Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

🔓 OPEN ACCESS | Check for updates
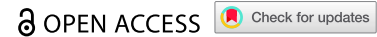
Parvimonas micra, an oral pathobiont associated with colorectal cancer, epigenetically reprograms human colonocytes

Emma Bergsten[a,b], Denis Mestivier[b,c], Francoise Donnadieu[a], Thierry Pedron[a,d], Caroline Barau[e], Landry Tsoumtsa Meda[f], Amel Mettouchi[f], Emmanuel Lemichez[f], Olivier Gorgette[g], Mathias Chamaillard[h], Amaury Vaysse[i], Stevenn Volant[i], Abiba Doukani[j], Philippe J. Sansonetti[a,k]*, Iradj Sobhani[b,l]*, and Giulia Nigro [ID][a,m]*

[a]Unité de Pathogénie Microbienne Moléculaire, INSERM U1202, Institut Pasteur, Paris, France; [b]Équipe universitaire EC2M3-EA7375, Université Paris- Est (UPEC), Créteil, France; [c]Plateforme de Bio-informatique, Institut Mondor de Recherche Biomédicale (IMRB/INSERM U955), Université Paris-Est, Créteil, France; [d]Unité Bactériophage, Bactérie, Hôte, Institut Pasteur, Paris, France; [e]Plateforme de Ressources Biologiques, CHU Henri Mondor Assistance Publique Hôpitaux de Paris (APHP), Créteil, France; [f]Unité des Toxines Bactériennes, Université Paris Cité, CNRS UMR6047, INSERM U1306, Institut Pasteur, Paris, France; [g]Plateforme de Bio-Imagerie Ultrastructurale, Institut Pasteur, Université Paris Cité, Paris, France; [h]Laboratory of Cell Physiology, INSERM U1003, University of Lille, Lille, France; [i]Bioinformatics and Biostatistics Hub, Institut Pasteur, Université Paris Cité, Paris, France; [j]Sorbonne Université, Inserm, Unité Mixte de Service Production et Analyse de données en Sciences de la Vie et en Santé, Paris, France; [k]Chaire de Microbiologie et Maladies Infectieuses, Collège de France, Paris, France; [l]Service de Gastroentérologie, CHU Henri Mondor Assistance Publique Hôpitaux de Paris (APHP), Créteil, France; [m]Microenvironment and Immunity Unit, INSERM U1224, Institut Pasteur, Paris, France
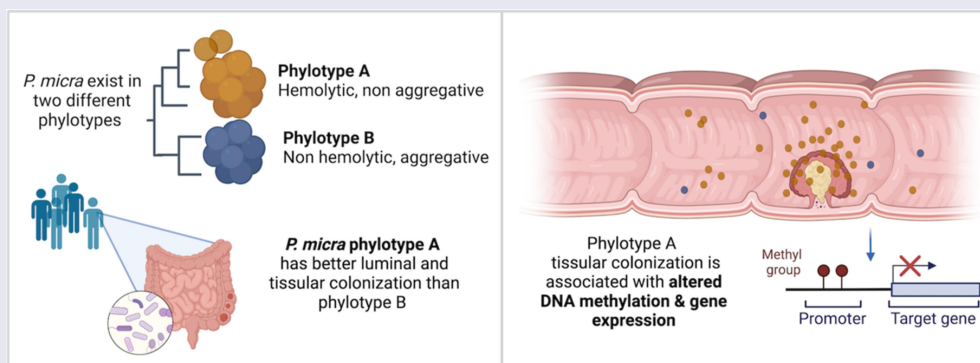
ABSTRACT

Recently, an intestinal dysbiotic microbiota with enrichment in oral cavity bacteria has been described in colorectal cancer (CRC) patients. Here, we characterize and investigate one of these oral pathobionts, the Gram-positive anaerobic coccus *Parvimonas micra*. We identified two phylo-types (A and B) exhibiting different phenotypes and adhesion capabilities. We observed a strong association of phylotype A with CRC, with its higher abundance in feces and in tumoral tissue compared with the normal homologous colonic mucosa, which was associated with a distinct methylation status of patients. By developing an *in vitro* hypoxic co-culture system of human primary colonic cells with anaerobic bacteria, we show that *P. micra* phylotype A alters the DNA methylation profile promoters of key tumor-suppressor genes, oncogenes, and genes involved in epithelial–mesenchymal transition. In colonic mucosa of CRC patients carrying *P. micra* phylotype A, we found similar DNA methylation alterations, together with significant enrichment of differentially expressed genes in pathways involved in inflammation, cell adhesion, and regulation of actin cytoskeleton, providing evidence of *P. micra's* possible role in the carcinogenic process.

Introduction

Colorectal cancer (CRC) is a multifactorial disease due to various genetic and environmental factors that contribute to tumor formation and disease development. Genetic predispositions for CRC, such as Lynch syndrome (also called human no polyposis colonic cancer-HNPCC) or familial adenomatous polyposis (FAP), only account for a minority of CRC, representing less than 5% of all cases.[1] The overwhelming majority of CRCs are sporadic cancers, i.e., the cause is unknown. Several lifestyle environmental factors increase the risk of CRC including physical inactivity,[2] overweight[3] and high consumption of red meat, processed meat, and unsaturated fatty acids.[4] The identification of lifestyle factors supports the hypothesis that the increase in CRC incidence is strictly related to environmental changes. Among environmental factors, the role of microorganisms in cancer has been increasingly recognized, and a global imbalance, called dysbiosis, of gut microbiota compared to healthy individuals was observed in several cancers such as biliary, hepatic, breast cancers, or CRC.[5] Thus, the gut microbiota has emerged as an important carcinogenic environmental factor, in particular for CRC, due to the gut microbiota proximity and constant crosstalk with the intestinal epithelium.[6] Different bacterial species, such as *Bacteroides fragilis*, *Fusobacterium nucleatum (F. nucleatum)*, pathogenic strains of *Escherichia coli* (CoPEC), or *Streptococcus gallolyticus*, have been associated with CRC development, and mechanistic studies have tried to understand roles of these bacteria in carcinogenesis.[7–10] Recently, our team and others have identified, by metagenomic studies from human feces, not only the intestinal bacterium *F. nucleatum* associated with CRC but also different bacteria belonging to the oral microbiota, such as *Gemella morbillorum*, *Solobacterium moorei*, *Porphyromonas gingivalis, and Parvimonas micra* (*P. micra*). The overabundance of these bacteria in the colonic niche may be an etiological factor in CRC,[11] but their role in carcinogenesis remains to be established. Here, we choose to focus on the candidate bacterium, *P. micra.*

P. micra, the only described species within its genus, is a Gram-positive anaerobic coccus and commensal of the oral cavity.[12] *P. micra* could be considered as a pathobiont[13] because it is often isolated from oral polymicrobial infections associated with periodontitis,[14] from superficial polymicrobial infections, including wounds, ulcers, and skin abscesses,[15–19] as well as from deep tissue infections in the brain, lung, and reproductive organs.[20] *P. micra* has been poorly studied due to difficulties in its cultivation and laboratory identification by traditional methods and the polymicrobial nature of the infective zones from which it is usually isolated.

The aim of this work was to better characterize *P. micra*, define its carriage in CRC patients, analyze *in vitro* the nature of the crosstalk between this oral pathobiont and colonic epithelial cells, and determine if the observed impact on primary epithelial cells was relevant in CRC patients.

Results

P. micra exists in two phylotypes with phenotypic and genetic diversity

To explore *P. micra* diversity, we began to study the 27 clinical isolates already available in biobanks, which were collected from various types of infections (blood, systemic, and oral cavity) (Supplementary Table S1). *P. micra* colonies on blood agar plates appear heterogeneous. A contact hemolytic activity on blood agar plates was observed in 20 isolates (20/27) and *P. micra* colonies were found to be either "compact" (7/27) or "non-compact" (20/27) (Figure 1a, Supplementary Fig. S1a). In addition, sedimentation assays on two representative isolates of each phenotype: ATCC 33270 ("non-compact") and HHM BlNa17 ("compact"), (named hereafter *PmA* and *PmB*), showed differences in the aggregation rate (Supplementary Fig. S1b).

Phenotypical differences observed among the 27 isolates could reflect genetic diversity. Therefore, we performed the full-length sequencing of the 16S rDNA on these isolates and compared them with the 22 rDNA16S sequences available on NCBI database. This revealed two distinct phylogenetic groups: A and B (Figure 1a). *Finegoldia magna (F. magna)*, the most closely related bacterial species to *P. micra*, was used to set the root of the phylogenetic tree. To confirm the existence of the
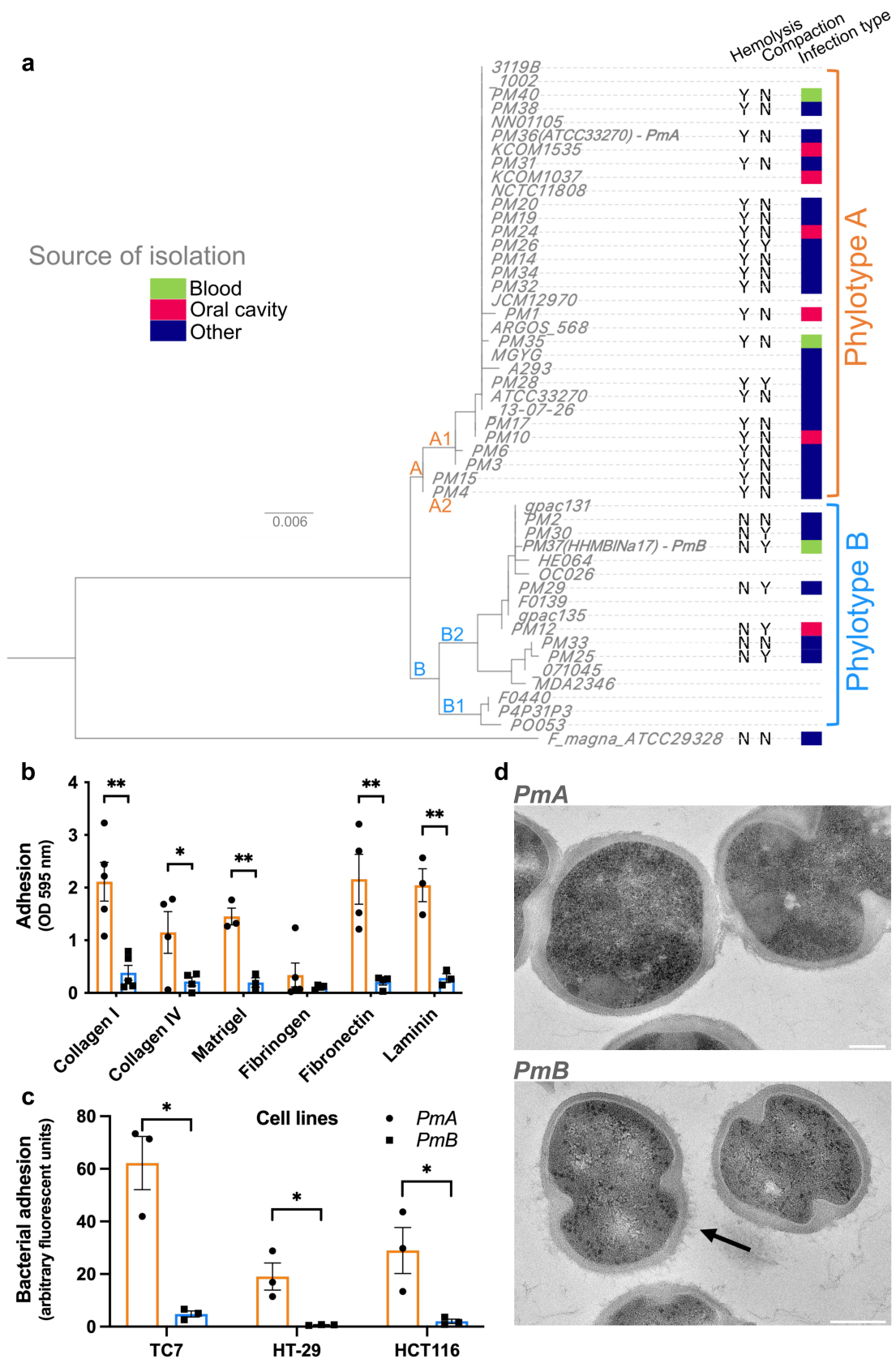
Figure 1. *P. micra* exists in two phylotypes with phenotypic and genetic diversity. (a) Phylogenetic tree based on full-length 16S rDNA sequences of 27 *P. micra* clinical isolates (named *PMX*) and 22 reference sequences from the NCBI database. Root: *Finegoldia magna* ATCC 29328. The infection types from which the 27 clinical isolates are derived are indicated by a color code; hemolysis and colony compaction abilities are indicated as present (Y) or absent (N). (b) Adhesion capacity of *P. micra* ATCC 33270 (*PmA*) and HHM BlNa17 (*PmB*) to extracellular matrix proteins. Optical density at 595 nm represents measurement of bacterial adhesion. (c) Adhesion capacity

two phylotypes, we have clustered whole-genome sequences, available on NCBI, of 10 of the 22 *P. micra* strains used for the 16S rDNA alignment, on the basis of the presence/absence of genes. This analysis confirmed the 16S rDNA clustering and the existence of phylotypes A and B (Supplementary Fig. S1c). Interestingly, all *P. micra* isolates from phylotype A were hemolytic, whereas isolates belonging to phylotype B were not. Likewise, most phylotype B isolates were "compact", whereas phylotype A isolates were not. These data provide genetic evidence of two phenotypically distinct phylotypes in *P. micra* whose respective impact on colonocytes and distribution in CRC had to be defined.

Clinical studies using 16S rDNA sequencing from tissue have shown the presence of *Parvimonas* genus sequences in tumoral and normal homologous colonic mucosa of CRC patients, suggesting that this microorganism may be able to adhere to colonocytes or to the colonic cellular matrix.[21,22] We therefore measured the ability of a representative isolate from each phylotype, *PmA* (phylotype A) and *PmB* (phylotype B), to adhere to different extracellular matrix (ECM) proteins and three human colonic cell lines. *PmA* adhered to the ECM proteins collagen I, collagen IV, fibronectin, and laminin, as well as to Matrigel® but not to fibrinogen, whereas *PmB* did not adhere to any of these ECM proteins (Figure 1b). Similarly, *PmA* adhered 6-, 2-, and 3-fold more to TC7, HT-29, and HCT116, respectively, as compared to *PmB* ($p < .05$) (Figure 1b), indicating heterogeneity of adhesion capacity in *P. micra* phylotypes. To further investigate the adhesion capacity of phylotypes, we focused on *P. micra* isolated from the oral cavity, its endogenous niche. The isolate *Pm12*, belonging to phylotype B, neither adhered to Matrigel® (Supplementary Fig. S1d) nor to the human colonic cell line TC7 (Supplementary Fig. S1e), while those of phylotype A, *Pm1*, *Pm24*, and *Pm10*, did ($p < .01$).

To explore whether the aggregative phenotype could be due to specific bacterial surface structural differences, ultrathin sections of the two phylotypes were analyzed by transmission electron microscopy (TEM). *PmB* (non-hemolytic, "compact") presented spiky surface structures rarely observed in *PmA* (hemolytic, "non-compact") (Figure 1d). Similar spiky structures were observed in phylotype B oral isolate *Pm12*, but not in the phylotype A oral isolate *Pm10* (Supplementary Fig. S1f).

These data show that two phenotypically and genetically different phylotypes can be described for *P. micra*.

The adherent *P. micra* phylotype A is associated with CRC

Differences in the *in vitro* adhesion capacity of the two phylotypes might reflect a different colonization ability *in vivo*. To assess this hypothesis, we first improved the resolution of our previous whole-genome sequencing (WGS) metagenomic analysis,[23] focusing on the taxonomic assignation of raw data, in order to precisely investigate the occurrence and abundance of *P. micra* of CRC patients' feces. We observed an enrichment of *P. micra* occurrence in the feces of CRC patients as compared to control healthy individuals (27% versus 1.1%). Three taxa of *P. micra* were detected: "*Parvimonas micra*" from phylotype A, "*Parvimonas* sp. oral taxon 393 - F0440" and "*Parvimonas* sp. oral taxon 110 - F0139" from phylotype B. All *Parvimonas* positive patients carried phylotype A in their feces, except in one patient where reads of both phylotypes were detected.

We then analyzed *Parvimonas* carriage, using 16S rDNA sequencing of the V3-V4 region, in feces of CRC patients and control individuals, as well as from paired tumor and normal homologous tissues (adjacent to the tumor), ranking the tumoral stages according to the international TNM (tumor-node-metastasis) staging system and named as early (CRC I&II) and late (CRC III&IV).

---

of *PmA* and *PmB* to different human colonic cell lines (TC7, HT-29, and HCT116) after 1 h of co-culture. Bacterial adhesion was quantified by the analysis of fluorescent images of cells co-cultured with the two phylotypes and is reported as arbitrary fluorescent units (x100). Data are represented as mean ± SEM from three independent experiments. Mann–Whitney test **$p < .01$; *$p < .05$. (d) Transmission electron microscopy on ultrafine sections of *PmA* (top) and *PmB* (bottom). Spiky surface structures (arrows) were observed only in *PmB* and not in *PmA*. Scale: 200 nm. See also Supplementary Fig. S1 and Supplementary Table S1.

The relative abundance of *Parvimonas* in feces of patients with adenomas vs. control individuals (0.0014%) was not statistically different, while it was enriched in those of CRC patients, with a median abundance of 0.0124% and 0.0063% for early and late stages, respectively (Figure 2a). *Parvimonas* prevalence in feces followed a similar trend with no differences between patients with adenoma (44.7%) and controls (60%) and a significant increase in CRC (early and late stages together) (83%) (Supplementary Fig. S2a). Within the early and late stages, no differences in abundance or prevalence neither in feces (Figure 2a and Supplementary Fig. S2a) or in tumoral tissues (Figure 2b and Supplementary Fig. S2b) were observed.

In 78.8% of CRC patients, *Parvimonas* sequences were detected in either tumoral or normal homologous samples (Supplementary Fig. S2c), with a significant abundance enrichment in the tumoral tissues ($p < .01$) (Figure 2c).

The V3-V4 regions of the 16S rDNA was then analyzed to discriminate between *P. micra* phylotypes A and B. An increase in the abundance of phylotype A, but not of phylotype B, was observed in CRC feces compared to controls ($p < .01$) (Figure 2d), with 61% of the CRC patients carried the phylotype A in their feces as compared to 44% of control individuals, whereas no difference in the prevalence of phylotype A was observed (22% and 27%, respectively) (Supplementary Fig. S2d,e). Furthermore, an increase in the abundance and prevalence of phylotype A, but not phylotype B, was also observed in tumoral tissues compared to homologous normal tissues ($p < .05$) (Figure 2e and Supplementary Fig. S2e). Moreover, we investigated *P. micra* phylotypes carriage in large, non-EU cohort data already published in Baxter et al. 2016, and observed similar results, with an increased abundance and prevalence of phylotype A, but not phylotype B, in CRC feces compared to control individuals ($p < .0001$) (Supplementary Fig. S3).

Previously, we demonstrated a strong correlation between *Parvimonas micra* carriage in feces and a positive cumulative methylation index (CMI) in the blood based on the methylation levels of promoters in the WIF1 gene, involved in carcinogenesis, as well PENK and NPY, two neuromediator genes.[24] Thus, we wondered whether the phylotypes herein described, were correlated with the CMI. Analysis of 16S rDNA sequencing data from feces showed that phylotype A was enriched in CMI positive patients ($p < .01$) but this was not the case for phylotype B (Figure 2f). *Parvimonas* carriage in tissues was also associated with a positive CMI ($p < .05$) (Figure 2g), although the number of samples was not sufficient to statistically differentiate between the two phylotypes (Figure 2h).

These results indicate that *Parvimonas micra* phylotype A sub-species is associated with a CRC and a positive CMI.

P. micra's impact on human colonic primary epithelial cells

Colorectal tumor is defined as a group of anarchically proliferating cells with DNA mutations and aberrant DNA methylation. To determine whether *P. micra* has an impact on host cells physiology, we first developed a compatible co-culture model between this oxygen-sensitive anaerobic microorganism and primary human colonic epithelial cells. Colonic samples were obtained from normal sites (distant from tumor tissue) of two colonic tumor resections. Colonic crypts were isolated and cultured as organoids (Figure 3a), allowing amplification of initial material through stem cell proliferation. To develop a system where bacteria could be in contact with the apical pole of primary cells, reflecting a physiological situation, organoid fragments were seeded onto a Transwell® insert system and grown to confluency up to a polarized monolayer (Figure 3b), as previously described.[25] Cell growth and formation of a tight monolayer were monitored over time by phase-contrast microscopy and transepithelial electrical resistance (TEER) measurements (Figure 3c, d). Under these culture conditions, monolayers were fully confluent 3 days after seeding organoid fragments onto the inserts (Figure 3d).

P. micra has been described as an anaerobic microorganism[26] and little is known about its oxygen tolerance. To determine compatible oxygenation conditions with eukaryotic cell growth requirements, *P. micra* was cultivated at various oxygen concentrations (0%, 2%, or 21% O_2) and viability was assessed by Colony Forming Units (CFU) counting at different time points. *P. micra's* viability dropped by 70% in less than 10 min in aerobic conditions (21% O_2), whereas in hypoxic conditions (2% O_2) it was able to grow,
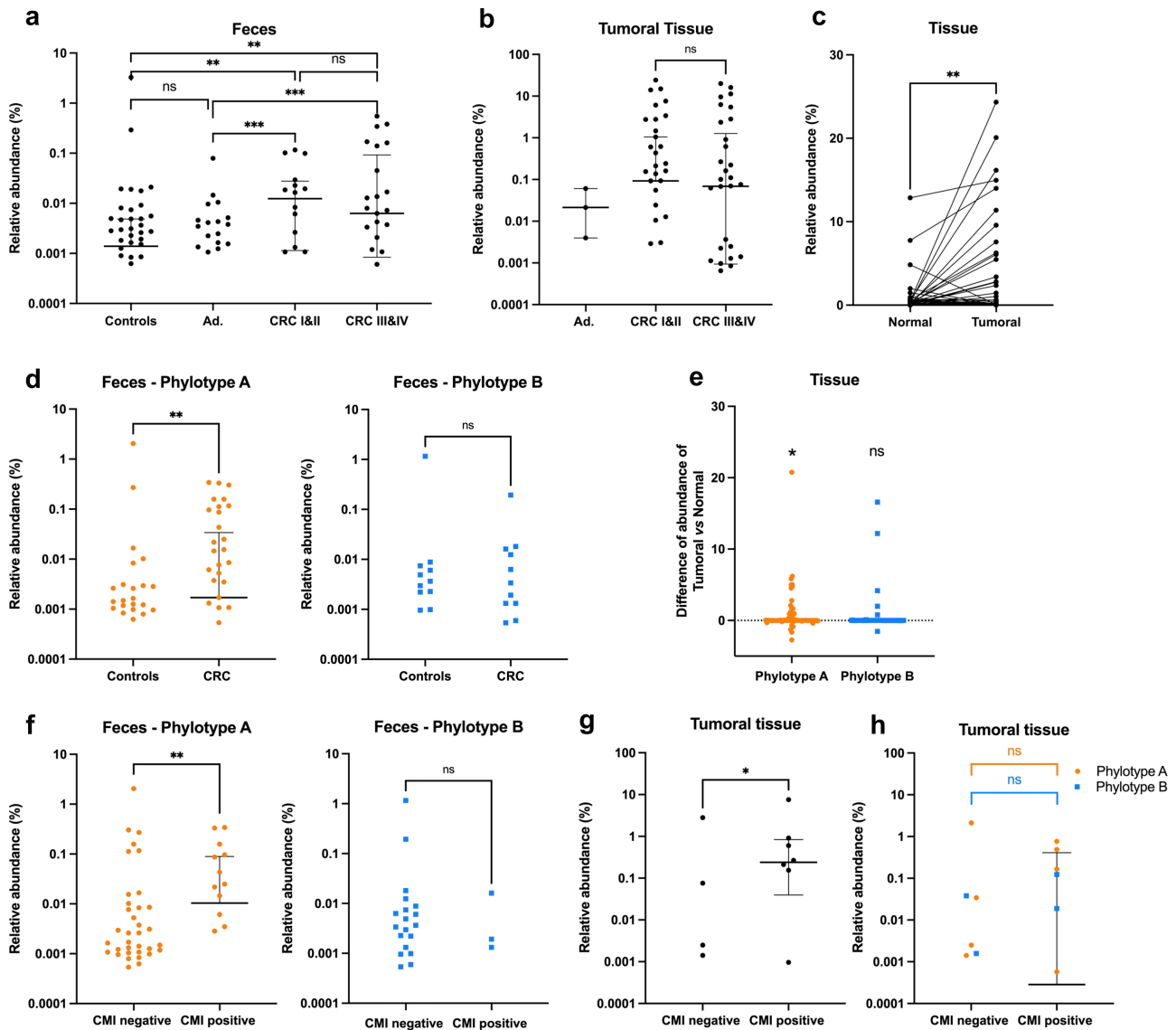
Figure 2. *P. micra* phylotype A is associated with CRC. *Parvimonas* carriage, determined by 16S rDNA sequencing of the V3-V4 region, in controls, adenoma (Ad.), or sporadic CRC patients at early (CRC I&II) and late (CRC III&IV) stages of carcinogenesis. (a) *Parvimonas* relative abundance in feces. (b) *Parvimonas* relative abundance in colonic tumoral tissue of CRC patients. (c) *Parvimonas* relative abundance in paired homologous mucosa and tumoral tissue samples. (d) *P. micra* A and B phylotypes relative abundances in control and CRC patients' feces. (e) Differences of relative abundance of each phylotype between tumoral and normal tissues. (f) *P. micra* phylotypes association with CMI methylation score of the WIF1, PENK, and NPY genes promoters in controls and sporadic CRC patients' feces. (g) *Parvimonas* or (h) *P. micra* phylotypes associated with CMI methylation score in tumoral tissue of CRC patients. The relative abundance is expressed as a percentage of counts (number of sequences assigned to *Parvimonas* per number of total bacterial sequences X100) and results are presented as median with interquartile range. Statistical analysis: Mann–Whitney test, ***$p < .001$; **$p < .01$, *$p < .05$ and Wilcoxon Rank Test for paired samples, **$p < .01$. See also Supplementary Table 2 and Supplementary Figure S2.

albeit at half the rate compared to full anaerobic conditions (Supplementary Fig. S4a). Thus, primary cells were grown for 4 days in aerobic conditions to allow monolayer formation, followed by 3 days in hypoxia (2% O_2). The primary cell monolayer's integrity was maintained in these conditions (Figure 3c, d) and showed all characteristics of a differentiated epithelium as assessed by the presence of tight junctions, differentiated epithelial cells, colonocytes, goblet cells, and enteroendocrine cells (Figure 3e).

To investigate the possible involvement of *P. micra* in early oncogenic processes, we cocultivated primary cell monolayers with two
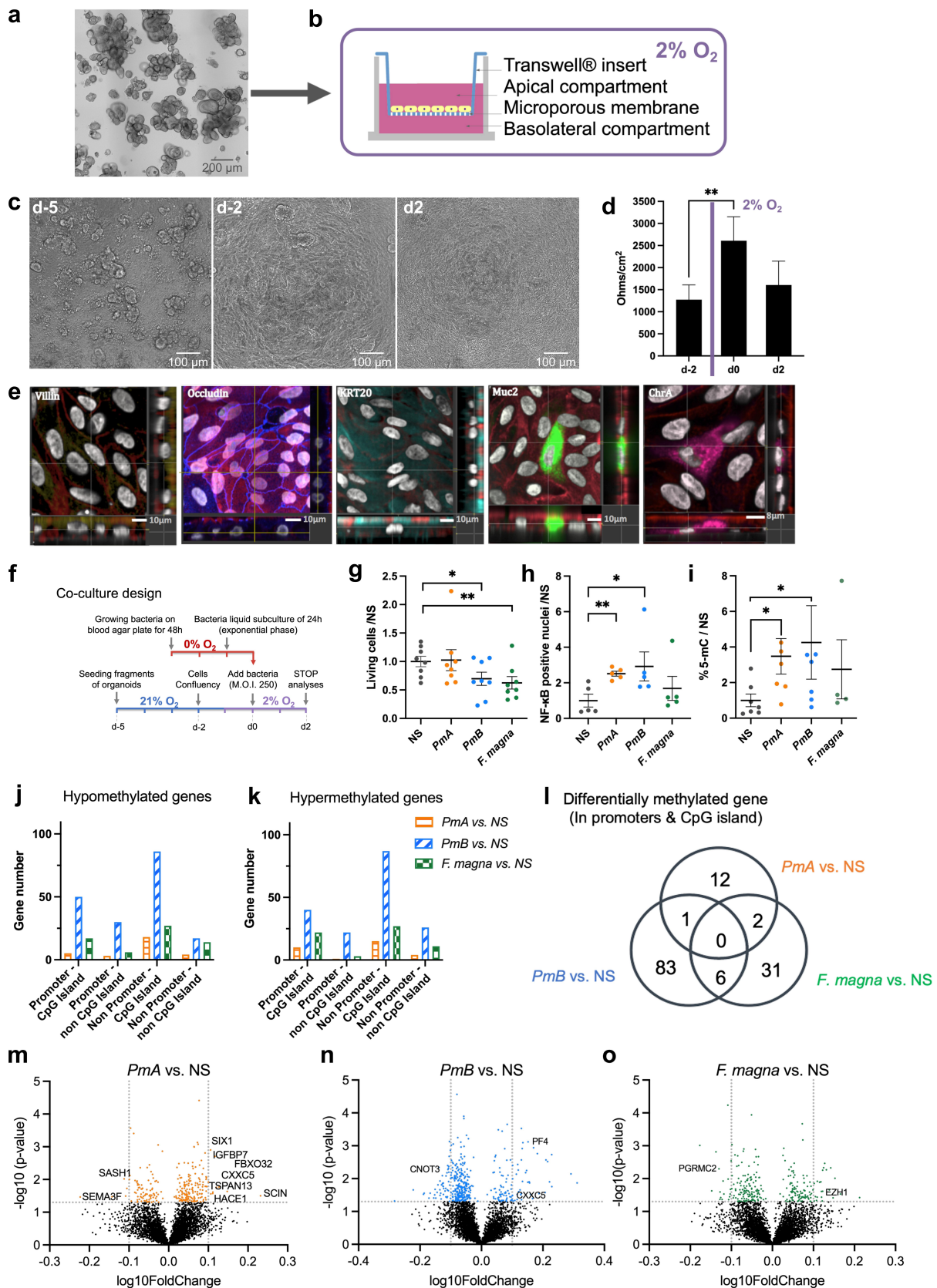
Figure 3. *P. micra* impact on human colonic primary epithelial cells. (a) Colonic organoids obtained from normal tissue of CRC patients and cultured in Matrigel® with growth factors-enriched medium. (b) Schematic representation of Transwell® permeable insert used to cultivate monolayers of primary colonic epithelial cells derived from organoids. (c) Representative phase contrast images of organoid monolayers at seeding (d-5), after 3 days of culture in aerobic conditions (d-2) and after 7 days of culture, including 4 days in aerobic conditions and 3 days in hypoxic conditions (d2). (d) Transepithelial electrical resistance (TEER) at different days of culture. TEER

reference isolates, *PmA* and *PmB*, as well as *F. magna* as a control, which is the closest phylogenetic taxon to *P. micra*. One day prior to co-culture, cells were placed in hypoxic conditions (2% O_2) to allow pre-adaptation. On the fourth day of culture, cells were co-cultured with bacteria using a multiplicity of infection (MOI) of 250 and incubated at 2% hypoxic conditions for 48 h prior to analysis (Figure 3f). Bacterial growth was monitored over time by CFU counting of co-culture supernatants. After 48 h of co-culture, *PmA* showed a 5-fold increase but did not impact cell viability, while *PmB* and *F. magna* showed no significant growth but had a more toxic effect, inducing a 30% decrease in cell viability (Figure 3g). Cell proliferation was measured by Ki-67 quantification upon immunofluorescent staining, and no significant differences were observed between the various groups (Supplementary Fig. S4c, f). Because goblet cells are known to be impacted by intestinal microbes[27], we quantified their proportion and no significant differences were observed (Supplementary Fig. S4d, g). Moreover, because virulent bacteria can stimulate NF-κB and induce DNA breaks[28], the proportion of cells with DNA double-strand breaks was measured and was found unchanged (Supplementary Fig. S4e, h). Yet, as compared to the NS condition, a significant increase in nuclear NF-κB was observed with both

PmA (p <.01) and *PmB* (p <.05), but not *F. magna*, (Figure 3h). Thus, in the time frame of the experiments, *P. micra* did not induce DNA breaks while triggering activation of the NF-κB master regulator of inflammatory and anti-apoptotic responses. These results suggest that *P. micra* might contribute to host epithelial cell transformation through a NF-κB pathway-mediated effect.

As we previously showed a correlation between *Parvimonas* carriage in feces and a positive CMI in patients[24], global DNA methylation was measured using a 5-methyl-cytosine dosage assay *in vitro*. Co-cultures of both *P. micra* phylotypes showed a significant increase in global DNA methylation of primary cells as compared to NS condition (p < .05) or to *F. magna* (Figure 3i). To identify the affected genes, a genome-wide DNA methylation analysis was performed on human colonic primary cells exposed or not to *P. micra* phylotypes or *F. magna*, using the Infinium Methylation EPIC BeadChip from Illumina. The methylation status of >850,000 CpG sites was assessed. Distributions of CpG sites were then classified by context (in or outside of CpG islands) and regions: body genes (3' or 5' UTR, first exon, exon bond, and body), promoters (200 or 1,500 bases upstream of the transcriptional start site [TSS200 and TSS1500]), or intergenic. Using the LIMMA paired test, we identified hyper- and hypo-methylated CpG sites from colonic primary cells challenged with bacteria

experiments were performed on cells from two donors, for at least four experiments per donor. The results are expressed as a mean of Ohm per cm^2 ± SEM. Mann–Whitney test **p < .01. (e) The differentiation of the monolayers after 4 days in aerobic conditions and 3 days in hypoxic conditions (2% O_2) was assessed by confocal microscopy. Nuclei (DAPI) are shown in gray, actin (phalloidin) in red, microvilli (anti-villin) in yellow, tight junctions (anti-occludin) in dark blue, Colonocytes (anti-KRT20) in cyan, goblet cells (anti-muc2) in green, and enteroendocrine cells (anti-ChrA) in magenta. Main panels, XY projection; right panels, YZ projection; bottom panels, XZ projection. (f) Co-culture experimental design with *P. micra* ATCC 33270 (*PmA*), HHM BlNa17 (*PmB*), and *F. magna*. (g) Cell viability after 48 h of co-culture under hypoxic conditions at 37°C in 5% CO_2. Living cells were quantified by flow cytometry and presented as the ratio of living cells in stimulated versus non-stimulated samples (NS). (h) Quantification of NF-κB positive nuclei upon immunofluorescent labeling with an anti-p65 antibody and nuclear staining. The results are expressed as ratios of p65-positive nuclei over the total number of nuclei and normalized by the NS condition. (i) Global DNA methylation dosage by quantification of 5-methyl-cytosine (5 m-C) residues compared to total cytosines, after 48 h of co-culture of bacteria with human colonic monolayers. The results are expressed as the percentages of 5-mC normalized to the non-stimulated samples (NS). In g, h, and i, the results are from cells derived from two donors with at least three independent experiments per donor and the data are represented as mean ± SEM. Mann–Whitney test ***p < .001, **p < .01, *p < .05. See also Supplementary Fig. S4. (j, k) Gene number and distribution of hypo- and hyper-methylated genes between *P. micra* ATCC 33270 (*PmA*), *P. micra* HHM BlNA17 (*PmB*) phylotypes, or *F. magna* ATCC 29328 and the NS condition. Differentially methylated genes were obtained by calculating a combined p-value based on differentially methylated CpG sites located at specific genomic regions (TSS, in gene promoters; or non-TSS, not at promoter sites) and regions within CpG islands or not. (l) Venn diagram of differentially methylated genes when only promoters (TSS) and CpG islands were considered. (m, n, and o) Volcano plots of differentially methylated genes (hypo- and hyper-methylated) in the comparison of *PmA, PmB*, or *F. magna* bacterium- *versus* non-stimulated condition, respectively. CpG sites only in promoters (TSS) and in CpG islands were considered. Genes belonged to oncogenes, tumor suppressor genes (TSGs) or epithelial–mesenchymal transition (EMT) databases are highlighted by their names. See also Table 1.

compared to non-stimulated cells. CpG sites were then sorted into four genomic categories: CpGs in promoters and within CpG islands; CpGs in promoters and outside CpG islands; CpGs in body genes (promoters region excluded) and within CpG islands, as well as CpGs in body genes outside of CpG islands. Because aberrant DNA methylation in promoters of tumor suppressor genes (TSGs) is a hallmark of CRC, combined p-values of differentially methylated CpG sites within the same gene were determined based on these four categories. We found that DNA methylation modifications were induced by all three bacteria, while *PmB* promoted the highest number of differentially methylated genes (DMGs) (Figure 3j, k). Focusing on CpG sites located in promoters and within CpG islands, we identified 15, 90, and 39 DMGs between non-stimulated cells and cells co-cultivated with *PmA*, *PmB*, and *F. magna*, respectively. The Venn diagram revealed six genes in common between *PmB* and *F. magna*, only one in-between the two phylotypes of *P. micra*, and two genes for *PmA* and *F. magna* (Figure 3l). Strikingly, most of the differentially methylated promotors in *PmA* co-culture cells were either found in oncogenes, TSGs, or genes involved in epithelial–mesenchymal transition (EMT) when compared to *PmB* or *F. magna* co-cultures (Figure 3m–o). Cells co-cultivated with *PmA* presented hyper-methylation of several TSGs promoters such as SCIN, HACE1, TSPAN13, FBXO32, IGFBP7, SIX1, or CXXC5. Except for the KIAA0494 gene that codes for an uncharacterized protein, all DMGs induced by *PmA* are involved in carcinogenesis, particularly in EMT processes or cytoskeleton remodeling (Table 1). These results suggest that *P. micra* phylotype A might alter expression of a set of genes, including well-characterized carcinogenesis regulators, through epigenetic promoter gene methylation.

P. micra *phylotype A carriage in CRC patients is associated with DNA methylation modifications and gene expression variations*

To confirm that the differences in phylotypes might influence colon cancer progression, we put our efforts into isolating *Parvimonas* from colonic tumoral biopsies. We eventually successfully isolated one clone that we could cultivate *in vitro*. The sequencing of the full-length 16S rDNA (Figure 4a) as well as the phenotypic properties (e.g. hemolysis, density of colonies, and adhesion) indicated that this new colonic tumoral isolate, named *PmG5*, belongs to phylotype A. Indeed, *PmG5* was hemolytic, "noncompact" on blood agar plates, and able to adhere to Matrigel® (Figure 4b) and to the colonic cell line TC7 *in vitro* (Figure 4c). No spiky structures were observed on *PmG5* surface by TEM (Figure 4d). Thus, *PmG5* showed all the features of the *PmA* isolate we used for the *in vitro* assay.

Having shown that *P. micra* phylotype A is more prevalent in CRC and associated with cancer tissue, we wondered if its presence was also associated with DNA methylation changes in host colonic cells, as observed *in vitro*. We thus performed reduced-representation bisulfite sequencing (RRBS) analysis on colonic tumoral tissues either negative or highly colonized by *P. micra* from three patients in each group. We observed 135 differentially methylated genomic regions (DMRs), mostly located in promotor regions (45%) and in the CpG island context (68%) (Figure 4e, f). Albeit we did not identify the same genes as in our *in vitro* study, differentially methylated regions were found in numerous TSGs, oncogenes, and genes involved in EMT (Figure 4g, Supplementary Tables S3).

To identify altered gene expression patterns linked with *P. micra* colonization of the mucosa, RNA-seq was performed on colonic tumoral tissues, and transcriptome comparisons of either *Parvimonas*-positives ($n = 30$) or *P. micra* phylotype A-positive patients ($n = 20$) versus negative patients ($n = 12$) were analyzed. We found 483 and 336 differentially expressed genes (DEGs) for *Parvimonas* and *P. micra* phylotype A, respectively (Log2FC > 1, $p < .05$) (Figure 4h, Supplementary Tables 3). To uncover the potential signaling pathways that are impacted by *P. micra* colonization, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (Figure 4i) on overexpressed genes in *P. micra* positive patients. Over-

Table 1. *PmA* induces DNA methylation changes in promoters of genes involved in carcinogenesis.

Names, Alias	Described role	Link with CRC or cancers	References
SCIN (Scinderin)	Ca^{2+}-dependent actin-severing and capping protein. Regulation of actin cytoskeleton.	Overexpressed in CRC. Inhibits cell proliferation and tumorigenesis.	Lin 2019
HACE1 (HECT domain E3 ubiquitin protein ligase 1)	E3 ubiquitin ligase involved in specific tagging of target proteins, leading to their subcellular localization or proteasomal degradation.	TSG. Down-regulated by DNA methylation in CRC. Loss or knockout of HACE1 enhanced tumor growth, invasion, and metastasis; in contrast, the overexpression of HACE1 can inhibit the development of tumors.	Li 2019, Hibi 2008, Zhang 2007
TSPAN13 (Tetraspanin 13)	Transmembrane signal transduction protein that plays a role in the regulation of cell development, activation, growth, motility and invasion.	TSG. Downregulation inhibits proliferation of CRC cells.	Lou 2017
FBXO32 (F-Box Protein 32) Astogin-1	Component of a SCF E3 ubiquitin-protein ligase complex which mediates the ubiquitination and subsequent proteasomal degradation of target proteins.	TSG. Under-expressed in CRC. Induce cell differentiation. Upstream regulator of EMT.	Yuan 2018, Sahy 2017
IGFBP7 (Insulin Like Growth Factor Binding Protein 7)	Protein coding gene that regulate IGFs. Stimulates prostacyclin production and cell adhesion. Promotes cancer cell growth and migration.	TSG. Silencing induce metastasis.	Ruan 2007, Suzuki 2010
KIAA0494 (EF-Hand Calcium Binding Domain 14)	Uncharacterized protein. Predicted membrane protein.	-	
DIAPH3 (Diaphanous Related Formin 3)	Protein coding gene involve in actin remodeling. Regulate cell movement and adhesion.	Deficiency enhances cell motility, invasion and metastasis in many cancers.	Hager 2012, Rana 2018
SIX1 (sine oculis homeobox 1)	Transcription factor involve in regulation of cell proliferation, apoptosis, embryonic development and tumorigenesis.	Oncogene. Overexpressed in CRC, overexpression of Six1 dramatically promotes CRC tumor growth and metastasis in vivo.	Wu 2014, Xu 2017
KCNB1 (Potassium Voltage-Gated Channel Subfamily B Member 1)	Voltage-gated potassium channel Kv2.1. Contributes to the pronounced pro-apoptotic potassium current surge during neuronal apoptotic cell death in response to oxidative injury.	KCNB1 polymorphisms correlate to CRC treatment and patient's outcome. Kv2.1 occasionally forms complexes with other voltage-gated potassium alpha-subunits (i.e. Kv9.3), witch its silencing potently inhibiting proliferation in human colon cells.	Li 2015, Fahra 2020
SASH1 (SAM and SH3 domain-containing protein 1)	Scaffold protein involved in the TLR4 signaling. Stimulate cytokine production and endothelial cell migration in response to binding pathogens	TSG. Downregulated in CRC. Downregulation expression was correlated with the formation of metachronous distant metastasis. Loss of SASH1 induces EMT. SASH1 inhibits metastasis formation *in vivo*.	Rimkus 2006, Franke 2019
ZNF282 (Zinc Finger Protein 282)	Transcription factor binding the U5RE (U5 repressive element) of HLTV-1 (human T cell leukemia virus type 1) with a repressive effect. Co-activator of the estrogen receptor alpha.	Knockdown reduced migration, invasion and tumorigenesis of ESCC *in vitro* and reduced the tumorigenicity of ESCC xenograft in nude mouse.	Yeo 2014
SEMA3F (Semaphorin 3F)	Secreted signaling protein that are involved in axon guidance during neuronal development. Act in an autocrine fashion to induce apoptosis, inhibit cell proliferation and survival. Regulator of actin cytoskeleton.	TSG. Down-regulated by DNA methylation in CRC tissues and CRC cell lines. Associated with progressive phenotypes of CRC. Overexpression reduced proliferation, adhesion, and migratory capability of colon cancer cells. SEMA3F-overexpressing cells exhibited diminished tumorigenesis when transplanted in nude mice and reduced liver metastases.	Gao 2015, Wu 2011, Bielenberg 2004

CRC: colorectal cancer; TSG: tumor suppressor gene; EMT: epithelial–mesenchymal transition; IGFs: insulin-like growth factors; SCF: SKP1-CUL1-F-box protein; TRAPP: Transport protein particle; ESCC: esophageal squamous cell carcinoma.

expressed genes in *P. micra* phylotype A-positive patients revealed significant enrichment in several pathways related to inflammation, such as phagosome, Toll-like receptor signaling pathway, cytokine–cytokine receptor interaction, chemokine signaling pathway, or intestinal immune network for IgA production, the latter demonstrating a localized inflammatory response to mucosal colonization (Figure 4i). Affected pathways also included transcriptional misregulation in cancer, hematopoietic cell lineage, and PI3K-Akt signaling pathway, a pathway well-known to be involved in cell growth, survival, cell-cycle progression, and differentiation. In addition, DEGs enriched in *P. micra* phylotype
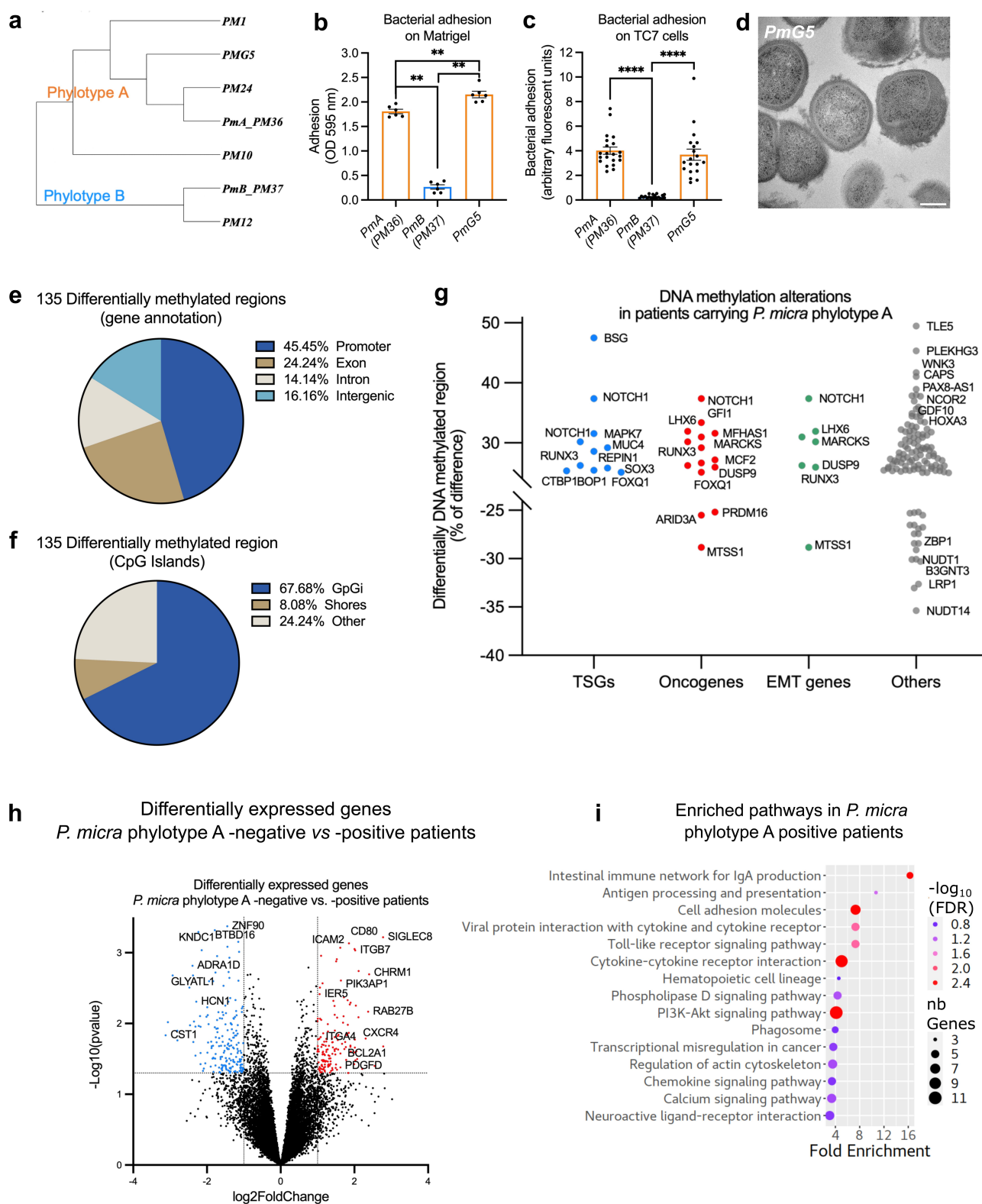
Figure 4. *P. micra* phylotype a carriage in CRC patients is associated with DNA methylation modifications and gene expression variations. (a) Phylogenetic tree based on full-length 16S rDNA sequences of *P. micra PmA, PmB,* other oral clinical isolates and of the new isolate from tumoral colonic tissue *PmG5*, showing that it belongs to phylotype A. (b) Adhesion capacity of *PmA, PmB,* and *PmG5* to Matrigel®. Optical density at 595 nm represents measurement of bacterial adhesion. (c) Adhesion capacity of *PmA*, *PmB*, and *PmG5* to human colonic cell line TC7 after 1 h of co-culture. Bacterial adhesion was quantified by the analysis of fluorescent images of cells co-cultured with the two phylotypes and is reported as arbitrary fluorescent units. Mann–Whitney test ****p<.0001; **p<.01. (d) Transmission electron microscopy on ultrafine sections of *PmG5* (Scale: 200 nm). (e) Genomic localization and (f) CpG context of the 135 differentially methylated regions between three *Parvimonas*-negative and three *Parvimonas*-positive CRC patients, using reduced-

A-positive patients were involved in cell adhesion and regulation of actin cytoskeleton (Figure 4i, Supplementary Tables S3).

These findings revealed that tissues colonized with *P. micra* phylotype A display altered DNA methylation patterns and expression in key genes involved in carcinogenic events.

Discussion

In this study, we investigated the oral pathobiont *P. micra*, a bacterium that has rarely been isolated and poorly studied, and identified two phenotypically and molecularly distinct phylotypes (A and B). Notably, we observed a robust association of the adherent and hemolytic phylotype A with CRC. To simulate physiologically relevant conditions, we developed a low-oxygen *in vitro* co-culture model using primary human colonic intestinal cells, and we showed the capacity of *P. micra* phylotype A to induce DNA methylation promoters' modifications. Moreover, our analysis of the colonic mucosa from CRC patients carrying *P. micra* phylotype A revealed not only DNA methylation alterations in promotors, but also changes in expression of tumor suppressor genes, oncogenes, and genes involve in inflammation, cell adhesion, and cytoskeleton rearrangements.

Recently, we and others observed an epidemiological association between CRC and several oral bacteria detected in feces using 16S rDNA sequencing[21,34] or metagenome analysis.[24,35,36] They mainly belong to species such as *Fusobacterium nucleatum*, *Porphyromonas gingivalis*, *Solobacterium moorei*, *Peptostreptococcus stomatis*, *Gemella morbillorum*, and *Parvimonas micra*. Some of these oral bacteria were also associated with colonic tissues of CRC patients[21,33,37] being more abundant in tumors than in normal adjacent tissues.[22,37] Interestingly, we found that these bacteria were often found in co-occurrence in the mucosa of patients (data not shown), suggesting that these oral microorganisms could live in close community in association with the colonic mucosa, partially reproducing an ecosystem similar to the oral cavity. In line with this hypothesis, *P. micra, F. nucleatum, P. stomatis*, and *G. morbillorum* have been observed in biofilm-like structures at the surface of the colonic mucosa of CRC patients and healthy subjects.[38–43] Therefore, pathobionts might synergize with other members of the microbiota or engage in intricate interactions with the human host, potentially leading to tumoral transformation. In our "simplified" *in vitro* model, we focused on one-bug/one-tissue interactions, observing a direct impact of *P. micra* on the epithelium. However, the presence of multiple bacteria in the gut does not preclude the possibility of a different effect (either weaker or stronger) on the host. Utilizing murine models to compare various phylotypes could offer insights into specific factors that regulate microbe–microbe interactions, as well as the crosstalk between the epithelium and the immune system.

In the present study, we quantitatively measured *P. micra* carriage in both feces and colonic tissues in a large cohort from Henri Mondor Hospital – Assistance Publique des Hôpitaux de Paris (Directed by Prof. I. Sobhani) that includes normal colonoscopy, adenomatous, and CRC patients. We observed, like others,[44] that *P. micra* is present but not significantly enriched in patients with adenomas, these benign tumors being considered as precursors of CRC,[45] indicating that *P. micra* is more likely to be involved in accelerating and/or exacerbating the carcinogenic processes rather than being a primary driver bacterium, according to the "driver-passenger" model.[46] In two recent studies, tumoral colonic cell lines (HT-29 and LoVo) were infected with either *P. micra* or *P. micra-*

representation bisulfite sequencing (RRBS). (g) Identity of genes with differentially methylated regions between *P. micra* phylotype A – negative versus – positive CRC patients. Genes belonging to oncogenes, tumor suppressor genes (TSGs) or epithelial–mesenchymal transition (EMT) databases are highlighted in red, blue, and green, respectively. (h) Volcano plot of differentially expressed genes (DEGs) between *P. micra* phylotype A – negative versus – positive CRC patients. Blue and red dots represent significantly down- and up-expressed genes, respectively (p < .05; |log2FoldChange|>1). (h) The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for up-regulated genes in *P. micra* phylotype A positive CRC patients. The x-axis indicates the degree of KEGG pathway enrichment. The y-axis indicates the name of the KEGG pathway. The dot size means the gene number. The dot color indicates the value of the False Discovery Rate (FDR). See also Supplementary Tables S3.

conditioned medium, leading to an increase in the proliferation rate upon stimulation.[47,48] Similarly, a parallel pattern emerged in the context of a *P. micra* oral infection of Apc$^{Min/+}$mice, where an augmentation in cell proliferation was observed.[48] In contrast, in our *in vitro* model, we did not observe similar results. This discrepancy could potentially be due to the fact that we have used primary cells and not cell lines, as in the published works. Possibly the observed induction of cell proliferation occurred within a context of cell lines that are immortalized and therefore already prone to proliferation. This might be explained by the fact that *P. micra* is a pathobiont, which might cause or promote disease (e.g. increase in proliferation) when specific genetic or environmental factors within the host are altered, as in the described models.

In CRC patients, *P. micra*, like other oral pathobionts, has been shown to be associated with the CMS1 subtype, which makes up 14% of all CRC cases,[49] and with an over-activation of genes involved in immune responses. The CMS1 tumors, also called "immune subtype," are characterized by a strong immune cell infiltration by CD8+ cytotoxic T cells, CD4+ T helper cells, and natural killer cells.[50,51] Therefore, a key aspect of *P. micra* involvement in colon carcinogenesis is inflammation. And indeed, mice orally gavaged with *P. micra* exhibited altered immune responses and enhanced inflammation.[47,48] In our *in vitro* model of colonic epithelial cells, we found that *P. micra* induced the activation of the transcription factor NF-κB, involved in cytokine and chemokine production, therefore potentially represent the first step in the recruitment of immune cells.

Apart from specific somatic mutations that characterize cancerous cells, epigenetic DNA modifications, particularly methylation, of TSGs promotors are the main contributors in colonic carcinogenesis. We recently showed that CRC-associated dysbiotic feces transplanted to mice caused epigenetic changes similar to those observed in human tumors and the occurrence of murine colonic crypt aberrations.[24] Furthermore, the CMS1 CRC subtype, in which *P. micra* and other oral bacteria enrichment has

been observed,[49] is associated with the phenotype of CpG islands hypermethylation in TSG promoters (high CIMP).[50] Consistently, Xia *et al.* observed an association between enrichments of *F. nucleatum*, *Parvimonas spp.* or other bacteria, and hypermethylation of promoter in several TSGs in CRC tumoral tissue, although they did not focus on *Parvimonas* subspecies. Moreover, they observed that *F. nucleatum* was able to up-regulate DNA methyltransferase activity *in vitro*.[52] In a previous study, we have shown in a large CRC patients' cohort that *P. micra* was significantly associated with methylation of the *WIF1* promoter, a very well-known TSG.[24] We now report for the first time that *P. micra* increases global DNA methylation of target host cells using an *in vitro* co-culture model of human primary colonic cells. Notably, by comparing different *P. micra* phylotypes, we established a *P. micra* phylotype A-dependent signature of methylated promoters whose gene functions converge toward the regulation of the cytoskeleton (Table 1) and include TSGs or genes involved in EMT (Table 1). For example, Scinderin gene (*SCIN*) coding for a Ca^{2+}dependent actin-severing and capping protein is involved in the regulation of actin cytoskeleton and known to be overexpressed in CRC;[29] Tetraspanin 13 (*TSPAN13*) is a TSG coding for a transmembrane signal transduction protein that regulates cell development, motility, and invasion;[31] *DIAPH3* gene, is a major regulator of actin cytoskeleton involved in cell motility and adhesion;[32] Semaphorin 3F (*SEMA3F*), is a TSG coding a secreted protein involved in cytoskeletal collapse and loss of migration;[53] and *SASH1* is a TSG coding for a scaffold protein involved in the TLR4 signaling, and known to interact with the actin cytoskeleton to maintain stable cell–cell adhesion.[54] Notably, the hypermethylation of the TSG *HACE1* gene promoter was also observed upon co-culture with *P. micra* phylotype A. HACE1 is an E3 ubiquitin ligase controlling RAC1 stability, a small GTPase of the Ras superfamily involved in cell motility[55] and the TSG *HACE1*'s promoter was

previously shown to be hypermethylated in CRC[30]. Hence, phylotype A of *P. micra* could contribute at several stages of CRC development by causing DNA methylation modifications enhancing cell transformation through cytoskeleton rearrangement. By comparing the methylome and transcriptome of *P. micra*-negative versus -positive patients, we observed differentially methylated regions in several other TSGs, oncogenes, or genes involved in EMT, together with significant enrichment of DEGs involved in cell adhesion, regulation of cytoskeleton, and PI3K-Akt signaling that promotes cell-cycle progression, cell survival, migration, and tumor progression.

In conclusion, here we describe the *P. micra* phylotype A as the most prevalent CRC-associated *Parvimonas*, characterized by its hemolytic capacity and adherent properties, able to colonize the colonic mucosa and induce DNA methylation changes. This could create a proper ground for carcinogenesis by potentially facilitating the accumulation of additional genomic alterations, inducing enhancement of cell migration, invasion, and/or metastatic events, consequently impacting tumor aggressiveness.

Clearly, the next steps require moving from association to causality. This involves comprehending the molecular differences between the two phylotypes, elucidating the basis for the tumor-association of phylotype A, exploring potential interactions with other microbiota members – such as biofilm formation – as well as unraveling the molecular mechanisms that underlie cancer promotion.

Materials and methods

Patient recruitment at the Créteil Henri Mondor hospital

Participants were selected from a large cohort recruited with informed consent between 2004 and 2018 by the endoscopy department at Henri Mondor hospital (Créteil) where patients had been referred for colonoscopy, a detailed description in Sobhani et al., 2011.[56] Participants with previous colon or rectal surgery, colorectal cancer, inflammatory bowel diseases, or with a genetic form of CRC were excluded. Individuals exposed to antibiotics or probiotics within 4 weeks before collection or suffering from acute diarrhea were also excluded from the study. Tumor Node Metastasis (TNM) stages of colonic neoplasia were determined by radiological examinations and analyses of the surgical specimen by the anatomopathological department of Henri Mondor Hospital. Clinical parameters of the patients, such as body mass index (BMI), age, sex, and disease history were referenced. The cumulative methylation index (CMI) score was previously determined.[56] This score was calculated from the methylation status of three genes (*wif1*, *penk*, and *npy*) involved in colorectal carcinogenesis and was considered negative for CMI < 2 or positive for CMI ≥ 2. Sample description is summarized in Supplementary Table S2.

Fecal and tissue samples

Fresh feces were collected between 2 weeks and 3 days prior to colonoscopy and 10 g were frozen at −20°C for 4 h and then stored at −80°C until use. Paired samples of colorectal tumor tissue and homologous normal mucosa (more than 15 cm from the margin of tumor resection) were collected within 30 min after surgical resection and immediately frozen in liquid nitrogen and stored at −80°C until further use.

DNA extraction and quantification for sequencing

Fecal DNA extraction was performed using the G'NOME DNA isolation kit® (MP Biomedicals) according to the supplier's instructions, with modifications as described in.[57] Tissue DNA extraction was performed from eight 50 μm cryosections of nitrogen frozen tissue using the QIAamp PowerFecal DNA Kit® (Qiagen) following the manufacturer's instructions with the following modification: 0.1 mm diameter silica beads were added to the lysis solution provided and shaking was performed at maximum speed for 10 min in a vibratory shaker. Fecal and tissue DNA concentrations were determined by Qubit® fluorometer (ThermoFisher) and stored at −20°C until use.

Whole genome sequencing (WGS) on fecal samples

Metagenome sequencing was performed using the high-throughput platforms BIOMICS (Institut Pasteur, Paris, France) and EMBL (Heidelberg, Germany). Sequencing was performed in pairs, using the HiSeq 2000/platform 2500 equipment, over a length of 100 bp of DNA and at a sequencing depth of 5 Gbp.[23] The raw data have already been reported in several papers.[23,35,56] For this study, 166 fecal samples from the CCR1 and DETECT cohorts were considered. The Diamond/MEGAN6 bioinformatics pipeline[58] was used for metagenomic assignment. Sequences were filtered for an average quality (Phred score) greater than 20 over a window of two consecutive bases and a length greater than 100 bp using Trimmomatic software (version 0.35). Good-quality sequences were translated into six possible reading frames and aligned to the NR reference library (RefSEQ non-redundant protein database) in which proteins/peptides that are more than 99% similar are combined into a single organism-associated group with a specific identifier.[59] A sequence can be assigned to several taxa. The MEGAN6 LCA (Lowest Common Ancestor) algorithm[60] was used to resolve multi-mapping reads: when a sequence was assigned to two (or more) taxa of different phylogeny, it was assigned to the top taxonomic level in common.

16S rDNA sequencing of fecal and tissue samples

For this study, 88 controls and 78 CRC patient's fecal samples and 71 pairs of homologous tissue samples (matched tumor and healthy tissue) were analyzed. After amplification of the 16S rRNA gene V3-V4 region, pairwise sequencing was performed to a length of 250 bp on the Illumina MiSeq platform. The resulting raw database was cleaned of sequences corresponding to human or phage sequences. Adapters and primers, as well as 5 and 3' ends with a Phred quality score over a 2 bp sliding window of less than 30 were removed using Trimmomatic (version 0.35) software. Paired sequences were merged using the FLASH2 tool (version 2.2.00). The taxonomic assignments were performed using MALT/MEGAN6 bioinformatic pipeline[60] with default parameters and the SILVA rRNA database version 123.[61]

For the already published data set, "Baxter,"[62] data were retrieved from the NCBI, cleaned, and analyzed as previously described. For abundance and prevalence computations, reads assigned to *Parvimonas* were extracted and blasted against phylotype A or B reference sequences: *Parvimonas micra* KCOM1535, 16S ribosomal RNA gene, partial sequence for phylotype A and *Parvimonas* sp. oral taxon 393 strain F0440 16S ribosomal RNA gene, a partial sequence for phylotype B.

RNA sequencing and pathway enrichment analysis

Total RNA was extracted from tumoral colonic mucosa samples using TRIzol®-chloroform extraction method. RNA was sent to the NOVOgene company (China) that carried out the quality control, library preparation, and sequencing. The reads quality was assessed using FastQC (v0.11.4 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) combined with MultiQC (v1.6 http://multiqc.info/).[63] We performed splice-aware alignment on RNAseq using the STAR transcriptome aligner (v2.5.0)[64] with human genome version GRCh38 from Ensembl release 99 (Homo_sapiens.GRCh38.dna.primary_assembly.fa). After alignment, featureCounts (subread v 1.6.1)[65] was used to obtain the matrix of counts of fragment by genes from sorted BAM files. We removed samples for which less than 50% of the fragments mapped to genes and used blast+ (2.9.0) to align the reads on the nt collection and confirm that the lack of alignment to human genes was due to an excess of reads mapping to non-human organism, suggesting a contamination. Overall, 88 samples were retained after quality control. We filtered gene level raw counts by removing all genes with no fragment mapped in more than half of the samples. Normalized counts for further use were generated by applying a quantile normalization using the voom procedure[66] implemented in the limma package (v3.46.0).[67] The 55 tumor tissue samples from which we had information both for *Parvimonas* carriage, assessed by 16S rDNA sequencing, and

host gene expression, were kept for differential expression analysis. Patients with no matching reads (dp = 0) in neither *Parvimonas genus*, *P. micra* phylotype A or *P. micra* B sequences, were considered as negative (n = 12). Patients with at least 10 matching reads in *Parvimonas* sequence were considered as *Parvimonas*-positive (n = 30). Patients with at least 10 matching reads in *P. micra* phylotype A sequence and representing at least 70% of all *Parvimonas* reads were considered as *P. micra* phylotype A-positive patients (n = 20). Differential analysis was not performed for phylotype B as only three patients were *P. micra* phylotype B-positive. The linear modeling followed by empirical Bayes statistics approach implemented in limma[68] was used to assess differential expression. Differentially expressed genes (DEGs) were identified as $p < .05$ and log2 fold change (Log2FC) > 1, and subjected to functional classification analysis, using ShinyGO (version 0.77)[69,70] and Kyoto Encyclopedia of Genes and Genomes (KEGG) database.[71] Pathway enrichment results were double-checked using DAVID (Database for Annotation, Visualization, and Integrated Discovery) version 2021.[72,73]

Bacterial strains

Escherichia coli pks+ strain IHE3034 was used as a positive control for double-strand DNA breaks. *Shigella flexneri* 5a BS176, lacking the virulence plasmid, was used as a negative control for bacterial adhesion and autoaggregation-negative control. The reference strains ATCC 33270 of *P. micra (PmA)* and ATCC 29328 of *Finegoldia magna* were obtained from the Pasteur Institute Collection (CIP). In this study, 27 clinical isolates of *P. micra*, including HHM BlNa17 (*PmB*), were collected from the bacteriology departments of three Parisian hospitals: Henri Mondor, Cochin, and Pitié Salpetrière (Supplementary Table S1). The identification of the different isolates was confirmed by mass spectrometry, and 16S rDNA sequencing analysis (Eurofins Genomics) and sequences were deposited on GenBank (Supple mentary Table S1).

16S rDNA sequencing for bacterial identification

16S rDNA whole sequence from positions 112 to 1,302 was aligned with the ClustalW program (EMBL), and the phylogenetic tree was calculated using the PhyML program. Pan-genomic comparison: available complete *P. micra* genomes were downloaded from the NCBI database. Analysis of pan-genomes and comparison of common or strain-specific gene carriage was performed using the Roary program[74] with a 95% identity threshold and the use of default parameters. Roary outputs a multi-FASTA alignment of all of the core genes and a matrix of presence or absence of each gene in each genome. The maximum likelihood phylogenetic tree from the alignment of all of the core genes was inferred using Fastree with the GTR +CAT model. An additional script called roary_-plots.py was used to visualize the tree against the presence and absence of core and accessory genes.

Bacterial culture

Bacteria were grown under anaerobic conditions in TGY-V enriched broth (Trypticase peptone 30 g/L, D-glucose 10 g/L, yeast extract 20 g/L, L-cysteine-HCl 0.5 g/L, hemin 5 mg/L, vitamin B12 5 µg/L, menadione sodium bisulfite 500 µg/L, Thiamine 1 mg/L, nicotinic acid 1 mg/L, riboflavin 500 µg/L, p-aminobenzoic acid 100 µg/L, biotin 25 mg/L, calcium pantothenate 1 mg/L, pyridoxamine 2HCl 500 µg/L, folic acid 500 µg/L) or on Columbia agar solid support enriched with 5% horse blood (COH) (Biorad). Anoxic conditions were generated in Gaspack (BD) or anaerobic cabinet (Don Whitley DG250 Workstation). Hemolysis ability and colony compactions were assessed by observing CFUs on COH plates after 48 h at 37°C in anaerobic conditions: i) contact hemolytic activity was defined by the presence of a hemolytic zone located underneath the colony and slightly beyond (Supplementary Fig. S1a) and ii) colony compaction was defined as "non-compact" or "compact": the "non-compact" colonies spread out on agar upon contact with a loop, whereas the compact colonies remained associated in clumps suggesting a strong interbacterial aggregative property. For the

sedimentation assay, bacteria were grown on blood agar plates under anaerobic conditions at 37°C for 48 h, resuspended in PBS, and incubated at 4°C statically. The optical density (600 nm) in the upper half of the test tubes was measured periodically (Supplementary Fig. S1b).

P. micra *isolation from tumoral colonic biopsy*

Biopsies of tumoral colonic mucosa from colorectal cancer patients were recovered at Henri Mondor Hospital. After colonoscopy, biopsies were immediately placed in 0.5 mL Schaedler liquid media and maintained at room temperature and in anaerobic condition for transport. On the same day, biopsies were placed inside an anaerobic cabinet (0% of O_2) where all steps of the isolation procedure were achieved. Biopsies were transferred in 0.5 mL of Schaedler media complemented with EDTA-free Protease Inhibitor Cocktail (Sigma) and 0.1% triton (Sigma), containing 0.2-mm-diameter glass beads (EMSCO) and tissue was disrupted upon bead beating agitation carried out at 20/s frequency for 5 min using a Beadbeater (Retsch MM400). Anti-*Parvimonas* serum (on order by Covalab), able to recognize both phylotypes, was added to the mixture at 1:100 and incubated for 30 min with agitation. After three PBS washes of tissue fragments, 10^6 magnetic beads in 0.2 mL PBS (Dynabeads ProteinA, Invitrogen) were added and incubated 15 min at room temperature upon agitation. Using a magnet, the beads were washed three times and further incubated for 48 h at 37°C in TGY-V media. Several dilutions of the growth culture media were plated on blood agar plates, and CFUs were identified using PCR (*P. micra*; PF: 5'-GACGGGCGGTGWGTRCA-3'; PR: 5'-AGAGTTTGATCCTGGGTCAG-3').

Transmission electron microscopy

Bacteria from 4-day-old cultures were fixed by adding glutaraldehyde (2.5% final concentration) to the culture for 1 h at room temperature (RT). The bacteria were washed twice with HEPES 0.1 M, pH 7.5 and postfixed with 1% osmium tetroxide in HEPES 0.1 M pH 7.5 for 1h30 at RT. After three washes in distilled water, the bacteria were dehydrated in a graded ethanol series (25%, 50%, 75%, 95%, and 100% ethanol) and gradually infiltrated in Epon resin. Ultrathin sections (60 nm) were obtained on a FC6/UC6 ultramicrotome (Leica). Sections were transferred to 200 Mesh Square Copper grids (CF-200-CuO, Delta Microscopy) formvar and carbon coated, stained with 4% uranyl acetate, and counterstained with Reynold's lead citrate for 20 min. The images were recorded with TECNAI SPIRIT 120 Kv, with a bottom-mounted EAGLE 4KX4K Camera.

Colonic cell cultures

Cancer cell lines: human colon cell lines HT-29, HCT116, and TC7 (subclone of Caco-2) were grown in DMEM (Dulbecco Modified Eagle Medium) 1 g/L glucose (Gibco), supplemented with 20% (vol/vol) decomplemented fetal bovine serum (FBS) (Gibco), 1% non-essential amino acids (Gibco), 1X GlutaMAX (Gibco), at 37°C in 10% CO_2.

Bacterial adhesion

Adhesion to extracellular matrix proteins: MaxiSorp 96-well flat-bottom plates (ThermoFisher) were coated overnight at 4°C under agitation with 1 μg per well, in triplicate, with the following extracellular matrix proteins: collagen I (Gibco), collagen IV (Millipore), fibronectin (Gibco), fibrinogen (Sigma), laminin (Sigma), and Matrigel® (consisting of a mixture of proteins, mainly laminin, collagen IV, entactin, and heparan sulfate proteoglycan) (Corning). Bacteria were grown on COH plates for 48 h at 37°C under anaerobic conditions, collected and suspended in PBS at an OD of three units. The bacterial suspension was placed in the wells, and the plate was centrifuged at 300 g for 5 min. After incubation for 1 h at RT, the wells were washed three times with PBS to remove non-adherent bacteria and then stained with 0.1% crystal violet (Sigma). The plates were incubated for 30 min at RT and washed three more times with PBS. The dye was dissolved with 20:80 acetone-ethanol, and the optical density at 595 nm was read by a spectrometer (Infinite M200Pro-TECAN).

P. micra adhesion to colonic cell lines: human colonic cell lines (TC7, HT-29, and HCT116) were grown on glass coverslips in 24-well plates to confluence. In parallel, *P. micra* was grown on COH plates for 48 h at 37°C under anaerobic conditions. Bacteria were washed in PBS, resuspended in DMEM with 1 g/L glucose (Gibco), and placed on the cells using an MOI of 1. The plates were centrifuged for 5 min at 300 g and incubated for 1 h at 37°C. After incubation, the cells were washed three times in PBS, fixed in 4% paraformaldehyde (PFA) (Electron Microscopy Sciences) for 30 min at RT, washed again three times in PBS, and then stored in 0.1% PBS-BSA at 4°C until labeling. Non-specific sites were blocked with 1% PBS-BSA for 1 h at RT. The cells were then incubated with rabbit anti-*Parvimonas* serum (on order by Covalab) at 1:1000 in 1% PBS-BSA for 30 min at room temperature. After washing, the cells were permeabilized with PBS- 0.2% triton for 30 min. Anti-rabbit secondary antibody, coupled to Alexa Fluor 488 (ThermoFisher) at 1:400, and phalloidin, coupled to Alexa Fluor 568 (Invitrogen) at 1:200, were used in 1% PBS-BSA for 45 min at RT. After washing, the nucleic acids were labeled with 1 µg/mL DAPI (ThermoFisher) for 5 min. The coverslip with the cells was washed again and mounted on a slide with Prolong Gold (ThermoFisher). The images were taken using a slide scanner with 40X objective (AxioScan, Zeiss) and processed using Fiji[75]. Bacterial adhesion was quantified as the area of the anti-*Parvimonas* fluorescence signal over an entire field normalized by the total number of nuclei and reported as arbitrary fluorescent units x100.

Human colonic organoids

Human colonic surgical resection specimens were obtained from Henri Mondor Hospital from two patients who had undergone colectomy surgery from rectal adenocarcinoma and had given their informed consent (agreement N°2012–37). Tissue samples from the normal site (far from tumor tissue) were sterilely washed using PBS supplemented with gentamicin (50 µg/mL), Normoxin (1 mg/mL) and amphotericin B (2 µg/mL), to obtain bacteria-free normal colonic mucosa and kept in 0.1%

PBS-BSA at 4°C for the duration of transport (approximately 4 h). Crypts were isolated according to the protocol of Sato et al.,[76] with some modifications. The epithelium was stripped of underlying muscularis and serosa, washed several times in cold PBS until the supernatant was clear, and cut into 5 mm^2 fragments. The fragments were incubated for 20 min on ice in a cold chelation buffer consisting of 5.6 mmol/L Na_2HPO_4, 8.0 mmol/L KH_2PO_4, 96 mmol/L NaCl, 1.6 mmol/L KCl, 44 mmol/L sucrose, 54.8 mmol/L D-sorbitol, and 0.5 mmol/L DL-dithiothreitol in distilled water plus 2 mmol/L EDTA. After removal of the supernatant and addition of cold chelation buffer without EDTA, the tissue fragments were vigorously resuspended by several passes through a 10 mL serological pipette. After sedimentation of the tissue fragments, the supernatant, containing the crypts, was recovered. The resuspension/sedimentation procedure was repeated twice. The tissue fragments were again incubated in a chelation buffer with EDTA for 15 min and the resuspension/sedimentation procedure was repeated three times. Fractions containing crypts were pooled and centrifuged at 300 g for 5 min. The crypt pellet was resuspended in Matrigel® growth factor-reduced medium (Corning) diluted to 75% in culture medium (see composition below). Four 25 µl drops of the Matrigel-crypt mixture were placed per well in 12-well plates with approximately 100 crypts per drop. The plates were incubated for 15 min at 37°C to allow polymerization of the Matrigel, then 800 µL of culture medium was added, and the plates were incubated at 37°C and 5% CO_2. The culture medium consisted of AdvancedDMEM/F12 (Gibco), 10 mM HEPES (Gibco), 1X GlutaMAX (Gibco), 100 u/mL penicillin, 100 µg/mL streptomycin (Gibco), 1X N2 (ThermoFisher), 1X B27 (ThermoFisher), 1 mM N-acetyl-L-cysteine (Sigma), 100 ng/mL Noggin (R&D systems), 50 ng/mL recombinant human EGF (R&D systems), 150 ng/mL Wnt-3A (R&D systems), 1 µg/mL recombinant human R-spondin-1 (R&D systems), 500 nM A83–01 (R&D systems), 10 mM nicotinamide (Sigma), 10 µM SB202190 (Sigma), 10 nM Gastrin I (Sigma), 3 µM CHIR99021 (Biogems), 10 µM Y-27632 (Sigma) and 10% decomplemented fetal bovine serum (vol/vol) (Gibco). The culture medium was changed every 2 days, without Y-

27632. Each week, organoids were split with a 1:3 ratio. The medium was replaced by cold fresh medium, and the organoids embedded in matrigel were dissociated by pipetting several times through a 200 µl tip and collect in a tube. After a spin for 5 min at 300 g at 4°C, the pellet was washed with fresh cold medium, resuspended with a p200 pipette several times, and spun again. The pellet of organoid fragments was resuspended in Matrigel® as described above. After expansion, organoids were frozen and stored at −80°C in freezing medium composed of Advanced DMEM/F12, 10% fetal bovine serum, 10% DMSO (PAN Biotech), 10 µM Y-27632, and 3 µM CHIR99021.

Generation of human colonic epithelial monolayer

Monolayers of organoid-derived primary cells were generated as previously described.[25] Briefly, the 24 well plates with 0.4 µm pores (Corning) Transwell® insert system were used as culture support for primary cells in monolayers. Inserts were previously coated with 50 µl of human collagen IV at 30 µg/ml (Millipore) overnight at 4°C, then washed with DMEM. Approximately 300 organoids at 5 days of growth were dissociated into fragments (using Cell Recovery (Corning)), resuspended in 200 µl of culture medium and put on each insert. The culture medium consisted of DMEM/F12 (Gibco), 10 mM HEPES (Gibco), 1X GlutaMAX (Gibco), 100 u/mL penicillin, 100 µg/mL streptomycin (Gibco), 1X N2 (ThermoFisher) and 1X B27 (ThermoFisher) supplements, 1 mM N-acetyl-L-cysteine (Sigma), 100 ng/mL noggin (R&D systems), 50 ng/mL recombinant human EGF (R&D systems), 150 ng/mL Wnt-3A (R&D systems), 1 µg/mL recombinant human R-spondin-1 (R&D systems), 500 nM A83–01 (R&D systems), 10 mM nicotinamide (Sigma), 10 µM SB202190 (Sigma), 10 nM Gastrin I (Sigma), 3 µM CHIR99021 (Biogems), 10 µM Y-27632 (Sigma) and 10% decomplemented fetal bovine serum (vol/vol) (Gibco). Monolayers were incubated at 37°C in 5% CO_2. After 3 days of culture, the medium was changed without addition of Y-27632. Cell growth and monolayer closure were monitored daily by observation with an inverted brightfield microscope (IX81, Olympus) and by recording the transepithelial electrical resistance (Millipore).

Co-culture of colonic cells and P. micra

Bacteria were grown in TGY-V medium under anaerobic conditions at 37°C for 48 h, followed by a 1:10 subculture for 24 h to reach exponential phase. Bacterial density was adjusted to an MOI of 250 in cell culture medium and then put on the cells. Twenty-four hours prior to bacterial exposure, the colonic cells were placed in 2% O_2 in H35 hypoxic cabinet (Don Whitley) to allow acclimatization. Co-cultures with primary cells were performed for 48 h at 37°C under hypoxic conditions (2% O_2, 5% CO_2).

Flow cytometry

Cell monolayers were dissociated with TrypLE™ Express Enzyme (Gibco) at 37°C for 10 to 20 min. The cells were recovered in DMEM/F12 at RT and then centrifuged for 5 min at 300 g. After washing the pellet in PBS, cells were stained with the Live/Dead fixable cell stain kit (ThermoFisher) for 20 min on ice, following the supplier's instructions. Cells were analyzed using a FACS Attune (ThermoFisher), and data were analyzed using FlowJo software (v10.1).

Ki-67, γH2a×, Muc2, and NF-κB immunofluorescence

Following 48 h co-culture with *P. micra*, the cells were washed with PBS, fixed (4% PFA, 30 min), washed again, and stored (0.1% PBS-BSA) at 4°C until labeling. Cells were permeabilized with 0.5% Triton-PBS for 30 min at RT, washed in 0.1% PBS-Triton, and blocked with 1% bovine serum albumin (BSA)-0.1% Triton for 30 min. The cells were then incubated either i) with a 1:200 dilution of an antibody against the proliferation marker Ki-67 coupled to Alexa Fluor 666 (clone SolA15, eBioscience) and a 1:500 dilution of an antibody against the phosphorylated form of H2A histone family member X (γH2Ax) (clone JBW301, Merck) and 1:200 phalloidin coupled to Alexa Fluor

568; ii) a 1:50 dilution of an antibody against nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) p65 (ab7970 Abcam) and 1:200 phalloidin coupled to Alexa Fluor 568; or iii) with a 1:100 dilution of an anti-mucin2 (Muc2) antibody (clone Ccp58, Abcam) and 1:200 phalloidin coupled to Alexa Fluor 568, in 0.1% Triton-BSA overnight at 4°C. After washing in 0.1% PBS-Triton, the cells were incubated with 1:400 secondary antibodies coupled to Alexa Fluor 488 (ThermoFisher) in 0.1% PBS-Triton for 45 min at RT. Cells were washed again in 0.1% PBS-Triton and the nucleic acids were labeled with 1 μg/mL DAPI for 5 min before cells were mounted on a slide with ProlongGold (ThermoFisher). Fluorescence images were taken using a confocal fluorescence microscope (Zeiss) equipped with an Opterra system (Bruker), and the number of cells was analyzed as the number of positive cells/total cells using Fiji[75] or Imaris 9.8 software (URL: http://www.bitplane.com/imaris/imaris).

Global DNA methylation level

Human cells co-cultured with *P. micra* were subjected to DNA extraction using the DNeasy Blood & Tissue Kit® (Qiagen) according to the manufacturer's instructions, and DNA was quantified using a Qubit® fluorometer. MethylFlash methylated DNA Quantification Fluorometer Kit (Epigentek) was used to detect the global methylation level in DNA isolated from primary colonic cells co-cultured with *P. micra* ATCC 33270 (*PmA*), HHM BlNa17 (*PmB*) and *F. magna*. Assays were performed in duplicate. As instructed by the manufacturer, 100 ng of the isolated genomic DNA was bound to the assay well. The capture antibody, detection antibody, and enhancer solution were then added consecutively to the wells before the fluoro-developing solution was added, and the relative fluorescence units (RFU) were measured (TECAN). The percentage of 5-methylcytosine (5-mC) relative to the total amount of cytosines in the sample was calculated to represent the global methylation dosage and was reported as a ratio to the non-stimulated samples.

DNA methylation profiling

DNA methylation profiling was performed on eight independent co-cultures of *PmA*, *PmB*, and *F. magna* with human primary colonic cells (four experiments per patient tissue). The Infinium Methylation EPIC BeadChip Kit (850K) (Illumina, San Diego, CA, USA) was used for a genome-wide methylation profiling to determine the DNA methylation status of >850,000 CpG sites.[77] A total of 500 ng of genomic DNA from cells in co-cultures were bisulfite-treated using the ZymoResearch EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA). The Infinium HD Methylation Assay (bisulfite modification, amplification, fragmentation, precipitation, hybridization, washing, extension, staining, and imaging) was performed according to the manufacturer's explicit specifications. The quality was supported by multiple quality control (QC) measures, including tests for proper bisulfite conversion, staining, and specificity of the internal controls, as determined by Illumina GenomeStudio software. Average beta values (proxy for methylated DNA level between 0, unmethylated, and 1, fully methylated) were normalized to internal controls and corrected by background subtraction. Non-autosomal CpGs ($n = 14,522$) and CpG probes with suboptimal detection ($p < .05$ in 80% of the samples) ($n = 29,316$), as well as single nucleotide polymorphisms (SNPs) associated CpGs ($n = 59$), were removed from our analyses. Beta values were transformed into M-values as described by[78], and differential probes were assessed using the limma statistical model[79] package R limma, version 4.0.2. The four conditions (Non-stimulated NS; *PmA*; *PmB* and FM) and individuals were kept as variables. Pairwise comparisons were obtained from a contrast matrix. No adjustment method was used. For each gene ($n = 25,560$), a t-test on log-fold-change was performed on four categories of probes depending on their CpG context and gene localization: probes in CpG island context and 1/localized in gene promoters (TSS1500 +TSS200) or 2/in body gene; and probes out of CpG island context 3/in promoters or 4/in body gene. Genes with >2 probes, p-value <.05, and with a log-FC > 0.1 (absolute value) were considered significant. Using three different databases, the genes were classified as oncogenes, tumor suppressor genes, or involved in Epithelial–Mesenchymal Transition (ONGene,[80] TSGene 2.0,[81] or dbEMT 2.0).[82]

Reduced representation bisulfite sequencing (RRBS)

Genomic DNA was isolated from Formalin-fixed par-affin-embedded (FFPE) tissue blocks of tumoral colonic tissue, using the Maxwell RSC® DNA FFPE kit (Promega). DNA concentration was measured with the Qubit® dsDNA HS Assay Kit (ThermoFisher), and DNA quality was assessed by gel electrophoresis. Each experimental group (*P. micra* – negative and – positive) contained three different CRC patients. DNA Methylation Profiling (RRBS Service) was performed using Active motif (Carlsbad); DNA digestion, adaptor ligation, bisulfite conversion, library amplification, sequencing, and data processing. Briefly, Bismark v0.23.0 (https://www.bioinformatics.babraham.ac.uk/projects/bismark/) was used to align reads to a human reference genome (hg38). Methylation analysis was performed with the MethylKit 1.18.0 software.[83] Pairwise comparisons using Fisher's test were conducted to identify differentially methylated regions (DMRs), consisting of 1,000 bp regions. The SLIM method (sliding window model) was used to correct for multiple comparisons and convert p-values to Adjusted p-value (q-values). Significant DMRs were defined as having a methylation fold change difference greater than 25% and q-value <0.1. Gene name, genomic region, and location in relation to CpG islands were identified for each DMR.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Giulia Nigro 🔟 http://orcid.org/0000-0002-1014-7394

Author contributions

Clinical and translational concept IS; basic conceptualization of *in vitro* study PS, GN, and EB; methodology GN, EB, AM, EL; investigation CB, FD, TP, GN, and EB; formal analysis AD, DM, SV, AV, LTM, GN, and EB; writing – original draft GN and EB; writing – review and editing GN, PS, and IS; funding acquisition and resources PS, MC, IS AM, and EL; supervision PS, IS, and GN.

Data availability statement

Data from Infinium MethylationEPIC BeadChip from Illumina.
Bergsten, Emma; Nigro, Giulia (2023), "Dataset DNA Methylation," Mendeley Data, V1, doi: 10.17632/g4nzmtdpwr.1.

References

1. Jiao S, Peters U, Berndt S, Brenner H, Butterbach K, Caan BJ, Carlson CS, Chan AT, Chang-Claude J, Chanock S, et al. Estimating the heritability of colorectal cancer. Hum Mol Genet. 2014;23(14):3898–3905. doi:10.1093/hmg/ddu087.
2. Samad AKA, Taylor RS, Marshall T, Chapman MAS. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. Colorect Dis. 2005;7(3):204–213. doi:10.1111/j.1463-1318.2005.00747.x.
3. Kyrgiou M, Kalliala I, Markozannes G, Gunter MJ, Paraskevaidis E, Gabra H, Martin-Hirsch P, Tsilidis KK. Adiposity and cancer at major anatomical

sites: umbrella review of the literature. BMJ. 2017;477:365. doi:10.1136/bmj.j477.

4. Chan DSM, Lau R, Aune D, Vieira R, Greenwood DC, Kampman E, Norat T, Tomé D. Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. PloS One. 2011;6(6):e20456. doi:10.1371/journal.pone.0020456.

5. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer. 2013;13(11):800–812. doi:10.1038/nrc3610.

6. Soderholm AT, Pedicord VA. Intestinal epithelial cells: at the interface of the microbiota and mucosal immunity. Immunology. 2019;158(4):267–280. doi:10.1111/imm.13117.

7. Goodwin AC, Shields CED, Wu S, Huso DL, Wu X, Murray-Stewart TR, Hacker-Prietz A, Rabizadeh S, Woster PM, Sears CL, et al. Polyamine catabolism contributes to enterotoxigenic bacteroides fragilis-induced colon tumorigenesis. Proc Natl Acad Sci USA. 2011;108(37):15354–15359. doi:10.1073/pnas.1010203108.

8. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-Cadherin/β-catenin signaling via its FadA adhesin. Cell Host & Microbe. 2013;14(2):195–206. doi:10.1016/j.chom.2013.07.012.

9. Pasquereau-Kotula E, Martins M, Aymeric L, Dramsi S. Significance of streptococcus gallolyticus subsp. gallolyticus association with colorectal cancer. Front Microbiol. 2018;9:614. doi:10.3389/fmicb.2018.00614.

10. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, Gurjao C, Manders F, Dalmasso G, Stege PB, et al. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature. 2020;580(7802):269–273. doi:10.1038/s41586-020-2080-8.

11. Saffarian A, Mulet C, Regnault B, Amiot A, Tran-Van-Nhieu J, Ravel J, Sobhani I, Sansonetti PJ, Pédron T. Crypt- and mucosa-associated core microbiotas in humans and their alteration in colon cancer patients. mBio. 2019;10(4):e01315–19. doi:10.1128/mBio.01315-19.

12. Rams TE, Feik D, Listgarten MA, Slots J. Peptostreptococcus micros in human periodontitis. Oral Microbiol Immunol. 1992;7(1):1–6. doi:10.1111/j.1399-302X.1992.tb00011.x.

13. Kamada N, Chen GY, Inohara N, Núñez G. Control of pathogens and pathobionts by the gut microbiota. Nat Immunol. 2013;14(7):685–690. doi:10.1038/ni.2608.

14. Socransky SS, Haffajee AD, Ximenez-Fyvie LA, Feres M, Mager D. Ecological considerations in the treatment of actinobacillus actinomycetemcomitans and Porphyromonas gingivalis periodontal infections. Periodontol 2000. 1999;20(1):341–362. doi:10.1111/j.1600-0757.1999.tb00165.x.

15. Mousa HA-L. Aerobic, anaerobic and fungal burn wound infections. J Hosp Infect. 1997;37(4):317–323. doi:10.1016/S0195-6701(97)90148-1.

16. Bowler PG, Davies BJ. The microbiology of infected and noninfected leg ulcers: microbiology of leg ulcers report. Int J Dermatol. 1999;38(8):573–578. doi:10.1046/j.1365-4362.1999.00738.x.

17. Bowler PG, Davies BJ. The microbiology of acute and chronic wounds. Wounds. 1999;11:72–78.

18. Edmiston C Jr., Gohr C, Walker AP. The nonpuerperal breast infection: aerobic and anaerobic Microbial recovery from acute and chronic disease. J Infect Dis. 1990;162:695–699. doi:10.1093/infdis/162.3.695.

19. Sanderson PJ, Wren MW, Baldwin AW. Anaerobic organisms in postoperative wounds. J Clin Pathol. 1979;32(2):143–147. doi:10.1136/jcp.32.2.143.

20. Murdoch DA, Mitchelmore IJ, Tabaqchali S. Peptostreptococcus micros in polymicrobial abscesses. Lancet. 1988;331(8585):594. doi:10.1016/S0140-6736(88)91393-1.

21. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, O'Riordain M, Shanahan F, O'Toole PW. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017;66(4):633–643. doi:10.1136/gutjnl-2015-309595.

22. Gao R, Kong C, Huang L, Li H, Qu X, Liu Z, Lan P, Wang J, Qin H. Mucosa-associated microbiota signature in colorectal cancer. Eur J Clin Microbiol Infect Dis. 2017;36(11):2073–2083. doi:10.1007/s10096-017-3026-4.

23. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 2014;10(11):766. doi:10.15252/msb.20145645.

24. Sobhani I, Bergsten E, Couffin S, Amiot A, Nebbad B, Barau C, De'angelis N, Rabot S, Canoui-Poitrine F, Mestivier D, et al. Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures. Proc Natl Acad Sci U S A. 2019;116:24285–24295. doi:10.1073/pnas.1912129116.

25. Noel G, Baetz NW, Staab JF, Donowitz M, Kovbasnjuk O, Pasetti MF, Zachos NC. A primary human macrophage-enteroid co-culture model to investigate mucosal gut physiology and host-pathogen interactions. Sci Rep. 2017;7(1):45270. doi:10.1038/srep45270.

26. Kremer BHA, Magee JT, van Dalen PJ, van Steenbergen TJM. Characterization of smooth and rough morphotypes of Peptostreptococcus micros. Int J Syst Bacteriol. 1997;47(2):363–368. doi:10.1099/00207713-47-2-363.

27. Deplancke B, Gaskins HR. Microbial modulation of innate defense: goblet cells and the intestinal mucus layer. Am J Clin Nutr. 2001;73(6):1131S–1141S. doi:10.1093/ajcn/73.6.1131S.

28. Nougayrède J-P, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J,

Dobrindt U, Oswald E. *Escherichia coli* induces DNA Double-strand breaks in Eukaryotic cells. Sci. 2006;313 (5788):848–851. doi:10.1126/science.1127059.

29. Lin Q, Li J, Zhu D, Niu Z, Pan X, Xu P, Ji M, Wei Y, Xu J. Aberrant Scinderin expression correlates with liver metastasis and poor prognosis in colorectal cancer. Front Pharmacol. 2019;10:1183. doi:10.3389/fphar.2019.01183.

30. Zhang L, Anglesio MS, O'Sullivan M, Zhang F, Yang G, Sarao R, Nghiem MP, Cronin S, Hara H, Melnyk N, et al. The E3 ligase HACE1 is a critical chromosome 6q21 tumor suppressor involved in multiple cancers. Nat Med. 2007;13(9):1060–1069. doi:10.1038/nm1621.

31. Lou J, Huang J, Dai X, Xie Y, Dong M, Chen B, Zhao J, Zhou B, Yu H. Knockdown of tetraspanin 13 inhibits proliferation of colorectal cancer cells. Int J Clin Exp Med. 2017;10:6387–6395.

32. Rana MK, Aloisio FM, Choi C, Barber DL, Gardel M. Formin-dependent TGF-β signaling for epithelial to mesenchymal transition. MBoC. 2018;29(12):1465–1475. doi:10.1091/mbc.E17-05-0325.

33. Gao Z, Guo B, Gao R, Zhu Q, Qin H. Microbiota dysbiosis is associated with colorectal cancer. Front Microbiol. 2015;6:20. doi:10.3389/fmicb.2015.00020.

34. Baxter NT, Zackular JP, Chen GY, Schloss PD. Structure of the gut microbiome following colonization with human feces determines colonic tumor burden. Microbiome. 2014;2(1):20. doi:10.1186/2049-2618-2-20.

35. Schmidt TS, Hayward MR, Coelho LP, Li SS, Costea PI, Voigt AY, Wirbel J, Maistrenko OM, Alves RJ, Bergsten E, et al. Extensive transmission of microbes along the gastrointestinal tract. eLife. 2019;8:e42693. doi:10.7554/eLife.42693.

36. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 2017;66(1):70–78. doi:10.1136/gutjnl-2015-309800.

37. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, Ng SC, Tsoi H, Dong Y, Zhang N, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun. 2015;6(1):8727. doi:10.1038/ncomms9727.

38. Koliarakis I, Messaritakis I, Nikolouzakis TK, Hamilos G, Souglakos J, Tsiaoussis J. Oral bacteria and intestinal dysbiosis in colorectal cancer. Int J Mol Sci. 2019;20(17):4146. doi:10.3390/ijms20174146.

39. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, Peterson SN, Snesrud EC, Borisy GG, Lazarev M, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. Proc Natl Acad Sci U S A. 2014;111(51):18321–18326. doi:10.1073/pnas.1406199111.

40. Dejea CM, Sears CL. Do biofilms confer a pro-carcinogenic state? Gut Microbes. 2016;7(1):54–57. doi:10.1080/19490976.2015.1121363.

41. Drewes JL, White JR, Dejea CM, Fathi P, Iyadorai T, Vadivelu J, Roslani AC, Wick EC, Mongodin EF, Loke MF, et al. High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. NPJ Biofilms Microbiomes. 2017;3(1):34. doi:10.1038/s41522-017-0040-3.

42. Li S, Konstantinov SR, Smits R, Peppelenbosch MP. Bacterial biofilms in colorectal cancer initiation and progression. Trends Mol Med. 2017;23(1):18–30. doi:10.1016/j.molmed.2016.11.004.

43. Horiuchi A, Kokubu E, Warita T, Ishihara K. Synergistic biofilm formation by Parvimonas micra and Fusobacterium nucleatum. Anaerobe. 2020;62:102100. doi:10.1016/j.anaerobe.2019.102100.

44. Périchon B, Lichtl-Häfele J, Bergsten E, Delage V, Trieu-Cuot P, Sansonetti P, Sobhani I, Dramsi S. Detection of streptococcus gallolyticus and four other CRC-Associated bacteria in patient stools reveals a potential "driver" role for enterotoxigenic bacteroides fragilis. Front Cell Infect Microbiol. 2022;12:794391. doi:10.3389/fcimb.2022.794391.

45. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759–767. doi:10.1016/0092-8674(90)90186-I.

46. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol. 2012;10 (8):575–582. doi:10.1038/nrmicro2819.

47. Zhao L, Zhang X, Zhou Y, Fu K, Lau HC-H, Chun TW-Y, Cheung AH-K, Coker OO, Wei H, Wu W-K-K, et al. Parvimonas micra promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients. Oncogene. 2022;41(36):4200–4210. doi:10.1038/s41388-022-02395-7.

48. Chang Y, Huang Z, Hou F, Liu Y, Wang L, Wang Z, Sun Y, Pan Z, Tan Y, Ding L, et al. Parvimonas micra activates the Ras/ERK/c-Fos pathway by upregulating miR-218-5p to promote colorectal cancer progression. J Exp Clin Cancer Res. 2023;42(1):13. doi:10.1186/s13046-022-02572-2.

49. Purcell RV, Visnovska M, Biggs PJ, Schmeier S, Frizelle FA. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. Sci Rep. 2017;7(1):11590. doi:10.1038/s41598-017-11237-6.

50. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350–1356. doi:10.1038/nm.3967.

51. Becht E, de Reyniès A, Giraldo NA, Pilati C, Buttard B, Lacroix L, Selves J, Sautès-Fridman C, Laurent-Puig P, Fridman WH. Immune and stromal classification of colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. Clin Cancer Res. 2016;22(16):4057–4066. doi:10.1158/1078-0432.CCR-15-2879.

52. Xia X, Wu WKK, Wong SH, Liu D, Kwong TNY, Nakatsu G, Yan PS, Chuang Y-M, Chan MW-Y, Coker OO, et al. Bacteria pathogens drive host colonic epithelial cell promoter hypermethylation of tumor suppressor genes in colorectal cancer. Microbiome. 2020;8(1):108. doi:10.1186/s40168-020-00847-4.

53. Shimizu A, Mammoto A, Italiano JE, Pravda E, Dudley AC, Ingber DE, Klagsbrun M. ABL2/ARG Tyrosine Kinase Mediates SEMA3F-induced RhoA Inactivation and cytoskeleton collapse in human glioma cells. J Biol Chem. 2008;283(40):27230–27238. doi:10.1074/jbc.M804520200.

54. Martini M, Gnann A, Scheikl D, Holzmann B, Janssen K-P. The candidate tumor suppressor SASH1 interacts with the actin cytoskeleton and stimulates cell–matrix adhesion. Int J Biochem Cell Biol. 2011;43:1630–1640. doi:10.1016/j.biocel.2011.07.012.

55. Torrino S, Visvikis O, Doye A, Boyer L, Stefani C, Munro P, Bertoglio J, Gacon G, Mettouchi A, Lemichez E. The E3 ubiquitin-ligase HACE1 Catalyzes the ubiquitylation of active Rac1. Dev Cell. 2011;21 (5):959–965. doi:10.1016/j.devcel.2011.08.015.

56. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, Corthier G, Van Nhieu JT, Furet JP, Pied S. Microbial Dysbiosis in Colorectal Cancer (CRC) patients. PloS One. 2011;6(1):e16393. doi:10.1371/journal.pone.0016393.

57. Furet J-P, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S, Doré J, Corthier G. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR: human and farm animal faecal microbiota. FEMS Microbiol Ecol. 2009;68 (3):351–362. doi:10.1111/j.1574-6941.2009.00671.x.

58. Bağcı C, Beier S, Górska A, Huson DH. Introduction to the analysis of environmental sequences: metagenomics with MEGAN. In: Anisimova M, editor. Evolutionary genomics methods in molecular biology. Springer New York;2019. pp. 591–604. doi:10.1007/978-1-4939-9074-0_19.

59. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1): D733–D745. doi:10.1093/nar/gkv1189.

60. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R, Poisot T. MEGAN community edition - interactive exploration and analysis of large-Scale microbiome sequencing data. PLoS Comput Biol. 2016;12(6):e1004957. doi:10.1371/journal.pcbi.1004957.

61. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012;41(D1): D590–D596. doi:10.1093/nar/gks1219.

62. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med. 2016;8(1):37. doi:10.1186/s13073-016-0290-3.

63. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32 (19):3047–3048. doi:10.1093/bioinformatics/btw354.

64. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29 (1):15–21. doi:10.1093/bioinformatics/bts635.

65. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30 (7):923–930. doi:10.1093/bioinformatics/btt656.

66. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29. doi:10.1186/gb-2014-15-2-r29.

67. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47–e47. doi:10.1093/nar/gkv007.

68. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann Appl Stat. 2016;10 (2):946. doi:10.1214/16-AOAS920.

69. Ge SX, Jung D, Yao R, Valencia A. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36(8):2628–2629. doi:10.1093/bioinformatics/btz931.

70. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29(14):1830–1831. doi:10.1093/bioinformatics/btt285.

71. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–D551. doi:10.1093/nar/gkaa970.

72. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57. doi:10.1038/nprot.2008.211.

73. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13. doi:10.1093/nar/gkn923.

74. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3693. doi:10.1093/bioinformatics/btv421.

75. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9 (7):676–682. doi:10.1038/nmeth.2019.

76. Sato T, Stange DE, Ferrante M, Vries RGJ, van Es JH, van den Brink S, van Houdt WJ, Pronk A, van Gorp J, Siersema PD, et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology. 2011;141 (5):1762–1772. doi:10.1053/j.gastro.2011.07.050.

77. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389–399. doi:10.2217/epi.15.114.

78. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinform. 2010;11 (1):587. doi:10.1186/1471-2105-11-587.

79. Wettenhall JM, Smyth GK. limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics. 2004;20(18):3705–3706. doi:10.1093/bioinformatics/bth449.

80. Liu Y, Sun J, Zhao M. Ongene: a literature-based database for human oncogenes. J Genet Genomics. 2017;44 (2):119–121. doi:10.1016/j.jgg.2016.12.004.

81. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. Tsgene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 2016;44(D1):D1023–D1031. doi:10.1093/nar/gkv1268.

82. Zhao M, Liu Y, Zheng C, Qu H. dbEMT 2.0: an updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. J Genet Genomics. 2019;46(12):595–597. doi:10.1016/j.jgg.2019.11.010.

83. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012;13(10):R87. doi:10.1186/gb-2012-13-10-r87.