tr>
<td>Eubacterium hallii [1597]</td>
<td>1,942.6</td>
<td>2.2e + 00</td>
<td>1.14</td>
<td>0.045</td>
</tr>
<tr>
<td>Clostridium sp. L2-50 [1593]</td>
<td>1,306.6</td>
<td>3.8e + 00</td>
<td>1.954</td>
<td>0.004</td>
</tr>
<tr>
<td>Coprococcus comes [1616]</td>
<td>1,306.5</td>
<td>2.1e + 00</td>
<td>1.107</td>
<td>0.022</td>
</tr>
<tr>
<td>Dialister invisus [1259]</td>
<td>1,023</td>
<td>6.2e + 00</td>
<td>2.65</td>
<td>0.004</td>
</tr>
<tr>
<td>Coprococcus eutactus [1592]</td>
<td>941</td>
<td>2.9e + 00</td>
<td>1.569</td>
<td>0.042</td>
</tr>
<tr>
<td>Bacteroides eggertii [1097]</td>
<td>938</td>
<td>2.9e − 01</td>
<td>−1.745</td>
<td>0.003</td>
</tr>
<tr>
<td>Ruminococcus obeum [1619]</td>
<td>714</td>
<td>2.3e + 00</td>
<td>1.246</td>
<td>0.002</td>
</tr>
<tr>
<td>Clostridium bolteae [1598]</td>
<td>478</td>
<td>3.1e − 01</td>
<td>−1.686</td>
<td>0.002</td>
</tr>
<tr>
<td>Bacteroides sp. D2 [1094]</td>
<td>367.5</td>
<td>3.8e − 01</td>
<td>−1.384</td>
<td>0.042</td>
</tr>
<tr>
<td>Enterococcus faecalis [1363]</td>
<td>240</td>
<td>6.6e + 00</td>
<td>2.723</td>
<td>0.042</td>
</tr>
<tr>
<td>Mitsuokella multacida [1653]</td>
<td>220</td>
<td>2.3e − 01</td>
<td>−2.12</td>
<td>0.003</td>
</tr>
<tr>
<td>Peptostreptococcus stomatist [1530]</td>
<td>110.5</td>
<td>2.0e − 01</td>
<td>−2.26</td>
<td>0.002</td>
</tr>
<tr>
<td>Veillonella atypica [1260]</td>
<td>84.9</td>
<td>1.9e − 01</td>
<td>−2.34</td>
<td>0.005</td>
</tr>
<tr>
<td>Streptococcus equinus [1381]</td>
<td>60.66</td>
<td>1.8e − 01</td>
<td>−2.42</td>
<td>0.042</td>
</tr>
<tr>
<td>Gemella morbillorum [1302]</td>
<td>53</td>
<td>1.3e − 01</td>
<td>−2.921</td>
<td>0.00004</td>
</tr>
<tr>
<td>Parvimonas sp. oral taxon 110 [1506]</td>
<td>35.6</td>
<td>2.7e − 01</td>
<td>−1.656</td>
<td>0.006</td>
</tr>
<tr>
<td>Parvimonas sp. oral taxon 393 [1507]</td>
<td>35</td>
<td>3.2e − 01</td>
<td>−1.622</td>
<td>0.026</td>
</tr>
</tbody>
</table>

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). Log2 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.

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 fasta-join from ea-utils (https://expressionanalysis.github.io/ea-utils). Taxonomic assignments 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 MetagenomSeq 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 Illuma 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 promoter 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 S12). 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 (Q5-MSAP) and quantitative multiplex methylation-specific PCR (QM-MSAP) 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 sub-cohort) 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 \( \chi^2 \) test for binary variables and Student's \( t \) test for continuous variables. Blood methylation data (CMI and methylation of each of its 3 components, \( \text{WIT}, \text{PENK}, \text{and NPY} \)) 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, 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 FIT 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 C57B6N 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-M) and CRC microbiota (CRC-M) donors, were informed about the experimental study, and accepted giving additional stools, needed for the experiment during the study period. Fresh stools were given by oral gavage, as follows: \( n = 53 \) in 7-wk study (CRC-M transfer, \( n = 30 \) and N-M transfer, \( n = 23 \)) and \( n = 132 \) in 14-wk study (CRC-M transfer, \( n = 66 \) and N-M transfer, \( n = 66 \)). In addition to the FIT, 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 challenge of environmental exposure measurement in molecular epidemiology. The main endpoint was identifying those individuals with advanced neoplasia (invasive carcinoma or carcinoma in situ or specific carcinoma in 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 \( \chi^2 \) test for binary variables and Student's \( t \) test for continuous variables. Blood methylation data (CMI and methylation of each of its 3 components, \( \text{WIT}, \text{PENK}, \text{and NPY} \)) 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, 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.

**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. Mozzer, A. Rosenbaum, Ph. Capelle, D. Levoir, F. Maille, Ph. Lebourgeois, Ph. De Land, E. Chanteloup, M. F. Mal, and F. Iggidr. We thank also Drs. J. Tran Vanthieu and M. L. Auriel for pathology analyses in human and animals; Prof. S. Loric for biochemistry analysis in mice; Prof. T. Simon, A. Touati, J. Tap, V. Jarrouse, A. Bado, and J. P. Fouret for animal experiments and help managing. Acknowledgments for revising the English; S. Peyvandi for assisting with the animal and molecular experiments; E. Guery and L. Segaux for their statistical contributions; C. Viatte for data managing; A. Caidia (Bioinformatics Core Laboratory) for 16S RNA 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 M. Sharifi for assistance with the analysis of inflammation in mouse colon 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 NCT01270360.

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).