Microbiome Analysis of Stool Samples from African Americans with Colon Polyps

Hassan Brim[1]*, Shibu Yooseph[2], Erwin G. Zoetendal[4], Edward Lee[1], Manolito Torralbo[3], Adeyinka O. Laiyemo[1], Babak Shokrani[1], Karen Nelson[3], Hassan Ashktorab[1]

1 Department of Pathology, Department of Medicine and Cancer Center, Howard University, College of Medicine, Washington, District of Columbia, United States of America, 2 JCVI, San Diego, California, United States of America, 3 JCVI, Rockville, Maryland, United States of America, 4 Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

Abstract

Background: Colonic polyps are common tumors occurring in ~50% of Western populations with ~10% risk of malignant progression. Dietary agents have been considered the primary environmental exposure to promote colorectal cancer (CRC) development. However, the colonic mucosa is permanently in contact with the microbiota and its metabolic products including toxins that also have the potential to trigger oncogenic transformation.

Aim: To analyze fecal DNA for microbiota composition and functional potential in African Americans with pre-neoplastic lesions.

Materials & Methods: We analyzed the bacterial composition of stool samples from 6 healthy individuals and 6 patients with colon polyps using 16S ribosomal RNA-based phylogenetic microarray; the Human intestinal Tract Chip (HITChip) and 16S rRNA gene barcoded 454 pyrosequencing. The functional potential was determined by sequence-based metagenomics using 454 pyrosequencing.

Results: Fecal microbiota profiling of samples from the healthy and polyp patients using both a phylogenetic microarraying (HITChip) and barcoded 454 pyrosequencing generated similar results. A distinction between both sets of samples was only obtained when the analysis was performed at the sub-genus level. Most of the species leading to the dissociation were from the *Bacteroides* group. The metagenomic analysis did not reveal major differences in bacterial gene prevalence/abundances between the two groups even when the analysis and comparisons were restricted to available *Bacteroides* genomes.

Conclusion: This study reveals that at the pre-neoplastic stages, there is a trend showing microbiota changes between healthy and colon polyp patients at the sub-genus level. These differences were not reflected at the genome/functions levels. Bacteria and associated functions within the *Bacteroides* group need to be further analyzed and dissected to pinpoint potential actors in the early colon oncogenic transformation in a large sample size.

Introduction

Colorectal cancer (CRC) is the third most prevalent cancer [1-3]. Its world distribution is heterogeneous with a predominance in the Western world. This heterogeneous distribution is also taking place within the Western societies [4,5]. African Americans (AAs) have a high incidence of, and mortality from this disease [6,7]. Several factors have been proposed and investigated, including genetics, epigenetics, diet, socioeconomic status and access to health [8-15]. However, in comparison with their African counterparts, with whom they share the same genetic background and who have

a very low burden of the disease [16], the higher incidence in AAs seems to be caused primarily by environmental factors.

Emerging data suggest an essential, mutualistic relationship between the host and their colonic microbiota [17-19]. Elegant experiments demonstrated, for example, that a single commensal, *Bacteroides thetaiotamicron* induces colonic mucosal gene expression, angiogenesis and immune responses revealing a broader extent of microbe-mucosal communication and cross-regulation than previously recognized [20]. Similar findings were also obtained with an enterotoxigenic *Bacteroides fragilis* [21,22].

Among body sites normally hosting a community of microbes, the human colon harbors the greatest number and diversity of organisms, primarily bacteria. Comprised of 500 to 1,000 bacterial species with two to four million genes, the gut microbiome contains about 100-fold more genes than the human genome and the estimated 10^{14} bacterial cells in the gut exceed by 10-fold the total ensemble of human cells [23]. Molecular analysis of the colonic luminal and mucosal microbiota indicates that individuals harbor unique microbiotas that are fairly stable along the colonic axis. However, the mucosal microbiota is either distinct or contains only a subset of the bacterial phylotypes identified in the luminal fecal samples [24,25]. Although mechanisms accounting for the composition and assemblage of the gut microbiota are incompletely understood, it appears that diet, host genetics, disease state (e.g. obesity, Inflammatory Bowel Disease (IBD)) as well as likely additional environmental factors influence the composition and the function of the colon microbiota [26,27].

The environmental exposures proposed to promote the development of human CRC have been primarily dietary agents. However, the local environment to which the colonic mucosa is exposed is created by the microbiota of the colon and their metabolic products that include beneficial components such as short chain fatty acids as well as harmful ones including toxins. Although it has been hypothesized for decades that the colonic microbiota influence CRC pathogenesis, neither specific bacteria nor mechanisms have been delineated [28-30]. Linkage of specific bacteria, their toxins and/or toxic metabolites (including mutagens) to CRC pathogenesis has been hampered by limited knowledge of the colonic microbiota and changing bacterial classification schemes over the last 40 years. While two divisions of bacteria (*Bacteroidetes* and *Firmicutes*) are considered dominant in the cultured colonic microbiota, up to 80% of the colonic microbiota has not yet been cultured [24,25]. Actinobacteria were also reported as prevalent in the intestinal tract but their presence has been underestimated in PCR based approaches [31].

Recent advances in sequencing have set the ground for sequence-based metagenomic studies that target the genomic diversity within an ecosystem Indeed, several studies have set the framework for metagenomic studies in general and for the gut microbiota in particular [24,25,32-36]. Huge databases for 16S rRNA genes as well as for gut microbiota functions have been established as a resource for other studies in the field.

We here performed a microbiome analysis of stool samples from healthy and colon polyp patients to elucidate bacterial changes that might induce or accompany early oncogenic transformation of the colon in African American patients with the goal of defining such markers for non-invasive screening protocols.

Materials and Methods

Ethics statement

The present study was approved by the Howard University Institutional Review Board. Written consent forms were obtained from all participants.

Samples collection and preparation

Fecal samples were obtained from 6 healthy (AFR001-006) and 6 colon polyp (AFR007-012) African American individuals who were informed about the study's goals and consented to participate according to Howard University IRB approved protocol. Patients with family history of colon cancer or inflammatory bowel diseases were excluded from this study. We included in the study adult patients who underwent screening colonoscopy and for whom a pathological diagnosis was established.

Each patient was given a kit with instructions for sample collection in sterile containers. The stool samples were collected at least two months after colonoscopy after a full microbiota restoration [37]. The patients were provided FedEx enveloppes to send samples immediately after bowel movements. The samples were delivered to us on the same day of collection to minimize changes in the microbiota composition. The samples were aliquoted and stored at -80°C. Colonoscopy and pathology reports were reviewed and used to select healthy and colon polyp patients for stool samples collection. The healthy and polyp patients were matched for demographic parameters to reduce the effect of cofounders on microbiota's differences. All polyp patients have colonic lesions of hyperplastic histology. The Healthy group had a mean age (SD) 59 years (9.4) and a BMI (SD) of 30.8 (4.41) while the polyp patients mean age (SD) was 59 years (8.4) and their BMI (SD) was 31.6 (3.56). There were 3 males and 3 females in both groups that were age matched across the two groups of patients.

DNA extraction

DNA from the stool samples was extracted using the QIAamp Stool DNA extraction Kit (Qiagen, Inc). The extracted DNA was used for 16S rRNA gene based barcoded 454 pyrosequencing and phylogenetic microarraying for microbiota profiling using the Human Intestinal Tract Chip (HITChip) [38] as well as for the metagenomic analysis.

HITChip analysis

DNA was used for phylogenetic profiling using the Human Intestinal Tract Chip (HITChip), a phylogenetic microarray which contains a duplicate set of 3,631 probes based on 16S rRNA gene sequences covering more than 1,100 intestinal bacterial phylotypes [38]. Briefly, 20 ng of DNA from each sample (n=12) was used to amplify the nearly full 16S rRNA genes. PCR products were in-vitro transcribed into RNA,

labeled with Cy3 and Cy5 and subsequently fragmented. Hybridizations were performed in duplicate and data were extracted from microarray-scanned images using Agilent Feature Extraction software version 10.7.3.1 (http://www.agilent.com). Array normalization was performed using a set of R-based scripts (http://r-project.org) in combination with a custom designed relational database which runs under the MySQL database management system (http://www.mysql.com).

Hierarchical clustering of probe profiles was carried out by calculating a distance matrix between the samples based on the squared difference between each pair of profiles (Euclidian distance). The distance matrix was used in the hclust implementation in R of a hierarchical clustering algorithm. The agglomeration method used in this algorithm was Ward's minimum variance method. The bacterial composition was compared at the phylum level (divided into class level for the *Firmicutes*) and at the genus-like level (131 phylogenetic groups with 90% or more 16S rRNA gene sequence similarity) using the Wilcoxon signed-rank test that was corrected for multiple comparisons (q value), in which q<0.05 was considered significantly different. The diversity was determined calculated using the Shannon index of diversity on probe signal intensities. Principle component analysis based on probe profiles was performed using CANOCO 4.5 software package (Biometrics, Wageningen, the Netherlands).

16S rRNA profiling by 454 pyrosequencing

DNA from each of the 12 stool samples was amplified using primers that targeted the V1-V3 regions of the 16S rRNA gene [39]. These primers included the A and B adaptor sequences for 454 pyrosequencing as well as a unique 12 bp barcode incorporated onto the reverse primer such that each sample receives its own unique barcode. The barcode sequences for all 12 samples are provided in File S2. This method of incorporating the A and B adaptors onto the primers at the PCR stage provided minimal loss of sequence data when compared to previous methods that would ligate the A and B adaptors to every amplicon after amplification. This method also allows to generate sequence reads which are all in the same 5'-3' orientation. Using approximately 100ng of extracted DNA, the amplicons were generated with Platinum Taq polymerase (Invitrogen, CA) and by using the following cycling conditions: 95°C for 5min for an initial denaturing step followed by 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec for a total of 35 cycles followed by a final extension step of 72°C for 7 min then stored at 4°C. Once the PCR for each sample was completed, the amplicons were purified using the QIAquick PCR purification kit (Qiagen Valencia, CA), quantified, normalized, and then pooled in preparation for emulsion PCR followed by 454 sequencing using Titanium chemistry (Roche, Basel Switzerland) following the manufacturer's protocol. In the first step of data processing, the generated sequence data were deconvolved using the sample barcodes to identify sequences from each of the samples. Barcode, primer, and adaptor sequences were also trimmed as part of this step. PCR artifacts "chimeras" were identified using the ChimeraSlayer program (http://microbiomeutil.sourceforge.net; reference http://

genome.cshlp.org/content/21/3/494.long), and removed prior to downstream analysis. The resulting deconvoluted and filtered sequence data were assigned taxonomy (to the genus level) using the Ribosomal Database Project (RDP) classifier [40] and the genera classifications were used to generate a sample-genus count matrix. Operational Taxonomic Unit (OTU) analysis of these sequences was performed as follows: sequences were processed (trimmed) using the Mothur software [41] and subsequently clustered at 97% sequence identity [42] using cd-hit [43] to generate OTUs. The OTU memberships of the sequences were used to construct a sample-OTU count matrix. The samples were clustered at genus and OTU levels using the sample-genus and sample-OTU count matrices respectively. For each clustering, Morisita-Horn dissimilarity was used to compute a sample distance matrix from the initial count matrix, and the distance matrix was subsequently used to generate a hierarchical clustering using Ward's minimum variance method. The Wilcoxon Rank Sum test was used to identify OTUs that had differential abundance in the healthy and polyp sample groups.

Metagenomic analysis

Prepared DNA from the 12 stool samples was processed using the Genomiphi MDA kit (GE-Healthcare) using the manufacturer's suggested protocol, in preparation for library construction for 454 sequencing. The sample preparation process in this system involves fragmentation of MDA amplified genomic DNA, followed by ligation of MID barcodes and 454 adaptor sequences. Each sample was then normalized, pooled then loaded into a half picoliter plate for 454 sequencing. The samples were then amplified using emulsion PCR followed by 454 sequencing using Titanium chemistry. The metagenomic sequence data were first processed to remove 454 artifacts (replicate reads arising from the emulsion PCR process) [44] and then genes were identified on reads using a frameshift tolerant gene finder (FragGeneScan) [45] so as to overcome any 454 homopolymer problems. These genes were then searched against the KEGG database to identify kegg ortholog counts [46,47].

All of the 16S and metagenomic data generated in this study have been deposited in the NCBI's Sequence Read Archive (http://www.ncbi.nlm.nih.gov/bioproject/222611).

Results

HITChip based taxonomy analysis

The HITChip analysis revealed the bacterial profile within each of the analyzed samples as well as the clustering of these samples based on profiles' similarities/dissimilarities. The clustering based on the generated profiles did not reveal a clear separation between the polyp and healthy samples' profiles. The samples in the generated clustering were intermixed with 3 polyp (AFR011, 010 & 007) and 3 healthy samples (AFR005, 001 & 006) on one side of the dendrogram located distantly from the other polyp (AFR008, 009 & 012) and healthy samples (AFR002, 003 & 004) (Figure 1).

Bacteroidetes and *Firmicutes* were prevalent in all samples and totaling about 92% of the total detected bacteria. Another

Hierarchical clustering, oligolevel, raw

as.dist(1 - cor(10^(dat) - 1, use = "complete.obs"))
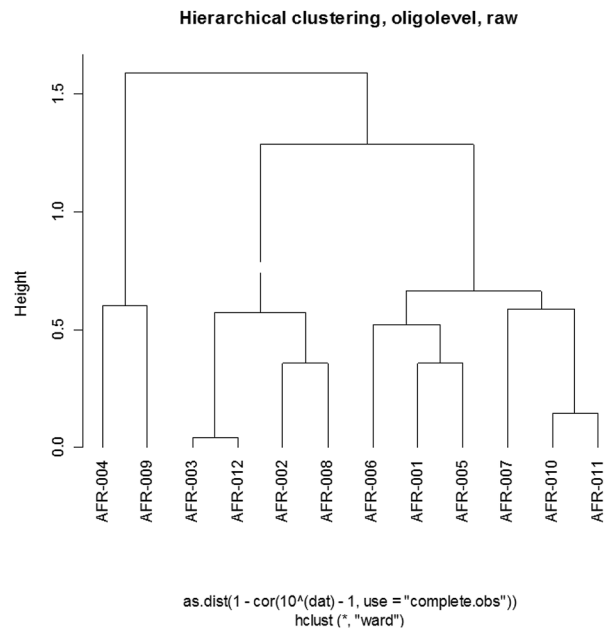hclust (*, "ward")

Figure 1. Dendrogram of the samples' clustering based on the global HITChip oligonucleotide probe signal intensities (AFR001-006; Healthy patients' samples, AFR007-012: Polyp patients' samples).
doi: 10.1371/journal.pone.0081352.g001

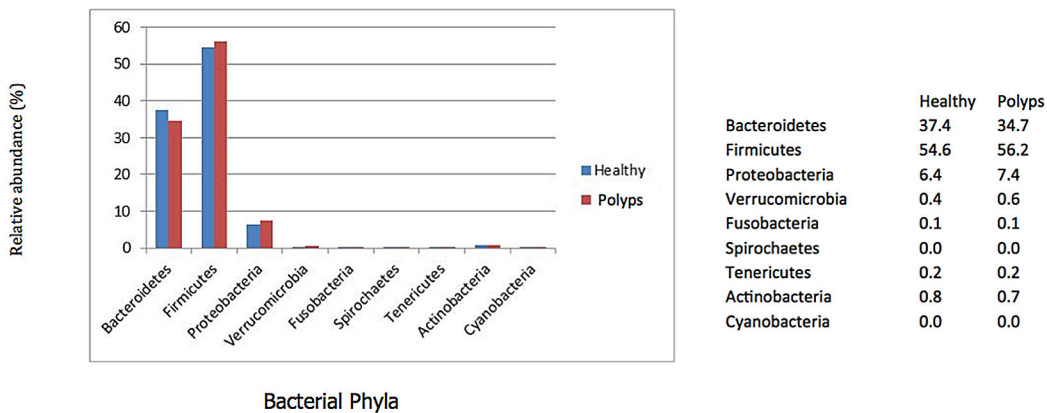|	Healthy	Polyps
Bacteroidetes	37.4	34.7
Firmicutes	54.6	56.2
Proteobacteria	6.4	7.4
Verrucomicrobia	0.4	0.6
Fusobacteria	0.1	0.1
Spirochaetes	0.0	0.0
Tenericutes	0.2	0.2
Actinobacteria	0.8	0.7
Cyanobacteria	0.0	0.0

Figure 2. Barplots showing the relative abundances of bacterial phyla in healthy and colon polyps patients' samples determined by HITChip profiling.
doi: 10.1371/journal.pone.0081352.g002

important group was Proteobacteria which was represented by ~7% of the probes in both sets of samples. Healthy individuals had a relatively higher prevalence of *Bacteroides* (37.4 vs. 34.7%) while polyp samples had a higher prevalence of *Firmicutes* (56.2 vs. 54.6%). The prevalence of Proteobacteria was 6.4 and 7.4 in healthy and polyp samples, respectively (Figure 2). At the genus level, there were several bacteria that displayed relative abundances in the analyzed sets of samples as depicted in Figure 3 and associated File S1.

454 pyrosequencing taxonomic analysis

In parallel to the HIT Chip analysis, we analyzed the same sets of samples using 454 pyrosequencing where V1-V3 variable region of the 16S rDNA sequence was PCR amplified and sequenced. The generated sequences were analyzed using the Ribosomal Database project data and sequence identification was established (File S2). *Bacteroides* and *Firmicutes* represented 80 to 85% of the sequences in both sets of samples with a higher prevalence of *Bacteroides* in healthy samples vs polyps and a higher prevalence of *Firmicutes* in polyp samples vs. healthy ones. Proteobacteria
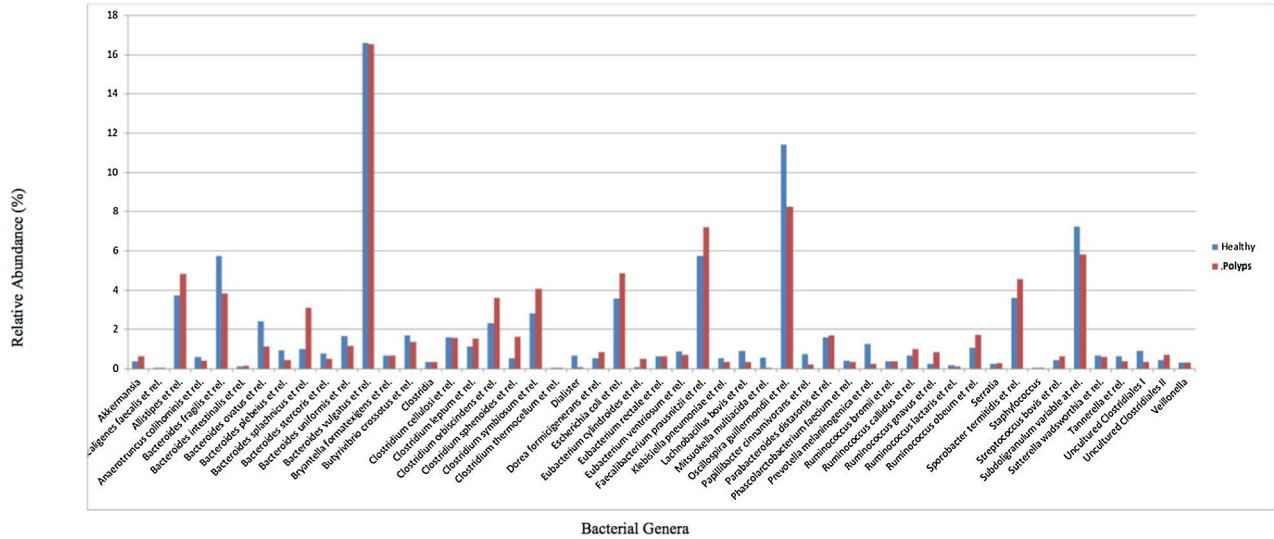
Figure 3. Barplots showing the relative abundances of bacterial groups at genus-like level in healthy and colon polyps patients' samples determined by HITChip profiling. Bacterial groups that are at least 1% in one of the samples are shown.
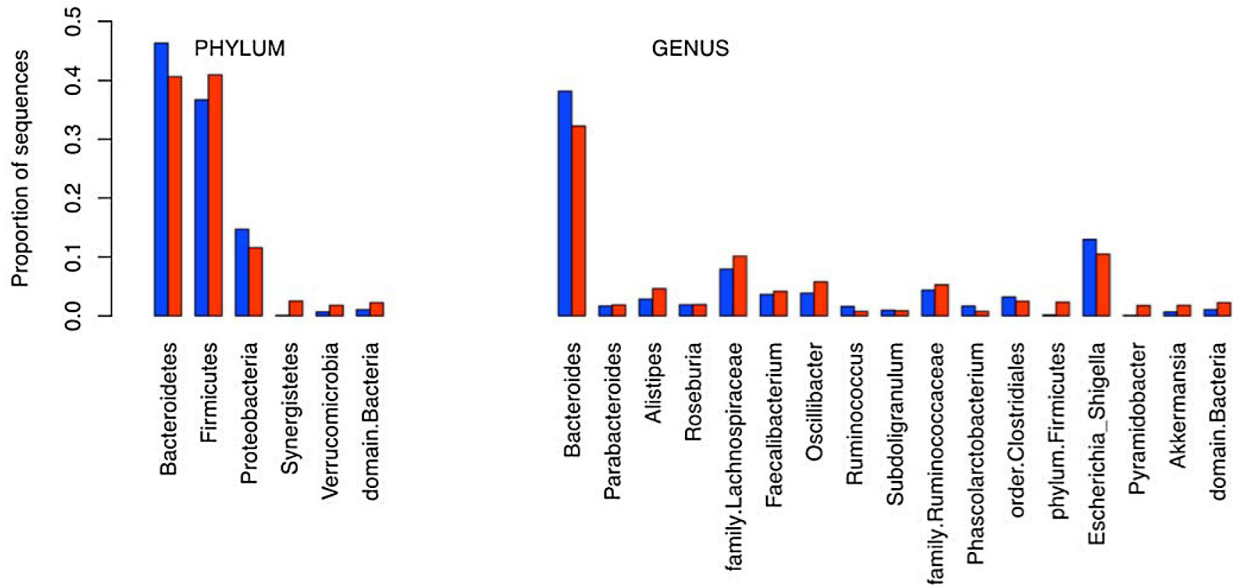
Figure 4. Bacterial phyla and genera distribution in healthy (Blue) and colon polyps (Red) patients' samples based on 454 pyrosequencing data.

were the third major group of bacteria accounting for ~10% of the sequences (Figure 4). A clustering of the samples based on the generated sequences at the genus level led to the dendrogram depicted in Figure 5, in which the healthy and polyp samples were intermixed. Polyp samples AFR011, 010 & 007 clustered with healthy samples AFR002 & 006 while polyp

samples AFR008, 009 & 012 clustered with healthy samples AFR001, 003, 004 and 005 (Figure 5).

Further clustering at the Operational Taxonomic Units level (OTU: sub-genus level) led to a different repartition of the analyzed samples (Figure 6). In this figure, a better resolution was obtained with polyp samples AFR008, 010, 011 & 012 clustering together with healthy sample AFR003 while polyp

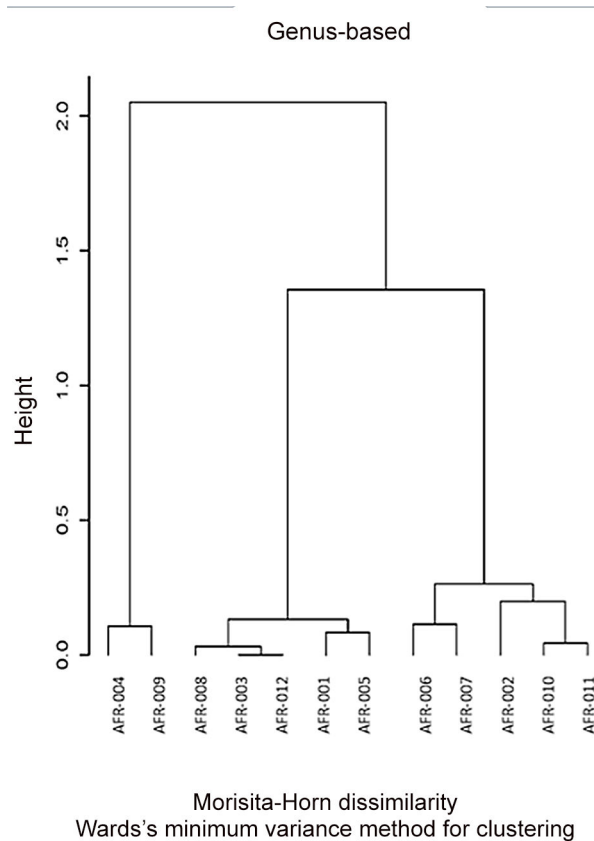**Figure 5. Dendrogram reflecting the healthy (AFR001-006) and colon polyps (AFR007-012) samples clustering based on the 454 pyrosequencing data analyzed at the genus level.**

samples 006 and 007 clustered with the other healthy samples (Figure 6). More importantly, the height (branching distance) obtained at the sub-genus comparison (Figure 6) was bigger than that obtained in the genus level comparison (Figure 5) pointing to much more differences between the two groups at the OTU rather than the genus level. Seven out of 11 OTUs that led to this separation consisted of *Bacteroides* (Table 1).

Metagenomic analysis

The raw count for Kegg Orthologs (KOs) were used to calculate the proportion of KOs in each of the twelve metagenomic datasets (Files S2, S3 & S4) using a method that accounts for difference in gene lengths [46,47]. These proportions were used in a Wilcoxon Rank Sum test to identify kegg orthologs that were differentially abundant in the healthy and disease sample groups. Overall, after correcting for multiple testing, we did not see statistically significant groups (Kegg Orthologs) using a false discovery rate of 5%. Because most of the OTUs leading to a higher resolution of the analyzed samples were from the *Bacteroides* group, we performed an analysis of the metagenomic data against available *Bacteroides* genomes (File S5). All reads from each sample were searched against available *Bacteroides* genomes. A read was considered as mapping to a genome if its match to the genome had ≥80% identity and covered ≥80% of the read sequence. Overall, the healthy samples have a slightly higher mean proportion of recruitment to *Bacteroides*, though this seems to be driven primarily by sample AFR005. The IDs of all *Bacteroides* genome sequences used in this analysis are included in File S5.

Discussion

The microbiota have long been overlooked and nowadays there is an increasing interest in studies that seek to define their role in health and disease [48-50]. The most important site for these studies is the gut since the intestinal microbiota plays major roles in nutrition, metabolism and immunity [48-50]. It has been well documented that many intestinal diseases have bacterial components such as in the case of ulcerative colitis and Crohn's disease [51]. Their role in triggering or promoting colon oncogenic transformation has yet to be established even though many publications reported the potential of many individual bacteria to induce tumorigenesis in germ free mice [21,22].
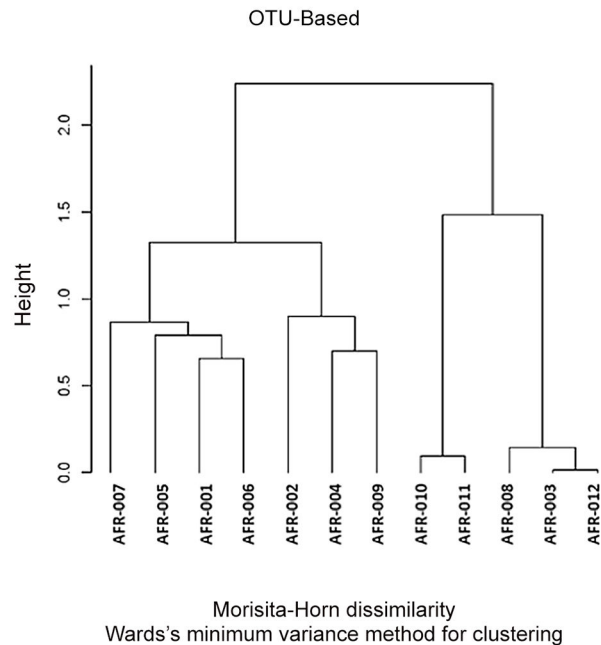
Figure 6. Dendrogram reflecting the healthy (AFR001-006) and colon polyps (AFR007-012) samples clustering based on the 454 pyrosequencing data analyzed at the Operational Taxonomic Unit (OTU: sub-genus) level.
doi: 10.1371/journal.pone.0081352.g006

Table 1. Differential OTUs leading to the separate clustering of the healthy and colon polyps samples the sub-genus level data analysis.

•OTU	p-value	Taxonomy				
•OTU_1776	0.002165	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
•OTU_1807	0.002165	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	
•OTU_96	0.016059	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
•OTU_1775	0.016711	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
•OTU_338	0.028441	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
•OTU_485	0.030348	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	
•OTU_1312	0.037941	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
•OTU_224	0.037941	Bacteria	Firmicutes	Clostridia	Clostridiales		
•OTU_1127	0.040253	Bacteria	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Coprobacillus
•OTU_1554	0.041596	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
OTU_177	0.043826	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroide

doi: 10.1371/journal.pone.0081352.t001

Here we report experiments where we show that there is a trend of specific bacterial profiles in colon polyp patients when compared to healthy individuals. We used two technologies to investigate the microbiota's profiles in our sets of samples. Both the phylogenetic microarraying using the HITChip and 454 pyrosequencing generated similar results and similar bacterial groups distribution. Similar findings were reported by Van den Bogert et al. who reported comparable results from stool and small intestinal samples when the two technologies were compared even when different sets of 16S rRNA primers were used [52].

The genus based clustering using data from both technologies was similar which reflects the strength of both technologies in microbiota profiling. However, no clear resolution of the two sets of samples was obtained at the genus level. Further analysis at the sub-genus level, led to much clearer separation of the samples with 4 of the polyp patients bacterial profiles clustering on one side of the generated dendrogram with bigger branching distance separating the two clusters in the sub-genus dendrogram (Figures 5 & 6). This finding is to be expected since changes in the microbiota in the oncogenic transformation are thought to

be taking place within the existing microbiota where previously less represented bacterial strains become dominant or unexpressed bacterial functions become induced in response to some environmental stressors [53,54]. Diet is known to be an important effector of both microbiota composition and colorectal cancer risk. The magnitude and nature of its effects can only be assessed in well controlled prospective studies.

Bacteroidetes and *Firmicutes* were the most prevalent groups of bacteria detected in all samples. Proteobacteria corresponded to the next group of relevance in our samples.

The sub-genus bacteria that led to the new clustering were predominantly from the *Bacteroides* group in the polyp patients' samples (7 out of 11). Yoshino et al. have reported that *Bacteroides* bacteremia as strongly associated with colorectal cancer in Japanese patients [55]. Also, Wu et al. have shown that strains from the *Bacteroides fragilis* species carrying the bft gene (toxin) are able to promote colon oncogenic transformation in mice models of colon cancer through pSTAT3 pathway [21,22]. Also, Toprak et al. did report a 38% prevalence of bft in colon cancer stool samples vs. 12% in control patients [56].

In an immunohistochemistry experiment using a pSTAT3 antibody on tissue microarray using both African American normal and adenoma tissue samples, adenoma samples were strongly stained when compared to normal samples (data not shown). pSTAT3 is the preferential pathway induced by bft toxin and other bacterial antigens [57]. While other bacterial toxins, from *Bacteroides* bacteria or others, might trigger the pSTAT3 pathway as well, our results do still point to the need to further dissect such functions in *Bacteroides* strains.

The metagenomic analysis in our study was more descriptive of the potential of the microbiome in African American microbiota but did not reveal any statistically significant markers or functions when the samples were compared either individually or as two groups. Overall, the usual bacterial functions were detected in the analyzed samples (See Files S2, S3 & S4). A second analysis of the metagenomic data was done in comparison to all known *Bacteroides* genomes (n=90). This analysis did not lead to significant differences between the two sets of samples as well.

It is noteworthy that recent publications have reported the prevalence of *Fusbacterium* spp and *Fusobacterium nucleatus* [58-60] in colon cancer tumors when compared to normal colon samples. Such was not the case in the analyzed stool samples. This finding points probably to the importance of analyzing and establishing bacterial markers of colon oncogenic transformation in colon tissues and subsequent validation in stool samples. Indeed, adherent bacteria might be more prone to affect gene expression in colon mucosal cells than transient bacteria that are flushed in the fecal samples. Large studies of stool and colon tissue samples from different stages of colon cancer development are needed to establish strong bacterial markers of oncogenic transformation.

Supporting Information

File S1. HITChip bacterial genera distribution in the analyzed samples.
(DOCX)

File S2. 454 Pyrosequencing barcodes and data for all samples.
(XLS)

File S3. Number of metagenomic reads analyzed for all 12 samples.
(XLSX)

File S4. KEGG analysis of the metagenomic data in the two sets of samples.
(XLSX)

File S5. *Bacteroides* strains against which genomes the metagenomic data was analyzed.
(XLSX)

Author Contributions

Conceived and designed the experiments: HB. Performed the experiments: HB MT. Analyzed the data: HB SY EZ EL AL BS KN HA. Contributed reagents/materials/analysis tools: HB SY EZ MT KN HA. Wrote the manuscript: HB SY EZ HA.

References

1. Sinicrope PS, Goode EL, Limburg PJ, Vernon SW, Wick JB et al. (2012) A population-based study of prevalence and adherence trends in average risk colorectal cancer screening, 1997 to 2008. Cancer Epidemiol Biomarkers Prev 21: 347-350. doi:10.1158/1055-9965.EPI-11-0818. PubMed: 22144500.
2. Friedenberg FK, Singh M, George NS, Sankineni A, Shah S (2012) Prevalence and distribution of adenomas in black Americans undergoing colorectal cancer screening. Dig Dis Sci 57: 489-495. doi:10.1007/s10620-011-1952-z. PubMed: 22052446.
3. Blumenstein I, Tacke W, Bock H, Filmann N, Lieber E et al. (2013) Prevalence of colorectal cancer and its precursor lesions in symptomatic and asymptomatic patients undergoing total colonoscopy: results of a large prospective, multicenter, controlled endoscopy study. Eur J Gastroenterol Hepatol 25: 556–61. PubMed: 23283303.
4. Nelson RL, Persky V, Turyk M (1999) Carcinoma in situ of the colorectum: SEER trends by race, gender, and total colorectal cancer. J Surg Oncol 71: 123-129. doi:10.1002/(SICI)1096-9098(199906)71:2. PubMed: 10389871.
5. Smith RA, von Eschenbach AC, Wender R, Levin B, Byers T, et al. (2001) American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers. Also: update 2001--testing for early lung cancer detection. CA: a cancer journal for clinicians 51: 38-80; quiz
6. Agrawal S, Bhupinderjit A, Bhutani MS, Boardman L, Nguyen C et al. (2005) Colorectal cancer in African Americans. Am J Gastroenterol 100: 514-523; discussion 10.1111/j.1572-0241.2005.41829.x. PubMed: 15743345.
7. Nouraie M, Hosseinkhah F, Brim H, Zamanifekri B, Smoot DT et al. (2010) Clinicopathological features of colon polyps from African-Americans. Dig Dis Sci 55: 1442-1449. doi:10.1007/s10620-010-1133-5. PubMed: 20225129.
8. Ashktorab H, Belgrave K, Hosseinkhah F, Brim H, Nouraie M et al. (2009) Global histone H4 acetylation and HDAC2 expression in colon adenoma and carcinoma. Dig Dis Sci 54: 2109-2117. doi:10.1007/s10620-008-0601-7. PubMed: 19057998.

9. Ashktorab H, Green W, Finzi G, Sessa F, Nouraie M et al. (2012) SEL1L, an UPR response protein, a potential marker of colonic cell transformation. Dig Dis Sci 57: 905-912. doi:10.1007/s10620-011-2026-y. PubMed: 22350780.

10. Ashktorab H, Nguza B, Fatemi M, Nouraie M, Smoot DT et al. (2011) Case-control study of vitamin D, dickkopf homolog 1 (DKK1) gene methylation, VDR gene polymorphism and the risk of colon adenoma in African Americans. PLOS ONE 6: e25314. doi:10.1371/journal.pone.0025314. PubMed: 22022386.

11. Ashktorab H, Schäffer AA, Daremipouran M, Smoot DT, Lee E et al. (2010) Distinct genetic alterations in colorectal cancer. PLOS ONE 5: e8879. doi:10.1371/journal.pone.0008879. PubMed: 20126641.

12. Brim H, Kumar K, Nazarian J, Hathout Y, Jafarian A et al. (2011) SLC5A8 gene, a transporter of butyrate: a gut flora metabolite, is frequently methylated in African American colon adenomas. PLOS ONE 6: e20216. doi:10.1371/journal.pone.0020216. PubMed: 21687703.

13. Brim H, Lee E, Abu-Asab MS, Chaouchi M, Razjouyan H et al. (2012) Genomic aberrations in an African American colorectal cancer cohort reveals a MSI-specific profile and chromosome X amplification in male patients. PLOS ONE 7: e40392. doi:10.1371/journal.pone.0040392. PubMed: 22879877.

14. Brim H, Mokarram P, Naghibalhossaini F, Saberi-Firoozi M, Al-Mandhari M et al. (2008) Impact of BRAF, MLH1 on the incidence of microsatellite instability high colorectal cancer in populations based study. Mol Cancer 7: 68. doi:10.1186/1476-4598-7-68. PubMed: 18718023.

15. Kumar K, Brim H, Mokarram P, Naghibalhossaini F, Saberi-Firoozi M, et al. (2009) Distinct BRAF (V600E) and KRAS mutations in high microsatellite instability sporadic colorectal cancer in African Americans. Clin Cancer Res. In press: 68

16. Graham A, Adeloye D, Grant L, Theodoratou E, Campbell H (2012) Estimating the incidence of colorectal cancer in Sub-Saharan Africa: A systematic analysis. J Glob Health 2: 20404. PubMed: 23289079.

17. Sears CL (2005) A dynamic partnership: celebrating our gut flora. Anaerobe 11: 247-251. doi:10.1016/j.anaerobe.2005.05.001. PubMed: 16701579.

18. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005) Host-bacterial mutualism in the human intestine. Science 307: 1915-1920. doi:10.1126/science.1104816. PubMed: 15790844.

19. Hooper LV, Gordon JI (2001) Commensal host-bacterial relationships in the gut. Science 292: 1115-1118. doi:10.1126/science.1058709. PubMed: 11352068.

20. López-Boado YS, Wilson CL, Hooper LV, Gordon JI, Hultgren SJ et al. (2000) Bacterial exposure induces and activates matrilysin in mucosal epithelial cells. J Cell Biol 148: 1305-1315. doi:10.1083/jcb.148.6.1305. PubMed: 10725342.

21. Wu S, Morin PJ, Maouyo D, Sears CL (2003) Bacteroides fragilis enterotoxin induces c-Myc expression and cellular proliferation. Gastroenterology 124: 392-400. doi:10.1016/S0016-5085(03)81986-8. PubMed: 12557145.

22. 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. doi:10.1038/nm.2015. PubMed: 19701202.

23. Savage DC (1977) Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 31: 107-133. doi:10.1146/annurev.mi.31.100177.000543. PubMed: 334036.

24. Human Microbiome Project Consortium (2012) A framework for human microbiome research. Nature 486: 215-221. doi:10.1038/nature11209. PubMed: 22699610.

25. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486: 207-214. doi:10.1038/nature11234. PubMed: 22699609.

26. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L et al. (2005) Diversity of the human intestinal microbial flora. Science 308: 1635-1638. doi:10.1126/science.1110591. PubMed: 15831718.

27. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ et al. (2006) Metagenomic analysis of the human distal gut microbiome. Science 312: 1355-1359. doi:10.1126/science.1124234. PubMed: 16741115.

28. Huycke MM, Gaskins HR (2004) Commensal bacteria, redox stress, and colorectal cancer: mechanisms and models. Exp Biol Med (Maywood) 229: 586-597. PubMed: 15229352.

29. Lax AJ, Thomas W (2002) How bacteria could cause cancer: one step at a time. Trends Microbiol 10: 293-299. doi:10.1016/S0966-842X(02)02360-0. PubMed: 12088666.

30. McGarr SE, Ridlon JM, Hylemon PB (2005) Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. J Clin Gastroenterol 39: 98-109. PubMed: 15681903.

31. Zoetendal EG, Rajilic-Stojanovic M, de Vos WM (2008) High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. Gut 57: 1605-1615. doi:10.1136/gut.2007.133603. PubMed: 18941009.

32. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG et al. (2012) Human gut microbiome viewed across age and geography. Nature 486: 222-227. PubMed: 22699611.

33. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA (2012) The application of ecological theory toward an understanding of the human microbiome. Science 336: 1255-1262. doi:10.1126/science.1224203. PubMed: 22674335.

34. Balter M (2012) Taking stock of the human microbiome and disease. Science 336: 1246-1247. doi:10.1126/science.336.6086.1246. PubMed: 22674333.

35. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464: 59-65. doi:10.1038/nature08821. PubMed: 20203603.

36. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T et al. (2011) Enterotypes of the human gut microbiome. Nature 473: 174-180. doi:10.1038/nature09944. PubMed: 21508958.

37. Lyra A, Forssten S, Rolny P, Wettergren Y, Lahtinen SJ et al. (2012) Comparison of bacterial quantities in left and right colon biopsies and faeces. World J Gastroenterol 18: 4404-4411. doi:10.3748/wjg.v18.i32.4404. PubMed: 22969206.

38. Rajilić-Stojanović M, Heilig HG, Molenaar D, Kajander K, Surakka A et al. (2009) Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. Environ Microbiol 11: 1736-1751. doi:10.1111/j.1462-2920.2009.01900.x. PubMed: 19508560.

39. Jeraldo P, Chia N, Goldenfeld N (2011) On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys. Environ Microbiol 13: 3000-3009. doi:10.1111/j.1462-2920.2011.02577.x. PubMed: 21910812.

40. 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. Nucleic Acids Res 37: D141-D145. doi:10.1093/nar/gkp353. PubMed: 19004872.

41. 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. doi:10.1128/AEM.01541-09. PubMed: 19801464.

42. Bond PL, Hugenholtz P, Keller J, Blackall LL (1995) Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. Appl Environ Microbiol 61: 1910-1916. PubMed: 7544094.

43. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659. doi:10.1093/bioinformatics/btl158. PubMed: 16731699.

44. Schütze T, Rubelt F, Repkow J, Greiner M, Erdmann VA et al. (2011) A streamlined protocol for emulsion polymerase chain reaction and subsequent purification. Anal Biochem 410: 155-157. doi:10.1016/j.ab.2010.11.029. PubMed: 21111698.

45. Rho M, Tang H, Ye Y (2010) FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res 38: e191. doi:10.1093/nar/gkq747. PubMed: 20805240.

46. Bodaker I, Sharon I, Suzuki MT, Feingersch R, Shmoish M et al. (2010) Comparative community genomics in the Dead Sea: an increasingly extreme environment. ISME J 4: 399-407. doi:10.1038/ismej.2009.141. PubMed: 20033072.

47. Sharon I, Bercovici S, Pinter RY, Shlomi T (2011) Pathway-based functional analysis of metagenomes. J Comput Biol 18: 495-505. doi:10.1089/cmb.2010.0260. PubMed: 21385050.

48. Pick M (2012) Gut flora on a crusade for good. A brief overview of probiotics. Adv NPs PAs 3: 35-36.

49. Berer K, Krishnamoorthy G (2012) Commensal gut flora and brain autoimmunity: a love or hate affair? Acta Neuropathol 123: 639-651. doi:10.1007/s00401-012-0949-9. PubMed: 22322994.

50. Goto Y, Kiyono H (2012) Epithelial barrier: an interface for the cross-communication between gut flora and immune system. Immunol Rev 245: 147-163. doi:10.1111/j.1600-065X.2011.01078.x. PubMed: 22168418.

51. Chandran P, Satthaporn S, Robins A, Eremin O (2003) Inflammatory bowel disease: dysfunction of GALT and gut bacterial flora (I). Surgeon 1: 63-75. doi:10.1016/S1479-666X(03)80118-X. PubMed: 15573623.

52. van den Bogert B, de Vos WM, Zoetendal EG, Kleerebezem M (2011) Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal

samples. Appl Environ Microbiol 77: 2071-2080. doi:10.1128/AEM. 02477-10. PubMed: 21257804.

53. Boleij A, Dutilh BE, Kortman GA, Roelofs R, Laarakkers CM et al. (2012) Bacterial responses to a simulated colon tumor microenvironment. Mol Cell Proteomics 11: 851-862. doi:10.1074/mcp.M112.019315. PubMed: 22713208.

54. Marteau P, Chaput U (2011) Bacteria as trigger for chronic gastrointestinal disorders. Dig Dis 29: 166-171. doi: 10.1159/000323879. PubMed: 21734380.

55. Yoshino Y, Kitazawa T, Ikeda M, Tatsuno K, Yanagimoto S et al. (2012) Clinical features of Bacteroides bacteremia and their association with colorectal carcinoma. Infection 40: 63-67. doi:10.1007/s15010-011-0159-8. PubMed: 21773761.

56. 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. Clinical Microbiology and Infection: the

57. Official Publication of the European Society of Clinical Microbiology and Infectious Diseases 12: 782-786.

Samavati L, Rastogi R, Du W, Hüttemann M, Fite A et al. (2009) STAT3 tyrosine phosphorylation is critical for interleukin 1 beta and interleukin-6 production in response to lipopolysaccharide and live bacteria. Mol Immunol 46: 1867-1877. doi:10.1016/j.molimm. 2009.02.018. PubMed: 19299019.

58. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M et al. (2012) Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res 22: 299-306. doi:10.1101/gr. 126516.111. PubMed: 22009989.

59. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F et al. (2012) Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res 22: 292-298. doi:10.1101/gr. 126573.111. PubMed: 22009990.

60. Ray K (2011) Colorectal cancer: Fusobacterium nucleatum found in colon cancer tissue--could an infection cause colorectal cancer? Nat Rev Gastroenterol Hepatol 8: 662. doi:10.1038/nrgastro.2011.208. PubMed: 22083120.