



**HAL**  
open science

## Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures

Iradj Sobhani, Emma Bergsten, Séverine Couffin, Aurelien Amiot, Biba Nebbad, Caroline Barau, Nicola De'angelis, Sylvie Rabot, Florence Canouï-Poitrine, Denis Mestivier, et al.

### ► To cite this version:

Iradj Sobhani, Emma Bergsten, Séverine Couffin, Aurelien Amiot, Biba Nebbad, et al.. Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116 (48), pp.24285-24295. 10.1073/pnas.1912129116 . pasteur-02551921

**HAL Id: pasteur-02551921**

**<https://pasteur.hal.science/pasteur-02551921>**

Submitted on 5 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures**

2

3 Iradj Sobhani<sup>a,b,1</sup>, Emma Bergsten<sup>a,c</sup>, Séverine Couffin<sup>a</sup>, Aurélien Amiot<sup>a,b</sup>, Biba Nebbad<sup>d</sup>,  
4 Caroline Barau<sup>e</sup>, Nicola de' Angelis<sup>f</sup>, Sylvie Rabot<sup>g</sup>, Florence Canoui-Poittrine<sup>h</sup>, Denis  
5 Mestivier<sup>a,i</sup>, Thierry Pédrón<sup>c</sup>, Khashayarsha Khazaie<sup>j,2</sup>, Philippe J. Sansonetti<sup>c,k,1,2</sup>

6

7 **Running title:** *Methylation Pathway-Related CRC Dysbiosis*

8 *Affiliations*

9 <sup>a</sup>EA7375 (EC2M3 research team) from Université Paris Est Creteil (UPEC)

10 <sup>b</sup>APHP, Hôpital Henri Mondor, Service de Gastroenterologie, Creteil, France

11 <sup>c</sup>Unité de Pathogénie Microbienne Moléculaire INSERM U1202-Institut PASTEUR,

12 Paris

13 <sup>d</sup>Service de Microbiologie Hôpital Henri Mondor-APHP Paris

14 <sup>e</sup>Plateforme de Ressources biologique, Hôpital Henri Mondor, Creteil

15 <sup>f</sup>Service de Chirurgie Digestive, Hôpital Henri Mondor-APHP Paris

16 <sup>g</sup>Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay

17 <sup>h</sup>Service de santé publique, Hôpital Henri Mondor-APHP Paris

18 <sup>i</sup>Bioinformatics Core Lab, Institut Mondor de Recherche Biomedicale (IMRB),

19 INSERM U955, Créteil, France

20 <sup>j</sup>Department of Immunology Mayo Clinic Rochester USA

21 <sup>k</sup>Chaire de Microbiologie et Maladies Infectieuses, Collège de France, Paris

22

23 **ClinicalTrials.gov registration number** NCT 01270360

24

25 **Classification: Biological Sciences, Microbiology**

26

27

28 <sup>1</sup>**Correspondence:** Prof. Iradj Sobhani, Service de Gastroenterologie

29 Hôpital Henri Mondor 51 Av Mal de Lattre de Tassigny 94010 CRETEIL France

30 **Email:** iradj.sobhani@aphp.fr

31 Tel: 33-149814358 or 33-149812362;

32 Fax: 33-149812352

33 Prof. Philippe J. Sansonetti, Unité de Pathogénie Microbienne Moléculaire, Institut Pasteur,

34 28 rue du Docteur Roux, 75724 Paris cedex 15

35 Email: [philippe.sansonetti@pasteur.fr](mailto:philippe.sansonetti@pasteur.fr)

36 Tel: 33-145688342

37

38 <sup>2</sup>K.K. and P.J.S. contributed equally to this work.

39

40 **Abstract**

41

42 Sporadic colorectal cancer (CRC) is a result of complex interactions between the host and its  
43 environment. Environmental stressors act by causing host-cell DNA alterations implicated in  
44 the onset of cancer. Here, we investigate the stressor ability of CRC-associated gut dysbiosis  
45 as causal agent of host DNA alterations. The epigenetic nature of these alterations was  
46 investigated in humans and in mice. Germ-free mice receiving fecal samples from subjects  
47 with normal colonoscopy or from CRC patients were monitored for 7 or 14 weeks. Aberrant  
48 crypt foci, luminal microbiota and DNA alterations (colonic exome sequencing and  
49 methylation patterns) were monitored following human feces transfer. CRC-associated  
50 microbiota induced higher numbers of hypermethylated genes in murine colonic mucosa (vs.  
51 healthy controls' microbiota recipients). Several gene promoters including SFRP1,2,3,  
52 PENK, NPY, ALX4, SEPT9, WIF1 promoters were found hypermethylated in CRC but not  
53 in normal tissues or effluents from fecal donors. In a pilot study (n=266), the blood  
54 methylation levels of three genes (*Wif1*, *PENK* and *NPY*) were shown closely associated  
55 with CRC dysbiosis. In a validation study (n=1000), the cumulative methylation index  
56 (CMI) of these genes was significantly higher in CRCs than in controls. Further, CMI  
57 appeared as an independent risk factor for CRC diagnosis as shown by multivariate analysis  
58 that included fecal immunochemical blood test. Consequently, fecal bacterial species in  
59 individuals with higher CMI in blood were identified by whole metagenomic. Thus, CRC-  
60 related dysbiosis induces methylation of host genes and corresponding CMI together with  
61 associated bacteria are potential biomarkers for CRC.

62

63

64

65 **Significance Statement**

66 This study advances our appreciation and understanding of the role of colon dysbiosis in the  
67 pathogenesis of colorectal cancer. In a human pilot study of 266 individuals, greater  
68 epigenomic (methylation) DNA alterations correlated with CRC and microbiota  
69 composition. Beyond this correlative evidence, when germ-free mice received fresh feces  
70 from CRC patients and their healthy controls, the former animals developed colon epithelial  
71 renewal, more precancerous lesions, and increased tissue and blood DNA methylation in  
72 intestinal tissues. Confirmation was obtained in a larger cohort of 1000 patients, indicating  
73 that CRC-associated dysbiosis may promote colon carcinogenesis via epigenome  
74 dysregulation. Gene methylation can therefore serve as a marker for CRC and likely for  
75 predicting efficacy of prebiotic supplementation in average-risk individuals.

76

77

78 **Key words:** Colon. Cancer. Microbiota. Gene methylation. *Wif1*, *NPY*, *PENK* Biomarker.

## 79 **Introduction**

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

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

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

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

104 free mice in order to assess the potential effect of this microbiota as an environmental  
105 stressor for host DNA. Secondly, we investigated the links between dysbiosis and host gene  
106 methylated markers suggested by animal experiments in a pilot human study and validated in  
107 a large human cohort.

108

109

## 110 **RESULTS**

111

### 112 **Human fecal microbiota transfer (FMT) to germ-free mice: clinical and routine** 113 **biochemical parameters among groups.**

114 We performed human fecal microbiota transfer (FMT) to germ-free mice utilizing  
115 material from patient donors with CRC (n=9) and healthy control donors (n=9) as illustrated  
116 in “SI Appendix, Fig. S1”. In addition, mice received the carcinogen azoxymethane (AOM)  
117 or saline control for 10 weeks in order to assess potential effects of chemically induced  
118 stress. As shown in “SI Appendix, Tables S1A-S1B”, FMT had no effect on the general  
119 blood parameters of the recipient mice as no significant differences were observed among  
120 any of the groups (a total of 185 mice) at 7 and 14 weeks. At 14 weeks, mice that received  
121 both AOM and colorectal cancer microbiota (CRC- $\mu$ ) showed a lower mean body weight  
122 compared to the normal microbiota (N- $\mu$ ) controls. This was accounted to reduced food  
123 intake observed in the CRC- $\mu$  group when compared to the N- $\mu$  controls ( $1.5\pm 0.1$  g/100 g  
124 food intake vs.  $2.0\pm 0.1$  g/100 g food intake, respectively;  $P<0.01$ ). Interestingly, this weight  
125 effect disappeared when mice were given PEG in their drinking water, which is known to  
126 prevent bacterial colonization of the gut (SI Appendix, Fig. S2). In line with this, the luminal  
127 bacterial load, adherence of luminal bacteria to the colonic mucosa and short-chain fatty acid  
128 (SCFA) concentrations in the caecum were significantly lower in mice that received PEG as

129 compared to the controls (SI Appendix, Fig. S2A-S2B). These results suggest that human  
130 CRC FMT may induce systemic changes in mice mediated by the interaction of bacteria with  
131 the colonic mucosa.

132

### 133 **Dysbiosis-related histological alterations in colonic tissue after human FMT in mice**

134 In order to delineate the predominant modifications occurring in murine colonic tissues  
135 following FMT, the number of aberrant crypt foci (ACF), i.e. clusters of enlarged thick  
136 crypts visualized by methylene blue at low magnification, were enumerated. Whereas  
137 histological features appeared normal in the group of mice receiving only N- $\mu$ , mild  
138 inflammation and ACF were seen at significant levels in the groups given CRC- $\mu$ , AOM, or  
139 both (Figure 1A-D; H-I). After 7 or 14 weeks, ACF numbers were highest in the group given  
140 CRC- $\mu$  + AOM, compared to all other groups and were also significantly higher in the  
141 groups given CRC- $\mu$  alone or AOM alone, compared to the N- $\mu$  only group (Figure 1H-I).  
142 No ACFs were seen in mice given PEG, in which the microbiota was efficiently reduced in  
143 the mucus layer proximal to the mucosa. These latter samples were therefore not further  
144 examined. According to the colon length, numbers of ACFs became significantly higher in  
145 groups given CRC- $\mu$  versus N- $\mu$ , when only 20 mm of the left colonic mucosa was  
146 examined ( $13 \pm 3$  versus  $5 \pm 1$ ;  $P < 0.05$ ; Figure 1I).

147

### 148 **Human fecal microbiota transfer (FMT) to germ-free mice: effect on epithelial** 149 **proliferation and gene expression**

150 Epithelial-cell proliferation in the colonic mucosa was visualized by KI67 staining (as  
151 illustrated in Figure 1G). At 7 weeks following FMT, a significant increase of proliferation  
152 was observed in the groups given AOM as compared to all other groups (Figure 1H).  
153 However, at week 14 cell proliferation was significantly greater in the CRC- $\mu$ , as compared

154 to N- $\mu$  recipients with AOM exerting an additional effect (Figure 1H-I) and ACFs increased  
155 with the rate of epithelial-cell proliferation. These results are consistent with those reported  
156 after FMT from only 5 CRC patients as compared to 5 control individuals' microbiota (13).  
157 Further, to investigate the involvement of *Wnt* and *Notch* pathways that control intestinal cell  
158 fate, mRNA of key mediators was quantified. qPCR analysis of mRNA from the colonic  
159 mucosa of recipient mice revealed an increase of 1.7-fold, 1.9-fold, and 1.9-fold expression  
160 levels of the transcription factors *HES1*, *KLF4*, and *ELF3*, respectively but not of *MATH1* in  
161 CRC- $\mu$  compared to N- $\mu$  recipients (SI Appendix, Table S2).

162

### 163 **Human fecal microbiota transfer (FMT) to germ-free mice: mucosal inflammation**

164 Early after FMT (7 weeks), a mild inflammatory response reflected by higher IL1 $\beta$ ,  
165 IL6, and MIP2 $\alpha$  levels and lower IL10, IL23 and INF $\gamma$  levels, was detected in CRC- $\mu$   
166 recipient animals as compared to N- $\mu$  controls. AOM treatment tended to to amplify this  
167 stimulatory effect as assessed by mRNA cytokine quantification (Figure 1J). In addition,  
168 semi-quantitative evaluation of mucosal myeloid cells along the entire histologically-  
169 examined intestine (ten fields per sample, 3 samples/mice) showed a trend towards increased  
170 numbers in human CRC- $\mu$  as compared to N- $\mu$  recipients, but without reaching statistical  
171 significance at any timepoint (Figure 1K).

172

### 173 **Microbiota characterization following human FMT in mice**

174 To assess for the impact of the microbial communities on the observed differences in  
175 the mucosal phenotypes, murine fecal samples were analyzed by 16S rRNA gene  
176 sequencing. At baseline, the fecal microbiota of the CRC- $\mu$  (n = 9 donors) group contained a  
177 higher proportion of *Fusobacteria*, *Parvimonas*, *Butyrivibrio*, *Gemella*, and *Akkermansia*,

178 and lower proportions of *Ruminococcus*, *Bifidobacterium*, *Eubacteria* and *Lachnospira*, as  
179 compared to stools from the N- $\mu$  (n=9 donors) group (SI Appendix, Table S3 and Fig. S3).

180 During the follow-up, all groups (n=6 per group and at each time point) showed a  
181 moderate decrease in *Coccoides*, *Clostridium leptum*, and *Bifidobacterium* on day 7 and  
182 stability in the community of bacteria afterward, as estimated by qPCR (SI Appendix, Table  
183 S4). Based on 16S rRNA gene sequencing of mouse stools (n=6 in each group at each  
184 timepoint of the follow-up), 85% of OTUs remained unchanged over time. Overall, the  
185 bacterial species most associated with significant histopathological alterations, *i.e.*  
186 precancerous lesions (e.g., ACF), were *Firmicutes*, *Clostridia*, and *Clostridiales*. These  
187 histological changes were also associated with lower counts of genera with an anti-  
188 inflammatory effect (*i.e.* *Faecalibacterium*, *Eubacterium*) and of butyrate-producing bacteria  
189 (*Firmicutes* species) (SI Appendix, Fig. S4). Interestingly, the abundance of *Coprococcus*  
190 was lower and of *Bacteroides* was higher in CRC- $\mu$  vs. N- $\mu$ . Co-variations of these two  
191 genera with histological changes characterized by ACF numbers, and with mRNA levels of  
192 the previously quantified transcription factors in mice colonic parameters were analyzed. Co-  
193 inertia analysis revealed that these parameters were associated with variation of the fecal  
194 bacteria. The first component of the analysis showed a significant separation between N-  $\mu$   
195 and CRC-  $\mu$  recipients. This first component was significantly associated linearly with  
196 *Bacteroides* and *Coprococcus* genera abundances in mice stools (SI Appendix, Fig. S5).  
197 These allowed a significant separation between N- $\mu$  and CRC-  $\mu$  recipients regarding  
198 bacteria and mRNA levels of key transcription factors involved in *Wnt* and *Notch* pathways.

199

## 200 **Human FMT to germ-free mice alters DNA**

201 **DNA mutations.** In order to analyze the potential of FMT to induce DNA modifications in  
202 the host, whole-genome sequencing of murine colonic mucosal tissues (n=12, three in each

203 group of recipients) was performed, covering 220,000 exons within 24,000 genes. Overall,  
204 the incidence of mutations at global exon/intron level was significantly higher in intestinal vs  
205 spleen tissues. Results of the colonic mucosa showed a trend toward increased DNA  
206 alterations in exons or introns of CRC- $\mu$  compared to N- $\mu$  recipients, but this did not reach  
207 significance (Figure 2A). In addition, mice given CRC- $\mu$  and AOM showed the highest  
208 levels of DNA alterations in exons or introns showing a trend to an additive effect of CRC- $\mu$   
209 as compared to N- $\mu$ . Interestingly, PCA scatter diagrams separated animal subgroups  
210 according to the total DNA changes in single-nucleotides in the colonic mucosa or spleen  
211 samples suggesting a link between gene mutation and type of microbiota (N or CRC) or  
212 treatment (Saline or AOM) given to mice (Figure 2B-C).

213 In order to investigate the local oncogenic potential of CRC- $\mu$ , we went on to perform in-  
214 depth analysis of the following selected gene pathways (number of genes analyzed): Wnt  
215 and  $\beta$ -catenin (19 genes), Notch (4 genes), PPAR (3 genes), SMAD (2 genes), TGF- $\beta$  (2  
216 genes), ACRV (2 genes), DKK (4 genes), TCF (2 genes), MYC (1 gene), SOCS (1 gene).  
217 Changes were most prominent in pivotal *Wnt* pathway genes (with indel and single  
218 nucleotide polymorphisms in 11 *Wnt* genes), with no significant differences between the  
219 CRC- $\mu$  and N- $\mu$  groups. When all DNA mutations were pooled, however, the number of  
220 DNA changes was highest in the CRC- $\mu$  + AOM recipients vs. the N- $\mu$  recipients (Figure  
221 2D). Again, DNA changes were significantly more numerous in colonic than in spleen  
222 tissues. Interestingly, in the spleen, DNA changes were similar among AOM-exposed  
223 animals that received CRC- $\mu$  or N- $\mu$ . Finally, an unsupervised analysis between mutated  
224 genes (including only selected genes as indicated above) and recipient status showed that  
225 changes in *Wnt*, *PPAR $\gamma$* , and *Notch* pathway genes were associated with CRC- $\mu$  recipient,  
226 with an additive effect of AOM treatment (Figure 2C, SI Appendix, Fig. S6 and Table S2).

227

228 **DNA methylation/demethylation.** We isolated colonic mucosal tissue from the colon (n=24;  
229 6 in each group) and evaluated epigenetic changes using the EPIC microarray to compare  
230 methylation rates in 63,987 probes reflecting 12,600 gene being expressed. For all probes  
231 pooled, both mean and median methylation rates decreased from the N- $\mu$ +NaCl group to the  
232 CRC- $\mu$ +NaCl and CRC- $\mu$ +AOM groups (Figure 2E), with significantly lower methylation  
233 levels in the CRC- $\mu$ /AOM group than in the N- $\mu$ +NaCl group ( $P<0.01$ ). Overall, one third of  
234 the probes appeared unmethylated in all experimental conditions ( $<0.2$  beta-value  
235 methylation mEPIC) in the colonic tissue after FMT (Figure 2F). The number of methylated  
236 probes ( $>0.8$ ) varied with N- $\mu$ , CRC- $\mu$ , and AOM exposure, although the ratio of  
237 methylated/unmethylated probes was not significantly different (3.3% and 3.1%,  
238 respectively ( $P=0.25$ , Fisher's test)) (Figure 2G). The number of unmethylated probes was  
239 11% higher in the CRC- $\mu$ +AOM group when compared to the N- $\mu$ +NaCl group ( $p=0.013$ )  
240 and the number of fully methylated probes was higher in CRC- $\mu$  and AOM recipients  
241 (Figure 2H). In summary, human CRC- $\mu$  induced greater epigenetic alterations in murine  
242 colonic mucosal tissues when compared to N- $\mu$ . The ratio of methylation-shifted probes  
243 (methylated to unmethylated and vice versa) were higher in the CRC- $\mu$  group than in the N-  
244  $\mu$  group, with AOM. When we limited the analysis to *in silico* specific EPIC probes, two  
245 thirds of the probes with hypermethylation levels (0.8 or more) in the CRC- $\mu$ +AOM group  
246 were unmethylated ( $<0.2$ ) in the N- $\mu$  group, corresponding to 46 genes in the CRC- $\mu$ +AOM  
247 group being silenced *vs.* N- $\mu$  group. In contrast, various probes in *Wnt* and *Notch* gene  
248 families appeared unmethylated in the CRC- $\mu$ +AOM group but were methylated in the N- $\mu$   
249 group. Overall, significant DNA epigenetic rather than mutation changes in several genes  
250 were associated with CRC- $\mu$  human FMT in the mouse colonic mucosa. Pooling of total  
251 genes with expected DNA alterations (i.e. hypermethylation, or mutation in exon) revealed

252 the involvement of pathways implicated in cell growth, signal transduction, nucleic acid  
253 binding, protein synthesis, channel and carrier protein (SI Appendix, Fig. S7).

254

### 255 **Gene methylation in human colonic samples: from bench to bed**

256 To investigate whether the gene methylation patterns observed in the mouse were also  
257 associated with CRC dysbiosis in humans, we first investigated methylated genes in CRC  
258 tissues and effluents such as blood and stool (n=9) and in normal tissue effluent controls  
259 (n=9), both obtained from the same individuals as for the FMT in mice (Figure 3A). In order  
260 to develop a quick and easy methylation test, panels of genes were selected, based on the  
261 difference between “normal” and “CRC” in various effluents and tissues. Among the various  
262 genes classified as hypermethylated in the individual samples (tissue, stool, blood) from the  
263 same individuals, only 8 (*Wif1*-regulating gene and *SEPT9*, *SFRP1,2,3*, *PENK*, *NPY*, and  
264 *ALX4* genes) of these were common to all samples and were therefore subjected to further  
265 characterization (Figure 3B and SI Appendix, Fig. S8-S11). We selected *Wif1* to represent  
266 the *Wnt* pathway and *NPY*, *PENK*, two other tissue CRC-associated hypermethylated genes  
267 for methylation testing, as at least either one was found hypermethylated in all tumor tissues.  
268 The sum of the methylation levels of the three selected genes was found to serve best when  
269 discriminating between effluents from normal and cancer individuals (details in SI  
270 Appendix). Finally, to conceive easy and reproducible biological testing for large cohort  
271 screening, a housekeeping gene plus selected genes (i.e. *ALB*; *WIF1*; *NPY*; *PENK*) were  
272 combined in a multiplex procedure for qPCR measurement in blood. A cumulative methyl  
273 index (CMI) was determined in blood samples from participants with normal colonoscopy or  
274 with a cancer diagnosed (32 patients with CRC and 46 with extra colonic cancer; see SI  
275 Appendix). To validate its performance, CMI assessment was performed in a pilot study  
276 including 266 individuals (SI Appendix, Table S5) from a well-characterized cohort

277 designed as CCR1 (SI Appendix, Table S6) and in a final validation study including 999  
278 individuals (Table 1) from a second cohort designed as CCR2 (Table S7). The specificity  
279 and the sensitivity of CMI>2 in blood, was 95% and 59% respectively in the pilot study  
280 when those with positive Fecal Occult Blood Test (FOBT) in stool were 97% and 43%,  
281 respectively. In the validation study half of individuals were enrolled through mass CRC-  
282 screening programs and had a new fecal immunochemical test to directly measure human  
283 hemoglobin in stools (FIT) and all had a blood with a threshold of CMI> 2 for positivity.

284

### 285 **Validation of a cumulative methylation index (CMI)**

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

291 A CMI>2 correlated significantly with CRC in 999 individuals in the validation study.  
292 Interestingly, a tendency of CMI increase with age was noted, however this observation did  
293 not reach significance. By keeping specificity of both tests superior to 95%, the negative  
294 predictive value (mean, extremes) of CMI (set at >2) test was better than FIT (set at >150  
295 ng/mL): 84.09 [81.5-86.4] and 81.64 [77.5-85.3], respectively (SI Appendix, Table S8). A  
296 multivariate model including all individuals with and without neoplasia adjusted for age,  
297 gender and FIT results, showed CMI>2 was independently associated with CRC (Table 1).  
298 Regarding CMI test and keeping with threshold for positivity set at  $\geq 2$ , the sensitivity for  
299 the detection of CRC could reach 37% in the validation study.

300 After stratification on TNM-ICAA classification (SI Appendix, Table S9), the CMI  
301 value levels increased significantly and linearly with increasing tumor stages.

302

### 303 **Dysbiosis and associations with methylation of genes**

304 16S rRNA gene sequencing on stool samples (n=513) indicated a great  $\beta$ -diversity in  
305 patients with CRC and in those with a CMI>2 ( $P<0.05$ ). Furthermore, CMI levels were  
306 significantly correlated with dysbiosis. In the validation cohort, we hierarchically clustered  
307 bacterial phylotypes on the genus level based on the similarity of their dynamics in patients  
308 with CRC and in those with CMI>2 (Figure 4).

309 Interestingly, Principal Coordinate Analysis on the microbiota when incorporating the  
310 diagnosis and the level of blood methylation test (Figure 4A), separated into two main  
311 clusters which differed significantly between participants with versus without a CMI>2,  
312 regardless of whether they did or did not have CRC (Figure 4B). A single cluster differed  
313 significantly in abundance between controls with and without a CMI>2 ( $P<0.05$ , Wilcoxon  
314 signed-rank test; Figure 4C).

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

322

### 323 **DISCUSSION**

324 The data reported here constitute the first evidence that the relative abundance of some  
325 bacterial taxonomic groups within the microbiota in CRC is significantly associated with  
326 methylation/demethylation of host genes. In our study, we show clear phenotypical

327 differences between mice receiving fecal transplants from CRC patients vs. healthy controls.  
328 The CRC- $\mu$  group presented with lower fecal short-chain fatty acid (SCFA) concentrations  
329 and had significant colonic mucosal changes including higher ACF numbers and marked  
330 epigenetic alterations independent AOM. However, two findings in the mouse study indicated  
331 an additive effect of AOM to the CRC microbiota: (i) the number of DNA alterations was  
332 greater in colonic than in spleen tissues under AOM (Figure 2D, 2G), and (ii) the number of  
333 spleen-tissue DNA alterations was similar in the CRC- $\mu$  and N- $\mu$  recipients. Interestingly,  
334 16S rRNA gene sequencing of fecal microbiota from CRC patients and normal controls  
335 showed that CRC-associated dysbiosis was significantly associated with hypermethylation of  
336 several gene promoters, including *NPY* and *PENK* from the brain gut system and *Wif1* from  
337 the *Wnt* pathway. A CMI test constructed from these genes and performed in blood detected  
338 more (vs. controls) CRC patients even in asymptomatic individuals (n=32) who were  
339 recruited through mass screening program and had early CRC at colonoscopy. However, the  
340 CMI showed higher sensitivity rate in the pilot study likely due to advanced CRCs being at  
341 stage III or IV in this series (SI Appendix, Table S9). Indeed, CRCs in the validation study  
342 were consistently at an early stage (0, I or II). Nevertheless, more than 50% of CRC patients  
343 with symptoms and more than 35% of asymptomatic individuals presenting with an early  
344 CRC in the validation study showed a  $CMI \geq 2$  contrasting with only 4% in individuals with  
345 normal colonoscopy.

346 In keeping with earlier reports and with two very recent meta-analyses (10, 13-16),  
347 microbiota from CRC patients in the current study contained higher proportions of  
348 *Fusobacterium*, *Parvimonas*, *Butyrivibrio*, *Gemella*, *Fusobacteria*, and *Akkermansia*  
349 contrasting with lower proportions of *Ruminococcus*, *Bifidobacterium*, *Eubacteria*, and  
350 *Lachnospira*, compared to human control microbiota. Despite some early changes after  
351 microbiota transfer in germ-free mice, possibly due to the FMT itself as reported (17), most

352 of the human microbiota components subsequently remained stable over time. The mouse  
353 fecal bacteria that showed the greatest decline over time (SI Appendix, Fig. S4) were  
354 *Clostridia* and *Clostridiales*, as analyzed at the class or the genus levels which might account  
355 for food uptake and weight differences among mouse groups. Numerous species belonging  
356 to this bacterial community have been associated with an increase in food intake and  
357 upregulated production of proinflammatory molecules (18-19). SCFAs, the main end-  
358 products of butyrate-producing bacteria (19) have been shown to inhibit intestinal  
359 inflammation and modulate immune responses (20), maintain barrier function (21), decrease  
360 precancerous lesions due to DNA damage (22-24) and regulate DNA methylation (25).

361 Of interest, the observed dysbiosis was correlated with the histological and DNA  
362 findings in the animals while the putative role of dysbiosis and inflammation in the initiation  
363 of cancer in the colonic mucosa remained unclear. We used AOM as a co-factor together  
364 with microbiota that could induce DNA alterations (22, 25). DNA alterations were more  
365 numerous in the colonic mucosa than in splenic tissue, effects which were abolished in the  
366 colon by PEG in CRC- $\mu$ +AOM recipients. Furthermore, transcripts of factors such as HES,  
367 KLF4, and ELF3 involved in the *Wnt* and *Notch* pathways, were more abundant in CRC- $\mu$   
368 than in N- $\mu$  mice recipients, depending on higher *Bacteroides* and lesser *Coprococcus* in  
369 mouse feces after human FMT (SI Appendix, Fig. S5). Thus, our findings suggest that  
370 histological alterations in the colonic mucosa might be due to an imbalance in microbiota  
371 composition with CRC- $\mu$  being associated with greater DNA damage and/or gene  
372 methylation/demethylation changes in the colonic mucosa. Although there was globally a  
373 greater number of unmethylated probes in the CRC- $\mu$ +AOM group than in the N- $\mu$  group,  
374 few genes (i.e. *Wif1*) might be methylated after FMT. This is suggested by the observed  
375 human CRC- $\mu$  induced DNA methylation/demethylation imbalance in mouse colonic  
376 mucosa with an additive effect of AOM. These findings are consistent with previous

377 evidence of gene expression silencing by gene methylation in overall one third of human  
378 tumor tissues (5) and significant associations between abundance of *Parvimonas micra* and  
379 *Bacteroides fragilis* with highly methylated tumors (16). To evaluate this hypothesis, we  
380 confirmed that *Parvimonas micra* species, was over abundant in the microbiota of those  
381 patients presenting a higher methylation gene index in the blood first by analyzing 16sRNA  
382 and then by using whole metagenome analysis for the confirmation at the species levels  
383 (Table 3). Thus, one would suggest that genera such as *Parvimonas* and *Proteobacteria* may  
384 use epigenetic pathways for adaptation to environmental factors (4, 26-30), with  
385 methylation/demethylation as a pivotal mechanism (31-32).

386 In our cohorts, the *Prevotella* genus, a leading source of inter-individual gut  
387 microbiota variation associated with long-term fiber intake (8), was more abundant in  
388 individuals with CMI $\leq$ 2. Of interest, *Parvimonas* and *Parasutterella* were the bacteria most  
389 closely associated with an animal-based diet and were related to a high fat intake (8), both of  
390 which probably result in greater bile acid release and higher enteric deoxycholic acid  
391 concentrations. These two bacteria co-exclude anti-inflammatory bacteria such as  
392 *Faecalibacterium* and *Eubacterium* as seen currently and reported by us and others (10, 12-  
393 15). Of interest, was also, *SFRP2* hypermethylation in tumor tissue and blood (SI Appendix,  
394 Fig. S10A&C) which was associated with *Bilophila*, another proinflammatory genus (Tables  
395 S3 and S10), and that co-excluded currently *Faecalibacterium* in the CRC microbiota.  
396 Increased abundance of the *Bilophila* genus was associated with *SFRP2* gene promoter  
397 demethylation after black raspberry supplementation (33) used as probiotics.

398 Our present results might be affected by various study limitations. Firstly, in our prospective  
399 validation trial (NCT01270360), the blood CMI was determined at a single laboratory, and  
400 bias may therefore have occurred. The blood CMI was performed on serum and was  
401 negative in some CRC patients. Different results might have been obtained using plasma

402 instead of serum. Thus, methylation was further investigated in random samples by  
403 performing a digital PCR test known to be more sensitive than routine qPCR as previously  
404 reported (34). Even with dPCR test, sensitivity was higher in the symptomatic CRC patients  
405 likely due to the more advanced CRC in symptomatic patients as compared to asymptomatic  
406 CRC individuals. This would suggest a blood CMI positivity test in CRC patients with  
407 symptoms was likely due to the presence of tumor cells in the bloodstream rather than to  
408 systemic plasticity of peripheral blood cells such as mononuclear cells. A second limitation  
409 of this study is that only three genes were incorporates in the blood CMI. Additional genes  
410 such as *SFRP2* might enhance sensitivity of the test and might yield different bacterial  
411 clustering patterns to those reported here. Furthermore, while viruses and CRC are still  
412 subject to discussion, we cannot rule out the hypothesis that in CRC patients particular  
413 phages modulate the gut microbiota of patients, inducing a dysbiotic profile compared to  
414 healthy controls (35).

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

422

## 423 **Individuals & Materials & Methods**

### 424 *Recruitment of participants and collection of samples*

425         Patients referred to University hospitals for colonoscopy were enrolled in several  
426 prospective cohorts. Effluents and tissues from participants were used for experimental,

427 proof-of-concept, and for a pilot study that included 266 individuals from a cohort named  
428 CCR1 (SI Appendix, Table S6). A second cohort, named CCR2 (SI Appendix, Table S7)  
429 was constituted for validation; this was composed of two sub-cohorts: one including only  
430 symptomatic (named Valihybritest study) and the second including only asymptomatic  
431 individuals (named Vatnimad) recruited from mass screening programs. All individuals  
432 underwent colonoscopy due to symptoms or due to a positive fecal blood test (FOBT). Both  
433 sub-cohort studies were registered on ClinicalTrials.gov (NCT01270360) and 1000  
434 consecutive colonoscopies were to be enrolled, among them 500 individuals (Vatnimad)  
435 without any digestive symptoms. The study protocol was approved by the ethics committee  
436 of CPP Paris Est-Henri Mondor #10-006 on 2010). All participants signed an informed  
437 consent.

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

443

#### 444 **Studies of microbiota from human participants**

##### 445 *16S rRNA gene and whole metagenomic sequencing on stools*

446 Stool samples were collected in donors and patients during experimental and  
447 validation studies, respectively; samples of stool were collected and stored within 4 hours  
448 for DNA extraction using the GNOME<sup>®</sup> DNA Isolation Kit (MP Biomedicals, Santa Ana,  
449 CA) as previously described (9-10, 12). After amplification by PCR of the V3-V4 region of  
450 the 16S rRNA gene, sequencing was performed using a 250-bp paired-end sequencing  
451 protocol on the Illumina MiSeq platform. Raw FASTQ files were demultiplexed, quality-

452 filtered using Trimmomatic (Sliding windows of 2 with a quality score of 20), and merged  
453 using fastq-join from ea-utils (<https://expressionanalysis.github.io/ea-utils>). Taxonomic  
454 assignments were performed using Qiime2 (no quality filtering; default parameters) (36)  
455 with the SILVA-123 database; OTU were constructed using UCLUST (threshold of 97% of  
456 similarity), Chimera Slayer for chimera removing and SILVA 16S rRNA database (version  
457 123) for taxonomical assignment. The intergroup high similarity and intragroup low  
458 similarity of microbiota were assessed by  $\beta$ -diversity, PCoA (generated by Qiime using  
459 unweighted unfrac metrics). We subjected study populations to two Principle Coordinates  
460 Analyses, independently of other data sets, and we investigated separation of CRC  
461 microbiota from control microbiota in donors in the experimental study as well as in controls  
462 versus CRC patients in the clinical trials. All microbiome statistical tests were produced  
463 using the MetagenomeSeq packages or the Shaman Webserver that used the DESeq2  
464 packages for differential expression analyses; therefore every statistical analysis for  
465 microbiome have been corrected according to gender, age and BMI and adjusted for multiple  
466 testing.

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

469

## 470 **Studies of methylation from human participants**

### 471 *DNA isolation and bisulfite (BS) conversion*

472 DNAs from colonic tissues (QIAamp DNA Mini Kit, Qiagen), blood and stool samples  
473 (QiAamp DNA stool Mini Kit, Qiagen) were extracted using the ZR Serum DNA kit  
474 (Ozyme, Montigny-le-Bretonneux, France) according to the manufacturer's protocol. The  
475 DNA samples were then exposed to sodium bisulfite at 50°C in the dark for 16 hours (EZ  
476 DNA Methylation kit, Zymo Research) to convert unmethylated cytosine nucleotides into

477 uracil nucleotides (subsequently converted to thymidine nucleotides during PCR cycling)  
478 without changing the methylated cytosines as detailed elsewhere (10, 12).

479

#### 480 ***DNA methylation using Illumina Golden Gate methylation bead arrays***

481 The GoldenGate Methylation Cancer Panel I (Illumina, San Diego, CA) was used to  
482 probe 500 ng of each bisulfite-converted DNA sample of human (n=18, 9 CRC patients and  
483 9 controls) including tissue (n=18), stool (n=18), and blood (n=18), the stool samples were  
484 used for fecal microbiota transfer to germ free mice (SI Appendix, Table S11). Methylation  
485 levels ranged from 0% to 100.0% were used for the calculation of the ratio of the methylated  
486 signal intensity. The strategy for hierarchical clustering of gene candidates are further  
487 described in Figure 3 and in SI Appendix, Fig. S8-S11. Briefly, comparisons across tissue,  
488 stool, and blood samples identified genes with CpG loci methylation levels in the promotor  
489 above the expected number. The CMI was computed by addition of the methylation values  
490 of the three genes generated by the discovery study (characteristics of individuals in the pilot  
491 study in SI Appendix, Table S12). Primers targeting all genes, (including albumin gene-  
492 ALB, devoid of CpG sites and used as a housekeeping gene and for normalization of DNA  
493 amounts), are reported elsewhere (SI Appendix, Table S13).

494

#### 495 ***Quantitative methylation-specific PCR amplification and verification of the specificity***

496 Bisulfite-converted universal human methylated DNA standard (Zymo Research)  
497 served as a calibrator and positive control and DNA-free distilled water as a negative  
498 control. Quantitative, single-gene methylation-specific PCR (QS-MSP) and quantitative  
499 multiplex methylation-specific PCR (QM-MSP) were applied. The relative methylation level  
500 was determined using the  $2^{-\Delta\Delta C_t}$  formula. Briefly, for each PCR run, a KAPA PROBE master

501 mix (Kapa Biosystems, Wilmington, MA) was prepared with pre-defined concentrations of  
502 genes candidates.

503

#### 504 ***Statistical analyses of blood CMI and FIT data in the validation cohort***

505 With the type I error set at 0.05 and assuming that CRC would be found in 8% of  
506 asymptomatic and 25% of symptomatic individuals, to detect at least 60% of CRCs and 20%  
507 of polyps with 90% specificity in average-risk (asymptomatic and aged  $\geq 50$  years) and high-  
508 risk (history of polyps or sibling with CRC) individuals with 80% power, 1000 participants  
509 (with at least 400 in each sub cohort) were needed. Characteristics of study populations were  
510 described using number (%) for qualitative variables and mean  $\pm$  SD for quantitative  
511 variables.

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

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

525

526 **Studies from experiments in mice**

527 Fresh stool samples were obtained from 9 females and 9 males in CCR1 cohort, for fecal  
528 microbiota transfer (FMT) to germ-free mice. They were 9 with CRC and 9 with normal  
529 colonoscopy (SI Appendix, Table S11).

530

531 ***Fecal microbiota transfer (FMT) experiments***

532 FMT was performed in male C3H/HeN germ-free 8-week-old mice (Design of the  
533 experimental study in SI Appendix, Fig. S1). Mice were maintained in gnotobiotic isolators,  
534 *ad lib* for 1 week of acclimation to the laboratory conditions followed by fecal microbiota  
535 transfer (FMT, day 0) then by 7 or 14 weeks of follow-up, after which the mice were  
536 sacrificed. The fecal microbiota donors (SI Appendix, Table S11) were 9 consecutive  
537 individuals with normal and 9 consecutive patients with CRC at colonoscopy from cohorts.  
538 They were considered in the current study as normal microbiota (N- $\mu$ ) and CRC microbiota  
539 (CRC- $\mu$ ) donors, were informed about the experimental study and accepted giving additional  
540 stools, if needed, for animal experimentation during the study period. Fresh stools were  
541 given by oral gavage, as follows: n=53 in 7-week study (CRC- $\mu$  transfer, n=30, and N- $\mu$   
542 transfer, n=23) and n=132 in 14-week study (CRC- $\mu$  transfer, n=66, and N- $\mu$  transfer, n=66).  
543 In addition to the FMT, the mice were given intraperitoneal injections of either, the  
544 carcinogen azoxymethane (AOM; Sigma; 8 mg/kg body weight once a week for three or 10  
545 weeks), here chosen as a potential environmental exposure, or saline. In the 14-week study,  
546 24 animals were given also polyethylene glycol (PEG3350) at a non-laxative dosage (1%) in  
547 their drinking water to impede adherence of bacteria to the gut mucosa.

548 Associations between mice colonic tissue events, type of transferred microbiota (CRC- $\mu$  vs.  
549 N- $\mu$ ), chemical product given (AOM vs. saline), and bacterial composition in stools, were  
550 assessed to evaluate potential associations by using Spearman's tests. The experimental

551 procedures were approved by the local ethics committee (Committee-reference number 45;  
552 approval #12/076).

553 ***Cytokine and chemokine and cell signal measurements in mouse colonic mucosa in vivo***

554 cDNA synthesis from total RNA extracted from colon mucosa scrapings of mice, were  
555 quantified using SYBR Green PCR Master Mix cDNA with the Light Cycler 480 System  
556 (Roche Diagnostics) according to the manufacturer's instructions. The mRNA levels each  
557 group (n=3) quantified and transcripts involved in cell renewal and those of involved in  
558 inflammation (TNF $\alpha$ , MIP2, IL10, IL1 $\beta$ , IL6, IL17 and IL17R) were analyzed. Three  
559 housekeeping genes (GADPH, HRPT1, and TBP) were used. The comparative  $\Delta\Delta$ CT-  
560 method was used for relative mRNA quantification of target genes, normalized to GAPDH  
561 and a relevant control equal to  $2^{-\Delta\Delta$ CT. Relative levels of genes involved in cell renewal  
562 (*Kfl4*, *Elf3*, *MATH1* and *Hes1*) or encoding proinflammatory cytokines and chemokines  
563 (TNF $\alpha$ , MIP2, IL10, IL1 $\beta$ , IL6, IL17, and IL17R) were analyzed using quantitative real-time  
564 PCR (qRT-PCR) as previously described (38).

565

566 **Contribution:**

567 Iradj Sobhani: clinical trials, data collection, data analyses, funding, manuscript preparation  
568 (writing and revising); corresponding author  
569 Emma Bergsten: animal and experimental data, bacterial cultures and analysis, and  
570 manuscript revising  
571 Severine Couffin: animal and experimental data, histology, and manuscript revising  
572 Aurelien Amiot: clinical data, monitoring, and manuscript revising  
573 Biba Nebbad: microbiology, cultures, and manuscript revising  
574 Caroline Barau: tissue and body fluid collection and quality assessment, manuscript revising  
575 Nicola de'Angelis: surgery and post-surgical monitoring and clinical data, and manuscript  
576 revising  
577 Sylvie Rabot: animal experiments, data collection, biochemical assays, and manuscript  
578 revising  
579 Florence Canoui Poitrine: clinical data and statistical analysis, manuscript revising  
580 Denis Mestivier: bioanalysis of data, bioinformatics analysis and manuscript reviewing  
581 Thierry Pédrón: bacterial and cellular studies and manuscript revising  
582 Khashayarsha Khazaie: inflammatory and immunology screening in mice and human tissues  
583 and manuscript revising  
584 Philippe Sansonetti: head of the laboratory, microbial function and manuscript revising

585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
  
625

### Acknowledgments

We thank all patients for their participation and physicians who invited them to participate; they are 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; 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 and V. Jarrousse, A. Bado, JP Fouret for animal experiments and help managing; A. Wolfe for revising the English; and 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 Lab) for 16S rRNA analysis. Prof. Sylvain Loric for blood test monitoring in mice; and Catherine Philippe for SCFA analyses in mice All technicians and scientific consultants from Faculté de Médecine site Pitié Salpêtrière (Paris-APHP), UMS 29 Omic Platform p3s for technical help on EPIC methylation array study in mice and all members of clinical monitoring group from URC EST- CHU St Antoine, Paris 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. Data availability: All data discussed in the paper will be made available to readers.

**Funding:** French Institut of Cancer and Ministry of Health (grant PHRC 2011-VatnimadaOM09268) and French Society of Gastroenterology -SNFGE (grant for fecal test screening); LNCC (Ligue Nationale Contre le Cancer for fecal test screening) and INCA (Cancéropôle Ile de France, Grant for Microbiota and CRC), INSERM partially financed under ITMO program.

*Sponsor:* Assistance Publique Hôpitaux de Paris (APHP), INSERM

### Conflict of interest statement

None of authors have any conflicts of interest to declare regarding this study; some of authors share rights in three patents: EP B31120, EP2635705, and EP 2955232 A1 20151216 based on methods for diagnosing adenomas and/or colorectal cancer (CRC).

### References

1. K.Z. Guyton, D. Loomis, Y. Grosse, F. El Ghissassi, L. Benbrahim-Tallaa, N. Guha, C. Scocciati, H. Mattock, K. Straif; International Agency for Research on Cancer Monograph Working Group, IARC, Lyon, France. Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate. *Lancet Oncol.* **16**(5), 490-491 (2015). doi: 10.1016/S1470-2045(15)70134-8.

- 631 2. G. López-Abente, J. García-Pérez, P. Fernández-Navarro, E. Boldo, R. Ramis. Colorectal  
632 cancer mortality and industrial pollution in Spain. *BMC Public Health*. **12**, 589 (2012). doi:  
633 10.1186/1471-2458-12-589.
- 634 3. C.P. Wild. Complementing the genome with an "exposome": the outstanding challenge of  
635 environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol*  
636 *Biomarkers Prev.* **14**(8), 1847-1850 (2005).
- 637 4. A. Ghantous, H. Hernandez-Vargas, G. Byrnes, T. Dwyer, Z. Herceg. Characterising the  
638 epigenome as a key component of the fetal exposome in evaluating in utero exposures and  
639 childhood cancer risk. *Mutagenesis*. **30**(6),733-742 (2015). doi: 10.1093/mutage/gev010.
- 640 5. J. Guinney, R. Dienstmann, X. Wang A. de Reyniès, A. Schlicker, C. Soneson, L. Marisa,  
641 P. Roepman, G. Nyamundanda, P. Angelino, B.M. Bot, J.S. Morris, I.M. Simon, S. Gerster,  
642 E. Fessler, F. De Sousa E Melo, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru,  
643 G.C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J.W. Gray, D.  
644 Hanahan, J. Tabernero, R. Bernards, S.H. Friend, P. Laurent-Puig, J.P. Medema, A.  
645 Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, S. Tejpar. The consensus  
646 molecular subtypes of colorectal cancer. *Nat Med.* **21**(11), 1350-1356 (2015). doi:  
647 10.1038/nm.3967.
- 648 6. S.J. O'Keefe, J.V. Li, L. Lahti, J. Ou, F. Carbonero, K. Mohammed, J.M. Posma, J.  
649 Kinross, E. Wahl, E. Rude, K. Vipperla, V. Naidoo, L. Mtshali, S. Tims, P.G. Puylaert, J.  
650 DeLany, A. Krasinskas, A.C. Benefiel, H.O. Kaseb, K. Newton, J.K. Nicholson, W.M. de  
651 Vos, H.R. Gaskins, E.G. Zoetendal. Fat, fibre and cancer risk in African Americans and rural  
652 Africans. *Nat Commun.* **6**, 6342(2015). doi: 10.1038/ncomms7342.
- 653
- 654 7. E.D. Sonnenburg, S.A. Smits, M. Tikhonov, S.K. Higginbottom, N.S. Wingreen, J.L.  
655 Sonnenburg. Diet-induced extinctions in the gut microbiota compound over generations.  
656 *Nature.* **529**(7585), 212-215 (2016). doi: 10.1038/nature16504.
- 657
- 658 8. L.A. David, C.F. Maurice, R.N. Carmody, D.B. Gootenberg, J.E. Button, B.E. Wolfe,  
659 A.V. Ling, A.S. Devlin, Y. Varma, M.A. Fischbach, S.B. Biddinger, R.J. Dutton, P.J.  
660 Turnbaugh. Diet rapidly and reproducibly alters the human gut microbiome. *Nature.*  
661 **505**(7484), 559-563 (2014). doi: 10.1038/nature12820
- 662
- 663 9. I. Sobhani, J. Tap, F. Roudot-Thoraval, J.P. Roperch, S. Letulle, P. Langella, G. Corthier,  
664 J. Tran Van Nhieu, J.P. Furet. Microbial dysbiosis in colorectal cancer (CRC) patients.  
665 *PLoS One.* **6**(1), e16393 (2011). doi: 10.1371/journal.pone.0016393.
- 666
- 667 10. G. Zeller, J. Tap, A.Y. Voigt, S. Sunagawa, J.R. Kultima, P.I. Costea, A. Amiot, J.  
668 Böhm, F. Brunetti, N. Habermann, R. Hercog, M. Koch, A. Luciani, D.R. Mende, M.A.  
669 Schneider, P. Schrotz-King, C. Tournigand, J. Tran Van Nhieu, T. Yamada, J. Zimmermann,  
670 V. Benes, M. Kloor, C.M. Ulrich, M. von Knebel Doeberitz, I. Sobhani, P. Bork. Potential of  
671 fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol.* **10**, 766 (2014).  
672 doi: 10.15252/msb.20145645.
- 673

- 674 11. C.M. Dejea, P. Fathi, J.M. Craig, A. Boleij, R. Taddese, A.L. Geis, X. Wu, C.E.  
675 DeStefano Shields, E.M. Hechenbleikner, D.L. Huso, R.A. Anders, F.M. Giardiello, E.C.  
676 Wick, H. Wang, S. Wu, D.M. Pardoll, F. Housseau, C.L. Sears. Patients with familial  
677 adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science*.  
678 **359**(6375), 592-597 (2018). doi: 10.1126/science.aah3648.  
679
- 680 12. A. Amiot, H. Mansour, I. Baumgaertner, J.C. Delchier, C. Tournigand, J.P. Furet, J.P.  
681 Carrau, F. Canoui-Poitaine, I. Sobhani; CRC group of Val De Marne. The detection of the  
682 methylated Wif-1 gene is more accurate than a fecal occult blood test for colorectal cancer  
683 screening. *PLoS One*. **9**(7), e99233 (2014). doi: 10.1371/journal.pone.0099233.  
684
- 685 13. S.H. Wong, L. Zhao, X. Zhang, G. Nakatsu, J. Han, W. Xu, X. Xiao, T.N.Y. Kwong, H.  
686 Tsoi, W.K.K. Wu, B. Zeng, F.K.L. Chan, J.J.Y. Sung, H. Wei, J. Yu. Gavage of Fecal  
687 Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-  
688 Free and Conventional Mice. *Gastroenterology*. **153**(6):1621-1633.e6 (2017). doi:  
689 10.1053/j.gastro.2017.08.022.
- 690 14. A.M. Thomas, P. Manghi, F. Asnicar, E. Pasolli, F. Armanini, M. Zolfo, F. Beghini, S.  
691 Manara, N. Karcher, C. Pozzi, S. Gandini, D. Serrano, S. Tarallo, A. Francavilla, G. Gallo,  
692 M. Trompetto, G. Ferrero, S. Mizutani, H. Shiroma, S. Shiba, T. Shibata, S. Yachida, T.  
693 Yamada, J. Wirbel, P. Schrotz-King, C.M. Ulrich, H. Brenner, M. Arumugam, P. Bork, G.  
694 Zeller, F. Cordero, E. Dias-Neto, J.C. Setubal, A. Tett, B. Pardini, M. Rescigno, L. Waldron,  
695 A. Naccarati, N. Segata. Metagenomic analysis of colorectal cancer datasets identifies cross-  
696 cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med*.  
697 **25**(4):667-678 (2019). doi: 10.1038/s41591-019-0405-7.
- 698 15. J. Wirbel, P.T. Pyl, E. Kartal, K. Zych, A. Kashani, A. Milanese, J.S. Fleck, A.Y. Voigt,  
699 A. Palleja, R. Ponnudurai, S. Sunagawa, L.P. Coelho, P. Schrotz-King, E. Vogtmann, N.  
700 Habermann, E. Niméus, A.M. Thomas, P. Manghi, S. Gandini, D. Serrano, S. Mizutani, H.  
701 Shiroma, S. Shiba, T. Shibata, S. Yachida, T. Yamada, L. Waldron, A. Naccarati, N. Segata,  
702 R. Sinha, C.M. Ulrich, H. Brenner, M. Arumugam, P. Bork, G. Zeller. Meta-analysis of fecal  
703 metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat*  
704 *Med*. **25**(4):679-689 (2019). doi: 10.1038/s41591-019-0406-6.  
705
- 706 16. R.V. Purcell, M. Visnovska, P.J. Biggs, S. Schmeier, F.A. Frizelle. Distinct gut  
707 microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Sci*  
708 *Rep*. **7**(1), 11590 (2017). doi: 10.1038/s41598-017-11237-6.  
709
- 710 17. S.I. Halkjær, A.H. Christensen, B.Z.S. Lo, P.D. Browne, S. Günther, L.H. Hansen, A.M.  
711 Petersen. Faecal microbiota transplantation alters gut microbiota in patients with irritable  
712 bowel syndrome: results from a randomised, double-blind placebo-controlled study. *Gut*.  
713 **67**(12), 2107-2115 (2018). doi: 10.1136/gutjnl-2018-316434.  
714
- 715 18. V.C. Lombardi, K.L. De Meirleir, K. Subramanian, S.M. Nourani, R.K. Dagda, S.L.  
716 Delaney, A. Palotás. Nutritional modulation of the intestinal microbiota; future  
717 opportunities for the prevention and treatment of neuroimmune and neuroinflammatory  
718 disease. *J Nutr Biochem*. **61**, 1-16 (2018). doi: 10.1016/j.jnutbio.2018.04.004.

719

720 19. T.O. Keku, S. Dulal, A. Deveaux, B. Jovov, X. Han. The gastrointestinal microbiota and  
721 colorectal cancer. *Am J Physiol Gastrointest Liver Physiol.* **308**(5), G351-G363 (2015). doi:  
722 10.1152/ajpgi.00360.2012.

723

724 20. R.F. Schwabe, C. Jobin. The microbiome and cancer. *Nat Rev Cancer.* **13**(11), 800-812  
725 (2013). doi: 10.1038/nrc3610.

726

727 21. K.B. Islam, S. Fukiya, M. Hagio, N. Fujii, S. Ishizuka, T. Ooka, Y. Ogura, T. Hayashi,  
728 A. Yokota. Bile acid is a host factor that regulates the composition of the cecal microbiota  
729 in rats. *Gastroenterology.* **141**(5), 1773-1781 (2011). doi: 10.1053/j.gastro.2011.07.046.

730

731 22. P. Rosignoli, R. Fabiani, A. De Bartolomeo, F. Spinuzzi, E. Agea, M.A. Pelli, G.  
732 Morozzi. Protective activity of butyrate on hydrogen peroxide-induced DNA damage in  
733 isolated human colonocytes and HT29 tumour cells. *Carcinogenesis.* **22**(10), 1675-1680  
734 (2001).

735

736 23. J. Winter, L. Nyskohus, G.P. Young, Y. Hu, M.A. Conlon, A.R. Bird, D.L. Topping,  
737 R.K. Le Leu. Inhibition by resistant starch of red meat-induced promutagenic adducts in  
738 mouse colon. *Cancer Prev Res (Phila).* **4**(11), 1920-1928 (2011). doi: 10.1158/1940-  
739 6207.CAPR-11-0176.

740

741 24. D.L. Worthley, V.L. Whitehall, R.K. Le Leu, N. Irahara, R.L. Buttenshaw, K.A. Mallitt,  
742 S.A. Greco, I. Ramsnes, J. Winter, Y. Hu, S. Ogino, G.P. Young, B.A. Leggett. DNA  
743 methylation in the rectal mucosa is associated with crypt proliferation and fecal short-chain  
744 fatty acids. *Dig Dis Sci.* **56**(2), 387-396 (2011). doi: 10.1007/s10620-010-1312-4.

745

746 25. J. Chen, X.F. Huang. The signal pathways in azoxymethane-induced colon cancer and  
747 preventive implications. *Cancer Biol Ther.* **8**(14), 1313-1317 (2009).

748

749 26. S. Wu, K.J. Rhee, E. Albesiano, S. Rabizadeh, X. Wu, H.R. Yen, D.L. Huso, F.L.  
750 Brancati, E. Wick, F. McAllister, F. Housseau, D.M. Pardoll, C.L. Sears CL. A human  
751 colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell  
752 responses. *Nat Med.* **15**(9):1016-1022 (2009). doi: 10.1038/nm.2015.

753

754 27. I.P. Pogribny, I. Rusyn. Environmental toxicants, epigenetics, and cancer. *Adv Exp Med*  
755 *Biol.* **754**, 215-232 (2013). doi: 10.1007/978-1-4419-9967-2\_11.

756

757 28. L. Hou, X. Zhang, L. Tarantini, F. Nordio, M. Bonzini, L. Angelici, B. Marinelli, G.  
758 Rizzo, L. Cantone, P. Apostoli, P.A. Bertazzi, A. Baccarelli. Ambient PM exposure and  
759 DNA methylation in tumor suppressor genes: a cross-sectional study. *Part Fibre Toxicol.* **8**,  
760 25 (2011). doi: 10.1186/1743-8977-8-25.

761

- 762 29. M.J. Blow, T.A. Clark, C.G. Daum, A.M. Deutschbauer, A. Fomenkov, R. Fries, J.  
763 Froula, D.D. Kang, R.R. Malmstrom, R.D. Morgan, J. Posfai, K. Singh, A. Visel, K.  
764 Wetmore, Z. Zhao, E.M. Rubin, J. Korlach, L.A. Pennacchio, R.J. Roberts. The Epigenomic  
765 Landscape of Prokaryotes. *PLoS Genet.* **12**(2), e1005854 (2016). doi:  
766 10.1371/journal.pgen.1005854.  
767
- 768 30. Y.D. Cho, P.J. Kim, H.G. Kim, Y.J. Seol, Y.M. Lee, Y. Ku, I.C. Rhyu, H.M. Ryoo.  
769 Transcriptomics and methylomics in chronic periodontitis with tobacco use: a pilot study.  
770 *Clin Epigenetics.* **9**, 81 (2017). doi: 10.1186/s13148-017-0381-z.  
771
- 772 31. R. Feil, M.F. Fraga. Epigenetics and the environment: emerging patterns and  
773 implications. *Nat Rev Genet.* **13**(2), 97-109 (2012). doi: 10.1038/nrg3142.  
774
- 775 32. K.J. Howell, J. Kraiczy, K.M. Nayak, M. Gasparetto, A. Ross, C. Lee, T.N. Mak, B.K.  
776 Koo, N. Kumar, T. Lawley, A. Sinha, P. Rosenstiel, R. Heuschkel, O. Stegle, M. Zilbauer.  
777 DNA Methylation and Transcription Patterns in Intestinal Epithelial Cells From Pediatric  
778 Patients With Inflammatory Bowel Diseases Differentiate Disease Subtypes and Associate  
779 With Outcome. *Gastroenterology.* **154**(3), 585-598 (2018). doi:  
780 10.1053/j.gastro.2017.10.007.  
781
- 782 33. L. Chen, B. Jiang, C. Zhong, J. Guo, L. Zhang, T. Mu, Q. Zhang, X. Bi.  
783 Chemoprevention of colorectal cancer by black raspberry anthocyanins involved the  
784 modulation of gut microbiota and SFRP2 demethylation. *Carcinogenesis.* **39**(3), 471-481  
785 (2018). doi: 10.1093/carcin/bgy009.  
786
- 787 34. S. Garrigou, G. Perkins, F. Garlan, C. Normand, A. Didelot, D. Le Corre, S. Peyvandi,  
788 C. Mulot, R. Niarra, P. Aucouturier, G. Chatellier, P. Nizard, K. Perez-Toralla, E. Zonta, C.  
789 Charpy, A. Pujals, C. Barau, O. Bouché, J.F. Emile, D. Pezet, F. Bibeau, J.B. Hutchison,  
790 D.R. Link, A. Zaanani, P. Laurent-Puig, I. Sobhani, V. Taly. A Study of Hypermethylated  
791 Circulating Tumor DNA as a Universal Colorectal Cancer Biomarker. *Clin Chem.* **62**(8),  
792 1129-1139 (2016). doi: 10.1373/clinchem.2015.253609.  
793
- 794 35. G.D. Hannigan, M.B. Duhaime, M.T. Ruffin, C.C Koumpouras, P.D. Schloss.  
795 Diagnostic potential and interactive dynamics of the colorectal cancer virome. *mBio*  
796 **9**:e02248-18 (2018). doi.org/10.1128/mBio.02248-18  
797
- 798 36. J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello,  
799 N. Fierer, A.G. Peña, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E.  
800 Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R.  
801 Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R.  
802 Knight. QIIME allows analysis of high-throughput community sequencing data. *Nat*  
803 *Methods.* **7**(5), 335-336 (2010). doi: 10.1038/nmeth.f.303.  
804
- 805 37. Y. Huang, M.S. Pepe, Z. Feng. Logistic regression analysis with standardized markers.  
806 *Ann Appl Stat.* **7**(3) (2013). doi: 10.1214/13-AOAS634SUPP.  
807

- 808 38. M. Abolhassani, N. Aloulou, M.T. Chaumette, T. Aparicio, N. Martin-Garcia, H.  
809 Mansour, S. Le Gouvello, J.C. Delchier, I. Sobhani. Leptin receptor-related immune  
810 response in colorectal tumors: the role of colonocytes and interleukin-8. *Cancer Res.* **68**(22),  
811 9423-3942 (2008). doi: 10.1158/0008-5472.CAN-08-1017.  
812
- 813 39. M. Needhamsen, E. Ewing, H. Lund, D. Gomez-Cabrero, R.A. Harris, L. Kular, M.  
814 Jagodic. Usability of human Infinium MethylationEPIC BeadChip for mouse DNA  
815 methylation studies. *BMC Bioinformatics.* **18**(1), 486 (2017). doi: 10.1186/s12859-017-  
816 1870-y.  
817

818 **Figures legends**

819

820 **Figure 1. Histological patterns of murine colonic mucosa following FMT**

821 After the intestine was removed *en bloc* from caecum to anus, mucosa was carefully pinned  
822 flat, without folds, to examine the totality of the colonic mucosa which were stained with  
823 0.2% methylene blue (left slides of couple slides A to D) or HES (right slides of couple  
824 slides A to D. Numbers of mice were as follows: n=53 in 7-week study (CRC- $\mu$  transfer,  
825 n=30, and N- $\mu$  transfer, n=23) and n=132 in 14-week study (CRC- $\mu$  transfer, n=66, and N- $\mu$   
826 transfer, n=66). Bar scale: 50  $\mu$ m.

827 **A)** After FMT from healthy human controls (N- $\mu$ ), no aberrant crypt foci (ACF) were visible  
828 in the colonic mucosa (here after 14 weeks in left) and pattern of crypts was normal by HES  
829 staining (right).

830 **B)** Elevated numbers of ACF were observed after FMT from patients with colorectal cancer  
831 (CRC- $\mu$ ) as compared to N- $\mu$  with multiple ACF (arrow in left slide) as verified by HES  
832 (right slide).

833 **C)** ACF counts were higher in the animals given azoxymethane (AOM) as compared to N- $\mu$ .  
834 Arrows indicate double ACF under bleu coloration (left slide) and illustration on HES slide  
835 of colonic mucosa.

836 **D)** The combination of CRC- $\mu$  and AOM increased the ACF count with dysplasia in rare  
837 cases (arrow).

838 **E)** No ACF no inflammatory cell infiltrate was visible in mice given N- $\mu$  under NaCl (HES)  
839 although FISH staining showed density of bacteria trapped in the mucus layer (arrow, left  
840 slide).

841 **F)** No ACF, no inflammatory nor injury were noticed when mice received polyethylene  
842 glycol (PEG) in their drinking water as shown by H&S staining (HES) when FISH staining  
843 shows clear decrease in density of bacteria trapped in the proximal mucus layer (arrow)

844 **G)** Representative pictures of KI67 staining after human FMT from patients with colorectal  
845 cancer (CRC- $\mu$ ) or from N- $\mu$  recipients.

846 **H)** Cell Proliferation assessed by Ki67 staining and ACF quantification 7 and 14 weeks after  
847 FMT in mice in the intestinal mucosa were higher in CRC- $\mu$  than in N- $\mu$  recipients.

848 **I)** The number of ACF counts were enhanced depending on the length of mucosa examined  
849 and the mice subgroups.

850 **J)** Comparative transcriptional levels of a set of inflammatory cytokines in the colonic  
851 mucosa assessed by murine cytokine qPCR quantification showing a trend to higher IL1,  
852 IL6, MIP2, IL17 and lower IFN $\gamma$ , IL10 and IL23 in mice given CRC- $\mu$  alone (fold vs N- $\mu$   
853 given mice) with AOM boosting this effect that reached significance for IL6, TNF $\alpha$ , IL10 in  
854 mice given CRC- $\mu$ +AOM (vs. N- $\mu$ +AOM).

855

856 **K)** Inflammatory cell infiltrate in the colonic mucosa as assessed by semi quantification on  
857 HES stained slides (ten consecutive fields) under optic microscope magnification 20: a  
858 pathologist blinded to animal groups used a semi-quantitative score to evaluate myeloid cell  
859 infiltrate in the colonic mucosa as 0, 1, and 2 indicating absence, scarce, and numerous  
860 inflammatory cells, respectively. The groups were compared by one-way ANOVA followed  
861 by the Tukey-Kramer multiple comparisons post-hoc test. And no significant difference in  
862 between mice groups was observed

863 \* $P$ <0.05; NS: not significant

864

## 865 **Figure 2. DNA changes in mice after human FMT**

866 Whole genome sequencing of total DNA extracted from colonic mucosa (n=12) and spleen  
867 samples (n=6) were performed. **A.** All single-nucleotide polymorphisms (SNPs) in the colon  
868 and spleen samples are compared to the reference mouse genome (GRCm38) and mutations  
869 levels within gene segments are indicated. Mice given AOM showed the highest levels of  
870 DNA alterations in exons or introns with a trend of additive effect of CRC- $\mu$  as compared to  
871 N- $\mu$ . **B)** Distribution of animal subgroups according to the total gene mutations. DNA  
872 changes in single-nucleotides with Pca scatter diagrams for colonic or spleen samples in the  
873 groups of mice; the groups are identified by the type of human microbiota received (CRC- $\mu$   
874 or N- $\mu$  for CRC patients' and controls' stool, respectively) and type of treatment (AOM or  
875 NaCl for azoxymethane or saline, respectively). **C)** Correlation circle of targeted gene  
876 mutations in the colonic mucosa and spleen tissues according to Pca; vector length reflects  
877 targeted gene mutation weight in the first two component analyses; targeted mutated genes  
878 are indicated (those of Wnt pathway in red color). **D)** When mutations in all Wnt genes were  
879 pooled together, rates of mutations were significantly higher in mice given AOM with an  
880 additive effect of CRC- $\mu$ . The total number of mutations in *Wnt* pathway genes in both  
881 colonic mucosa and spleen (Sp) was the highest in the animals given the CRC- $\mu$  and AOM

882 combination (see also Figure S6). The number of mutations was greater in colonic mucosa  
883 (but not in spleen tissues) with AOM combined with CRC microbiota compared to AOM  
884 combined with control microbiota. There was no significant effect in between colonic  
885 mucosa due to CRC- $\mu$  alone as compared to N- $\mu$  alone. Col, colonic mucosa; Sp, spleen.

886 **E)** DNA epigenetic changes were investigated by using mEPIC array (39). The methylation  
887 level of probes (n=63 987) were estimated after bisulfite modification of DNAs. Changes  
888 based on the methylation of probes were investigated on DNAs from colon samples (n=16; 4  
889 mice from each experimental group): the level of methylated probes was quantified as  
890 reported (39) and ranged from 0 (not methylated) to 1 (fully methylated). DNAs were  
891 classified as unmethylated if the methylation value was  $<0.2$  and as hypermethylated if the  
892 methylation value was  $>0.799$ . Overall, mean and median values of all probes pooled  
893 (n=63 987) in each group of mice showed lowest values in the group of CRC- $\mu$  + AOM  
894 recipients and highest in control microbiota recipients.

895 **F)** Most of the probes were unmethylated in all animal groups; elevated numbers of both  
896 hypomethylated and hypermethylated probes were observed in the mice given CRC- $\mu$  +  
897 AOM.

898 **G)** Probes whose methylation level changed or remained unchanged under AOM in mice  
899 given N- $\mu$ - or CRC- $\mu$ . Mice receiving CRC- $\mu$  had a greater number of genes with changed  
900 methylation levels

901 **H)** Number of hypermethylated probes in each group showing highest level in mice given  
902 CRC- $\mu$  and AOM combination compared to all other groups.

903 \*\* $P < 0.01$ , \* $P < 0.05$ ; NS: not significant

904

905

906 **Figure 3. Identification of hypermethylated genes related to fecal microbiota in human**

907 Overview on the strategy from experimental approach for the validation of gene methylated  
908 targets in human based on microbiota donors (CRC patients or controls) in germ-free mice  
909 experiments. Methylated genes in CRC-associated tissues and fluids were identified based  
910 on their power for showing differences between normal colonoscopy individuals and CRC  
911 patients. **A)** Human tissues and effluents were submitted to methylation gene array. Based on  
912 significant differences of methylation values in CpG probes between controls (n=9) and  
913 CRC patients (n=9) donors, genes were selected according to the promoter segments  
914 hypermethylated in CRC patients. **B)** Bidimensional (right panel) and tridimensional (left

915 panel) distribution of genes regarding the difference in methylation values are indicated; in  
916 red color are indicated 7 selected more discriminant genes regarding CRC patients and  
917 controls; D: difference.

918

919 **Figure 4. Distribution of bacteria in fecal microbiota from patients with colorectal**  
920 **cancer (CRC) and controls with normal colonoscopy findings**

921

922 Overall, 348 individuals from CCR2 cohort (173 asymptomatic individuals enrolled via a  
923 mass CRC-screening program, and 165 patients from “Vatnimad” and “symptomatic” sub-  
924 cohorts, respectively) enrolled. Invasive carcinoma, carcinoma in situ or specific carcinoma  
925 on either flat mucosa or within a polyp in the rectum or colon were defined as CRC (n=177);  
926 controls had no malignancy or significant polyp visible by full colonoscopy (n=171).

927 **A)** Pattern of microbiota clustering according to the diagnosis as assessed by Principal  
928 Coordinate Analysis. The genus-level analysis based on distance matrix variances showed  
929 significant differences between CRC patients and controls. Fecal DNA was subjected to  
930 metagenomic sequencing of the conserved V3-V4 region of the 16S rRNA gene. The  
931 amplicons were purified, quantified, and pooled then sequenced on an Illumina MiSeq  
932 platform. For the analysis of 16S rRNA gene sequences, raw MiSeq FASTQ files were  
933 demultiplexed, quality-filtered using Trimmomatic, and merged. Taxonomic assignments  
934 were performed using Qiime2 (default parameters) with the SILVA-123 database. The  
935 statistical analysis was done with MetagenomeSeq (36).

936 **B)** Pattern of microbiota clustering according to the blood methylation test as assessed by  
937 Principal Coordinate Analysis. Analysis of variance using distance matrices on 789 OTUs  
938 (metagenomeSeq\_1.16.0) from 362 individuals (175 with normal colonoscopy findings and  
939 187 with advanced neoplasia) demonstrated a significant difference between the groups with  
940 positive and negative blood CMI values ( $\geq 2$  and  $\leq 2$ , respectively).

941 **C)** Distribution of genera in fecal microbiota in the groups with positive and negative blood  
942 CMI values ( $\geq 2$  and  $\leq 2$ , respectively); Shaman c3bi platform; Institut Pasteur, Paris, France;  
943 <http://shaman.c3bi.pasteur.fr/>). A maximum likelihood phylogenetic tree was tested (see  
944 Methods). Brown-to-red colors indicate negative CMI results ( $\leq 2$ ) and blue colors indicate  
945 positive CMI results ( $> 2$ ). Note that diversity was less in the group with a positive CMI  
946 compared to the group with a negative CMI.

**Table 1. Characteristics of participants in the validation CCR2 cohort<sup>a</sup>**

		CRC (n=187)	No CRC (n=812)	<i>P</i> value <sup>b</sup>
		Number (%) or mean $\pm$ SD		
Age, years (n=981)		63 [57-71]	60 [53-68]	<0.001
Gender (n=981)	Female	77 (41.2)	392 (48.8)	0.06
Body mass index, kg/m <sup>2</sup>		25.31 [22.86-28.09]	25.31 [22.39-28.7]	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
<b>Blood methylation test</b>				
Cumulative ( <i>Wif1</i> + <i>PENK</i> + <i>NPY</i> ) methylation index	>2	36 (20.3)	36 (4.6)	<0.001
<i>Wif1</i> + <i>PENK</i> + <i>NPY</i>	<i>Continuous</i>	0.16 [0.01-1.16]	0.12 [0.01-0.40]	<0.001
<i>Single gene</i>				
<i>Wif1</i> (n=956)	>2	20 (11.4)	21 (2.7)	<0.001
	<i>Continuous</i>	0.06 [0-0.49]	0.07 [0-0.24]	0.002
<i>PENK</i> (n=956)	>2	3 (1.7)	3 (0.4)	0.045
	<i>Continuous</i>	0 [0-0.04]	0 [0-0.02]	0.052
<i>NPY</i> (n=956)	>2	11 (2.3)	12 (1.5)	<0.001
	<i>Continuous</i>	0 [0-0.03]	0 [0-0]	0.03
CMI <sup>†</sup> in multivariate analysis Adjusted OR* [95% CI]	>2	4.92 [2.79-0.68]		0.005
FIT limited to the Vatnimad sub-cohort (n=468)	>150 ng**	45 (37.8)	20 (5.7)	<0.001
FIT limited to the Vatnimad sub-cohort (n=468)	<i>Continuous</i>	61.66 [1-379]	1 [0-9.67]	<0.001
FIT <sup>†</sup> in multivariate analysis Adjusted OR* [95% CI]				
Negative		reference		
Positive				

Not done	8.69 [4.66-16.21] 0.50[0.33-0.75]	<0.001 0.24

<sup>a</sup> The cohort included 981 individuals from two 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. <sup>b</sup>Student's *t* test and chi-square test for continuous and binary variables, respectively. \*adjusted for age, gender and all variables listed in the table; †Positivity cutoffs were 2.0 for the CMI and 150 ng/mL for the FIT; OR, odds ratio; 95%CI, 95% confidence interval. \*\*concentration of hemoglobin for FIT is given per device.

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

**Comparisons between controls and CRC patients according to tumour staging and cohorts**

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

<i>Shewanella</i>	-11.2	1.13E-3			
<i>Phenylobacterium</i>	11.4	3.71E-03			

\*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.

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= logarithmic value; P values are given after adjustment on "age, gender and BMI" using Bonferroni-corrected Mann-Whitney U test.

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

Species	Mean at baseline	Fold change	Log <sub>2</sub> fold change	P value
Faecalibacterium prausnitzii [1574]	2480.5	2.2e+00	1.159	0.042
Ruminococcus sp. SR1/5 [1621]	2022	2.5e+00	1.354	0.004
Eubacterium hallii [1597]	1942.6	2.2e+00	1.14	0.045
Clostridium sp. L2-50 [1593]	1306.6	3.8e+00	1.945	0.004
Coprococcus comes [1616]	1306.5	2.1e+00	1.107	0.022
Dialister invisus [1259]	1023	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	53	1.3e-01	-2.921	0.00004

---

[1302]

---

Parvimonas sp. oral taxon 110 [1506]	35.6	2.7e-01	-1.86	0.006
---	------	---------	-------	-------

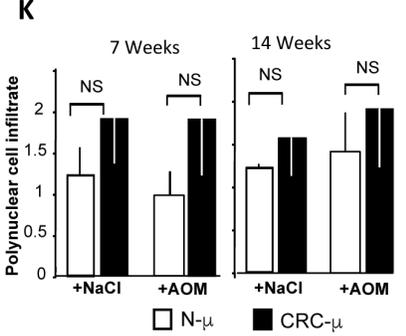
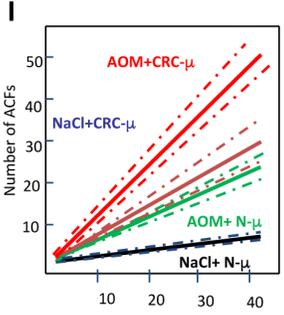
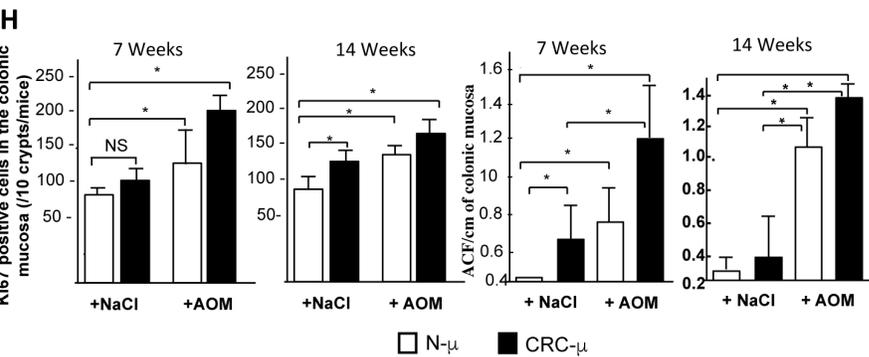
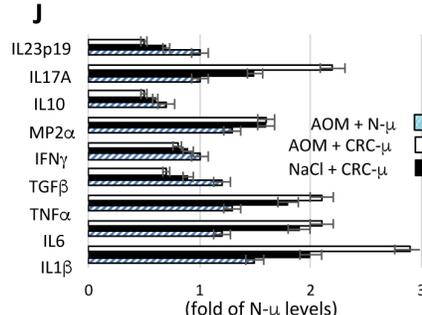
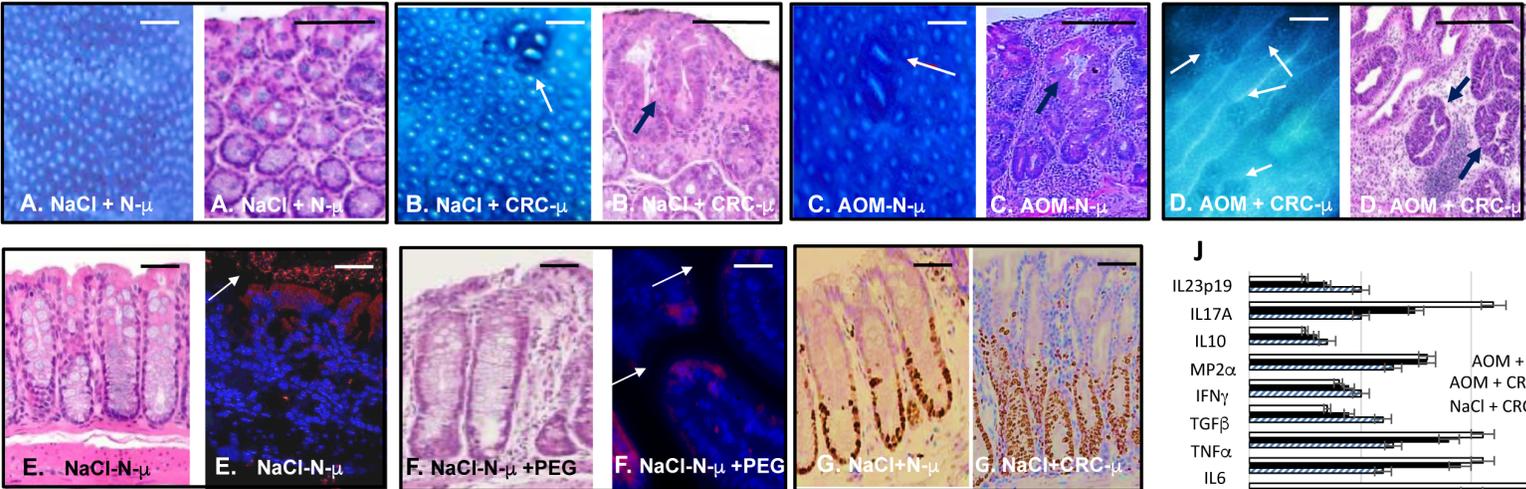
---

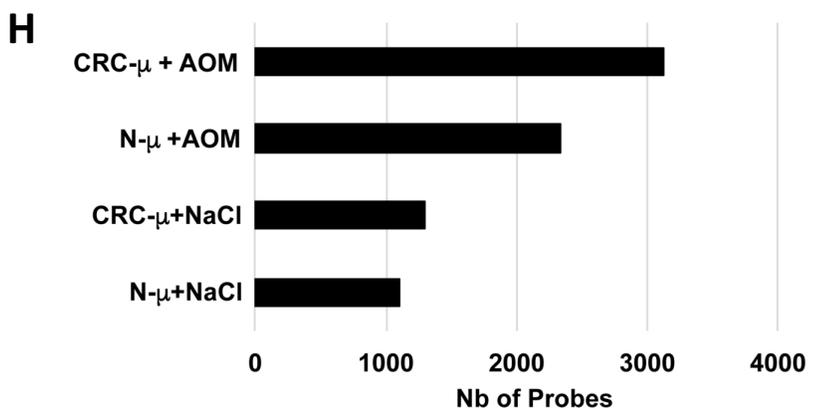
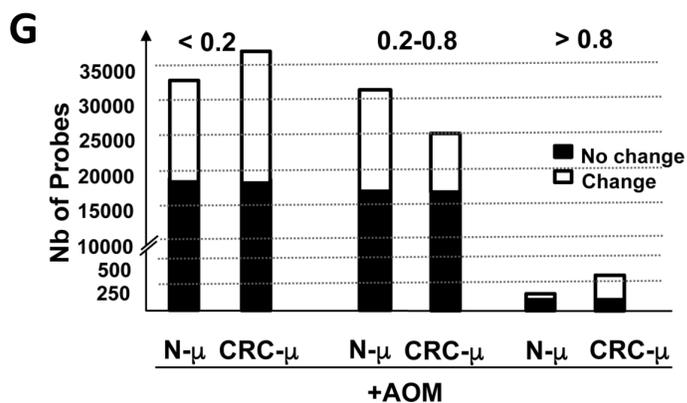
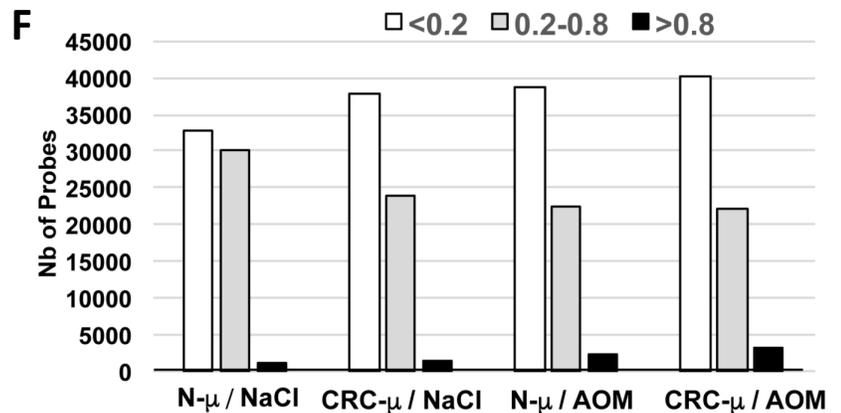
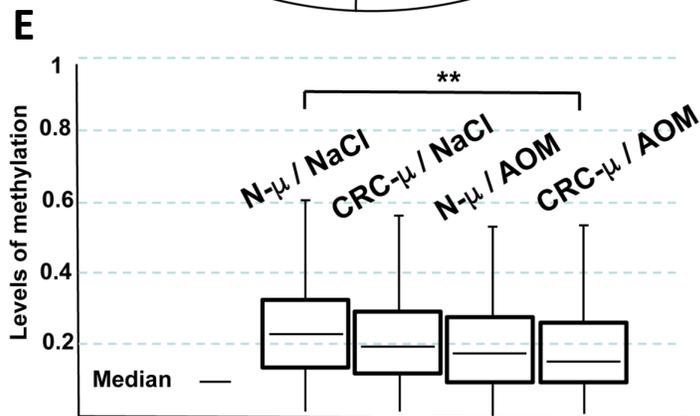
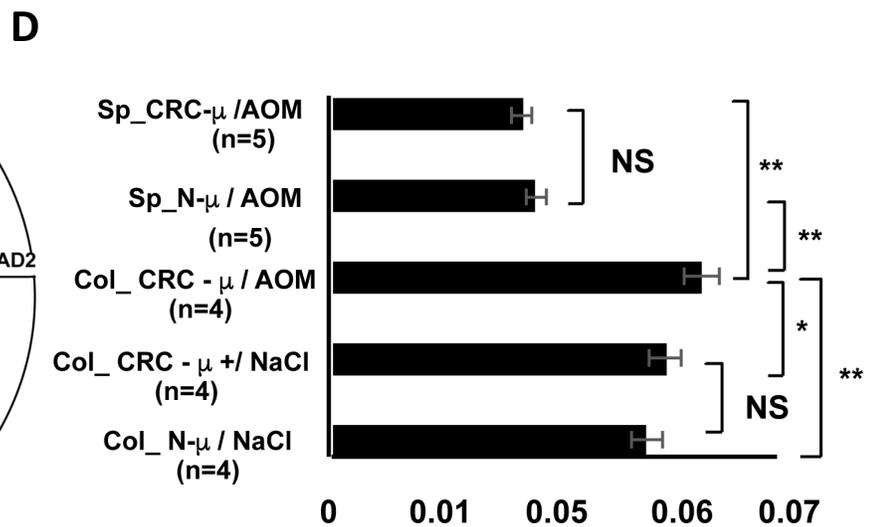
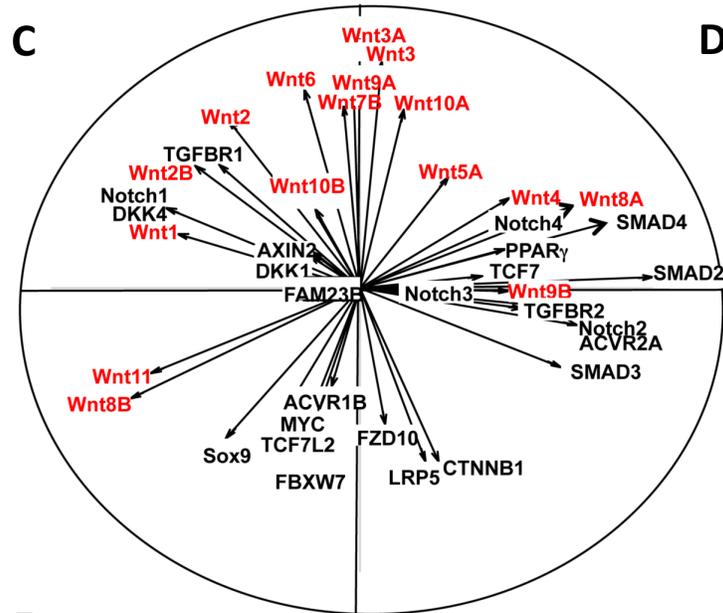
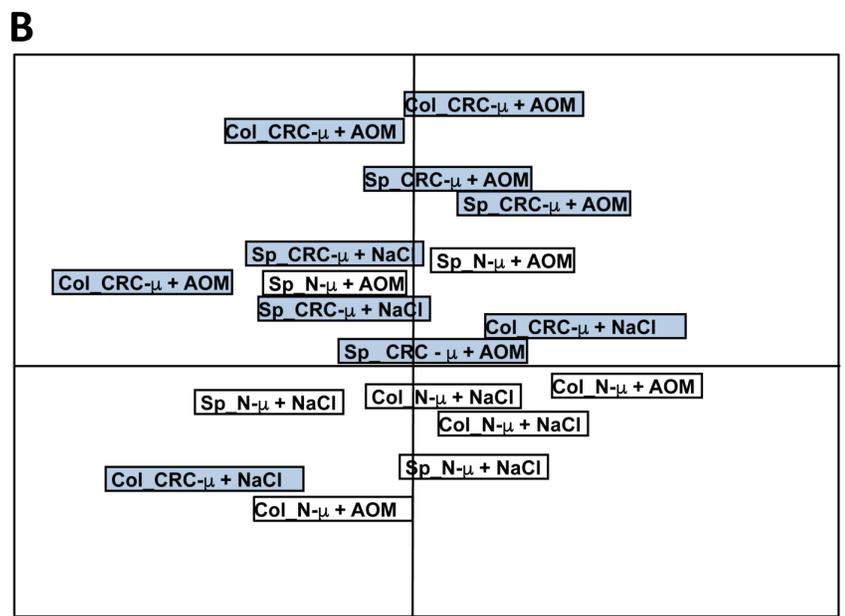
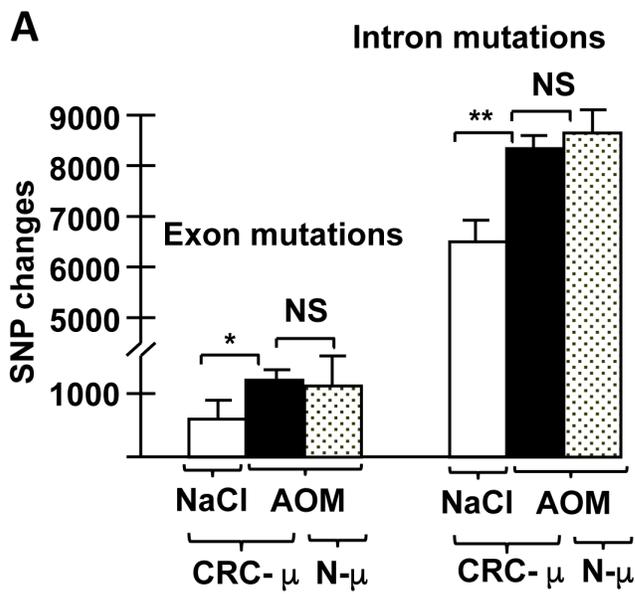
Parvimonas sp. oral taxon 393 [1507]	35	3.2e-01	-1.622	0.026
---	----	---------	--------	-------

---

---

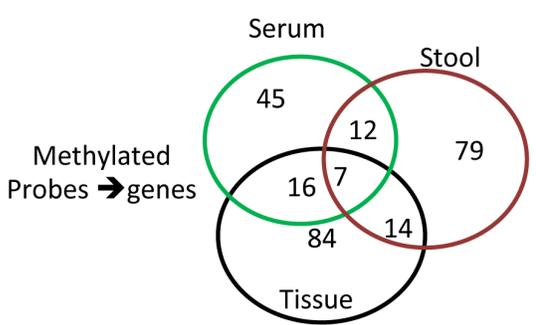
Stool samples were submitted to whole metagenomic sequencing of fecal bacteria DNA (Controls, n=61 and CRC patients, n=53) according to CMI test and results were generated using Shaman C3bi from Institut Pasteur de Paris (<http://shaman.c3bi.pasteur.fr/>). (for methods see Ref 9). log<sub>2</sub>=logarithmic 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 “-“ meaning the bacteria is higher in CRC patients’ microbiota and “+” meaning that the bacteria abundance was higher in controls.





**A** Human Fecal Microbiota Transfer in mice → methylation/demethylation genes  
 Back to the donors' tumours and effluents → using DNA methylation-array

1505 CpG probes → (statistical analysis CRC vs Control) → 195 CpG Probes highly methylated in CRC tissues or effluents



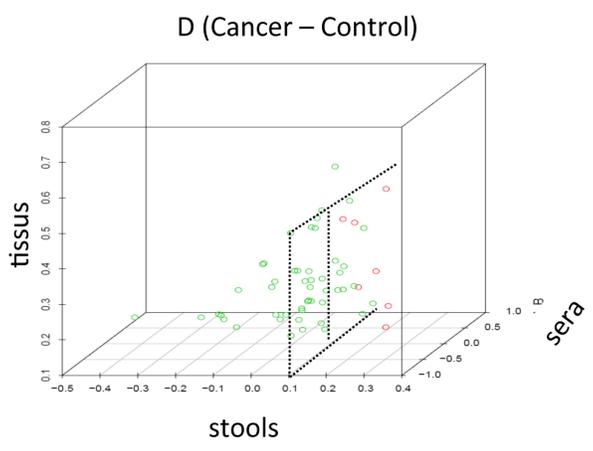
→ Selection for Gene candidates (see B)

- ALX4
- NPY
- PENK
- SEPT9
- SFRP1
- SFRP2
- SFRP3
- Wif1

→ Final selection for pilot and validation studies

- NPY
- PENK
- Wif1

**B**



D (Cancer - Control)

