

Towards the Human Colorectal Cancer Microbiome

Julian R. Marchesi¹, Bas E. Dutilh^{2,3}, Neil Hall⁴, Wilbert H. M. Peters⁵, Rian Roelofs⁶, Annemarie Boleij⁶, Harold Tjalsma^{6*}

1 School of Biosciences, Cardiff University, Cardiff, United Kingdom, **2** Centre for Molecular and Biomolecular Informatics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **3** Departments of Computer Science and Biology, San Diego State University, San Diego, California, United States of America, **4** Centre for Genomic Research, School of Biological Sciences, University of Liverpool, Liverpool, United Kingdom, **5** Department of Gastroenterology, Nijmegen Institute for Infection, Inflammation and Immunity (N4i) & Radboud University Centre for Oncology (RUCO) of the Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **6** Department of Laboratory Medicine, Nijmegen Institute for Infection, Inflammation and Immunity (N4i) & Radboud University Centre for Oncology (RUCO) of the Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Abstract

Multiple factors drive the progression from healthy mucosa towards sporadic colorectal carcinomas and accumulating evidence associates intestinal bacteria with disease initiation and progression. Therefore, the aim of this study was to provide a first high-resolution map of colonic dysbiosis that is associated with human colorectal cancer (CRC). To this purpose, the microbiomes colonizing colon tumor tissue and adjacent non-malignant mucosa were compared by deep rRNA sequencing. The results revealed striking differences in microbial colonization patterns between these two sites. Although inter-individual colonization in CRC patients was variable, tumors consistently formed a niche for *Coriobacteria* and other proposed probiotic bacterial species, while potentially pathogenic *Enterobacteria* were underrepresented in tumor tissue. As the intestinal microbiota is generally stable during adult life, these findings suggest that CRC-associated physiological and metabolic changes recruit tumor-foraging commensal-like bacteria. These microbes thus have an apparent competitive advantage in the tumor microenvironment and thereby seem to replace pathogenic bacteria that may be implicated in CRC etiology. This first glimpse of the CRC microbiome provides an important step towards full understanding of the dynamic interplay between intestinal microbial ecology and sporadic CRC, which may provide important leads towards novel microbiome-related diagnostic tools and therapeutic interventions.

Citation: Marchesi JR, Dutilh BE, Hall N, Peters WHM, Roelofs R, et al. (2011) Towards the Human Colorectal Cancer Microbiome. PLoS ONE 6(5): e20447. doi:10.1371/journal.pone.0020447

Editor: Niyaz Ahmed, University of Hyderabad, India

Received: February 18, 2011; **Accepted:** April 22, 2011; **Published:** May 24, 2011

Copyright: © 2011 Marchesi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: AB was supported by the Dutch Cancer Society (KWF; Project KUN 2006-3591). Dutch Digestive Diseases Foundation (MDLS; project WO 10-53). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: H.Tjalsma@labgk.umcn.nl.

Introduction

The human intestinal tract contains about 10^{14} bacteria, comprising $\sim 10^3$ species, which are essential for digestion of food, the control of intestinal epithelial homeostasis, intestinal development and human health [1]. Conversely, a large body of evidence supports a relationship between infective agents and human cancers [2] and suggests that certain mucosa-associated bacterial species play an important role in the pathogenesis of colorectal cancer (CRC; [3,4,5]). Moreover, clinical associations between bacterial infection and CRC have been described for many decades, the most prominent of which concern infections with *Streptococcus bovis* [6,7] and *Clostridium septicum* [8]. However, the co-incidence of these infections with CRC is very low (<1%) since such low-grade opportunistic pathogens can only become clinically manifest in compromised patients. Correspondingly, serological data have shown an increased exposure to *S. bovis* antigens in early stage CRC patients without clinical signs of bacterial infection [9]. Based on this, it has been suggested that specific gut bacteria have a competitive advantage in the CRC microenvironment, whereas opportunistic infections remain repressed by the active immune system in the majority of patients.

Recent publications have provided mechanistic evidence for the involvement of gut bacteria in the development of CRC, which comprises i) production of DNA damaging superoxide radicals, ii) production of genotoxins, iii) T helper cell-dependent induction of cell proliferation, iv) Toll-like receptor mediated induction of pro-carcinogenic pathways [10–14]. Despite this vast body of circumstantial evidence, however, no clinical data have thus far been available to directly show distinct bacterial colonization patterns in CRC patients. In fact, the molecular nature of the complex intestinal community was largely unexplored prior to the moment that Eckburg and coworkers [15] revealed the presence of ~ 400 bacterial species by sequencing prokaryotic ribosomal RNA gene sequences from multiple colonic mucosal sites and feces of healthy subjects. Further investigations revealed high intra-individual variation of intestinal microbiomes in the human population, whereas the microbial colonization of the mucosa within adult individuals is relatively stable throughout the colon [16–18]. Based on the latter observations we hypothesized that the in-depth analysis of a relatively small number of paired on/off-tumour tissue samples from CRC patients could disclose bacterial species that might be implicated in CRC etiology. To achieve this goal, we used deep pyrosequencing of bacterial rRNA to compare CRC tumor microbiomes to that of adjacent non-malignant

mucosa across six patients. The data provided the first high-resolution image of the human CRC microbiome and showed that CRC is associated with quite dramatic shifts in the adherent intestinal microbiota.

Materials and Methods

Patient Material

Six patients (labeled A–F, Table 1) underwent resections for primary colon adenocarcinoma at the Radboud University Nijmegen Medical Centre. After resection, the colonic specimens were extensively rinsed with sterile water after which the specimens were examined by an oncological pathologist. Disease was staged according to the Tumor-Node-Metastasis (TNM) classification [19]. From each colonic specimen, biopsies were taken from the tumor site (“on-tumor”, A_{on} - F_{on}) and from adjacent non-malignant tissue (“off-tumor”, A_{off} - F_{off}) on the luminal side of the colonic wall (distance about 5–10 cm). Tissue specimens were disrupted by mechanical shearing after which total DNA was extracted using the AllPrep DNA/RNA kit (Qiagen). All samples were stored at -80°C until use.

Ethics Statement

Research was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the by the Medical Ethical Committee of the district Arnhem-Nijmegen (The Netherlands); patients provided written informed consent for the collection of samples and subsequent analysis when required.

Denaturing Gradient Gel Electrophoresis (DGGE) and Ribosomal Intergenic Spacer Analysis (RISA)

Using total DNA from the 12 colonic biopsies as a template, bacterial 16S rRNA genes were amplified by a nested approach [20] using the primers pairs 27f/1492r [21] and L1401r/968f-GC [22,23] in two subsequent PCR reactions (Table S1). DGGE was performed on the resulting PCR mixture as described previously [24]. It should be noted that visible bands (see Figure 1) represent bacterial species that have an abundance of at least 1–10% of the total community, whereas low abundant species will not result in a detectable bands by this approach. To confirm DGGE data, bacterial ribosomal intergenic spacer regions were amplified with primers 1406f and 23Sr using the same DNA as template (Table S1; [24]). RISA was performed as described previously [23].

FLX 454 titanium pyrosequencing

In the second step of the nested PCR approach, we amplified the V1–V3 region of the bacterial 16S rRNA gene using primer

pairs tagged with 12 distinct **M**etagenome **I**dentification (MID) tags (Table S1). 454 sequencing was performed at the University of Liverpool’s Advanced Genomics Facility. Sequences are available on request.

Read processing and community diversity

All partial 16S rRNA gene sequences were processed initially using the Pyro-pipeline at the Ribosomal database project (RDP, [25]; Release 10) to trim and remove primers from the partial ribotags and to limit sequences to >400 bp and ≤ 500 bp, sequences were processed using the pre.cluster command which minimizes errors introduced by the pyrosequencing platform [26]. This step provided the datasets for analysis (Table S2) with the read length histograms shown in Figure S2A. The data from all the samples was processed using MOTHR [27] to generate indices of diversity, rarefaction curves (Figure S2B) and to undertake the Libshuff analysis of sample similarity. MOTHR was run using the computational facilities of the Advanced Research Computing @ Cardiff (ARCCA) Division, Cardiff University. Comparisons of the libraries from an individual was performed using the RDP’s Library compare tool. Analysis of the ribotags was also performed using MEGAN [28] for which the input was the csv output from the RDP’s classifier pipeline (using default settings and a confidence level of 50%). The comparison tool was selected and reads normalized between samples and Bonferroni correction used to highlight differences between samples. An alignment independent analysis of the data was also undertaken using 5-mers and frequency landscape distribution (fLAND) analysis [29–31]. The generation of the 5-mers was performed using a bespoke PERL script (written by BED; available on request) and PCA analysis was undertaken in MATLAB on ARCCA, the fLAND analysis was performed using the software fLAND.

Consistency analysis

Biases in microbiota between the on-tumor and off-tumor samples across patients were summarized in order to identify taxa which were either consistently enriched or consistently depleted in the two niches. All pyrosequencing reads were first mapped to the SILVA comprehensive database of aligned, quality checked 16S/18S rRNA sequences >300 nt (version SSUParc_100; [32]) using BLAT v34 with default parameters and cutoffs [33]. We assumed that each read was derived from a different micro-organism and that the sampling of reads represented the taxonomic distribution within the intestinal microbiota. For each sample, every read was assigned to its most similar sequence in the SILVA database and a summary of the taxonomic annotations of the detected database sequences was generated. Each taxonomic clade was assessed to determine whether it showed a higher fraction of reads in off-tumor or on-tumor samples for every patient, and a consistency score was calculated by counting “+1” if the clade was higher off-tumor, “-1” if the clade was higher on-tumor, and “0” if the fraction of reads on- and off-tumor was identical (e.g. if the clade was not measured in this patient). Finally, these scores were summed, yielding an overall consistency score between -6 and $+6$ that reflects how consistently the clade was enriched or depleted across all patients. Note that each sequence in the SILVA database has two taxonomic annotations, i.e. EMBL and RDP.

Results

DGGE fingerprinting

As a first exploration, profiles of bacterial 16S rRNA genes were generated using DGGE for tumor and matching adjacent non-

Table 1. Patient characteristics.

Patient	Gender	Age	Colon Tumor	
			Stage ¹	Localisation
A	m	67	T ₂ N ₀ M ₀	sigmoid
B	m	61	T ₂ N ₀ M ₀	rectum/sigmoid
C	m	49	T ₃ N ₁ M ₀	sigmoid
D	m	71	T ₂ N ₀ M ₀	rectum
E	m	67	T ₄ N ₀ M ₀	cecum
F	f	66	T ₂ N ₀ M ₀	rectum

¹T, tumor growth into the wall of the intestine; N, spread to nearby lymph nodes; M, metastases in other organs; numbers 0–4 indicate increasing severity.

doi:10.1371/journal.pone.0020447.t001

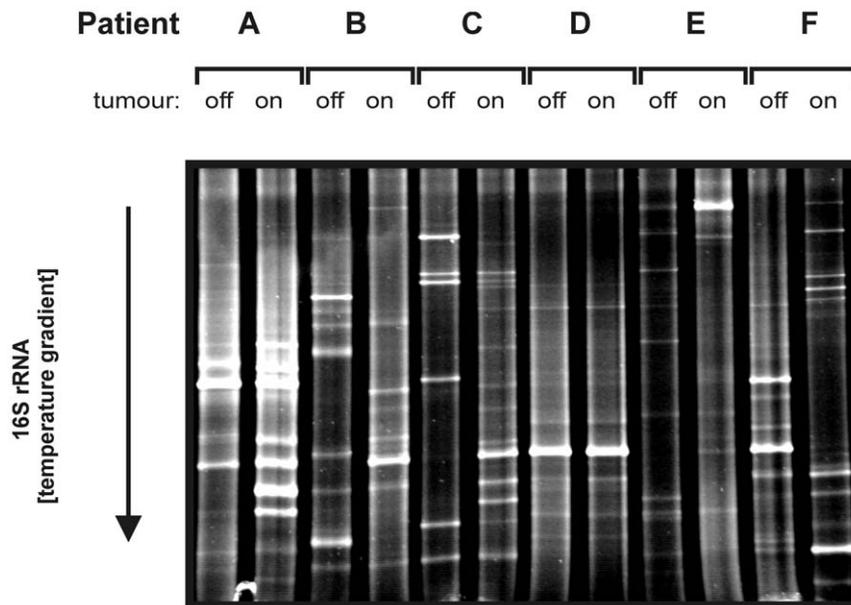


Figure 1. DGGE Fingerprinting of CRC Tissue and Non-malignant Adjacent Mucosa. An internal fragment (~450 bps) of the bacterial 16S rRNA gene was amplified from colon tissue-extracted DNA by a broad-range PCR approach after which these amplicon mixtures were applied to DGGE. Patient characteristics can be found in Table 1; *off*, non-malignant tissue; *on*, tumor tissue. doi:10.1371/journal.pone.0020447.g001

malignant (off-tumor) mucosa from 6 CRC patients (Table 1). As shown in Figure 1, the microbial communities of tumor tissue and adjacent “off-tumor” mucosa were strikingly different and similar results were obtained when RISA fingerprinting was applied to the same samples (Figure S1). Notably, this result sharply contrasts with the previous observations that the colonic mucosal microbiota is almost identical at adjacent sites in healthy subjects [15,16,34–36]. Inspired by this striking observation, we next aimed for a microbiome sequencing approach to map the microbiome changes at a high resolution rather than sequencing of individual distinguishing bands that only represent the abundant species in a certain sample.

FLX 454 titanium pyrosequencing

To define the colon tumor microbiome at a deep level, we amplified and sequenced the V1–V3 region of the bacterial 16S rRNA genes (Table S1), which resulted in a total of 193,880 ribotags of length 401–500 bp. The data showed high coverage values (>88%) and rarefaction curves indicated satisfactory sampling of the communities at 90% identity (Figure S2B; Table S2). Libshuff analysis indicated that all on- and off-tumor communities were significantly different ($p < 0.0001$) from each other (Table S3). Importantly, both alignment-dependent and independent methods supported the observation of these altered tumor microbiomes (Figure 2, Figure S2C and D; Table S3). Moreover, while DGGE and RISA only showed minor differences for patient D, subtle, but significant, differences could be identified by the deep sequencing approach. This clearly exemplifies the superior resolution that can be obtained with the latter technology. The data showed a general tendency of more Bacteroidetes and less Firmicutes from in tumor tissue compared to matching off-tumor mucosa (Figure S3). However, as could be expected the observed microbiome shifts showed a high level of variability among patients (Figure S3A–F) and in certain cases were even contrary to the general tendency (Figure S3G). Although *S. bovis* or *C. septicum* infections have a known clinical association with CRC, only very few sequences mapped to rRNA of these two species and no dependable colonization of CRC tissue was observed. This could be explained by the fact that such

opportunistic pathogens are predominantly present in the transient adenoma stage of CRC [37].

Consistency analysis

To pinpoint the most imperative microbiome changes during formation of CRC, consistency across the patients was calculated for each taxon individually by giving it a score of “–1” if the normalized number of sequence reads of that taxon in the tumor tissue was higher than in matching off-tumor mucosa, “+1” if the taxon was more abundant in healthy mucosa and “0” if it was not detected at all in that patient. Summing these scores across patients resulted in a consistency score from –6 to +6 for each taxon (Tables 2 and S4, Figure S4A and B). It should be noted that some sequences are better annotated by EMBL than by RDP, and *vice versa* (see Figure S4C), so we report both rather than preferentially trusting either of these taxonomic annotations. This approach showed that CRC tissue was consistently associated with overrepresentation of the subclass of *Coriobacteridae*, especially the genera *Slackia* and *Collinsella*, which can be regarded as gut commensals. On the other hand, members of the *Enterobacteriaceae*, such as *Citrobacter*, *Shigella*, *Cronobacter*, *Kluyvera*, *Serratia* and *Salmonella* spp. (scores between +4 and +6; Table 2; Figure S4A and B) were underrepresented in CRC tissue. Although these findings were consistent, the relative abundance of these taxa differed considerably between on and off tumor mucosa from the investigated patients as depicted in Figure 3.

Discussion

The most striking observation from our current study was the dramatically different microbiomes in CRC tissue and adjacent non-malignant mucosa in 5 of the 6 investigated patients. To our surprise, however, we found no consistent overrepresentation of potential pathogenic bacteria in CRC tissue. In contrast, overrepresented species concerned members of the genera *Coriobacteridae*, *Roseburia*, *Fusobacterium* and *Faecalibacterium*, which are generally regarded as gut commensals with pro-biotic features.

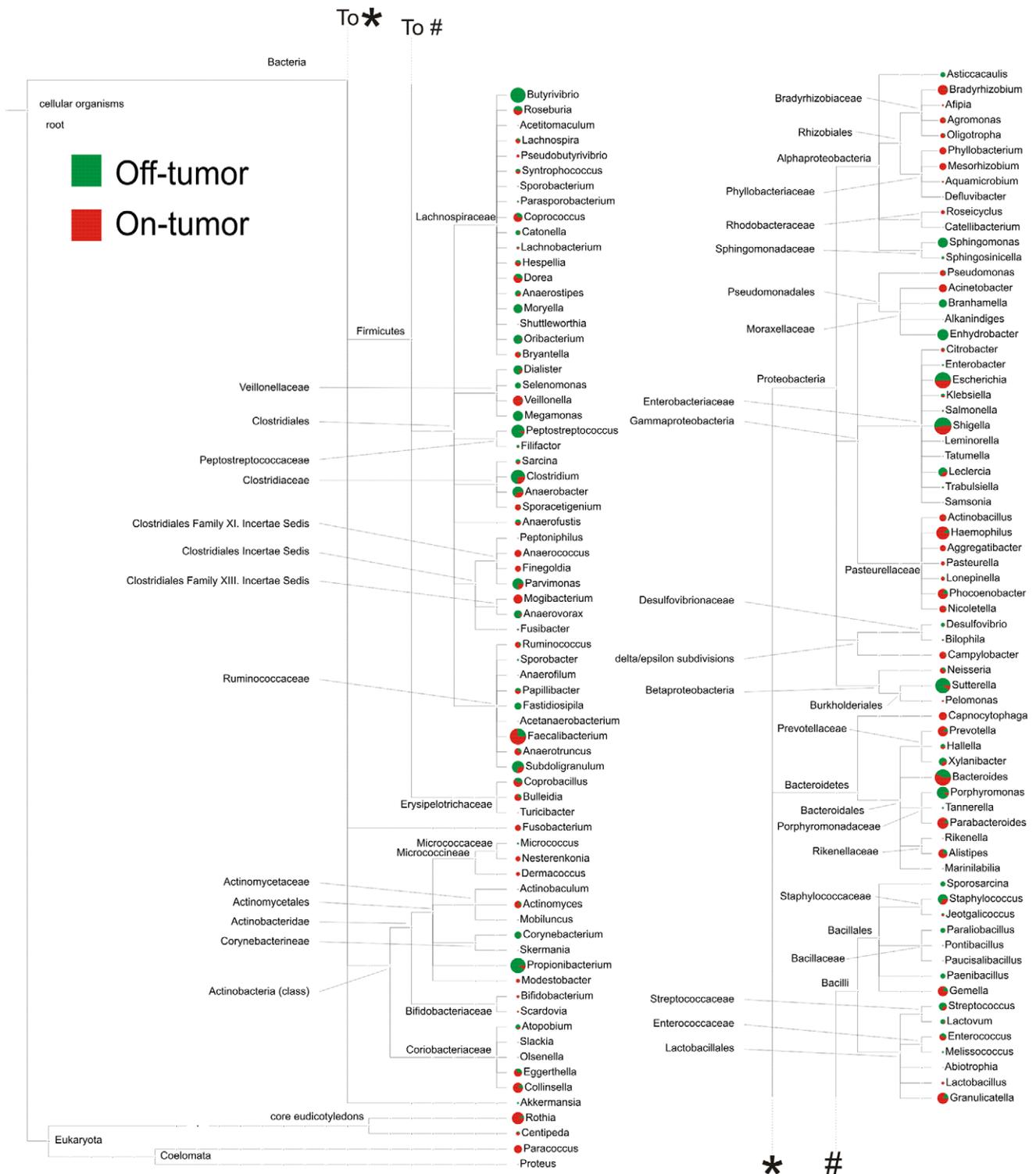


Figure 2. Phylogenetic Analysis of Altered Microbiomes. The 454 sequencing data were normalized in MEGAN (Huson et al. 2009) and parsed through the RDP pyro-pipeline classifier tool (Cole et al. 2009) to generate a csv file of taxonomic abundance. This file was used as input for MEGAN to visualize in which families differences between non-malignant tissue (*off-tumor*) and CRC tissue (*on-tumor*) communities are present. A high-resolution image of this Figure for “zoom-in” purposes can be downloaded from Figure S2E. doi:10.1371/journal.pone.0020447.g002

This suggests that the observed microbial shifts are caused by the quite dramatic physiological and metabolic alterations that result from colon carcinogenesis itself [38–40], and that these species

may be regarded as CRC passengers. In fact, recent metabolomics studies revealed extremely altered nutritional conditions in the CRC tumor microenvironment compared to non-malignant

Table 2. Consistent CRC-associated Microbiome Shifts¹.

Terminal Clade	consistency score		Bacterial order
	EMBL	RDP	
<i>Overrepresented in CRC tissue</i>			
<i>Slackia</i>	.	−6	<i>Coriobacteriales</i>
<i>Collinsella</i>	.	−5	<i>Coriobacteriales</i>
<i>Eubacterium</i> ;environmental samples	−5	.	<i>Clostridiales</i>
<i>Coriobacterium</i> ;environmental samples	−4	.	<i>Coriobacteriales</i>
<i>Roseburia</i>	−4	.	<i>Clostridiales</i>
<i>Clostridiaceae</i> ;environmental samples	−4	.	<i>Coriobacteriales</i>
<i>Fusobacterium</i>	−3	−3	<i>Fusobacteriales</i>
unclassified_ <i>Coriobacteriaceae</i>	.	−3	<i>Coriobacteriales</i>
unclassified_ <i>Peptostreptococcaceae</i>	.	−3	<i>Clostridiales</i>
<i>Erysipelotrichaceae</i> Incertae Sedis	.	−3	<i>Erysipelotrichales</i>
<i>Fusobacteria</i> ;environmental samples	−3	.	<i>Fusobacteriales</i>
<i>Faecalibacterium</i> ;environmental samples	−3	.	<i>Clostridiales</i>
<i>Bacteroidales</i> ;environmental samples	−3	.	<i>Bacteroidales</i>
unclassified_Firmicutes	-	−3	N/A
<i>Underrepresented in CRC tissue</i>			
<i>Citrobacter</i>	6	4	<i>Enterobacteriales</i>
<i>Shigella</i>	4	4	<i>Enterobacteriales</i>
<i>Cronobacter</i>	4	.	<i>Enterobacteriales</i>
<i>Kluyvera</i>	4	.	<i>Enterobacteriales</i>
<i>Serratia</i>	4	.	<i>Enterobacteriales</i>
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Saintpaul	4	.	<i>Enterobacteriales</i>
<i>Eubacteriaceae</i> ; environmental samples	4	.	<i>Clostridiales</i>
<i>Anaerovorax</i>	.	4	<i>Clostridiales</i>
unclassified_Ruminococcaceae	.	4	<i>Clostridiales</i>
unclassified_Bacteria	.	4	N/A
<i>Microbacterium</i>	3	.	<i>Actinomycetales</i>
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi B	3	.	<i>Enterobacteriales</i>
<i>Clostridium</i>	.	3	<i>Clostridiales</i>
<i>Peptoniphilus</i>	.	3	<i>Clostridiales</i>

¹only terminal clades with consistency scores ≤ -3 or ≥ 3 that were found using the EMBL or RDP 16S ribosomal databases as template are shown. Full consistency data are provided as Table S4 and Figure S4. Both annotations were included because well-annotated sequences by RDP are often not well annotated by EMBL and vice versa (Figure S4C).

doi:10.1371/journal.pone.0020447.t002

mucosa [41]. The most prominent and consistent findings concerned a drastic decrease in glucose and pyruvate and an increase in lactate (low pH), amino acids, lipids and fatty acids. The intra-patient variability in microbiome alterations could possibly reflect the intra-individual variability in the CRC tumor microenvironment [42], resulting in the preferential recruitment of different classes of intestinal species. Notably, reduced numbers of *Collinsella* spp. and *Roseburia* spp. have previously been found in elderly subjects using non-steroidal anti-inflammatory drugs compared to non-users (NSAID [43]; suggesting that these bacteria need inflammatory niches to optimally colonize the bowel wall. Interestingly, the genera *Roseburia*, *Fusobacterium* and *Faecalibacterium*, which are moderately enriched in tumors belong to the major butyrate producing intestinal bacteria. Butyrate is thought to be protective against CRC by inducing a p21-dependent cell cycle arrest resulting in an increased apoptosis rate of carcinogenic cells [44]. The effects of butyrate are however

still under debate as tumor inhibition may for instance be restricted to the early phases of carcinogenesis. Markedly, it has been shown that *Faecalibacterium prausnitzii* secretes anti-inflammatory factors that block NF- κ B activation and IL-8 production in an experimental animal model for Crohn's disease [45]. Moreover, patients with inflammatory bowel disease, who are at increased risk for CRC, have been associated with lower numbers of *F. prausnitzii* in the intestinal population [46]. Finally, *Slackia* spp. are known to convert dietary isoflavones into more potent antioxidants [47] capable of inducing apoptotic pathways in tumor cells [48]. Thus in view of our current data, we could draw the conclusion that the CRC microenvironment is preferably colonized by intestinal bacteria with anti-tumorigenic and anti-carcinogenic properties, which thereby may prevent rapid progression of this disease. However, one could also argue that for instance butyrate provides an additional energy source for tumor cells, while dampening the inflammatory response stops the

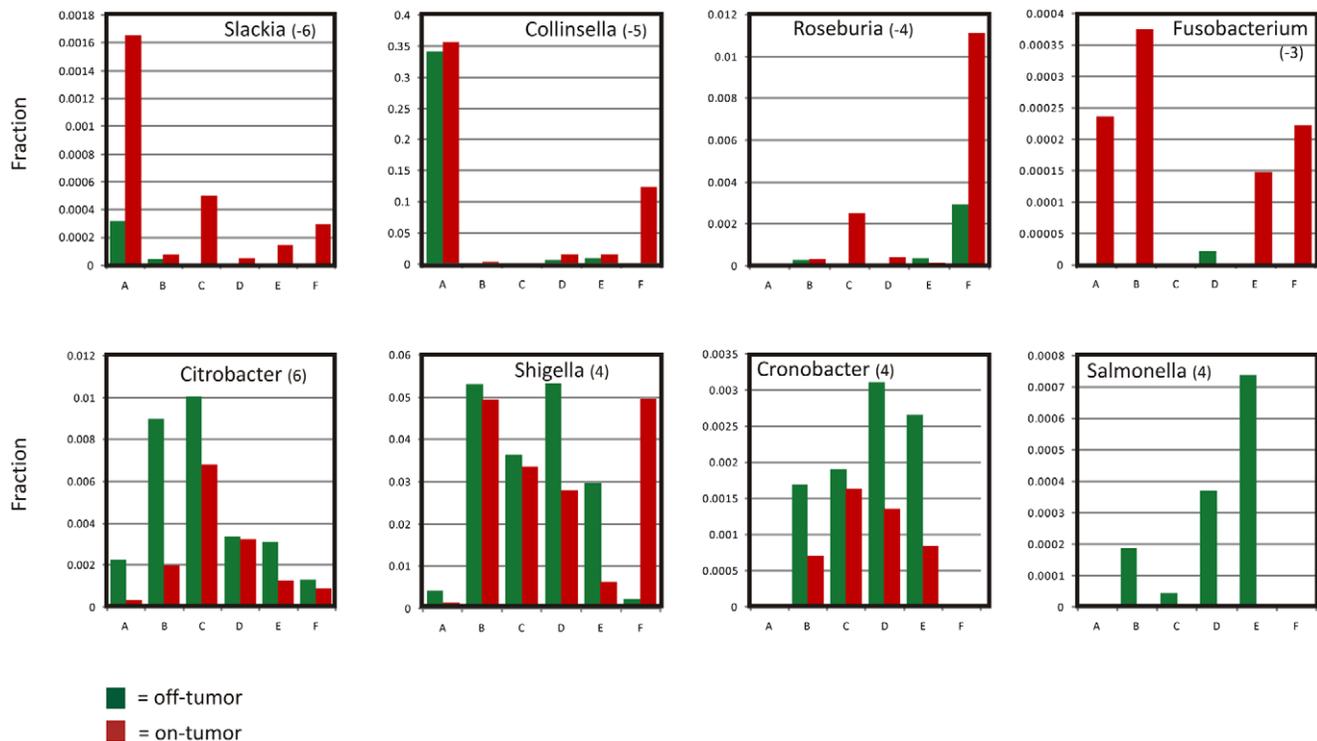


Figure 3. Consistent Biases in Microbiota CRC Tissue and Non-malignant Adjacent Mucosa. Relative distribution of selected CRC over/under represented taxa was calculated as the fraction annotated sequences of the total number of reads in that specific sample. Consistency score for the indicated taxon (see Table 2) is given between brackets and reflects how consistently clades were enriched across patients A–F; green bars indicate the fraction in *off-tumor* and red bars indicate fraction of this taxon *on-tumor*. doi:10.1371/journal.pone.0020447.g003

innate immune system from attacking the nascent tumor. Thus, both tumor suppressing or tumor promoting scenarios may be possible outcomes of the differential colonization of CRC tissue and further, detailed investigations, will be required to elucidate this issue.

Another remarkable observation concerned the decreased presence of members of the *Enterobacteriaceae*, such as *Citrobacter*, *Shigella*, *Cronobacter*, *Kluyvera*, *Serratia* and *Salmonella* spp. in tumour tissue of the investigated CRC patients. This data may suggest that these bacteria are part of the intrinsic microbiome of CRC patients, but outcompeted by the above mentioned commensal-like bacteria upon disease progression. Although we realize that in the absence of a large reference database of mucosal microbiomes from healthy individuals it is difficult to draw conclusions on this observation, we would like to take the opportunity to shortly review why Enterobacterial intestinal colonization could be associated with an increased risk for CRC. First, metagenomic inventories of the human intestinal microbiome showed that *Salmonella*, *Citrobacter*, and *Cronobacter* were among the low abundant intestinal species or were even completely absent in healthy individuals [15,49], which is fully in-line with their pathogenic character [50]. Contrarily, this bacterial family was detectably present in non-malignant colonic mucosa samples from CRC patients [36], while Shen and colleagues recently showed that *Shigella* spp displayed an increased abundance in the intrinsic (non-malignant) microbiome of adenoma patients [51]. Importantly, the potential of Enterobacteria to initiate CRC has already been shown for *Citrobacter* species in an animal model [52], and it is thought that this increased susceptibility for CRC is caused by an asymptomatic, but chronic, inflammatory response in the colonic mucosa [53]. Additionally, several Enterobacterial strains produce

DNA damaging genotoxins [54] and may thereby actively contribute to the accumulation of mutations that characterize the adenoma-carcinoma sequence [55]. In this context, our data may further suggest that upon CRC progression, the tumor microenvironment changes in such a way that Enterobacteria are replaced by commensal-like species or bacteria with proposed probiotic properties that have increased access to and/or can more efficiently forage in the altered tumor microenvironment. The disappearance of CRC-driving pathogenic bacteria from advanced CRC tumor tissue may be analogous to what has been reported for *Helicobacter pylori* during gastric cancer progression [56].

Altogether, our study provides an important first glimpse of the CRC-associated microbiome and indicates a highly dynamic relationship between intestinal bacteria and developing tumours. Nonetheless, many open questions need to be addressed in future studies, including deep microbiome analysis of extended groups of tumor samples from different disease stages, including adenomas and biopsies from a large set of non-cancer patients to serve as reference database. Furthermore, for diagnostic purposes it will be important to investigate how the local tumour-associated microbiome shifts relate to the fecal microbiota composition [57]. A more detailed analysis of CRC (meta-)genomes and bacterial transcriptomes is needed to pinpoint the genes that cause differential colonization in the tumor microenvironment and to better define high-risk microbial populations. Subsequently, high-risk and low-risk bacterial populations should be validated in (animal) models for sporadic CRC, and mechanisms of bacterial interference in CRC have to be unraveled in more detail. All-in-all, this sets the agenda for a new exciting era in colorectal cancer research, integrating microbiology and microbial ecology with tumor biology. This will lead us towards an increased under-

standing of the driving forces of CRC, as well as novel microbiome-related diagnostic tools and therapeutic interventions.

Supporting Information

Figure S1 RISA fingerprinting of CRC tissue and non-malignant adjacent mucosa. The intergenic spacer region between the 16S and 23S rRNA genes was amplified with conserved primer pairs (Table S1) and analyzed using an Agilent Bioanalyser. Patient characteristics can be found in Table 1; off, non-malignant tissue; on, tumor tissue.

(PDF)

Figure S2 **A**, Read lengths per sample, with X_{off} and X_{on} coming from off-tumor and on-tumor samples in subject X, the data was derived from the read lengths post-processing via the RDP pyropipeline. **B**, Rarefaction curves for each sample, the sample key is the same as use for Figure S1. These curves were generated using MOTHRUR, cut off values are shown. **C**, Venn diagram for paired samples generated using MOTHRUR, each diagram shows the OTUs (at 0.03% cut-off) shared and those unique to each sample from on and off tumor. **D**, Principle component analysis of the annotated 16S rRNA sequence data generated for on and off tumor samples of each patient (A–F), was plotted as an 0.5×0.5 interval density distribution. The color coding shows the natural logarithm of the densities in each segment. Where there are significant differences ($P < 0.05$) between tumor tissue and adjacent off-tumor mucosa, a white cross is shown in that segment. The taxonomic groups contributing to the most densely populated segments are shown (and the numbers of sequences contributing are shown in parentheses).

(PDF)

Figure S3 Taxonomic affiliation of the 16S rRNA gene reads for each paired sample set from subjects **A–F** and the combined samples (**G**). The figures were generated using MEGAN and show in which PHYLX (piecharts) and families (barcharts) the main alterations of levels of reads are found. In the barchart only families which were greater than 1% of the community are shown.

(PDF)

References

- Hooper LV, Gordon JI (2001) Commensal host-bacterial relationships in the gut. *Science* 292: 1115–1118.
- Parkin DM (2006) The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 118: 3030–3044.
- Mager DL (2006) Bacteria and Cancer: Cause, Coincidence or Cure? *J Transl Med* 4: 14.
- zur Hausen H (2006) *Streptococcus bovis*: causal or incidental involvement in cancer of the colon? *Int J Cancer* 119: xi–xii.
- Rowland IR (2009) The role of the gastrointestinal microbiota in colorectal cancer. *Cur Pharm Des* 15: 1524–1527.
- Waisberg J, Mathews CdeO, Pimenta J (2002) Infectious endocarditis from *Streptococcus bovis* associated with colonic carcinoma: case report and literature review. *Arq Gastroenterol* 39: 177–180.
- Boleij A, Schaeps RM, Tjalsma H (2009) Association between *Streptococcus bovis* and colon cancer. *J Clin Microbiol* 47: 516.
- Seder CW, Kramer M, Long G, Uzieblo MR, Shanley CJ, et al. (2009) *Clostridium septicum* aortitis: Report of two cases and review of the literature. *J Vasc Surg* 49: 1304–1309.
- Tjalsma H, Schöller-Guinard M, Lasonder E, Ruers TJ, Willems HL, et al. (2006) Profiling the humoral immune response in colon cancer patients: diagnostic antigens from *Streptococcus bovis*. *Int J Cancer* 119: 2127–2135.
- Toprak NU, Yagci A, Gulluoglu BM, Akin ML, Demirkalem P, et al. (2006) A possible role of *Bacteroides fragilis* enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect* 12: 782–786.
- Wang X, Allen TD, May RJ, Lightfoot S, Houchen CW, et al. (2008) *Enterococcus faecalis* induces aneuploidy and tetraploidy in colonic epithelial cells through a bystander effect. *Cancer Res* 68: 9909–9917.
- Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, et al. (2009) A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med* 15: 1016–1022.
- Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, et al. (2010) *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A* 107: 11537–11542.
- Lee SH, Hu LL, Gonzalez-Navajas J, Seo GS, Shen C, et al. (2010) ERK activation drives intestinal tumorigenesis in *Apc*(min/+) mice. *Nat Med* 16: 665–670.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, et al. (2005) Diversity of the Human Intestinal Microbial Flora. *Science* 308: 1635–1638.
- Green GL, Brostoff J, Hudspeth B, Michael M, Mylonaki M, et al. (2006) Molecular characterization of the bacteria adherent to human colorectal mucosa. *J Appl Microbiol* 100: 460–469.
- Dethlefsen L, Eckburg PB, Bik EM, Relman DA (2006) Assembly of the human intestinal microbiota. *Trends Ecol Evol* 21: 517–523.
- Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, et al. (2009) Bacterial community variation in human body habitats across space and time. *Science* 326: 1694–1697.
- Scott-Conner CE, Christie DW (1995) Cancer staging using the American Joint Committee on Cancer TNM System. *J Am Coll Surg* 181: 182–188.
- Boon N, De Windt W, Verstraete W, Top EM (2002) Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. *FEMS Microbiol Ecol* 39: 101–112.
- Lane DJ, Stackebrandt E, Goodfellow M (1991) 16S/23S rRNA Sequencing. *Nucleic Acids Techniques in Bacterial Systematics*. Chichester: John Wiley & Sons. pp 115–175.

Figure S4 Consistent biases in microbiota between the on- and off-tumor samples. The overall consistency scores between +6 (green) and –6 (red) reflects how consistently clades were enriched across six patients. Trees were visualized with iTOL **A**, Consistent clades derived from the EMBL annotation of SILVA sequences. **B**, Consistent clades derived from the RDP annotation of SILVA sequences. **C**, Differential annotation depth of the SSU rRNA sequences in the SILVA database by EMBL and RDP. Low annotation depth means little resolution: many sequences are either well annotated by RDP (bottom right) or by EMBL (top left).

(PDF)

Table S1 Primers used in this study.

(PDF)

Table S2 MOTHRUR diversity indices of bacterial communities in samples on- and off-tumor (X_{on} and X_{off} , respectively).

(PDF)

Table S3 Genus level comparison generated with the RDP library compare tool. Values indicate the number of 16S rRNA pyrosequencing reads that map to the listed genus. Only significant differences are shown ($P < 0.05$). A–F: patients; X_{off} : off-tumor tissue; X_{on} : on-tumor tissue.

(PDF)

Table S4 Consistency scores for CRC associated microbiome shifts.

(XLS)

Acknowledgments

We wish to thank the patients for their samples and Martijn A. Huynen, Dorine W. Swinkels, Alistair Darby (454 protocol), Renée M.J. Schaeps, Hennie Roelofs and Tanja Schölin for stimulating discussions, technical support and/or sample selection.

Author Contributions

Conceived and designed the experiments: JRM HT AB. Performed the experiments: JRM RR. Analyzed the data: JRM BED HT. Contributed reagents/materials/analysis tools: NH WHMP. Wrote the paper: JRM BED AB HT.

22. Nubel U, Engelen B, Felske A, Snaird J, Wieshuber A, et al. (1996) Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *J Bact* 178: 5636–5643.
23. Scanlan PD, Shanahan F, Clune Y, Collins JK, O'Sullivan GC, et al. (2008) Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. *Environ Microbiol* 10: 1382.
24. Borneman J, Triplett EW (1997) Molecular microbial diversity in soils from eastern Amazonia: Evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* 63: 2647–2653.
25. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl Acids Res* 37: D141–145.
26. Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* 12: 1889–1898.
27. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur: Open Source, Platform-independent, Community-supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* 75: 7537–7541.
28. Huson DH, Richter DC, Mitra S, Auch AF, Schuster SC (2009) Methods for comparative metagenomics. *BMC Bioinformatics* 10 Suppl 1: S12.
29. Rudi K, Zimonja M, Kvenshagen B, Rugtveit J, Midtvedt T, et al. (2007) Alignment-Independent Comparisons of Human Gastrointestinal Tract Microbial Communities in a Multidimensional 16S rRNA Gene Evolutionary Space. *Appl Environ Microbiol* 73: 2727–2734.
30. Rudi K, Zimonja M, Trosvik P, Naes T (2007) Use of multivariate statistics for 16S rRNA gene analysis of microbial communities. *Int J Food Microbiol* 120: 95–99.
31. Rudi K, Zimonja M, Naes T (2006) Alignment-independent bilinear multivariate modelling (AIBIMM) for global analyses of 16S rRNA gene phylogeny. *Int J Syst Evol Microbiol* 56: 1565–1575.
32. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188–7196.
33. Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12: 656–664.
34. Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben Amor K, Akkermans AD, et al. (2002) Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* 68: 3401–3407.
35. Palmer C, Bik EM, Eisen MB, Eckburg PB, Sana TR, et al. (2006) Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Research* 34: e5.
36. Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, et al. (2007) Mucosa-Associated Bacterial Diversity in Relation to Human Terminal Ileum and Colonic Biopsy Samples. *Appl Environ Microbiol* 73: 7435–7442.
37. Boleij A, Roelofs R, Schaeps RM, Schülin T, Glaser P, Swinkels DW, et al. (2010) Increased exposure to bacterial antigen Rpl7/L12 in early stage colorectal cancer patients. *Cancer* 116: 4014–22.
38. Sansonetti PJ (2004) War and peace at mucosal surfaces. *Nat Rev Immunol* 4: 953–964.
39. Corfield AP, Myerscough N, Longman R, Sylvester P, Arul S, et al. (2000) Mucins and mucosal protection in the gastrointestinal tract: new prospects for mucins in the pathology of gastrointestinal disease. *Gut* 47: 589–594.
40. Dhawan P, Singh AB, Deane NG, No Y, Shiou SR, et al. (2005) Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer. *J Clin Invest* 115: 1765–1776.
41. Hirayama A, Kami K, Sugimoto M, Sugawara M, Toki N, et al. (2009) Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Res* 69: 4918–4925.
42. Righi V, Durante C, Cocchi M, Calabrese C, Di Febo G, et al. (2009) Discrimination of healthy and neoplastic human colon tissues by ex vivo HR-MAS NMR spectroscopy and chemometric analyses. *J Proteome Res* 8: 1859–1869.
43. Makivuokko H, Tiihonen K, Tynkkynen S, Paulin L, Rautonen N (2009) The effect of age and non-steroidal anti-inflammatory drugs on human intestinal microbiota composition. *Br J Nutr* 103: 227–234.
44. Bordonaro M, Lazarova DL, Sartorelli AC (2008) Butyrate and Wnt signaling: a possible solution to the puzzle of dietary fiber and colon cancer risk? *Cell Cycle* 7: 1178–1183.
45. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, et al. (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 105: 16731–16736.
46. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, et al. (2009) Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm Bowel Dis* 15: 1183–1189.
47. Jin JS, Kitahara M, Sakamoto M, Hattori M, Benno Y (2009) *Slackia equalifaciens* sp. nov., a human intestinal bacterium capable of producing equol. *Int J Syst Evol Microbiol* 4: 4.
48. Choi EJ, Ahn WS, Bae SM (2009) Equol induces apoptosis through cytochrome c-mediated caspases cascade in human breast cancer MDA-MB-453 cells. *Chem Biol Interact* 177: 7–11.
49. Qin JJ, Li RQ, Raes J, Arumugam M, Burgdorf KS, et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464: 59–U70.
50. DuPont HL (2009) Clinical practice. Bacterial diarrhea. *N Engl J Med* 361: 1560–1569.
51. Shen X, Rawls J, Randall T, Burcal L, Mpande C, et al. (2010) Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes* 1: 138–147.
52. Newman JV, Kosaka T, Sheppard BJ, Fox JG, Schauer DB (2001) Bacterial infection promotes colon tumorigenesis in *Apc(Min/+)* mice. *J Infect Dis* 184: 227–230.
53. Maggio-Price L, Treuting P, Bielefeldt-Ohmann H, Seamons A, Drivdahl R, et al. (2009) Bacterial infection of *Smad3/Rag2* double-null mice with transforming growth factor-beta dysregulation as a model for studying inflammation-associated colon cancer. *Am J Pathol* 174: 317–329.
54. Nougayre JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, et al. (2006) *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 313: 848–851.
55. Vogelstein B, Kinzler KW (1993) The multistep nature of cancer. *Trends Genet* 9: 138–141.
56. Kang HY, Kim N, Park YS, Hwang JH, Kim JW, et al. (2006) Progression of atrophic gastritis and intestinal metaplasia drives *Helicobacter pylori* out of the gastric mucosa. *Dig Dis Sci* 51: 2310–2315.
57. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, et al. (2011) Microbial Dysbiosis in Colorectal Cancer (CRC) Patients. *PLoS One* 6(1): e16393.