Front. Med. 2019, 13(4): 461–470
https://doi.org/10.1007/s11684-019-0695-7

RESEARCH ARTICLE

Altered intestinal microbiota associated with colorectal cancer

Hong Zhang[1,*], Ying Chang[2,*], Qingqing Zheng[2], Rong Zhang[1], Cheng Hu (✉)[1], Weiping Jia[1]

[1]*Shanghai Diabetes Institute, Shanghai Key Laboratory of Diabetes Mellitus, Shanghai Clinical Center for Diabetes, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China;* [2]*Digestive Endoscopic Center, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China*

Abstract **The gut microbiota plays an important role in the development and progression of colorectal cancer (CRC). To learn more about the dysbiosis of carcinogenesis, we assessed alterations in gut microbiota in patients with CRC. A total of 23 subjects were enrolled in this study: 9 had CRC (CRC group) and 14 had normal colons (normal group). The microbiome of the mucosal–luminal interface of each subject was sampled and analyzed using 16S rRNA gene amplicon sequencing. We also used Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) to predict microbial functional profiles. The microbial composition of the mucosal lumen differed between the groups, and the presence of specific bacteria may serve as a potential biomarker for colorectal carcinogenesis. We identified a significant reduction in *Eubacterium,* which is a butyrate-producing genera of bacteria, and a significant increase in *Devosia* in the gut microbiota of CRC patients. Different levels of gut microflora in healthy and CRC samples were identified. The observed abundance of bacterial species belonging to *Eubacterium* and *Devosia* may serve as a promising biomarker for the early detection of CRC.**

Keywords colorectal cancer (CRC); gut microbiota; intestinal; *Eubacterium*; *Devosia*

Introduction

Colorectal cancer (CRC) is one of the most common malignant tumors in the Western world and, second to lung cancer, causes the greatest number of deaths from cancer [1,2]. The incidence of CRC in China has recently increased.

Epidemiological studies have shown that the occurrence and development of sporadic CRC is a multifactorial and multi-step process that is jointly affected by environmental and genetic factors [3]. However, certain populations are at higher risk of developing the disease than others, including those with a family history of CRC and colon polyps, a personal history of CRC or polyps, patients with chronic inflammatory bowel disease, and those over 50 years of age. Accumulating evidence shows links between microbial populations and different stages of tumor development. For example, *Helicobacter pylori* is correlated with the development of gastric cancer [4].

Alterations in the gut microbiota, whether caused by lifestyle, diet, environmental factors, or infection, can change the symbiotic relationship between a host organism and its environment and have increasingly been associated with the occurrence of CRC [5,6]. No single bacterial species has yet been determined to be a risk factor for CRC, but pathogenic bacterial species are directly responsible for approximately 15% of all CRC cases [7]. *Fusobacterium* is a common bacterium significantly enriched in the gut microbiota of CRC patients compared with that in normal controls [8–10]. The *Clostridium leptum* and *Clostridium coccoides* subgroups are specific to CRC in fecal microbiota [11]. The pathogenesis of CRC is not well understood, and whether specific intestinal bacteria are associated with the dysbiosis of CRC remains unclear. Microbiota from mucosal samples represent the underlying dysbiosis and seem to be more appropriate for detecting shifts in microbial composition than fecal samples [11]. However, previous studies mainly focus on differences in fecal microbiota between CRC and healthy controls. Therefore, whether the intestinal microbiome plays a major role in the early stages of CRC remains unclear. Moreover, most of the relevant studies are conducted in Western populations, the lifestyles and dietary habits of which are quite different from those of Chinese populations.

Received April 15, 2018; accepted March 13, 2019

Correspondence: Cheng Hu, alfredhc@sjtu.edu.cn

*These two authors contributed equally to this work.

In this study, we determined the microbial composition of samples from normal and CRC-diagnosed subjects. Most published studies use fecal samples to detect gut microbiota because these samples are convenient to collect; however, in the present work, we used colonic mucosal-luminal interface samples and analyzed them using 16S rRNA sequencing. The role of butyrate-producing bacteria, which present distinct interactions with the host at mucosal surfaces, in the dysbiosis of CRC is a current area of research. Some butyrate-producing bacteria, such as *Eubacterium* and *Faecalibacterium*, are well suited to colonize intestinal mucosal surfaces, thereby indicating that mucosal-luminal interface samples, rather than stool samples, may be more suitable for detecting these important bacteria. Here, we compared the bacterial community structures of CRC with those of normal colons and identified several potential bacterial genera and species associated with the dysbiosis of CRC.

Materials and methods

Subject enrolment

A total of 23 subjects, 9 with CRC and 14 with healthy colons, were selected randomly from Shanghai Jiao Tong University Affiliated Sixth People's Hospital, China. Written informed consent was obtained from all participants, and the study protocol was approved by the Research Ethics Board of the Hospital. Endoscopy is the gold standard for the diagnosis of colorectal diseases. All eligible adult patients who were scheduled to undergo colonoscopies and were not diagnosed with other diseases except CRC were included in the study. Patients with intact colons served as normal controls. The exclusion criteria were related to known conditions affecting the intestinal microbiota composition including: (1) irritable bowel syndrome, (2) use of any antibiotics or probiotics within the past 30 days before colonoscopy, or (3) infectious gastroenteritis within the past 60 days.

Sample collection

Mucosal-luminal interface samples were collected from the ascending colon using an operation method similar to that previously published [12]. Preparations for the colonoscopy were completed within a day based on standard protocols. During colonoscopy, when the cecum and proximal ascending colon were reached, any loose fluid and debris that were present were aspirated. Sterile water was flushed onto the mucus to remove the mucus layers from the mucosal epithelial cells, and the mixture of water, mucus, and intestinal cells was aspirated into sterile containers through the colonoscope. These mucosal-luminal interface samples were instantly frozen in liquid nitrogen and then stored at −80 °C for analysis.

DNA extraction and 16S rRNA sequencing

DNA extraction from the mucosal-luminal interface samples was performed using the QIAamp DNA Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. DNA integrity was assessed using 1% agarose gel electrophoresis, and purity and concentration were assessed using a NanoDrop2000 UV spectrophotometer (Thermo Scientific, USA). Barcoded amplicons were generated covering variable region 6 (V6) of the 16S rRNA gene using universal primers (1048F, 5′-GTGSTGCAYGGYYGTCGTCA-3′; 1194R, 5′-ACGTCRTCCMCNCCTTCCTC-3′) and incorporating the Illumina paired-end sequencing adapters and barcode sequences. PCR was conducted using a GeneAmp® PCR System 9700 thermal cycler (Applied Biosystems) with the following parameters: 3 min of initial denaturation at 94 °C, followed by 25 cycles of 94 °C for 10 s, 55 °C for 15 s, and 72 °C for 30 s, and a final extension at 72 °C for 7 min. Each PCR reaction mixture (25 μL) contained 10 ng of genomic DNA, 2.5 μL of 10× Ex Taq Buffer (Takara), 0.5 μL of each primer, 1 μL of dNTP (2.5 mmol/L each), and 0.1 μL of Ex Taq DNA polymerase (Takara). The amplicons of each sample were separated on a 1% agarose gel and purified using an Agencourt AMPure XP kit (Beckman Coulter, USA). The concentration of purified DNA for each reaction was measured using the Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA). Finally, amplicon libraries from all samples were pooled at equal molar concentrations and then sequenced using a 500-cycle MiSeq reagent kit via the paired-end (2 × 251 bp) method on the Illumina MiSeq platform.

Bioinformatics analysis of sequencing data

Raw sequences were assigned to different samples according to their barcodes. After removing adapters and reads with an average quality value lower than 20 or any sequence where the longest homopolymer was greater than 8 nt, only reads longer than 350 bp were considered. Unique clean reads were aligned in accordance with the SILVA database [13] using k-mer-based methods. The resulting sequences were processed and analyzed using Mothur software (v.1.39.5) [14]. Potential chimeric sequences were screened using the chimera.uchime command and deleted by the remove.seqs command of Mothur. The sequences were then classified using the Bayesian classifier with the classify.seqs command, and undesirables were removed by the remove.lineage command.

We combined the nearest neighbor algorithm in Mothur

with a 0.03 distance unit cut-off to cluster the reads. For high-quality sequences, operational taxonomic units (OTUs) were clustered at 97% similarity. The bacterial community composition of each sample was counted at different taxonomic levels by comparison with the Greengenes 13.5 database. Thereafter, we analyzed the richness estimator (Chao), diversity index (Shannon), and species composition at various taxonomic levels (i.e., phylum, class, order, family, genus, and species). In the present data, sequences $> 97\%$ identical to each other were considered to correspond to the same OTUs, representing a group of reads presumably belonging to the same species.

We also performed unweighted UniFrac principal coordinate analysis (PCoA) based on the matrix of distance. By using a nonparametric Kruskal–Wallis rank sum test, linear discriminant analysis effect size (LEfSe) [15] analysis allows for a quick comparison between multiple groups; species with significant differences in abundance were identified as biomarkers.

We performed Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) [16] analysis to predict the functional potential of bacteria communities based on the 16S rRNA data. To do so, pick_closed_reference_otus.py script was used to cluster the reads into a collection of OTUs sharing 97% sequence identity. The OTUs were normalized based on 16S rRNA gene copy numbers by employing the normalize_by_copy_number.py script. The normalized OTU table was used as input data to predict microbial community metagenomes with the predict_metagenomes.py script, and metagenome prediction was further categorized into Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways at levels 2 and 3. The edgeR package in R was applied to identify KEGG pathways that were significantly different between groups.

Statistical analyses

Data were analyzed using the Kruskal–Wallis test or Mann–Whitney test (two-sided) for continuous variables and Fisher's exact test for categorical variables using GraphPad Prism 6 or R (version 3.4.0). The threshold of statistical significance was set at $P \leqslant 0.05$.

Data access

The 16S rRNA gene sequencing data generated in this study were submitted to the NCBI Sequence Read Archive under accession number SRP139052.

Results

Eligible subjects and samples

A total of 23 subjects were included in this study: 14 healthy individuals in the normal group (NG) and 9 individuals in the CRC group (CG). The mean age of the subjects was 62.6 ± 8.9 years in CG and 44.1 ± 15 years in NG. Seven males (50%) were in NG, while six males (67%) were in CG (Table 1). CG had more males and older participants than NG. Details of the clinical information of the patients are shown in Supplementary Table 1.

Characteristics of 16S rRNA gene sequencing

After removing sequences shorter than 350 bp and sequences with ambiguous bases or homopolymers longer than 8 nt, a total of 540 119 high-quality sequences were obtained from 23 samples. Samples showed an average of 23 483 sequences, and no significant difference between NG and CG was found (Fig. 1A, $P = 0.27$). The Chao community richness estimator is shown in Fig. 1B. Although no significant difference in community richness between the two groups was found, CG tended to have lower richness than NG ($P = 0.08$). The Shannon index (Fig. 1C) suggested that CG has lower bacterial diversity than NG. To compare the types and amounts of microflora between the two groups, we conducted unweighted UniFrac PCoA based on the OTUs of each sample. The overall microbial communities of the CRC and normal groups were similar according to PC1 and PC2 (CG =

Table 1 Summary of population information

	NG ($n = 14$)	CG ($n = 9$)	P value
Male/female	7/7	6/3	0.6693
Age (year), mean±SD	44.1±15	62.6±8.9	0.0041
BMI (kg/m^2), mean±SD	22.03±2.68	22.37±2.17	0.7527
Family history of CRC, %	0	22.2	0.0001

P values are based on the Mann–Whitney test (two-sided) for continuous variables and Fisher's exact test for categorical variables. BMI, body mass index; NG, normal group; CG, CRC group.

8.2% of the variance explained; NG = 6.52% of the variance explained; Supplementary Fig. 1); however, CG exhibited lower diversity than NG.

Comparisons of gut microbiota at different levels

We analyzed the bacterial communities in the mucosal-luminal interface samples taken from NG and CG participants. The overall microbial structures at the phylum, class, order, and family levels were assessed by taxonomic assignment for each group (Fig. 2). The dominant phyla in both groups were Firmicutes, Proteobacteria, Bacteroidetes, and Fusobacteria (Fig. 2A). Firmicutes was the highest contributor to the bacterial populations in NG (50.48%) and CG (45.08%). Proteobacteria was the second highest contributing phylum and accounted for 33.55% and 43.76% of the bacteria in NG and CG, respectively. Bacteroidetes was the third highest contributing phylum and accounted for 13.05% and 6.95% of the bacteria in NG and CG, respectively. Fusobacteria was the fourth most abundant phylum, contributed 1.45% of the bacteria in NG and 3.06% of those in CG. Within individual study participants, the composition of the microbes was highly variable. For example, Firmicutes

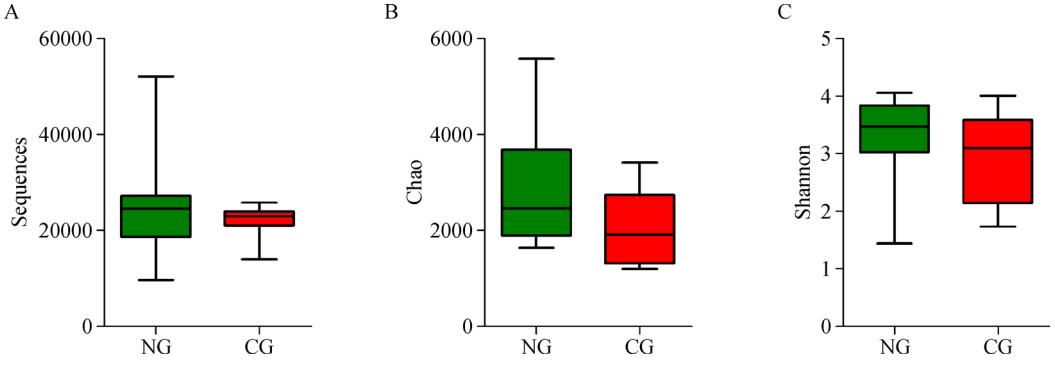

Fig. 1 Quality control of the sequencing data. (A) Sequence number, (B) Chao richness estimator, and (C) Shannon index of the 16S rRNA sequencing data for the normal (NG) and CRC (CG) groups.

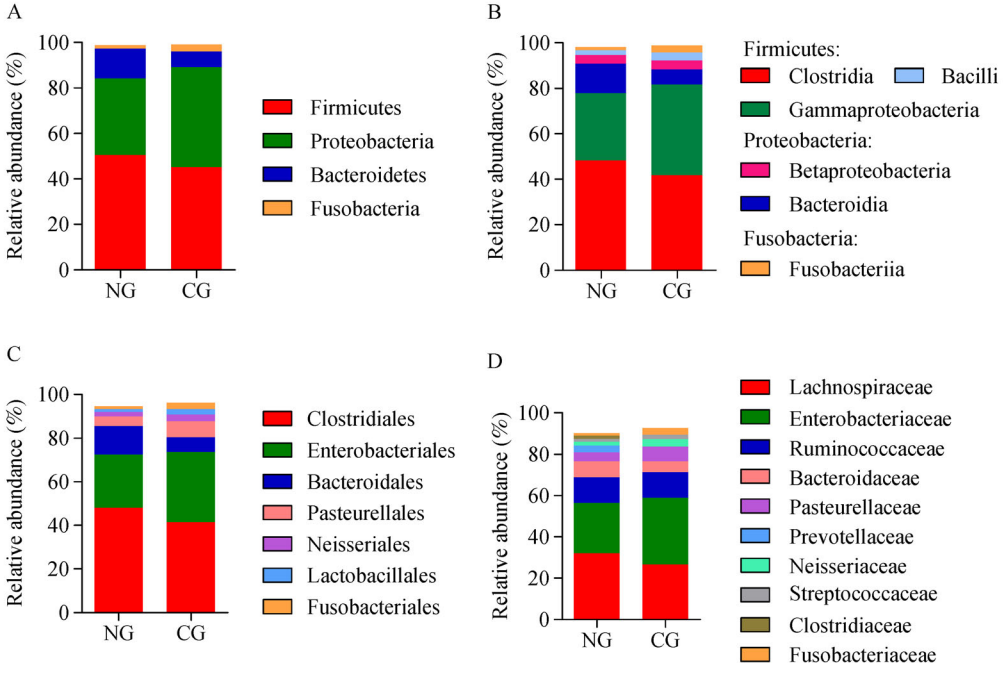

Fig. 2 Bacterial taxonomic groups discriminating between NG and CG. Relative abundances of dominant bacterial taxonomic groups discriminating between NG and CG. (A) Phylum level, (B) class level, (C) order level, and (D) family level.

accounted for 5.09%–85.99%, Bacteroidetes contributed 0.47%–45.17%, and Fusobacteria provided 0.11%–15.52% of the microbial composition in all study participants (Supplementary Fig. 2). In addition, the numbers of Bacteroidetes were higher whereas the numbers of Proteobacteria were lower in the gut microbiota of the NG samples compared with those of the CG samples. Although the distribution of these phyla did not differ significantly between the groups, Fusobacteria increased in CG relative to that in NG. This trend was also observed at the class, order, and family levels (Fig. 2B–2D). Clostridia was the dominant class and order, and a trend, although not statistically significant, toward decreasing numbers from NG to CG was found. The compositions of the dominant genera in the samples are shown in Supplementary Fig. 3. *Blautia*, *Salmonella*, *Faecalibacterium*, *Bacteroides*, *Dorea*, *Hemophilus*, *Coprococcus*, *Prevotella*, *Neisseria*, and *Streptococcus* were the 10 most abundant genera in NG. In CG, the 10 most abundant genera were *Salmonella*, *Faecalibacterium*, *Blautia*, *Bacteroides*, *Coprococcus*, *Hemophilus*, *Dorea*, *Neisseria*, *Fusobacterium*, and *Streptococcus*. These genera constituted over 60% of the total bacteria in each group, and nine of these genera were common to both groups. *Salmonella* was the most abundant genus in CG (21.7% vs. 10.9% in NG, $P = 0.07$), while *Prevotella* was higher in NG than in CG. Only the 20 most abundant species in each group are shown in Supplementary Fig. 4. Although not statistically significant, *Clostridium perfringens* was higher in NG (0.16%) than in CG (0.0042%).

Identification of key contributors for structural segregation of the groups

LEfSe software presents a powerful identification function that can identify high-dimensional biomarkers and reveal genomic features through statistically significant biological differences. The algorithm emphasizes statistical significance and biological correlations, allowing researchers to identify different features and related categories. In this study, LEfSe was used to identify specific bacterial phylotypes for which the abundance was significantly different between the groups (Fig. 3); the details are shown in Supplementary Table 2. *Devosia* (class Alphaproteobacteria, order Rhizobiales, family Hyphomicrobiaceae) was enriched in the CG. The species *stercorea* (genus *Prevotella*, family Prevotellaceae), *copri* (genus *Prevotella*, family Prevotellaceae), and *producta* (genus *Blautia*, family Lachnospiraceae; the genera *Eubacterium* (family Erysipelotrichaceae), *02d06* (family Clostridiaceae), *Phascolarctobacterium* (family Veillonellaceae), *Leptotrichia* (family Leptotrichiaceae), and *Klebsiella* (family Enterobacteriaceae); and the family Desulfovibrionaceae (class Deltaproteobacteria, order Desulfovibrionales) were enriched in NG (Table 2 and Fig. 3).

Functional properties predicted by PICRUSt

We performed PICRUSt analysis of the 16S rRNA sequences to determine whether the taxonomic differences between the groups corresponded to functional changes

Table 2 Relative abundance of significantly enriched bacterial taxa from different groups

Group enriched	Taxonomic rank	Bacteria	NG (%)	CG (%)	*P* value
CG	Genus	*Devosia*	0.006 (0.003–0.010)	0.023 (0.009–0.029)	0.008
	Family	Hyphomicrobiaceae	0.006 (0.003–0.010)	0.023 (0.009–0.029)	0.008
	Order	Rhizobiales	0.008 (0.003–0.012)	0.027 (0.019–0.034)	0.012
	Class	Alphaproteobacteria	0.010 (0.006–0.017)	0.031 (0.027–0.039)	0.020
NG	Species	*stercorea*	0.063 (0.042–0.138)	0.004 (0–0.035)	0.001
	Species	*copri*	0.126 (0.051–0.993)	0.009 (0.004–0.115)	0.006
	Family	Prevotellaceae	0.225 (0.172–1.310)	0.042 (0.023–0.170)	0.006
	Genus	*Eubacterium*	0.025 (0.016–0.123)	0.005 (0–0.017)	0.009
	Genus	*Klebsiella*	0.020 (0.010–0.090)	0.008 (0.005-0.012)	0.014
	Order	Desulfovibrionales	0.127 (0.049–0.441)	0.037 (0.021–0.104)	0.014
	Class	Deltaproteobacteria	0.127 (0.049–0.446)	0.037 (0.025–0.104)	0.017
	Family	Desulfovibrionaceae	0.126 (0.045–0.368)	0.037 (0.017–0.104)	0.017
	Genus	*02d06*	0.044 (0.015–0.074)	0.009 (0–0.025)	0.018
	Species	*producta*	0.189 (0.033–0.287)	0.014 (0.008–0.059)	0.023
	Genus	*Phascolarctobacterium*	0.062 (0.005–0.125)	0.005 (0–0.014)	0.027
	Family	Clostridiaceae	0.269 (0.164–0.843)	0.122 (0.033–0.182)	0.038
	Genus	*Leptotrichia*	0.028 (0.021–0.132)	0.010 (0.005–0.016)	0.038

Data are shown as medians and interquartile ranges (IQR). Statistical analysis was conducted via the Mann–Whitney test. NG and CG data reflect the relative abundance (percentage) of all sequences in each group. Taxonomic rank indicates the level of the taxon in taxonomic hierarchy.

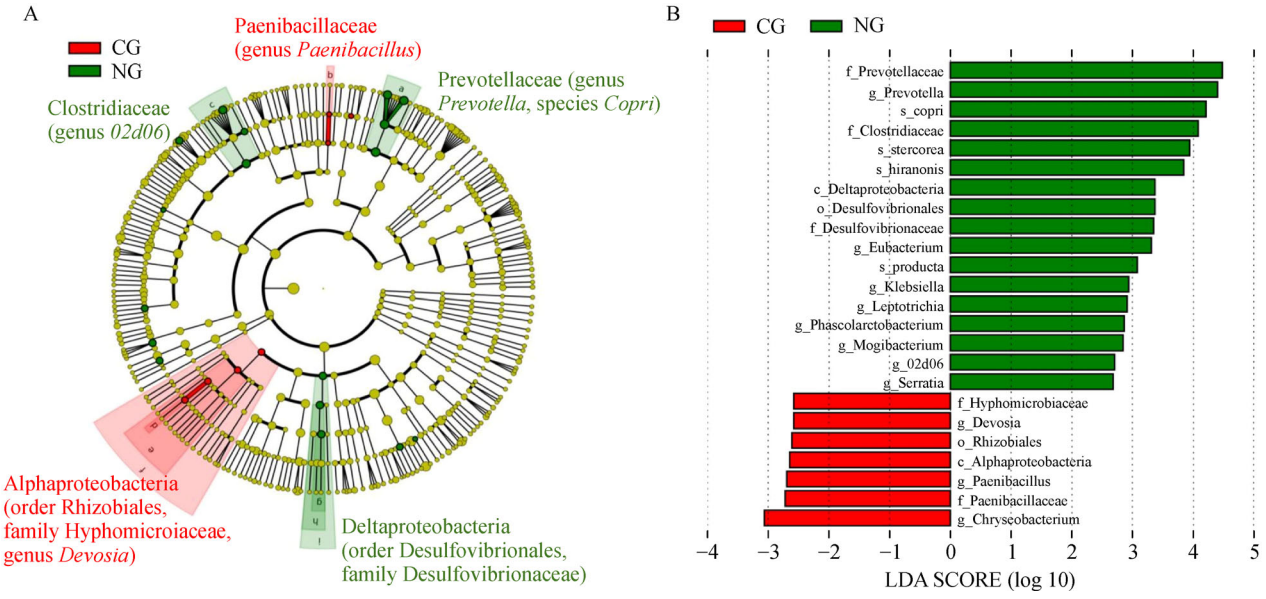

Fig. 3 Key contributors of the structural segregation of different groups identified using LEfSe. (A) A cladogram of the phylotypes that differed between the groups displayed according to effect size. Differences are represented by the color of the most abundant class (red = CRC group; green = normal group). Significant bacterial taxonomic groups are labeled, with the genus, family, species, or order in parentheses. (B) LDA scores of enriched bacterial taxa (LDA > 2 of LEfSe). Significantly enriched bacterial taxa from different groups are clustered on different sides (NG = right; CG = left) and labeled with different colors. NG = normal group; CG = CRC group; LDA = linear discriminate analysis.

(Table 3). Compared with normal group, the CRC group showed a remarkably larger abundance of KEGG pathways affiliated with cancers (small cell lung cancer, colorectal cancer, $P < 0.05$), infectious diseases (influenza A, toxoplasmosis, $P < 0.05$), cardiovascular diseases (viral myocarditis, $P < 0.05$), cell growth and death (p53 signaling pathway), transport and catabolism (endocytosis, $P < 0.05$), immune system (FcγR-mediated phagocytosis, $P < 0.05$), and endocrine system (GnRH signaling pathway, $P < 0.05$), as well as a significantly lower abundance of the KEGG pathway affiliated with *Vibrio cholerae* infection ($P < 0.05$).

Discussion

Many studies have shown that the gut microbiota and their metabolites are associated with CRC; however, changes in the populations of intestinal microbes, their effect on CRC, and the related mechanism(s) remain unclear. Although no microorganism that is pathogenic to any host from a microecological point of view has yet been uncovered, specific perturbations to the intestinal microbiome are indicative of some disease states, and specific gut bacteria are known to be involved in the pathogenesis of CRC. In this work, we compared microbial populations in mucosal-luminal interface samples from study participants with CRC versus normal controls. Previous studies report that the diversity of the gut microbiota in CRC patients is reduced compared with that in healthy controls, which may be caused by inhibition of the immune response in the pathological tissue. In the current study, PCoA results exhibited a slight difference in the overall microbiota structure between CG and NG (Supplementary Fig. 1) using the first two principal component scores, which indicates that changes in the gut microbiota may contribute to the pathogenesis of CRC. In addition, no distinct clustering of the points for the two groups based on visual inspection of the PCoA score plot was found, which is likely related to the low power of the study because of its small sample size.

The human microbiome contains over 1000 microbial species; the gastrointestinal tract itself harbors up to 100 trillion bacteria [17]. Although many types of bacteria are found in the gut, the amounts of each species vary widely. Over 99% of the gut microbiome is composed of 30–40 types of bacterial species [18]. These bacteria can be divided into three broad groups according to their different physiological functions in the gut: commensal bacteria, conditional pathogens, and pathogenic bacteria. Commensal bacteria occupy more than 99% of the gut microbiota, produce beneficial substances, and protect human health [19]. Compared with commensal bacteria, fewer conditional pathogens are found in the gut; however, allowing

Table 3 Selected main microbial pathways grouped into level-3 functional categories determined by PICRUSt

KEGG pathways	NG	CG	*P* value
Cancers			
Small cell lung cancer	0.10 (0.07–0.18)	0.30 (0.13–0.40)	**0.016**
Colorectal cancer	0.10 (0.07–0.18)	0.30 (0.13–0.40)	**0.017**
Bladder cancer	30.26 (11.85–69.91)	42.23 (29.89–109.37)	0.441
Renal cell carcinoma	55.01 (22.16–123.0)	69.56 (37.40–201.82)	0.443
Infectious diseases			
Vibrio cholerae infection	0.90 (0.54–3.02)	0.57 (0.37–0.79)	**0.005**
Influenza A	0.10 (0.07–0.18)	0.30 (0.13–0.40)	**0.016**
Toxoplasmosis	0.10 (0.07–0.18)	0.30 (0.13–0.40)	**0.016**
Pertussis	541.81 (190.97–942.52)	596.60 (481.92–1893.19)	0.256
African trypanosomiasis	37.04 (20.87–72.41)	43.24 (32.99–109.60)	0.400
Chagas disease (American trypanosomiasis)	31.72 (20.66–72.18)	43.15 (32.76–109.52)	0.403
Cardiovascular diseases			
Viral myocarditis	0.10 (0.07–0.18)	0.30 (0.13–0.40)	**0.016**
Cell growth and death			
p53 signaling pathway	0.10 (0.07–0.18)	0.30 (0.13–0.40)	**0.016**
Meiosis-yeast	5.88 (2.38–8.91)	1.77 (1.17–5.87)	0.088
Apoptosis	1.79 (1.56–2.06)	2.27 (0.82–6.38)	0.295
Transport and catabolism			
Endocytosis	0	0.04 (0–0.05)	**0.044**
Lysosome	621.65 (352.94–720.52)	428.15 (241.70–638.47)	**0.282**
Biosynthesis of other secondary metabolites			
Stilbenoid, diarylheptanoid, and gingerol biosynthesis	3.89 (1.75–5.86)	3.67 (1.58–6.24)	0.085
Betalain biosynthesis	0.06 (0.03–0.14)	0.05 (0.02–0.24)	0.199
Butirosin and neomycin biosynthesis	601.70 (441.17–668.97)	356.68 (328.17–538.20)	0.451
Flavonoid biosynthesis	26.09 (22.30–43.97)	24.90 (10.00–35.72)	0.485
Endocrine system			
GnRH signaling pathway	0	0.04 (0–0.05)	**0.043**
Melanogenesis	0.01 (0–0.05)	0.03 (0–0.12)	0.061
Proximal tubule bicarbonate reclamation	124.57 (83.98–164.06)	101.69 (95.73–231.88)	0.453
Immune system			
FcγR-mediated phagocytosis	0	0.04 (0–0.05)	**0.043**
RIG-I-like receptor signaling pathway	30.90 (13.48–42.25)	14.64 (7.41–26.66)	0.065

Data are shown as medians and interquartile ranges (IQR). Significant differences were defined as a P value < 0.05 (marked in bold).

these pathogens to reproduce under certain conditions can have deleterious effects on the body. Pathogenic bacteria, such as *Salmonella*, *Shigella*, *Proteus*, *Escherichia coli*, and *Jerson*, directly cause disease. *Salmonella* was the most abundant genus in CG and found in higher amounts in CG than in NG, thus indicating that intestinal flora disorder is potentially more serious in the former than in the latter. In both groups, the intestinal microbiome was composed mainly of Firmicutes, Bacteroidetes, and Proteobacteria, followed by far less abundant Fusobacteria and Actinobacteria. On average, participants in CG had higher amounts of Proteobacteria and Fusobacteria and lower amounts of Bacteroidetes and Firmicutes than those in NG. These results are consistent with previous studies of human samples and animal models of CRC [20,21]. Fusobacteria is a small group of Gram-negative bacteria commonly found in the digestive tract that can cause some diseases. Large numbers of *Fusobacterium* are associated with CRC, but its role in disease development is unclear [8,22,23]. Although the average number of *Fusobacterium* in the mucosal-luminal interface samples was higher in CG than that in NG, further studies with larger sample sizes are needed to confirm this trend. Previous studies have shown that the *C. leptum* and *C. coccoides* subgroups are specific to CRC [24], but neither species was detected in the present study. However, we found that *Clostridium perfringens* tended to decrease in CG, which is similar to a finding by Sasada *et al*. from mouse models [25].

Firmicutes includes a large group of bacteria, most of which are Gram positive. Firmicutes is highly enriched in the intestinal lumen and can enhance energy harvesting from the diet [26]; the phylum consists of several genera, including *Clostridium*, *Blautia, Coprococcus, Dorea,* and *Streptococcus*, and the butyrate producers *Eubacterium* and *Faecalibacterium*. Butyric acid is an important short-chain fatty acid that is an energy source for colonic epithelial cells and regulates gene expression, inhibits inflammation, and prevents tumorigenesis [27]. The *in vivo* supply of butyrate strongly depends on butyrate-producing bacteria, which mainly exist in the cecum and colon. Here, we identified a significant reduction in *Eubacterium* in the gut microbiota of CRC patients; this genus was further identified as a biomarker by LEfSe analysis. Balamurugan *et al*. reported that *Eubacterium* and *Faecalibacterium* decreased by approximately 4-fold in CRC patients compared with those in healthy control volunteers [28]. In the present study, *Faecalibacterium* comprised 10.5%–13.0% of the microbiota in the different patient groups with no significant differences between groups. A randomized clinical trial showed that butyrylated starch intake can prevent red meat-induced O6-methyl-2-deoxyguanosine adducts in human rectal tissue [29]. In another dietary intervention study, the abundance of *Eubacterium* showed a strong positive correlation with fecal butyrate concentrations in response to carbohydrate intake [30], thus revealing the importance of *Eubacterium* in the production of butyrate *in vivo*. Many animal experiments and clinical studies have shown that butyrate plays an important role in the repair of intestinal mucosa and the treatment of esoenteritis. According to a study by Sengupta *et al*., the delivery of adequate butyrate to the appropriate sites appears to protect against early tumorigenic events [31]. Therefore, our results, together with previous reports, suggest that *Eubacterium* may play an important role in protecting hosts from colorectal carcinogenesis. However, a study with greater power requires a larger sample size; further research is also needed to explore the underlying mechanisms.

We observed significant increases in the richness of *Devosia* in CG compared with that in NG to CG. *Devosia* is a type of mycotoxin deoxynivalenol-degrading bacterium that belongs to the family Hyphomicrobiaceae [32]. The proportion of *Devosia* bacteria increases as the degree of radiation in polluted soils raises, and *Devosia* is one of the most abundant genera in propylene oxide saponification wastewater treatment plants [32]. Consistent with these data, we observed high levels of *Devosia* in the CRC samples, as well as a negative correlation between enterobacterial and *Devosia* levels. These findings suggest that the genus could be a promising biomarker for tumorigenesis. Nevertheless, further investigation is required to fully understand and evaluate its role in the progress of CRC.

According to the predictive functional profiles of microbial communities determined by PICRUSt analysis, the most abundant functions were infectious diseases, cancers, transport and catabolism, biosynthesis of other secondary metabolites, and endocrine system (Table 3). Disorders of the KEGG pathways of colorectal cancer, small cell lung cancer, and the p53 signaling pathway may be related to the carcinogenic mechanism of CRC. Increases in the endocytosis pathway indicated increased communication between cells and CRC environments. Alterations in the KEGG pathways of influenza A, toxoplasmosis, viral myocarditis, the GnRH signaling pathway, and FcγR-mediated phagocytosis may explain the serious adverse health effects caused by CRC.

Because of the difficulties associated with obtaining samples from the intestinal mucosa, one limitation of the current study is the sample size of each group; this limitation inhibits the drawing of definitive conclusions related to the changes to the intestinal microbiota during the progression of CRC based solely on the results of this study. In addition, this research only focused on characterizing the observed microbiota; intestinal metabolites were not analyzed. We believe that further studies including the analysis of intestinal metabolites will provide a better understanding of the mechanisms leading to CRC.

In conclusion, in this case-control study, the abundance of gut microbiota at different levels were detected in CRC patients and healthy volunteers. The presence of exclusive microbial genera in each group indicates the existence of potential biomarkers for the disease. Our results reveal that the observed abundance of species belonging to *Eubacterium* and *Devosia* may act as a promising biomarker for the early detection of CRC. Nonetheless, more detailed information on these taxa and further exploration of intestinal lumen-associated microbiota remain essential. Further research on the relationship between different organisms and the etiology of CRC could lead to the monitoring of individual microbial strains for the early detection of intestinal cancer. Eventually, future work may establish new treatments based on the application of probiotic strains.

Acknowledgments

This current study was supported by the Shanghai Sixth People's Hospital Grant (No. YNLC201725), the National Natural Science Foundation of China (No. 81800708), Outstanding Academic Leaders of Shanghai Health System (No. 2017BR008) and Yangtze River Scholar.

Compliance with ethics guidelines

Hong Zhang, Ying Chang, Qingqing Zheng, Rong Zhang, Cheng Hu, and Weiping Jia have declared no conflict of interest. Written informed consent was obtained from all participants, and the study

protocol was approved by the Research Ethics Board of Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China.

Electronic Supplementary Material Supplementary material is available in the online version of this article at https://doi.org/10.1007/s11684-019-0695-7 and is accessible for authorized users.

References

1. Cunningham D, Atkin W, Lenz HJ, Lynch HT, Minsky B, Nordlinger B, Starling N. Colorectal cancer. Lancet 2010; 375 (9719): 1030–1047
2. Regula J, Rupinski M, Kraszewska E, Polkowski M, Pachlewski J, Orlowska J, Nowacki MP, Butruk E. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. N Engl J Med 2006; 355(18): 1863–1872
3. Azcárate-Peril MA, Sikes M, Bruno-Bárcena JM. The intestinal microbiota, gastrointestinal environment and colorectal cancer: a putative role for probiotics in prevention of colorectal cancer? Am J Physiol Gastrointest Liver Physiol 2011; 301(3): G401–G424
4. Horiuchi Y, Fujisaki J, Ishizuka N, Omae M, Ishiyama A, Yoshio T, Hirasawa T, Yamamoto Y, Nagahama M, Takahashi H, Tsuchida T. Study on clinical factors involved in *Helicobacter pylori*-uninfected, undifferentiated-type early gastric cancer. Digestion 2017; 96(4): 213–219
5. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Pettersson S. Host-gut microbiota metabolic interactions. Science 2012; 336(6086): 1262–1267
6. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human gut microbiome viewed across age and geography. Nature 2012; 486(7402): 222–227
7. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer 2006; 118(12): 3030–3044
8. Amitay EL, Werner S, Vital M, Pieper DH, Höfler D, Gierse IJ, Butt J, Balavarca Y, Cuk K, Brenner H. *Fusobacterium* and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. Carcinogenesis 2017; 38(8): 781–788
9. Yamaoka Y, Suehiro Y, Hashimoto S, Hoshida T, Fujimoto M, Watanabe M, Imanaga D, Sakai K, Matsumoto T, Nishioka M, Takami T, Suzuki N, Hazama S, Nagano H, Sakaida I, Yamasaki T. *Fusobacterium nucleatum* as a prognostic marker of colorectal cancer in a Japanese population. J Gastroenterol 2018; 53(4): 517–524
10. Repass J, Maherali N, Owen K, Reproducibility Project: Cancer Biology. Registered report: *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. eLife 2016; 5:e10012
11. Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, Moris F, Rodrigo L, Mira A, Collado MC. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol 2015; 50(2): 167–179
12. Mottawea W, Chiang CK, Mühlbauer M, Starr AE, Butcher J, Abujamel T, Deeke SA, Brandel A, Zhou H, Shokralla S, Hajibabaei M, Singleton R, Benchimol EI, Jobin C, Mack DR, Figeys D, Stintzi A. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. Nat Commun 2016; 7(1): 13419
13. 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 2013; 41(Database issue): D590–D596
14. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009; 75(23): 7537–7541
15. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. Genome Biol 2011; 12(6): R60
16. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 2013; 31(9): 814–821
17. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 2006; 124(4): 837–848
18. Kimura K, McCartney AL, McConnell MA, Tannock GW. Analysis of fecal populations of bifidobacteria and lactobacilli and investigation of the immunological responses of their human hosts to the predominant strains. Appl Environ Microbiol 1997; 63(9): 3394–3398
19. Sears CL. A dynamic partnership: celebrating our gut flora. Anaerobe 2005; 11(5): 247–251
20. Zhu Q, Jin Z, Wu W, Gao R, Guo B, Gao Z, Yang Y, Qin H. Analysis of the intestinal lumen microbiota in an animal model of colorectal cancer. PLoS One 2014; 9(6): e90849
21. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, Jia W, Cai S, Zhao L. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J 2012; 6(2): 320–329
22. Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, Jelinek J, Yamano HO, Sugai T, An B, Shureiqi I, Toyota M, Kondo Y, Estécio MR, Issa JP. *Fusobacterium* in colonic flora and molecular features of colorectal carcinoma. Cancer Res 2014; 74(5): 1311–1318
23. McCoy AN, Araújo-Pérez F, Azcárate-Peril A, Yeh JJ, Sandler RS, Keku TO. *Fusobacterium* is associated with colorectal adenomas. PLoS One 2013; 8(1): e53653
24. Scanlan PD, Shanahan F, Clune Y, Collins JK, O'Sullivan GC, O'Riordan M, Holmes E, Wang Y, Marchesi JR. Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. Environ Microbiol 2008; 10(3): 789–798
25. Sasada T, Hinoi T, Saito Y, Adachi T, Takakura Y, Kawaguchi Y, Sotomaru Y, Sentani K, Oue N, Yasui W, Ohdan H. Chlorinated water modulates the development of colorectal tumors with chromosomal instability and gut microbiota in APC-deficient mice. PLoS One 2015; 10(7): e0132435
26. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased

capacity for energy harvest. Nature 2006; 444(7122): 1027–1031

27. Scharlau D, Borowicki A, Habermann N, Hofmann T, Klenow S, Miene C, Munjal U, Stein K, Glei M. Mechanisms of primary cancer prevention by butyrate and other products formed during gut flora-mediated fermentation of dietary fibre. Mutat Res 2009; 682(1): 39–53
28. Balamurugan R, Rajendiran E, George S, Samuel GV, Ramakrishna BS. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, *Desulfovibrio* and *Enterococcus faecalis* in the feces of patients with colorectal cancer. J Gastroenterol Hepatol 2008; 23(8 Pt 1): 1298–1303
29. Le Leu RK, Winter JM, Christophersen CT, Young GP, Humphreys KJ, Hu Y, Gratz SW, Miller RB, Topping DL, Bird AR, Conlon MA. Butyrylated starch intake can prevent red meat-induced O6-methyl-2-deoxyguanosine adducts in human rectal tissue: a randomised clinical trial. Br J Nutr 2015; 114(2): 220–230
30. Duncan SH, Belenguer A, Holtrop G, Johnstone AM, Flint HJ, Lobley GE. Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. Appl Environ Microbiol 2007; 73(4): 1073–1078
31. Sengupta S, Muir JG, Gibson PR. Does butyrate protect from colorectal cancer? J Gastroenterol Hepatol 2006; 21(1 Pt 2): 209–218
32. Sato I, Ito M, Ishizaka M, Ikunaga Y, Sato Y, Yoshida S, Koitabashi M, Tsushima S. Thirteen novel deoxynivalenol-degrading bacteria are classified within two genera with distinct degradation mechanisms. FEMS Microbiol Lett 2012; 327(2): 110–117